Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 258 (2006) 58–73
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

SpectraMiner, an interactive data mining and visualization software for single particle mass spectroscopy: A laboratory test case

Alla Zelenyuk [a,*], Dan Imre [b], Yong Cai [a], Klaus Mueller [c], Yiping Han [c], Peter Imrich [c]

[a] *Pacific Northwest National Laboratory, Richland, WA 99354, USA*
[b] *Imre Consulting, Richland, WA 99352, USA*
[c] *State University of New York at Stony Brook, Stony Brook, NY 11794, USA*

Received 20 April 2006; received in revised form 19 June 2006; accepted 21 June 2006
Available online 26 July 2006

Abstract

Single particle mass spectrometers are sophisticated instruments designed to measure the sizes and compositions of a wide range of individual particles in situ, in real-time. They characterize hundreds of thousands or millions of particles, generating vast amounts of rich and complex data, the proper mining of which requires dedicated state of the art tools. The analysis of individual particle mass spectra is particularly difficult because of their high dimensionality—each data point, representing a single particle, includes the 450 mass spectral peak intensities, particle size, and time of detection. The first step is to organize the data; a process typically accomplished by grouping particles of similar attributes. Since the common assumption is that the data should be reduced to become manageable, they are typically classified into a small number of clusters (∼10), each of which is represented by an average/representative spectrum. Our approach is quite different. We have developed a data mining and visualization software package we call SpectraMiner that makes it possible to handle hundreds of clusters, limiting loss of information and thus overcoming the boundaries set by traditional statistical data analysis approaches. Data, which often include over 1 million particle spectra, are organized using K-mean clustering algorithm. The clusters are merged into nodes by sequentially combining similar clusters. The final structure is displayed in a hierarchical dynamical tree or circular dendogram. This interactive dendogram is the visual interface that allows for real-time data exploration and mining. Clicking on any of the clusters/nodes in the dendogram reveals the detailed information about the particles that reside at that position. At each step the scientist is in control of the level of detail and the visualization format, rapidly switching between them while running the program on a PC.

Here we present a study that puts the classification aspect of SpectraMiner to the test. Twelve types of laboratory generated particles are carefully chosen to test some of the difficult aspects of single particle mass spectroscopy. We quantify the degree of particle identification and separation at a number of levels and demonstrate how the visualization tools that SpectraMiner provides can be used to refine, steer and control the data mining process.

Keywords: Single particle mass spectrometer; Data classification; Data visualization

1. Introduction

Single particle mass spectrometers (SPMSs) are presently widely used to provide real-time, in situ information on the sizes and compositions of individual aerosol particles. The path from instrument design and construction to data acquisition and analysis is long and demanding. The goal is to use SPMSs to generate high quality, reproducible, easy to assign individual particle mass spectra (IPMS). In reality IPMS that are generated by laser ablation tend to exhibit very large particle-to-particle variations, making the data mining process a daunting task. The steady drive to improve the instrumental aspects of SPMSs represents great challenges and remains at the center of a significant research and development effort in the field. It is important to realize that the immensity and complexity of the rich data that are produced by these sophisticated instruments requires comparable, dedicated state of the art analytical tools that afford the user the opportunity to extract as much knowledge as the data can offer. The focus of this paper is on the approach we have developed to analyze the vast amounts

* Corresponding author. Tel.: +1 5093767696.
E-mail address: alla.zelenyuk@pnl.gov (A. Zelenyuk).

1387-3806/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2006.06.015

of complex and highly detailed data that are generated by SPMSs.

Analysis of IPMS, even if they are nearly reproducible, is difficult because of their high dimensionality—each data point, representing a single particle, comprises a long vector of attributes, which in the present applications include: the 450 mass spectral peak intensities, particle size, and time of detection. In addition the analysis of IPMS requires means to account for the large particle-to-particle variations, which are an integral part of experimental data in general. In the case of IPMS it is important to keep in mind that in laser ablation variability is often very high and it takes on a number of forms. The simplest involves an apparently consistent fragmentation pattern but with large particle to particle variations in overall and/or in relative peak intensities. Another common finding is that particles of a given composition produce a number of different well-defined fragmentation patterns. Which pattern dominates may correlate, to some degree, with particle size. And the cases that are the most difficult to analyze are those in which particles of different compositions produce nearly identical mass spectra. Under most applications of SPMSs the compositions of the individual particles exhibit a very wide range and they vary significantly from one particle to the next. The probabilities of generating ions from the different compounds in atmospheric particles can vary by orders of magnitudes, worse yet, ion production varies not only from one component to the next, but also on the basis of what other compounds happen to be present in the same particle. This partial list is intended to illustrate some of the complexity inherent in the mining IPMS datasets.

The first step of mining datasets containing hundreds of thousands of individual particle data points is to organize the datasets into groups of particles of similar composition. This process of classification is typically accomplished by grouping particles with similar mass spectral attributes together using statistical methods to define, compare and finally partition the IPMS into a number of clusters. Several clustering algorithms have been applied to the analysis of data produced by SPMSs [1–5]. Three of these methods treat the IPMS as multidimensional vectors and calculate their proximity in *N*-D space, which can be expressed by Euclidean distances, dot products, correlation coefficients, Mahanalobis distances, and other metrics. Two research groups [1,4] are currently using an artificial neural network, ART-2A to organize the IPMS into classes. This algorithm groups particles according to the dot products of the normalized particle vectors. Murphy et al. [2] utilize a hierarchical clustering analysis to classify data acquired with their SMPS. They also use the dot products of normalized individual particle vectors as a measure of spectra similarity and compare it to a set threshold. This process yields typically a large number of clusters, which are then hierarchically combined until stopping conditions are met. What is interesting is that in this approach there is an option for expert knowledge input to play a role in determining the final outcome of the classification. To account for the complex internal mixtures of individual atmospheric aerosol particles Trimbom et al. [3] use a fuzzy classification algorithm, in which individual particles can belong to more than one class, with varying degree of membership. Another method—an algorithm for discriminant analysis of mass spectra (ADAMS) [5] classifies aerosol mass spectra into predetermined classes and groups rare particles into an outlier class. Each of the defined classes is characterized by discriminant chemical markers that are assigned to it on the basis of prior mass spectral knowledge.

While the methods, metrics and threshold distances used as criteria for particles to belong to the same class can be different for the different data classification methods, a common assumption for all of them is that the data must be organized into a manageable number of classes, where manageable refers to a number that can be easily comprehended and handled by the scientist. In most cases this means that the data are classified into a small number of clusters (∼10), each of which is then represented by an average/representative spectrum. The results are typically presented in pie charts, bar graphs, or 2D and 3D plots of time and/or, size and class [1,6,7]. It is important to keep in mind that once the data are organized and reduced there is no convenient path back to the original data or to a higher level of details.

Our aim is to develop a software package that provides the user with the option and tools to push data mining to the point where the limits are determined by the data and not by the mining process. Because the information content in these datasets is often overwhelming, we provide the user ease and flexibility to choose at any stage the depth and details of the data mining process. In this study we show that even the very limited dataset, containing only 12 particle types, cannot be properly represented by a very small number of statistical classes.

At the root of our approach [8] is a requirement that the statistical classification algorithm is used to order the data for mining with a minimal loss of information. Thus, we classify our data, which often include over 1 million IPMS, into hundreds of groups, which we call clusters. The clusters are merged into nodes by sequentially combining similar clusters and the nodes are further merged into larger nodes. The final structure is displayed in a hierarchical dynamical tree portrayed in a space efficient polar format, or circular dendogram. Most importantly, the dendogram serves as the visual interface that allows the user to navigate through the complexities of the individual particle mass spectral datasets with ease, taking advantage of the fact that the human brain is capable of comprehending visual information that is orders of magnitude more complex than text, speech, or tabulated numbers. Hence, we have overcome the need to reduce the data to ∼10 clusters by providing novel visualization and data mining tools. This task is accomplished with the data mining and visualization software package we call SpectraMiner. With SpectraMiner the user can explore the data on any level: from nodes that include hundreds of thousands of particles down to the individual particle level with speed and ease, never having to permanently disregard even the smallest fraction of details or data.

Here, we present a study that illustrates some aspects of the data organization and identification part of SpectraMiner. Although this study yields a detailed quantification of the particle identification process, our goal is not to tabulate the number of particles that are correctly or incorrectly classified. Instead, we use this study to identify some of the common difficulties of analyzing IPMS and describe our approach to curtail information

loss and illustrate specific solutions. To this end we generated under controlled settings IPMS of 12 types of particles, whose composition is specifically chosen to illustrate the problems we face in atmospheric science, and demonstrated the contribution that SpectraMiner makes to resolve them.

2. Experimental

2.1. Particle types and particle generation

The 12 types of particles used in this study are listed in Table 1 with their corresponding abbreviations. These compounds were chosen to represent a very small sample of the types of particle compositions we routinely encounter in the atmosphere. They were selected in a manner that reasonably reproduces some of the fundamental categories of complexities encountered in single particle mass spectroscopy.

