Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytica Chimica Acta 549 (2005) 179–187
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

Predicting bulk ambient aerosol compositions from ATOFMS data with ART-2a and multivariate analysis

Weixiang Zhao [a], Philip K. Hopke [a,*], Xueying Qin [b], Kimberly A. Prather [b]

[a] *Department of Chemical Engineering, and Center for Air Resources Engineering and Science, Clarkson University, P.O. Box 5708, Potsdam, NY 13699-5708, USA*
[b] *Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0314, USA*

Received 11 March 2005; received in revised form 31 May 2005; accepted 6 June 2005
Available online 15 July 2005

Abstract

The aerosol time-of-flight mass spectrometry (ATOFMS) has not generally been used to provide a quantitative estimation of chemical compositions of ambient aerosols. In an initial study, the possibility of developing a calibration model to predict chemical compositions from ATOFMS data was demonstrated, but because of the limited number of samples (only 12), the ability of the calibration model was not fully realized. In this study, 50 samples were created to further test the prediction ability of the calibration model. The conceptual framework is to relate the mass concentrations of the particles in the identified classes to the average aerosol compositions for each sampling time interval using a calibration model based on ART-2a and multivariate analysis. There may be some non-linearity between cluster mass concentrations and ambient species concentrations because of measurement errors, the scaling equations used to estimate particle mass and various assumptions required for building the model. Thus, in this study, PLS regression was integrated with radial basis functions (RBF-PLS) to obtain better prediction effects and compared to partial least square (PLS) regression alone. Compared with an earlier study, these results provide better and a more convincing demonstration of the ability of the calibration model to estimate the chemical compositions from ATOFMS data. The results also suggest that the model would be able to provide carbon data and thus substitute for thermal optical reflectance (TOR) measurements. Additionally, the calibration model based on RBF-PLS showed more accurate predictions in the cases with some non-linearity. Some of the key steps in the modeling effect are also discussed in detail.

Keywords: ATOFMS; Aerosol; Calibration model; Adaptive resonance theory; Aritificial neural networks; ART-2a; Partial least squares; PLS; Radial basis functions; RBF

1. Introduction

Ambient aerosols have been proved to have adverse effects on environment quality and human health [1,2]. Motor vehicle exhaust, road dust, industrial emissions, biogenic emissions and other pollution sources make the exposure to ambient aerosols unavoidable. New techniques and data analysis tools have been applied to study ambient particles [3,4]. First developed in 1994, the aerosol time-of-flight mass spectrometry (ATOFMS) measures the size and composition of individual aerosol particles in real time [5]. This technique provides information to understand the size and composition distribution of atmospheric particles [3,6,7]. However, it is extremely difficult for ATOFMS to provide a quantitative estimation of the species bulk mass concentrations of particles. Bulk chemical compositions are very helpful for studying the relationship between ambient aerosols and human diseases and for designing pollution control strategies. Solving this estimation problem will make the use of the ATOFMS more advantageous, extend the application fields of ATOFMS (an sufficiently accurate ATOFMS calibration model can reduce the experiment work needed to measure chemical species concentrations), and enhance the substantial investment in developing this important aerosol monitoring instrument.

* Corresponding author. Tel.: +1 315 268 3861; fax: +1 315 268 4410.
E-mail address: hopkepk@clarkson.edu (P.K. Hopke).

0003-2670/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2005.06.012

Fergenson et al. [8] initially studied the development of a calibration model to estimate the ambient aerosol chemical composition from ATOFMS data. However, because of the limited samples for that research (only 12 samples), the ability of the multivariate calibration model to predict bulk chemical compositions was not sufficiently demonstrated. Thus, the goals of the present study are: (1) to fully prove the feasibility and effect of the multivariate calibration model based on adaptive resonance theory (ART) neural networks and partial least square regression (PLSR) on estimating bulk aerosol chemical compositions from ATOFMS data, (2) to discuss the influences of the non-linearity caused by measurement errors (of both ATOFMS and species concentrations), the employed experiential equations and various assumptions on the accuracy of the calibration model [9] and (3) accordingly to provide a method with better prediction effect on the cases with some non-linearity.

2. Method description

The whole data analysis process consisted of two major parts. First, the individual particles were clustered based on their individual mass spectrum and the mass concentration of the particles in each cluster was estimated. Then, the mass concentrations of the identified classes were used to predict bulk aerosol compositions with the multivariate calibration model. The conventional methods for these two parts have been described in detail elsewhere [8,10] and only a brief introduction to them are presented in the subsequent sections.

2.1. Description of ART-2a

Various cluster analysis methods, typically ART-2a, have been applied for the on-line particle composition analysis. There are a number of reports involving the use of ART-2a for the classification of single particle mass spectrometry data [10,11]. In the classification of ATOFMS data, the inputs of ART-2a are the positive and negative ion mass spectral data for each particle and the output is the index of the class each particle belongs to. Compared with most clustering methods, the significant advantage of ART-2a is the ability to add a new cluster without disturbing any existing clusters, and thus, it has the potential to be used for on-line data analyses.

Suppose the sample set to be clustered is denoted by $\{\mathbf{x}_i | i = 1, \ldots, n\}$, where $\mathbf{x}_i$ is the sample vector and n is the number of samples. The training algorithm for ART-2a is briefly described below. The details are provided in literatures [12,13].

