

# Nonlinear relationships between atmospheric aerosol and its gaseous precursors: Analysis of long-term air quality monitoring data by means of neural networks

I. B. Konovalov

Institute of Applied Physics, Russian Academy of Sciences, Nizhny Novgorod, Russia

Received: 6 December 2002 – Published in Atmos. Chem. Phys. Discuss.: 14 February 2003

Revised: 30 May 2003 – Accepted: 30 May 2003 – Published: 5 June 2003

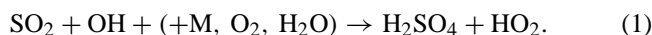
**Abstract.** The nonlinear features of the relationships between concentrations of aerosol and volatile organic compounds (VOC) and nitrogen oxides ( $\text{NO}_x$ ) in urban environments are revealed directly from data of long-term routine measurements of  $\text{NO}_x$ , VOC, and total suspended particulate matter (PM). The main idea of the method is development of special empirical models based on artificial neural networks. These models, that are basically, the nonlinear extension of the commonly used linear statistical models provide the best fit for the real (nonlinear) PM- $\text{NO}_x$ -VOC relationships under different atmospheric conditions. Such models may be useful in the context of various scientific and practical problems related to atmospheric aerosols. The method is demonstrated on an example of two empirical models based on independent data-sets collected at two air quality monitoring stations at South Coast Air Basin, California. It is shown that in spite of a rather large distance between the monitoring stations (more than 50 km) and thus substantially different environmental conditions, the empirical models demonstrate several common qualitative features. Specifically, under definite conditions, a decrease in the level of  $\text{NO}_x$  or VOC may lead to an increase in mass concentration of aerosol. It is argued that these features are due to the nonlinear dependence of hydroxyl radical on VOC and  $\text{NO}_x$ .

## 1 Introduction

Atmospheric aerosol is an important environmental factor that has an impact on human health, affects visibility, and plays a significant role in atmospheric chemistry and climate. The mass concentration of aerosols about the surface is regulated by national air quality standards, implementation of which requires reliable scientific assessment of possi-

ble strategies for its reduction. Such an assessment, however, represents an extremely difficult problem because the evolution of aerosol is driven by complex physical and chemical processes which, in turn, are influenced by a number of environmental factors.

In particular, major components of typical aerosols in the boundary layer such as sulfates, nitrates, and organic compounds may reach a particulate phase as a result of photochemical processes involving certain gaseous species (aerosol precursors) such as  $\text{NO}_x$ ,  $\text{SO}_2$ ,  $\text{NH}_3$ , and volatile organic compounds (VOC) (e.g. Senfeld and Pandis, 1998). Specifically, the important potential source of the airborne particulate matter (PM) is formation of sulfuric acid by a gas-phase chain mechanism which can be represented by the combined reaction (Stockwell et al., 1990)



Another potential source of atmospheric PM is the reversible reaction of ammonia with nitric acid



where (g) and (p) denote gas and particulate phases, respectively.  $\text{HNO}_3$ , in turn, is formed in the gas-phase reaction



Condensable compounds may also be formed in reactions of certain VOCs with OH,  $\text{O}_3$ , or  $\text{NO}_3$  (Odum et al., 1996; 1997).

It is obvious from the aforesaid that formation of PM is driven to a significant extent by the hydroxyl radical (OH), the behavior of which is known to be determined by a large number of chemical processes (see, e.g. Stockwell et al., 1997). It is clear also that due to the crucial role of OH in processes of ozone formation, the gas-to-particle conversion processes are closely linked with complex photochemistry of ozone (e.g. Meng et al. 1997).

Correspondence to: I. B. Konovalov  
(konov@appl.sci-nnov.ru)

During the two last decades much effort has been given to modeling evolution of aerosols from the first principles; that is, attempts have been made to describe explicitly all important processes affecting formation and removal of airborne particles (see, e.g. Pilinis and Seinfeld, 1988; Wexler et al., 1994; Binkowski and Shankar, 1995; Jacobson, 1997; Meng et al., 1998; Ackermann et al., 1998; Sun and Wexler, 1998; Nenes et al., 1999). Nevertheless, in spite of the great progress in understanding and representation of aerosol evolution, the comparisons of model results with measurements bear evidence that models still contain significant uncertainties. Moreover, it still remains unknown to which extent the chemical-transport (CT) models are accurate in predictions of responses of aerosol to changes of emissions of potential PM precursors.

Taking into account close chemical coupling between aerosol and ozone, it is useful to note that evaluation of adequacy of CT model predictions of effectiveness of various emission reduction scenarios represents a rather difficult problem in researches of surface ozone as well (e.g. Sillman et al., 1999). In ozone studies, that problem arises, principally, due to the essential non-linearity of dependence of ozone production on concentrations of its precursors (National Research Council (NRC), 1991). Looking for solution of this problem, several methods for assessing ozone sensitivity to changes of its precursors have been developed recently (see, e.g. Kleinman et al., 2000) which are based predominantly on the use of data of ambient measurements. On the other hand, similar methods for aerosol have not been proposed yet. Moreover, although some nonlinear features of PM-NO<sub>x</sub>-VOC relationships have been discussed (e.g. Meng et al., 1997), it still remains unclear to which extent these features are general.

Thus, at the present, there exists a definite need in better understanding of relationships between aerosol and its gaseous precursors, as well as the need in developing additional methods for their identification in the real environment. This paper presents results of the first attempt to study nonlinear features of PM-NO<sub>x</sub>-VOC relationships directly on the basis of data of ambient measurements. The essence of the method which is used for the analysis is creation of empirical nonlinear statistical models based on artificial neural networks. Although neural networks have been widely used in atmospheric researches, up to our knowledge, the first successful attempt to use them with the purpose to derive interpretable information on relationships between observed atmospheric characteristics was undertaken only very recently in our study devoted to the analysis of nonlinear relationships between surface ozone and its precursors (Konovalov, 2002). Note that there is a common belief (see, e.g., the overview by Gardner and Dorling, 1998) that the "black box" nature of the neural networks seriously limits their utility, where the main goal of their uses is to increase the knowledge of a physical system. Nonetheless, as it was shown in our previous work (Konovalov, 2002), it is still possible to obtain useful

knowledge about physical or chemical properties of a natural system if the neural networks are trained and analyzed properly.

In this study, neural networks are used for analysis of data of routine long-term (1980–1995) measurements of total suspended particulate matter (TSP) in two air quality monitoring stations in California. The method of the analysis employed here is very similar to that of our previous study (Konovalov, 2002). Although TSP is sometimes used as an indicator of harmful effects of air pollution (e.g., Delucchi et al., 2002), the health and visibility are affected, primarily, by fine fractions of aerosol, and hence the analysis of PM<sub>2.5</sub> or PM<sub>10</sub> data could be of most interest from the practical point of view. Unfortunately, at the present time available records of PM<sub>2.5</sub> and PM<sub>10</sub> measurements accompanied with simultaneous measurements of NO<sub>x</sub> and VOC are not long enough for building neural network models aimed at retrieving PM-NO<sub>x</sub>-VOC relationships. Nevertheless, as TSP encompasses all aerosol components which may exhibit nonlinear dependence on concentrations of NO<sub>x</sub> and VOC, the analysis of TSP data may also provide deeper insight into the processes of aerosol formation. Moreover, we expect that nonlinear features of PM-NO<sub>x</sub>-VOC relationships discussed in this paper are fairly robust and applicable to all size fractions of aerosol. Note that these features are most likely associated with the particulate matter of the secondary origin, contribution of which to TSP at the chosen sites is discussed below in Sect. 2.2. As for the substances of primary origin contributing to TSP (e.g., fugitive dust or cement), they may hardly demonstrate any regular dependence on concentrations of gaseous aerosol precursors except trivial linear correlation which is not in the focus of this study. It is hoped also that the method proposed in this paper may be successfully applied to measurements of the fine fractions of aerosol in the near future.

## 2 Data

### 2.1 General description

This study utilizes data-sets of measurements of TSP, O<sub>3</sub>, NO<sub>x</sub>, and total volatile organic compounds (this quantity is referred to below as VOC) made at two air quality monitoring stations at South Coast Air Basin, California during 1980–1995 (California Air Quality Data, 1998). The stations were selected a priori based on the announced spatial scale of observations (only stations with the urban scale (4–50 km) of ozone and NO<sub>x</sub> observations were considered) and availability of synchronous measurements of all components noted above for all years from 1980 to 1995. The urban scale means that both distribution of sources of pollutants and topographical situation in the neighborhoods of the station are such that the appearance of strong spatial gradients of the corresponding pollutants is unlikely. The only

stations satisfying these criteria turned out to be Riverside-Rubidoux (33°58'55" N, 117°22'21" W) located at Riverside County and Azusa (34°08'10" N, 117°55'26" W) located at a suburb of Los Angeles. Note that the data series used in this analysis ended in 1995, as the measurements of VOC at South Coast Air Basin were cancelled that year. The measurements of TSP (they present daily values of mass concentration) were conducted rather irregularly. The most frequent interval of subsequent measurements was 6 days; however, there is a number of much more extended gaps in the data series. The measurements of gaseous species were much more regular and were conducted each hour as a rule. However, corresponding data series also contain a number of gaps. The hourly data for gaseous species were averaged from 10 a.m. to 18 p.m. each day of measurements. Such averaging was performed in order to get a unique value of each component per day and to suppress effects of possible random variations of species concentrations within the course of the day. Note also that since the aim of this study is to analyze manifestations of photochemical processes which are most active during the light period of a day, it is natural to consider the dependence of PM on the day-time rather than night-time or daily mean concentrations of gaseous pollutants. No doubt, it would be advantageous to use TSP data for the same period of a day as NO<sub>x</sub> and VOC data, but they are not available. Nevertheless, it is reasonable to expect that neural networks themselves are capable of retrieving part of variations of daily mean TSP caused by variations of day-time concentrations of NO<sub>x</sub> and VOC.

