Available online at www.sciencedirect.com
SCIENCE DIRECT
Analytica Chimica Acta 555 (2006) 47–56
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

Cluster analysis of single particle mass spectra measured at Flushing, NY

Liming Zhou, Philip K. Hopke*, Prasanna Venkatachari

Center for Air Resources Engineering and Science, Department of Chemical Engineering, Clarkson University, P.O. Box 5708, Potsdam, NY 13699-5708, USA

Received 3 March 2005; received in revised form 23 August 2005; accepted 24 August 2005
Available online 29 September 2005

Abstract

ART-2a and a density based cluster method, density based spatial clustering of application with noise (DBSCAN), have been used for classification of the single particle mass spectra measured at New York City. Using too large of a vigilance factor in ART-2a leads to many similar clusters with overlap, and thus a low vigilance factor was used in this study. The DBSCAN method can identify clusters with complex shapes and various sizes, and representative spectra are chosen to identify different particle types within each cluster. The cluster structure of the single particle mass spectra were examined by DBSCAN. Both methods found that the major clusters were sea salt and anthropogenic combustion emissions. The continua in sulfate, potassium and OC particles were found by DBSCAN and a large cluster was formed, while ART-2a broke it into several small clusters without finding this continuum. A detailed discussion of the cluster analysis results including representative mass spectra, size distributions and temporal behavior will be provided.

Keywords: Aerosol; Single particle mass spectrum; Cluster analysis; Representative Object; ART-2a; DBSCAN

1. Introduction

Airborne particles play an important role in influencing regional visibility and global climate [1]. High particulate matter (PM) levels have also been found to be associated with increased morbidity and mortality [2,3]. Measurement of atmospheric PM, including the size distribution and chemical composition of particles, is fundamental for further understanding these issues.

Many instruments have been used to measure the size distribution and bulk phase chemical composition, and in the past several years, real time single particle mass spectrometry has developed rapidly [4]. These kinds of instruments measure the size and composition of single particles simultaneously and have provided insight into particle sources, chemical transformations between particles and gases and distributions among or within particles. Different single particle mass spectrometry techniques have been developed, including particle analysis by laser mass spectrometry (PALMS), aerosol time-of-flight mass spectrometry (ATOFMS) and rapid single particle mass spectrometry (RSMS). These instruments were deployed and compared at the Atlanta supersite as well as an aerosol mass spectrometer (AMS), which measures ensembles of particles instead of single particles [5].

The single particle measurement generates huge amounts of data. In a one month field study, >200,000 mass spectra can be obtained. These data cannot be analyzed manually and cluster analysis methods have been applied, including ART-2a neural network [6,7] and hierarchical regression tree analysis [8]. Murphy et al. [8] compared ART-2a and regression tree. His conclusion is that ART-2a has the advantage of online analysis and less computing time, and that regression tree has the advantages of convergence and generating clusters with less overlap than ART-2a.

Another cluster method, density based spatial clustering of application with noise (DBSCAN) has recently been introduced into the chemistry field [9,10]. Since this method can identify clusters with arbitrary shapes and sizes, it produces clusters without overlap and the continua between different types of particles caused by internal mixing may be found. For separating overlapping clusters, partition methods can be used inside the DBSCAN

* 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.2005.08.061

clusters. It was demonstrated that DBSCAN could find clusters with complex and even concave shapes while a partition method such as *K*-means failed [9,10].

In this study, the single particle mass spectra measured in Flushing, New York City, are clustered with both ART-2a and DBSCAN. These two methods will be compared.

2. Experiment

During the $PM_{2.5}$ Technology Assessment and Characterization Study (PMTACS-NY) winter intensive, single mass spectra were measured by an ATOFMS (TSI 3800) at the Queens College site in Flushing NY (latitude: 40.74, longitude: 73.82), January 10–12, 14 and 15, 2004. The ATOFMS has been described in a number of previous publications [11–13] and only a brief description will be given here.

Air is drawn into the ATOFMS instrument and a narrow particle beam is formed through a converging nozzle followed by skimmers. The particle then enters a sizing stage. The time-of-flight is obtained when the particle crosses two laser beams and can be converted to size through a size calibration file. A pulsed Nd:YAG laser is triggered by detection of particle velocity and the particle is then ablated and ionized. The ions produced by the ablation are analyzed by time-of-flight mass spectrometry. For each particle, positive and negative ion mass spectra are recorded at the same time.

A total of 59,989 single particle spectra were measured by ATOFMS and 46,676 of them have both negative and positive ion signals. The time-of-flight for aerodynamic sizing was also recorded for each particle. The ablation laser broke before an on-site size calibration could be done, so a factory calibration was used to convert the flight times to sizes. As a result, the sizes obtained may have some errors and can only be used to compare particles measured in this study in a qualitative way. All of the mass spectra were converted to a list of peaks with in-house software. The mass to charge ratio (m/z) range is from −350 to 350 Da. The measured particle number per hour is presented in Fig. 1. The highest number/hour values were on the morning of January 12.

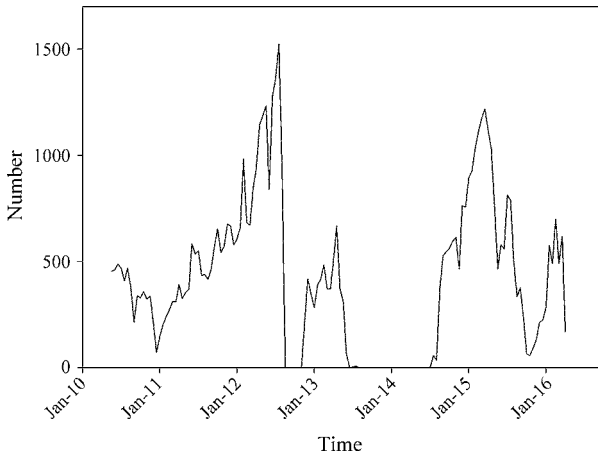

Fig. 1. Temporal profile of total sampled particle number with 1 h resolution. When there is no sampling, the particle number drops to zero.

3. Data analysis

First, the positive and negative ion mass spectra were normalized separately; the missing negative spectra were taken as all zeros. Next, the positive and negative spectrum from each particle were combined into one vector and normalized. For all of the algorithms used in this study, the distance is defined by the dot product of two normalized spectra vectors.

