Journal of Aerosol Science 123 (2018) 1–16

Contents lists available at ScienceDirect

Journal of Aerosol Science

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

A model based two-stage classifier for airborne particles analyzed with Computer Controlled Scanning Electron Microscopy

Mario Federico Meier[a,c,*], Thoralf Mildenberger[b], René Locher[b], Juanita Rausch[a,c], Thomas Zünd[c], Christoph Neururer[a], Andreas Ruckstuhl[b], Bernard Grobéty[a]

[a] *Department of Geosciences, University of Fribourg, Chemin du Musée 6, 1700 Fribourg, Switzerland*
[b] *Institute for Data Analysis and Process Design (IDP), Zürich University of Applied Sciences (ZHAW), 8400 Winterthur, Switzerland*
[c] *Particle Vision GmbH, c/o Fri Up, Annexe 2, Passage du Cardinal 11, 1700 Fribourg, Switzerland*

ARTICLE INFO

Keywords:
Particle classifier
CCSEM/EDX
Single particle analysis
Cluster analysis
Source apportionment

ABSTRACT

Computer controlled scanning electron microscopy (CCSEM) is a widely-used method for single airborne particle analysis. It produces extensive chemical and morphological data sets, whose processing and interpretation can be very time consuming. We propose an automated two-stage particle classification procedure based on elemental compositions of individual particles. A rule-based classifier is applied in the first stage to form the main classes consisting of particles containing the same elements. Only elements with concentrations above a threshold of 5 wt% are considered. In the second stage, data of each main class are isometrically log-ratio transformed and then clustered into subclasses, using a robust, model-based method. Single particles which are too far away from any more densely populated region are excluded during training, preventing these particles from distorting the definition of the sufficiently populated subclasses. The classifier was trained with over 55,000 single particles from 83 samples of manifold environments, resulting in 227 main classes and 465 subclasses in total. All these classes are checked manually by inspecting the ternary plot matrix of each main class. Regardless of the size of training data, some particles might belong to still undefined classes. Therefore, a classifier was chosen which can declare particles as unknown when they are too far away from all classes defined during training.

1. Introduction

Computer controlled scanning electron microscopy (CCSEM) is a powerful method for characterizing individual airborne particles. It has the advantage that thousands of particles can be characterized individually in short time by localizing particles in backscattered electron images and measuring their composition with energy dispersive X-ray spectroscopy (EDS). Morphological parameters of individual particles can be obtained as additional characteristics of the particles. CCSEM as well as other methods for automated single particle analysis like Raman microscopy (Huang et al., 2013; Ivleva, McKeon, Niessner, & Poschl, 2007) or aerosol mass spectroscopy (Lanz et al., 2007; Richard et al., 2011) are particularly useful for apportionment studies of ambient aerosols. An important step of the apportionment procedure is the classification of the particles. Comparing ambient particle compositions with a library of reference emission particles allows assigning the particles to corresponding sources.

Many different approaches have been proposed for the classification process, including rule-based expert systems (e.g. Willis,

---

* Corresponding author at: Department of Geosciences, University of Fribourg, Chemin du Musée 6, 1700 Fribourg, Switzerland.
E-mail address: mario.meier@unifr.ch (M.F. Meier).

https://doi.org/10.1016/j.jaerosci.2018.05.012
Received 13 February 2018; Received in revised form 21 May 2018; Accepted 22 May 2018

Blanchard, & Conner, 2002), hierarchical (e.g. *Osán et al.*, 2001; Huang et al., 2013) and non-hierarchical cluster analysis (e.g. Bernard & Van Grieken, 1986), artificial neural networks (e.g. ART-2a) (Hopke, 2008) and other statistical approaches like principal component analysis (PCA) (e.g. Genga et al., 2012) and Partial Least Squares Discriminant Analysis (PLS-DA) (Tan, Malpica, Evans, Owega, & Fila, 2002).

Rule-based classifiers using chemical boundary conditions (CBC) are widely used and are suitable for simple situations with well-known sources (Ebert et al., 2002; Ebert, Weinbruch, Hoffmann, & Ortner, 2000; Ebert, Weinbruch, Hoffmann, & Ortner, 2004; Kandler et al., 2007; Kang, Hwang, Park, Kim, & Ro, 2008; Ro, Kim, & Van Grieken, 2004). For example, Lorenzo, Kaegi, Gehrig, and Grobéty (2006) studied particles originated from railway traffic by using a simple but efficient rule-based class building system, based on EDS net intensities of Fe, Si, Al, S and Ca.