1. Randomly select an input vector and scale it into unit length.

$$\mathbf{p}_i = \frac{\mathbf{x}_i}{||\mathbf{x}_i||} \tag{1}$$

2. Contrast enhancement: transfer all elements of $\mathbf{p}_i$ through a non-linear transfer function.

$$q_{ij} = \begin{cases} p_{ij}, & \text{if } p_{ij} > \theta \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where θ is a threshold value for discriminating against noise [13]. Signals smaller than the threshold are set to 0. Generally, θ is set to a value between 0 and $1/\sqrt{d}$, where d denotes sample dimensions [13]. In this study, the range of mass-to-charge (m/z) of the spectral sample is [−350, +350], so d equals 700. Finally, θ was set to 0.005, since it provided a reliable clustering result for the calibration model.

3. Rescale $\mathbf{q}_i$ to unit vector $\mathbf{r}_i$.
4. Compare the resonances between the input vector and the cluster vectors of all existing l output neurons and determine the neuron with the largest resonance as "winner". The resonance is represented as the dot product of the input vector and the existing cluster vector.

$$\rho_k = \mathbf{r}_i \mathbf{w}_k (k = 1, 2, \ldots, l) \quad \text{and} \quad \rho_{\text{win}} = \max(\rho_k) \tag{3}$$

5. If the resonance of the winner neuron is larger than the pre-defined vigilance limit ρ_{vig} (in this study, ρ_{vig} was 0.6), modify the cluster vector of the winner neuron toward the input vector according to the following procedure (equations (4)–(7)). Vigilance is a key parameter to control the cluster number. The larger the vigilance, the more the classes. An over-large vigilance would result in an "over-fine" clustering result (the extreme case is one cluster for one sample), while an over-small vigilance would result in an "over-coarse" result. There is no generalized rule to determine vigilance value. In this study, ρ_{vig} was set to 0.6, since it provided a feasible clustering solution for the calibration model.

$$\mathbf{v}_{ij} = \begin{cases} \mathbf{r}_{ij}, & \text{if } \mathbf{w}^{\text{old}}_{(\text{win})ij} > \theta \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where θ is as the same as defined in equation (2).

$$\mathbf{u}_i = \frac{\mathbf{v}_i}{||\mathbf{v}_i||} \tag{5}$$

$$\mathbf{t}_i = \mathbf{w}^{\text{old}}_{\text{win}} + \eta(\mathbf{u}_i - \mathbf{w}^{\text{old}}_{\text{win}}) \tag{6}$$

$$\mathbf{w}^{\text{new}}_{\text{win}} = \frac{\mathbf{t}_i}{||\mathbf{t}_i||} \tag{7}$$

where η is learning rate. In general, η should be smaller than 0.5 (in this study it was 0.1). Otherwise, create a new cluster as below.

$$\mathbf{w}_{\text{new}} = \mathbf{r}_i \tag{8}$$

Repeat the above steps for all the input vectors, which are defined as a cycle. In ideal cases, the criterion for stopping the training of ART-2a is when the change between the cluster vectors of two consecutive cycles is zero or smaller than the pre-defined criterion value. However, in ATOFMS studies,

it is almost impossible to reach the above ideal criteria, so in this study the criterion was to set a pre-defined number of cycles. The initial cluster vectors were randomly selected from the clustering sample set and scaled. The clustering results were used to estimate the mass concentrations of the particle classes in each time interval. The detailed process will be explained in the following sections.

2.2. Description of PLSR

PLS regression is a generalization of multiple linear regression (MLR) [14]. The significant advantage of PLSR over traditional MLR is that PLSR can analyze strongly collinear and noisy data, and also simultaneously model a number of response/dependent variables [14,15]. In general, a linear regression model can be written as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{9}$$

where **X** and **Y** are independent and dependent variables, respectively, **B** contains the regression coefficients and **E** is the residual matrix. In PLSR, **X** can be transformed to

$$\mathbf{X} = \mathbf{TC}^{\mathrm{T}} \tag{10}$$

where **T** is the matrix of PLS components and **C** is the loading matrix. Let $\Lambda = \mathbf{C}^{\mathrm{T}}\mathbf{B}$, then the PLS regression model can be written as

$$\mathbf{Y} = \mathbf{T}\Lambda + \mathbf{E} \tag{11}$$

It can be seen that the regression of **X** against **Y** is turned into the regression of PLS components T against **Y**. The detailed process can be found in Hoskuldsson [16]. In this study, **X** and **Y** denote the mass concentrations of all particle classes and the bulk species concentrations in each sampling time interval, respectively. Thus, a ATOFMS calibration model is available. However, the calibration model could be affected by various types of errors, such as measurement error, the estimate of particle density and the experimentally measured inlet efficiency. These errors could render the relationship between the mass concentrations of clusters and the ambient species concentrations non-linear. Thus, the next section introduces a novel method to solve the possible non-linearity problems in the calibration model.

