STATISTICAL ANALYSIS OF AEROSOL MASS SPECTRA: CHEMOMETRICS AND CHEMICAL SPECIATION

MIKKO ÄIJÄLÄ

Institute for Atmospheric and Earth System Research (INAR) / Physics
Faculty of Science
University of Helsinki
Helsinki, Finland

Academic dissertation

To be presented, with the permission of the Faculty of Science
of the University of Helsinki, for public criticism in auditorium B123,
Exactum, Pietari Kalmin katu 5, on February 1st, 2019, at 12 o'clock noon.

Helsinki 2019

Author's Address: Institute for Atmospheric and
                     Earth System Research / Physics
                     P.O. Box 64
                     FI-00014 University of Helsinki
                     mikko.aijala@helsinki.fi

Acknowledgements

Statistical analysis of aerosol mass spectra: chemometrics and chemical speciation

Mikko Yrjö Erik Äijälä

University of Helsinki, 2019

Abstract

Climate change and degraded air quality rank among the most severe environmental problems that humanity currently faces. Atmospheric aerosols - the fine particles suspended in air – play a significant role in both of them. Aerosols affect Earth's radiation balance via their interactions with radiation (scattering and absorption of light) and atmospheric water vapour (cloud formation, lifetime, brightness, and rainfall). Overall, aerosols have a cooling effect on climate. When inhaled, aerosols can cause a range of adverse health effects, from allergies to lung disease and increased rates of cancer. Air pollution is estimated to result in millions of premature deaths each year in polluted environments. All of these aerosol effects and interactions significantly depend on their chemical composition.

This thesis focuses on aerosol chemical composition, measured by mass spectrometry. To make sense of the vast diversity of chemical compounds present in the atmospheric aerosol particles, we rely on chemical classifications (speciations). In this work, different types of traditional aerosol mass spectrometric chemical speciations are evaluated by their ability to explain measured aerosol physicochemical properties, such as volatility, hygroscopicity, and optical properties.

While the parameter estimates for aerosol physicochemical properties derived from traditional chemical classifications do correlate with experimental data, the estimates fail to adequately capture the observed variabilities. Whether this is due to deficiencies in speciation schemes, flaws in assumptions or models, or measurement uncertainties, is not apparent. However, it seems evident that there is room for improvement in all of the above-mentioned fields.

During the course of this work, chemometric methods, (i.e. the application of mathematical and statistical methods to the analysis of chemical measurements) were implemented in the analysis of aerosol mass spectra. The results clearly show the value of statistics-based, machine learning oriented methods for feature extraction and chemical classification. Such chemometric approaches maximise the amount of information available for an analyst of aerosol mass spectrometric results, and enable making better informed interpretations. Combining methods such as data factorisation and clustering can overcome typical limitations of these methods when used as stand-alone techniques. Chemometrics-based, machine-learning-like methods thus show considerable potential for advancing aerosol chemical analysis.

Finally, chemometric tools were shown to be capable of producing new, comprehensive, mathematically and statistically robust chemical classifications, which again contribute to our understanding of the atmospheric aerosol properties, interactions, and effects.

Keywords: atmospheric aerosols, mass spectrometry, chemometrics, chemical speciation

Contents

List of publications

This thesis consists of an introductory review, followed by six research articles. In the introductory part, the Papers are cited according to their roman numerals. All the Papers are reproduced under the Creative Commons Attribution Licence.

I Häkkinen, S., Äijälä, M., Lehtipalo, K., Junninen, H., Backman, J., Virkkula, A., Nieminen, T., Vestenius, M., Hakola, H., and Ehn, M.: Long-term volatility measurements of submicron atmospheric aerosol in Hyytiälä, Finland, Atmospheric Chemistry and Physics, 12, 10771-10786, 2012.

II Zieger, P., Aalto, P., Aaltonen, V., Äijälä, M., Backman, J., Hong, J., Komppula, M., Krejci, R., Laborde, M., and Lampilahti, J.: Low hygroscopic scattering enhancement of boreal aerosol and the implications for a columnar optical closure study, Atmospheric Chemistry and Physics, 15, 7247-7267, 2015.

III Hong, J., Kim, J., Nieminen, T., Duplissy, J., Ehn, M., Äijälä, M., Hao, L., Nie, W., Sarnela, N., Prisle, N. J. A. C.: Relating the hygroscopic properties of submicron aerosol to both gas-and particle-phase chemical composition in a boreal forest environment, Atmospheric Chemistry and Physics, 15, 11999-12009, 2015.

IV Hong, J., Äijälä, M., Häme, S. A., Hao, L., Duplissy, J., Heikkinen, L. M., Nie, W., Mikkilä, J., Kulmala, M., and Prisle, N. L.: Estimates of the organic aerosol volatility in a boreal forest using two independent methods, Atmospheric Chemistry and Physics, 17, 4387-4399, 2017.

V Äijälä, M., Heikkinen, L., Fröhlich, R., Canonaco, F., Prévôt, A. S., Junninen, H., Petäjä, T., Kulmala, M., Worsnop, D., and Ehn, M.: Resolving anthropogenic aerosol pollution types–deconvolution and exploratory classification of pollution events, Atmospheric Chemistry and Physics, 17, 3165-3197, 2017.

VI Äijälä, M., Daellenbach, K. R., Canonaco, F., Heikkinen, L., Junninen, H., Petäjä, T., Kulmala, M., Prévôt, A. S. H., and Ehn, M.: Constructing a data-driven receptor model for organic and inorganic aerosol – a synthesis analysis of eight mass spectrometric data sets from a boreal forest site, Atmospheric Chemistry and Physics Discussions, https://doi.org/10.5194/acp-2018-1069, in review, 2018.

1 Introduction

1.1 Aerosols, climate, and air quality

The atmosphere of our world is integral to most of the life forms on our planet, and breathing the same air is something that connects all of us despite of our differences. Thus, the many environmental concerns related to Earth's atmosphere are also common to all of us.

Anthropogenic climate change certainly ranks among the greatest human challenges of our time, extending its effects to many areas of life, and it shapes the future for human societies and lives of other species alike (IPCC, 2018). Another threat which is less existential, but harder to disregard in the daily lives of people, is the pollution of our atmosphere. While the most severe outcomes of global warming will not unfold until in the seemingly distant future, air pollution is the leading contemporary cause of environment-related illnesses (Lelieveld et al., 2015; Shiraiwa et al., 2017) and prematurely lost lives, resulting in seven million premature deaths every year – five times the mortality rate related to traffic injuries (WHO, 2016). Addressing and mitigating the effects of climate change and degraded air quality require grasping the scale and severity of these issues and gathering the political momentum and scientific knowledge required to act on them. Atmospheric research is integral in solving these grand challenges (Rockström et al., 2009; Kulmala et al., 2016), as science needs to provide reliable information for the basis of informed decisions, ways to evaluate and predict the effectiveness of the possible courses of action, and to provide new solutions in the form of tools and technologies.