3.1. Clustering by ART-2a

In ART-2a, a sample spectrum is chosen randomly and the inner products of the spectrum with the weight vectors are computed. The weight vector with the highest inner product is updated if this product exceeds a criterion, the vigilance factor (VF) and the sample spectrum is assigned to this weight vector. Otherwise, the sample spectrum is added as a new weight vector. The whole process is repeated until the convergence criterion is reached. The ART-2a algorithm was described elsewhere [14,15].

3.2. Density based clustering

A density based cluster method, DBSCAN, was developed by Ester et al. [16] and introduced into the chemistry field by Daszykowski et al. [9]. This method determines a cluster by investigating density, the number of objects in the neighborhood space of the current object that are within a given radius ε. If the density is greater than a criterion value, i.e. the number of objects k within the radius ε, each current object is thought to be a core object. Otherwise, the current object can be either a border object or an outlier object if there is no core object in the neighborhood defined by representative object the k and ε parameters representative object. Only two parameters, ε and k, need to be assigned for the DBSCAN algorithm. In this study, they were chosen to be 0.9 and 20, respectively. These two parameters were decided empirically. When the dot product of two mass spectra is greater than 0.9, they can be thought to be identical [8]. Thus, this neighborhood is considered to be sufficiently close. A value of 20 was suggested by Daszykowski et al. [9] for large data sets. Actually, the DBSCAN algorithm is not as sensitive to these parameters as ART-2a is to the VF. When using large k values the clusters will "shrink" as compared to those obtained using small k values, more outliers may be produced, and some large clusters may be broken into smaller ones. When using lower k values, more clusters may be combined. For clusters without connection and with similar sizes and densities, DBSCAN and other partition methods find similar clusters. The presence of continua are one of the difficulties of the partition methods since the partition methods separate them by hiding the continua. In DBSCAN, the continua are defined by the ε and k values. Partition methods can be used to find representative objects within a large cluster for better understanding the different particle types (not clusters) and also provides an understanding as to why there are continua for these particles.

Table 1
Summary of the ART-2a classification results with various vigilance factors ρ_{max}; N_T is the total number of classes; and N_{99} and N_9 are the number of classes explaining 99 and 90% of the total mass, respectively

ρ_{max}	0.3	0.4	0.5	0.6
N_T	100	165	303	750
N_{99}	56	81	152	337
N_9	26	34	61	121

For DBSCAN, the nature of the initiation process does not influence the final results. The DBSCAN algorithm is deterministic and there are no iteration and convergence problems. For ART-2a, the initiation, in terms of the order in which the objects are presented to the program, may have some effect since convergence may be achieved at slightly different locations depending on the order in which the particles are analyzed.

The DBSCAN algorithm can be described as follows:

a. At the beginning all objects are flagged as unprocessed.
b. Randomly choose an unprocessed object as the current object, mark it as processed. If the current object is a core, create a new cluster and assign the current object to it and go to 'c'. If not, move to the next unprocessed object. When all the objects are processed, terminate the algorithm.
c. Find neighbors of the current object within the distance ε, assign them to the cluster created in step 'b', mark as processed, and transfer them to 'seeds'. If all objects have been processed, then continue to step 'd', otherwise return to step 'b'.
d. Take each object in 'seeds' as the current object and perform 'c'. When all objects are processed, go back to 'b'.

Since some clusters found by DBSCAN may be very large and the center of the cluster is not sufficient to represent all the objects, a partition method like *K*-means was suggested as a post-processing method [10]. In this study, the Kennard and Stone algorithm [10,17] is used to select a representative object within each individual cluster found by DBSCAN. The first representative is chosen as the object that is the nearest to the mean of the current cluster. Subsequent representative object are chosen as the most distant object from the previous representative object. The distance of an object from previous

Table 2
Summary of the classification results when $\rho_{max} = 0.4$

Class no.	Major component of negative spectra	Major component of positive spectra	Volume fraction	Number
1	NO_3^-, NO_2^-	Na^+	0.3008	10949
2	NO_3^-, NO_2^-	Mg^+	0.1096	3689
3	–	Na^+	0.0663	2263
4	$(NO_3)_2^-$, HCO_3^-	Mg^+	0.0354	1035
5	HSO_4^-	K^+ + OC	0.0285	4479
6	NO_3^-, NO_2^-	K^+	0.0284	2474
7	–	K^+ + OC	0.0269	3816
8	–	K^+	0.0246	1542
9	HSO_4^-	OC	0.0246	3244
10	–	Na^+, $C_2H_3^+$/Al^+	0.0234	278
11	NO_2^-	Ca^+	0.0229	991
12	NO_2^-	Mg^+, NaH_2O^+	0.0215	523
13	NO_3^-, NO_2^-	–	0.0174	923
14	–	Si^+	0.0145	170
15	NO_3^-	K^+ + OC	0.014	1325
16	NO_3^-, NO_2^-	Ca^+	0.0135	867
17	NO_3^-, NO_2^-	CH_3CO^+/AlO^+	0.0125	1049
18	–	Na^+, K^+	0.0123	561
19	–	Fe^+	0.0107	617
20	NO_2^-, Cl^-	Mg^+	0.0097	179
21	PO_4^-, $H(NO_2)_2^-$	Na^+	0.0097	277
22	–	Mn^+	0.0079	158
23	HSO_4^-, NO_2^-	K^+	0.0078	1048
24	–	OC	0.0068	1578
25	HSO_4^-	EC	0.0062	1587
26	NO_2^-	$C_2H_3^+$/Al^+	0.0061	168
27	–	$CaOH^+$	0.0059	153
28	–	Al_2^+	0.0056	75
29	SO_4^-	OC	0.0056	1320
30	NO_3^-	OC	0.0052	851
31	HSO_4^-	OC	0.0051	247
32	PO_3^-	K^+ + OC	0.0051	205
33	HNO_2^-	–	0.0049	99
34	PO_4^-	K^+ + OC	0.0048	983

These classes account for 90% of the total mass.

representative object is defined as the greatest dot product of the object with all representative object and the distance thus defined will ensure the object found is the farthest from all the previous representative object. When this dot product is greater than 0.7, the selection of representative object stops. Thus, all the members in one cluster can be assigned to a representative with a dot product over 0.7. After this post-processing step, the objects within each cluster are assigned to the representative that yields the largest dot product. The number of representative object is also a measure of cluster size. More representative object indicates large cluster representative object sizes.

