Tellus (1992), *44B*, 581–592

A statistical examination of the chemical differences between interstitial and scavenged aerosol

By KEVIN J. NOONE[1], JOHN A. OGREN[2], ANNELI HALLBERG*, *Department of Meteorology, Stockholm University, S-106 91 Stockholm, Sweden*; HANS-CHRISTEN HANSSON, ALFRED WIEDENSOHLER and ERIK SWIETLICKI, *Department of Nuclear Physics, Lund University, Sölvegatan 14, S-223 62 Lund, Sweden*

(Manuscript received 26 November 1991; in final form 15 July 1992)

ABSTRACT

The difference in chemistry between interstitial aerosol particles and particles that were scavenged into fog droplets is examined using multivariate statistical techniques. 15 trace elements (P, S, Cl, K, Ca, Cr, Mn, Fe, Ni, Cu, Zn, Se, Br, Pb, EC) were used in the analysis. There was a significant difference in composition between the two types of particles. S, Fe, Mn, and Cu were among the elements best describing the scavenged aerosol, while the interstitial aerosol was best described by elemental carbon (EC).

1. Introduction

Nucleation scavenging in clouds and fogs leads to a partitioning of material into two reservoirs: the cloud droplet reservoir and the interstitial reservoir. This partitioning has consequences both in terms of cloud microphysics and cloud chemistry (Twomey, 1977). The number of aerosol particles scavenged will determine the initial number of cloud droplets present in the cloud. The mass of aerosol scavenged will determine the initial concentration of material in the cloud drops. The chemical composition of the scavenged aerosol will determine, to a large extent, the initial chemical composition of the cloud droplets.

It is well known that various chemical species are distributed in different ways over the atmospheric aerosol (Lannefors et al., 1983; Hasan and Dzubay, 1987; Heintzenberg and Covert, 1987). Some species may be present primarily in the sub-micrometer aerosol, while other species are primarily in the super-micrometer aerosol. The chemical composition of the aerosol can vary even within a given size range (i.e., the accumulation mode or the coarse mode [Infante and Acosta, 1991]).

In most cases, not all of the aerosol present in the atmosphere is scavenged into cloud droplets. Both the number and mass scavenging fractions vary widely between different cloud types and for different aerosol loadings. The scavenging fractions will depend upon particle size and composition; usually larger particles are scavenged more efficiently than smaller ones (Noone et al., 1992). We can define scavenging fraction as the amount of aerosol material incorporated in the cloud droplets divided by the sum of the interstitial aerosol and the material in the droplets.

The combination of non-uniform aerosol composition with size and a scavenging efficiency that is also a function of particle size can lead to different species being scavenged in clouds with varying efficiencies. Some species may be found preferentially in the interstitial reservoir, while others may be found primarily in the cloud droplet reservoir.

A consequence non-uniform scavenging in clouds would be that it cannot be assumed that

Present addresses: [1] Center for Atmospheric Chemistry Studies, School of Oceanography, University of Rhode Island, Narragansett, RI 02882-1197, USA. [2]NOAA/CMDL R/E/CG, 325 Broadway, Boulder, CO 80303-3328, USA.

* Contribution No. 668.

composition of the aqueous phase in a cloud is a simple dilution of the bulk pre-cloud aerosol. Here we use *aqueous phase* to refer to the activated cloud droplets. This is not strictly true; even unactivated particles contain some water. A droplet is said to be activated if it has grown beyond its critical radius (a function of the mass and composition of the aerosol particle upon which the droplet formed and the supersaturation in the fog or cloud (Köhler, 1921)). The major fraction of the liquid water in a cloud, however, is associated with activated droplets.

There are several ways of investigating these potential differences in scavenging. One is by examining the mass fraction of different species before and after cloud formation both in the interstitial aerosol and the cloud droplets. Changes in mass fractions would indicate that the various species are scavenged with different efficiencies. This approach is taken in a companion paper (Hallberg et al., 1992). Another approach is to use multivariate statistical techniques to examine the difference in chemistry between the two reservoirs. In this case, the cloud system is being allowed to "speak for itself" in the sense that what is known are the chemical concentrations of certain elements for a given sample and that some samples came from the cloud droplet reservoir and other samples came from the interstitial aerosol reservoir. The statistical techniques are used to determine whether these two reservoirs had different compositions, and which species were responsible for the difference.

The purpose of this paper is to test the hypothesis that there is a difference in composition between the interstitial aerosol and the aerosol that is scavenged into cloud droplets. We shall examine three specific questions: (1) Can this postulated difference be discerned using statistical methods?; (2) Is this difference statistically significant?; (3) Which elements can be attributed to the two reservoirs?

2. Techniques and methods

2.1. Description of sampling techniques

The experiment described here was part of the EUROTRAC project *Ground-based cloud experiments* (GCE). It took place in November 1989 in the Po Valley near San Pietro Capofiume, Italy. Five fog episodes were sampled during the course of the experiment. A detailed description of the site and an overview of the experiment itself are presented in Fuzzi et al., this issue.

The data presented in this paper are samples of the interstitial aerosol and the material that was scavenged into fog droplets. The interstitial and cloud droplet reservoirs were defined in terms of size. Interstitial particles are defined as having sizes below 5 μm diameter, while cloud droplets are defined as having sizes larger than 5 μm diameter.

2.2. Interstitial inlet

The interstitial inlet used an inertial impactor to remove droplets larger than 5 μm diameter (Martinsson et al., this issue). Particles smaller than this size were pulled through a 25 mm OD polyethylene tube into a sampling container. Once in the container, the interstitial sample air stream was distributed through a plenum to various instruments. Interstitial particles were collected on Nuclepore filters (25 mm, 0.4 μm pore size) for subsequent analysis. Because the temperature in the container was usually 10–20°C higher than the ambient temperature, the particles will have been dried before they were collected on the filter.