The basic principle of how increased carbon dioxide levels in the atmosphere result in global warming was understood already more than 120 years ago (Arrhenius, 1896; Jacobson, 2012). Global warming of the late 20th century was rather accurately predicted already in the 1970's (Sawyer, 1972), and its consequences and the necessity of changes spelled out at that time (Mesarovic and Pestel, 1974; Meadows, 2014). Recently, the remaining deficiencies in our understanding of climate change relate to the role and importance of various temperature feedbacks of our planet, e.g. heat absorbed in oceans, decreased albedo due to snow melt at high latitudes, and the effects of atmospheric aerosols, the fine particles suspended in air. Specifically, atmospheric aerosols interact with (i) incoming sunlight by scattering and absorbing radiation, as well as (ii)

atmospheric water vapour; aerosols act as "seeds" for cloud droplets, affecting cloud formation, properties (brightness, lifetime), and rainfall. All of these processes modify the temperature of our planet, but the knowledge on the magnitudes of the effects remains inadequate (IPCC, 2013; Burkart et al., 2011; McFiggans et al., 2006).

Along with reactive trace gases, such as nitrous oxides (NO_x) and tropospheric ozone (O_3), aerosols are also among the most important air pollutants causing adverse health effects (McNeill, 2017). Understanding the nature, effects, and interactions of atmospheric aerosols is thus an important piece in the atmospheric puzzle we are hoping to solve.

1.2   Aerosol sources, effects, and interactions

While the need to label and classify things is profoundly human, attempts to dichotomously classify the extremely diverse assortment of gases and aerosol particles is difficult. We can try to divide the particles according to their formation mechanisms - based on if they were injected to the atmosphere (primary aerosols) - or formed therein via gas-to-particle conversion (secondary aerosols). However, it is usually impossible to ascertain where the seed of the aerosol particle came from and, for example, many combustion aerosols are considered primary, even when actually formed in the emission plume outside of a smokestack or a tailpipe (so called "delayed primary aerosol"; e.g. Rönkkö et al., 2017). Likewise, the aerosol sources can be from human activity (anthropogenic) or natural, with the latter often dubbed as biogenic aerosols. This can cause confusion as well, for e.g. agricultural aerosol emissions or those created out of cutting down a forest, or aerosols from burning of biomass, all of which can be thought of as both biogenic (originating from biological processes) and anthropogenic. To be able to discuss and describe different types of aerosols, use of labels is, however, certainly useful and necessary – it is just important to note that most of the labels and classifications are not mutually exclusive.

Aerosol particles originate from a variety of sources. Globally the largest natural aerosol sources in terms of particle mass are suspended soil dust and sea salt, due to their typically large particle sizes. Atmospheric secondary aerosol formation also contributes significantly to aerosol mass, and is a dominant source in terms of aerosol number concentration (Merikanto et al., 2009). Other natural sources include botanical debris and pollen, volcanic dust and sulphates, as well as smoke from forest fires. Main

anthropogenic sources include gas-to-particle conversion as well as primary particles from e.g. combustion engines (transportation), coal burning in electricity production and heating, cement manufacturing, metallurgy and waste incineration. (IPCC, 2002).

Atmospheric aerosol particles come in many shapes and sizes. The smallest of them only consists of a few molecules clumped together, with sizes of few nanometres (nm), while the coarse particles visible to the eye can have sizes comparable to the diameter of a human hair (around 100 μm). Similar variability exists in their shape and composition, which can be anything from long agglomerate chains of black carbon soot, to crystallised salt, to liquid droplets of organics. It is thus understandable that the different aerosols also exert different effects when interacting with light and water vapour of the atmosphere, or when deposited inside human respiratory organs (Atkinson et al., 2015; Shiraiwa et al., 2017). In physics, such climate-relevant aerosol interactions are often described with physicochemical parameters, such as hygroscopicity (expressing the water affinity of a substance), solubility (how easily a compound is dissolved in e.g. water), volatility (the tendency to evaporate or condense), and scattering and absorption coefficients (what happens to light or photons when they encounter a surface). For health effects, the aerosol qualities of interest are their redox activity (capacity to cause reductive and oxidative stress) as well as toxicity, allergenicity, and carcinogenicity (whether they are poisonous or cause allergy or cancer). Inhalation of aerosol particles and their deposition into the respiratory tract can also result in physical scarring (lung fibrosis; Shiraiwa et al., 2017), and small nanoparticles have been shown to be capable of crossing the epithelial and blood-brain-barriers to enter blood circulation and even the brain, where they may potentially influence the central nervous system (Semmler et al., 2004; Oberdörster et al., 2004), and contribute to a wide range of chronic and acute diseases.

1.3 Mass spectrometry, chemometrics, and chemical speciation

Of the various aerosol particle quantities, basic physical qualities such as size, shape, and electric charge, are often the easiest parameters to measure. Studying the chemical composition of particles has traditionally proven much harder, due to the miniscule particle mass available for chemical analysis. In extremely polluted environments such as megacities, the total mass of aerosol particles can reach a few milligrams per cubic

meter, but even these extremely high particle concentrations only represent approximately one millionth of the mass of air in the same volume. Fortunately, refinement of mass spectrometric methods in the recent decades has enabled detailed, on-line chemical analysis of these extremely small amounts of mass. Essentially, a mass spectrometer measures the elemental or molecular composition of a substance, typically yielding a spectrum of signals over a range of atomic masses, from which chemical composition of the analysed sample can be inferred.

While the principle is straightforward for samples containing only a few types of molecules, an atmospheric aerosol can contain dozens of inorganic and tens to hundreds of thousands of different organic compounds. An atmospheric aerosol mass spectrum thus represents the sum of all these components. Although mass spectrometry has been an established method already from the early 1900's, the interpretation of mass spectra has long relied on the expertise of analytical chemists dedicated to the topic. However, even the most perceptive and experienced analyst can only process a fraction of all the information present in very complex mass spectra. The problem can be seen in the context of "big data", an over-abundance of information for manual analysis, with interdependencies too complex for human comprehension to pick up from the raw numbers or typical sample statistics values.

Since the 1990's, the scene has considerably changed – modern computer capacity and adoption of advanced statistical methods formerly too cumbersome to run effectively has made it possible to produce advanced and complex analyses of enormous amounts of data with relative ease. As a result, the scientific field of chemometrics, *"using mathematical and statistical methods to provide maximum chemical information by analyzing chemical data",* (Kowalski, 1975) seems to have entered a golden age with no end in sight.

As the title of this work suggests, one of the focuses of this thesis is the topic of "chemical speciation". Here I define it broadly as the process of classifying chemical components to specific classes, "chemical species". Analogous to the word "classification", speciation can equally refer to the classifying process or the resulting system of classes.

Chemical speciation is a practical necessity: with the currently available experimental techniques and theoretical frameworks it is not possible to construct an exact description of what atmospheric aerosol consists of. Nevertheless, as often is the case in nat-

ural sciences, this need not deter us from constructing approximate descriptions that serve as useful tools for understanding the phenomena, despite their in-built, known limitations.

In this thesis, I approach the topic of chemical speciation both from the traditional viewpoint of accumulation of chemical knowledge, as well as from an information theoretical perspective, where I essentially consider it as an exercise in statistical classification.

1.4   Machine learning - new tools for analysis of scientific data

Modern buzzwords like neural networks, machine learning, and artificial intelligence are frequently and increasingly appearing not only in commercial applications of big data on consumer behaviour and social networks, but also in the context of scientific data analysis. Early adopters of advanced statistical tools and machine learning methods have for a time been found within the practitioners of biological and medical sciences, who similarly wrestle with problems of large data volumes. A scientific big sister to chemometrics - bioinformatics, i.e. application of advanced statistical, mathematical, and machine-learning methods for biological data - has already revolutionised research in biological sciences, especially in genetics (Diniz and Canduri, 2017). Although not as of yet equally widespread, aerosol sciences and aerosol mass spectrometry have also seen some emergent chemometric applications that begin to qualify as machine learning methods. So far, they have mostly been limited to data dimensionality reduction (i.e. mapping large sets of data into a more manageable amount of variables with minimal loss of information), but it is increasingly evident that the potential uses of advanced statistical methods are much more diverse.

