*Anal. Chem.* **1999,** *71,* 860−865

# Classification of Single Particles Analyzed by ATOFMS Using an Artificial Neural Network, ART-2A

Xin-Hua Song and Philip K. Hopke*

*Department of Chemistry, Clarkson University, Potsdam, New York 13699-5810*

David P. Fergenson and Kimberly A. Prather

*Department of Chemistry, University of California, Riverside, California 92521*

**Aerosol particles have received significant public and scientific attention in recent years due to studies linking them to global climatic changes and human health effects. In 1994, Prather et al. (Prather, K. A.; Nordmeyer, T.; Salt, K. *Anal. Chem.* 1994, *66,* 1403−1407) developed aerosol time-of-flight mass spectrometry (ATOFMS), the first technique capable of simultaneously determining both size and chemical composition of polydisperse single particles in real time. ATOFMS can typically analyze between 50 and 100 particles/min under typical atmospheric conditions. This significant volume of data requires automated data analysis for efficient processing. This paper reports the successful analysis of ATOFMS data acquired during a 1996 field study in Southern California using an adaptive resonance theory-based neural network, ART-2a. The ART-2a network revealed particle categories consistent with those obtained previously by manual analysis. The classification was accomplished in less time than the acquisition, rendering it possible to develop a data acquisition system using an on-line ART-2a that classifies particles as they are acquired.**

Since Dockery et al.[1] published their study in 1993 linking aerosol particles to human morbidity and mortality, researchers have been paying increased attention to the sources and effects of atmospheric particulate matter. First developed in 1994, aerosol time-of-flight mass spectrometry (ATOFMS), a new technique capable of measuring the precise size and chemical composition of individual aerosol particles in real time, offers several advantages over conventional particulate analysis techniques.[2,3] Because ATOFMS analyzes individual particles, it can be used to discern between internally and externally mixed particulate populations. Also, through information on both the aerodynamic diameter and associated single particle composition, one can gain insight into the sources, chemistry, and relevant health effects of tropospheric particles.

ATOFMS instruments are theoretically capable of determining the size and chemical composition of 600 particles/min, with rates of 50−100 particles/min common under normal atmospheric concentrations. In fact, ATOFMS presents the problem of acquiring data at a rate exceeding that required currently to manually analyze the data.[4] It is a generally difficult task to find a method of automated data analysis that can handle data sets as large as those produced by ATOFMS over the course of several hours let alone continuously over a period of weeks with multiple instruments sampling in parallel.

In this paper, we present the successful implementation of the ART-2a algorithm for the analysis of ATOFMS single-particle data. Adaptive resonance theory (ART)[5,6] neural networks are efficient methods of intelligently finding clusters in data sets with large numbers of variables. ART has the ability to add a class should a data point fall outside of a preset proximity to all existing classes. In this way, it can identify new categories of data as they emerge without disturbing the existing categories. The most recent incarnation of the method is a particularly efficient execution called the ART-2a algorithm.[7] The ART-2a algorithm was previously used for the analysis of data from off-line particulate-measuring techniques where it identified different particle classes on the basis of shape and elemental composition as determined by scanning electron microscopy.[8,9]

## EXPERIMENTAL SECTION

**Method of Data Collection.** A more thorough description of the field-portable ATOFMS systems appears elsewhere.[10] For completeness, the operating principles are presented briefly here. Particle-laden air is drawn into the ATOFMS instrument through a converging nozzle. As the particles traverse the nozzle, they are differentially accelerated as a function of their aerodynamic

