Talanta 101 (2012) 420–427

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Chemical boundary conditions for the classification of aerosol particles using computer controlled electron probe microanalysis

Willemien Anaf*, Benjamin Horemans, René Van Grieken, Karolien De Wael

Department of Chemistry, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium

ARTICLE INFO

Article history:
Received 8 May 2012
Received in revised form
12 September 2012
Accepted 22 September 2012
Available online 30 September 2012

Keywords:
Classification
Airborne particulate matter
Single particle analysis
EPMA

ABSTRACT

A method for the classification of individual aerosol particles using computer controlled electron probe microanalysis is presented. It is based on chemical boundary conditions (CBC) and enables quick and easy processing of a large set of elemental concentration data (mass%), derived from the X-ray spectra of individual particles. The particles are first classified into five major classes (sea salt related, secondary inorganic, minerals, iron-rich and carbonaceous), after which advanced data mining can be performed by examining the elemental composition of particles within each class into more detail (e.g., by ternary diagrams). The CBC method is validated and evaluated by comparing its results with the output obtained with hierarchical cluster analysis (HCA) for well-known standard particles as well as real aerosol particles collected with a cascade impactor. The CBC method gives reliable results and has a major advantage compared to HCA. CBC is based on boundary conditions that are derived from chemical logical thinking and does not require a translation of a mathematical algorithm output as does HCA. Therefore, the CBC method is more objective and enables comparison between samples without intermediate steps.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Automated chemical analysis of individual aerosol particles with computer controlled electron probe microanalysis (CC-EPMA) produces a vast body of numerical data which has to be translated to a physical reality. The analytical scientist can rely on his ability of logical reasoning and decision making. However, his limited processing power makes him rapidly outnumbered when he is analyzing huge particle numbers. Luckily, the discipline of chemometrics has facilitated his task by offering some mathematical tools for data mining.

Clustering analysis is a popular chemometric technique which has been frequently used for the classification of aerosol particles in groups of chemical similarity [1–3]. Since clustering is based on mathematic algorithms, it is necessary to interpret the chemical relevance of the extracted clusters [4,5]. This interpretation depends on the experience of the user, and therefore introduces some degree of subjectivity and irreproducibility. Expert systems can overcome such problems by mimicking the human ability of decision making. Such computer systems combine a knowledge base and an inference engine (computer programme). Ro et al. [6] developed an expert system for the automatic recognition and classification of chemical species in individual aerosol particles. As an output, the system gives an overview of the chemical species present in every single particle and the class to which the particle belongs, based on the chemical composition. Every particle is labelled with an ID code, composed of the formula of the chemical species with an abundance of more than 10%. Since particles are rarely in a pure form but contain some internal degree of heterogeneity; this leads to a very specific and complex particle classification scheme. Other chemical-based classification methods start from the net count rates from energy-dispersive X-ray microanalysis, combining the chemical composition with morphological information [7–9]. The chemical criteria are simple, giving a very basic classification output.

In the present article, an alternative classification method is presented, which can be considered as an elementary system based on an introductory knowledge base. The idea of the system is different from the system developed by Ro et al. [6]. In first instance, it uses chemical boundary conditions (CBC) for the automated classification of particles into five major classes (sea salt related, secondary inorganic, minerals, iron-rich and carbonaceous). In this way, particles are easily and rapidly grouped into general classes. Afterwards, the elemental composition of particles within each class can be examined into more detail, depending on the goals and needs of the research.

---

* Corresponding author. Tel.: +32 32653411; fax: +32 32652376.
E-mail addresses: willemien.anaf@ua.ac.be (W. Anaf),
benjamin.horemans@ua.ac.be (B. Horemans),
rene.vangrieken@ua.ac.be (R. Van Grieken),
karolien.dewael@ua.ac.be (K. De Wael).

0039-9140/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2012.09.051

The main aim of this study is to evaluate the classification performance of the CBC method. To this end, the classification results will be compared to that of hierarchical clustering analysis (HCA), which has been frequently used for single particle analysis [10,11]. Finally, it will be demonstrated how—in first instance—the rough classification of the CBC method could be advantageous for advanced data mining.

2. Materials and methods

2.1. Samples

Ten standard particle types were selected in order to evaluate the developed classification method, as compared to HCA performance. The standard particle types represent the common major inorganic components of an airborne particulate matter (PM) sample: NaCl, $NaNO_3$, Na_2SO_4, NH_4NO_3, $(NH_4)_2SO_4$, $CaCO_3$, $CaSO_4 \cdot 2H_2O$ (Merck), SiO_2 (Balzers), Fe_2O_3 (Aldrich). Furthermore, graphite (Johnson Matthey Chemicals Limited) was used to represent carbonaceous (soot) particles. All standard particles were prepared from analytical grade chemicals.

In order to make the standard ammonium and sodium salt particles, 0.5 M stock solutions were prepared from relevant reagents and ultra-pure water (Milli-Q). These solutions were sprayed on pre-cleaned silver foil substrates with an air brush (BADGER AIR-BRUSH model 200, Chicago, US) using nitrogen-gas as a carrier. The other particle types (water insoluble or poorly water soluble) were prepared by grounding the pure chemicals with an agate mortar, and suspending the resulting micrometre-sized particles in a water–acetone solution (50:50 vol%). A 250 μL droplet of each suspension was pipetted on a pre-cleaned silver substrate, after which it was left to dry under stable conditions in a laminar flow hood. All samples were measured within 3 days.

Real aerosol samples were collected in the Portuguese National Tile Museum in Lisbon. Samples were collected on silver substrates with a 9-stage Berner impactor. When operating the low vacuum pump at $30\,L\,min^{-1}$, this impactor has a 50% collection efficiency for particles with an aerodynamic diameter (AD) of 0.0625, 0.125, 0.25, 0.5, 1, 2, 4, 8 and 16 μm for stages 1–9, respectively.