The theme of this dissertation relates to this paradigm shift in the analysis of aerosol chemical data, for aerosol mass spectrometry in particular. The first publications (Papers I to IV) present how traditional chemical classifications are used in explaining aerosol physicochemical properties. While the attempts did meet some success, it seems evident that there is a gap between the estimates for physicochemical parameters derived from currently widely used chemical speciations and the actual, experimentally measured values of those quantities. Although the reasons for these discrepancies are likely to be many, I suggest in this work that the gap might be partly bridged by using more relevant and robust chemical classifications. The latter portion of this work

explores the possibilities of advanced, statistical models and machine learning tools in deriving an improved description of aerosol chemical composition.

In the end, this thesis forms but one link on the chain from (1) the experimental measurements of the atmospheric aerosols, to (2) describing their nature, composition, and interactions, and finally (3) understanding how they affect Earth's atmosphere and human health. The value of specific results derived from this work, such as the improved chemical speciation I propose in Papers V and VI of this thesis, is determined by their utility in future studies on physicochemical properties, in the theme of Papers I to IV of this thesis. However, what seems already evident is that advanced statistical and machine-learning applications will certainly have an important role to play in further advancement of modern aerosol science.

1.5 Aims of this thesis

The objectives of this thesis are:

1. To derive and develop aerosol chemical classifications from aerosol mass spectrometric data, to enable connecting chemical composition to aerosol physicochemical properties regarding aerosol

   (a) volatility (Papers I, IV)
   (b) hygroscopicity (Paper III)
   (c) light scattering (Paper II).

2. To incorporate chemometric and machine learning methods and to combine them in new ways in the analysis of aerosol mass spectra, in order to

   (a) overcome limitations in traditional aerosol speciations (Papers V, VI)
   (b) reduce subjective bias in interpretation of mass spectra (Papers V, VI).

3. To develop an improved, objective and statistically robust chemical speciation to

   (a) maximise chemical information attainable from aerosol mass spectrometric data (Paper VI)
   (b) provide better chemically speciated data for use in future studies of aerosol physicochemical properties (Paper VI).

2 Scientific background

The first part of this thesis work concentrates on testing the usability of traditional aerosol chemical classifications (speciations) in predicting the observed physical qualities and properties, such as aerosol response to temperature changes and their interactions with light and atmospheric water. In particular, the focus is on the propensity of particles to take up water (hygroscopicity), as well as their evaporation (volatility) and light scattering properties.

The latter part of the work concentrates on developing more robust and consistent, statistics-based methods for aerosol chemical speciations, endeavouring to encompass a maximal amount of chemical information. Besides exploring chemical classifications, an equally important theme of the final part of this thesis is to highlight the usefulness and potential of the selected statistical methods, characteristic of machine learning, implemented to analysis of aerosol mass spectra.

2.1 Aerosol physicochemical properties

2.1.1 Basic properties of particles: number, size, charge, shape, and phase

Traditionally, the easiest parameters to measure describing the aerosol particles are physical ones – their number, size, mass, shape, and electric charge. Measuring the number of particles in a volume of air is an old branch of aerosol science since the 19th century (Aitken, 1889) and has since evolved to comprise of several techniques (McMurry, 2000b). Real-time particle counting methods are typically based on direct optical detection (optical particle counter, (OPC; McMurry, 2000a) or additionally growing particles via condensation of vapour prior to their optical counting (condensation particle counter, CPC; McMurry, 2000a). If the particles are electrically charged, or neutralised to their naturally occurring charge distribution (Fuchs, 1963), their detection is relatively easy, as techniques for ion detection are plentiful (e.g. Baron and Willeke, 2011).

The size of an aerosol particle is one of the most fundamental quantities necessary for their characterisation, and it strongly influences their interactions with other particles and surfaces, surrounding carrier gas, as well as radiation, such as visible light (Seinfeld and Pandis, 2012). While aerosol particles can be spherical, allowing unambiguous

11

definition of their diameter, they can equally have other geometries, complicating the definition of their size, thus there exist several metrics for defining particle size, deriving from e.g. their aerodynamic properties or mobility in an electric field (Baron and Willeke, 2011). Measurement of particle size, or in case of an aerosol particle population, a distribution of particle sizes, is relatively straightforward for the largest particles (diameter > 1 μm), which can be detected optically (e.g. APS; Remiarz et al., 1983). For smaller particles (< 1 μm), electrical charging or neutralisation and observing their flight tracks in an electric field by e.g. a differential mobility particle sizer (DMPS; Aalto et al., 2001) is a typical technique.

Analysis of particle shape and phase state is possible for large particles, which can be characterised optically using various applications of microscopy (Baron and Willeke, 2011). Similar direct, physical characterisation of smaller particles is much more difficult. While some information of their shape can be derived from their aerodynamic properties relative to their size (Hinds, 2012), and clues to their phase state obtained from e.g. their bouncing or adhesion behaviour when impacting a surface (Baron and Willeke, 2011), determining the phase and shape of small particles remains problematic even up to date. Studies have shown that in addition to fully liquid and solid forms, with one or more liquid or solid layers (Buajarern et al., 2007), particle-phase aerosols can contain several phases (e.g. solid core particles with a liquid coating; Sobanska et al., 2014). Particles can also exhibit crystal-like or glassy forms (Virtanen et al., 2010), and can form long chain-like structures for e.g. soot (Bond et al., 2013).

2.1.2 Volatility

In Papers I and IV of this thesis, the focus is on aerosol volatility, i.e. the tendency of a substance to shift between gas and liquid or solid phases. Volatility determines whether a compound will prefer gas or particle phase in given environmental conditions and therefore connects together atmospheric trace gases and aerosols.

Low-volatile compounds, for example organics with high molecular weights (e.g. organic acids) typically have a very low saturation vapour pressure and thus prefer partitioning to the particles over the gas phase. Conversely, highly volatile compounds such as methane, toluene, and monoterpenes with high saturation vapour pressures strongly favour partitioning to gas phase in atmospheric temperatures and pressures. These predominantly gas-phase compounds, nitrogen (N_2) and oxygen (O_2) molecules excluded,

are often called atmospheric trace gases. Compounds that feature saturation vapour pressures between these extremes are typically labelled semi-volatile. Depending on their concentrations, prevailing temperature, and pressure, they partition in significant amounts to both the particle and gas phases. For the semi-volatiles, this partitioning is often sensitive to small changes in temperature, so that they may e.g. reside in the particle phase during the night when the air temperature is low, and volatilise to the gas phase during the day when the temperature is higher.

Due to the diversity of chemical compounds present in the atmospheric aerosol, the task of theoretically describing aerosol volatility is difficult. The physical parameter describing energy required to transfer substance from a liquid to gas phase is enthalpy of vapourisation (ΔH_{vap}; also commonly known as 'latent heat of evaporation'). While there exists an experimental formula connecting vapour pressure to enthalpy of vapourisation for organic compounds of aerosols (Epstein et al., 2009), a practical problem is, again, that the exact organic compounds present in atmospheric aerosol particles are counted in thousands (or probably tens to hundreds of thousands; Donahue et al., 2011). Even if it would be possible to somehow to count exactly the amounts of each of these compounds, which remains but a dream for an aerosol chemist, the enthalpies for the majority of these compounds are not available from measurements. Thus, the only way to currently reliably obtain aerosol volatility is to measure the actual vapourisation behaviour of the aerosols.