(1) Dockery, D. W.; et al. *N. Engl. J. Med.* **1993**, *329*, 1753−1759.
(2) Prather, K. A.; Nordmeyer, T.; Salt, K. *Anal. Chem.* **1994**, *66*, 1403−1407.
(3) Salt, K.; Noble, C. A.; Prather, K. A. *Anal. Chem.* **1996**, *68*, 230−234.
(4) Fergenson, D. P.; Liu, D.-Y.; Silva, P. J.; Prather, K. A. *Chemom. Intell. Lab. Syst.* **1997**, *37*, 197−203.
(5) Grossberg, S. *Biol. Cybernetics* **1976**, *23*, 121−134.
(6) Grossberg, S. *Biol. Cybernetics* **1976**, *23*, 187−202.
(7) Carpenter, G. A.; Grossberg, S.; Rosen, D. B. *Neural Networks* **1991**, *4*, 493−504.
(8) Xie, Y.; Hopke, P. K.; Wienke, D. *Environ. Sci. Technol.* **1994**, *28*, 1921−1928.
(9) Hopke, P. K.; Song, X.-H. *Anal. Chim. Acta* **1997**, *348*, 375−388.
(10) Gard, E. E.; et al. *Anal. Chem.* **1997**, *69*, 4083−4091.



10.1021/ac9809682 CCC: $18.00 
Published on Web 01/13/1999

diameter. Excess air is pumped away through three stages of differential pumping as the particles are collimated into an aerosol beam by skimmer cones separating each stage. Each particle not removed by the skimmers passes through two continuous wave laser beams separated by 6 cm. Light scattered from each particle as it passes through the lasers is detected by photomultiplier tubes. The exact time difference between the two light-scattering pulses is recorded and can be used to obtain the velocity of the particle, which is a function of its aerodynamic diameter. The position and velocity of the particle, as measured by the scattering pulses, is further used to actuate the firing of a high-energy pulsed Nd:YAG laser to coincide with the arrival of the particle in the source region of a time-of-flight mass spectrometer. The laser desorbs and ionizes molecules from the particle, and the mass spectrometer collects a mass spectrum of the generated ions. In this manner, the composition of the aerodynamically sized individual particle is determined.

The particle data used in this study were collected over a period of three days in Long Beach, CA, as part of a larger field study conducted between September 23 and October 2, 1996. A field-portable ATOFMS system was placed in a laboratory on the fifth floor of a building located at California State University at Long Beach. Ambient aerosols were introduced into the system continuously through a $^3/_8$-in. copper tube protruding through an open window connected directly to the interface of the instrument. The system was operated continuously over the entire period presented here.

**Data Description and Pretreatment.** The data consist of approximately 43 000 particle positive ion mass spectra and their associated velocities collected between 6:00 AM PST on September 23 and 8:00 PM PST on September 25, 1996. Each spectrum contains 16 000 8-bit data points. The particle mass spectra were calibrated using software developed in-house. A list containing the area and exact mass-to-charge ratio of all peaks in each particle mass spectrum was generated. Peaks less than 10 arbitrary units above baseline on a 256-unit scale or with areas of less than 30 arbitrary units were rejected. The revised lists were compiled into matrices in MATLAB where each spectrum's data was stored as one row in a matrix, with the value at the *n*th column representing the area under any peak within 0.5 Da of *n*. The matrices were organized into 2-h blocks, beginning and ending on even-numbered hours. Each particle mass spectrum is represented numerically as a vector of 170 numbers such that the *i*th number corresponds to the area under all significant peaks falling within 0.5 mass-to-charge unit within *I*, and each vector is normalized to unit length according to the Pythagorean theorem.

**Application of the ART-2a Algorithm.** The ART-2a program was written and executed in MATLAB 5.2 on an IBM-compatible microcomputer running Microsoft WindowsNT 4.0 on an AMD-K6 microchip. General descriptions of the ART-2a algorithm appear elsewhere.[7–9] Only a description focusing on its application to the mass spectral data will be presented here.