2.2. Particle elemental composition

The standard and real particles were analyzed by means of CC-EPMA on a JEOL 733 instrument (Tokyo, Japan), equipped with a super-atmospheric thin-window energy-dispersive X-ray detector (Oxford, Abingdon, UK). Detailed information on the instrument and analytical procedure was previously described by Ro et al. [12]. In total, about 75 particles were analyzed for each standard type. Substrates with beam sensitive standard particles (e.g., $NaNO_3$, NH_4NO_3, $(NH_4)_2SO_4$) were mounted on a sample holder cooled with liquid nitrogen [13]. For real aerosols, only samples from stages 3–8 were used for analysis (instrumental limit for particles < 0.3 μm). About 1200 particles were analyzed per stage. Stages 3–6 were analyzed using cryo-conditions.

The obtained X-ray spectra from the standard and real aerosol particles were evaluated by non-linear least squares fitting with the AXIL code [14]. The semi-quantitative elemental particle composition was calculated with an in-house developed programme which is based on reversed iterative Monte Carlo simulations [15].

2.3. Classification methods

Preceding particle classification, erroneous spectra were omitted from the data set when the sum of peak intensities (substrate not included) was below 500 counts.

The CBC classification method was based on the procedure as reported by Kandler et al. [16]. The focus of their study, Saharan mineral dust, is reflected in the specified particle classes: iron-rich, titanium-rich, carbonates, other calcium-rich, gypsum, halite, quartz, silicates, sulphate silicate mixtures, sulphates and carbonaceous particles. However, in a wide range of applied aerosol studies, also other particle compounds are of importance, such as hygroscopic sea salt related particles and secondary inorganic aerosols (SIA). On the other hand, an accurate distinction between the different mineral particles is not always relevant. Therefore, in the classification method presented here, some classes are appended, other existing classes are combined. This more general classification is applicable in a wide variety of studies; the procedure of Kandler et al. [16] could be considered as a specific application. An overview of all criteria can be found in Table 1. The conditions are partly based on stoichiometry. However, since aerosol particles have an internally mixed structure and the relative abundance of compounds within a class could vary strongly, the concentrations and elemental ratios were defined less strictly as based on the stoichiometry of pure substances. Moreover, nitrogen concentrations are often underestimated due to spectral interferences. In order to account for these difficulties, the conditions were empirically adapted to improve the classification.

The group of the sea salt related particles is split up into genuine (NaCl) and aged ($NaNO_3$ and Na_2SO_4) sea salt components. All sea salt criteria are mainly based on the presence of sodium. Though, sea salts also contain minor concentrations of magnesium, sulphur, calcium and potassium, of which magnesium is most abundant. According to Cotruvo [17], the average molar ratio of Mg-to-Na in bulk sea water is 1:9. Hence, only minor concentrations of pure magnesium-related sea salts are expected. Therefore, compounds such as $MgCl_2$, $Mg(NO_3)_2$ and $MgSO_4$ are not explicitly incorporated in the classification method. Genuine sea salt particles contain high amounts of chlorine. Ageing induces a chlorine loss and an increase in oxygen, nitrogen and/or sulphur. Since the boundary conditions imply a strict classification, only from a certain degree of aging ($Cl/Na < 0.2$), sea salt particles will be classified as either $NaNO_3$ or Na_2SO_4. SIA refers to particulate ammonium salts. Although these particles are generally mixtures of different compounds, only NH_4NO_3 and $(NH_4)_2SO_4$ subclasses are differentiated. Sea salt and mineral related elements (Na, Mg, Al, Si, P, Ca) can only be present in limited concentrations. Mineral particles were divided into different classes, namely carbonates, gypsum, silicates and mixed minerals. The carbonate class includes calcium as well as magnesium carbonates, and combinations of both (e.g. dolomite). The gypsum class contains all particles with a formula related to $CaSO_4 \cdot xH_2O$. Within the class of the silicates, pure quartz as well as aluminium silicates are classified. Finally, the class of the mixed minerals contains all other particles that have clear mineral characteristics, but do not fit to the conditions of the other mineral classes. Especially the calcium to silicon ratio is an important condition since there is a gap between the silicon to calcium ratios of the silicates ($Si/Ca > 2$) and carbonates ($Si/Ca < 1/2$). Particles in this class can for example comprise internal mixtures of $CaCO_3$ and SiO_2 with some minor traces of iron. The class of iron-rich particles contains all iron oxides and hydroxides. As a last particle type, carbonaceous particles were selected. These particles are characterized by a high carbon concentration ($> 40\%$) and, except for oxygen, low concentrations of all other elements. Calcium and magnesium are explicitly mentioned in the criteria to avoid an overlap with carbonates. This class contains elemental carbon particles (soot) as well as biogenic particles, organic carbon etc. A class “others” designates the particles that could not be classified in one of the above mentioned classes.

All conditions were programmed in an excel-document with the use of logical functions. This pre-formatted document enables

Table 1
Aerosol classes and chemical boundary conditions.