There exist theoretical models to describe evaporative response to the heating of aerosol particles (e.g. Riipinen et al., 2010) and a framework to describe the organic aerosol volatility distribution (Donahue et al., 2006; Donahue et al., 2011; Donahue et al., 2012). Furthermore, a connection between organic aerosol volatility and aerosol oxidation level has been shown in several studies probing aerosol response to heating (Massoli et al., 2010). A typical estimate from these examinations is, that on the order of 20 of organic aerosol mass is volatilised when heated from 0 to 50 °C (Huffman et al., 2009a; Huffman et al., 2009b; Raatikainen et al., 2010; Cappa and Jimenez, 2010).

Interestingly, the same studies also show that at the other end of the volatility distribution, there exist some extremely low-volatile aerosols. Even after heating to a very high temperature (around 280 °C), a considerable fraction of aerosol mass remains in the particle phase (e.g. Paper I). Although some of this mass is likely soot, also termed black carbon (BC), with a vapourisation temperature around 4000 K, black carbon does not fully account for these observations. The nature and composition of these

practically non-volatile, organic aerosol species remains largely unknown, which is why aerosol chemical examinations like the one exemplified in Paper I, are necessary.

2.1.3 Hygroscopicity

The changes in cloud cover are one of the main uncertainties in predicting climate change, and aerosols play an important role in it: they act as seeds for growing cloud droplets in the atmosphere. The fraction of particles that can begin to grow (also termed 'to activate') by condensation of atmospheric water vapour on their surface, are termed cloud-condensation nuclei (CCN). Whether or not a particle will activate for growth depends on several parameters, related both to properties of the particle (size, phase, composition, and layering) as well as to the environment (water vapour saturation).

The main parameter determining if a particle will activate as CCN is its size (Junge and McLaren, 1971; Dusek et al., 2006) – the larger the particle, the lower the vapour saturation (proportional to relative humidity, RH) it will need to activate. Although there exist estimates that the particle size distribution could explain 84 to 96 % of the variability in CCN concentrations (Dusek et al., 2006), this may only be the case for polluted air masses, as it has later been reported that predictions for CCN concentrations based on size distributions alone can be often be inaccurate (Burkart et al., 2011; Hudson, 2007). Specifically, the activation of smaller particles is sensitive to chemical composition (e.g. Chang et al., 2017). Thus, for better CCN estimates, particle composition needs to also be accounted for.

In order to fully understand interactions between aerosols and atmospheric water, a description of how particles respond to water vapour saturation is needed; that is whether or not water condenses or evaporates from their surface at the given environmental conditions. This hygroscopic property can be approached from an experimental point of view, by measuring the change in particle size when environmental RH is changed: the change in particle size due to condensed water on its surface can be readily measured. The physical parameter associated with particle growth by water condensation is the hygroscopic growth factor (HGF) defined as the ratio between the particle size at high humidity (RH = 90 %) relative to its size when dried (RH < 15%).

The hygroscopic properties of inorganic particles are generally well known. Typical inorganic salts in atmospheric aerosol, such as ammonium nitrate and sulphate and

sea-salt, are typically hydrophilic, i.e. highly hygroscopic. On the other hand, black carbon (BC) is typically hydrophobic and repels water. The hygroscopicity properties of organics are generally more complex, and vary strongly by compound (Rastak et al., 2017; Pajunoja et al., 2015). Like in the case of volatility, a connection has been made between oxidation level and hygroscopicity: oxidised, highly aged particles have been found to be more hygroscopic than aerosols consisting of fresh, less oxidised organics (Massoli et al., 2010). To date, the hygroscopicity models are very sensitive to organic aerosol hygroscopicity estimates (Rastak et al., 2017), and further research is needed to complete the picture on how both chemical composition and size could be best accounted for to produce more accurate models on hygroscopic growth and CCN activation. In Paper III of this work, we touch upon the topic by examining the connection between experimentally measured HGF and the aerosol chemical composition.

2.1.4 Optical properties

Optical properties of aerosols are important for climate and visibility, as aerosol particles scatter and absorb incoming radiation. The effect that aerosol particles have on attenuation of light in the atmosphere is a million times larger than the effect of similar mass concentration of water vapour molecules and a thousand times more than the equivalent amount of water present in raindrops. Therefore, aerosol particles can affect visibility in much lower concentrations (Hinds, 2012). Atmospheric aerosol particles can impair visibility locally or regionally, in environments with high aerosol loadings (e.g. desert dust suspended by winds, haze formed from biogenic precursor gases in forested areas; Went, 1960) or anthropogenic aerosols in polluted cities (White, 1976). They also affect Earth's radiative balance by scattering incoming sunlight and to a lesser extent, outgoing infrared radiation.

Almost all of the atmospheric aerosol particles scatter away incoming sunlight, generally cooling the atmosphere, but some dark coloured aerosols (mainly BC, brown carbon, and dust) can also absorb light, and thus cause atmospheric warming (Moosmüller et al., 2009). These dark aerosol particles continue to absorb light even after their deposition, which can decrease surface albedo for e.g. snow. The deposition of BC aerosol especially affects radiation budget in Arctic regions (Bond et al., 2013).

Important parameters for aerosol optical effects are the scattering coefficient (σ_a), which describes how effectively the aerosol particles scatter incoming light and the

absorption coefficient (σ_a), which similarly describes the light absorption effect by the particles. These coefficients strongly depend on particle size. Very small particles, with diameters less than 10 % of radiation wavelength λ (e.g. 40 - 70 nm particles for visible light $\sim 400 - 700$ nm), interact with radiation in a different way compared to larger particles, and scatter light quite similarly to molecules (Rayleigh scattering).

Very large particles, cloud, and rain droplets ($d > \lambda$, $\sim 100\,\mu$m), on the other hand, can be treated by laws of geometrical optics. The intermediary range, which incidentally encompasses a large fraction of aerosol particles, is the domain of the Mie theory, which describes the interaction between the electromagnetic (plane) fields of photons and spherical objects (aerosol particles), starting from Maxwell equations (Hinds, 2012). Although difficult to estimate theoretically due to uncertainties in aerosol refractive indeces and size distributions, scattering coefficient for a polydisperse aerosol population can be readily measured experimentally using a nephelometer (e.g. Anderson et al., 1996), an aerosol instrument described in Materials and methods.

Beside their size, light scattering by aerosols additionally depends on their refractive index, albedo, phase, and internal structure. These properties are strongly affected by their chemical composition. While the size-dependencies of aerosol optical phenomena are generally well understood, the effects of chemical composition are more ambiguous. An important optical effect for atmospheric aerosols is the scattering enhancement by water vapour condensing on aerosol particles. Since water is often abundant in the troposphere, this process can have pronounced effects for the aerosol scattering properties due to hygroscopic growth (Malm and Day, 2001; Paper III). In Paper III, we specifically focus on this RH dependent enhancement of aerosol light scattering, $f(RH)$, and its relation to particle chemical composition.