2.3. Description of RBF-PLS

One effective method to solve non-linear problems is to convert them into linear problems. As a kernel function, radial basis function (RBF) can transfer the input space to a transitional linear space. Thus, linear methods can be applied to build the relationship between the transitional space and output space. This process is the principle of conventional radial basis function networks (RBFN) [17,18]. The advantages of integrating RBF with PLS over conventional RBFN are to make full use of the information of all the samples and to solve the problems of determining the radial bases (such as the local optima of the radial basis vectors obtained through a K-means algorithm). The integration of RBF and PLS extend the application of PLS to non-linear problems [19,20]. A typical radial basis function is:

$$a_{ij} = \exp\left(\frac{-(||\mathbf{x}_i - \mathbf{c}_j||^2)}{\sigma_j^2}\right) \tag{12}$$

where $\mathbf{x}_i$ is the input vector, $\mathbf{c}_j$ the radial basis vector, σ_j the radial basis width and a_{ij} is the output of radial basis j on input vector i. In the RBF-PLS approach, each input sample (cluster mass concentration vector in this study) is a radial basis vector. Suppose there are n samples, thus one will have an $n \times n$ transitional matrix according to equation (12).

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{13}$$

Then, PLS is applied to build a relation between the transitional matrix **A** and outputs (species concentrations in this study). The details of this process can be found in Walczak and Massart [19] and Zhao et al. [20].

3. Data treatment and analysis

In this study, the ATOFMS data were collected in Fresno, CA. The sampling period was from December 1, 2000 to February 3, 2001. Both the positive ions and the negative ions were provided in the ATOFMS data, so the range of mass-to-charge (m/z) for this study was set to [−350, +350]. The input to the ART-2a analysis was a 700 variable vector. The bulk aerosol species concentrations were measured as part of the California Regional Particulate Air Quality Study [21]. The California Regional $PM_{10}/PM_{2.5}$ Air Quality Study is a comprehensive public/private sector collaborative program with two main goals: (1) to provide an improved understanding of particulate matter and visibility in central California and (2) to provide decision-makers with the tools needed to identify equitable and efficient control methods. The species concentration data of $PM_{2.5}$ for the Fresno site were collected from December 15, 2000 to February 3, 2001. Only on 11 days (December 15, 16, 17, 18, 26, 27 and 28, January 31 and February 1, 2 and 3) were both ATOFMS data and species concentration data available. The species concentrations for each day were collected in five time intervals 0:00–5:00, 5:00–10:00, 10:00–13:00, 13:00–16:00 and 16:00–24:00 h. Thus, the number of possible time periods for the calibration model was $11 \times 5 = 55$. The total number of measured particles in these 55 periods was 230,432.

It was a problem to cluster such a large number of particles with ART-2a networks. In Fergenson et al. [8], a total of 12,479 particle samples were grouped into 12 cohorts, which finally generated 12 samples for the calibration model.

Each cohort was classified individually by ART-2a but the weight matrix (cluster vectors) was preserved from one analysis to the next. Any particles that did not fit into the existing classes nucleated their own classes. The feature of ART-2a that ART-2a can create new classes without disturbing the existing classes permits this procedure to function, but in terms of system completeness, it would be better to cluster the samples at the same time. In this study, 230,432 particles were clustered at the same time. The vigilance factor for this study was 0.6. Initially, a total of 1339 classes were created. However, most of the 1339 classes contained very few particles. Twenty-eight classes accounting for 80% of the total particle mass were retained for further analysis. The selected classes were the top 28 in terms of particle mass and each of the rest classes accounted for less than 1% of total particle mass.

The cluster results provided the number of the particles in each class in each time interval. ATOFMS provides the aerodynamic diameter of each particle that can then be used to estimate the physical diameter, so the particle mass concentrations of the identified classes in each time interval can be estimated. The ATOFMS instruments do not detect particles of all aerodynamic diameters equally. Larger particles are detected with a higher efficiency than smaller particles, so a scaling equation was applied to relate the particle detection efficiency to the aerodynamic diameter of a particle [8]. The detection efficiency as a function of particle size can be expressed as

$$N = \alpha D_{\mathrm{a}}^{\beta} \tag{14}$$

where N is the number of particles in a given volume of air per particle observed by ATOFMS in that volume, D_{a} the aerodynamic diameter in micrometer of the particle and α and β are the coefficients that were determined through calibration experiments. In this study, the coefficients were set to be 1383.12 and −4.312, respectively. The density of each spherical particle was assumed to be $1.3\,\mathrm{g\,cm^{-3}}$ [8]. For further details on the data pre-treatment process, see Fergenson et al. [8]. In addition, in order to ensure the statistical reliability of the clustering results of 55 time periods, the periods that contained less than 1000 particles were excluded from analysis. Thus, 50 time periods were retained for final analysis, i.e., 50 mass concentration vectors (28 dimensions) were available to build a calibration model.

In the species concentration data, the species whose missing or below detection limit measurements were more than one third of the total measurements were excluded from