The two Na-containing particle types test our ability to differentiate these very different particle types whose mass spectra are dominated by the Na^+ ion peak. Similarly, distinguishing between the pure ammonium nitrate particles and the internally mixed particles that are composed of ammonium nitrate and lauric acid is difficult since the IPMS for both of these particle types are dominated by the NO^+ peak. Ammonium sulfate and ammonium nitrate are some of the most common compounds found in tropospheric aerosols. Yet in similar studies by other researchers [2,4] it was found that these two particle types are difficult to differentiate with SPMSs. It was important for us to test our ability to detect ammonium sulfate and properly classify these two particle types. Our experience sampling atmospheric particles revealed that a large fraction of them are composed of sulfates or nitrates, which are internally mixed with organics. The AN/LA and the three mixtures of AS with SA provide an opportunity to test the performance of the instrument and the software on these types of internally mixed particles. We included three very different types of organic particles in this study: SA is a small dicarboxylic acid of four carbon atoms, LA is a 12 carbon long-chain fatty acid and PY is a four-ring polyaromatic-hydrocarbon (PAH). The soot data is a small subset of our diesel exhaust emissions characterization research project and was included in this study because soot is commonly found in the atmosphere, where it plays an important climatic and public health role. Moreover, for the present project it was important to include soot since during laser ablation a fraction of organic particles can fragment to the point at which they are difficult to distinguish from soot.

Table 1
Listing of the 12 particle compositions and their abbreviations

Abbreviation	Composition[a]
SC	Sodium chloride
SN	Sodium nitrate
AN	Ammonium nitrate
AN/LA	Ammonium nitrate/lauric acid (1:1)
LA	Lauric acid
PY	Pyrene
AS	Ammonium sulfate
0.8AS	Ammonium sulfate/succinic acid (4:1)
AS/SA	Ammonium sulfate/succinic acid (1:1)
0.8SA	Ammonium sulfate/succinic acid (1:4)
SA	Succinic acid
ST	Soot

[a] The compositions of all binary mixtures refer to the weight fraction ratios of the corresponding components.

All particle types, except soot, were generated by aerosolizing them from solutions using an atomizer (TSI Inc., Model 3076). Aerosol flow was first dried to remove solvent by two diffusion dryers (TSI Inc., Model 3062), connected in series, and then diluted and further dried by mixing with dry compressed air at a ~50:1 ratio in a large volume mixing/drying chamber.

Soot particles were sampled from a Mercedes 1.7L A-Class diesel engine during the deployment of our single particle laser ablation time-of-flight mass spectrometer (SPLAT) at the National Transportation Research Center at the Oak Ridge National Laboratory. The raw exhaust was dried and diluted by a factor of 2500 by mixing with dry air. For the present study we have chosen exhaust particles sampled under engine operating conditions when 96% of the particulate emission was determined to be composed of pure soot.

The entire dataset presented here contains 36,000 IPMS and was constructed by combining 3000 individual particle mass spectra from each of the 12 particle types.

2.2. Individual particle mass spectra

Polydisperse aerosol particles were sampled by SPLAT, a detailed description of which is given in [9]; here we give a brief description only. Particles enter the instrument through a 100 μm orifice into an aerodynamic lens inlet. The lens is used to focus entrained particles into a narrow, low divergence particle beam and transmit the particles into the vacuum chamber with high efficiency. Two stages of optical detection placed along the well-defined particle path provide aerodynamic velocity and size information for the individual particles in the particle beam. Pulse from an excimer laser, operated at 193 nm, is timed to arrive coincident with the particle at the ionization region of the time-of-flight-mass spectrometer (TOF-MS) and generates ions by ablation. IPMS are subsequently acquired by measuring the ions time-of-flights in the reflectron TOF-MS.

The signal from the TOF-MS microchannel plates is digitized at a rate of 50 MHz using 8-bit A/D card for PCI bus (Gage Applied Technologies, Inc., Model CompuScope 8500). The digitized IPMS are paired with the corresponding particle size information, bundled to form files, each containing five IPMS to match Windows allocation unit size, compressed on the fly and written to the hard drive. At present the sampling rate of ~20 particles per second is limited by the rate with which the data could be transferred and written to the computer hard drive.

To prepare the IPMS for classification the raw data are first processed and reduced: spectra are decompressed, a baseline is subtracted, ions' time-of-flights are converted to the corresponding mass-to-charge (m/z) ratios, and the integrated area of each of the IPMS and the areas under each of the peaks are calculated by integrating the intensities within 0.5 Da of each of the 450 m/z values. "Hits" are separated from "misses" on the basis of a

comparison between the total integrated mass spectral intensity and a preset threshold, which is set to be slightly larger than the average integrated area of the particle-free background mass spectrum. At the conclusion of this process two files are generated: one contains the measured aerodynamic diameter and the time of detection for each of the detected particles and the second file lists for each particle that was classified as a "hit" the aerodynamic diameter, time of detection, total integrated mass spectrum area and the 450 mass spectral peak intensities, one for each of the *m*/*z* values.

2.3. Data classification

To assure fast response by SpectraMiner to the user commands during the data mining process we have taken a two-tier approach: an off-line data clustering process and a real-time data mining and visualization.

We first run an off-line process based on K-means clustering that organizes the data into hundreds or a few thousands of representative groups we call clusters, by combining together particles with very similar mass spectra. We treat each of the IPMS as vectors in 450-D space and group them into clusters based on their proximity in the 450-D space. The first, randomly selected particle mass spectrum serves as the first seed and the distance between each of the subsequently picked mass spectra and that seed is calculated. If the calculated distance is less than a threshold distance, the particle is placed into that cluster, if the distance is greater than the threshold; that mass spectrum is set to be an additional seed. Each of the IPMS is added to the cluster that it is nearest to. Once the entire data has been clustered, the average mass spectrum for each of the clusters is calculated and the distance between each of the IPMS and these average mass spectra are recalculated and compared and particles are placed in the clusters they are nearest to. This iterative process is designed to take into account the fact that the cluster center shifts during the classification and drive the process to convergence.

As a measure of the distance between mass spectra we use the value of $(1 - r)$, where r is the Pearson correlation coefficient. In this study the IPMS are not normalized, but the distance between two proportional mass spectra is very small and therefore they are always grouped together.

At the end of this off-line procedure an output file with the calculated statistical properties of each cluster, like the cluster mean mass spectrum and the covariance matrix, is created. In addition the information about each particle size, time of detection and the mass spectral peak intensities are included. The final listing of "particles" and the clusters they belong to is later utilized by the interactive data mining and visualization program described in the sections below.

The dataset presented in this study consists of 36,000 IPMS–3000 IPMS of each particle type. Classification of this dataset, i.e., the off-line portion of the data analysis, takes ~2 min on an office PC. For a distance threshold of 0.3 the classification produces a total of 583 clusters, 62 of which contain more than 20 particles each, accounting for 98% of all the particles. It is important to note that although our analysis from this step forward will be focused on the 62 clusters, the information on the remaining 2% of particles will not be lost and can be instantaneously visualized. Even though this feature of the software is not very important for the present study, it is crucial for measurements that depend on the ability to identify and track a few "gold nuggets".

K-mean clustering is only one out of a number of available classification methods that can be applied to our data. We have tested a number of them on our SPLAT data and have come to the conclusion that as long as the IPMS are organized in a reasonable manner the classification algorithm plays only a minor role. The critical point is the ability to visualize and explore the data in an efficient manner.

2.4. The interactive dendogram: design and construction

Dendograms are an effective and established way to convey results of a hierarchical clustering or classification process. Here we describe an advanced incarnation of the dendogram paradigm, which combines a more space-efficient polar or circular layout with a number of interactive features to facilitate the exploration of large data hierarchies. Hence, we call our approach the interactive dendogram. In our polar dendogram layout we have chosen the equispaced arrangement of clusters on the outer-most circle. This is the most appropriate placement since each cluster stands for a partition of the original (large, preprocessed) dataset, and all partitions (or clusters) have been chosen in such a way that their similarities are equal or smaller to a preset distance or similarity threshold, max Sim. As the requirement on the similarity between IPMS is sequentially relaxed, similar clusters are merged together to form larger groups, we call nodes, until eventually all IPMS in the dataset are merged together, forming the root node. In our polar dendogram layout the radius R of a concentric circle onto which a node with similarity node Sim is placed is given by

$$R = \left(\frac{\min \mathrm{Sim} - \mathrm{node\,Sim}}{\min \mathrm{Sim} - \max \mathrm{Sim}} \right) \max \mathrm{Rad} \quad (1)$$

max Sim is the partitioning threshold of the K-means preprocessing algorithm (or some other metric), min Sim is the similarity of the root node, and max Rad is the radius of the outer-most circle.

In this configuration, however, when the number of nodes in a sub-tree is large, or the tree is highly imbalanced (as is the case in our application), sub-tree edges that connect a cluster or node on circle R_1 to a node on circle R_2, which is smaller than R_1, may pass through the circle with R_3, which is smaller than R_2, which could lead to a cluttered display. To avoid this we use curved arcs instead of edges to connect the tree branches. In addition edges and curved arcs on the dendogram are colored using a rainbow colormap to indicate the number, or percentage of particles they carry.

The clusters are placed along the circumference of the dendogram in such a way that no crossings between branches occur. No other rules apply. It is for that reason that there is no unique solution to the dendogram structure. But, since its construction follows a specific and reproducible order, the end results tend to be similar.

Often, the classification of mass spectral data produces many clusters which contain only few particles. To avoid the visual clutter that they can create in SpectraMiner the user interactively selects the minimum number of particles that clusters need to have in order to be displayed. Modifying this threshold redraws the dendogram with the full rainbow color spectrum mapped to the active population interval. The user has the ability to expand the node into its full sub-tree and expand or collapse the sub-tree.

From the perspective of data exploration and mining, setting the dendogram to suit user preferences is only the beginning. The most important features of this software are ease, speed, and versatility with which data exploration can be performed. Some of these features will be presented in the sections below describing the data mining process in this paper. Other features, specifically designed to be used with ambient data that are acquired as a function of time and make it possible to search for correlations between particle types and their relationships to other chronological observation, will be demonstrated in the another publication.