2.2 Analysis of aerosol chemistry

2.2.1 Chemical composition and chemical speciation

Chemical characterisation of aerosol particles has historically been difficult, as most chemical analysis methods require samples with larger mass than is commonly available in atmospheric aerosol. The traditional solution has been of prolonged (off-line) sampling of particles for analysis. However, this can result in measurement artefacts via physical and chemical processes that alter the sample after collection (such as

16

evaporation and chemical reactions; e.g. Kristensen et al., 2016). Fortunately, during recent decades the rapid evolution of aerosol instrumentation has finally enabled on-line chemical characterisation of aerosols using in-situ mass spectrometry (McMurry, 2000b; Canagaratna et al., 2007). Even so, off-line methods continue to provide complementary information, such as quantification of functional groups present in molecules (e.g. Fourier transform infrared spectroscopy, FTIR; Griffiths and De Haseth, 2007; Corrigan et al., 2013) via spectroscopic methods.

It is currently impossible to describe aerosol chemical composition on an exact, compound-by-compound basis, due to (1) the abundance of chemical species (likely to span at least tens of thousands of organic compounds; Goldstein and Galbally, 2007), and (2) the lack of a universal, quantitative measurement technique applicable for all the compound types. Thus, cruder classifications to chemical "species" and groups of compounds are inevitable in practice.

A very typical basic chemical speciation is the dichotomy of 'organic' and 'inorganic' compounds. This division is also commonly used for atmospheric aerosols, often with the additional sub-classes for different types of inorganics, such as sulphates (SO_4), nitrates (NO_3), ammonium (NH_4), and black carbon. BC (also sometimes termed EC for 'elemental carbon') consists of aggregates of small spheres, strongly absorbs visible light, and vapourises around 4000 K (Bond et al., 2013). Mineral dust species are usually classified separately, as are sea spray aerosols whose particles mainly consist of sea salt. Additionally, primary biogenic aerosol particles (such as pollen, spores, and plant detritus from the biosphere (Després et al., 2012), are sometimes considered separately.

The portion of aerosol chemical compounds containing organics is usually termed organic aerosol (OA), and the carbon in organic molecules as organic carbon (OC), in contrast to inorganic aerosol and elemental carbon (EC), respectively. There are several ways to divide organics into sub-species and no universal convention exists. However, as one of the main processes of shaping aerosol chemical composition in the atmosphere is oxidation, i.e. the chemical processes resulting in an addition of oxygen to molecules when exposed to atmospheric conditions, it is natural that many of the sub-speciations relate to this process. Although the rate of oxidation depends on many parameters (e.g. oxidant concentrations, chemical identity of the precursors, reaction rates, atmospheric composition, temperature, radiation), oxidation level, often described by the ratio between oxygen and carbon molecules in the compound (O:C; Aiken et al.,

2008), generally increases upon aerosol age, which is the time spent in the atmosphere after formation. Therefore, various oxidation-based classifications - for example if the aerosol is fresh or aged - are often used. Notably, as discussed in the sections above, oxidation level also correlates with aerosol physical properties, making the organic aerosol oxidation level based classification a useful tool for estimating physicochemical properties.

2.2.2 Aerosol mass spectrometry

Mass spectrometry is a traditional analytical chemistry method for chemical composition measurement. As the name suggests, it involves measuring a spectrum of detected signal from ions and molecules over a range of atomic mass units, which chemically characterises the measured sample. A mass spectrum is analogous to a light spectrum, which presents light intensity over a range of wavelengths, and characterises the source of the light (and/or the medium it was passing through).

The dictionary characterisation (Merriam-Webster online dictionary, 2018) of mass spectrometry reads:

> "*: an instrumental method for identifying the chemical constitution of a substance by means of the separation of gaseous ions according to their differing mass and charge*"

which implies three important stages common to all aerosol mass spectrometry applications. (1) The particle-phase sample needs to be converted from particles to gaseous phase, which can be accomplished by heating it (for example with a hot surface or a laser). (2) the sample gas needs to be ionised, for which there are many methods (e.g. bombarding the molecules with atoms, ions, electrons or radiation; or relying on ionisation taking place upon thermal decomposition of molecules). (3), the mass-to-charge-ratio for the ions resolved by observing movement, which is induced by an electric field. Within these boundary conditions, there exist a variety of specific methods of mass spectrometry (Gross, 2006).

Historically, the research field of mass spectrometry arguably already began with electromagnetic deflection of ion streams by Wien (1898), and resulted in first veritable mass spectrometers by Thomson (1922) and Aston (1919). The field, now at the ripe

age of a century, has since emerged as one the indispensable tools of analytical chemistry (Griffiths, 2008). Advances such as the quadrupole (Paul and Steinwedel, 1953) and time-of-flight mass analysers (Stephens, 1946) and electron [impact] ionisation (EI; Bleakney, 1929), developed during the 20th century have since enabled new applications, such as the aerosol mass spectrometer (AMS; Canagaratna et al., 2007) used in this work. The technical details of the AMS used in this work are described in Materials and methods.

2.3   Statistical methods

2.3.1   Chemometrics

The feat of quantitatively measuring thousands of unknown compounds, however, produces data in very large quantities, presenting a huge challenge for any early adopters of mass spectrometry. Even today, the management, storage, and especially analysis and interpretation of this overabundance of information still challenge an analyst of aerosol mass spectrometric results. Recognising the issue, Bruce R. Kowalski wrote in his 1980 article titled "Chemometrics" (Kowalski, 1980):

> "[S]ince the measurements made by analytical chemists are associated with some degree of uncertainty and an analytical result is usually derived from a mathematical formula, it is difficult to conceive of a more perfect marriage than analytical chemistry and statistics and applied mathematics. Unfortunately, at this time the relationship between these fields can be characterized only as an engagement."

Since then, the rapid development of computer capacity and mathematical analysis tools has allowed the bridging of the gap between experiments and their scientific interpretation in many fields of mass spectrometry (Belu et al., 2003), leading to a variety of statistics applications and forming one of the oldest branches of chemometrics (Kowalski, 1981; Vandeginste, 1982; Kowalski, 1975; Shoenfeld and DeVoe, 1976). Some of the chemometrics themes outlined originally (Kowalski, 1980) involve de-convolution, pattern recognition, parameter estimation, optimisation (of analysis procedures and experiments), information theory, wave transform and image analysis. These recurring themes seem as relevant even currently, nearly 40 years later, with some obvious additions such as neural nets (again, originally an idea from 1940's and

50's and revived in the late 1980's; Hebb, 1949; Rosenblatt, 1958, Rumelhart and Mc-Clelland, 1986). This thesis encompasses several different chemometric applications; among others parameter estimation (Papers II, III and VI), de-convolution, pattern recognition and optimisation (Papers V, VI).

2.3.2 Machine learning

The field of machine learning lies at the intersection of the fields of statistics and computing. It also closely relates to chemometrics. In a previously cited chemometric article, Kowalski (1975) also notes:

> *"In 1969 a series of papers concerned with applications of learning machines [. . .] to the determination of molecular structural features directly from spectral data [. . .] appeared in Analytical Chemistry. Shortly thereafter, a review described early results of such applications [. . .]. Thus began a search for new mathematical methods to solve multivariant problems in chemistry."*

Originally arising from a study of whether a computer could play a game of checkers (Samuel, 1959), machine learning is generally understood as the study of how computers can be utilised in data analysis, by e.g. making predictions, classifications, or performing dimensionality reduction. It literally implies that a machine (i.e. computer machine) learns to perform a task (e.g. among the above mentioned), with or without human supervision. While is it closely tied to more applicative fields such as chemometrics or bioinformatics, machine learning itself is not limited to a particular domain of applicability.