The air quality data were matched by daily mean meteorological data from the Climate Prediction Center, National Ocean and Atmospheric Administration (<http://www.cpc.ncep.noaa.gov/>). Again, it is reasonable to assume that the use of the meteorological data averaged over the same period of day as gaseous species data would provide a better quality of the empirical models. Unfortunately, free hourly meteorological data were not found.

Finally, the data-sets analyzed in this study include 1880 and 1032 days of measurements at Azusa and Riverside-Rubidoux stations, respectively. The data have not been pre-selected in any way.

## 2.2 Preliminary analysis

Inasmuch as this study is focused on the analysis of the relationships between secondary aerosol and its gaseous precursors, it is useful to ascertain that the aerosol mass of secondary origin contributes substantially to the measure of aerosol concentration, TSP, which is used in this study. While it is hardly possible to quantify primary and secondary aerosol fractions precisely without detailed modelling of sources and transport of aerosols, it seems still possible to roughly estimate the lowest bound for the contribution of the secondary aerosol. The idea underlying such estimation is rather simple: it is reasonable to expect that the rate of forma-

tion of secondary PM is highest during summer photochemical smog events, and, consequently, the difference between summer and winter mean aerosol concentration can reflect, in particular, the secondary aerosol formed during hot and sunny summer days. However, in order to identify the contribution of the secondary aerosol to seasonal variations of TSP, it is necessary to analyze all other factors which could have a strong impact on these variations. To do that, let us consider the averaged concentrations of aerosol, VOC, and NO<sub>x</sub> for winter and summer conditions separately. The corresponding data are presented in Table 1. Note, first of all, that the mean aerosol concentration (see the lines "total") is considerably higher in summer than in winter at both sites. In principle, this difference may be due to either higher emission rates of primary PM in summer, or seasonal changes in atmospheric circulation and precipitation, or higher concentrations of secondary PM in summer months. Let us assess now plausibility of all these potential reasons.

In accordance to data of emission inventories at South Coast Air Basin (see, e.g., <http://www.arb.ca.gov/app/emsinv/fcemssumcat.html>), the total emissions of particulate matter are indeed somewhat larger in summer than in winter. However, the seasonal difference in the emissions is much less than the seasonal difference in TSP. To be specific, in accordance to data of emission inventory for 1990, the relative differences between total emissions of PM in summer and winter are of 14.8% at Los Angeles County and 13.4% at Riverside County, while the seasonal differences in TSP are of 81.1% and 70.4%, respectively. To eliminate the impact of seasonal changes in total emissions on the seasonally averaged TSP values, we have multiplied the winter TSP values by the ratio of the PM emissions in summer to that in winter. Similar scaling has been performed for VOC and NO<sub>x</sub>. The results are given in Table 1. Note that the relative seasonal differences in PM emissions are very similar for all other years which are represented in the emission database cited above and are relevant for this study (1980, 1985 and 1995). Thus, in view of the available emission inventory data, the first explanation for the observed seasonal variation in TSP seems insufficient.

Substantial seasonal changes of atmospheric circulation in California (e.g., changes of the average wind direction and speed, the height of the mixing layer, the turbulent diffusion rates) are undeniable and could be responsible for the observed changes in aerosol concentration. In particular, the changes in wind direction could lead to significant changes of the observed aerosol concentrations, if sources of primary PM were distributed strongly asymmetrically around the given site. However, the data presented in Table 1 bear evidence that it is not a case for the considered sites. Indeed, the summer TSP dominates strongly over the winter TSP irrespectively of the wind direction. Furthermore, while in accordance to the emission inventories the major emission sources of PM, VOC, and NO<sub>x</sub> are associated with transport, and thus they are basically common, it seems reasonable to

**Table 1.** Some seasonally averaged characteristics used in the analysis of seasonal variations of TSP

|                    | winter |                                |                    |                     |                          | summer |                                |              |                     |                          |
|--------------------|--------|--------------------------------|--------------------|---------------------|--------------------------|--------|--------------------------------|--------------|---------------------|--------------------------|
|                    | $N_d$  | $TSP (\mu\text{g}/\text{m}^3)$ | $VOC(\text{ppmC})$ | $\text{NO}_x$ (ppb) | $T$ ( $^\circ\text{C}$ ) | $N_d$  | $TSP (\mu\text{g}/\text{m}^3)$ | $VOC$ (ppmC) | $\text{NO}_x$ (ppb) | $T$ ( $^\circ\text{C}$ ) |
| Azusa              |        |                                |                    |                     |                          |        |                                |              |                     |                          |
| $U > 0$            | 69     | 78.2                           | 2.31               | 64.3                | 12.1                     | 756    | 140.0                          | 2.47         | 55.3                | 27.2                     |
| $U < 0$            | 173    | 76.0                           | 2.3                | 71.2                | 11.6                     | 11     | 147.9                          | 2.25         | 53.8                | 26.7                     |
| $V > 0$            | 67     | 51.1                           | 2.16               | 59.0                | 12.8                     | 456    | 139.0                          | 2.44         | 53.6                | 27.3                     |
| $V < 0$            | 175    | 87.3                           | 2.36               | 73.3                | 11.4                     | 292    | 141.7                          | 2.52         | 57.8                | 27.0                     |
| <i>Total</i>       | 245    | 77.3                           | 2.31               | 69.2                | 11.8                     | 768    | 140.2                          | 2.47         | 55.3                | 27.2                     |
| (a)                | 245    | 88.7                           | 2.22               | 62.3                | 11.8                     | 768    | 140.2                          | 2.47         | 55.3                | 27.2                     |
| (b)                | 95     | 96.0                           | 2.31               | 79.6                | 12.2                     | 674    | 143.2                          | 2.48         | 56.1                | 27.7                     |
| (c)                | 95     | 110.2                          | 2.40               | 88.3                | 12.2                     | 674    | 143.2                          | 2.48         | 56.1                | 27.7                     |
| Riverside-Rubidoux |        |                                |                    |                     |                          |        |                                |              |                     |                          |
| $U > 0$            | 62     | 91.5                           | 2.20               | 40.3                | 11.9                     | 289    | 169.0                          | 2.19         | 24.8                | 27.0                     |
| $U < 0$            | 147    | 100.1                          | 2.26               | 48.1                | 11.8                     | 9      | 142.0                          | 2.07         | 23.5                | 25.9                     |
| $V > 0$            | 61     | 59.9                           | 2.21               | 42.9                | 12.6                     | 176    | 159.3                          | 2.19         | 23.8                | 27.0                     |
| $V < 0$            | 149    | 114.3                          | 2.25               | 47.2                | 11.6                     | 115    | 181.4                          | 2.17         | 25.8                | 26.9                     |
| <i>Total</i>       | 212    | 98.7                           | 2.24               | 46.2                | 11.9                     | 298    | 168.2                          | 2.19         | 24.7                | 27.0                     |
| (a)                | 212    | 111.9                          | 2.17               | 42.4                | 11.9                     | 298    | 168.2                          | 2.19         | 24.7                | 27.0                     |
| (b)                | 66     | 123.4                          | 2.29               | 48.6                | 12.5                     | 252    | 171.8                          | 2.20         | 24.6                | 27.5                     |
| (c)                | 66     | 139.9                          | 2.35               | 52.9                | 12.5                     | 252    | 171.8                          | 2.20         | 24.6                | 27.5                     |

(a): scaled for seasonal changes in emissions

(b): obtained after exclusion of days affected by precipitation

(c): both scaled for seasonal changes in emissions and corrected for precipitation

$N_d$ , the number of days

expect that the changes in atmospheric circulation should be reflected not only in observations of TSP but in observations of VOC and  $\text{NO}_x$  as well. Meanwhile, the mean summertime concentrations of VOC are only 6% higher at Azusa, and even 2.3% lower at Riverside than the mean wintertime concentrations. The fact that the seasonal changes in VOC and  $\text{NO}_x$  concentrations are so small when compared with the changes in TSP can hardly be accounted for by somewhat larger emissions of VOC at winter: to be specific, the differences between emissions of VOC in summer and in winter are only about 3.9% at Los Angeles County and 3.0% at Riverside County (see also Table 1). As for the mean concentrations of  $\text{NO}_x$ , they are considerably higher in winter than in summer (25% at Azusa and 87% at Riverside). Such clearly defined “inverse” seasonal profile of  $\text{NO}_x$  (as compared to TSP) also cannot be accounted for by the 11% and 9.0% increase of  $\text{NO}_x$  emissions at Los Angeles and Riverside Counties, respectively, in winter and, most probably, is due to large photochemical losses of  $\text{NO}_x$  in summer.

It is conceivable also that the lower level of TSP in winter may be due to washing out aerosols from the atmosphere by rains which are much more frequent in winter than in summer. The effect of precipitation on aerosol depends on life times of aerosol particles in the dry atmosphere, and it may, in principle, last for several days. However, in urban areas where the emission rates of primary PM are high, the

losses of aerosols during rains may be restored rather rapidly. Indeed, successive exclusions from the database of all days with precipitation rates greater than zero, then days following these days, and so on shows that even an impact of “yesterday” rains on “today” aerosol concentrations is quite small. Table 1 presents mean winter and summer concentrations obtained after exclusion of both “today” and “yesterday” precipitation events. It is seen that although such correction has led to rather significant increase of the PM concentrations in winter at both sites, they are still considerably lower than the respective values in summer.