All the outliers will be assigned to their nearest cluster if the distance between an outlier and any cluster member is over 0.7. These outliers are called the "cloud" of the cluster they are assigned to. For a cluster, the density is usually high at the center but low at the border. It is possible that the outlier near a cluster still belong but is not identified since a certain density criterion is used. The "cloud" is only for reference and was not used to find representative object.

Because of the complexity of the data structure and the algorithms, as well as the large amount of data, a special class was constructed to store all the original spectra, variables and functions used in the cluster analysis, and the object oriented (OO) property of the C++ language was utilized.

4. Results and discussion

4.1. ART-2a

Table 1 shows the number of classes with different vigilance factors. One of the difficulties in applying ART-2a is that when using large VFs like 0.7, too many clusters were found and there are significant overlaps among the clusters. When using a small VF of 0.4, the clusters have much less overlap, but the spectra inside each cluster may be dissimilar and the weight vector of each cluster may not be representative. This problem seems to be caused by large cluster sizes. Pastor et al. [13] found similar clusters even with VF = 0.5 and combined those similar clusters manually. In Table 2, using a VF of 0.4, the results of the 34 classes which account for 90% of the total mass (<2.5 μm) are summarized.

The major particle classes include sodium-containing, magnesium-containing, potassium + OC, calcium-containing and OC. The sodium-containing classes are thought to be from sea salt and previous cluster analysis of ATOFMS measurements also found these classes [6]. The *m/z* ratios of −62 (NO_3^-) and −46 (NO_2^-) are markers of nitrate [7] and the largest classes are sodium nitrate and magnesium nitrate. $NaNO_3$ is formed from sea salt through reaction with HNO_3 resulting in the depletion of chloride. The reaction between sea salts and HNO_3 was previously observed using the ATOFMS technique [18]. This class explains the largest particle number and volume of all the classes.

Magnesium particles are also thought to be from sea salt. As shown in Fig. 2, Class 2 has nearly the same negative spectrum as Class 1, suggesting that the reactions these two

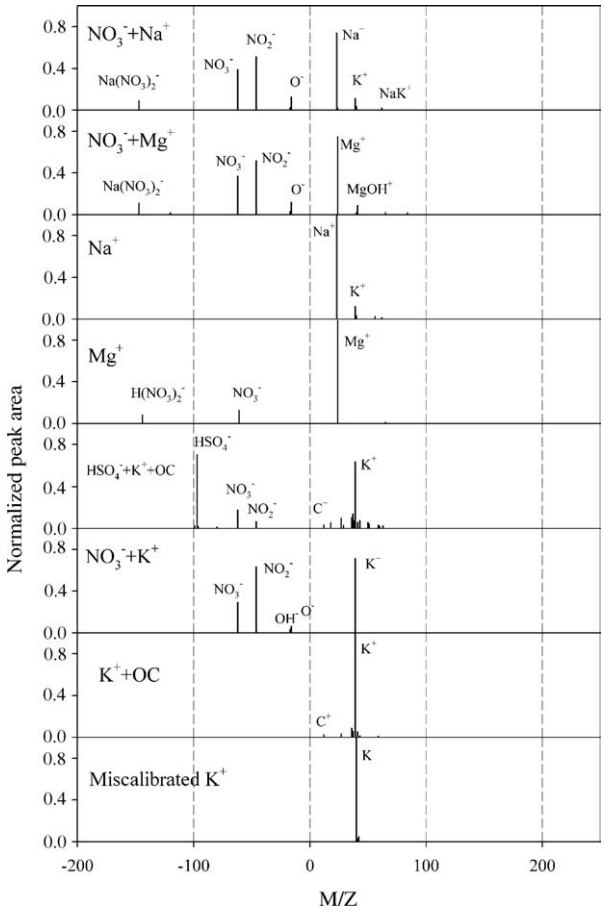

Fig. 2. Mass spectra of Classes 1–8 by ART-2a.

kinds of sea salt particles experience during transport are similar.

Fig. 3 indicates that most of the sea salt particles (Classes 1–4) have sizes around 1 μm. These four classes account for half of the total volume. These sea salt particles are smaller than those reported elsewhere [6,13], and this discrepancy may be due to an inappropriate size calibration.

Class 3 has no negative ions like sulfate and nitrate, and may represent sea salt particles that have not been processed by the atmosphere. This lack of processing may explain the larger numbers of smaller particles in Fig. 3 as compared to the processed sea salt classes. However, it is also possible that low ablation efficiencies for sulfate and nitrate in the smaller particle sizes are responsible for the lack of negative ions.

Potassium plus organic fragments is found in Classes 5 and 7, and also several minor classes as indicated in Table 2. These classes are associated with emissions from biomass burning and Fig. 3 indicates that most of the particles in Classes 5 and 7 are smaller than 0.5 μm. Class 6 contains potassium but no organic fragments. It may also be from biomass burning.

Although the ion signal peaks at 40 Da for class 8 in Fig. 2, suggesting Ca^+, it is likely that this is a mis-calibrated K^+ peak considering that the size and temporal distribution of class 8