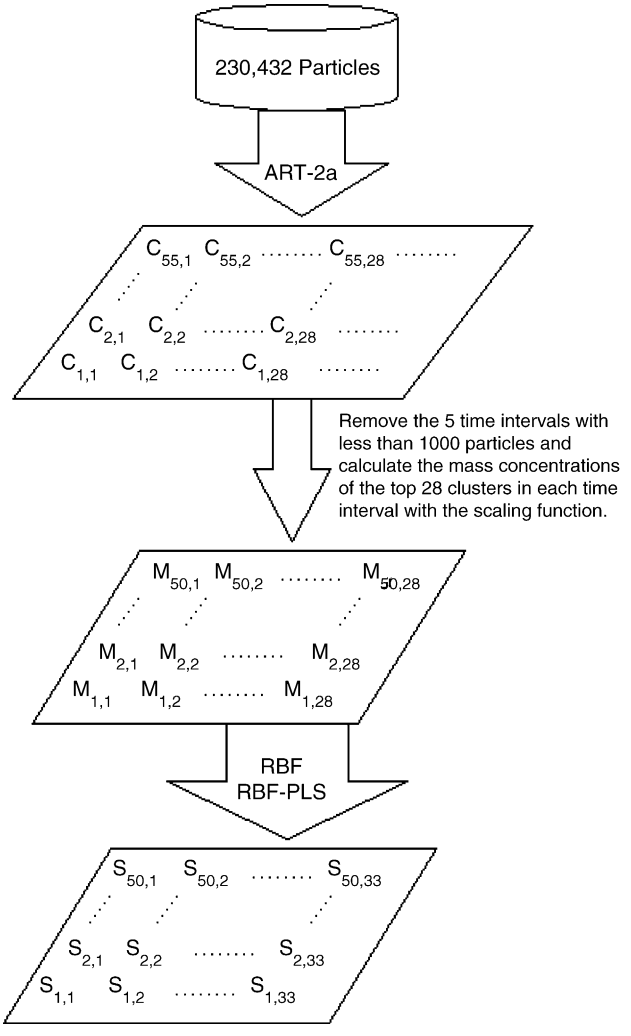

Fig. 1. Flowchart of the data pre-treatment process for the calibration model (C, cluster; M, mass concentration of cluster; S, species concentration).

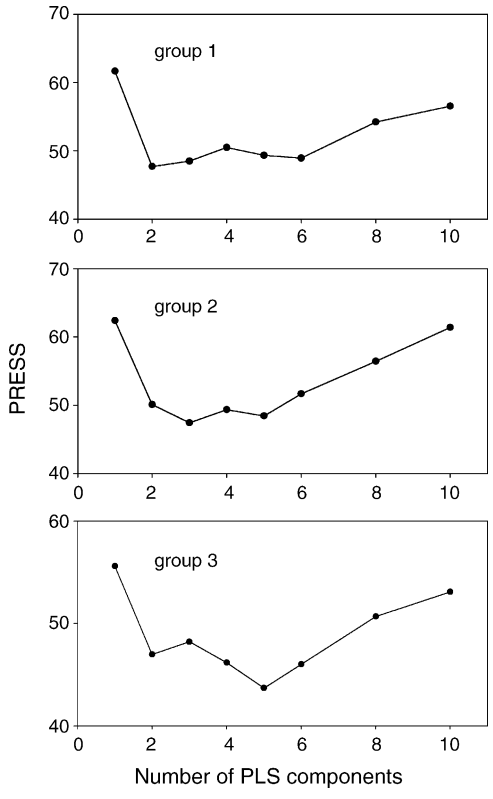

Fig. 2. PRESS vs. PLS component number.

analysis. Finally, 33 species (1: Cl^-, 2: NO_3^-, 3: SO_4^{2-}, 4: NH_4^+, 5: Na^+, 6: K^+, 7: OC1 (OC: organic carbon), 8: OC2, 9: OC3, 10: OC4, 11: OC (total organic carbon), 12: EC1 (EC: element carbon), 13: EC2, 14: EC (total element carbon), 15: TC (total carbon), 16: Na, 17: Mg, 18: Al, 19: Si, 20: S, 21: Cl, 22: K, 23: Ca, 24: Mn, 25: Fe, 26: Ni, 27: Cu, 38: Zn, 29: As, 30: Se, 31: Br, 32: Rb and 33: Pb) each of which had 50 measurements were retained to build the calibration model. The ions were measured by ion chromatography (IC). The carbon fractions (OCs and ECs) were measured by quartz filters and thermal optical reflectance (TOR). This protocol volatilizes organic carbon (OC) in four temperature steps in a helium atmosphere: OC1 at 120 °C, OC2 at 250 °C, OC3 at 450 °C and OC4 at 550 °C. OC4 responses return to constant values. Pyrolyzed organic carbon (OP) is oxidized at 550 °C in a mixture of 2% oxygen and 98% helium atmosphere until the return of filter's reflectance to its initial value. Then, three elemental carbon fractions are measured in an oxidizing atmosphere: EC1 at 550 °C, EC2 at 700 °C and EC3 at 850 °C [22]. All the other species were measured by X-ray fluorescence spectroscopy (XRF). Thus, both independent and dependent variables for the calibration model were available. The whole pre-treatment process is summarized in Fig. 1.