3. Results and discussion

3.1. Setting-up the present dendogram

Fig. 1 shows one of SpectraMiner's annotated visual displays. At the center is the hierarchical tree, in which the classified and organized data are displayed. It is generated by loading the output of the off-line data clustering process. What makes this tree different from typical hierarchical trees is that it is transformed into a polar format, called the circular dendogram. In the table in the top left corner the scientist can select the time period to be covered and the particles size range to be displayed. In the present dendogram we include particles with sizes from 50 nm to 3 μm. Other dendogram display parameters are defined in the tree type table below. The "# of levels" defines the number of distinct concentric levels on the dendogram, on which nodes are located. The radius is a true measure of similarity: the further from the center a node is the more similar the particles within it are. Hence, moving from the circumference towards the center amounts to reducing the similarity requirements for particles to be grouped together. The # of levels can therefore be viewed as defining the resolution scale, with which the tree branching are displayed. Here we have chosen to divide the dendogram radius into 30 resolution elements and display them on logarithmic scale, to accommodate the fact that many of the clusters in this dataset merge together at high similarity levels. "Flat" defines the overall wrap angle of the dendogram layout and "empty" sets the angle of the white wedge, whose role is to help define beginning and end clusters. "Min # particles", in this case set to 20, defines the minimum size of visible clusters or nodes to be represented in the dendogram.

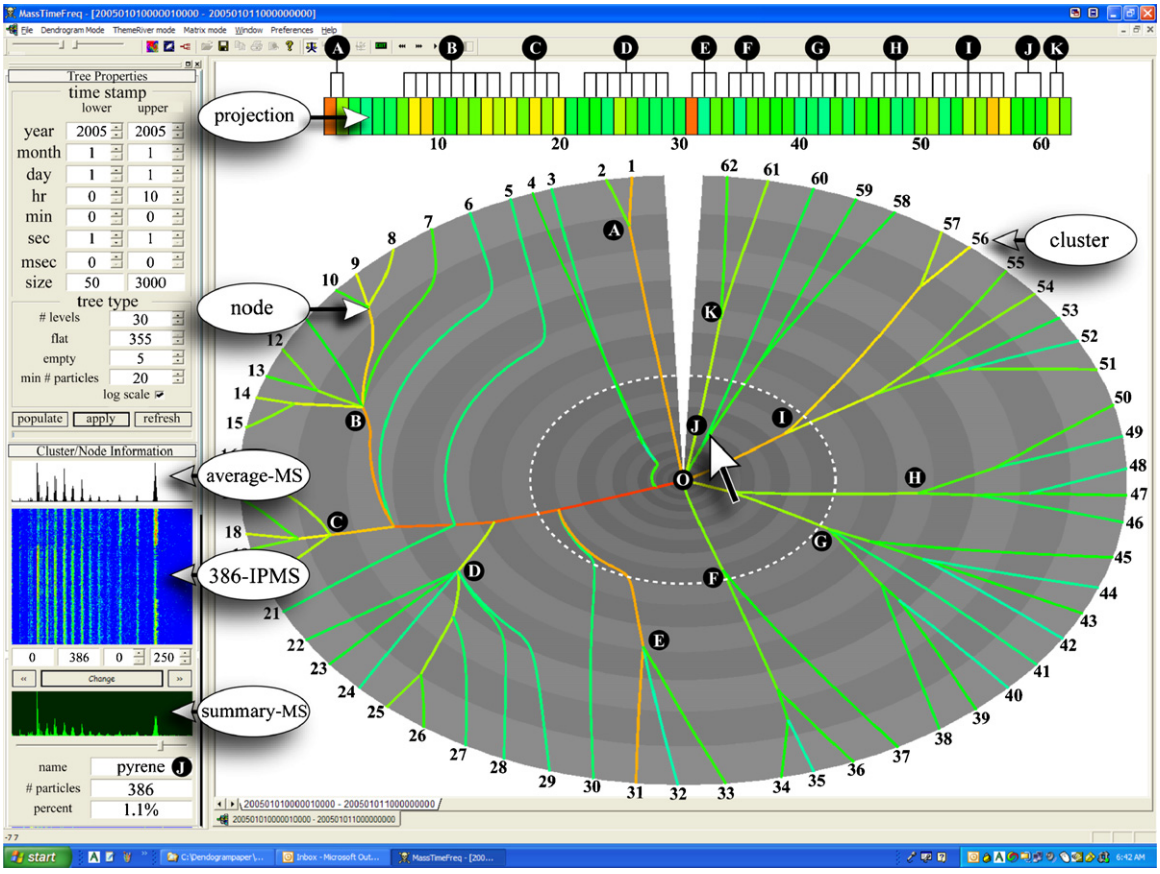

Fig. 1. An annotated screen capture of a SpectraMiner visual display. At the center is the dendogram and above it is a linear projection of clusters populations. On the top left is a table describing the dendogram input parameters and below is a graphic display of the mass spectral information of node J.

At the circumference of the dendogram in Fig. 1 we find 62 numbered clusters, each of which contains more than 20 particles. Each of the edges and arcs on the dendogram, "tree branches", are colored using a rainbow colormap to indicate the number, or percentage of particles they carry, with red being high, blue being low and black (not present in Fig. 1) corresponding to zero fraction of the particle population. Here we use a log(log(population)) color scale, resulting in a display that makes it easy to recognize the presence of even sparsely populated clusters.

The linear strip above the dendogram is the projection of clusters' populations. Each of the small colored rectangles represents, with identical color scale, the population of the corresponding cluster at the circumference of the dendogram. For this presentation we have labeled the projections of the clusters' populations by the corresponding cluster number to make the connection with the dendogram straightforward. The structure above the projection was also added specifically for this study and is intended to illustrate which clusters are merged to form the nodes that are alphabetically labeled in the dendogram in Fig. 1. Like the dendogram, the projection is interactively explorable.

The merged clusters form nodes, similar nodes merge together to form larger nodes, and the colored lines mark the merging pathways, with their colors representing particle population they carry. Clusters with similar particles merge rapidly and clusters containing very different particles do not merge until the center of the circle, root node, is reached.

An all encompassing view of the dendogram in Fig. 1 reveals 11 main branches. The nodes at which we chose to define the important bifurcation points in each of the 11 branches are alphabetically labeled. These nodes were chosen on the basis of a brief exploration and personal experience and preference, but can easily be changed to adapt to the questions at hand.

3.2. Exploring the dataset

3.2.1. Connecting dendogram points with particle mass spectra

Our interactive dendogram is tightly coupled with what we call the cluster/node content browser. It creates for the researcher the connection between the clusters/nodes, in which particles reside, and their mass spectra and hence the particle compositions. We begin data mining by examining the 11 key nodes labeled alphabetically in Fig. 1. It is important to note that the 11 nodes, which we chose as examples in this study, represent points on the classification tree that have a wide range of radii, or similarity parameters. In other words the visual, interactive, hierarchical classification tree guides the researcher in making choices and provides the possibility to make them with a flexible similarity requirement for each of the tree branches. If we had instead attempted to force the entire dataset into less than 15 fixed clusters with a single similarity threshold we would need to set that threshold at the level shown in Fig. 1 by the dashed white ellipse. Instead we have the flexibility to explore nodes that appear to be important branching points and have significant particle population despite the fact that they happen to be positioned at different radii.

To browse the content of a cluster/node one needs simply click on the corresponding point on the dendogram and the program generates the graphic and tabular information shown in the lower left side of Fig. 1. Here we illustrate the outcome of clicking the mouse on the node labeled J. Under the heading "Cluster/Node Information" the program displays the average mass spectrum of all the particles in this node. The J node mass spectrum shown here is easily assigned to pyrene by making use of the interactive display readout of the peaks positions.

Underneath the average mass spectrum the program displays a scrollable pixel-map of all the 386 mass spectra, one for each of the particles in this node. It is constructed from 386 horizontal lines, one for each of the IPMS, and the mass spectral peaks intensities are indicated by color, again using rainbow colormap, with red being high and blue being low. In this display the color scheme uses a log(log(I)) scale to enhance the presence of low intensity peaks and to make it easy to observe particle-to-particle variations. Clicking on any horizontal line displays the corresponding particle size and its time of detection. An overall view of the "mosaic" in this panel reveals three distinct regions that correspond to three different fragmentation patterns, one for each of the three clusters, 58, 59 and 60. These clusters merge to form node containing many of the are pyrene particles. The program is presently set to display up to 500 IPMS at a time and provides the option to scroll through the entire node/cluster content.

The last, green mass spectrum is a dynamic cluster summary view. By examining this mass spectrum the user can gain an additional level of understanding that is not observable from the average mass spectrum. The cluster summary mass spectrum is generated by superimposing the 386 IPMS on top of each other with an opacity that is determined by the slider bar below it. Changing the opacity allows the user to gradually bring up and observe the features that are present in fewer particles and learn more about the mass spectral peak intensity distribution. This display is a simplified version of parallel coordinates tuned to work on very high dimensional data. Note that all three views are aligned on the m/z scale, making it possible for the user to directly compare the different views.

The table below the three graphically displayed mass spectra summarizes the information about the node that is being explored. The node name is originally assigned by the scientist during dendogram exploration. Once assigned and saved, it can be reloaded, such that it will reappear anytime this node, or any unnamed node/cluster connected to it and positioned at larger radii is clicked. The other two cells in the table display the number of particles in the node and its fraction of the total number of particles represented in the dendogram.

In addition to the cluster/node information described above SpectraMiner easily generates other informative visualization formats of each cluster/node, which include, for example, 3D plots of particle population as a function of size and time of detection; individual particle mass spectra, linear plots of the number of particles in any clusters/nodes as a function of time or any other observable that was simultaneously measured, etc. The use of these commonly utilized visualization tools will be described in other publications.