Particle type	Class	Conditions	
Sea salt particles	NaCl	$\|Cl\| > 1$ $\|Na+Cl\| > 0.5$ $S/Na < 0.375$ $S/Cl < 0.5$ $Si/Cl < 0.5$	
	$NaNO_3$	$\|Na+Cl\| > 0.5$ $\|Na\| > 0.2$ $Cl/Na < 0.2$ $S/N < 0.2$	
	Na_2SO_4	$\|Na+S\| > 0.5$ $\|S\| > 0.2$ $Na/S > 1.25$ $Cl/Na < 0.2$ $Si/S < 0.5$ $Fe/S < 0.5$ $\|Mg+Al+Si+P+Ca\| < 0.35$ Not in class ‘gypsum’	
Secondary inorganic aerosols (SIA)	NH_4-rich particles	$(NH_4)_2SO_4$	NH_4NO_3
		$\|N+S\| > 0.35$ $Na/S < 0.2$ $N/S > 0.8$ $Si/S < 0.3$ $\|Na+Mg+Al+Si+P+Ca\| < 0.35$ Not in class ‘Na_2SO_4’ $\|Ca\| < 0.05$ $\|\|C\|\| < 0.45$	OR $\|N\| > 0.3$ $Na/N < 0.2$ $S/N < 0.25$ $Si/N < 0.3$ $\|Na+Mg+Al+Si+P+Ca\| < 0.35$ $Fe/N < 2$ $\|Ca\| < 0.05$ $\|\|C\|\| < 0.45$ (not in class $(NH_4)_2SO_4$)
Minerals	Carbonates	$\|Ca\| > 0.4$ $Mg/Ca < 0.33$ $Si/Ca < 0.5$ $S/Ca < 0.25$ $P/Ca < 0.15$ $\|\|Ca\|\| > 0.05$	OR $\|Mg+Ca\| > 0.45$ $0.33 < Mg/Ca < 3$ $\|\|Mg+Ca\|\| > 0.05$
	Gypsum	$\|Ca+S\| > 0.5$ $4 > Ca/S > 0.25$ $Na/Ca < 0.5$ $\|\|C\|\| < 0.5$	
	Silicates	$\|Si\| > 0.2$ $\|\|Si\|\| > 0.05$ $Na/Si < 0.7$ $Mg/Si < 1.33$ $Al/Si < 2.5$ $K/Si < 0.5$ $Ca/Si < 0.5$ $Fe/Si < 0.5$ $\|P+S+Cl\| < 0.2$	
	Mixed minerals	Not in any of the previous classes	
		$Ca/Si > 1.5$ $\|Na+S+Mg+Al+Si+K+Ca\| > 0.6$	OR $\|Na+S+Mg+Al+Si+K+Ca\| > 0.75$ $\|\|C\|\| < 0.4$
Iron-rich particles	Fe-oxides/hydroxides	$\|Fe\| > 0.3$ $Si/Fe < 1$	
Carbonaceous		Not in any other class $\|\|C+O\|\| > 0.6$ $\|\|C\|\| > 0.4$ $\|\|Ca\|\| < 0.2$ $\|\|Mg\|\| < 0.2$	

All elemental ratios are based on moles. $|X|$ denotes the total atomic content of element X relative to all other elements analyzed, except of carbon and oxygen. $\|X\|$ refers to the total atomic content of element X relative to all other elements, including carbon and oxygen.

quick and easy processing of a large set of elemental concentration data (mass%) for individual aerosol particles, without the necessity to interpret the chemical relevance of a mathematical output as would have been the case when applying a mathematical algorithm such as HCA.

The performance of the above described chemical classification method was evaluated and validated by using standard particles with a well-known composition and compared with the output results of HCA. By using the SPSS software package, HCA was performed with Ward's method (minimum variance) in combination with the Euclidean distance as a measure of particle dissimilarity. These conditions proved to be the most optimal for classifying individual aerosol particles [10]. Since the composition of the standard particle mixture is known, HCA was performed by fitting a fixed structure of 10 clusters (nr. of standard particle types). For real aerosols, the appropriate amount of clusters was chosen according to the average elemental composition of each cluster.

3. Results and discussion

3.1. Classification accuracy: standard particles

The set of standard particles was used to evaluate the accuracy of the CBC classification method. The results were compared with the classification output obtained with HCA. The CBC-based classification contains two more categories, namely 'mixed minerals' and 'others' (particles that could not be classified).

Table 2 shows the results of the two classification methods. The columns indicate the respective clusters/classes whereas the rows indicate the composition of the standard particles. In every class, 75 particles are expected, except for the classes 'other minerals' and 'others' where no particles are expected. When using the CBC method, almost 99% of all particles was classified in the appropriate class (Table 2). Since HCA is a method that cannot be supervised during the calculations, clusters do not necessarily correspond to the different standard particle types [10]. Yet, for the present data, the HCA clusters are analogous to the different particle types. Only graphite is partly (about 13%) classified in the same cluster as the ammonium-rich salts. In both methods, a proper distinction between NH_4NO_3 and $(NH_4)_2SO_4$ seems to be questionable. Therefore, these compounds were combined into one group as NH_4-rich particles. With this assumption, over 99% of the particles were correctly classified by the CBC method.

Since clustering of large data sets with an algorithm is never exact, it can be concluded that both methods have an equally valuable outcome. However, some advantages and disadvantages of both methods must be noted. A major advantage of the CBC classification is the objectivity. Each single particle is exposed to the same conditions by which it is classified. For example, every sea salt particle will be classified as aged sea salt when a fixed ratio of e.g. sodium to chloride is exceeded. This objectivity is not present in HCA. First of all, the applicant should estimate the most appropriate amount of clusters. Afterwards, the clusters have to be dedicated to a particle type depending on the average elemental composition of the cluster. These average compositions are often composed of several elements that should be related to a mixture of particle types. The interpretation is therefore more difficult and relies on the experience of the user. Moreover, the objectivity of the CBC method permits to directly compare different aerosol samples. With HCA, this is only possible after applying a supplementary algorithm [3,5], again time consuming and relying on the interpretation qualities of the user.

A minor disadvantage of the CBC classification is that the particles are divided into 'pure' classes. For all classes, a relatively high concentration of the characteristic elements is required. Mixed particles, or particles that overlap on the substrate during analysis, can contain a wide range of elements, which lowers the standardized concentration per element. These particles do not

Table 2
Classification accuracy for chemical boundary condition and hierarchical clustering algorithms.