The TSP values corrected for precipitation can further be scaled for seasonal changes of emissions (see the lines “(c)” in Table 1). However, it is reasonable to expect that if the rainy days were not taken into account in the emission inventories, the PM emissions in winter would be considerably higher than reported. The reason is that the largest (up to 70%) fraction of freshly emitted PM is identified in emission inventories as road dust which is, obviously, raised from dry roads much more intensively than from wet ones. Consequently, in order to estimate the fraction of secondary aerosols formed during summer months, the mean aerosol concentrations corrected for precipitation do not need, in our opinion, to be scaled further for seasonal changes in emissions. Thus, in accordance to our estimates, about 30% of the summer-mean TSP at Azusa and Riverside is of secondary

origin. Although these estimates are rather rude and say nothing about secondary aerosol fraction in winter, they indicate that measured TSP incorporates rather substantial fraction of secondary aerosol, and thus the following analysis of relationships between TSP and aerosol precursor gases makes sense indeed.

### 3 Methodology

The basic assumption in creation of the empirical models for aerosol is that mass concentration of TSP (this quantity is referred to below simply as PM) can be approximately represented as a sum of a nonlinear function ( $f$ ) of  $n$  other measured values  $x_i$ ,  $i = 1, \dots, n$  and some random value ( $\delta$ ), variations of which cannot be accounted for by variations of the arguments of  $f$ :

$$\text{PM} = f(x_1, \dots, x_n) + \delta. \quad (4)$$

The proper choice of the arguments  $x_i$  is the most critical point in providing usefulness and reliability of the empirical models. Obviously, the arguments should include, in particular, such values the dependence on which is to be studied. In this study, such values are gaseous precursors of PM, that is,  $\text{NO}_x$  and VOC. It would be of interest to consider other important gaseous precursors such as  $\text{NH}_3$  and  $\text{SO}_2$ , but adequate contemporary measurements of these species are not available. It is prudent also to take into account other important factors which, presumably, can significantly influence the studied relationships. However, the number of such factors should not be too large. Indeed, it is obvious that a larger number of arguments of a function, more data are needed to approximate it to the same accuracy. Taking into account the aforesaid, we have included in the PM empirical models only four meteorological values: temperature, specific humidity, and two horizontal components of wind speed. The empirical models with an additional argument, for example, total cloud cover that was included in our ozone empirical models (Konovalov, 2002) demonstrated higher diversity in the properties discussed in Sect. 4 than the models without this argument, which is indicative of the deficit of data. Thus, we believe that the models with seven arguments are less adequate and thus such models will not be considered here. However, it was found that empirical models with seven arguments retain the most important features of PM- $\text{NO}_x$ -VOC relationships retrieved from the models with six arguments.

It is assumed further that  $f$  can be approximated by means of a neural network of the perceptron type with one hidden layer (see, e.g. Bishop, 1995). Hornik et al. (1989) have proved that the perceptron with just one hidden layer (given a large enough number of neurons) is capable of approximating any measurable function to any desired degree of accuracy. Mathematically, the perceptron  $p(\mathbf{X})$  with one hidden

layer can be represented as follows:

$$p(\mathbf{X}) = \sum_j^{N_N} (w_j g_j(\mathbf{X}) + w_0),$$

$$g_j(\mathbf{X}) = \frac{1}{1 + \exp \left[ - \sum_{i=1}^n (\overline{w}_{ij} x_i + \overline{w}_{0j}) \right]}, \quad (5)$$

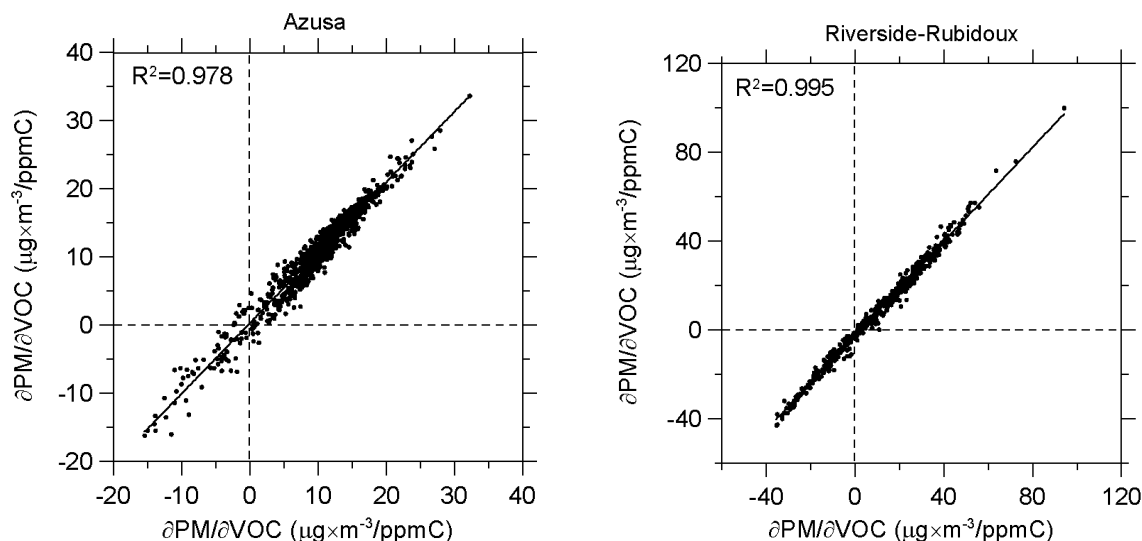
where  $\mathbf{X} = (x_1, \dots, x_n)$  is the vector of input values (that is, the vector of arguments of the function  $f$ ),  $w_j$ ,  $w_0$ ,  $\overline{w}_{ij}$ ,  $\overline{w}_{0j}$  are weight coefficients,  $N_N$  is the number of neurons in the network.

A procedure by which appropriate values of weights are determined is one of the key elements of the methodology of this study. Such a procedure is usually referred to as a training algorithm. Here a standard backpropagation algorithm (e.g. Reed and Marks, 1999) is used in combination with the early stopping method (Nelson and Illingworth, 1991; Sarle, 1995). The early stopping is usually used to provide such an important property of a neural network as generalization. For most applications generalization means that outputs of the network can approximate target values with given inputs that are not in the training set (Sarle, 2003). In our study, generalization means that a neural net learns the regular relationships ( $f$ ) that are valid for any arbitrarily chosen subset of the considered data-set and disregards noisy variations ( $\delta$ ). Implementation of early stopping in this study includes the following steps. First, the sequence of days in the data-sets was randomly mixed. Thereafter each of the data-sets was divided into two equal subsets: training and validation ones. The net is trained with the training subset but the error of prediction that is defined as

$$\varepsilon = \left[ \frac{2}{N_d} \sum_{i=1}^{N_d/2} (\text{PM}^{\text{ob}} - \text{PM}^{\text{n}})^2 \right]^{1/2}, \quad (6)$$

where  $\text{PM}^{\text{ob}}$  and  $\text{PM}^{\text{n}} = p(\mathbf{X})$  are observed and predicted PM levels, and  $N_d$  is the number of training cases (that is, the total number of days presented in the data-set), is controlled by the validation subset. The iterations are stopped as  $\varepsilon$  reaches minimum.

Direct application of this technique for creation of empirical models in the form of the single perceptron (5) meets two main problems. First, it cannot be neglected a priori that perceptrons with different numbers of neurons ( $N_N$ ) may give qualitatively different results. Second, since the training of neural networks is an iteration process, the results may depend on initial assumption regarding values of weights. To resolve these problems, we considered first, to what degree changes of  $N_N$  may affect the quality of approximations and important features of the relationships analyzed. It was found, in particular, that the dependence of  $\varepsilon$  (see Eq. 6) on  $N_N$ , averaged over a large number of single perceptrons with different random initial weights, is saturating such that systematic differences between perceptrons with different  $N_N$



**Fig. 1.** Evaluations of first derivatives of PM with respect to VOC made with the empirical models formed by neural networks with numbers of neurons ( $N_N$ ) ranging from 21 to 25 (axes of abscissas), versus the similar evaluations made with the empirical models with  $N_N$  ranging from 26 to 30 (axes of ordinates). Each point corresponds to one day in the training sub-set of the data-set. Solid lines represent best linear fits of the relationships.

greater than 20, at least, are quite negligible. The similar result was obtained for dependences of derivatives of PM with respect to  $\text{NO}_x$  and VOC, evaluated from Eq. (5), on  $N_N$ . Values and, especially, signs of these derivatives are the most important characteristics in the context of this study. Based on these results, it was concluded that the dependence of the approximations on  $N_N$  is not of any significance if  $N_N$  is greater than 20, at least. In order to minimize any hypothetical uncertainties associated with the choice of definite value of  $N_N$  to maximum degree, it was decided to include in the models perceptrons with different  $N_N$  ranging from 21 to 30.