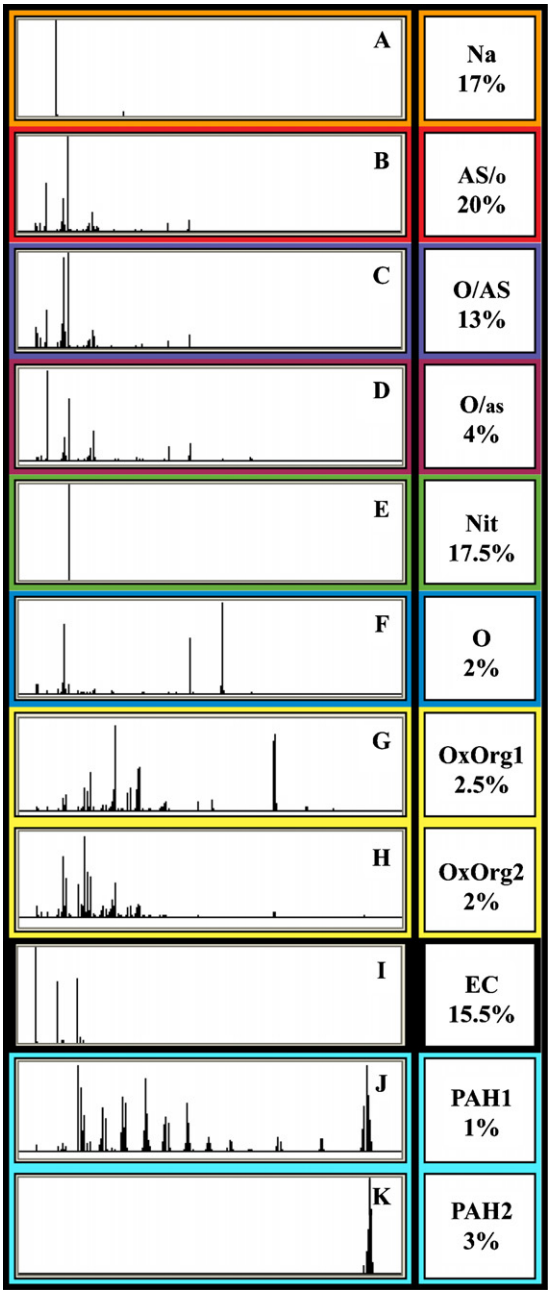

Fig. 2. Average mass spectra of the 11 major nodes (the *m*/*z* scale is 0–250), their tentatively assigned compositions and the fraction of particles in each node.

3.2.2. *Exploring the average mass spectra of the major nodes*

Shown in Fig. 2 are the average mass spectra, our tentative assignments and the fraction of particles in each of the selected 11 major nodes. The tentative assignment was carried out by examining all of the mass spectral information, while assuming no prior knowledge of particle composition. Table 2 provides a key to the shorthand notations used here and throughout the rest of the paper.

In Fig. 3a and b we present in bar graph format the relationship between the tentative assignments of the 11 nodes and the true particle compositions. Fig. 3a shows which particle types populate each of the 11 nodes. Note that even at this rather crude

Table 2
Listing of the tentative assignments of the 11 major nodes and their abbreviations

Abbreviation	Assignment
Na	Sodium
AS/o	AS with minor amount of small organic acid
A/AS	Small organic acid with AS
A/as	Small organic acid with minor amount of AS
Nit	Nitrate
O	Small organic acid
OxOrg1	Oxygenated organics
OxOrg2	Oxygenated organics
EC	Elemental carbon
PAH1	Pyrene
PAH 2	Pyrene

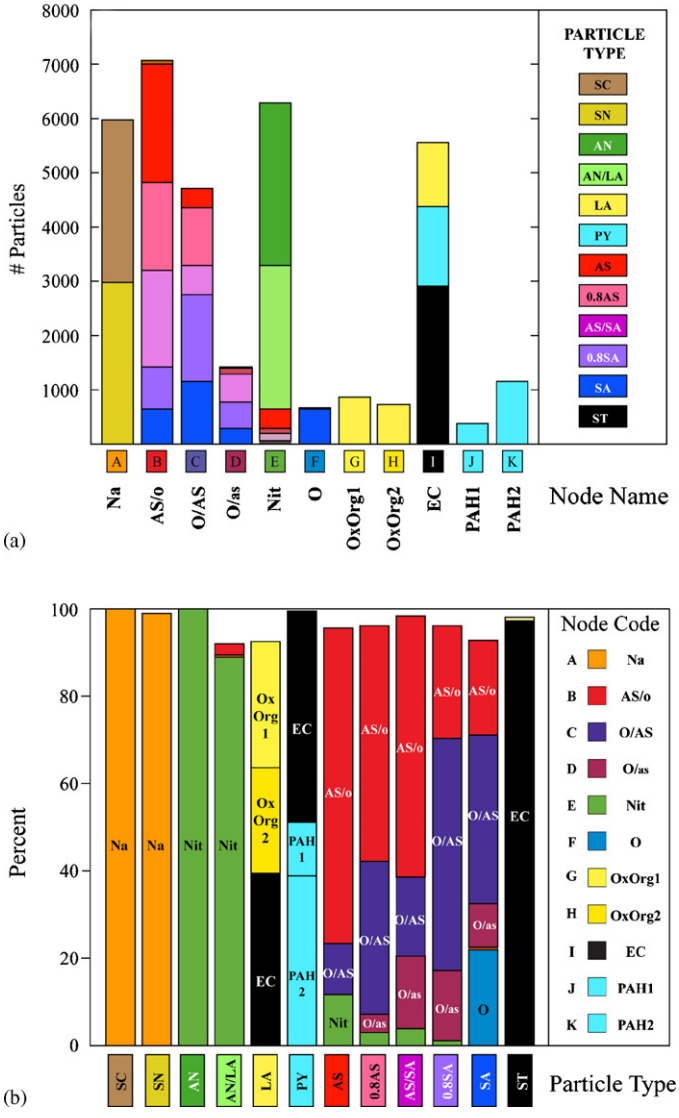

Fig. 3. (a) A bar graph display illustrating which particle types reside in which of the 11 nodes; (b) a bar graph display illustrating which nodes contain which of the 12 particle types.

level of classification AS particles (red) and AN particles (dark green) are for the most part distinguishable. In contrast, at this classification level there is insufficient detail to properly distinguish AS particles from internally mixed particles, composed of AS mixed with SA, or even from some of the pure SA particles. But a careful view of nodes B–D and F reveals that as we move from node B to C to D and to F the fraction of the particles with high AS content in the nodes decreases, while that of particles with high SA content increases. AN particles and AN/LA particles, the mass spectra of which are dominated by an intense NO^+ peak, share node E. In the same node we also find 11% of the AS particles. A 100% of the two sodium containing particle types; SN and SC, are in node A. Nodes A and E represent generic cases of particles, whose mass spectra are dominated by common intense peaks that greatly impact the classification. We will demonstrate an approach that can often resolve this type of a classification problem later in the paper.

The mass spectra of the particles in node I assigned as EC are dominated by a progression of C_n^+ peaks, indicating the presence or the formation during the ablation process of soot-like ionic fragments. As seen from Fig. 3a node I contains besides the real soot particles two other particle types, composed of organic molecules. The finding that a significant fraction of organic particles are indistinguishable from soot particles is another common feature of single particle mass spectroscopy. At its root is the high degree of fragmentation, often produced in ablation generated mass spectra, combined with a classification at coarse level. Ablation is the multiphoton process that is used to evaporate and create ions out of individual particles. It is not uncommon to find that ablation of smaller particles that contain organic compounds results in a high degree of fragmentation and produces mass spectra that are dominated by the C_n^+ progression. Often these mass spectra contain other, lower intensity peaks that contain the information that can be used to separate organic particles from soot, but get overlooked in a coarse classification.

In Fig. 3b we present the classification results from the particle type perspective. First we point to the fact that the 11 nodes contain 97.5% of all the particles (here we have not included clusters 3–6, 21 and 30). The first two particle types, with IPMS dominated by the Na^+ peak, not surprisingly are found in one class and assigned as sodium-containing particles. A similar situation is observed for the AN and AN/LA particle types with dominant NO^+ peak in their IPMS. The pattern produced by the other particle types provide a clear graphic display of the fact that the complex fragmentation pattern of organic molecules tends to spread their population into a number of nodes. LA and PY particles are each spread into three major nodes and some of their mass spectra are indistinguishable from the IPMS of soot particles.

In addition it is important to note that AS and internally mixed AS and SA particles are difficult to properly identify and segregate at the major node level.

Figs. 2 and 3a and b provide a graphic summary of the classification results at the level of 11 nodes. Were we to end the data mining process at this point, and treat it as one would IPMS of unknown ambient particles, our conclusions regarding the particle composition would be specified by the compositions assigned to the nodes with their corresponding average mass spectra and the relative particle populations as shown in Fig. 2. We would be forced to conclude that while we captured some of the properties of the sampled particles, we have missed too many others.

3.2.3. Intermediate comments on the results of the exploration of the 11 nodes

The common notion in the field is that distinguishing ammonium sulfate from ammonium nitrate particles on the basis of their positive ion mass spectra alone is very difficult. Our spectra and classification process do not exhibit the same limits even at the level of classification presented above. Instead we find in this and other studies that ablation generated mass spectra of particles containing organic compounds are the most difficult to properly classify.

The classification results thus far reveal that all the particle types composed of organics exhibit complex fragmentation patterns: LA populates nodes G, H, and I; SA inhabits nodes B–D, and F; and PY resides in nodes I–K. In addition, their fragmentation patterns exhibit particle size dependence. We will use PY as an example to illustrate some of the processes that create such a wide range of fragmentation patterns. We will also use it to demonstrate the visual tools that help with the mass spectral assignment process that are at our disposal.