The purpose of this application of the ART-2a algorithm is to identify the various classes of particles that exist without any a priori assumptions about their possible identities. To visualize a class, it is convenient to generate a "quintessential" particle to represent this class. The set of quintessential particles is called the weight matrix, and as such, each quintessential particle is

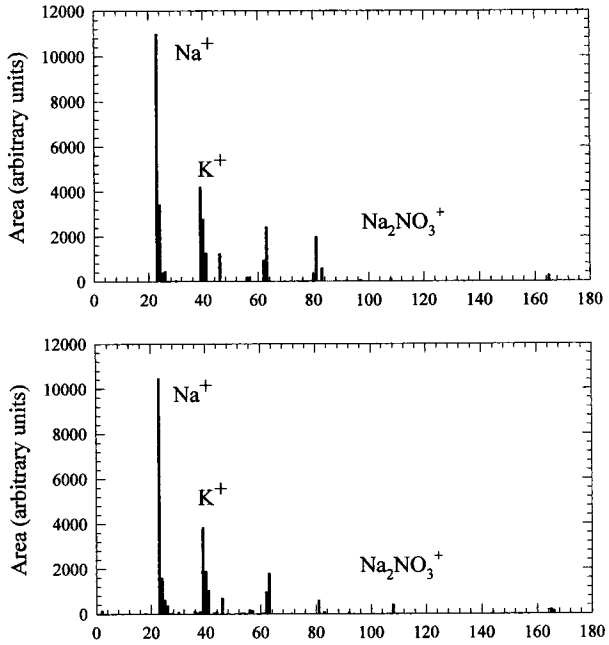


Figure 1. (top) Centroid mass spectra of the unreacted marine particle class; (bottom) centroid mass spectra of the reacted marine particle class.

referred to as a weight vector. The particles are selected in random order and compared to each weight vector, one at a time. The degree of similarity between the particle and the weight vector is calculated numerically by taking the dot product of the corresponding particle vector and weight vector.

Each particle is first compared to all of the existing weight vectors, and the winning weight vector with the largest degree of similarity is found; if this winner meets the vigilance criterion (predefined as the vigilance factor), then the particle vector is considered as belonging to the class the winning weight vector represents. If the particle vector does not fall within a vigilance factor of any weight vector, then the particle is considered to be representative of a previously unidentified class and its vector is added to the weight matrix. If it does fall within a vigilance factor of a weight vector, then the weight vector is moved in the direction of the particle vector associated with it by a certain amount called the learning rate. The first particle selected in the analysis will always become a weight vector because there are no existing weight vectors to which it could be considered close. In these experiments, the learning rate was set at 0.05 and the vigilance factor 0.7. It is interesting to note that the dot product of two vectors of unit length generated from random numbers would be $2^{1/2}/2$ which is greater than the vigilance factor used. A lenient vigilance factor can be employed because the data are not random.

The particles are selected, one at a time, until none are left unclassified. The process is repeated while the weight matrix is maintained between iterations. In this manner, the weight vectors will find themselves nucleating clusters of particles of more and more similar memberships from iteration to iteration. After the network is trained for a given number of cycles, in this application 20, the analysis is said to be complete. It was observed that after 18−20 cycles of training the values of network weight vectors did not change or only changed very little. Further training beyond 20 cycles would not improve the classification results significantly