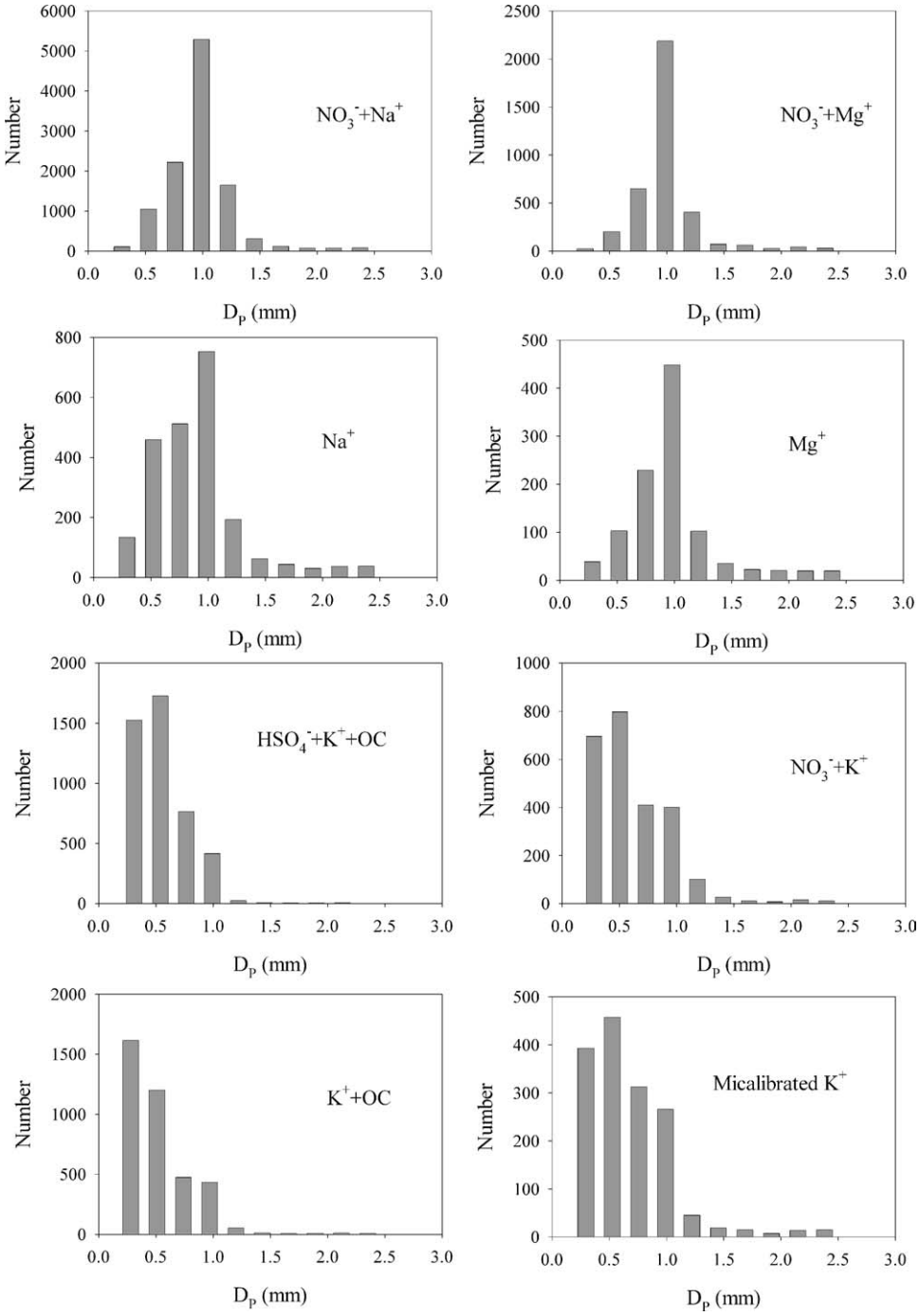

Fig. 3. Size distributions of Classes 1–8 by ART-2a.

is similar to class 6 and that these two ions are separated by only one *m*/*z* unit. In addition, the small peak at 42 Da (mis-calibrated 41 Da) indicates the isotope of K rather than Ca; the Ca isotope should be at 44 Da. Mis-calibrated spectra have also been reported by Pastor el al. [13].

In Fig. 4, it can be found that the temporal profile of Classes 1 and 2 are close, high on the 15th and 16th, suggesting they are from similar directions. The temporal profiles of the $K^+ + OC$ classes, Classes 5 and 7, have the highest number on the morning of the 13th.

4.2. DBSCAN

The clusters found by DBSCAN are presented in Table 3, where the representative object and their member numbers are given. The major clusters include sea salt (Class 1, $NO_3^- + Na^+$; Class 2, $NO_3^- + Mg^+$; Class 5, Na^+; also Classes 7 and 12), potassium (Class 4, $NO_3^- + K^+$), potassium and OC (Class 6, $K^+ + OC$; Class 3, $SO_4^- + K^+ + OC$), EC (Class 17), nickel (Class 22, Ni^+; Class 18, $HSO_4^- + Ni^+$) and vanadium (Class 13).

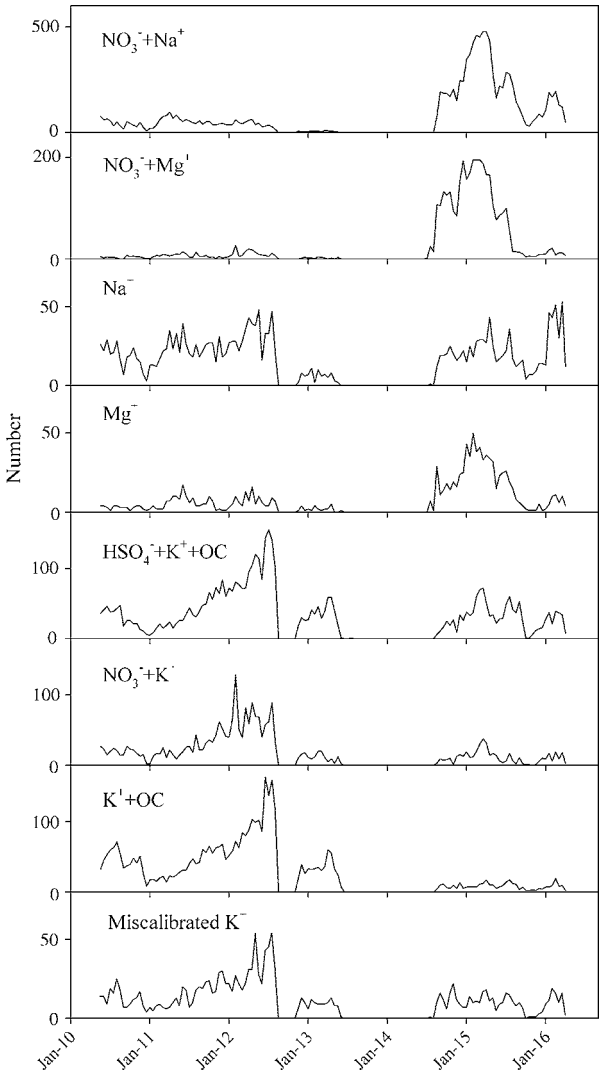

Fig. 4. Temporal profiles of particle number with 1 h resolution for Classes 1–8 by ART-2a.