In Fig. 4 we show in three separate frames the mass spectra of pyrene particles that reside in nodes I–K along with a listing of the fraction of pyrene particles in the corresponding node and the average size of its particle population. Each frame representing a node displays the three mass spectral views we mentioned earlier. Not surprisingly, we find that the mass spectra of the largest particles exhibit the least amount of fragmentation and their mass spectra are dominated by the parent ion peak. In contrast the smallest particles, with ~10 times lower mass, fragment to the point that their average mass spectrum is almost indistinguishable from that of elemental carbon. The observed dependence of the mass spectrum on particle size is not surprising considering the complexity and non-linearity of the ablation process. Examination of other particle types shows that the trend observed for pyrene is general: smaller particles typically exhibit larger degree of fragmentation.

In the laboratory settings, in which particle composition and size can be limited and are often known, it might be possible to reduce the ablation laser power and decrease the degree of fragmentation for the specific particle type and size. However, this sort of fine "tuning" is not applicable to the study of the ambient particles, whose compositions and sizes are unknown and range widely. The common effect of low ablation laser power is that important atmospheric particles like ammonium sulfate become undetectable.

It is worth noting that a careful examination of the data in Fig. 4 shows that the parent ion peak and the typical PAH progression can be observed in the pixel-map of the IPMS and in the green cluster summary mass spectra of each of the three nodes. In node I these peaks exhibit very low intensities in comparison with the three carbon ion peaks and are clearly insufficient to achieve separation from soot at this classification level. But the

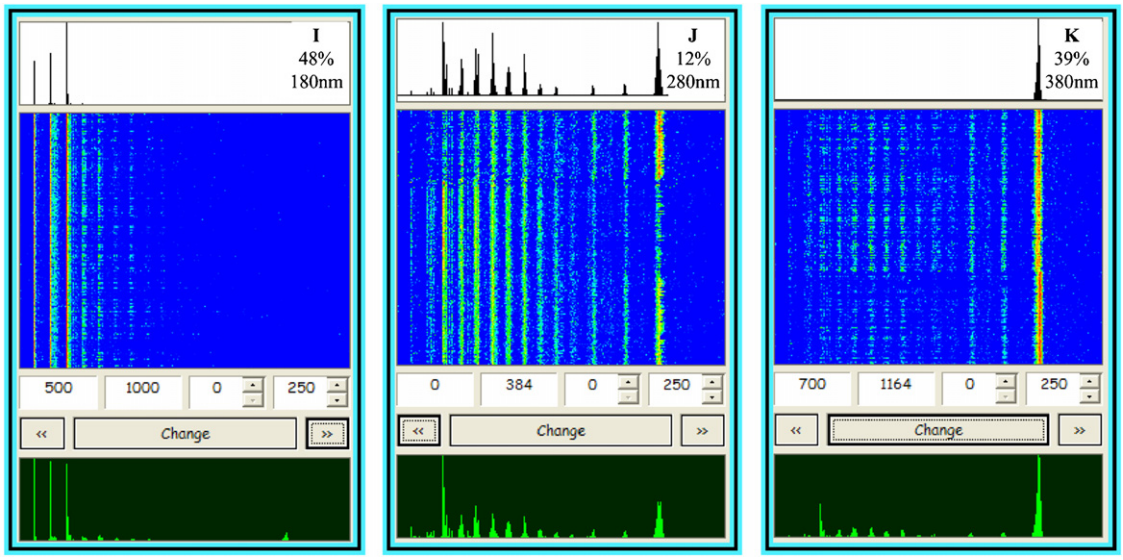

Fig. 4. The mass spectral information obtained for pyrene particles which were classified into the three labeled nodes. Note the relationship between particle size and the mass spectral fragmentation pattern.

fact that they can be observed in the majority of the pyrene IPMS suggests that it should be possible to take them into account to improve the classification.

Before we proceed to examine the classification results on a more refined level we find it worthwhile to visualize the patterns formed by clusters, nodes and branches of each of the particle types; the rudimentary structure of the classification output.

3.3. The 12 underlying dendograms

In Fig. 5 we provide a graphic presentation of the results of the classification separated according to the 12 particle types. This figure shows 12 dendograms and the corresponding projections of their clusters' populations, one for each of the particle types. To help orient the reader with respect to the original, all particle dendogram that is shown in Fig. 1, we have noted the cluster's numbers on the population projection and where relevant, indicated the same alphabetical node labels we have used throughout the paper.

An examination of the 12 individual particle type dendograms and their relationship to each other is very insightful. These dendograms reveal the complexity of the hierarchical patterns separated by particle type and show which clusters are populated by what particle type. We chose not to provide a detailed description of each of the dendograms in Fig. 5; instead, we will leave the reader to visually inspect the dendograms, use them as an information source for the discussion presented below and derive whatever specific information he/she finds to be useful.

3.4. Data exploration at the cluster level

3.4.1. Deconstructing node I into clusters 51 through 57

We chose to explore node I because it represents one the most important generic problems of ablation based single particle mass spectroscopy. Moreover, from an atmospheric science perspective, it is very important to be able to distinguish between particles composed of EC or soot and those containing the complex range of organic compounds. In the present case node I, which has tentatively been assigned to EC on the basis of its average mass spectrum, contains in addition to soot 39% of the LA particles and 48% of the PY particles.

Fig. 6 displays the mass spectral information for clusters 51, 54, and 57, three of the six clusters that merge to form node I. What is common to all these clusters is that the most intense peaks in the spectra are C_1^+, C_2^+ and C_3^+. But, an examination of the three mass spectral display formats makes it clearly apparent that the mass spectra of the particles in these three nodes are clearly quite different and that they can be unambiguously assigned as soot (cluster 57), PAH (cluster 54) and oxygenated organics (cluster 51) as annotated in the figure.

In Fig. 7 we return to the bar graph presentation of the nodes and their contents and extend it to include clusters 51 through 57. An examination of the compositions of the particles in these clusters shows that at the cluster level the only cluster that includes more than one particle type is cluster 56, in which 19% of the LA and 12% of the PY particles are indistinguishable from EC. In all other clusters the particles have been properly identified.

3.4.2. Deconstructing node B into clusters 7 through 15

Internally mixed particles composed of AS mixed with organics are some of the most common particle types found in the atmosphere. Since many of the properties of these particles strongly depend on the relative amounts of AS and the organics, it is clearly important to be able to accurately identify the particles internal composition. Node B, which on the basis of its average mass spectrum we tentatively assigned to AS with a minor fraction of a small organic acid (AS/o), contains five particle types: AS, 0.8AS, AS/SA, 0.8SA and pure SA.

Returning to Fig. 7, where we present the results of deconstructing this node into its clusters 7 through 15, we note that on the cluster level the picture for node B becomes significantly better resolved. The clusters that comprise node B can be separated

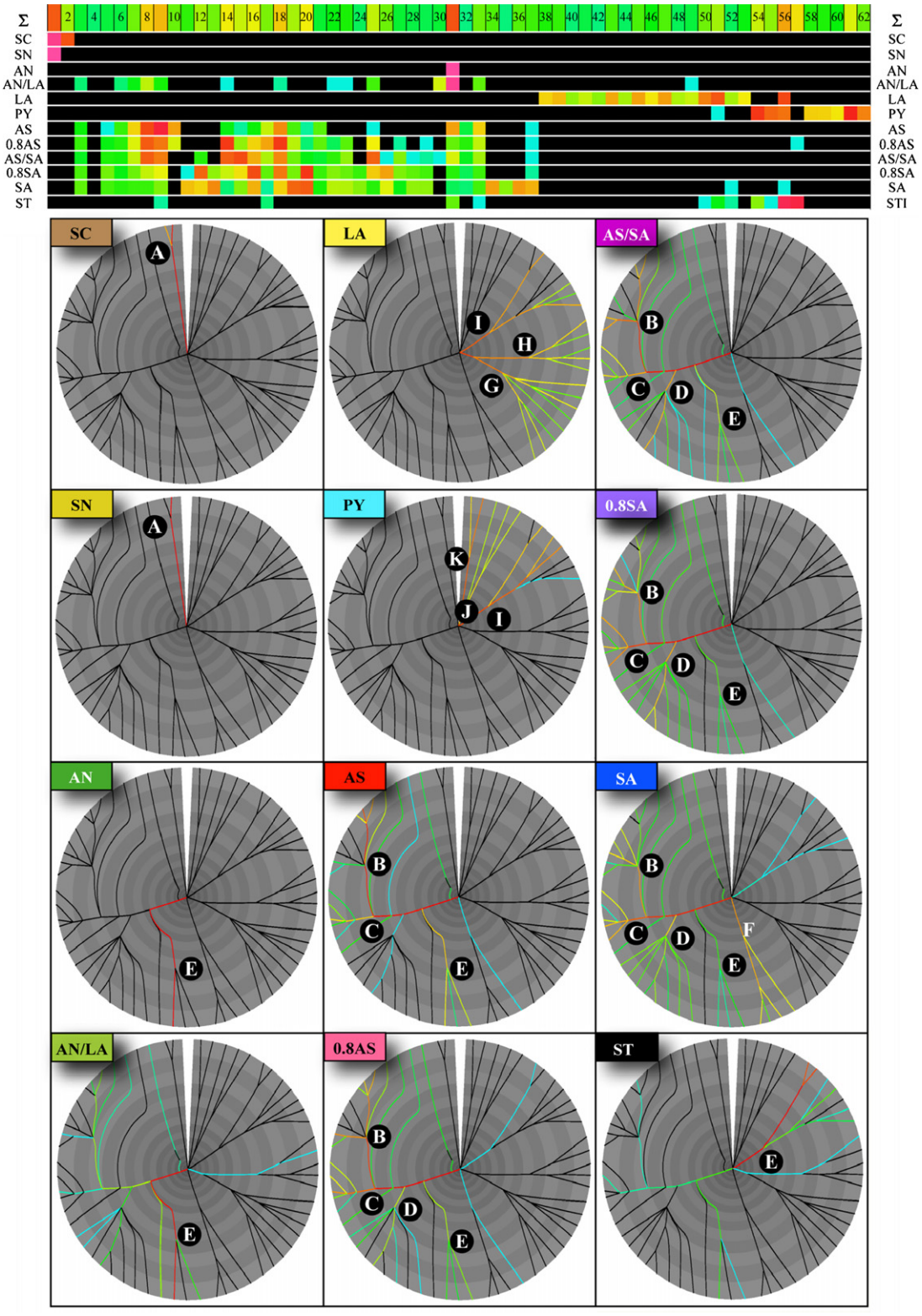

Fig. 5. A display of the particles distributions in the overall dendogram separated into each of the 12 particle types. The 13 colored horizontal lines above are projections of the particle cluster populations starting with the entire dataset (labeled Σ) and proceeding with a projection for each of the 12 particle types as marked.