Chemical boundary condition (CBC)

	Class										
Standard type*	NaCl	$NaNO_3$	Na_2SO_4	NH_4NO_3	$(NH_4)_2SO_4$	carbonates	gypsum	silicates	Fe-rich	carbonaceous	mixed min.	others
NaCl	75	–	–	–	–	–	–	–	–	–	–	–
$NaNO_3$	–	75	–	–	–	–	–	–	–	–	–	–
Na_2SO_4	–	–	72	–	–	–	–	–	–	–	–	3
NH_4NO_3	–	–	–	75	–	–	–	–	–	–	–	–
$(NH_4)_2SO_4$	–	–	–	4	69	–	–	–	–	1	–	1
$CaCO_3$	–	–	–	–	–	75	–	–	–	–	–	–
$CaSO_4$	–	–	–	–	–	–	73	–	–	–	1	1
SiO_2	–	–	–	–	–	–	–	75	–	–	–	–
Fe_2O_3	–	–	–	–	–	–	–	–	75	–	–	–
Graphite	–	–	–	–	–	–	–	–	–	75	–	–

Hierarchical clustering algorithm (HCA)

	Cluster									
Standard type*	1	2	3	4	5	6	7	8	9	10
NaCl	75	–	–	–	–	–	–	–	–	–
$NaNO_3$	–	75	–	–	–	–	–	–	–	–
Na_2SO_4	–	–	75	–	–	–	–	–	–	–
NH_4NO_3	–	–	–	62	13	–	–	–	–	–
$(NH_4)_2SO_4$	–	–	–	–	75	–	–	–	–	
$CaCO_3$	–	–	–	–	–	74	1	–	–	–
$CaSO_4$	–	–	–	–	–	–	75	–	–	–
SiO_2	–	–	–	–	–	–	–	75	–	–
Fe_2O_3	–	–	–	–	–	–	–	–	75	–
Graphite	–	–	–	–	10	–	–	–	–	65

* For each standard type, 75 particles were analyzed.

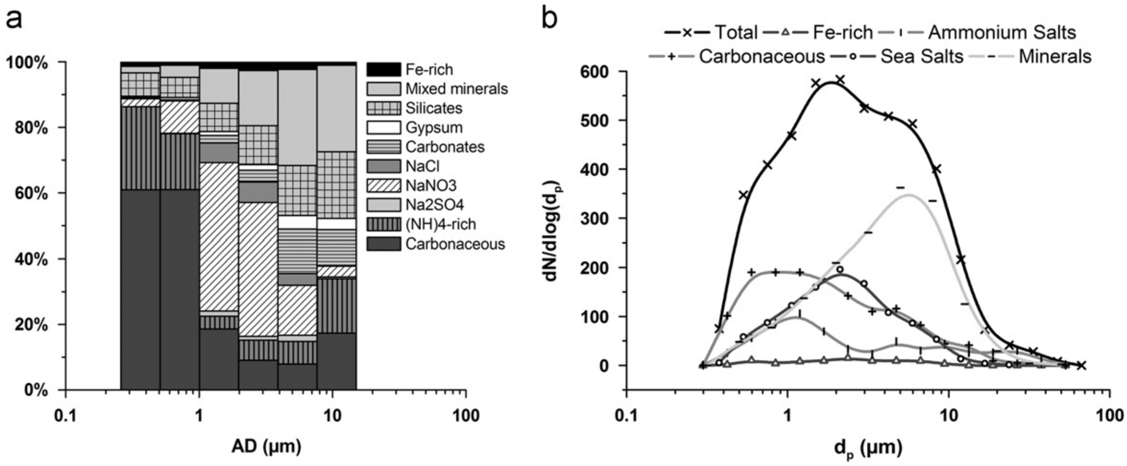

Fig. 1. Composition of a real aerosol sample as determined with CBC. (a) Relative abundance of each particle type in segregated fractions of aerodynamic diameter (AD) and (b) particle number size distribution for the main particle classes.

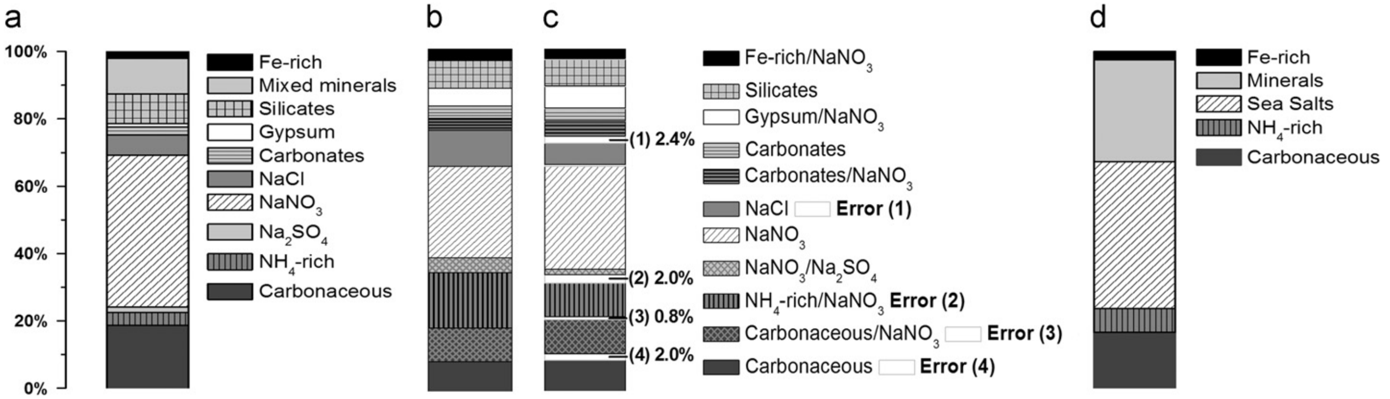

Fig. 2. Composition for particles with an aerodynamic diameter (AD) between 1 and 2 µm. (a) CBC classification, (b) HCA classification, (c) comparison between CBC and HCA and (d) k-nearest neighbour analysis. Error: amount of particles which are classified differently by CBC and HCA.

necessarily have a sufficiently high concentration of characteristic elements to be classified in one of the classes, and will therefore be categorized as ‘others’. The expert system of Ro et al. (2004) meets this drawback by identifying/classifying particles as mixed compositions. However, when a particle contains for example the elements Na, S, O, Ca and N in abundant concentrations, it is hard to correctly predict the exact composition. Calcium nitrates and/or sulphates, and sodium nitrates and/or sulphates all belong to the possibilities.

3.2. CBC versus HCA: real aerosols

To explore the reliability of CBC, the classification system was applied to real aerosol data and compared with the output of HCA. Classification was performed independently for each AD size fraction (Fig. 1a). In the smaller fractions (<1 µm), carbonaceous particles and ammonium salts predominate, whereas mineral particles only represent a small fraction. At the other hand, the mineral fraction increases with increasing particle size, while the number of carbonaceous particles decreases. The size range from 1 to 4 µm is characterized by a dominant fraction of (aged) sea salt particles.