To get rid of uncertainty associated with initial values of weights, a large enough set (50) of perceptrons with random initial weights were trained for each value of  $N_N$  and a few (5) “best” (having minimum  $\varepsilon$ ) networks selected from that set were included in the empirical model. Although a value of the cost function,  $\varepsilon$ , is not a direct indicator of the quality of the approximation, it is reasonable to expect that the networks having less  $\varepsilon$  provide, on the average, better approximations, and vice versa. For example, it is obvious that a network whose weight coefficients proved to be “trapped” by a local bad minimum of the cost function during the training, may remain seriously “undertrained” and exhibit spurious relationships between variables. The selection of a few networks having minimum  $\varepsilon$  out of the whole set of trained networks is performed in order to insure that the empirical models involve only the well trained networks. It was checked that the results presented below are not sensitive to the number of the “best” networks selected, as far as it remains much

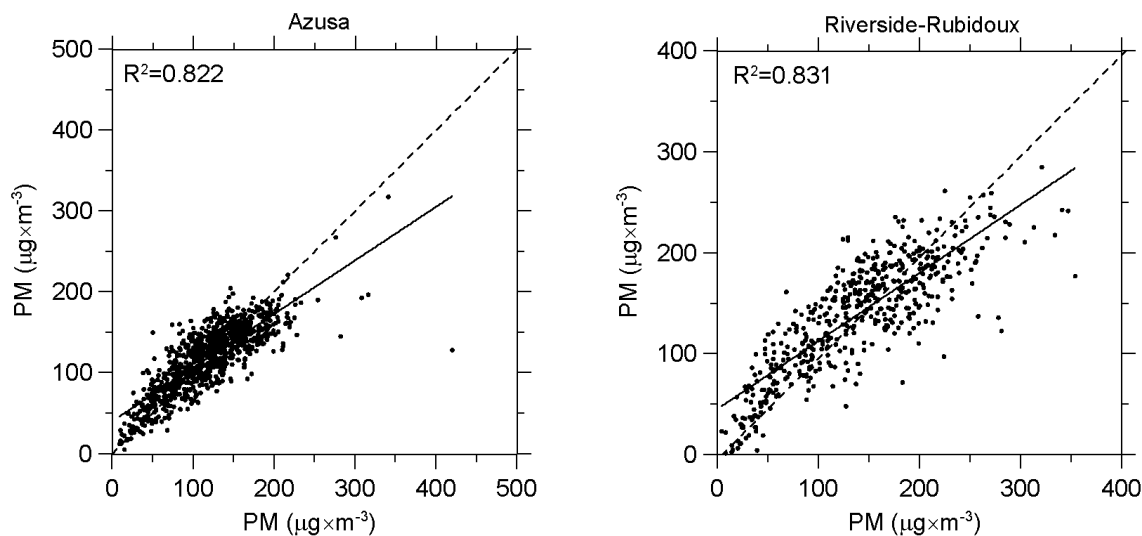
lesser than the total number of trained networks. Finally, the desired empirical model is represented as the mean of outputs of the set of 50 networks, which includes five “best” networks for each  $N_N$  from 21 to 30:

$$\text{PM} = \frac{1}{50} \sum_{N_N=21}^{30} \sum_{i=i_{b1}}^{i_{b5}} p(X, N_N, i), \quad (7)$$

where  $i_b$  are the indexes of the best nets.

Figure 1 illustrates that the uncertainties in results of the empirical models, associated with individual features of each of the networks involved in the models (7) are quite small indeed. This figure presents comparisons of evaluations of the first derivatives of PM with respect to VOC obtained from pairs of empirical models created in general accordance to Eq. (7) but with  $N_N$  ranging from 21 to 25 and from 26 to 30. Each point in the figure represents one day in the training sub-set of the data-sets. It is seen that each pair of independently created versions of the empirical models gives very similar results. In particular, fractions of days for which estimated signs of  $\partial\text{PM}/\partial\text{VOC}$  values are different in different versions are only 2.1% and 3.1% for the Azusa and Riverside models, respectively. It is reasonable to expect that the uncertainty of the full empirical models (7), associated with the discussed factors, should be even smaller than that of their component parts.

The empirical model performances in representation of variability of PM are demonstrated in Fig. 2 which shows measured values of PM (from the training sub-sets) versus those evaluated by the models. It is quite natural that the



**Fig. 2.** The empirical model evaluations of the PM concentrations (axes of ordinates) versus actually measured PM concentrations (axes of abscissas). Each point corresponds to one day in the training sub-set of the data-set. Solid lines represent best linear fits of the relationships.

models do not reproduce the measured values exactly since they are intended to reflect only part of total variability associated with changes of the factors represented by the model arguments. It is seen, however, that this part is quite substantial. This fact bears evidence that the model arguments, indeed, include values which strongly influence the PM level. The results for the validation subsets are only slightly worse.

The main idea of the analysis of the empirical models consists in the following. It is assumed that empirical models provide fairly precise fits of actual PM-NO<sub>x</sub>-VOC relationships under various conditions determined by values of the arguments. Qualitative features of these relationships under the given conditions may be learned via evaluation of various derivatives of PM with respect to VOC and NO<sub>x</sub>. For example, the positive first derivative PM with respect to VOC means that the increase in the level of VOC, most probably, would cause the increase in the level of PM under the given conditions; that is, the PM level demonstrates positive sensitivity to changes in the VOC level. It should be noted, however, that positive first derivatives of PM with respect to VOC or NO<sub>x</sub> may reflect simply the natural correlation between the levels of *primary* PM and gaseous pollutants as they mainly originate from the same emission sources. The positive second derivative of PM with respect to VOC under the same conditions means that the sensitivity would increase with the increase of VOC.

The standard practice of the neural network applications includes dividing the whole set of data into the training, validation and test subsets. However, as it has been noted above, we specify here only the training and validation subsets. Usually, the test subset is used to assess the quality of neural networks in terms of their prediction or classification capabilities. In such cases, definite measures of the network

performance are used, which assume comparison of neural network outputs with real observations. In our case, however, a similar measure of performance cannot be defined, as the real PM-NO<sub>x</sub>-VOC relationships are not known. Accordingly, introducing the test subset in this study would be meaningless. The derivatives of PM with respect to NO<sub>x</sub> or VOC, which are of interest in this study, may be evaluated, in principle, under any values of arguments in the range of the observed conditions. However, special investigation has shown that evaluations of individual nets forming the empirical models for the conditions that are not in the training subset, including those that are presented in the validation sub-sets, are much more dispersed. This is, presumably, due to the fact that each individual neural network included in the model best learned the conditions that are in the training subset but has more freedom under other conditions. Such freedom may result in larger uncertainty of evaluations of both the output value itself and its derivatives. Accordingly, it was decided to analyze only the evaluations made for the days presented in the training sub-sets. While the strict mathematical justification of this rather heuristic decision is beyond the scope of this paper, we would like to note that the serious evidence in favor of its legitimacy is provided by the fact that, as it is shown in the following sections, our two models created with different observational data exhibit many common features of PM-NO<sub>x</sub>-VOC relationships. If we had only one empirical model, it would not be possible to conclude then, whether the observed features are the indication of some real regularity, or it is just a result of poor quality of the model. However, when the same result is found with another model created using an independent data-set collected in a different environment, we have much more grounds to believe that this result is, indeed, a consequence of some general regularity. It

should be taken into account that neural networks untrained or trained with the random data-sets give random outputs and thus cannot demonstrate any significant correlation either between different output values, or between output values and values of arguments.

It is rather obvious that PM-NO<sub>x</sub>-VOC relationships in the real environment under windy conditions may be affected by peculiarities of distributions of sources of pollutants. Thus, after creation of the models, it was found to be useful to make the approximation of no-wind conditions. That is, all model evaluations were made with zero values of wind speed in Eq. (7). Such approximation is possible as similar values of pollutant concentrations, temperature and humidity can be associated with different magnitudes and directions of wind. No-wind conditions allow us to identify the effects associated with local processes of formation of secondary PM.

The obtained evaluations are further analyzed subsequently in order to find some associations between values (and signs) of different derivatives and argument values. For example, it is investigated, whether days with negative values of first derivatives of PM with respect to VOC are more associated with lower temperatures than days with positive derivatives, or with higher temperatures.

## 4 Results

### 4.1 PM-NO<sub>x</sub> relationship

The rate of gas-to-particle conversion due to the reactions (1)–(3) is proportional to concentration of hydroxyl radical. Thus it can be expected that the increase of the OH level under other constant conditions would lead to the increase of PM production, and vice versa. It is well known that OH itself depends on NO<sub>x</sub> in the essentially non-linear manner (see, for example, Kley et al., 1999). Specifically, OH increases with the increase of NO<sub>x</sub> under a low enough NO<sub>x</sub> level, but may decrease under a higher NO<sub>x</sub> level. Thus, there is a possibility that the decrease in NO<sub>x</sub> may actually lead to the increase of the level of PM. Although this idea is rather obvious, no experimental demonstration of this possibility, as far as we know, has been discussed in peer reviewed literature earlier.

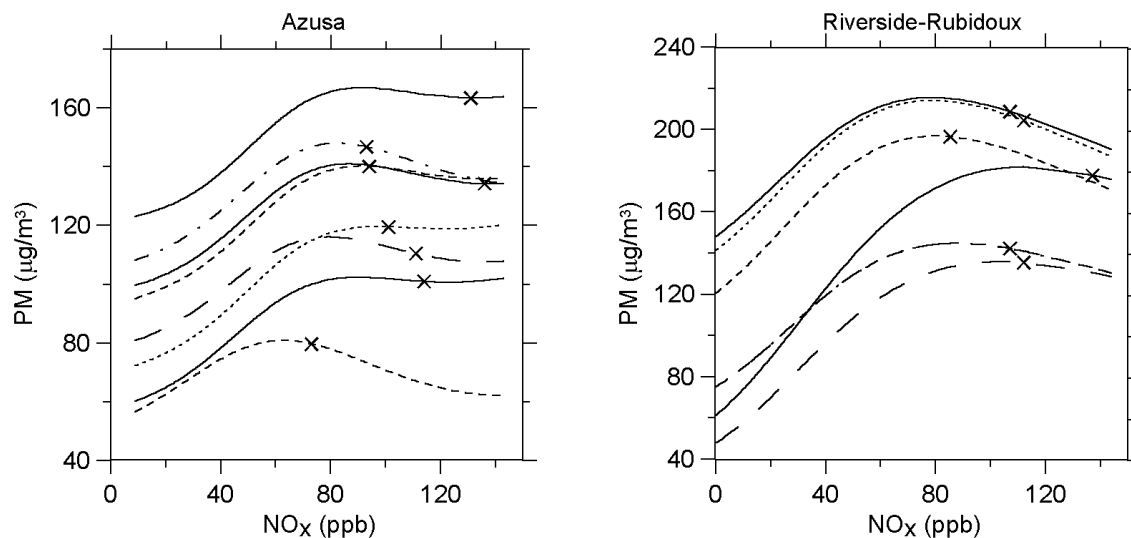
In order to check this idea, we have evaluated (using Eq. 7) the first derivatives of PM with respect to NO<sub>x</sub> ( $\partial\text{PM}/\partial\text{NO}_x$ ) for each day presented in the training sub-sets of the data-sets. It was found that although the majority of those days are associated with positive  $\partial\text{PM}/\partial\text{NO}_x$ , few of them (8 days of 940 in the training sub-set for the Azusa and 6 days of 516 for Riverside) are associated with  $\partial\text{PM}/\partial\text{NO}_x < 0$ . In order to understand the origin of negative  $\partial\text{PM}/\partial\text{NO}_x$  in the empirical models, some additional analysis was undertaken.