into three types: clusters 7–10 contain particles whose composition is dominated by AS but many contain small to medium amounts of SA. Clusters 11–13 contain particles whose composition is dominated by SA and clusters 14 and 15 are dominated by mixed particles that contain significant fractions of both AS and SA.

While it is clear that the mass spectra of these particles do not contain sufficient details to unambiguously resolve node B into

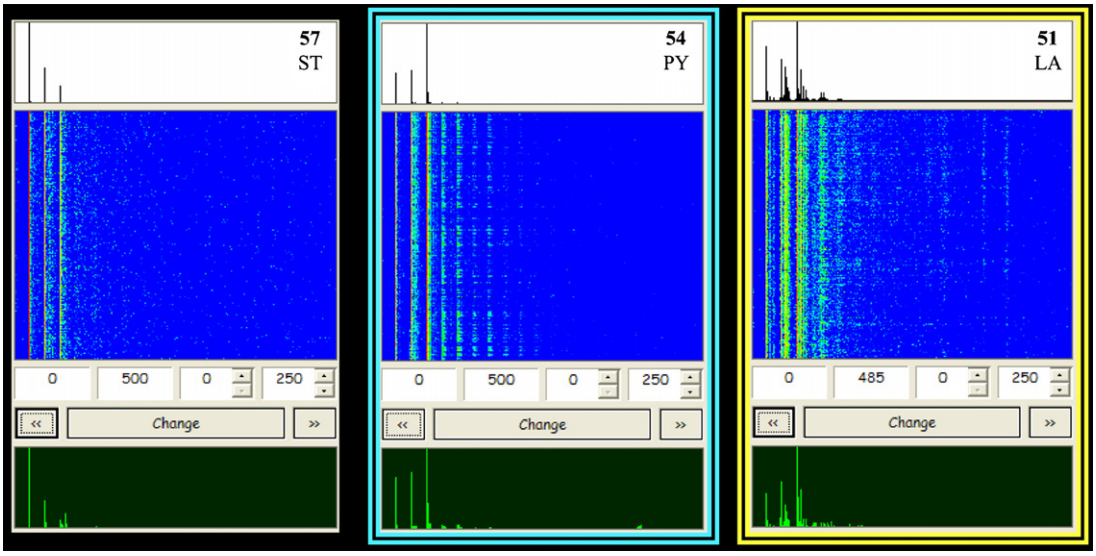

Fig. 6. The mass spectral information of the particles that are classified into clusters 51, 54 and 57 and their assignments.

all its particle types, we have demonstrated that at the cluster level we could separate the node into three types of internally mixed AS/SA particles in accord with their variable composition. Similar refinement can be achieved for nodes C and D.

3.4.3. *Deconstructing node A to the cluster level*

The laser ablation generated mass spectra of particles that contain species with low ionization potentials like alkali metals are often dominated by a few very intense mass spectral peaks of

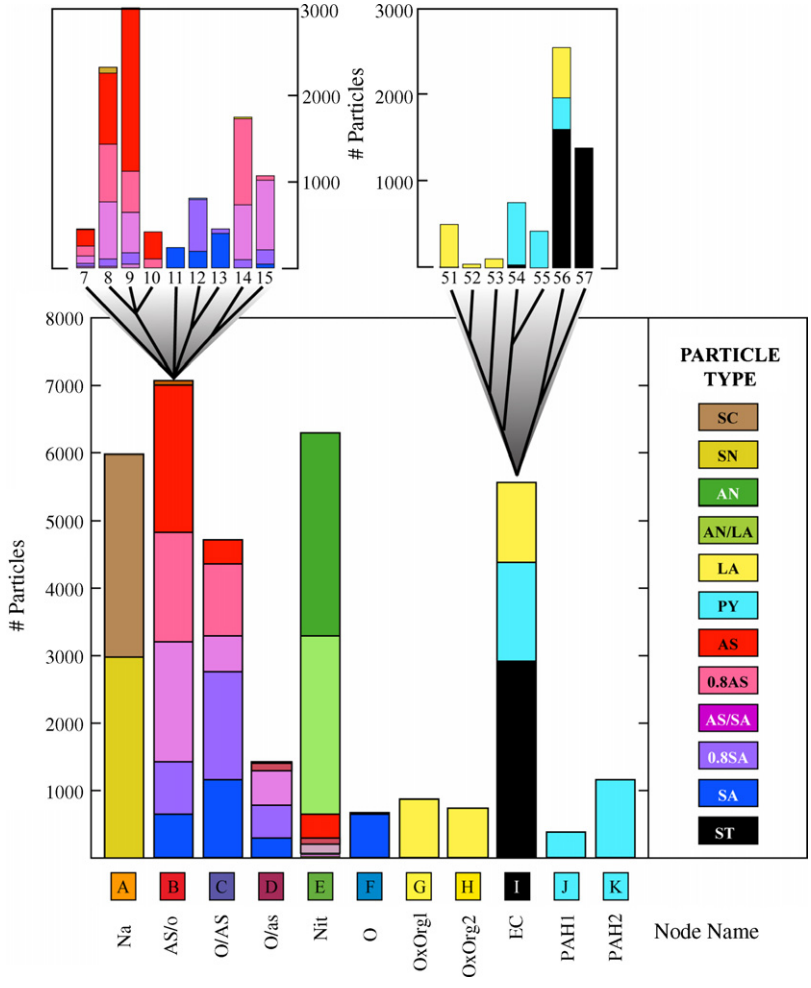

Fig. 7. A bar graph display illustrating the results of deconstructing nodes B and I into their clusters.

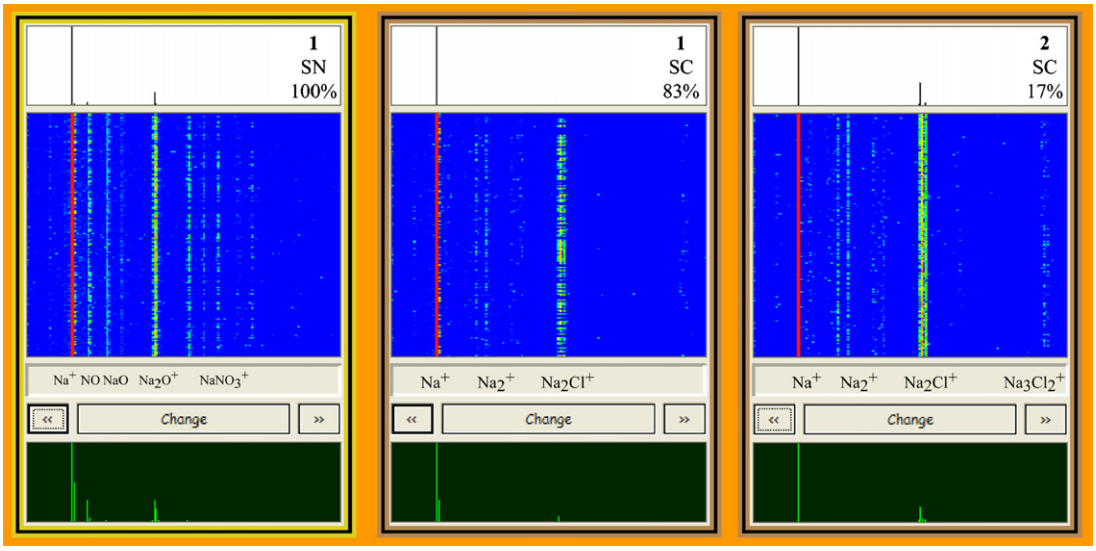

Fig. 8. The mass spectral information of the particles that are classified into clusters 1 and 2 separated according to the particle types.

these species. In a recent paper [10] we have demonstrated that the presence of sodium atoms can even act to suppress other ions through charge transfer processes. In cases where two or more particle types share very intense mass spectral peaks, a typical classification process groups these particles together.

In the present study we find that SC and SN particles are classified together in node A because of the intense Na^+ peak the two particle types share and the AN/LA and pure AN get grouped together in node E because of the presence of the intense NO^+ peak. Below we will explore two approaches to refine the classification and overcome some of these limitations using SC and SN particles as examples.

Fig. 8 shows the mass spectral information of SN and SC particles that reside in clusters 1and 2. An examination of these mass spectra shows that all of them exhibit an intense Na^+ peak. But we also note that the mass spectra of the SN and SC particles have other than Na^+ peaks, which are unique to each of the particle types. These weaker peaks are most easily observable on their IPMS pixel maps. These additional mass spectral peaks that are not shared by the two particle types can provide the information needed to sort these two particle types. To improve the classification there is a need to refine it by brining these other peaks to the forefront.