Simultaneously with the X-ray spectra, particle diameters were measured from the backscattered image (projected area). After CBC classification, the particle size distribution for the dominant classes was constructed (Fig. 1b). About 78% of the detected particles were larger than 1 µm, whereas 46% were larger than 2.5 µm. Although the accumulation and coarse mode cannot be clearly distinguished in the general distribution (total), the bimodal structure is evident from the individual distributions of the dominant particle classes. The coarse mode is dominated by mineral particles (silicates, carbonates, gypsum and mixed minerals), whereas the accumulation mode consists of ammonium salts, formed by secondary processes, and carbonaceous particles. Around 2.5 µm, another increased concentration of carbonaceous particles can be observed. This can be explained by larger biogenic particles or soot particles forming large but light agglomerates on the Ag-substrate. The size distribution of the sea salt related particles can be situated in between, with the maximum of the distribution around 2 µm.

In order to study the CBC particle classification into more detail, the particle fraction with AD between 1 and 2 µm was selected (Fig. 2a). Additionally, HCA was performed on this limited set of particles to make a comparison with the output of the CBC method (Fig. 2b). A total of 1000 particles was considered. Particles that were classified as “others” by the CBC method were not taken into account. Though we already compared the outcome of both classification methods for the set of standard particles of which a correct classification can easily be evaluated, a similar approach is much more complex for the data set of real aerosol particles. One does not know the intrinsically best classification, thus, comparison on the basis of correct scores is impossible [18]. We compared the classification outcome according to following criterion: when both methods classified a particle equally, the classification was considered to be valid. For clusters containing mixtures, both possibilities were taken to be correct. For example, in the cluster containing $NaNO_3$ and gypsum, CBC classification as $NaNO_3$ and gypsum particles were both considered to be valid. When the classification of a particle differed for both methods, it was labelled as "error" (Fig. 2c).

Taking the mixed clusters of the HCA output into account, the relative abundance of the dominant particle classes (carbonaceous, ammonium salts, sea salts, minerals and iron-rich) are similar for both classification methods. Within the mineral class, the abundance of gypsum and carbonate particles was overestimated by HCA, as compared to the CBC method. However, this is compensated by the class 'mixed minerals' in the CBC method, which are characterized by high carbon and oxygen fractions. Also for NaCl-rich particles, HCA tends to overestimate the total particle number, though the sum of all sea salt related particles is similar. To quantify the similarity between the two methods, the number of "erroneously" classified particles to the total particle number was considered. For the particle fraction with AD between 1 and 2 μm, about 91.5% of the particles were classified similarly. The other 8.5% were labelled as "erroneously". The "erroneously" classified particles were only numerous in the case of NH_4-rich/$NaNO_3$ (2.4%), NaCl (2.0%) and carbonaceous particles (2.0%). As for the NaCl-cluster, these wrongly classified particles were labelled as carbonaceous by the CBC method. The NH_4-rich/$NaNO_3$-cluster obtained by HCA contains also carbonaceous particles according to CBC. The "erroneously" classified particles in the carbonaceous cluster are due to a few particles that were classified as NH_4-rich and $NaNO_3$ particles in the CBC method. All other clusters have a minor "error" of less than 0.8%. From this comparison it can be concluded that both HCA and CBC give a similar result. The observed differences result from the way in which heterogeneous particles (internal mixtures/agglomeration on the substrate) are treated by both methods. The main difference is that the interpretation step of the CBC method is applied for each particle separately, while the HCA output has to be interpreted by the user, based on the average composition of the clusters in their entirety.

Since five major particles classes were set a priori, an additional comparison was made with *k*-nearest neighbour supervised clustering analysis. The algorithm's evaluation is based on the closest (Manhattan) distance of the unknown particle to its *k*-nearest neighbours ($k=35$). When the majority of *k* neighbours is member of class *X*, the unknown particle is assigned to *X*. The set of standard particles was used as training data to build up a five-class model (sea salts, ammonium-rich particles, minerals, iron-rich particles and carbonaceous particles). Afterwards, this model was used to assign the real aerosol particles (size fraction of 1–2 μm) into one of the pre-set classes. Fig. 2(d) represents the result of this clustering method. Comparing the outcome of each single particle, 77.3% of the particles are grouped within the same class as the CBC method. Good correspondence was found for the iron-rich particles, the minerals and the carbonaceous particles. For the sea salts and NH_4-rich particles, the outcome of the two methods was less similar, resulting in a higher abundance of minerals and NH_4-rich particles, and less sea salts for the *k*-nearest neighbour analysis compared to the CBC method. Although the *k*-nearest neighbour method is simple in use, the main disadvantage is that the particles are always classified in one of the pre-set classes. Even when a particles composition is far from corresponding to one of the classes, it will be classified according to its "nearest" neighbours. Since a class "others" as present in the CBC method can be very diverse from compositional point, one cannot make a set of training data that covers all "other" particles.

3.3. Advanced data mining for real aerosol data

3.3.1. Sea salt particles

It is well known that sea salts can age by reacting with atmospheric acidity. The extent of sea salt nitration and sulphation in aged sea salt particles can be evaluated by plotting the relative abundance of N, S and Cl in a ternary diagram. For the standard particles, considered as references of pure compounds, non-aged sea salt particles (NaCl) and the two types of aged sea salt particles ($NaNO_3$ and Na_2SO_4) can be clearly distinguished as separate clusters in the diagram (Fig. 3a). A similar plot for all sea salt related real aerosol particles as classified by the CBC method showed the same pattern (Fig. 3b). Though, a more fluent transition can be observed between the particles classified as NaCl and $NaNO_3$. Particles with a Cl-to-N ratio up to 0.5 are classified as NaCl, although most of these particles are already partly nitrated. On the opposite, most particles classified as $NaNO_3$ still contain (minor) concentrations of unreacted sea salt. This illustrates how a detailed classification into specific classes of 'pure' chemicals makes no sense, because a considerable fraction of the particles could be equally assigned to two (or more) groups. A relatively high amount of particles is found to contain a mixture of nitrates and sulphates. As a comparison, a ternary plot of HCA-classified sea salts (omitting mixed clusters) within the