First, we considered the dependences of PM on NO<sub>x</sub> for all days with  $\partial\text{PM}/\partial\text{NO}_x < 0$ . Such dependences are shown in Fig. 3. They were calculated from (7) assuming that all

model arguments except NO<sub>x</sub> have fixed values corresponding to the days considered. The general feature of all these dependences is the presence of increasing portions under lower levels of NO<sub>x</sub> and decreasing portions under higher levels of NO<sub>x</sub>. The same features are characteristic for the dependence of OH on NO<sub>x</sub> (Kley et al., 1999). It should be noted, however, that the analysis of the relationships between variables of our empirical models by plotting the corresponding dependences contains potentially more uncertainty than the same analysis performed through evaluating partial derivatives, as, in general, any results of empirical models obtained for input vectors of the model variables which do not coincide with any input vectors in the data sets are expected to be more uncertain than those obtained for the really observed conditions (see also the discussion in the end of Sect. 3). This remark is related, specifically, to the increasing portions of some of the dependences for Azusa under highest values of NO<sub>x</sub>. Most likely, this growth is an artifact of the empirical model and is caused by insufficient amount of data. Indeed, days with NO<sub>x</sub> > 120 ppb constitute less than 3% of the total number of days in the data-set for the Azusa station, while the decreasing portions of the dependences corresponds to the range of NO<sub>x</sub> values which are represented in the data-sets much more abundantly.

Second, it was found that at both stations, days with  $\partial\text{PM}/\partial\text{NO}_x < 0$  are associated, on the average, with significantly higher concentrations of NO<sub>x</sub> (107 ppb at the Azusa station and 110 ppb at the Riverside station), higher concentrations of VOC (2.43 ppmC and 2.61 ppmC) and lower temperatures (18.8° C and 14.6° C), as compared to the corresponding values averaged over the whole data sets (58.5 ppb and 31.6 ppb; 2.38 ppmC and 2.21 ppmC; 22.2° C and 20.4° C at the Azusa and Riverside stations, respectively). Regarding the influence of VOC level on values of  $\partial\text{PM}/\partial\text{NO}_x$  it was found additionally that at both stations all days except one at the Azusa station are associated with positive values of mixed second derivative  $\partial^2\text{PM}/\partial\text{VOC}\partial\text{NO}_x$ . This means that the increase of the VOC level is actually associated with less negative values  $\partial\text{PM}/\partial\text{NO}_x$ . Thus, the noted above association of days when  $\partial\text{PM}/\partial\text{NO}_x < 0$  with higher concentrations of VOC is, most probably, just the result of positive correlation between the ambient VOC and NO<sub>x</sub> levels. Any common, for both models, association between the signs of  $\partial\text{PM}/\partial\text{NO}_x$  and second order derivatives of PM with respect to temperature and NO<sub>x</sub> was not found. This fact, most probably, indicates significant uncertainty of evaluations of these derivatives. It is known from the theory (see, for example, Kleinman, 1994; Sillman, 1999) that higher NO<sub>x</sub> level, lower insolation (which can be expected to be associated with lower temperature) and lower VOC level are favorable for the NO<sub>x</sub>-rich regime, when radical level falls with the increase in NO<sub>x</sub>. Thus, the noted results of the empirical models, at least, do not contradict the hypothesis that negative sensitivity of PM to NO<sub>x</sub> has the photochemical origin.





**Fig. 3.** The dependencies of PM concentrations on  $\text{NO}_x$  obtained from the empirical models for the days with detected negative sensitivity of PM to  $\text{NO}_x$ .  $\text{NO}_x$  concentrations observed on these days are shown by crosses.

Finally, let us discuss the reasons of smallness of a fraction of days with  $\partial\text{PM}/\partial\text{NO}_x < 0$ . We see two possible reasons. On the one hand, as  $\text{NO}_x$  itself can be incorporated into particulate material according to the mechanism (2), (3), it is obvious that the ridge in the dependence of PM on  $\text{NO}_x$  should be shifted toward higher values of  $\text{NO}_x$ , in comparison with the ridge in the dependence of OH on PM under the same conditions, and thus a fraction of days with  $\partial\text{PM}/\partial\text{NO}_x < 0$  should be smaller than a fraction of days with  $\partial\text{OH}/\partial\text{NO}_x < 0$ . On the other hand, in the real environment, the level of  $\text{NO}_x$  may correlate with intensity of the primary emission of PM and/or with the levels of other precursors of PM, such as  $\text{NH}_3$  or  $\text{SO}_2$ , which are not included in the empirical models. Obviously, such correlation would lead to a persistent positive shift in  $\partial\text{PM}/\partial\text{NO}_x$  values evaluated by the empirical models. However, in our opinion even with regard to quantitative uncertainties of the empirical model evaluations, the possibility of negative sensitivity of PM level to a change in  $\text{NO}_x$  is established here rather firmly.

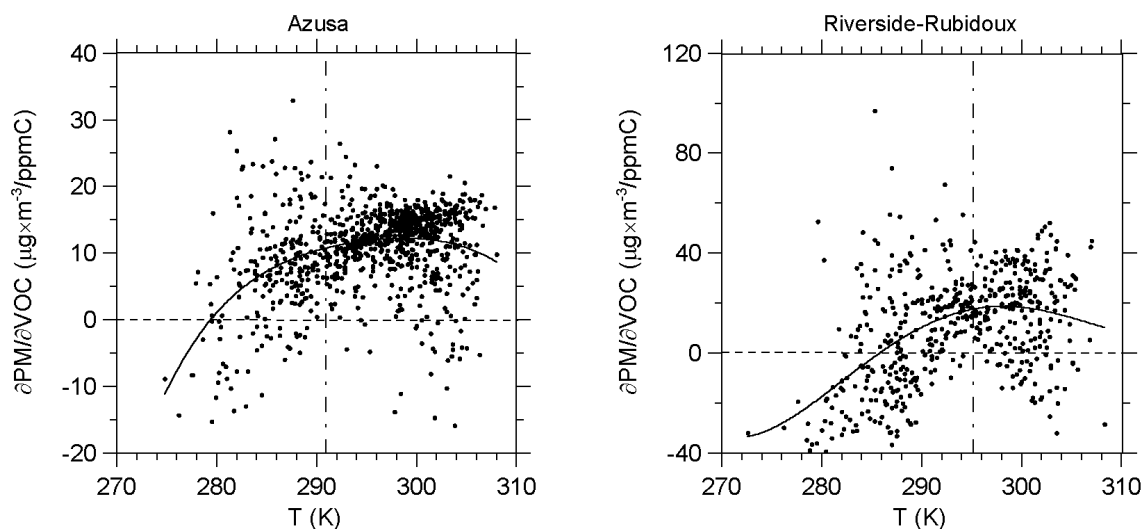
## 4.2 PM-VOC relationship

### 4.2.1 Empirical results

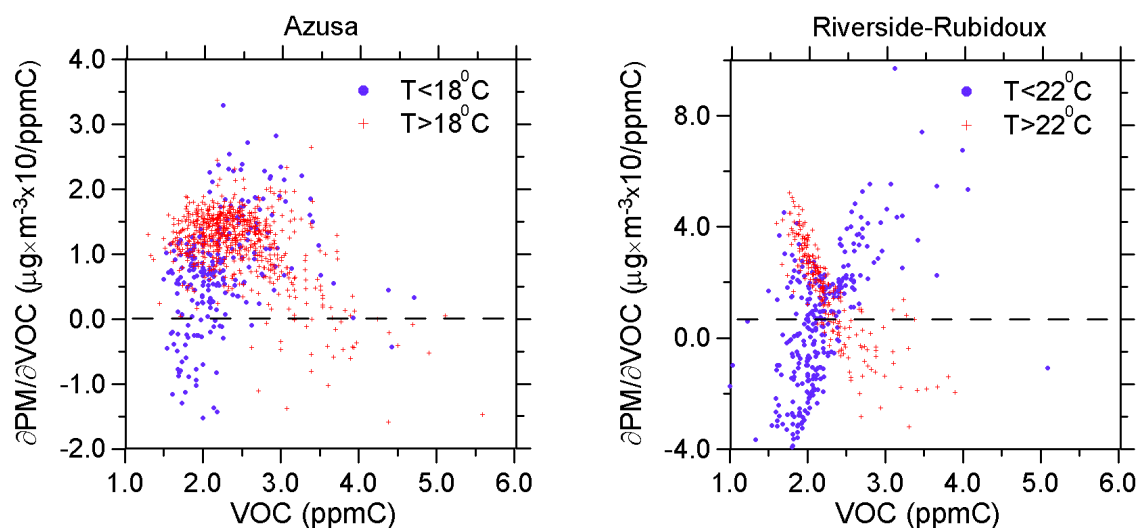
The first evidence for the essential non-linearity of PM-VOC relationship was found by Meng et al. (1997) in results of their PM-CT model simulation of changes of the  $\text{PM}_{2.5}$  level in response to changes in emission rates of  $\text{NO}_x$  and VOC at Riverside, California, under conditions of 28 August 1987. They found, in particular, that a decrease in VOC emission rate may lead either to an increase or a decrease in the  $\text{PM}_{2.5}$

level, depending on the base case for  $\text{NO}_x$  and VOC emission rates. Specifically, greater (by absolute value) negative sensitivity of the PM level to changes in the VOC level was found, in most cases considered, under higher levels of VOC. Unfortunately, fruitful comparison of the theoretical results by Meng et al. (1997) with our empirical results is seriously hindered for the following reasons. (i) The model by Meng et al. and our empirical models use different measures of PM ( $\text{PM}_{2.5}$  and TSP, respectively). (ii) The arguments of the empirical models are surface concentrations of  $\text{NO}_x$  and VOC. Thus, the empirical models cannot account, for example, for possible changes in  $\text{NO}_x$  surface concentration after changes in VOC emission rate due to chemical interaction between VOC and  $\text{NO}_x$ . (iii) The results of Meng et al. concern maximum 1-h average concentration of PM, whereas our empirical models deal with daily averaged level of PM. Nevertheless, analysis of the dependence of PM on VOC given by the empirical model for the Riverside station under conditions of 28 August 1987 (not presented here) reveals that this dependence is also essentially non-monotonous and that high enough values of VOC correspond to negative sensitivity of PM to VOC, and the lower ones to positive sensitivity.