The interstitial inlet was also used to sample aerosol outside fog. When fog was not present, the inlet sampled essentially all of the aerosol.

2.3. Cloud droplet inlet

Cloud droplets were sampled using the Counterflow Virtual Impactor (CVI: Ogren et al., 1985; Noone et al., 1988). The CVI samples cloud droplets larger than a given size; nominally 5 μm in the Po Valley 1989 experiment. Once the droplets are sampled by the CVI, they are evaporated. The liquid water and volatile material in the droplets are driven into the gas phase. The non-volatile material in the droplets remains behind as a *residual aerosol particle*. It is assumed that a single cloud droplet, when evaporated, leaves behind a single aerosol particle. These *residual* particles were also collected on Nuclepore filters.

A schematic diagram of the sampling system is shown in Fig. 1. Filter samples were taken in pairs (interstitial and cloud) on a one-hour basis in fog.

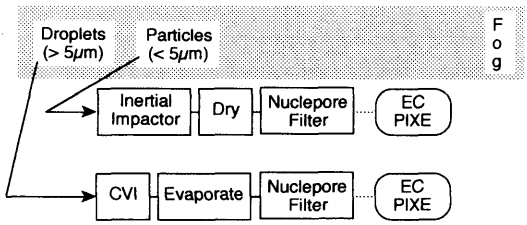

Fig. 1. Schematic diagram of sampling system. Fog droplets were defined as being >5 μm, while interstitial particles were defined as being <5 μm diameter.

2.4. Chemical analyses

After the field campaign, the aerosol samples were analyzed for elemental carbon (EC) and for elemental composition. A light absorption technique was used to estimate the EC concentration on the filters (Clarke et al., 1986; Heintzenberg, 1988). Particle Induced X-ray Emission (PIXE) was used to measure the elemental composition of the samples. Detection limits for PIXE analysis of aerosol samples on various substrates are discussed in Swietlicki (1989). The detection limit varies smoothly with atomic number and shows a minimum for elements with atomic numbers around 30.

While PIXE is capable of detecting elements with atomic numbers greater than 13 (Al), many of these elements were below detection limits for the samples discussed here. All elements where more than 10% of the samples were below detection limits were discarded. The following 14 elements (in addition to EC) were used in the analysis: P, S, Cl, K, Ca, Cr, Mn, Fe, Ni, Cu, Zn, Se, Br, and Pb.

2.5. Description of statistical techniques

Interesting structures and relationships in the data can often be discerned when the covariations between all the measured variables are taken into account simultaneously in a multivariate statistical analysis, rather than simply examining the univariate or bivariate properties of the variables (Green and Carroll, 1976; Afifi and Azen, 1979; Sharaf et al., 1986). For this reason, several multivariate statistical techniques were used to examine the cloud/interstitial aerosol samples. Two different data normalizations/transformations were used in the analysis (these are described in detail below). Normalization transformation of the data is necessary for a number of reasons. The ambient concentrations of the various elements varied from micrograms per cubic meter (e.g., S, EC) to ng/m^3 (e.g., Ni, Se). These pronounced differences in absolute concentrations also imply that the various elements will be given unequal weights in the subsequent statistical analysis, which is undesirable, since the information contained in the variations of the trace elements (with small concentrations) will then be suppressed by the much larger variations in the major aerosol components (having higher concentrations). Meteorological variability during sampling can also mask the inter-elemental relationships sought after, since it induces correlations in the data which are simply due to similarities in the dispersion and deposition of the elements during transportation from the sources of the aerosol to the receptor. In addition, some statistical tests are sensitive to the distribution of the data.

Elemental concentrations in the atmospheric aerosol are often non-normally distributed. Such distributions may influence the results of the analyses by giving high concentration samples undue weight. The statistical tests that were used here are not as sensitive to data distributions as are more "classical" tests (e.g., ANOVA) because multivariate solutions (scores, to be defined later) are usually more normally distributed than the variables themselves.

Three multivariate techniques will be discussed. They are principal components analysis (PCA), soft independent modelling of class analysis (SIMCA), and partial least-squares regression (PLS).

In the following discussion, the term sample will be used to denote a filter sample (e.g., interstitial filter sample #16), and the term variable will be used when referring to an element (e.g., S, EC). The data set can be seen as a matrix of m rows and n columns where the rows correspond to the different samples and the columns correspond to the different elements (variables). Each data point can be located by two indices: i and j. Here, x_{ij} denotes the concentration of variable (element) j in sample i. Both cloud and interstitial samples are included in the matrix; each element (column in the data matrix) contains both cloud and interstitial samples. The data matrix in this case is 15 columns (elements) by 66-rows (samples).

2.6. Normalization/transformation

The two different normalization/transformation techniques used will be referred to as the Lund normalization and the log normalization.