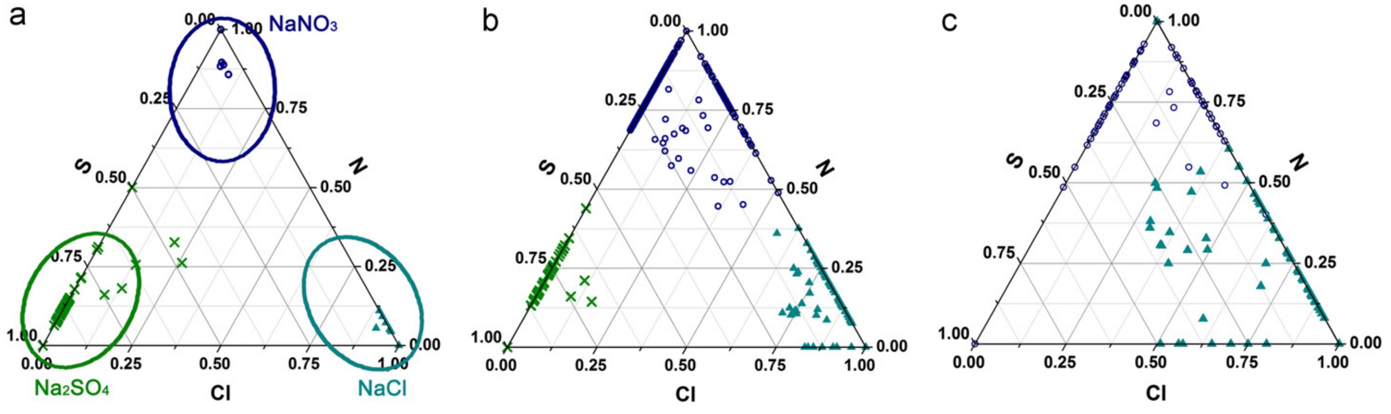

Fig. 3. Relative atomic abundance of N, S and Cl in sea salt particles. (a) CBC, standard particles; (b) CBC, real aerosol particles and (c) HCA, real aerosol particles with aerodynamic diameter (AD) between 1 and 2 μm. Green crosses: Na_2SO_4; blue open dots: $NaNO_3$; blue triangles: NaCl. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

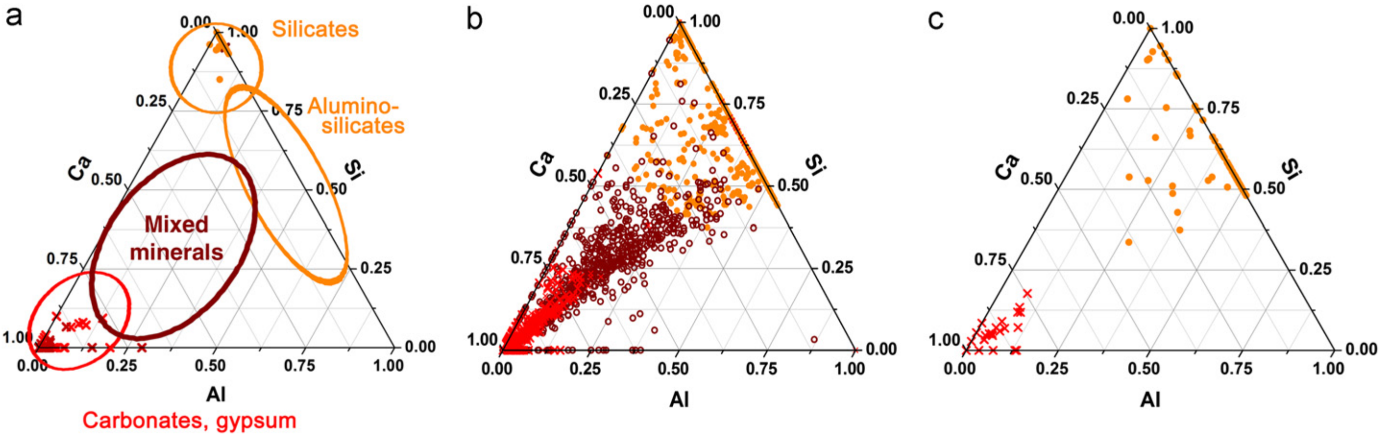

Fig. 4. Relative atomic abundance of Al, Ca and Si in mineral particles. (a) CBC, standard particles; (b) CBC, real aerosol particles and (c) HCA, real aerosol particles with aerodynamic diameter (AD) between 1 and 2 μm. Red crosses: carbonates/gypsum; dark red open dots: mixed minerals; orange closed dots: silicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

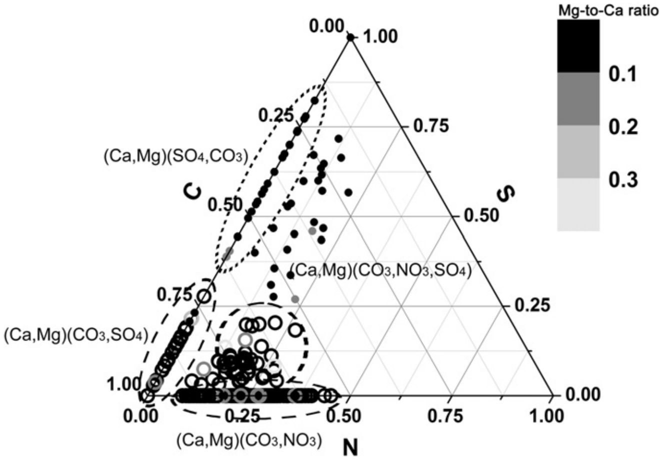

Fig. 5. Relative atomic abundance of N, C and S in calcium-rich particles. Open dots: carbonates; closed dots: gypsum.

aerodynamic size range of 1–2 μm is shown in Fig. 3(c). The sea salt related particles are more scattered in this plot. Some particles appear in the centre of the ternary diagram which were not recognized by the CBC method. The composition of these particles consists of less than 15% of the elements Na, Cl, S and N. Although these particles are classified as sea salts by HCA, the condition for (aged) sea salts in the CBC method requires a higher abundance for these elements. Though both HCA and CBC described similar performance qualities for the standard particles, the CBC method seems to display better results in the case of real aerosols.