Empirical model evaluations of the first derivative of PM with respect to VOC ( $\partial\text{PM}/\partial\text{VOC}$ ) show that some of the days presented in the data-sets are associated with negative sensitivity of PM to changes of VOC, and others with the positive one. These evaluations are plotted in Figs. 4 and 5 against temperature and concentration of VOC observed on the corresponding days, respectively. It is seen, in particular, that noticeable fractions of the days (namely, 6.7% at the Azusa station and 31.4% at the Riverside station) have



**Fig. 4.** The empirical model evaluations of the first derivatives of concentration of PM with respect to concentration of VOC, plotted against values of ambient temperature. The solid and dashed-dot lines represent 4<sup>th</sup> order polynomial fits of data and the boundary between two temperature modes, respectively. The dotted line marks the null level of the derivative values.



**Fig. 5.** The empirical model evaluations of first derivatives of concentrations of PM with respect to concentration of VOC, plotted against values of ambient concentration of VOC. Blue dots and red crosses correspond to low-temperature and high-temperature modes, respectively.

negative values of  $\partial\text{PM}/\partial\text{VOC}$ . Furthermore, the points with  $\partial\text{PM}/\partial\text{VOC} < 0$  in Fig. 4 are not distributed uniformly but tend “to avoid” definite ranges of moderately warm temperatures, thus forming a “butterfly-like” structures. Solid curves represent polynomial (of the 4<sup>th</sup> order) fits of the data and illustrate nonlinear character of dependence (in the statistical sense) of  $\partial\text{PM}/\partial\text{VOC}$  on temperature. These features allow us to hypothesize that different processes drive behavior of PM under low and high temperatures. This hypothesis finds further confirmation in Fig. 5, where the points are sep-

arated into high-temperature and low-temperature “modes”: the threshold temperatures are chosen to be equal to 18° C at the Azusa station and to 22° C at the Riverside station and marked in Fig. 4 by dash-dot vertical lines. Specifically, it is seen that, under low enough VOC level, the high-temperature mode is associated predominantly with positive  $\partial\text{PM}/\partial\text{VOC}$ , whereas the majority of the points with negative  $\partial\text{PM}/\partial\text{VOC}$  belong to the low-temperature mode. In contrast, under high enough VOC level, the majority of the points with  $\partial\text{PM}/\partial\text{VOC} < 0$  belong to the high-temperature

**Table 2.** Average characteristics of days with  $\partial\text{PM}/\partial\text{VOC} < 0$  and  $\partial\text{PM}/\partial\text{VOC} \geq 0$  for different temperature modes

|   | Azusa  |   |  |   | Riverside-Rubidoux                                 |   |  |   |
|---|--|---|--|---|--|---|--|---|
|   | $T < 18^\circ\text{C}$                             |   | $T \geq 18^\circ\text{C}$                          |   | $T < 22^\circ\text{C}$                             |   | $T \geq 22^\circ\text{C}$                          |   |
|   | $\frac{\partial\text{PM}}{\partial\text{VOC}} < 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} \geq 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} < 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} \geq 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} < 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} \geq 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} < 0$ | $\frac{\partial\text{PM}}{\partial\text{VOC}} \geq 0$ |
| Fraction of days (%)  | 17.2   | 82.8  | 4.0  | 96.0  | 41.9   | 58.1  | 17.0   | 83.0  |
| PM ( $\mu\text{g}/\text{m}^3$ )   | 68.7   | 79.8  | 138.5  | 133.0   | 72.2   | 130.6   | 196.9  | 177.4   |
| VOC (ppmC)  | 1.96   | 2.28  | 3.73   | 2.38  | 1.96   | 2.31  | 2.84   | 2.14  |
| NO <sub>x</sub> (ppb)   | 42.9   | 61.6  | 69.2   | 58.0  | 27.1   | 39.8  | 37.0   | 26.1  |
| T ( $^\circ\text{C}$ )  | 9.8  | 13.6  | 28.1   | 25.3  | 12.6   | 16.9  | 28.2   | 26.7  |
| RH (%)  | 52.7   | 68.8  | 45.8   | 39.7  | 73.2   | 50.5  | 31.1   | 36.7  |
| Fraction of days with $\partial^2\text{PM}/\partial\text{VOC}^2 < 0$                  | 0.42   | 41.4  | 3.3  | 62.1  | 3.8  | 38.7  | 13.5   | 79.9  |
| Fraction of days with $\partial^2\text{PM}/\partial\text{NO}_x\partial\text{VOC} < 0$ | 5.8  | 6.7   | 1.3  | 41.8  | 21.4   | 29.1  | 6.6  | 74.0  |
| Fraction of days with $\partial^2\text{PM}/\partial\text{T}\partial\text{VOC} < 0$    | 0.8  | 29.2  | 4.0  | 67.3  | 1.0  | 12.1  | 17.0   | 59.9  |

mode. The noted features give an idea to perform separate analysis of PM-VOC relationships for different temperature modes.

Average observed values of PM, VOC, NO<sub>x</sub>, T, and RH, calculated separately for high temperature and low temperature modes and separated additionally with regard to the sign of predicted value of  $\partial\text{PM}/\partial\text{VOC}$  in the respective days are listed in Table 2. The table also includes day fractions (of total number of days in the given temperature mode) satisfying certain conditions concerning values of temperature, the sign of  $\partial\text{PM}/\partial\text{VOC}$ , and the sign of the corresponding second order derivatives of PM. The following features general for both models are worthy of noting. First, in the low (high) temperature mode, the average values of PM, VOC, NO<sub>x</sub>, and T are less (greater) on the days with negative  $\partial\text{PM}/\partial\text{VOC}$  than on the days with positive  $\partial\text{PM}/\partial\text{VOC}$ . Second, the minority (majority) of days have negative values of  $\partial^2\text{PM}/\partial\text{VOC}^2$  and  $\partial^2\text{PM}/\partial\text{T}\partial\text{VOC}$  in the low (high) temperature mode. Values of relative humidity do not demonstrate changes general for both models with the change in the sign of  $\partial\text{PM}/\partial\text{VOC}$ . There may be only two reasons for that. First, at least one of the models did not “learn” the relations between  $\partial\text{PM}/\partial\text{VOC}$  and RH satisfactorily. Second, these relations are indeed qualitatively different in different environments. At the given state of knowledge of these relationships, it is impossible to determine which of two reasons is true. The situation is similar with the behavior of  $\partial^2\text{PM}/\partial\text{NO}_x\partial\text{VOC}$ . The only common feature is that the ratio of the number of days having both negative  $\partial^2\text{PM}/\partial\text{NO}_x\partial\text{VOC}$  and  $\partial\text{PM}/\partial\text{VOC}$  to the number of days with negative  $\partial\text{PM}/\partial\text{VOC}$  (this ratio can be obtained by dividing the number in the 8<sup>th</sup> row of the table by that in the 1<sup>st</sup> row) is significantly less than 0.5 in the high temperature mode. An attempt to identify physical and chemical reasons for the noted features is described in the following section.

#### 4.2.2 Discussion of the PM-VOC relationship

First of all, it should be noted that the discussion that will follow does not pursue the notion of exact identification of the processes which cause the noted above features of the PM-VOC relationships. Such identification is outside of the scope of this paper and would require detailed consideration of complex processes that drive evolution of PM at the given site. Here, we would like to formulate and substantiate, wherever possible, the most plausible hypotheses for the explanation of the empirical relationships.

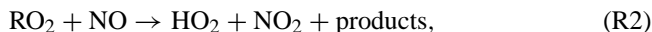
We believe that the main factor driving formation of secondary PM in the high-temperature mode is the behavior of OH radical. The behavior of OH radical is directly related to the behavior of ozone and underlines qualitative nonlinear features of O<sub>3</sub>-NO<sub>x</sub>-VOC relationships which were extensively studied in the two last decades (see, for example, NRC, 1991). Correspondingly, the key processes responsible for the essential general nonlinear features of the dependence of OH on VOC and NO<sub>x</sub> have also been identified earlier. In particular, those features were interpreted by Sillman (1990, 1995) who proposed the mechanistic model based on the equation of the radical budget. It is demonstrated below that the main empirical features of the PM-VOC relationships discussed above are closely associated with features of the OH-VOC relationships given by Sillman’s model. Note that the use of the mechanistic model in this study appears to be quite natural, as our desire here is to appeal to the most general qualitative properties of the tropospheric photochemistry. For the sake of clarity of the discussion, it is reasonable to give here brief description of that model with inessential modifications.