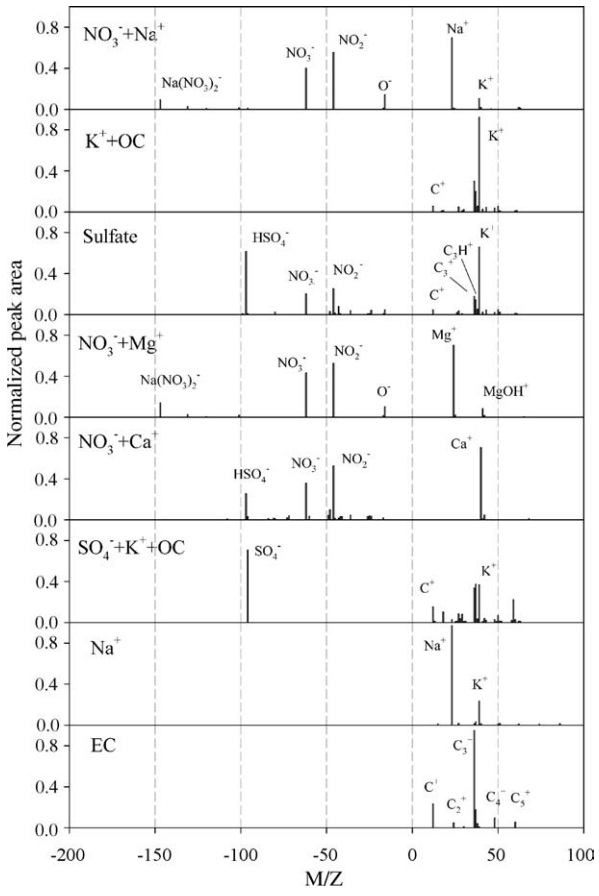

Fig. 5. Central spectra of the major clusters found by DBSCAN.

The centroid spectra of these clusters are shown in Fig. 5. The mass spectra of the sea salt clusters are similar to those found by ART-2a (Classes 1–3 in Table 2). Their sizes concentrate around 1 μm as indicated in Fig. 6. Fig. 7 suggests that most of the particles containing sodium and magnesium with nitrate appear on January 15 and 16, similar to the two classes found by ART-2a.

The potassium class (Class 4 in Table 3) has a wide size range from 0.4 to 1.0 μm and appears during most of the sampling period. Fig. 5 indicates that some sulfate is also present in these particles. When more sulfate is present, a new representative is formed as seen in Table 3.

Classes 4 and 14 are two clusters similar to those found by ART-2a. These are small particles with diameter ranging from 0.3 to 0.5 μm, consistent with emissions from combustion sources. Fig. 7 shows high particle numbers on the 13th for Class 4 but Class 14 does not show this pattern. Class 3 is composed of several different types of particles and the major component in these particles, potassium, organics, sulfate and nitrate, varies significantly. In the ART-2a analysis, this cluster is separated into several clusters according to the differences in their compositions as shown in Table 2. During transport from source to the receptor, some of these particles become internally mixed through condensation and/or coagulation. For each individual particle, the extent of internal mixing is different and thus different particle types become continuous. The continua between sulfate and organics were also found by Middlebrook et al. [5] and was regarded as one of the major difficulties in cluster analysis [5]. The size distribution and temporal behavior is similar to that of Class 5 in Table 2 found by ART-2a.

DBSCAN found two EC clusters that are different than the EC cluster found by ART-2a. The ART-2a EC cluster also contains sulfate, C_3H^+ and K^+. It seems to be a mixture of two kind of particles, EC and $HSO_4^+ + K^+$. Fig. 5 indicates that there is ion signal for carbon cluster ions ranging from C_1^+ to C_5^+, which are considered characteristic ions for EC. As shown in Fig. 6, these are the smallest particles of all the major clusters. The temporal behavior of the EC cluster is similar to Class 4 (K^+ + OC).

The vanadium cluster (Class 13) and two nickel clusters (Classes 18 and 22) are associated with oil combustion and it was found that nickel and vanadium particles are externally mixed. Classes 14 and 15 both have representative object of nickel, suggesting these two clusters are also associated with oil combustion emissions. Since a small VF was used, these small classes were not found by ART-2a and have likely been combined with other classes. One Fe class is found by ART-2a and two Fe classes

Table 3
Clusters found by DBSCAN and their representative object[a]

No.	Representative object and their member numbers without cloud	Without cloud	With cloud	Volume fraction without cloud[b] (<2.5 μm)
1	9710 (NO_3^-, NO_2^-, Na^+); 313 (O^-, Na^+, K^+); 13; 269 (NO_2^-, K^+); 903 (NO_3^-, Na^+)	11208	11616	0.315
2	3395 (NO_3^-, NO_2^-, Mg^+); 24; 228 (HNO_2^-, Mg^+); 12	3659	3847	0.112
3	6702 (HSO_4^-, K^+); 7; 56; 38; 678 (NO_3^-, Na^+, K^+); 169 (C_2^-, K^+); 348 (HSO_4^-, NO_3^-, C^+, C_3^+); 147; 306; 31; 1062 (HSO_4^-, OC); 183; 194; 112; 176; 332; 2059 (NO_3^-, NO_2^-, Mg^+, K^+); 568 (HSO_4^-, NO_2^-, O^-, C_3^+)	13168	13796	0.104
4	1902 (NO_3^-, NO_2^-, K^+); 377 (HSO_4^-, C_3^+, K^+); 99; 60; 162 (NO_3^-, K^+)	2600	3172	0.048
5	1302 (Na^+); 27; 127 (Na^+, K^+); 29; 7; 29; 22	1543	1544	0.035
6	3694 (C_3^+, C_3H^+, K^+); 401 (C_3^+, C_3H^+); 393 (C^+, C_3^+); 262 ($C_2H_3^+$, C_3H^+, $C_3H_2^+$, $C_3H_3^+/K^+$)	4750	4751	0.020
7	334 (PO_4^-, $H(NO_2)_2^-$, Na^+, Na_2Cl^+); 32	366	395	0.017
8	214 (NO_2^-, NaH_2O^+); 22 (HSO_4^-, NO_2^-, NaH_2O^+)	236	393	0.015
9	414 (Mg^+)	414	415	0.012
10	829 (Ca^+); 11(OC, Ca^+)	840	960	0.011
11	300 (NO_2^-, Fe^+)	300	406	0.008
12	348 ($Na(NO_3)_2^-$, NO_2^-, Na^+); 5	353	359	0.007
13	319 (V^+, Fe^+); 96 (V^+, VO^+); 26(V^+)	441	528	0.005
14	941 (SO_4^-, C_3^+, C_3H^+, K^+); 160 ($H_3NiO_3^-$, SO_4^-, Ni^+); 8 (SO_4^-, Na^+)	1109	1110	0.004
15	812 (PO_4^-, C_3^+, K^+); 106 (PO_4^-, Ni^+)	918	1108	0.004
16	211 (NO_3^-, K^+, OC)	211	487	0.004
17	1096 (C_3^+); 54 (C_3^+, C_5^+); 14 (C_3^+, C_4^+)	1164	1165	0.004
18	237 (HSO_4^-, Ni^+); 13 (HSO_4^-, NH_4^+, K^+); 5 (NO_3^-, Ni^+)	255	306	0.004
19	469 (C_2^-, C_3^+); 12 (NO_2^-, C_2^-, Na^+)	481	640	0.003
20	221 (NO_3^-, Na^+); 2	223	287	0.003
21	375 (NO_3^-, C_3^+); 19 (NO_3^-, NO_2^-, C_5^+); 40 (NO_2^-, C_3^+, C_3H^+); 14 (HSO_4^-, C_3^+)	448	449	0.002
22	269 (Ni^+); 13(C_3^+, C_3H^+, $C_3H_2^+$, $C_3H_3^+$, Ni^+)	282	364	0.002
23	330 (NO_3^-, K^+)	330	421	0.002
Sum	–	45299	48519	0.742