In the left part of Fig. 9, in bar graph format, we present the deconstruction of node A into its clusters 1 and 2. We note that for these particle types even at the cluster level of classification only 17% of the SC particles are separated from the rest of the Na-containing particles.

A straightforward approach, which could refine the IPMS classification, is to decrease the distance threshold parameter. In the right part of Fig. 9 we present the results of a classification run on the particles in node A only, with a distance threshold set to 0.1, showing that reducing the distance threshold, by a factor of three somewhat improves the results. In this case node A is deconstructed into four clusters, in which 69% of the SN and 26% of the SC particles reside in separate clusters.

3.5. Expert driven classification

3.5.1. Deconstructing node A with expert input

It is possible to take advantage of the knowledge gained by examining the mass spectral information, at the node and cluster levels, to refine the classification process and separate these mass spectrally similar yet very different particle types. The approach we adopt here is to apply our expert knowledge by inputting it into the classification to help in its refinement. We base our input on the conclusions we arrived at on the basis of an examination of the IPMS shown in Fig. 8. We have already noted above the existence of peaks, other than Na^+, that uniquely identify the two particle types, and concluded that if we were to base the classification on all, but the Na^+ peak, the two particle types could easily be separated.

There are a number of approaches one can take to sculpt the mass spectral features and "help" the classification process. One of them is to assign to peaks different "weights", which would allow the user to amplify certain mass spectrometric features. Here we took the simplest approach by assigning zero weight to the Na^+ peak, i.e., eliminating this peak in the IPMS of all the Na-containing particles altogether. Following the removal of the Na^+ peak the mass spectra of the 6000 particles that belong to node A are reclassified with a distance threshold of 0.3. As the result, aside from 11 (0.37%) SC particles that share some of the SN clusters, the two particle types have been successfully separated.

3.5.2. Expert driven classification of inorganic/organic particles

We have already mentioned above that the AN and AN/LA particles present a case similar to the SC and SN particles. A procedure, comparable to that used to separate the Na-containing particles can be applied to refine the classification of node E to separate the AN and the AN/LA particles, whose mass spectra are dominated by an intense NO^+ peak. Fig. 10 shows the mass spectral information of the three particle types in node E:

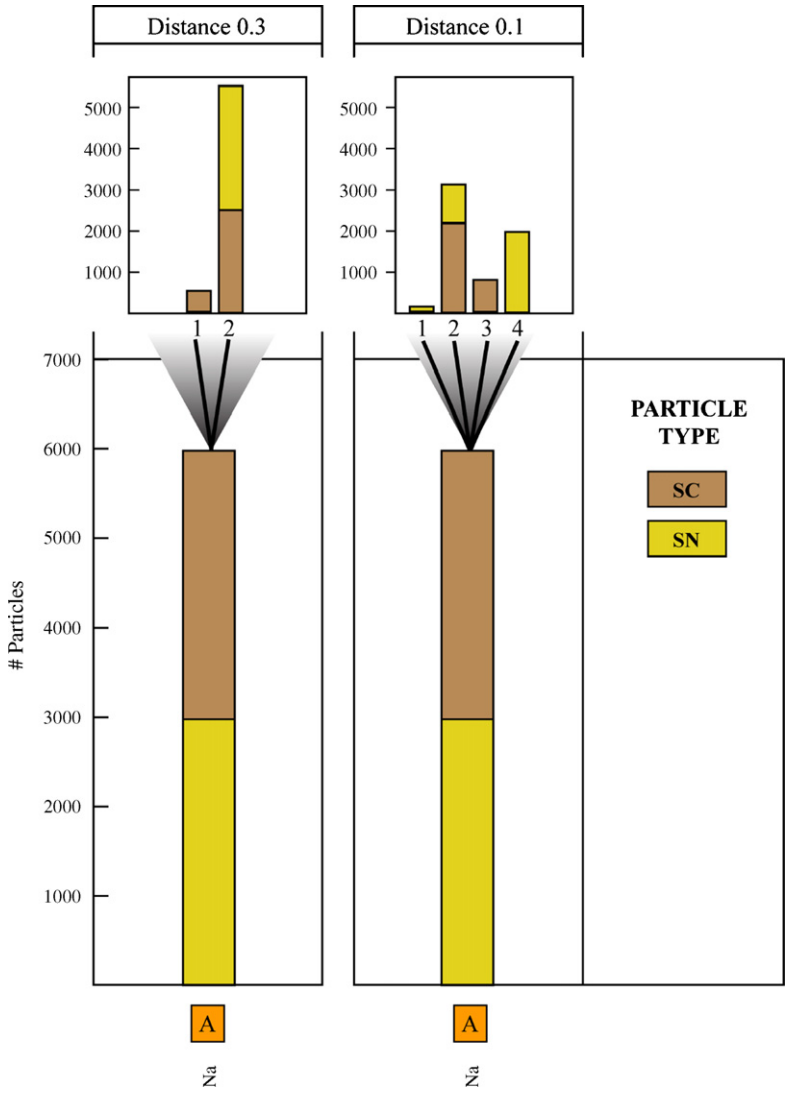

Fig. 9. A bar graph display illustrating the results of deconstructing node A into clusters 1 and 2 with a distance threshold of 0.3 and into four clusters with a distance threshold of 0.1.

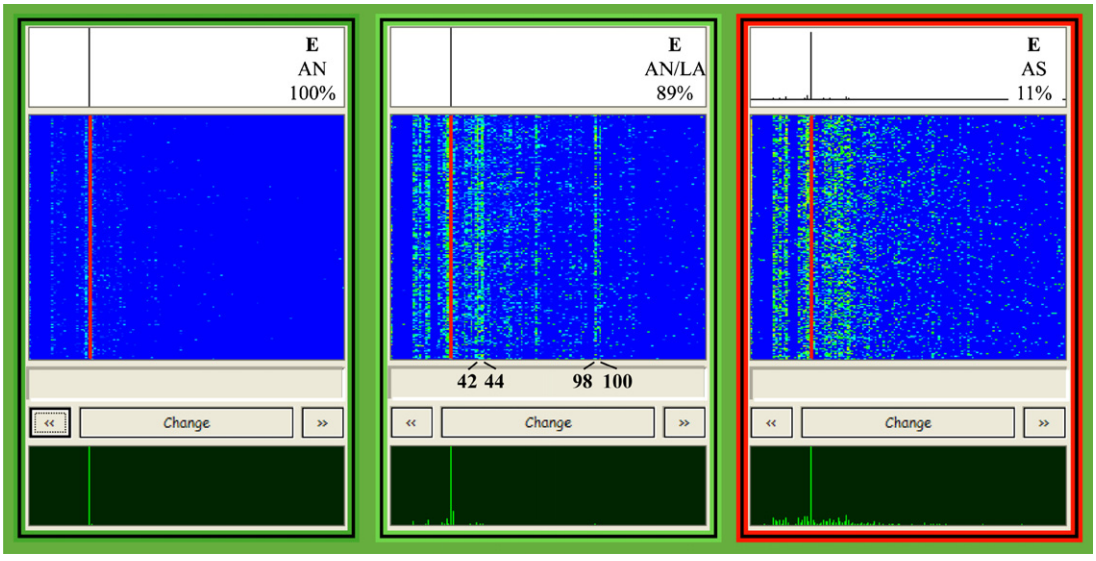

Fig. 10. The mass spectral information of the particles that are classified into node E separated according to the three particle types that populate the node.

100% of AN, 89% of AN/LA and 11% of AS. An examination of these mass spectra shows that it should be possible to separate the three particle types in this node by assigning lower weight to the dominant NO^+ peak. Note, however, that the mass spectra of the AS particles that reside in node E, are contaminated by the presence of a small amount of organics, which would make it difficult to completely resolve them from the AN/LA particles. The appearance of organic compounds in the mass spectra of "pure" AS has been reported by others [4]. It has been interpreted to be a result of the low ionization probability of pure ammonium sulfate [11,12] that makes it possible to detect even minute quantities of organic contaminants.

In the case of mixed AS/organics particles we can use a slightly different approach and enhance the importance of small but characteristic key mass spectral peaks that signify AS and specific organic compounds of interest. In some sense identifying key peaks and increasing their weights is similar to the approach used by the ADAMS [5] classification in which discriminant chemical markers are used to guide the classification. By assigning higher weights to key peaks we can decrease the importance that weak, background and/or impurity peaks play and guide the classification to more precisely identify all the particles containing AS and internally mixed AS particles with organics.

To test this approach, we have used peak with $m/z = 48$ as a marker for AS and peaks with $m/z = 43$, 100, and 118 as markers for SA and increased their weight (intensities) by factor of 10. Rerunning this expert steered classification of the weighted IPMS shows that all the AS, SA and AS mixed with SA particle types have been successfully separated from the rest of the dataset. A 87% of these 15,000 particles were classified into 19 clusters, whose populations are illustrated in Fig. 11a with the five projections, one for each particle type, and in the bar graph format in Fig. 11b. We note that this classification produced three main particle groups that can be assigned as mostly AS with a small amount of SA, AS mixed with SA and mostly SA with some AS. We find these results to be very encouraging and plan to apply the approach to the ambient atmospheric data we have acquired.