The model includes schematic representation of the sequence of essential processes determining the sources and sinks of the radicals OH, HO<sub>2</sub> and RO<sub>2</sub>. (The latter denotes

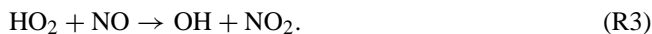
various organic radicals.) Specifically, RO<sub>2</sub> is formed in the reaction of a hydrocarbon (RH) with OH:



The instantaneous equilibrium of RO<sub>2</sub> radicals is provided by the reaction



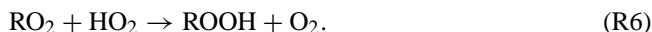
that also provides the source of HO<sub>2</sub> radical, the concentration of which is balanced by the reaction



The magnitude of the sum of concentrations of OH, HO<sub>2</sub> and RO<sub>2</sub> radicals (this sum is called “odd hydrogen”) is determined by the equation of the radical budget

$$S_H = (\text{R4})[\text{NO}_2][\text{OH}] + \left( (\text{R5}) \left( \frac{(\text{R1})[\text{RH}]}{(\text{R3})[\text{NO}]} \right)^2 + \frac{(\text{R6})(\text{R1})^2[\text{RH}]^2}{(\text{R3})(\text{R2})[\text{NO}]^2} \right) [\text{OH}]^2, \quad (8)$$

where the source of odd hydrogen, SH, is assumed to be balanced by its sinks provided by the reaction of OH with NO<sub>2</sub> (see Eq. 3) with the rate coefficient (R4), and the radical-radical reactions



The square brackets in Eq. (8), denote concentrations of species. SH may be further parameterized as

$$S_H = S_1 + S_2[\text{RH}], \quad (9)$$

where it is assumed that the main processes responsible for the radical production are photolysis of ozone and aldehydes.

Assuming that the dependence of the ratios of NO<sub>2</sub> and NO to NO<sub>x</sub> on VOC and NO<sub>x</sub> can be neglected, concentration of OH can be obtained by resolving the quadratic equation. With a little manipulation Eq. (8) can be represented in the most general, non-dimensional form as

$$1 + \beta x = \alpha y v + \frac{x^2}{y^2} v^2, \quad (10)$$

where

$$v = z[\text{OH}], \quad z = \frac{(\text{RH})_0(\text{R1})}{[\text{NO}]_0 S_1^{1/2}} \left( \frac{(\text{R5})}{(\text{R3})^2} + \frac{(\text{R6})}{(\text{R3})(\text{R2})} \right)^{1/2},$$

$$x = [\text{RH}]/[\text{RH}]_0, \quad y = [\text{NO}_x]/[\text{NO}_x]_0, \quad (11)$$

$$\alpha = \frac{(\text{R4})[\text{NO}_2]_0}{z S_1}, \quad \beta = S_2[\text{RH}]_0/S_1,$$

and the subscript “0” denotes some typical values of the species concentrations.

In typical polluted environments, the sources of radicals due to photolysis of ozone and aldehydes are of the same order of magnitude, so a typical value of  $\beta$  is equal approximately to unity. Thus, the dependence of OH on VOC is almost exclusively determined by parameter  $\alpha$ .

Specifically, when  $\alpha \gg 1$ , OH increases with an increase of VOC ( $\partial v/\partial x \approx \beta/\alpha y > 0$ ). However, as  $\alpha$  decreases,  $\partial \text{OH}/\partial \text{VOC}$  can become negative. For example, when  $\alpha \ll 1$ ,  $v \approx y(1 + \beta x)^{1/2}/x$  and  $\partial v/\partial x < 0$ . It is evident from Eq. (11) that a lesser value of  $\alpha$  corresponds to situations with greater overall reactivity of hydrocarbons (i.e. greater (R1)) and greater insolation and ozone concentration (i.e. greater S<sub>1</sub>). It is clear that in the real atmosphere, they are associated with higher, on the average, temperature. These features may be responsible for appearance of negative sensitivity of PM to changes of VOC in the high-temperature mode, the association of days with  $\partial \text{PM}/\partial \text{VOC}$  with higher temperatures and the dominance of negative values of  $\partial^2 \text{PM}/\partial T \partial \text{VOC}$  on those days.

It is also obvious from Eq. (10) that a similar “switch-over” from  $\partial v/\partial x > 0$  to  $\partial v/\partial x < 0$  under fixed parameter values can be caused by either an increase in  $x$ , or a decrease in  $y$ . This feature may be responsible for the empirical facts that the high-temperature mode is mostly associated with negative values of  $\partial^2 \text{PM}/\partial \text{VOC}^2$ , and days with  $\partial \text{PM}/\partial \text{VOC} < 0$  at the high-temperature mode are associated with higher values of VOC and, mostly, with positive values of  $\partial^2 \text{PM}/\partial \text{NO}_x \partial \text{VOC}$ . Note that  $x$  and  $y$  are involved in Eq. (10) asymmetrically. Specifically, when  $\alpha \gg 1$ , an increase in  $y$  leads to a decrease in  $\partial v/\partial x$ , that is  $\partial^2 v/\partial x \partial y < 0$ , in contrast to the situation with  $\alpha$  of about unity, when  $\partial^2 v/\partial x \partial y > 0$ . At the same time, it can be seen that  $\partial^2 v/\partial x^2$  is negative in both these situations and becomes positive only when  $\alpha \ll 1$ . This observation, with the additional assumption that in the real situations  $\alpha$  is greater or insignificantly less than unity, may account for the empirical fact that domination of  $\partial^2 \text{PM}/\partial \text{VOC}^2 < 0$  takes place independently of the sign of  $\partial \text{PM}/\partial \text{VOC}$  in the high temperature mode, whereas domination of  $\partial^2 \text{PM}/\partial \text{NO}_x \partial \text{VOC} > 0$  is found only on days with  $\partial \text{PM}/\partial \text{VOC} < 0$ .

We believe that conditions associated with the low-temperature mode generally correspond to  $\alpha \gg 1$ . Indeed, the CT model evaluations (see, for example, Milford et al., 1994) show that significant areas at South Coast Air Basin (SCAB) are under NO<sub>x</sub>-saturated conditions (that is, in terms of our model, with  $\alpha > 0$ ) even on the hot and sunny days. Thus, it is reasonable to expect that a significant decrease in insolation in colder seasons would drive the photochemistry of SCAB to conditions with  $\alpha \gg 0$ . Under these conditions the concentration of OH is relatively small and, as a result, the behavior of PM in the low-temperature mode is affected by non-chemical processes to much greater extent than that in the high-temperature mode. In particular, it can be speculated that dominating positive values of  $\partial^2 \text{PM}/\partial \text{VOC}^2$  in the low-temperature mode are associated, in part, with a

physically evident possibility that an increase in total concentration of hydrocarbons is associated with an increase of the number of individual semi-volatile hydrocarbons reaching the saturation. Consequently, the increase of measured total concentration of volatile hydrocarbons (that is, VOC) may be associated with the increase of the ratio of total concentration of hydrocarbons in condensed phase to VOC concentration, and as a result, with positive values of  $\partial^2\text{PM}/\partial\text{VOC}^2$ . However, the same effect can be caused by chemical processes of PM formation via oxidation of volatile hydrocarbons (Odum et al., 1996; 1997). The rate of formation of PM due to these processes is proportional to the product of concentrations of VOC and OH, that is, in terms of Eq. (10), it is proportional to  $xv$ . It is easy to see that for  $\alpha \gg 1$ ,  $\partial^2xv/\partial x^2 > 0$ . This process may also account also for substantial fraction of days with  $\partial^2\text{PM}/\partial\text{VOC}^2 > 0$  in the high-temperature mode.

The most speculative point of this discussion concerns the reasons for negative sensitivity of PM to VOC under low temperatures. We failed to find any chemical reason for this, consistent with the majority of available empirical facts. It can be assumed that the increase in VOC concentration associated with the increase in the concentration of condensed semi-volatile hydrocarbons, hinders condensation of inorganic semi-volatile species such as  $\text{NH}_4\text{NO}_3$  by providing a hydrophobic film on the surface of aerosol particles and thus leads effectively to the increase in saturation vapor pressure, or to the decrease the value of Henry's law constants. At the present, we are not aware of experimental or theoretical data that would allow us to estimate actual possibility and significance of this effect. Under higher temperatures, which are associated, presumably, with lesser  $\alpha$  and, consequently, larger values of  $\partial\text{OH}/\partial\text{VOC}$  ( $\partial v/\partial x \approx \beta/\alpha y$  when  $\alpha \gg 1$ ), this effect is suppressed, probably, by the more rapid increase in the concentration of semi-volatile inorganic species with the increase in VOC. This effect may also be suppressed due to the increase of the fraction of organic aerosol with the increase in VOC concentration under the same temperature. Anyway, it is evident that results of the empirical models concerning the low-temperature conditions pose very interesting questions for further researches.

## 5 Summary and conclusions

This paper presents results of the first attempt of analysis of nonlinear relationships between airborne particulate matter and its gaseous precursors based on observational data. The analysis of the long-term monitoring data at South Coast Air Basin, California, is performed using an original method which implies creation of empirical models based on artificial neural networks. Such models do not involve any a priori assumptions regarding the character of analyzed relationships. Reliability of the results of this study is confirmed by the fact that two empirical models developed using independent data-sets obtained at two different monitoring sta-

tions separated by a distance of more than 50 km manifest a number of common qualitative features of the relationships between PM, VOC and  $\text{NO}_x$ . The models are believed to provide the best fit for the real PM-VOC- $\text{NO}_x$  relationships under different observed values of temperature, relative humidity, and horizontal wind speed. In order to suppress the effects caused by transport of PM from different areas and to emphasize the features of the PM-VOC- $\text{NO}_x$  relationships determined by the local formation of PM, the approximation of zero wind speed is used in the models. Under this approximation, the models are used to evaluate different derivatives of PM with respect to VOC and  $\text{NO}_x$  under the conditions corresponding to a number of days presented in the data-sets. These evaluations are further used to identify general, in the statistical sense, regularities of PM-VOC- $\text{NO}_x$  relationships.