While the differentiation between statistical methods, chemometrics, and machine learning (even artificial intelligence, AI) are ambiguous and overlapping, the multivariate methods, feature extraction, and classification tasks described in Papers (II, III), IV, and V are certainly chemometrical. The (semi)supervised, iterative applications of multiple chemometric methods and algorithms (Papers V and VI) also fulfil the definition of what is often considered machine learning. For example, a pioneer of the field, Tom M. Mitchell (2006) describes the difference between statistics and machine learning as follows:

"Whereas Statistics has focused primarily on what conclusions can be inferred from data, Machine Learning incorporates additional questions about what computational architectures and algorithms can be used to most effectively capture, store, index, retrieve and merge these data, how multiple learning subtasks can be orchestrated in a larger system, and questions of computational tractability."

While the term AI is often very loosely used, I would associate it with deep learning (e.g. LeCun et al., 2015) and thus do not consider the term an appropriate label for the methods of this thesis.

3 Materials and methods

In this section, I have included short descriptions of the instruments and the main statistical tools used in this thesis work. To keep this article introductionary, I will refer to the Papers and their references as well as the journal articles mentioned below for the specific details and technicalities.

3.1 Instrumentation

3.1.1 Aerosol mass spectrometer

The main instrument in this work is the Aerosol Mass spectrometer (AMS; Jayne et al., 2000; Jimenez et al., 2003; Canagaratna et al., 2007). The AMS was developed in the late 1990's by Aerodyne Research Inc. (Billerica, MA, U.S.), and has since become a household name in aerosol chemical measurements. There exist several variants and development versions of the AMS, with various improvements from the early versions as well as modifications catering for specific measurement needs. Due to the difficulties of measuring the small particle mass amidst the carrier gas of air and trace gases, advanced technical and data analytical solutions are necessary. This makes the AMS a relatively complicated instrument, compared to most of the other aerosol instruments presented above.

The AMS variant used in most of the papers of this thesis (I, V, and VI) is described by Drewnick and co-workers in their 2005 article (Drewnick et al., 2005), and the particulars of the specific instrument we used are discussed in Paper V. Other AMS variants used during the course of this thesis work include the high-resolution AMS variant (HR-AMS; Papers III and VI; Decarlo et al., 2006) and the "simplified" AMS version for monitoring use, the Aerosol Chemical Speciation Monitor (ACSM; Ng et al., 2011). Below, a short description of the main operating principles of an AMS is given.

As in all spectrometers, the AMS interior contains a vacuum, where the chemical measurement takes place. The interface between the atmospheric pressure and the vacuum chamber consists of a critical orifice, which limits the inflow, and is followed by an aerodynamic lens system (Murphy and Sears, 1964). The multiphase lens system focuses the particles into a tight beam. The losses of large particles at the lens and losses

of smaller particles from the beam limit the size range of particle AMS can measure to approximately 40 to 600 nm (Liu et al., 2007). The particle beam from the lens is directed to impact a heated ceramic plate, the vapouriser, with a temperature of 600 °C. The temperature is one of the main factors defining which species can be detected by the AMS: the species that experience flash-vapourisation upon impacting the vapouriser are termed 'non-refractory', while the ones that have higher vapourisation temperatures and are not (quantitatively) detected are 'refractory' species. Typical refractory compounds include for example BC, NaCl, metals, and most minerals.

The resulting sample vapour is ionised by bombarding the gas with electrons (70 eV electron ionization, EI) from a tungsten filament. The side effect of this universal, well understood ionization method is extensive fragmentation of the sample molecules, which severely smears out the unique characteristics of the parent molecules. The positively charged ions and their fragments are then guided to the mass analyser by an electric field. In the modern ToF mass analyser the ions are orthogonally accelerated to a flight track by a pulsed electric field. The track is lengthened by curving it with additional electric fields, termed reflectors (the reflectron ToF analyser). At the end, the ions are counted and their arrival time registered using a microchannel plate detector (MCP; e.g. Wiza, 1979).

The signal measured by the detector then needs to be interpreted. A back calculation is needed through all the processing phases of the measurement. The main phases include the calculation of sample ion mass-to-charge ratios from the flight time results, as well as accounting for the fact that only a miniscule fraction of the original vapour molecules (typically 1 out of 10 million, 10-7) were ionised. These calculations require various calibrations, outlined in detail in AMS related scientific literature (Jimenez et al., 2003; Allan et al., 2003; Allan et al., 2004). Performing these calculations allows for quantification of mass, i.e. representing signal response of the instrument to a mass concentration of aerosol entering it.

The basic AMS data processing is the ion speciation, which is achieved using a "fragmentation table" containing a plethora of accumulated chemical information on fragmentation patterns, isotopes, and air composition (Allan et al., 2004). The basic ion speciation divides the ions of the mass spectrum to organics (org), sulphates (SO_4), nitrates (NO_3), ammonia (NH_4), and chlorides (chl). This speciation provides the basis on which further speciations are built, or a reference against which they can be compared, as done in the final paper of this thesis.

3.1.2 Aethalometer

The aerosol Aethalometer optically measures black carbon (Hansen et al., 1984; Magee Scientific, Berkeley, CA, U.S). It derives its name from classical Greek verb 'aethaloun', which translates as "to blacken with soot" (Hansen, 2005). The principle of operation is simple: an aerosol sample is collected on a spot on a roll tape of quartz filter, which is automatically rolled forward to the next spot at a set time interval. Optical analysis of the collected sample is continuously performed as the sample is being collected, allowing for continuous, automated operation. An aethalometer is very suitable for long, continuous measurements, and requires little maintenance. The optical measurement technique is based on light attenuation through the filter, and can typically be measured for seven different wavelengths. As aerosol is gradually collected on the filter, the sample spot is blackened, increasing the light attenuation, to which BC concentration is assumed linearly proportional (Hansen et al., 1984). From the measured intake flow of air and the rate of rise in light attenuation, BC mass concentration in (ng/m^3) of the sample can be calculated with a time resolution of 5 to 10 min (Bond et al., 1999; Virkkula et al., 2007).

3.1.3 Nephelometer

Another classical aerosol instrument, designed in the 1940's (Buettel and Brewer, 1949), the nephelometer also gets its name from classical Greek – the noun 'nephos' meaning 'cloud'. An integrating nephelometer measures the light scatter coefficient of aerosol particles by illuminating the sample aerosol with light of variable wavelengths and detecting how much of it is scattered, by integrating the scattered light signal over a half circle (Uin, 2016; Anderson et al., 1996). The scatter by carrier gas is subtracted from the result for total aerosol, yielding the light scattering coefficient (σ_s) due to particles only.

While a typical nephelometer measures σ_s for dried particles, the "WetNeph" instrument (Ecotech Pty Ltd., Aurora 3000; Fierz-Schmidhauser et al., 2010) used in in Paper II of this thesis was equipped with a humidifier unit, allowing for measurements of scattering coefficients for humidified as well as dried particles. This allows for calculation of scattering enhancement, i.e. ratio of σ_s for particles at dry and humidified conditions, $f(RH)$, which was at the focus of research in Paper II.