The Lund normalization is a three-step procedure (Hansson et al., 1984). First, the sample concentration is divided by the mean of each variable taken over all of the samples. Now, the distributions of the variables will be centered around 1. The 2nd step involves calculating the sum of the variables for each sample, and dividing the variables for the sample by this sum. The initial step of the normalization procedure ensures that the elements that are present in high concentrations will not dominate in the calculation of the elemental fractions. The second step has the effect of greatly reducing the influence of the specific meteorological history of each sample on the variability in the data. This is done by removing the direction in the n-dimensional measurement space in which all variables increase. The dimensionality of the data set is thus decreased by one. These two steps are shown in eqs. (1a) and (1b):

$$x'_{ij} = \frac{x_{ij}}{(1/m)\sum_{i=1}^{m} x_{ij}} \tag{1a}$$

$$x''_{ij} = \frac{x'_{ij}}{\sum_{j=1}^{n} x'_{ij}}. \tag{1b}$$

Here, x''_{ij} is the normalized data point. The last step in this procedure is to standardize the data: subtract the column (variable) mean (m''_j) of the normalized data and divide by the column standard deviation (s''_j). This transforms the data into z-scores (eq. 2); the subsequent analyses are performed on these z-scores:

$$z_{ij} = \frac{x''_{ij} - m''_j}{s''_j}. \tag{2}$$

A normalization of this kind can lead to closure problems. Johansson et al. (1984) suggest a method for minimizing closure problems. They state that the means and standard deviations of the variables should approximately equal, and that the number of variables in the data set be large. It will be seen in the comparison of the normalization schemes that closure did not seem to be a serious problem for these data.

The log transformation is also a three-step procedure. First, each data point is divided by the weighted mass on the filter (M_i). This step is similar to step (2) in the Lund normalization; it is intended to reduce the correlation in the data due to meteorological variations. Here the dimensionality of the data set is not reduced, since the weighed mass is an additional measurement. The next step is to log-transform the mass-normalized data. The original data distributions were somewhat skewed; the log-transformation makes the distributions more normal. These two steps are illustrated in eq. (3):

$$x'_{ij} = \log\left[\frac{x_{ij}}{M_i}\right]. \tag{3}$$

The third step is the same as in the Lund normalization; standardization of the mass-normalized, log-transformed data (eq. (2)).

2.7. *Principal components analysis*

We can think of each of the variables in our data set as a column vector with length m. We call the vector of the concentrations of a particular element (e.g., P) X_j. The data set can be thought of as a collection of n vectors $X_1, X_2, ..., X_n$, each vector representing a dimension in a n-dimensional space. The interdependency of the variables can be measured by the correlations and covariances between the X_j's. Principal components analysis (PCA) attempts to find linear combinations of the original data that can explain the variation in the original data set with a fewer number of dimensions. If the original data matrix had n columns (dimensions), then it is possible that most of the variation in the data could be explained by p principal components, where $p < n$.

Initially, the method finds n linear combinations

$$Y_1 = \sum_{j=1}^{n} \alpha_{1j} X_j, ..., Y_n = \sum_{j=1}^{n} \alpha_{nj} X_j, \tag{4}$$

such that the new vectors (Y_j's) are orthogonal ($\mathrm{cov}[Y_j, Y_k] = 0$ for $j, k = 1, ..., n$, $j \neq k$). The terms α in eq. (4) are the coefficients relating the original X data vectors to the new principal components (Y vectors). The new vectors are ranked by the amount of variance in the data set they account for: $V(Y_1) \geq V(Y_2) \geq \cdots \geq V(Y_n)$. In addition, the total amount of variance is conserved after the linear combinations are computed (if $p = n$):

$$\sum_{j=1}^{n} V(Y_j) = \sum_{j=1}^{n} \sigma_{jj}^2. \tag{5}$$

The left-hand side of eq. (5) is the sum of the variance for the new vectors and the right hand side is the total variance in the original data set. Normally, the first few principal components will explain most of the variance in the original data set. The PCA routine in the statistical software package SYSTAT (Wilkinson, 1987) was used here. PCA will be used to establish whether there is a difference in the chemistry of the scavenged and interstitial aerosol as reflected in the measured elements. If such a distinct difference exists, it will be clearly visualized when the data points are projected onto the plane spanned by the first two PCs, which is the 2-dimensional projection retaining most of the original variance in the data set.

2.8. *SIMCA analysis*

In our data set, we know that the samples come from one of two possible classes: the cloud class (scavenged aerosol particles) or the interstitial class. In the SIMCA method (Wold and Sjöström, 1977, Sharaf et al., 1986), separate principal components models are fitted to each class and confidence envelopes are constructed around each model to contain the data points belonging to each class. The distance in measurement space of each data point to the various class models can then be used to determine whether the points correspond to a particular class, to both classes, or to neither class. Strictly speaking, we do not specifically test for a Type II error (Mendenhall et al., 1981). While we can conclude that a sample *is not different from a particular class* (*at a certain confidence level*), we do not specifically test for class membership. In terms of class membership, we infer that a sample is a member of a class by concluding that it *is not different from that class.* The confidence envelopes also allow a statistical test to be performed showing whether the two class models are significantly different. In the following discussion, i is the sample index, j is the variable index, and k is the principal component index.

With SIMCA, a data point in class q can be expressed as

$$x_{ij} = \bar{x}_j + \sum_{k=1}^{NC} v_{jk}\, y_{ik} + e_{ij}, \tag{6}$$

where $\bar{x}_j$ is the class mean for variable j; v_{jk} are the loadings of the NC principal components used in the class model; y_{ik} are the component scores within the class model; and e_{ij} is the residual between the datum and its model description. Here, the *loadings* are the correlation coefficients between the variables and the principal components. The scores are the strength of that component in any given sample. The residual variance for class q (a measure of the dispersion of the data around the class model) can be written as

$$s_0^2 = \sum_{i=1}^{NS} \sum_{j=1}^{NV} \frac{(e_{ij})^2}{2(NS - NC - m)(NV - NC)}. \tag{7}$$

In the above equation, NS is the number of samples, NV is the number of variables, NC is the number of principal components, and m accounts for the proper number of degrees of freedom. For the Lund normalization, $m = 2$ while for the log normalization, $m = 1$.

The residual variance of each sample i about class model q is given by (Sharaf et al., 1986)

$$s_i^2 = \sum_{j=1}^{NV} \frac{(e_{ij})^2}{(NV - NC)}. \tag{8}$$

Using these two variances, an F-test can be performed on each data point to check whether it belongs to a class or whether it is an outlier to the class model.