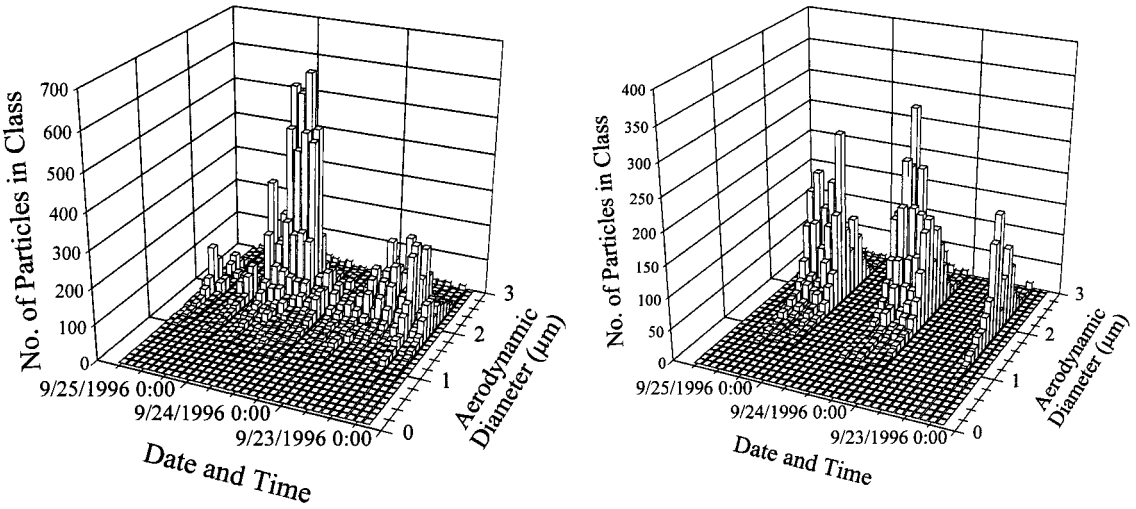


Figure 2. Temporal variations of the size distributions of the unreacted marine (left) and reacted marine (right) particle classes. Note that the two types are antiassociated in time.

and it would need longer computation time. All cycles were done repeatedly for all of the particle input vectors with random ordering. For the ART-type neural network, the ordering of the input presentation to the network could affect the classification results a little bit, but for the large data set treatment as in the present application, the ordering of the input presentation almost did not have any effect on the classification results, because the major classes were considered as described in the following context.

**Process of Analysis.** Any cluster containing more than 1% of all particles was considered important. The mass spectra of all of the particles in each important cluster were averaged to yield the centroid of that cluster. The classes produced by the ART-2a neural network were grouped by visual inspection into the seven particle types, which were named on the basis of their chemical content and apparent origin. Actually, the ART-2a network produced much more classes, with each class corresponding to an activated weight vector in the network weight matrix, but only seven classes contained more than 1% of all particles. There were a lot of small classes with very few particles. The seven major classes were considered important and further studied. The numbers of particles falling within each given class were compiled for each time bin and the aerodynamic diameter/number distributions of the particles were plotted versus time.

## RESULTS AND DISCUSSION

ART-2a can be adjusted to yield many or few particle classes, depending on the value of the vigilance parameter. At the extremes, it is possible to see each particle as an individual class or to view all of the particles as belonging to the same class. The vigilance parameter was optimized for the data collected in Long Beach, CA, by using laboratory-generated particles composed of species similar to those detected in the field. Particles made of sodium and potassium chloride with different sodium/potassium ratios were analyzed by ART-2a, and the vigilance parameter was altered until the number of classes found and the members of each of the classes were reasonable given the test data set. It was found that a vigilance parameter of 0.70 was appropriate for analyzing data from particles of marine origin. This result was confirmed when the data collected in Long Beach were analyzed at a vigilance factor of 0.60. Although many of the clusters in this analysis were consolidated, only an additional 3% of the particles were classified into clusters comprising more than 0.5% of all particles analyzed.

Using ART-2a, seven classes of particles were distinguished in the Long Beach data set. They are “unreacted marine”, “reacted marine”, “organic”, “ammonium nitrate I”, “ammonium nitrate II”, “calcium”, and “soil” particles. These particle types are consistent with those manually classified in previous studies of single particles using ATOFMS.[11]

Marine particles constituted 78% of all particles sampled during the time period and were distinguished by peaks representing sodium, potassium and sodium chloride, and sodium nitrate cluster ions (see Figure 1). Particles of marine origin consist mainly of sodium and potassium chloride, but when they are exposed to nitrogen oxides in the atmosphere, a reaction occurs that displaces the chloride ions from the particle and nitrate ions are retained in the particle as sodium nitrate.[12−14] In the positive ion mass spectra, the sodium nitrate product peak is represented by the appearance of mass-to-charge 108 ($Na_2NO_3^+$). As the 108 product peak begins to grow in, the reactant peak at mass-to-charge 81 ($Na_2Cl$) begins to disappear. As expected, the clusters of particles of marine origin were subtly different from time block to time block as the nitrate/chloride displacement reaction took place. Clusters of marine-derived particles with less than 3% of the area of the centroid from the mass 108 product peak were labeled “unreacted marine” while those with a higher intensity 108 peak were named “reacted marine”.