It is found, in particular, that there are conditions when PM demonstrates negative sensitivity to changes of  $\text{NO}_x$ , or VOC (that is, when the first derivatives of PM with respect to  $\text{NO}_x$  or VOC are negative). It is found also that negative sensitivity of PM to  $\text{NO}_x$  is facilitated by higher ambient  $\text{NO}_x$  concentrations and lower temperatures and VOC concentrations. These features are consistent with the hypothesis that the non-linearity of the PM- $\text{NO}_x$  relationships is caused by the corresponding non-linearity of the dependence of the concentration of hydroxyl radical (OH) on  $\text{NO}_x$ .

Qualitative features of the relationships between PM and VOC are found to depend on temperature. In view of this dependence it proved fruitful to divide analyzed days into two "modes" with respect to the value of temperature ( $T$ ) observed on those days: the high-temperature mode and the low-temperature one, with the threshold temperature of about 20° C. Consequently, the PM-VOC relationships are analyzed for each temperature mode separately. It is found that in the low-temperature mode, appearance of negative sensitivity of PM with respect to VOC is facilitated by lower temperatures and VOC concentrations, whereas positive and greater values of the sensitivity correspond to higher temperatures and VOC concentrations. In the high-temperature mode, the overall picture is inverse: appearance of negative sensitivity of PM with respect to VOC is facilitated by higher temperatures and VOC concentrations, whereas positive and greater values of the sensitivity correspond to lower temperatures and VOC concentrations.

It is argued that the most significant features of the PM-VOC- $\text{NO}_x$  relationships in the high-temperature mode are closely associated with the similar features of the OH-VOC- $\text{NO}_x$  relationships, whereas the behavior of aerosol in the low temperature mode is strongly perturbed by specific features of the processes of condensation of primary and secondary semi-volatile compounds. In our opinion, the empirical results of this study provide rich "food" for the further theoretical analysis.

The main condition for future successful applications of the methodology discussed in this paper is availability of long enough concurrent measurements of PM and its main

gaseous precursors. Up to now, there were only a few stations satisfying to this criterion. Moreover, unfortunately, the required measurements of VOC at air pollution monitoring stations in California were almost totally cancelled in 1995 and thus it is impossible to employ this methodology for assessment PM-NO<sub>x</sub>-VOC relationships in the present-day situation. Nevertheless, in the near future, this methodology may be used for analysis of data of PM<sub>2.5</sub> and PM precursors measurements at the Photochemical Assessment Monitoring Stations (PAMS) in US (US Environmental Protection Agency, 1994), which were started in the recent years. Anyway, the results of this study emphasize the importance and fruitfulness of concurrent and long-term measurements of coupled parameters of the atmospheric environment.

The results of this study indicate that elaboration of the effective strategy for reduction of the PM level may be a rather difficult problem in view of essential non-linearities of relationships between PM and its gaseous precursors. These non-linearities need further investigation.

*Acknowledgements.* This work was supported by the Russian Foundation for Basic Researches, grant 02-02-17080 and by the Programme for Basic Researches of Russian Academy of Sciences "Physics of the Atmosphere: Electricity Processes and Radio-physical Methods of Research", Project No. 4.1.

## References

- Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., and Shankar, U.: Modal aerosol dynamics model for Europe: Development and first applications, *Atmos. Environ.* 32, 2981–2999, 1998.
- Binkowski F. S. and Shankar, U.: The regional particulate matter model, 1. Model description and preliminary results, *J. Geophys. Res.*, 100, 26 191–26 209, 1995.
- Bishop, C. M.: *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.
- California Ambient Air Quality Data 1980–1997 (CAAQD): CD No. PTSD-98-010-CD, Air Resources Board, California Environmental Protection Agency, 1998.
- Delucchi, M. A., Murphy, J. J., and McCubbin, D. R.: The health and visibility cost of air pollution: a comparison of estimation methods. *J. Environ. Management*, 64, 139–152, 2002
- Finlayson-Pitts, B. J. and Pitts, J. N. Jr.: *Atmospheric Chemistry: Fundamentals and Experimental Techniques*, John Wiley and Sons, New York, NY, 1986.
- Jacobson, M. J.: Development and application of a new air pollution modeling system II. Aerosol module structure and design. *Atmos. Environ.*, 31, 131–144, 1997.
- Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, 359–366, 1989.
- Gardner, M. W. and Dorling, S. R.: Artificial neural networks (the multilayer perceptron)– A review of applications in the atmospheric sciences, *Atmos. Environ.*, 32, 2627–2636, 1998.
- Kleinman, L. I.: Low and high NO<sub>x</sub> tropospheric photochemistry, *J. Geophys. Res.*, 99, 16 831–16 838, 1994.
- Kleinman, L. I.: Ozone processes insights from field experiments – part II: Observation-based analysis for ozone production, *Atmos. Environ.*, 34, 2023–2033, 2000.
- Kley, D., Kleinmann, M., Sanderman, H., and Krupa, S.: Photochemical oxidants: state of science. *Environ. Pollution*, 100, 19–42, 1999.
- Konovalov, I. B.: Application of neural networks for studying nonlinear relationships between ozone and its precursors, *J. Geophys. Res.*, 107, D11, 10.1029/2001JD000863, 2002.
- Meng Z., Dabdub, D., and Seinfeld, J. H.: Chemical coupling between atmospheric ozone and particulate matter, *Science*, 277, 116–119, 1997.
- Meng Z., Dabdub, D., and Seinfeld, J. H.: Size-resolved and chemically resolved model of atmospheric dynamics, *J. Geophys. Res.*, 103, 3419–3435, 1998.
- Milford, J. B., Gao, D., Sillman, S., Blossy, P., and Russel, A. G.: Total reactive nitrogen (NO<sub>y</sub>) as an indicator of the sensitivity of ozone to reductions in hydrocarbon and NO<sub>x</sub> emissions, *J. Geophys. Res.*, 99, 3533–3542, 1994.
- National Research Council (NRC), Committee on Tropospheric Ozone Formation and Measurement: *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, National Academy Press, NY, 1991.
- Nelson, M. C. and Illingworth, W. T.: *A Practical Guide to Neural Nets*, Reading, MA: Addison-Wesley, 1991.
- Nenes, A., Pandis, S. N., and Pilinis, C.: Continued development and testing of a new thermodynamic aerosol module for urban and regional air quality models, *Atmos. Environ.*, 33, 1553–1560, 1999.
- Odum, J. R., Hoffmann, T., Bowman, F., Collins, D., Flagan, R. C., and Seinfeld, J. H.: Gas/particle partitioning and secondary organic aerosol yields, *Environ. Sci. Technol.*, 30, 2580–2585, 1996.
- Odum, J. R., Jungkamp, T. P. W., Griffin, R. J., Flagan, R. C., and Seinfeld, J. H.: The atmospheric aerosol-forming potential of whole gasoline vapor, *Science*, 276, 96–99, 1997.
- Pilinis, C. and Seinfeld, J. H.: Development and evaluation of an Eulerian photochemical gas-aerosol model, *Atmos. Environ.*, 22, 1985–2001, 1988.
- Reed, R. D. and Marks, R. J.: *Neural Smoothing: Supervised Learning in Feedforward Artificial Neural Networks*, Cambridge, MA: The MIT Press, 1999.
- Sarle, W. S.: Stopped Training and Other Remedies for Overfitting, *Proceedings of the 27<sup>th</sup> Symposium on the Interface of Computing Science and Statistics*, 352–360, 1995.
- Sarle, W. S.: Usenet newsgroup comp.ai.neural-nets, <ftp://ftp.sas.com/pub/neural/FAQ3.html>, 2003.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics: From air pollution to Climate Change*, John Wiley & Sons, New York, NY, 1998.
- Sillman, S., Logan, J. A., and Wofsy, S. C.: The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes, *J. Geophys. Res.*, 95, 1837–1851, 1990.
- Sillman, S.: The use of NO<sub>y</sub>, H<sub>2</sub>O<sub>2</sub>, and HNO<sub>3</sub> as indicators for ozone-NO<sub>x</sub>-hydrocarbon sensitivity in urban locations, *J. Geophys. Res.*, 100, 14 175–14 188, 1995.
- Sillman, S.: The relation between ozone, NO<sub>x</sub> and hydrocarbons in urban and polluted rural environments, *Atmos. Environ.*, 33, 1821–1845, 1999.
- Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X.: The second generation regional acid deposition model chemical

- mechanism for regional air quality modeling, *J. Geophys. Res.*, 95, 16 343–16 367, 1990.
- Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S.: A new mechanism for the regional atmospheric chemistry modeling, *J. Geophys. Res.*, 102, 25 847–25 879, 1997.
- Sun, Q. and Wexler, A.: Modeling urban and regional aerosols near acid neutrality Application to the 24–25 June SCAQS episode, *Atmos. Environ.*, 32, 3533–3545, 1998.
- US Environmental Protection Agency: Photochemical Assessment Monitoring Station Implementation Manual, EPA-454/B-93/051, Research Triangle Park, NC 27711, 1994.
- Wexler, A. S., Lurmann, F. W., and Seinfeld, J. H.: Modelling urban and regional aerosols I. Model development, *Atmos. Environ.*, 28, 531–546, 1994.