2.9. *Partial least squares regression*

Partial least-squares regression (PLS) is a technique for determining the relationships between blocks of data (data matrices) (Hansson et al., 1984; Lorber et al., 1987; Vong et al., 1988). As such, it is similar to multiple linear regression (MLR) or principal components regression (PCR). Like these techniques, PLS uses one block of data (the X-block, or data matrix) to predict another block of data (the Y-block, or the dependent variables). To obtain the predictions, the model must first be calibrated using a training data set, which contains data in both the X- and Y-blocks. The calibration is done by regressing the extracted components of the X-block on the components of the Y-block. A feature unique to PLS is that the components of each block are not computed independently. PLS exchanges information between the blocks of data during the computation of the components such that the predictive capability of the X-block to the Y-block is maximized (Geladi and Kowalski, 1986). In the case

discussed here, PLS will be used to discriminate between two classes of data: the interstitial and cloud reservoirs. Two-class discriminant PLS uses a single variable in the Y-block which is given one of two possible values (e.g., 0 and 1). The Y-block then denotes the class membership of each sample. For the analysis presented here, the primary difference between PCA and PLS is that the projection of the data points onto the plane spanned by the 1st 2 PCs. This projection shows the two-dimensional score plot that explains most of the variance in the whole data set (interstitial + cloud). The corresponding two-dimensional discriminant PLS score plot will show the projections of the data points onto the plane that shows the maximum separation between the two classes of data.

3. Results

Table 1 shows the average concentrations and standard deviations of the measured elements for the interstitial and cloud reservoirs. All concentrations are given in ng m^{-3} (STP: 1 atm, 273 K). The cloud and interstitial data come from 33 samples each. The dominant elements in both the cloud and interstitial reservoirs are S and EC. They are not present in the same ratio in both reservoirs; their scavenging characteristics were different for the fogs. The scavenging efficiency of aerosols in the fog events are discussed elsewhere (Noone et al., 1992; Hallberg et al., 1992).

Table 1. *Aerosol concentrations*

Element	Cloud avg.	Cloud SD	Inter. avg.	Inter. SD
P	35.8	15.0	227.6	117.3
S	320.6	172.1	1800.7	783.2
Cl	17.1	9.0	80.5	57.9
K	31.5	43.5	441.4	706.1
Ca	18.0	33.0	31.2	40.1
Cr	2.6	1.1	17.2	6.5
Mn	1.9	1.2	85.6	305.8
Fe	19.7	13.5	150.5	302.0
Ni	0.5	0.3	6.2	5.9
Cu	1.8	2.2	12.1	29.0
Zn	14.8	11.9	63.3	63.0
Se	0.3	0.2	3.6	1.4
Br	2.6	0.8	29.2	14.2
Pb	9.4	6.5	93.2	43.9
EC	137.8	100.6	3172.1	1995.8

Aerosol concentrations and standard deviations for the cloud and interstitial reservoirs. Also shown are the out-of-fog concentrations and standard deviations for comparison. All values are ng m^{-3} at STP.

3.1. *Principal components analysis*

A principal components analysis was carried out on the 66 cloud and interstitial samples using the Lund normalization. The first PCA result indicated that 5 of the samples (2 interstitial and 3 cloud; none were paired samples) were outliers, these had anomalously high Ca values. While the reason for the high Ca values was unknown, these values were removed. The PCA (and subsequent SIMCA) analysis is influenced by outliers. The analysis was then repeated on the data set, now with 61 samples.

The second PCA yielded 4 components with eigenvalues greater than unity, i.e., each of the four PCs explained more variance than a single original variable. The 1st 2 PCs explained 37.4% and 20.3% of the variance in the original data, respectively. Examination of the loadings for the components showed that the last 2 were single-element components; only one element had a high loading on these components. Component 3 had a large loading only for Pb, while component 4 contained only Ni. A VARIMAX rotation was then used, and only the first two of the PCs rotated. The loadings for the 2 rotated PCs are shown in Table 2.

Table 2. *VARIMAX-rotated principal component loadings; these two components explained 58% of the variance in the data before rotation*

Element	PC#1	PC#2
P	0.4489	0.8231
S	0.7941	0.3545
Cl	0.4377	0.7799
K	−0.1745	−0.2517
Ca	0.7071	0.0188
Cr	0.2842	0.7805
Mn	0.7128	0.1444
Fe	0.9539	−0.0763
Ni	−0.0170	0.3524
Cu	0.6666	−0.0266
Zn	0.8399	0.0749
Se	−0.3419	0.7857
Br	−0.3092	0.8130
Pb	−0.1701	−0.2453
EC	−0.7606	−0.3580

These 2 components define a plane in the 15-dimensional space defined by the original elements. The projections of the data points onto this plane can be examined to see whether there are interpretable groupings of the data. In addition, each element can be located on the plot by its loadings. Such a plot (called a scores/loadings plot) is shown in Fig. 2.

There is an obvious separation samples from the two reservoirs evident in the figure. The interstitial samples all have negative values on PC#1, while all of the cloud samples have positive values. A clear difference in the chemistry of these 2 groups (at least in terms of the measured elements) is apparent. The cloud samples show a larger spread in the plot than do the interstitial samples. The scavenged aerosol was apparently slightly more variable in composition than the aerosol not scavenged by the fogs.