In general, particles of marine origin were found in the diameter range of 1.1−2.5 $\mu$m (Figure 2). As expected, the reacted marine particles were slightly larger on average than the unreacted marine particles. Consistent with our previously reported data, the reacted marine particles were observed to alternate in

(11) Noble, C. A.; Prather, K. A. *Environ. Sci. Technol.* **1996**, *30* 2667−2680.
(12) Allen, H. C.; Laux, J. M.; Vogt, R.; Finlayson-Pitts, B. J.; et al. *J. Phys. Chem.* **1996**, *100*, 6371−6375.
(13) Finlayson-Pitts, B. J.; Pitts, J. N. *Chemistry: Fundamentals and Experimental Techniques*; Wiley: New York, 1986.
(14) Ravishankara, R. A. *Science* **1997** *276*, 1058.

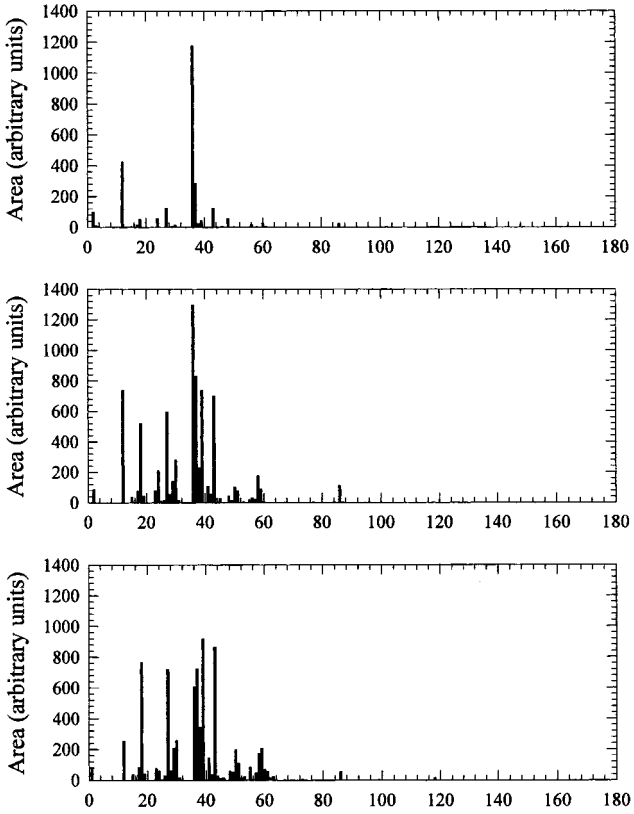


Figure 3. (top) Centroid mass spectra of the organic particle class; (middle) centroid mass spectra of the ammonium nitrate I particle class; (bottom) centroid mass spectra of the ammonium nitrate II particle class.

frequency diurnally, increasing in frequency at approximately 2:00 a.m. and declining around 2:00 p.m. over all three of the days that were analyzed. A more thorough exploration of the reaction between atmospheric nitrogen oxides and particles of marine origin as measured by ATOFMS appears elsewhere.[15]

A class of previously identified particles containing hydrocarbons with or without ammonium nitrate was observed (Figure 3).[11,16] The ammonium nitrate-containing particles are proposed to form as a result of ammonia gas and nitric acid reacting in the air and perhaps settling on the organic particles. Particles of this general variety were divided into three classes on the basis of the intensity of the peaks at $m/z$ 18, $NH_4^+$, and $m/z$ 30, $NO^+$. Classes where either $m/z$ 18 or 30 accounted for less than 5% of the total area of the cluster centroid were assigned to type organic. Classes in which both $m/z$ 18 and 30 accounted for more than 5% each of the total area but neither accounted for more than 30% were assigned to type ammonium nitrate I. Classes where either peak accounted for more than 30% of the total area of were assigned to type ammonium nitrate II. All three types of hydrocarbon particles were found primarily in the aerodynamic diameter range from 0.5 to 1.2 $\mu$m. The organic particles were generally smaller and arrived earlier than the ammonium nitrate I particles, which were generally smaller and arrived earlier than the ammonium nitrate II particles, suggesting the possibility that the process of ammonium nitrate coating organic particles was being

(15) Gard, E. E.; et al. *Science* **1988**, *279*, 1184−1187.
(16) Russell, A. G.; McCue, K. F.; Cass, G. R. *Environ. Sci. Technol.* **1988** *22*, 263−271.