In order to test the predicting ability of the calibration model, 20 samples were randomly selected from the 50 samples to build the model and the other 30 samples were for testing. This random selection was performed three times. The corresponding training/testing sample sets were called groups 1–3, respectively. In the 50 calibration samples, 11 samples did not contain any missing and below detection limit values in the species concentration data (they were

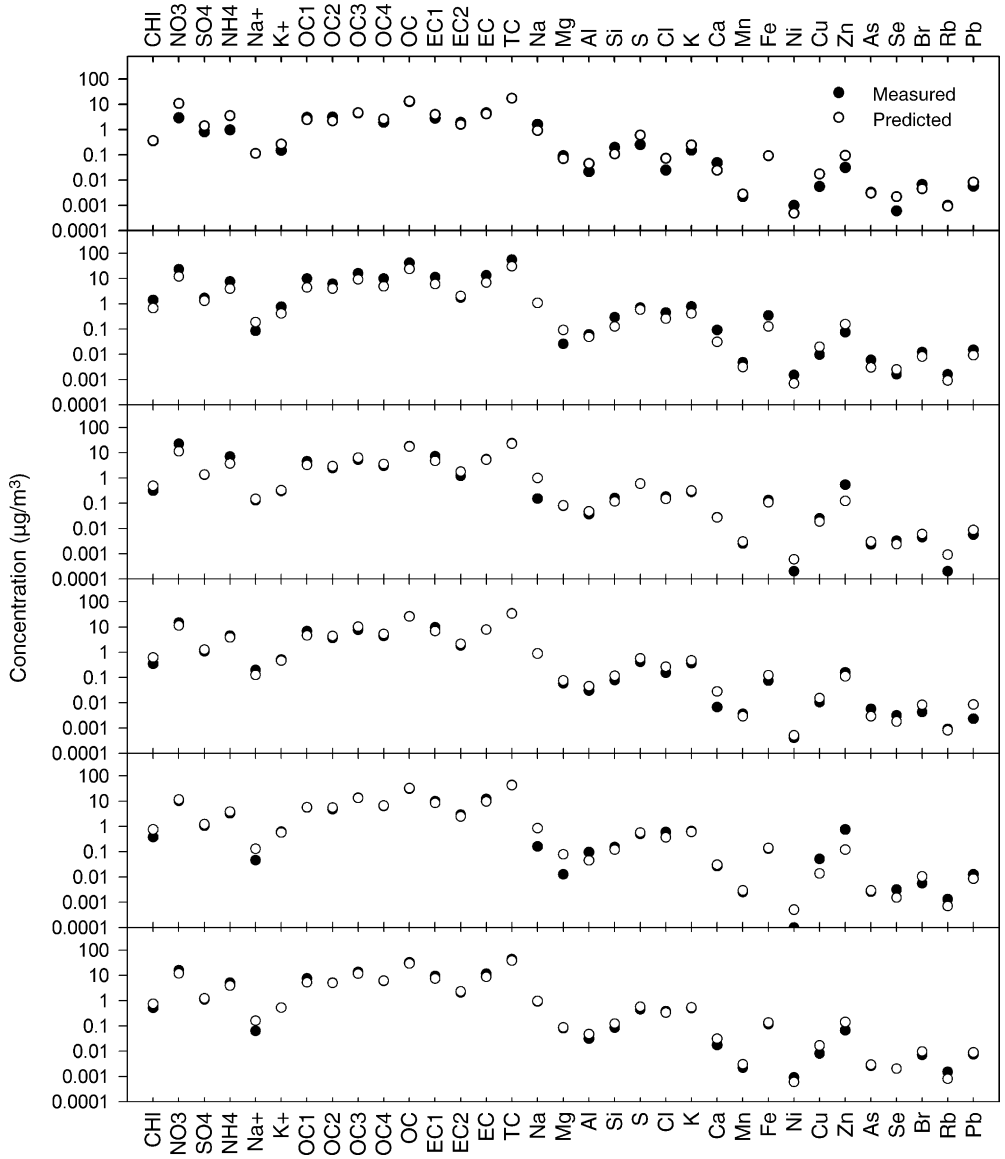

Fig. 3. Comparison between the predicted and measured species concentrations of six normal testing samples in group 1 (PLS).

called "normal samples" and were collected during 0:00–5:00 and 5:00–10:00 h on 12/15/00, 0:00–5:00 h on 12/18/00, 16:00–24:00 h on 12/28/00, 0:00–5:00 and 16:00–24:00 h on 2/1/01, 0:00–5:00 and 10:00–13:00 h on 2/2/01 and 0:00–5:00, 5:00–10:00 and 16:00–24:00 h on 2/3/01, respectively). The discussion will be focused on the normal samples. Groups 1–3 contained six to eight normal test samples, respectively. In addition, the independent and dependent variables were scaled and centered.

4. Results and discussion

In PLS modeling, the determination of the number of PLS components is one of the critical problems. The predictive error of sum of squares (PRESS) was used as the criterion for determining the PLS component number.

$$\text{PRESS} = \sum_{i=1}^{30}\sum_{j=1}^{33}(y_{ij} - \tilde{y}_{ij})^2 \tag{15}$$

where y_{ij} is the concentration of species j in testing sample i and is the estimation of $\tilde{y}_{ij}$. Different number of PLS components produced different PRESS values as shown in Fig. 2. First, the PRESS decreased and then stayed relatively flat, but finally increased with increasing PLS component numbers. This rise occurs because additional noise was being included in the model as extra PLS components were added into model. This behavior is called "over-fitting" that creates well-fit models with poor or no predictive ability. Thus, the proper PLS component number should lie in the relatively flat range. In this study, in addition to the PRESS, the prediction