[a] The number before each parenthesis is the member number of a representative. A new representative begins after a semicolon.
[b] All the above clusters explain 0.872 of the total volume.

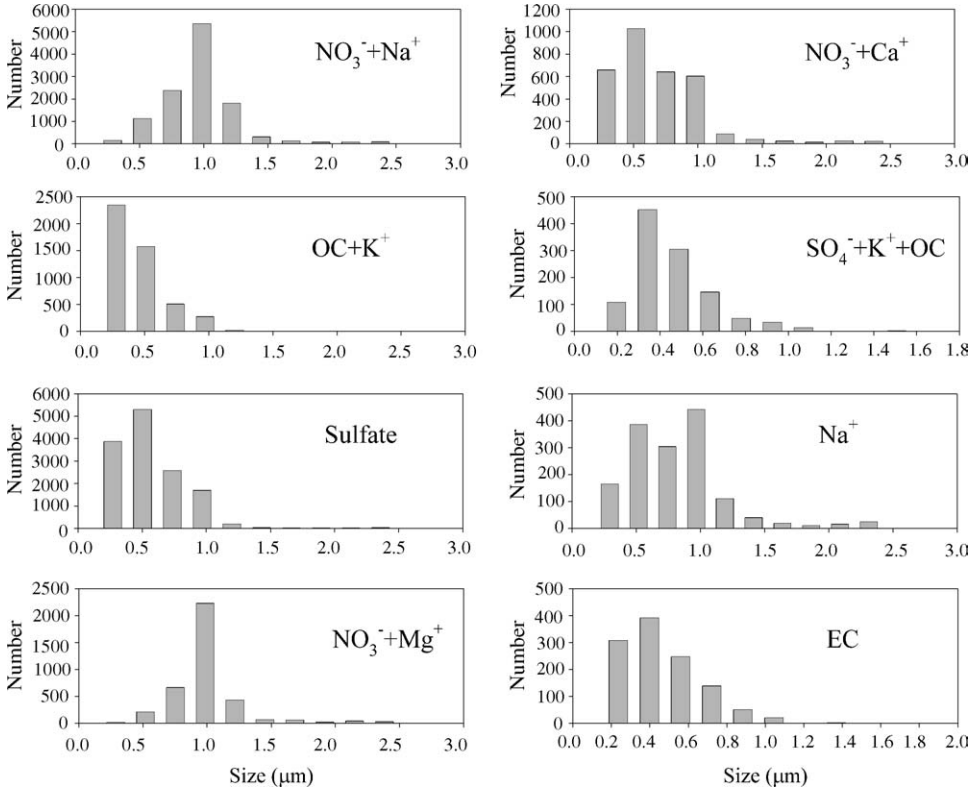

Fig. 6. Size distributions of major DBSCAN clusters.

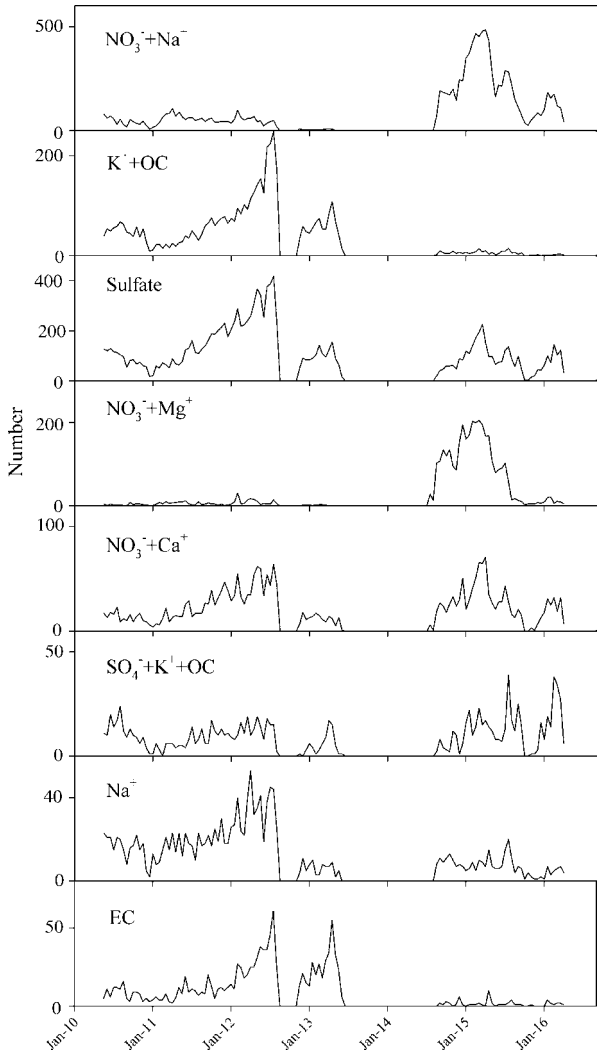

Fig. 7. Temporal profiles of particle number with 1 h resolution for major DBSCAN clusters.

were found by DBSCAN. It appears that the two Fe classes were combined by ART-2 and that the weight vector is dissimilar to both of them.