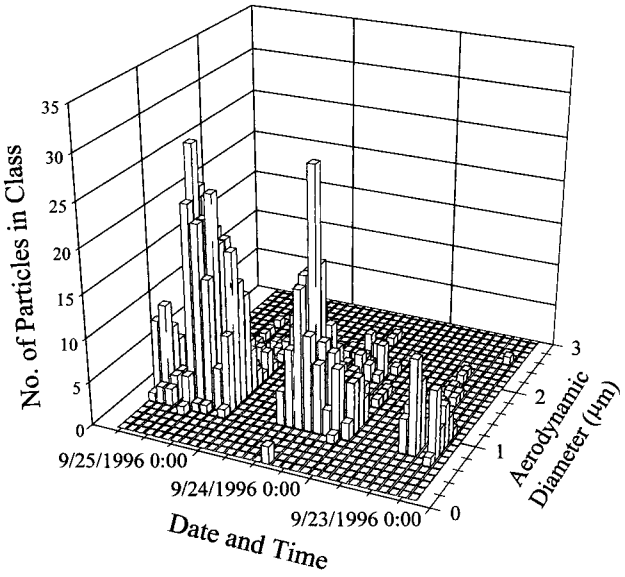


Figure 4. Temporal variations of the size distribution of particle type organic.

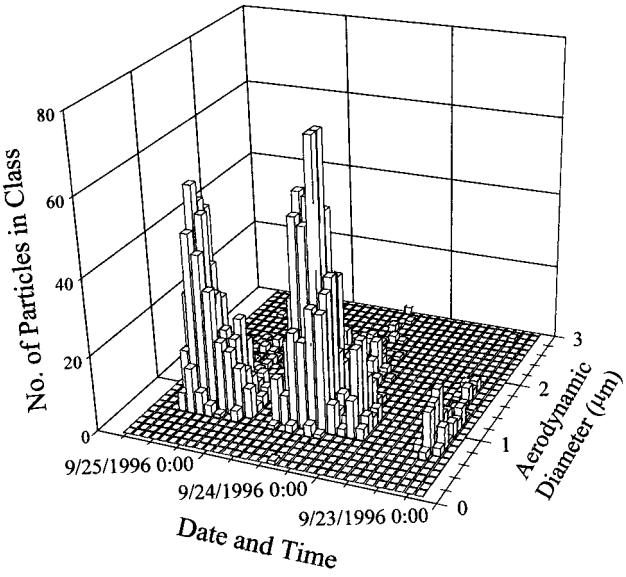


Figure 5. Temporal variations of the size distribution of particle type ammonium nitrate I.

observed (Figures 4−6). Although the three classes of particles only accounted for 10% of all particles observed, at certain times, they accounted for up to 28%.

A class of particles was identified that contained calcium, as evidenced by the $Ca^+$ ions and calcium oxide cluster ions (Figure 7, top). The calcium-containing particles only accounted for <1% of all particles but came primarily during two episodes beginning at 10:00 p.m. and lasting only 2 h on one night and 6 h on the next. Calcium particles were primarily 1.8−2.3 $\mu$m in diameter (Figure 8).