The elements loading on PC#1 best describe the difference *between* the 2 groups, while the elements loading on PC#2 seem to describe more the variation *within* the groups. Elemental carbon appears in the middle of the interstitial grouping. EC is clearly a specific component of the interstitial aerosol, as it is found in higher concentrations relative to the other measured elements in the interstitial aerosol than in the cloud reservoir. Sulfur is located near the center of the cloud grouping, along with several other elements. It is interesting to note that the elements Fe, Mn, and Cu are also found within the cloud grouping. These are trace metals that can catalyze the aqueous-phase oxidation of SO_2 to $SO_4^=$. These elements are characteristic of the cloud grouping. The separation of EC and S into different reservoirs is supported by the results of Hallberg et al. (1992), who show that $SO_4^=$ was scavenged more efficiently than EC in the fogs.

It appears from the PCA that there is a chemical difference between the scavenged and interstitial aerosol. We can now test whether this difference is statistically significant.

3.2. *SIMCA analysis*

A principal components model for each of the two groups was calculated with the SIMCA method. The distance of each sample to the two

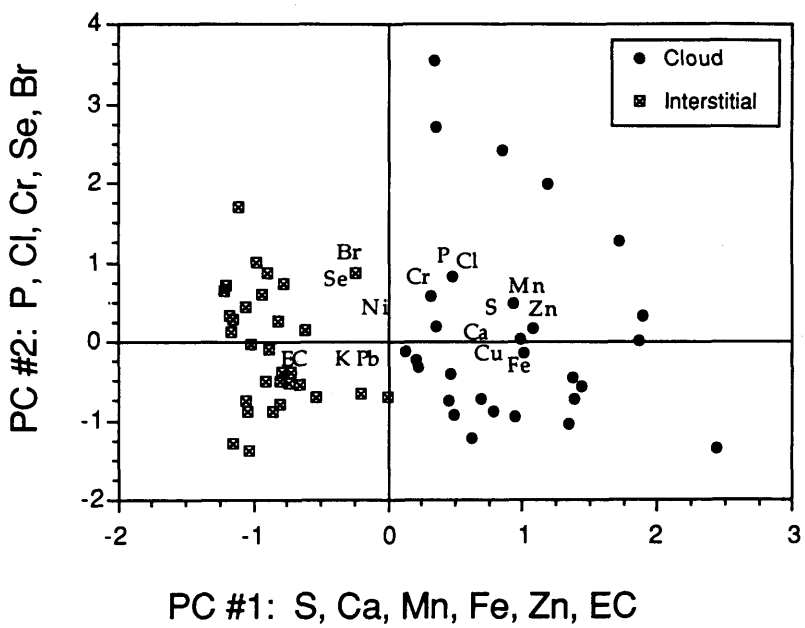

Fig. 2. Principal components plot. The 2 PCs explain 37.4% and 20.3% of the variance in the original data, respectively. Both scores and loadings are plotted. The cloud sample scores are plotted as solid circles, the interstitial sample scores as crossed squares. The elemental loadings are plotted with the element ID.

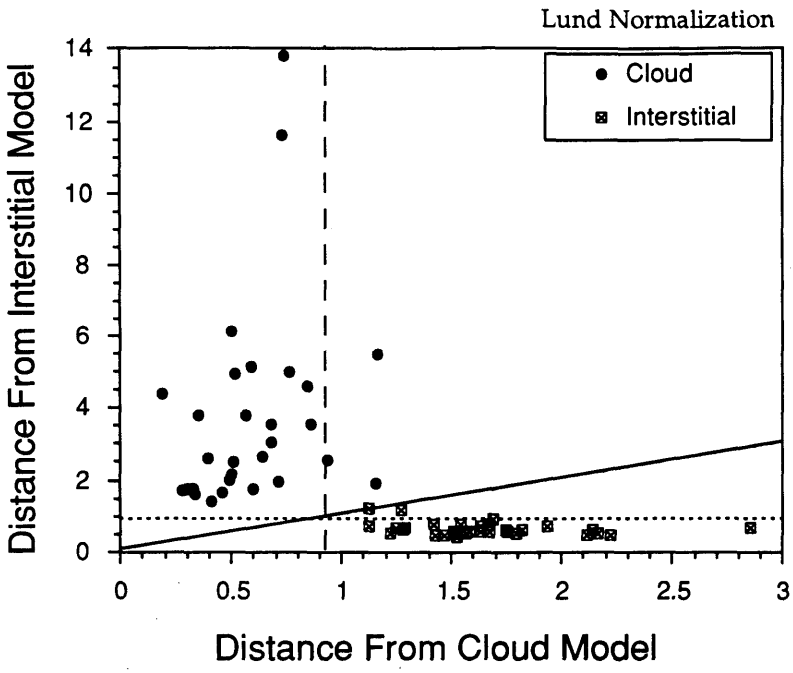

Fig. 3. SIMCA distance plot (Lund normalization). The distances of each data point to the SIMCA class model are shown. The vertical dashed and horizontal dotted lines are the 95% confidence limits for the cloud and interstitial models, respectively.