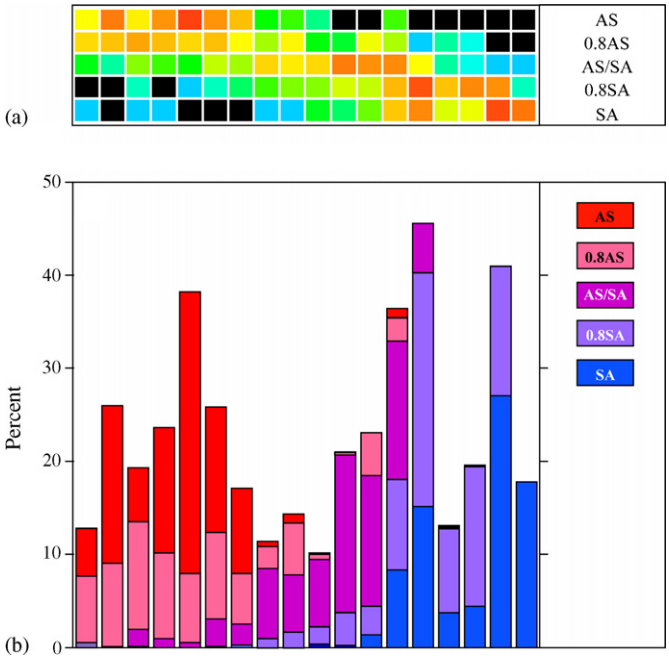

Fig. 11. An illustration of the results of the expert steered classification of the AS, SA and their three internally mixed particle types: (a) projections of the cluster populations for the 5 particle types; (b) bar graph format.

Another implication to ambient atmospheric data is an improved accuracy in the assignment of potassium-containing particles. Potassium-containing particles are frequently detected in the atmosphere and often present a case that is similar to the Na-containing particles discussed above. There are a number of well-known sources of atmospheric potassium, which can easily be distinguished from each other on the basis of the other compounds that are inevitably present and observed in the same particles. Potassium, like all other alkali metal atoms, tends to dominate the IPMS, making it difficult to identify and properly classify these often very different particles.

We have shown above that cluster visualization is a powerful strategy for the exploration of high-dimensional data in the absence of a-priori hypotheses or data classification models. But even though formal models may not exist, we have a vast amount of knowledge and intuition that we can bring to bear in this effort. The results of a non-supervised data classification rarely agree with this expert knowledge. SpectraMiner offers data visualization and mining tools that make it possible for the user to evaluate the results of the classification with ease and inject scientific domain knowledge and intuition to steer the clustering process. Moreover, SpectraMiner is structured in a manner that makes it easy to focus on subsets of the data that are of particular interest. In the present study, sculpting the 6000 mass spectra in node A or the 15,000 mass spectra of internally mixed AS and SA, particles using different expert driven approaches, was performed in less than a minute and significantly improved the representation of the sampled particles.

4. Conclusions

We presented a study, in which we put to test the classification aspect of SpectraMiner using IPMS of 12 particle types generated by SPLAT operated in the ablation mode only. These particle types were chosen in a manner that represents some of the generic problems that are often encountered by laser ablation based single particle mass spectroscopy of atmospheric particles. In our analysis of the data we have illustrated the approaches we have taken to bring about improvements and even solutions to some of the common problems that we face in this field. We have developed approaches that give the user the option to push the data analysis process to the point where the limits are set by the quality of the data. To accomplish this goal we developed the tools that make it possible to explore the data on variable levels of details with great ease.

The measurements of the organic content of ambient particles represent a particularly challenging and important aspect of SPMS. In this study our data and its analysis illustrated that two very different organic particle types can fragment to the point that, at coarse classification level, nearly half of their particles

are indistinguishable from EC. We showed that in these cases it was possible to significantly improve particle identification by simply exploring the results of classification at the cluster level.

Particles composed of sulfates internally mixed with organics represent a very significant fraction of atmospheric aerosols. Here we looked at the internally mixed AS and SA particles and concluded that at the major nodes level particle identification is rather poor, whereas at the cluster level we noted significant improvements.

We showed that the addition of expert input to sculpt the mass spectral data can greatly improve the outcome of the classification. This approach was demonstrated for a case in which two different particle types share a few dominant peaks. We showed how to visually analyze the mass spectral information and select and implement an expert driven classification procedure that eliminates the overlap between the particle types. We illustrated the application of expert steered classification to isolate and classify the AS, internally mixed AS and organics and SA particles with significant success.

We must conclude that while there are ways to improve the data analysis process the true limits we are presently facing stem from the quality of the mass spectra that are generated by laser ablation based SPMSs. To accomplish significant improvements we must address this aspect. One approach that we and others have deployed [9,13,14] is to use a two-laser, two-step process for ion generation: the first step relies on an infrared laser to evaporate the particles, and the second delivers a time-delayed UV or VUV pulse to ionize the resultant plume [9,13,14]. Another approach is to incorporate into classification additional measurements for the same individual particles. Most commonly this is accomplished by simultaneously measuring the positive and negative ion mass spectra [15–17]. It might be possible to improve the classification results by incorporating into it other simultaneously measured single particle properties like density [18–20].

We showed some of the features offered by SpectraMiner and demonstrated that it yields a data mining platform that offers great flexibility. This flexibility is apparent even at the node level, where the number of nodes being examined is comparable to that used by other data classification approaches. Except that we let the structure of the data determine the "natural" distances, at which clusters merged to form key junctions. Most importantly, the more detailed information at larger radii is always only a mouse-click away.

In the present study we pruned the dendogram data presentation to 20 out of 36,000 particles, or 0.06%. In applications that relate to global climate such detail may not be needed. In contrast, in homeland security applications or in studies of special atmospheric events or episodes we could be searching for the extremely rare "nuggets" that represent a minute fraction of all the sampled particles. SpectraMiner is designed to make the transformation between these two limits easy.

An essential aspect of this software is the speed with which the program responds. We felt that it was very important to develop software, in which the transformations we described here and all the other data manipulations involved in the data mining process would be fast, even when the software is executed on an office or laptop PC and the dataset contains ∼1 million particles.

It is important to mention that in this study we presented only a fraction of the SpectraMiner features. Many of the other features were specifically designed to analyze ambient data as function of time of their detection. SpectraMiner makes it possible to interactively view the time evolution of the particle composition data in a number of visual formats and search for correlations in the time evolution of particle clusters and nodes. SpectraMiner is also set to load any other data that is acquired in parallel, like gas phase pollutant concentrations, wind direction, or engine performance characteristics and analyze the relationship between these data and the particle composition. These aspects of the software will be presented in separate publications.

Acknowledgments

Part of this research was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research at Pacific Northwest National Laboratory (PNNL). PNNL is operated by the US Department of Energy by Battelle Memorial Institute under contract No. DE-AC06-76RL0 1830. This work was supported by the US Department of Energy Office of Basic Energy Sciences, Chemical Sciences Division.

References

[1] X.-H. Song, P.K. Hopke, D.R. Fergenson, K.A. Prather, Anal. Chem. 71 (1999) 860.
[2] D.M. Murphy, A.M. Middlebrook, M. Warshawsky, Aerosol Sci. Technol. 37 (2003) 382.
[3] A. Trimbom, K.-R. Hinz, B. Spengler, Aerosol Sci. Technol. 33 (2000) 191.
[4] D.J. Phares, K.P. Rhoads, A.S. Wexler, D.B. Kane, M.V. Johnston, Anal. Chem. 73 (2001) 2338.
[5] P.V. Tan, O. Malpica, G.J. Evans, S. Owega, M.S. Fila, J. Am. Soc. Mass Spectrom. 13 (2002) 826.
[6] D.J. Phares, K.P. Rhoads, M.V. Johnston, A.S. Wexler, J. Geophys. Res.-Atmos. 108 (D7) (2003) (Art. No. 8420).
[7] 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. Seinfeld, R.C. Flagan, J. Geophys. Res.-Atmos. 108 (D7) (2003) (Art. No. 8424).
[8] A. Zelenyuk, D.G. Imre, K. Mueller, P. Imrich, W. Zhu, R. Mugno, Eos Trans. AGU 83 (47) (2002) A12D.
[9] A. Zelenyuk, D. Imre, Aerosol Sci. Technol. 39 (2005) 554–568.
[10] Y. Cai, A. Zelenyuk, D. Imre, Aerosol Sci. Technol., submitted for publication.
[11] D.B. Kane, M.V. Johnston, Environ. Sci. Technol. 34 (2000) 4887–4893.
[12] R.J. Wenzel, D.Y. Liu, E.S. Edgerton, K.A. Prather, J. Geophys. Res. - Atmos. 108 (D7) (2003), Art. No. 8427.
[13] B.D. Morrical, D.P. Fergenson, K.A. Prather, J. Am. Soc. Mass Spectrom. 9 (1998) 1068.
[14] E. Woods, G.D. Smith, Y. Dessiaterik, T. Baer, R.E. Miller, Anal. Chem. 73 (2001) 2317.

[15] E. Gard, J.E. Mayer, B.D. Morrical, T. Dienes, D.P. Fergenson, K.A. Prather, Anal. Chem. 69 (1997) 4083.
[16] A.D. Lake, M.P. Tolocka, M.V. Johnston, A.S. Wexler, Environ. Sci. Technol. 37 (2003) 3268.
[17] K.P. Hinz, R. Kaufmann, B. Spengler, Aerosol Sci. Technol. 24 (1996) 233.
[18] D.M. Murphy, D.J. Cziczo, P.K. Hudson, M.E. Schein, D.S. Thomson, J. Aerosol Sci. 35 (2004) 135.
[19] A. Zelenyuk, Y. Cai, L. Chieffo, D. Imre, Aerosol Sci. Technol. 39 (2005) 972.
[20] A. Zelenyuk, D.G. Imre, J.-H. Han, S. Oatis, Eos Trans. AGU 84 (46) (2003) A12B.