Another class of particles with dominant peaks at either $m/z$ 38 or $m/z$ 39 was identified (Figure 7, bottom). A visual inspection of the particle mass spectra revealed that the breadth of a peak at $m/z$ 40 had caused it to be confused with the two mass-to-charge ratios. Because of their similarity to particles observed

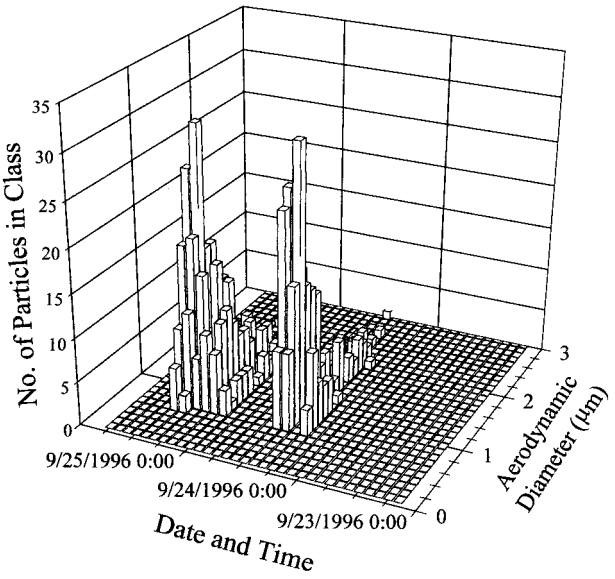


Figure 6. Temporal variations of the size distribution of particle type ammonium nitrate II.

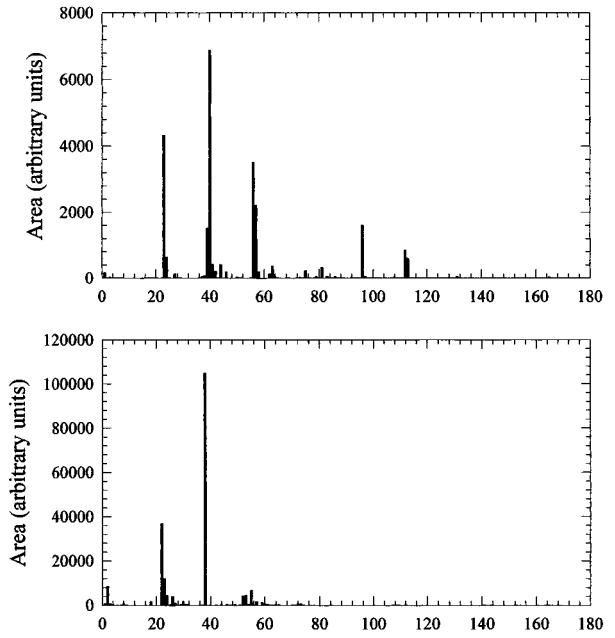


Figure 7. (top) Centroid mass spectra of the calcium particle class; (bottom) centroid mass spectra of the soil particle class.

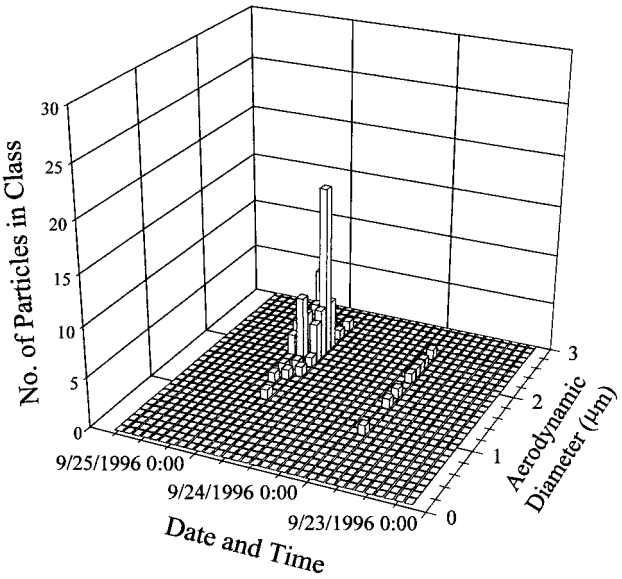


Figure 8. Temporal variations of the size distribution of particle type calcium.