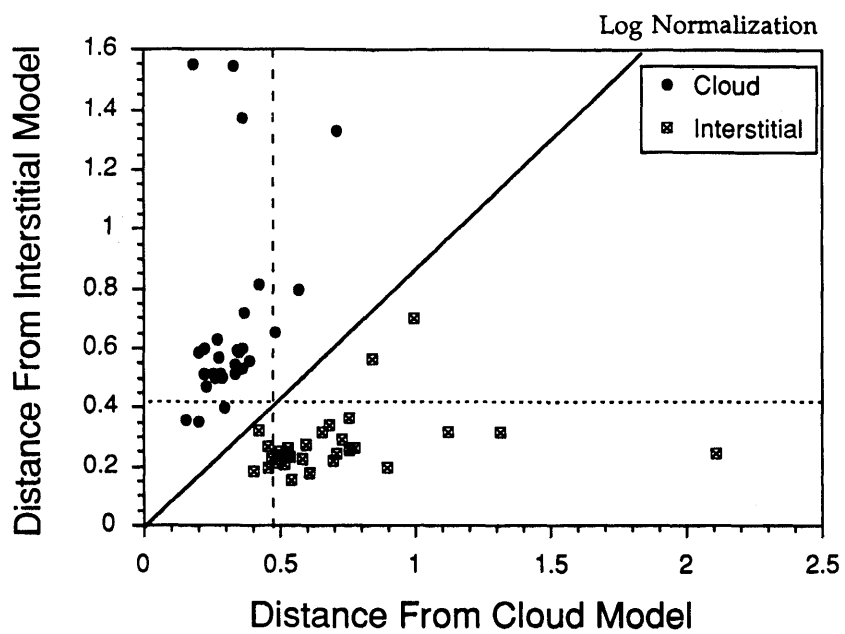

Fig. 4. SIMCA distance plot (log normalization). Similar to Fig. 3. Here, log-normalized data have been used.

models was then calculated. Additionally, a confidence limit for each model was calculated.

The results from the SIMCA classification are presented in Fig. 3 as a distance plot. The horizontal axis in Fig. 3 is the distance from the cloud model. Samples belonging to the cloud reservoir should have low values on this axis. The 95% confidence limit for the cloud model is plotted as the vertical dashed line. A sample actually belonging to the cloud model group that falls on this line has a 5% probability of being classified as not belonging to the group (type I error). The vertical axis in Fig. 3 is the distance from the interstitial model. Samples belonging to the interstitial reservoir should have low values on this axis. The 95% confidence limit for the interstitial model is plotted as the horizontal dotted line.

The two confidence limit lines divide the plot into 4 quadrants. The lower left quadrant is within the confidence limits for both the cloud and interstitial models. Points in this quadrant correspond to samples that could be classified as not being different from both the cloud and interstitial reservoirs. The lower-right quadrant is within the interstitial confidence limit, but outside the cloud model limit. Points in this quadrant correspond to samples classified exclusively as interstitial. The upper-left quadrant is outside the interstitial model limit, but within the cloud model confidence interval. Points in this area correspond to samples classified exclusively as cloud. The largest quadrant, at the upper right, is outside both the cloud and interstitial model limits. Points in this area cannot be classified by either model. However, since the type I error probability is 5%, a few samples are usually expected to fall immediately outside the confidence limits.

There are no points in the lower left quadrant. Also, none of the interstitial samples are found within the cloud model limits, and none of the cloud samples fall within the interstitial model limits. There are only five of the 61 samples in the upper right quadrant, outside the limits of both models. These samples cannot be classified at the 95% confidence interval. These samples represent only 8% of the data. There is then a clear difference in the two classes.

It is important to check whether the results are sensitive to the normalization and the distributions of the data. A parallel analysis was therefore carried out with the log-normalized data. Similar separation of the two reservoirs was evident on the PC score plot for these data, although the loadings for the two components were somewhat different.

SIMCA was used to calculate class models for the log-normalized data. A distance plot from this analysis is shown in Fig. 4 (note the difference in scale from Fig. 3). A similar picture is found when comparing the SIMCA results using the Lund-normalized data with the results using the log-normalized data. There are again 5 samples outside both model's confidence limits; these cannot be classified. Here, however, there are 8 samples that fall within the confidence limits of both models. They can be classified as not being different from either of the cloud or interstitial categories. The separation between classes is not as clear using the log transformation as compared to the Lund normalization. While the class separation is not as well defined, it is still clear that there is a substantial difference between the two reservoirs.

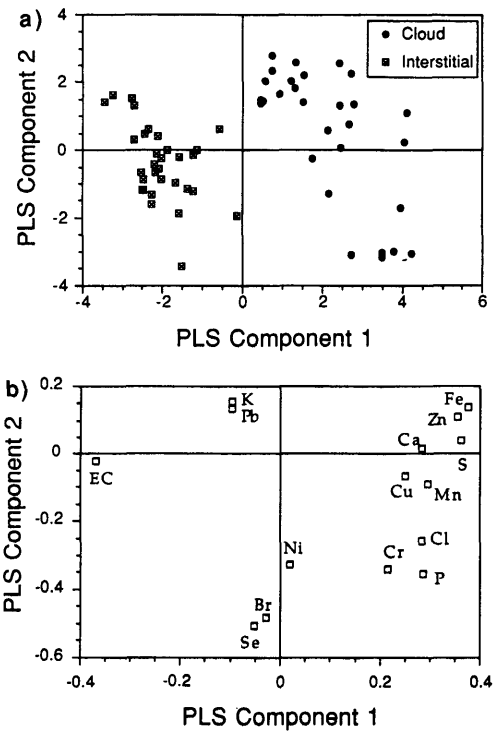

Fig. 5. (a) PLS scores plot showing the separation between the cloud and interstitial samples; (b) loadings plot. Overlaying (b) on (a) shows which elements are best in modelling the two different aerosol reservoirs.

3.3. PLS analysis

The PCA analysis showed that EC was a characteristic element of the interstitial reservoir, while S was one of the important elements for the cloud reservoir. Discriminant PLS regression will more clearly indicate which variables are best in modelling the difference between the reservoirs.

The results from the PLS analysis are presented in Fig. 5. As expected, both the PLS scores and loadings plots are very similar to the principal components plot shown in Fig. 2. Fig. 5a is a PLS components scores plot. In contrast to the PCA, the PLS components are computed with the additional constraint that they also maximize the separation between the two groups of data. Again, the separation between the two reservoirs is clearly seen. Fig. 5b is a loadings plot for the PLS components. It shows that EC is found to be enriched in relation to the other elements in the interstitial reservoir. The same is true for S in the cloud reservoir. Again, Fe, Mn, and Cu are associated with the cloud reservoir.