3.1.4 V/H-TDMA, V-DMPS

As can be deduced from the name alone, the volatility/hygroscopicity tandem differential mobility analyser (V/H-TDMA) is a complex assemblage of aerosol instrumentation. It is perhaps easier to begin understanding its function by investigating its components separately. A differential mobility analyser (DMA; Reischl, 1991; Knutson and Whitby, 1975) allows the classification of aerosol particles by their electrical mobility. The particles are typically dried and electrically charged and their flow is then led through a cylinder with an electric field. Only particles with the requested electrical mobility diameter survive without impacting the cylinder walls, and can subsequently be measured. This allows for aerosol size distribution measurement by combining with e.g. a CPC (McMurry, 2000a), and by scanning over the particle size range by altering the electric field strength.

In a volatility DMPS (V-DMPS; Ehn et al., 2007), the size distribution is measured twice: once for the unaltered atmospheric sample, and a second time after directing it through a thermodenuder (an oven) heated to a set temperature. The difference in the size distributions represents how much of the sample was vaporised. By altering the oven temperature, a volatility distribution can be obtained. This setup was used in Paper I. In a volatility tandem DMA (V-TDMA; Hakala et al., 2017), a certain particle size is selected with one DMA, after which the particles are passed through a thermodenuder. The resulting size distribution is measured using a second DMA and a CPC. The observed particle size is then a function of the volatility, as more volatile material will have evaporated, causing the particles to shrink. The decrease in diameter due to heating can be directly related to the volatility of the particles. Typically, the measurement is performed for several combinations of particle sizes and temperatures. Alternatively, instead of the thermodenuder, a humidifier can be installed and the sample flow analogously humidified to a set RH, resulting in a hygroscopicity tandem DMA (H-TDMA). The difference between the particle size before and after humidification represents particle hygroscopic growth, and a hygroscopic growth factor (HGF) can be calculated out of it. Both the volatilisation and hygroscopicity functionalities can also be combined in an instrument, forming the V/H-TDMA, which was used in Papers III and IV. Such a setup also allows for additional measurement options, such as subsequent volatilisation and humidification stages, resulting in increasingly complex technical solutions and data interpretation. For a non-redacted technical description please refer to Hakala et al. (2017).

3.2 Chemometric tools

3.2.1 Factor analysis and Positive Matrix Factorisation

The single most important advanced data analytical tool of this work is doubtlessly factor analysis (FA). This old statistical method was originally developed for use in psychology by Charles Spearman (1904), to study the psychological and physiological variables connected to human intelligence. Subsequently FA developed and expanded to include multiple factors to explain the observations (Burt, 1917; Carey, 1915; Vincent, 1953).

Statistically, factor analysis aims to describe variability in data, by constructing factors out of correlated variables. Since the number of factors (f) is usually much lower than the number of original variables (v), the factor model for a number of observations (n) contains less dimensions (f x n) than the original data (v x n). Thus, factor analysis is also a method for data dimensionality reduction. In essence, it is a method of compressing the information (on data variability) contained in a large number of variables, into a small number of factors, while minimising the loss of information in the process. As the number of factors (f) is typically a free parameter selected by the analyst, and due to lack of decisive mathematical diagnostics available, selecting the optimal f value for the factor model often forms one of the main challenges in factor analysis.

From another point of view, factor analysis can be understood as feature extraction, referring to the notion that the numerous changes we observe in multivariate data actually reflect changes in a small number of latent, underlying factors. Thus to understand and interpret complex observations with large numbers of variables, it suffices to understand a handful of factors – a much more humane task for the data analyst.

In aerosol mass spectrometry, the factor analysis of mass spectra measured from ambient aerosol would be understood as uncovering the aerosol types that arrive at the measurement site. An inherent property of factor analytical models is that the factorisation solution is typically not unique, meaning that there are typically multiple, statistically sound ways to describe the data with a different number and type of components. This common property of factor analytical solutions is often termed 'rotational ambiguity', as mathematically the alternative solutions can be described as rotations of the factor axes in the multidimensional variable space. The different descriptions

may for example reflect aerosol differences in relation to their sources, or equally their chemical characteristics or role in an atmospheric process (e.g. sources and sinks) – all of these descriptions may be "equally correct", mathematically and physically – they are simply alternative views to the scene described by the data. Along with the selection of f, choosing an appropriate rotation is often challenging and prone to analyst subjective decisions.

Another important aspect of factorisation is that it assumes that the observation is a linear sum of all the components, and the components and the observation must be commeasurable – in the same units. This limits the valid use of factorisation to solving problems where we can somehow think of the observation to consist of a sum of components in a rather concrete way. In environmental sciences most valid cases for factorisation involve positive quantities only (e.g. Paatero, 1997), which rules out many mathematically sound but physically unrealistic factorisation solutions. Integral to all experiments in natural sciences is also the concept of uncertainty or the error estimate of how reliable the observation is. Tailored especially to include these two characteristics typical of environmental sciences, Pentti Paatero and co-workers developed in the 1990's a specific factorisation model named Positive Matrix Factorsation (PMF; Paatero, 1994, 1999), which caters to the needs of environmental data. The main differences of PMF in comparison to most other factorisation methods are that the variables and factors are all constrained to positive values only. PMF also gives the observations with low relative uncertainty a higher weight in deciding the solution, and vice versa. This has the positive effect of erroneous or poor quality measurements not having a large influence on how a larger data set is factorised.

In Papers V and VI, I try to circumvent the traditional factor analytical issues of subjectivity in selection of f and the rotation (e.g. Paatero et al., 2012, 2014), by using statistical classification methods to evaluate and interpret the PMF output. As described in Papers V and VI posteriori classification often reveals which of the data structures uncovered by PMF are stable and interpretable, and which are mathematical artefacts or outliers.

3.2.2 Classification and clustering

Another important statistical method featured in this thesis is unsupervised, statistical classification of aerosol mass spectra. In the last two Papers (V and VI) of this work, clustering methods were applied to classify de-convolved mass spectra based on their chemical characteristics. Like factorisation, clustering has a long history, dating back to the 1930's (Blashfield and Aldenderfer, 1988), when it was first developed for anthropological sciences, for dealing with data on cultural relationships (Driver and Kroeber, 1932) and soon after adopted in psychology for classifying personality traits (Zubin, 1938).

Generally, clustering methods aim to classify data objects into a number of groups, 'clusters', based on their similarities. Analogous to the f parameter in factorisation, in clustering the number of clusters (k) also needs to be selected by the analyst. The groups are mutually exclusive in most of the applications, i.e. an object can only belong to one group, with the notable exceptions of "soft" clustering methods such as "fuzzy" clustering or Gaussian mixture methods. To function, clustering algorithms require as input the (dis)similarities (also "distances") between all the objects in a set of data (e.g. Kaufman and Rousseeuw, 2009) – in our case the (dis)similarities of mass spectra. This input data can be additionally pre-processed, to up-weight (or "scale") important, characteristic features in data, such as the less fragmented ions in a mass spectrum (Stein and Scott, 1994).