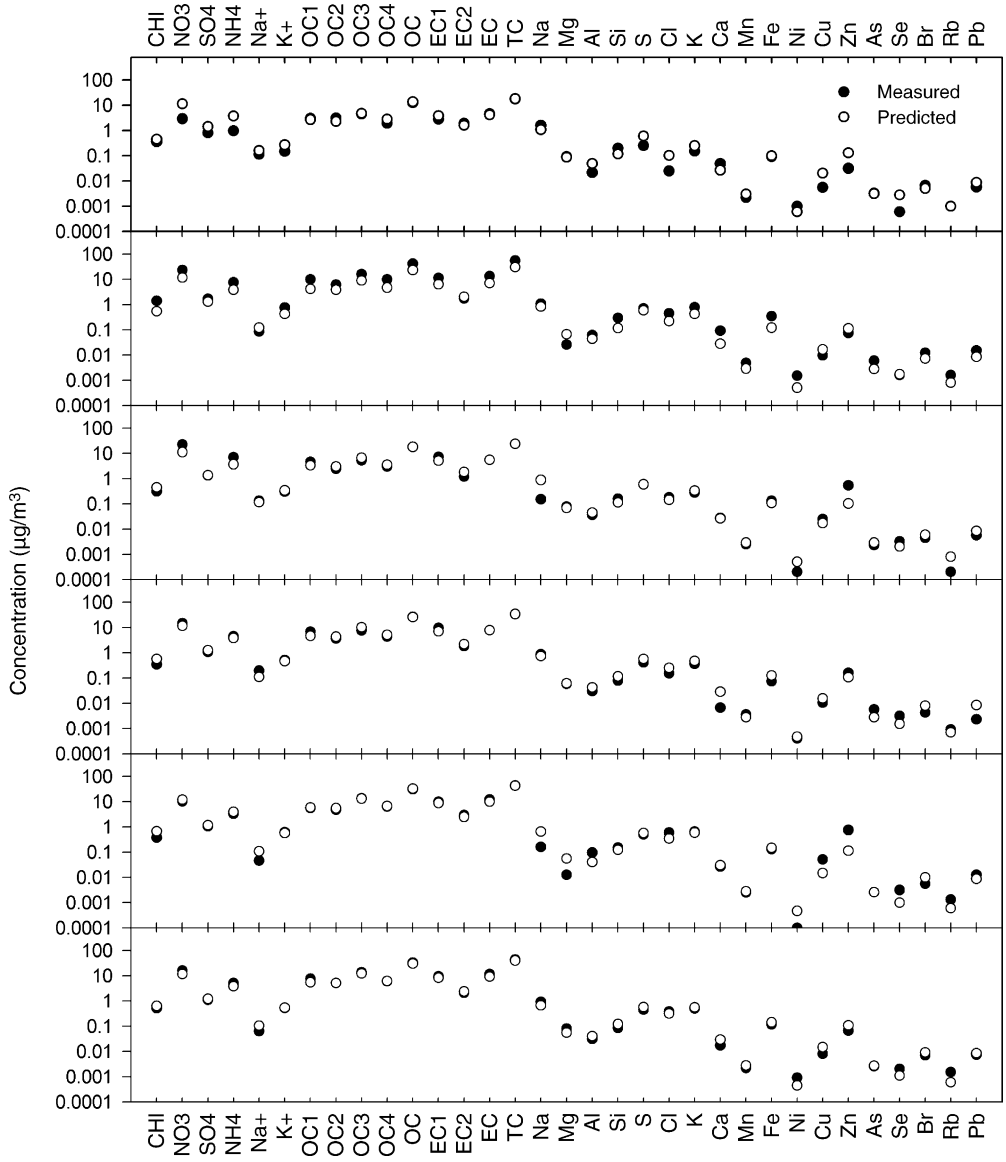

Fig. 4. Comparison between the predicted and measured species concentrations of six normal testing samples in group 1 (RBF-PLS).

Table 1
The correlation coefficients of the predicted and measured concentrations of all the species in the normal testing samples

	R^2 (PLS)	R^2 (RBF-PLS)
Group 1 (33 × 6 pairs)	0.91	0.91
Group 2 (33 × 7 pairs)	0.95	0.95
Group 3 (33 × 8 pairs)	0.93	0.94

errors exclusively of the thermal carbon fractions were also given particular concerns (the reason for the concerns will be explained later), so in order to get a "dual win" solution that makes the prediction errors for carbon fractions as small as possible without significantly changing or increasing the PRESS (the error for the whole system), the PLS component numbers for the three groups were set to 2, 5 and 5, respectively.

In the RBF-PLS model, in addition to the PLS component number, the width for radial basis function is an important parameter. As a Gaussian kernel function, the radial basis function controls its response region through its width. Too small a width could result in an overly sharp response such that there is almost no response outside a narrow range. Alternatively, too large a width could make the radial basis function yield the same response for all of the input samples. Clearly, both choices produce inappropriate models. In this study, the width for each radial basis was set as the same value 22.

One criterion for testing the prediction ability is the correlation coefficient R^2 between the predicted and measured values. Table 1 shows the R^2 values of the predicted and measured concentrations of all the species in the normal testing samples of each group. The numbers of the data pairs for calculating R^2 of groups 1–3 were 33×6, 33×7 and 33×8, respectively. Clearly, each group of this study has a better calibration effect than the initial study [8] where the R^2 value was 0.83. In addition, the RBF-PLS approach provided somewhat better results than PLS alone.