3.4. Analysis of paired data

Because simultaneous interstitial and cloud samples were taken, we can take advantage of the paired nature of the data to examine the covariation of individual elements in both phases. In this instance, the data matrix has separate columns for the elements in each phase. For instance, interstitial and cloud Pb are separate columns in this data matrix. Here, the matrix is 30 columns (elements in each phase) by 27 rows (samples, outliers were removed). In the prior analysis, the question was whether the cloud and interstitial were different in terms of the chemical elements we analyzed. In this analysis, we want to determine which elements behave similarly in the 2 reservoirs and which behave differently.

A principal components analysis was performed using the Lund-normalized data and the SIMCA statistical package. A principal components plot of the results is shown in Fig. 6. It is a plot of the 2nd and 3rd components. This plot was chosen (rather than the 1st 2 PCs) because the projection of the

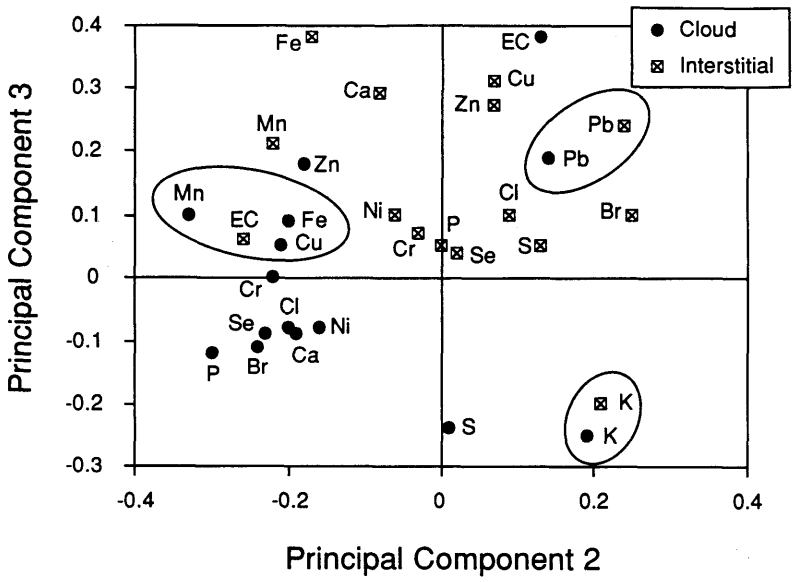

Fig. 6. Principal components plot for paired data. Plotted are the second and third principal components for the paired data. The circled groups are more obvious when plotted in three dimensions (PC 1, 2, and 3). Lead and potassium tend to vary in the same manner in both the interstitial and cloud reservoirs. Interstitial EC and cloud Mn, Cu and Fe vary in similar fashions.

data onto this particular plane best shows the groupings of the elements. There are a few interesting groupings on this plot which have been circled. These groupings become more evident when plotted in three dimensions; the projection shown here is not completely indicative of the proximity of the points in three dimensions. Potassium (lower-right quadrant) varies similarly in both phases, as does lead (upper-right quadrant). These 2 elements had similar scavenging fractions (9% for K and 11% for Pb; Hallberg et al., 1992). However, other elements with similar scavenging fractions (e.g., Ni, Br) do not behave in the same way. Apparently, this covariation is not a simple function of scavenging efficiency. There are also some interesting groupings of different elements evident in Fig. 6. Iron, manganese and copper in the cloud reservoir covary with interstitial EC. This grouping indicates that when the concentration of interstitial EC increased or decreased, so did the concentrations of cloud Fe, Cu, and Mn. Such groupings in the data often indicate a common source for the elements. Since we have no source profiles with which to compare, we can not draw any conclusions about possible sources. If we speculate, however, that EC, Fe, Cu, and Mn had a common source, it is interesting that the three metals tend to be found in the cloud reservoir and that EC tends to be found in the interstitial reservoir. Both particle size and composition were important factors in determining particle scavenging (Noone et al., 1992; Martinsson et al., 1992). It seems that either these elements had different size distributions (EC was associated with the smaller particles, Fe, Cu, Mn with the larger ones) or EC was associated primarily with hydrophobic particles while Fe, Cu, and Mn were associated with the hygroscopic particles.

4. Conclusions

We have investigated the difference in chemical composition (with respect to 15 chemical elements) between particles scavenged into fog droplets and those left behind in the interstitial air. Several multivariate statistical methods were used to examine the cloud/interstitial aerosol system. The following conclusions about the system can be drawn from the results presented here:

(1) there is a statistically significant difference in composition between the scavenged and non-scavenged aerosol;

(2) elemental carbon is the element that is found to be most enriched in the interstitial reservoir in relation to the other elements;

(3) sulfur, iron, manganese and copper are among the elements found to be most enriched in the scavenged reservoir in relation to the other elements;

(4) comparison of SIMCA results using two different data normalization/transformation procedures shows that a better separation between classes was obtained using the Lund transformation.

There are a few interesting speculations that arise from these conclusions. Elemental carbon in this polluted location apparently prefers to be in the interstitial aerosol. The fact the EC was found primarily in the interstitial aerosol would tend to increase its atmospheric lifetime with respect to (for instance) sulfur, which is found to be associated with fog droplets. Additionally, the radiative properties of these fogs may depend on whether or not EC (found in relatively high concentrations) was associated with the droplets.

It is interesting to note that the elements (Fe, Mn, Cu) that can catalyze aqueous-phase oxidation of S(IV) to S(VI) are associated with the fog droplets. Their solubility and oxidation states must be known, however, before an estimate of their influence on oxidation rates can be made.