3.3.2. Minerals

Most mineral particles are characterized by the presence of at least one of the elements Si, Al or Ca. The mass fraction for these three elements was set in ternary plots for all mineral particles (silicates, carbonates, gypsum and mixed minerals); the standard particles were again considered as a reference (Fig. 4a–c). Silicates, including aluminium silicates, are clearly characterized by a high abundance of silicon. The fractions of aluminium range from 0% (e.g., quartz) up to more than 50% with a calcium fraction up to 35% (e.g., feldspar). In this plot, gypsum and carbonate particles cannot be distinguished since both are characterized by a high calcium concentration and low concentrations of aluminium and silicon. Mixed mineral particles are found in the transition zone between silicates and Ca-rich particles (Fig. 4b). Classification results of HCA (silicates and carbonates) show a similar pattern.

3.3.3. Calcium-rich particles

About 92% of the Ca-rich particles were found with a Mg-to-Ca ratio below 0.1, of which another 20% does not contain any Mg. Analogously to sea salt particles, soil-derived carbonates can react with atmospheric acids. This results in internally mixed particles containing $Ca(NO_3)_2$, $CaSO_4$, $Ca(CO_3,NO_3)_2$, $Ca(NO_3,SO_4)$, $(Ca,Mg)NO_3$ and/or $(Ca,Mg)(NO_3,SO_4)$ [19]. Although this amount of detail cannot be obtained directly by the CBC classification, it can be visualized in a ternary plot. Fig. 5 demonstrates how a fraction of the carbonate particles (partly) reacted to calcium and magnesium nitrates and/or sulphates. Three main groups can be distinguished: $(Ca,Mg)(CO_3,SO_4)$, $(Ca,Mg)(CO_3,NO_3)$ and $(Ca,Mg)(CO_3,NO_3,SO_4)$. The Mg-to-Ca ratio is visualized by a grey-scale. A fraction of the gypsum particles is characterized by the absence of nitrogen, showing a clear gradient in CO_3- and SO_4-mixture. All other gypsum particles contain mixtures of $(Ca,Mg)(CO_3,NO_3,SO_4)$ or $Ca(CO_3,NO_3,SO_4)$. A small percentage of all carbonate and gypsum particles (6%) contains minor amounts of chloride, with a median Cl mass fraction of 2.1%.

3.3.4. Secondary inorganic aerosols

The separation of different NH_4-rich salts appeared to be difficult with HCA as well as CBC. Therefore, these compounds were classified as one group in the CBC method. Nonetheless, the distribution of different ammonium salts such as $(NH_4)_2SO_4$, NH_4NO_3 and NH_4Cl can be derived out of a ternary plot, giving the relative abundances of sulphur, nitrogen and chloride (Fig. 6a and b).

3.3.5. Iron-rich particles

For all particles classified as iron-rich, the Fe-to-O ratio was calculated. With EPMA, one cannot distinguish oxides and hydroxides. Here, we assumed all iron-rich particles to be oxides. One particle was characterized by a remarkably high ratio (17). This particle can be considered as metallic iron. Only few particles (6%) of the iron-rich particles had an Fe-to-O ratio between 0.5 and 1. These particles correspond to (mixtures of) iron and/or iron oxides such as FeO and Fe_2O_3. Two particles had a slightly higher Fe-to-O ratio (around 1.5), which could be linked to a mixture of iron oxides and metallic iron. All other iron-rich particles were characterized by a low Fe-to-O ratio down to 0.1. Examining

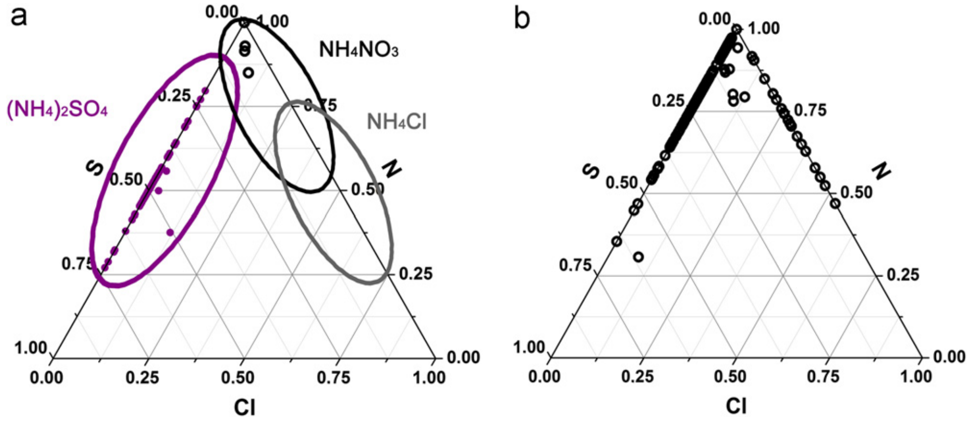

Fig. 6. Relative atomic abundance of Cl, S and N in secondary inorganic particles. (a) CBC, standard particles and (b) CBC, real aerosol particles.

additional elements in these particles shows that also the abundance of nitrogen and sodium is relatively high. Therefore, it can be concluded that the iron-rich particles with a low Fe-to-O ratio were mostly mixed with sodium nitrate salts, and possibly SiO_2 and $CaCO_3$. The latter two compounds could be found in iron-rich particles that are generated from the friction between train wheels and brake block composed of iron, $CaCO_3$ and glass fibres [20]. Since a railway station was only a few 10 m located from the sampling location, this is reasonable.