As an illustrative example, Figs. 3 and 4 show the predicted species concentrations (of PLS model and RBF-PLS model, respectively) and the measured ones of six normal testing samples in group 1. It can be seen that almost every black circle (measured) is covered by or overlaps with the white circle (predicted). Like the results of the initial study [8], this study also showed that the predictions of the higher concentration species were better than the low concentration variables. One possible reason was that the high concentration species exert greater influence on the ART-2a analysis, ensuring that the identified classes depend more heavily on their concentrations [8]. Fig. 5 shows the relative prediction errors of two methods for the thermal carbon fractions of six normal testing samples in group 1. The fact that PLS and RBF-PLS show the similar prediction error variation patterns for each species (i.e., when PLS yields a relatively large/small error RBF-PLS also yields a relatively large/small one) suggests the non-linearity in this calibration model is not too high. The comparison in quantity of the prediction abilities of two methods will be discussed in detail as below.

Table 2 shows the mean values of the prediction errors of the thermal carbon fractions in the normal test samples. The

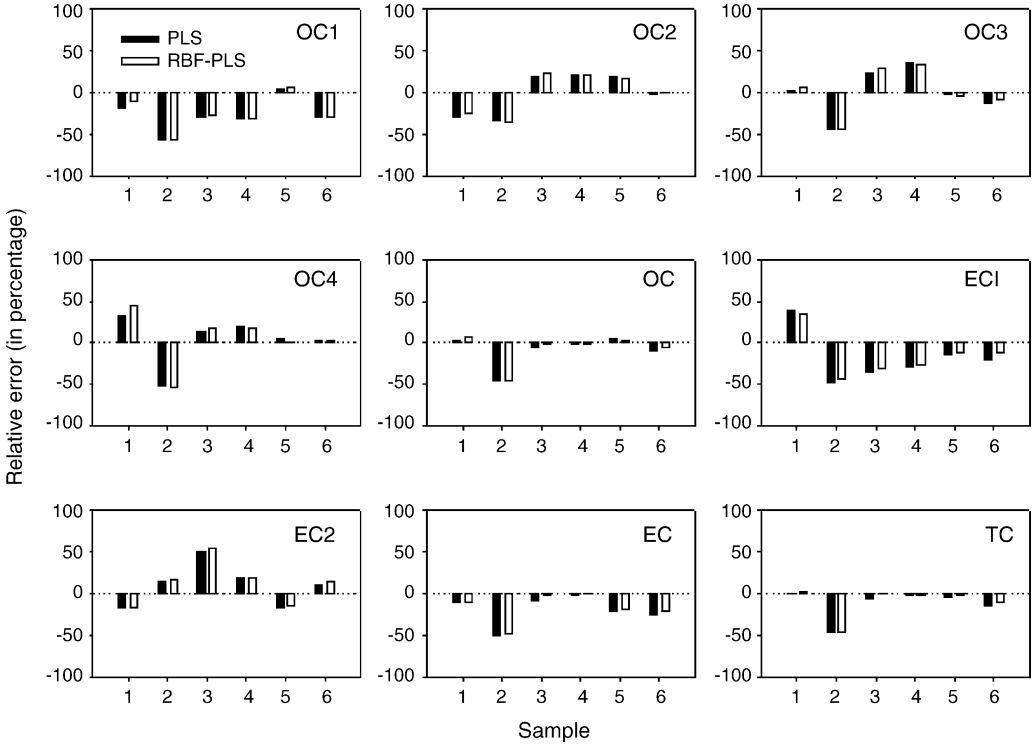

Fig. 5. Comparison between the prediction errors of PLS and RBF-PLS with respect to the thermal carbon fractions of six normal testing samples in group 1.

Table 2
Mean value of the relative prediction errors (in percentage) of each thermal carbon fraction of each group (for normal testing samples)

	Group 1		Group 2		Group 3	
	PLS	RBF-PLS	PLS	RBF-PLS	PLS	RBF-PLS
OC1	28.8	**27.4**	32.1	37	41.6	43.3
OC2	20.5	**20.2**	16.3	**13.7**	32.2	**27.2**
OC3	19.6	21.1	21.3	**18.3**	36.4	**34**
OC4	20.7	22.9	17.1	18	20.7	22.2
OC	11.3	**10.8**	14.1	**10**	23.8	**19.3**
EC1	30.7	**26.2**	24.3	**18.9**	32.1	**30.1**
EC2	21	22.7	34.4	37.2	25.9	28
EC	18.7	**16.5**	25.4	**21.8**	31.8	**31**
TC	11.7	**10.4**	14	**11.1**	24.1	**21.3**
Ave_all[a]	20.3	**19.8**	22.1	**20.7**	29.8	**28.5**
Ave_3[b]	14.2	**14.1**	16.5	**13.1**	28.1	**24.9**

[a] Ave_all: the average over the mean relative prediction errors of all the thermal carbon fractions.
[b] Ave_3: the average over the mean relative prediction errors of OC3, OC and TC.

bolded values highlight those variables that are better fit by the RBF-PLS as compared to the PLS model. The average mean errors for all the carbon fractions and those for OC3, OC and TC are also listed in Table 2. The carbon fractions were selected as examples since they have relatively large concentrations (especially, OC3, OC and TC) and they are the very important species for air quality studies. For example, they can assist in the identification/apportionment of gasoline and diesel emissions [23,24]. It can be seen in the results for both methods on sample groups1 and 2 that the average mean error of the carbon fractions is approximately 20%, and the average mean errors of OC3, OC and TC are ∼15%. The measurement uncertainties for the carbon fractions are typically larger than 15% [25]. Thus, both PLS and RBF-PLS showed good prediction capability for the carbon fractions. The prediction errors of group 3 were somewhat larger than for the other groups. One of the possible reasons could be the samples selected for group 3 are not as similar to those samples included in the calibration set as those in groups 1 and 2. The RBF-PLS shows better predictions than PLS suggesting there is some non-linearity in this system. It also shows the ability of RBF-PLS to deal with the non-linearity.