5. Acknowledgements

We are indebted to Richard Vong for his help with the PLS analysis of the paired data and his comments on the paper. The invaluable contribution of A. Correggiari, S. Miserocchi, D. Correggiari, C. Foschini and L. Tarozzi during the field work is gratefully acknowledged. L. Bäcklin, S. Å. Odh and S. Moritz helped prepare the CVI equipment. Funding for the experiment was supplied by the Swedish Environmental Protection Board, and the CNR-Progetto Strategico Deposizioni Atmosferiche Acide. The station of S. Pietro Capofiume is managed by Assessorato Agricoltura, Regione Emilia Romagna.

REFERENCES

Afifi, A. A. and Azen, S. P. 1979. *Statistical analysis: a computer-oriented approach.* New York: Academic Press.

Clarke, A. D., Noone, K., Heintzenberg, J., Warren, S. and Covert, D. S. 1986. Aerosol light absorption measurement techniques: analysis and intercomparisons. *Atmos. Environ. 21*, 1455–1465.

Geladi, P. and Kowalski, B. R. 1986. Partial least-squares regression: A tutorial. *Anal. Chim. Acta 185*, 1–17.

Green, P. E. and Carroll, J. D. 1976. *Mathematical tools for applied multivariate analysis.* New York: Academic Press.

Hallberg, A., Ogren, J. A., Noone, K. J., Heintzenberg, J., Berner, A., Solly, I., Kruisz, C., Reischl, G., Fuzzi, S., Facchini, M. C., Hansson, H.-C., Wiedensohler, A. and Svenningsson, I. B. 1992. Phase partitioning for different aerosol species in fog. *Tellus 44B*, 545–555.

Hansson, H.-C., Martinsson, B. G. and Lannefors, H. O. 1984. Long range aerosol transport in southern Sweden: an example of multivariate statistical evaluation methodology. *Nuclear instruments and methods in Physics Research 3*, 483–488.

Hasan, H. and Dzubay, T. G. 1987. Size distributions of species in fine particles in Denver using a microorifice impactor. *Aerosol Sci. Tech. 6*, 29–39.

Heintzenberg, J. 1988. A Processor-Controlled Multi-sample Soot Photometer. *Aerosol Sci. Tech. 8*, 227–233.

Heintzenberg, J. and Covert, D. S. 1987. Chemically resolved submicrometric size distribution and external mixing of the Arctic haze aerosols. *Tellus 39*, 374–382.

Infante, R. and Acosta, I. L. 1991. Size distributions of trace metals in Ponce, Puerto Rico air particulate matter. *Atmos. Env. 25B*, 121–131.

Johansson, E., Wold, S. and Sjödin, K. 1984. Minimizing effects of closure on analytical data. *Anal. Chem. 56*, 1685–1688.

Köhler, H. 1921. Zur Kondensation des Wassers in der Atmosphäre. *Meterologische Zeitschrift 38*, 168–171.

Lannefors, H., Hansson, H.-C. and Granat, L. 1983. Background aerosol composition in southern Sweden, fourteen micro and macro constituents measured in seven particle size intervals at one site during one year. *Atmos. Env. 17*, 87–101.

Lorber, A., Wangen, L. E. and Kowalski, B. J. 1987. A theoretical foundation for the PLS algorithm. *J. Chemometrics 1*, 19–31.

Martinsson, B. G., Swietlicki, E., Hansson, H.-C., Wiedensohler, A., Noone, K. J., Ogren, J. A. and Hallberg, A. 1992. Elemental composition of fog interstitial particle size fractions and hydrophobic fractions related to fog droplet nucleation scavenging. *Tellus 44B*, 593–603.

Mendenhall, W., Schaeffer, R. L. and Wackerly, D. D. 1981. *Mathematical Statistics with Applications.* Boston: PWS Publishers.

Noone, K. J., Ogren, J. A., Heintzenberg, J., Charlson, R. J. and Covert, D. S. 1988. Design and calibration of a counterflow virtual impactor for sampling of atmospheric fog and cloud droplets. *Aerosol Sci. Tech. 8*, 235–244.

Noone, K. J., Ogren, J. A., Hallberg, A., Heintzenberg, J., Ström, J., Hansson, H.-C., Svenningsson, B., Wiedensohler, A., Fuzzi, S., Facchini, M. C., Arends, B. G. and Berner, A. 1992. Changes in aerosol size- and phase distributions due to physical and chemical processes in fog. *Tellus 44B*, 489–504.

Ogren, J. A., Heintzenberg, J. and Charlson, R. J. 1985. In-situ sampling of clouds with a droplet to aerosol converter. *Geophys. Res. Lett. 12*, 121–124.

Sharaf, M. A., Illman, D. I. and Kowalski, B. R. 1986. *Chemometrics.* John Wiley & Sons.

Swietlicki, E. 1989. European source region identification of long range transported ambient aerosol based on PIXE analysis and related techniques. PhD Thesis, Lund University, Lund, Sweden.

Twomey, S. 1977. *Atmospheric aerosols.* Elsevier Scientific Publishing Co., Amsterdam.

Vong, R., Geladi, P., Wold, S. and Esbensen, K. 1988. Source contributions to ambient aerosol calculated by discriminant partial least squares regression (PLS). *J. Chemometrics 2*, 281–296.

Wilkinson, L. 1987. SYSTAT: *The system for statistics.* SYSTAT Inc., Evanston, IL USA.

Wold, S. and Sjöström, M. 1977. SIMCA: A method for analyzing chemical data in terms of similarity and analogy. In: *Chemometrics: theory and application* (ed. B. R. Kowalski). ACS symposium series no. 52, 243–282.