3.3.6. *Carbonaceous particles*

No additional relevant specifications could be extracted after classification of the carbonaceous particles from the current data set. This class contains a wide variety of components such as soot, biogenic particles (bacteria, pollen, algae,…) and organic particulate matter. Typical compositions such as the presence of nitrogen in combination with carbon and oxygen and some characteristic minor elements (e.g., Si, P, S, K, Ca) could be linked to a biogenic source [21]. Though, also ammonium or nitrate compounds associated with soot can cause a similar elemental composition. Hence, supplementary information such as morphological characteristics [19], the application of molecular analysis techniques such as μ-Raman spectroscopy [22], or combined techniques such as Energy-Dispersive Electron Probe X-ray Microanalysis/Attenuated Total Reflection Fourier Transform-Infrared Imaging (SEM-EDX/ATR-FT-IR) and Scanning Transmission X-ray Microscopy/Near Edge X-Ray Absorption Fine Structure Spectroscopy (STXM-NEXAFS) [21,23] could help with further differentiating within this particle type.

4. Conclusion

The present study has shown how the developed CBC method gives reliable results for the classification of aerosol particles. Although the classification results were comparable to those obtained with HCA, some differences should be underlined.

The CBC method is based on boundary conditions that are derived from chemical logical thinking. HCA, on the contrary, is based on a mathematical algorithm which has to be translated into particle chemistry. The most difficult task in HCA is the choice where to cut the dendrogram and whether the addition of an extra cluster would add value to the chemical relevance of the system. Moreover, the chemical meaning of a cluster is determined from the average elemental composition of all particles in that cluster. Generally, this is no problem for homogeneous particles. However, in reality particles are often internally mixed, which results in a more diffuse classification and mean elemental composition per cluster. Since the interpretation step of the CBC method is applied for each particle separately, the CBC method is considered to be more objective compared to HCA. This is immediately the main advantage of the CBC method: since it applies the same criteria on particles from different samples, it enables the comparison between samples without any intermediate step.

One can argue that the CBC process of boxing particles by their chemical composition implies similar difficulties as for HCA: 'What is the size of the boxes in which the particles have to be classified?', or in other words, 'What is the chemical detail one should reach?' In the presented work, it was chosen to take five major particle classes as the basis of the classification. Despite this rough classification, it was demonstrated how detailed information can be extracted by examining each class by means of elemental ratios and ternary diagrams.

References

[1] R.K. Xie, H.M. Seip, L. Liu, D.S. Zhang, Air Qual. Atmos. Health 2 (2009).
[2] R.M. Healy, I.P. O'Connor, S. Hellebust, A. Allanic, J.R. Sodeau, J.C. Wenger, Atmos. Environ. 43 (2009) 6408–6414.
[3] A. Worobiec, L. Samek, Z. Spolnik, V. Kontozova, E. Stefaniak, R. Van Grieken, Mikrochim. Acta 156 (2006) 253–261.
[4] B. Treiger, I. Bondarenko, H. Van Malderen, R. Van Grieken, Anal. Chim. Acta 317 (1995) 33–51.
[5] I. Bondarenko, B. Treiger, R. Van Grieken, P. Van Espen, Spectrochim. Acta Part B 51 (1996) 441–456.
[6] C.-U. Ro, H. Kim, R. Van Grieken, Anal. Chem. 76 (2004) 1322–1327.
[7] P. Fruhstorfer, R. Niessner, Mikrochim. Acta 113 (1994) 239–250.
[8] M. Ebert, S. Weinbruch, P. Hoffmann, H.M. Ortner, Atmos. Environ. 38 (2004) 6531–6545.
[9] M. Ebert, S. Weinbruch, P. Hoffmann, H.M. Ortner, J. Aerosol Sci. 31 (2000) 613–632.
[10] J. Osan, J. de Hoog, A. Worobiec, C.-U. Ro, K.Y. Oh, I. Szaloki, R. Van Grieken, Anal. Chim. Acta 446 (2001) 211–222.
[11] K. Gysels, F. Delalieux, F. Deutsch, R. Van Grieken, D. Camuffo, A. Bernardi, G. Sturaro, H.-J. Busse, M. Wieser, J. Cult. Heritage 4 (2004) 221–230.
[12] C.-U. Ro, J. Osan, R. Van Grieken, Anal. Chem. 71 (1999) 1521–1528.
[13] A. Worobiec, J. de Hoog, J. Osan, I. Szaloki, C.-U. Ro, R. Van Grieken, Spectrochim. Acta Part B 58 (2003) 479–496.
[14] B. Vekemans, K. Janssens, L. Vincze, F. Adams, P. Vanespen, X-Ray Spectrom. 23 (1994) 278–285.
[15] C.-U. Ro, J. Osan, I. Szaloki, J. de Hoog, A. Worobiec, R. Van Grieken, Anal. Chem. 75 (2003) 851–859.
[16] K. Kandler, N. Benker, U. Bundke, E. Cuevas, M. Ebert, P. Knippertz, S. Rodriguez, L. Schutz, S. Weinbruch, Atmos. Environ. 41 (2007) 8058–8074.
[17] J.A. Cotruva, Water Cond. Purif. 47 (2005) 13–17.
[18] D.L. Massart, L. Kaufman, The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, John Wiley & Sons, New York, 1983.
[19] S. Maskey, T. Kang, H.-J. Jung, C.-U. Ro, Indoor Air 21 (2011) 12–24.
[20] S. Kang, H. Hwang, Y. Park, H. Kim, C.-U. Ro, Environ. Sci. Technol. 42 (2008) 9051–9057.
[21] Y.-C. Song, J. Ryu, M.A. Malek, H.-J. Jung, C.-U. Ro, Anal. Chem. 82 (2010) 7987–7998.
[22] R. Escribano, J.J. Sloan, N. Siddique, N. Sze, T. Dudev, Vib. Spectrosc. 26 (2001) 179–186.
[23] R.C. Moffet, T. Henn, A. Laskin, M.K. Gilles, Anal. Chem. 82 (2010) 7906–7914.