The negative spectra of some particles were missing, so even if the positive spectra of two particles were identical, they were still classified into different clusters. In Table 3, some clusters without negative signal can be combined into those with complete spectra manually, such as Class1/Class 5, Class 2/Class 9 and Class 17/Class 19.

The two methods found similar small classes such as $PO_4^- + K^+ + OC$ (Class 34 in Table 2 and Class 15 in Table 3) and some minor classes were found only by one method. Since a small vigilance factor is used for ART-2a, the minor classes in Table 2 may not be as reliable as those in Table 3.

The daily bulk phase concentrations of $PM_{2.5}$ and some chemical species are listed in Table 4. The $PM_{2.5}$ concentration is highest on January 12 and lowest on January 16. The other major species and metals, except Na, also show this trend. Together, these data suggest that anthropogenic sources dominated PM measured on January 12 and that air parcels from the clean areas dominated on the 15th. The time series of the $K^+ + OC$ clusters and sea salt clusters are consistent with the aforementioned conclusion from the bulk phase measurements. The size dependence of the particle detection efficiency for the ATOFMS has been discussed by Allen et al. [19]. For particles from 0.32 to 1.8 μm in diameter, the detection efficiency was highest for the largest particles and declined by approximately 2 orders of magnitude for the smallest particles. Since the particles dominating January 12 were small and had low detection efficiencies, the actual particle number in the atmosphere on January 12 was the highest.

There is good agreement between the two methods on the size and temporal distributions of different particle types for the major classes, such as the sodium classes, magnesium classes and $K^+ + OC$ classes.

Two back trajectories (January 12, 2004, 10:00 a.m. and January 15, 2004, 3:00 a.m.) originating from the receptor site were computed by HYSPLIT [20,21] and are presented in Fig. 8. On the morning of January 12, the air mass traveled through a heavy residential area in New Jersey before reaching the receptor site and the high $K^+ + OC$ concentrations measured during this time were likely due to wood burning for heat. On the early morning of January 15, the trajectory was from relatively clean areas in Canada and New York State, consistent with the low $PM_{2.5}$ concentration around that time. After the sudden change of direction at the Connecticut coast, sea salt particles were collected and transported to the receptor site.

5. Discussion

When using ART-2a for classification, it is assumed that the same kind of particles are similar and their mass spectra are concentrated within a small space whose size is determined by the vigilance factor. However, this assumption may not be true for all the particle types. For example, when the organics are ablated, various fragment ions are produced such that their mass spectra may occupy a large space, and two organic spectra may not be similar. The ion spectra generated from the ablation are confined in a certain space, and when there are a large number of particles measured, the density inside the space will be significantly higher than outside. Because of this discontinuity in density, DBSCAN is able to identify the aforementioned space no matter what size and shape it has. Based on the classification results in this study, the assumption of DBSCAN that the spectra are located in a certain space with high density seems more realistic.

It can be found from Table 3 that some clusters are large and some are small, based on the number of representative object. Thus, using a constant VF is not effective to correctly classify all of the clusters. For the classes with continua, the concepts of cluster may not be applied since the transition between the different particle types is gradual. The representative object found within the cluster with continua can identify the different particle types for further physical interpretation.

Table 4
STN/ASRC (Speciated Trend Network/Atmospheric Science Research Center, State University of New York at Albany) measurement of bulk phase concentrations of $PM_{2.5}$ and some chemical species[a]

Date	January 10, 2004	January 11, 2004	January 12, 2004	January 13, 2004	January 14, 2004	January 16, 2004
$PM_{2.5}$	8.3	18.0	31.7	15.2	10.0	6.7
EC	0.530	−0.017	0.605	0.817	0.083	0.337
OC	1.40	0.527	4.26	2.02	−0.188	0.800
NH_4^+	0.952	2.112	4.226	2.213	0.894	0.651
K^+	0.050	0.054	0.100	0.051	0.026	0.000
Na^+	0.086	0.120	0.151	0.094	0.258	0.176
NO_3^-	1.560	3.932	7.190	3.128	1.879	1.002
SO_4^{2-}	1.682	2.831	5.472	3.843	1.685	1.649
Ca	0.0458	0.0456	0.123	0.064	0.040	0.0148
Cr	0.00099	0.00089	0.00275	0.00123	0.00046	0.00048
Cu	0.003	0.0035	0.0094	0.0040	0.0021	−0.0007
Fe	0.048	0.064	0.19	0.10	0.040	0.027
Pb	0.0022	0.0057	0.0212	0.0079	0.0034	0.0012
Mn	0.00069	0.00117	0.0019	0.0011	−0.0000	−0.0002
Ni	0.0321	0.0486	0.0947	0.058	0.025	0.037
K	0.0376	0.0445	0.11	0.045	0.032	0.026
Si	0.041	0.066	0.12	0.088	0.029	0.083
Ti	0.000178	0.00025	0.0063	0.0058	0.0011	−0.00063
V	0.01	0.01	0.0247	0.0147	0.0079	0.0064

[a] The units of all reported concentrations are $\mu g/m^3$.

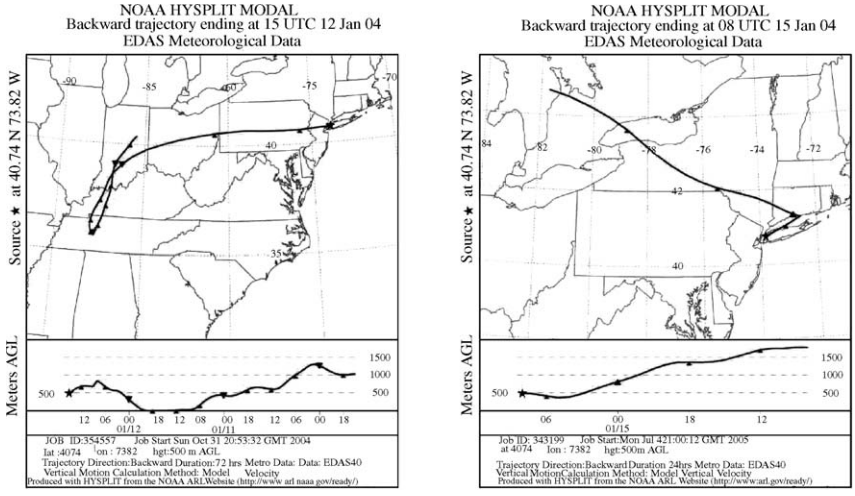

Fig. 8. Back trajectories computed by HYSPLIT on January 12th (left) and 15th (right), 2004.