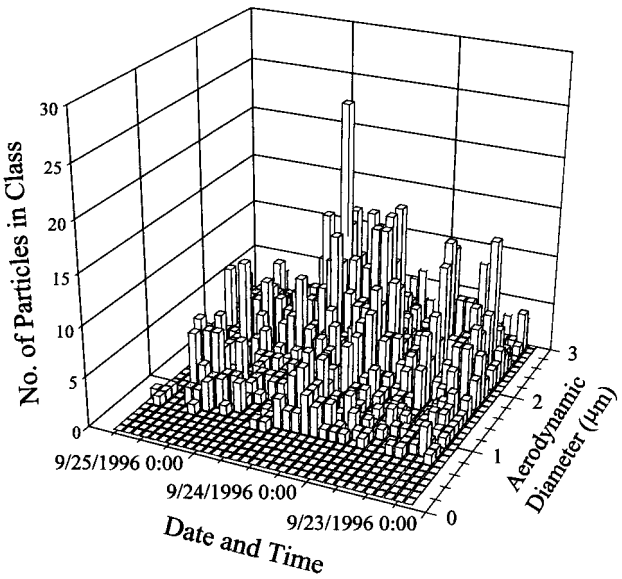


Figure 9. Temporal variations of the size distribution of particle type soil.

previously under controlled circumstances,[17] they were termed soil particles. The soil particles were observed primarily in the large size mode between 1.2 and 2.3 $\mu$m in aerodynamic diameter. (Figure 9). They constituted less than 4% of all of the observed particles.

Any class containing less than 1% of all particles within a given time bin was declared an outlier and was grouped with the other outliers. Overall, these outliers accounted for 5% of all particles. Any class of more than 1% of all particles that was not identifiable by the rules used to classify the other clusters was reported as unclassified and plotted as such. Overall, only 8% of all particles fell into either of these groups.

The observed periodicity and size distributions in Figures 2, 4−6, 8, and 9 lend credence to the particle classifications. Particles determined by ART-2a to be of similar composition were found to possess similar aerodynamic diameters. The fact that some particle classes are associated or are antiassociated in time is representative of sea salt chemistry occurring in the atmosphere over this time period that was also observed in manually classified and analyzed particles in this study.[15]

It is important to note that the relative percentages of particle populations are based strictly on the raw ATOFMS number counts. For this study, because no attempt was made to scale the particles to represent the actual atmospheric populations, the larger sized particle classes (i.e., marine and soil) are favored due to sampling biases of the ATOFMS instruments. To obtain a representative picture of the various particle types at any given time as they actually occur in the atmosphere, scaling factors are being determined from side-by-side measurements made during this study with cascade impactors and real-time sizing equipment.

(17) Silva, P. J.; Carlin, R. A.; Prather, K. A., unpublished.

## CONCLUSION

The ART-2a algorithm is uniquely suited to ATOFMS data because it groups particles in a manner consistent with their occurrence in the atmosphere—as groups of related particles with some degree of variability. The resulting classes are consistent with those manually observed in previous studies. However, the time it takes to arrive at these classes is substantially less and actually offers the potential that it can be done as the data are acquired. The ART-2a neural network is a promising approach to inferring possible particle emission sources and distinguishing between single particle types. The ART-2a network will be applied to a collection of data accumulated by three ATOFMS instruments during field studies in 1996 and 1997 to gain insight into tropospheric chemistry processes in Southern California.

## ACKNOWLEDGMENT

The authors acknowledge Dr. Eric Gard, Dr. Markus Gaelli, and Bradley D. Morrical for assisting with the collection of the data. The advice of Don-Yuan Liu, Christopher A. Noble, and Philip J. Silva was invaluable in identifying the particle classes. The work at Clarkson University was supported by the U.S. Environmental Protection Agency through Grant R822482 and by the California Air Resources Board through Contract 97-321. The work at the University of California, Riverside, was supported by the California Air Resources Board through Contracts 95-305 and 96-307 and the National Renewable Energy Laboratory through Contract ACI-17075-01.

Received for review August 27, 1998. Accepted December 4, 1998.

AC9809682