Willis et al. (2002) proposed in the guidelines for the application of Scanning Electron Microscopy (SEM) of the US Environmental Protection Agency (U.S. EPA) a rule-based classifier, using chemical composition, morphological aspect ratio, total X-ray counts and grayscale brightness value.

Anaf, Horemans, Van Grieken, and De Wael (2012) compared their CBC classifier, based on procedures proposed by Kandler et al. (2007), with a method for hierarchical clustering (HCL) (e.g. Osán et al., 2001). They concluded that CBC has advantages compared to HCL, as cutting the clustering dendrogram at different heights in the latter method is somewhat arbitrary. Furthermore, they pointed out that HCL has some difficulties with particles of mixed phases. These particles often have wide compositional ranges, which is difficult to map by hierarchical clustering. Nevertheless, HCL and non-hierarchical clustering are widely used for particle classification (Bein, Zhao, Wexler, & Johnston, 2005; Bernard & Van Grieken, 1986; Genga et al., 2012; Kim & Hopke, 1988; Moffet et al., 2013). Handling of outliers is a major problem in clustering. Most methods cannot properly deal with outliers, which have a strong impact on the cluster shapes.

In this paper, a new two-stage classifier for particles is presented. It consists of a combination of a rule-based first stage and a robust model-based second stage classifier. For each particle, element concentrations below 5 wt% are set to 0 wt% and the proportions of the remaining elements are rescaled to sum up to 100%. This elimination step is necessary because most spectra are noisy and, without thresholding, erroneous peak attributions are likely. For a system with better signal/noise ratio the threshold could be lowered. The set of the remaining elements defines the main class, which is subdivided in a second stage by a robust model-based clustering method. The resulting classes are hereafter called subclasses and are (after a transformation) described by (p-1)-dimensional ellipsoids with p being the number of elements present in the corresponding main class. Outliers are excluded automatically and hence have no effect on the shapes of the subclasses. Outliers found during classification are not assigned to any class and declared as unknown. As the classifier was trained with samples exposed to a broad range of sources, it is able to classify a broad range of airborne particles. If required, new classes can be easily included by further training. The classifier works for homogeneous particles, heterogeneous mixed particles and solid solutions.

2. Methods and materials

2.1. CCSEM

CCSEM for training and testing the classifier was performed with a FEI XL30 Sirion FEG Scanning Electron Microscope (SEM) equipped with an Energy Dispersive X-ray system (EDX, 10 mm^2 Lithium doped silicon detector by EDAX, with maximal energy resolution of 128 eV) at the Department of Geosciences of the University of Fribourg, Switzerland. All samples were coated with a 40 nm thick carbon layer for better electrical conductivity. The analyses were performed with an acceleration voltage of 20 keV and spot size 4 (beam current = 10 nA). Data were acquired using the particle analysis module of the EDAX GENESIS software. Eighteen elements (Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Zn, Ba and Pb) were used by default for classification. Lighter elements (C, N, O and F) were not taken into account for several reasons. As the sampling substrates (polycarbonate filters and carbon stubs) and the coating consist of O and/or C, they interfere with the C and O content of the particles. Many of the particles consisting of light elements only were not recognized properly in backscatter images. This is due to poor contrast of these particles deposited on carbon based substrates. Net intensities have been corrected for background, matrix, absorption and fluorescence effects (ZAF-correction) before converting them into concentrations. ZAF-correction neglects geometric effects but assumes semi-infinite flat samples. The ZAF procedure is thus not ideal for particles that have an irregular surface or which are too small. Correction factors obtained by Monte Carlo simulations can provide more accurate results for the particle compositions (Armstrong, 1991; Choël, Deboudt, Osán, Flament, & Van Grieken, 2005; Ro, Osàn, & Van Grieken, 1999). Geometric corrections, however, are time consuming and test measurements on a standard glass sample described below show that the concentrations obtained from ZAF-procedure without geometric corrections for particles larger than 1.5 μm (geometric diameter) are precise and accurate enough for classification. In our setting, accuracy is less important than precision as long as measurement biases, i.e. systematic errors, are size independent. Changing the instrument or acquisition parameters may change bias of measurements. In this case all samples have to be reanalyzed by CCSEM and retrained. Further details on the SEM/EDS settings used during data acquisition are given in Table 1.

Total number of X-ray counts emitted from a given area of the substrate that is free of particles was used to check measurement set-up. For carbon based substrates (carbon pad, polycarbonate filter etc.) and optimized measurement geometry the minimum was set to 6000 counts per second (CPS). This threshold and most settings described above are hardware specific and have to be adapted for other instrumental set-ups.