6. Conclusion

ART-2a and DBSCAN have been used for the classification of single particle mass spectra measured at the New York supersite. When using high vigilance factors, too many similar clusters were formed and there were areas of overlap between them. This problem can be partially solved by using low vigilance factors. However, for the clusters found by ART-2a with low VFs, more representative object than the center may be needed, since the objects in each cluster are not so similar.

For the first time, the cluster structure of single particle mass spectra is investigated by a density-based method. The representative object of the clusters found by DBSCAN were chosen by the Kennard and Stone algorithm for better understanding each cluster and also for measuring the sizes of the clusters. All the ART-2a clusters are spheres with the same sizes (determined by VF). The shapes and sizes of DBSCAN clusters are various. Some are small such as Classes 11and 16 in Table 3 and they are close to spheres. Some are large and may have several branches, such as Class 3 whose branches extend along HSO_4^+, K^+ and C^+/C_2^+ directions.

Similar sea salt clusters (sodium nitrate and magnesium nitrate) were found by the two methods. Among sulfate, potassium and OC particles, there are serious continua caused by internal mixing and a large cluster was found by DBSCAN. The clusters found by ART-2a, with different content of sulfate, potassium and OC, need to be combined.

The classification results indicate that a majority of the particles detected on January 12 were from anthropogenic combustion sources and that sea salt particles dominated on January 15 and 16.

Acknowledgments

This work was supported in part by the U.S. Environmental Protection Agency (EPA) cooperative agreement No. R828060010 and by EPA Science to Achieve Results Program through a subcontract from the University of Rochester PM and Health Center Grant R827354. Special thanks go to Kenneth Demerjian, James Schwab and others from SUNY Albany for being great hosts. The acquisition of equipment used in this study was supported by the New York State Office of Science, Technology, and Academic Research (NYSTAR). The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model and/or READY website (http://www.arl.noaa.gov/ready.html) used in this publication.

References

[1] J.H. Seinfeld, S.N. Pandis, Atmospheric Chemistry and Physics, John Wiley & Sons, New York, 1998 (Chapters 21 and 22).
[2] L. van Bree, F.R. Cassee, A Critical Review of Potentially Causative PM Properties and Mechanisms Associated with Health Effects, National Institute of Public Health and the Environment (RIVM) Research Report, Report No. 650010015, P.O. Box 1, 3720 BA, Bilthoven, the Netherlands, 2000.
[3] C.A. Pope, Aerosol Sci. Tech. 32 (2000) 4.
[4] P.H. McMurry, Atmos. Environ. 34 (2000) 1959.
[5] A.M. Middlebrook, D.M. Murphy, S.-H. Lee, D.S. Thomson, K.A. Prather, R.J. Wenzel, D.-Y. Liu, D.J. Phares, K.P. Rhoads, A.S. Wexler, M.V. Johnston, J.L. Jimenez, J.T. Jayne, D.R. Worsnop, I. Yourshaw, J.H. Senfeld, R.C. Flagan, J. Geophys. Res. 108 (D7) (2003) 8424, doi:10.1029/2001JD000660.
[6] X.-H. Song, P.K. Hopke, D.P. Fergenson, K.A. Prather, Anal. Chem. 71 (1999) 860.
[7] D.-Y. Liu, R.J. Wenzel, K.A. Prather, J. Geophys. Res. 108 (D7) (2003) 8426, doi:10.1029/2001JD001562.
[8] D.M. Murphy, A.M. Middlebrook, M. Warshawsky, Aerosol Sci. Tech. 37 (2003) 382.
[9] M. Daszykowski, B. Walczak, D.L. Massart, Chemom. Intell. Lab. Syst. 56 (2001) 83.
[10] M. Daszykowski, B. Walczak, D.L. Massart, Anal. Chim. Acta 468 (2002) 91.
[11] K.A. Prather, T. Nordmeyer, K. Salt, Anal. Chem. 66 (1994) 1403.
[12] E. Gard, J.E. Mayer, B.D. Morrical, T. Dienes, D.P. Fergenson, K.A. Prather, Anal. Chem. 69 (1997) 4083–4091.
[13] S.H. Pastor, J.O. Allen, L.S. Hughes, P. Bhave, G.R. Cass, K.A. Prather, Atmos. Environ. 37 (2003) S239.
[14] G.A. Carpenter, S. Grossberg, D.B. Rosen, Neural Netw. 4 (1991) 493.
[15] P.K. Hopke, X.-H. Song, Anal. Chim. Acta 348 (1997) 375.
[16] M. Ester, H. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, 1996, p. 226 (available from www.dbs.informatik.uni-muenchen.de/cgi-bin/papers?query=-CO).
[17] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137.
[18] E.E. Gard, M.J. Kleeman, D.S. Gross, L.S. Hughes, J.O. Allen, B.D. Morrical, D.P. Fergenson, T. Dienes, M.E. Galli, R.J. Johnson, G.R. Cass, K.A. Prather, Science 279 (1998) 1184.
[19] J.O. Allen, D.P. Fergenson, E.E. Gard, L.S. Hughesm, B.D. Morrical, M.J. Kleeman, D.S. Gross, M.E. Galli, K.A. Prather, G.R. Cass, Environ. Sci. Technol. 34 (2000) 211.
[20] R.R. Draxler, G.D. Rolph, HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (http://www.arl.noaa.gov/ready/hysplit4.html), NOAA Air Resources Laboratory, Silver Spring, MD, 2003.
[21] G.D. Rolph, Real-Time Environmental Applications and Display sYstem (READY) Website (http://www.arl.noaa.gov/ready/hysplit4.html), NOAA Air Resources Laboratory, Silver Spring, MD, 2003.