In this study, the width for each radial basis in RBF-PLS approach was set to the same value. It is likely that RBF-PLS could provide more accurate results if the radial basis widths are set individually (i.e., each radial basis determines its individual width according to the space distribution of its surrounding samples). A genetic algorithm could be used to obtain the optimal widths for the radial basis functions. These issues will be further explored in future work.

5. Conclusions

In this paper, 50 calibration samples were available to test the feasibility of developing a calibration model to predict the bulk aerosol chemical composition from ATOFMS single particle data. Compared with the initial study that had only 12 calibration samples [8], this study showed a better calibration model based on ART-2a and PLS/RBF-PLS. The comparison between the predictions of the calibration models (PLS and RBF-PLS) of the carbon fractions suggests the calibration model could provide comparable data to the thermal optical reflectance (TOR) measurements. In addition, the results of this study suggest that RBF-PLS is a better choice for non-linear calibration problems.

Acknowledgments

This work was supported in part by the California Air Resources Board under contract number 01-348 and by the U.S. Environmental Protection Agency through Science to Achieve Results (STAR) grant number R831083.

References

[1] D.W. Dockery, C.A. Pope, X.P. Xu, J.D. Spengler, J.H. Ware, M.E. Fay, B.G. Ferris, F.E. Speizer, N. Engl. J. Med. 329 (1993) 1753.
[2] J. Cyrys, J. Heinrich, G. Hoek, K. Meliefste, M. Lewne, U. Gehring, T. Bellander, P. Fischer, P. Van Vliet, M. Brauer, H.E. Wichmann, B. Brunekreef, J. Exposure Anal. Environ. Epidemiol. 13 (2003) 134.
[3] L.S. Hughes, J.O. Allen, M.J. Kleeman, R.J. Johnson, G.R. Cass, D.S. Gross, E.E. Gard, M.E. Gälli, B. Morrical, D.P. Fergenson, T. Dienes, C.A. Noble, D.-Y. Liu, P.J. Silva, K.A. Prather, Environ. Sci. Technol. 33 (1999) 3506.
[4] X.-H. Song, N.M. Faber, P.K. Hopke, D.T. Suess, K.A. Prather, J.J. Schauer, G.R. Cass, Anal. Chim. Acta 446 (2001) 329.
[5] K.A. Prather, T. Nordmeyer, K. Salt, Anal. Chem. 66 (1994) 1403.
[6] D.Y. Liu, R.J. Wenzel, K.A. Prather, J. Geophys. Res. 108 (2003) 8426.
[7] J.R. Whiteaker, K.A. Prather, Atmos. Environ. 37 (2003) 1033.
[8] D.P. Fergenson, X. Song, Z. Ramadan, J.O. Allen, L.S. Hughes, G.R. Cass, P.K. Hopke, K.A. Prather, Anal. Chem. 73 (2001) 3535.
[9] P.J. Gemperline, J.R. Long, V.G. Gregoriou, Anal. Chem. 63 (1991) 2313.
[10] X. Song, P.K. Hopke, D.P. Fergenson, K.A. Prather, Anal. Chem. 71 (1999) 860.
[11] D.J. Phares, K.P. Rhoads, A.S. Wexler, D.B. Kane, M.V. Johnston, Anal. Chem. 73 (2001) 2338.
[12] G.A. Carpenter, S. Grossberg, D.B. Rosen, Nerual Networks 4 (1991) 493.
[13] Y. Xie, P.K. Hopke, D. Wienke, Environ. Sci. Technol. 28 (1994) 1921.
[14] S. Wold, M. Sjöström, L. Eriksson, Chemom. Intell. Lab. Syst. 58 (2001) 109.
[15] P.D. Wentzell, L.V. Montoto, Chemom. Intell. Lab. Syst. 65 (2003) 257.
[16] A. Hoskuldsson, J. Chemom. 2 (1988) 211.
[17] S. Chen, C. Cowan, P.M. Grant, IEEE Trans. Neural Networks 2 (1991) 302.

[18] F. Lampariello, M. Sciandrone, IEEE Trans. Neural Networks 12 (2001) 1235.
[19] B. Walczak, D.L. Massart, Anal. Chim. Acta 331 (1996) 177.
[20] W. Zhao, D. Chen, S. Hu, Comput. Chem. Eng. 28 (2004) 1403.
[21] http://www.arb.ca.gov/airways/crpaqs/overview.htm.
[22] J.G. Watson, J.C. Chow, D.H. Lowenthal, L.C. Pritchett, C.A. Frazier, Atmos. Environ. 28 (1994) 2493.
[23] W. Zhao, P.K. Hopke, Atmos. Environ. 38 (2004) 5901.
[24] E. Kim, P.K. Hopke, J. Air Waste Manage. Assoc. 54 (2004) 733.
[25] http://vista.cira.colostate.edu/improve/.