In a review of different clustering algorithms and methods H. Späth (1980) list as the basic clustering problems *"selection of distance, selection of algorithm, number of clusters to be formed and the choice of variables, especially their scaling"*. This quote highlights the fact that any new application of clustering requires decisions on several metrics and approaches to reach optimal results. The selection of a clustering method is undoubtedly of high importance in the context of this type of an effort. Lacking clear precedents for AMS spectra, I decided to select a simple, general, partitive algorithm that has been proven to perform efficiently and successfully on many different classification problems - the 'k-means' clustering algorithm. While newer and potentially more powerful clustering methodologies exist, I decided to begin with the traditional, tried-and-tested and versatile (Jain, 2010) k-means algorithm.

I refer to A. Jain (2010) in his review article "Data clustering: 50 years beyond K-means", where he summarises the history and development of methods:

> *"In spite of the fact that K-means was proposed over 50 years ago and thousands of clustering algorithms have been published since then, K-means is still widely used. This speaks to the difficulty in designing a general purpose clustering algorithm and the ill-posed problem of clustering",*

and, after a comparison of various methods and algorithms, states as one of his main conclusions:

> *"In most applications, it may not necessarily be the best clustering algorithm that really matters. Rather, it is more crucial to choose the right feature extraction method that identifies the underlying clustering structure of the data".*

In most applications, it may not necessarily be the best clustering algorithm that really matters. Rather, it is more crucial to choose the right feature extraction method that identifies the underlying clustering structure of the data". To make the other important selections outlined by e.g. Späth (1980) and Jain (2010), I needed a way to evaluate the solution quality of the clustering outcome. After a brief literature review, I decided to refer to the work of P. Rousseeuw (1987) on the "silhouette" method of interpreting the clustering solution quality. The mathematical basis is described in more detail in Paper V, but in short, the silhouette describes in relative terms how dissimilar a clustered object is with the members of the same cluster versus the objects in the nearest neighbouring cluster. The major task of optimising the clustering methods for aerosol mass spectrometric data is discussed in more detail in Papers V and VI.

3.3   The SMEAR II measurement site

All of the data on atmospheric aerosols for this thesis was collected from an atmospheric measurement station at Juupajoki, Finland. The research station was established in 1995 alongside the historical Hyytiälä forestry station (established in 1910 as a field station of the Finnish university, for training forestry experts). The site, Station for Measuring Ecosystem-Atmosphere Relations II (SMEAR II; Hari and Kulmala, 2005)

has been in the forefront of comprehensive atmospheric and ecosystem measurements since, and is currently among the best-known atmospheric research sites in the world. At the centre of the research focus is the effect of changing climate on forest ecosystems, using long, continuous time series of energy and mass fluxes between the atmosphere, ecosystem and soil.

The site is situated in the sub-arctic, boreal forest, and is often considered representative of the vast Northern Eurasian taiga biome, the largest global forest biome. The site is generally surrounded by forests in every direction, with an estimated 94 % at local 5 km radius (and 90 % of the nearest 50 km) of the nearby land as forest (including seedling/saplings) (Williams et al., 2011). The SMEAR II site is situated approximately 500 meters away from the forestry station. The immediate vicinity is a rather level area with a (monoculture) pine stand planted in 1962 after a prescribed burn, so at the time of writing this thesis in 2018, it is 56 years of age, with tree canopy height at approximately 20 meters. The forest is representative of typical, Finnish economically managed forest.

The station is comprehensively equipped with an extensive suite of state-of-the-art atmospheric and ecosystem scientific measurements (Hari and Kulmala, 2005), with emphasis on meteorological, aerosol, and forest ecology instrumentation. An impressive amount of auxiliary data is available for interpreting observations from SMEAR II, a benefit that becomes evident in the research papers of this work as well. From aerosol measurement perspective, the important measurements cover the full aerosol size range, aerosol mass, optical properties, volatility, cloud-condensation nuclei activity, aerosol mass spectrometry, organic and elemental carbon measurements. Additional shorter-term measurement campaigns involving additional instruments were conducted periodically, and included e.g. the V/H-TDMA and humidified nephelometer instruments described above.

The local anthropogenic air pollution sources are limited in magnitude and breadth. The forestry station houses research personnel around the year, with intensive research and education activities taking place there in the summertime. Thus, intermittent cooking, biomass burning, and vehicle emissions are to be expected. The nearby county of Juupajoki is sparsely populated (5-10 inhabitants / km^2; Paper V). The nearest village of Korkeakoski is located 5 km south-east from the station and notably includes air pollution emission sources in the form of two sawmills, a pellet factory, and a district heating plant. The city of Orivesi, with some light industry and a population of 9 500

is located 20 kilometres to the south of Hyytiälä, and city of Tampere (pop. 213 000), 50 km south-west. The nearby regions also include scattered, small-scale agriculture in form of livestock and grain farms, as well as swamps and small lakes.

The dominant source of air pollution at SMEAR II is long range atmospheric transport of pollutants from southern, more densely populated and industrialised parts of Finland, as well as the highly industrial St. Petersburg region (Riuttanen et al., 2013; Kulmala et al., 2005). Transported air pollution has been shown to also arrive from continental Europe (Sogacheva et al., 2005; Niemi et al., 2009). The various anthropogenic pollution sources were identified in Paper V of this thesis, and a comprehensive model for aerosol types constructed in Paper VI.

4 Results and discussion

In this chapter, I will shortly summarise those main results from the research Papers that are relevant to this thesis, i.e. regarding the use of chemical speciation in explaining physicochemical properties (Section 4.1), and chemometric findings (4.2).

4.1 Aerosol chemical speciation and physicochemical properties

In Papers I through IV, measured aerosol physicochemical properties were interpreted in the light of chemical information available from the AMS and other instruments. The experimental results were additionally compared to theoretical predictions, whenever those were available. In the following, I present the main findings, arranged thematically to volatility (Section 4.1.1), hygroscopicity (4.1.2), and optical properties (4.1.3).

4.1.1 Volatility

In Paper I, Volatility of submicron aerosol was studied using a VDMPS, and the results on mass fraction remaining (MFR) were examined for different seasons at the boreal forest site of SMEAR II. On average, 19 % of aerosol at SMEAR II was not volatilised at the temperature of 280 °C, prompting the question of what the composition of this remaining mass was. Aethalometer-measured black carbon was found to explains a maximum of 55 % (winter) to 85 % (summer) of observed MFR, and for the submicron aerosols at SMEAR II there exist no obvious candidates, as sea salt contributions are low at this inland site (Saarikoski et al., 2005). Oxidised organics have also been observed to contribute to the non-volatile fraction (Huffman et al., 2009b), and to form organic polymers (Kalberer et al., 2004). In 2010 Virtanen and co-workers showed that biogenic SOA can adopt an amorphous solid, most likely glass-like structure.

Looking at AMS measured ion species, sulphates (SO4) and ammonia (NH_4) correlate markedly with high aerosol mass remaining after heating, at lower temperatures (below 200 °C), but instead show mild anti-correlation at higher temperatures (Figure 1). We interpreted this behaviour to indicate presence of the inorganic salt $(NH_4)_2SO_4$. For

total nitrate (NO_3), the temperature response was less clear, but statistically significant ($p < 0.05$) correlations were found when nitrates were split into organic and inorganic components using an ion balance scheme. Inorganic nitrate behaves as expected for NH_4NO_3, and seems to evaporate below 200 °C. However, the organic nitrate instead exhibits the opposite behaviour, and actually features the highest correlation with MFR at temperatures at 280 °C. This is one of the most interesting findings of this study yet further conclusions could not be drawn based on the data available. The organic species also seem to be included in the MFR at 280 °C.
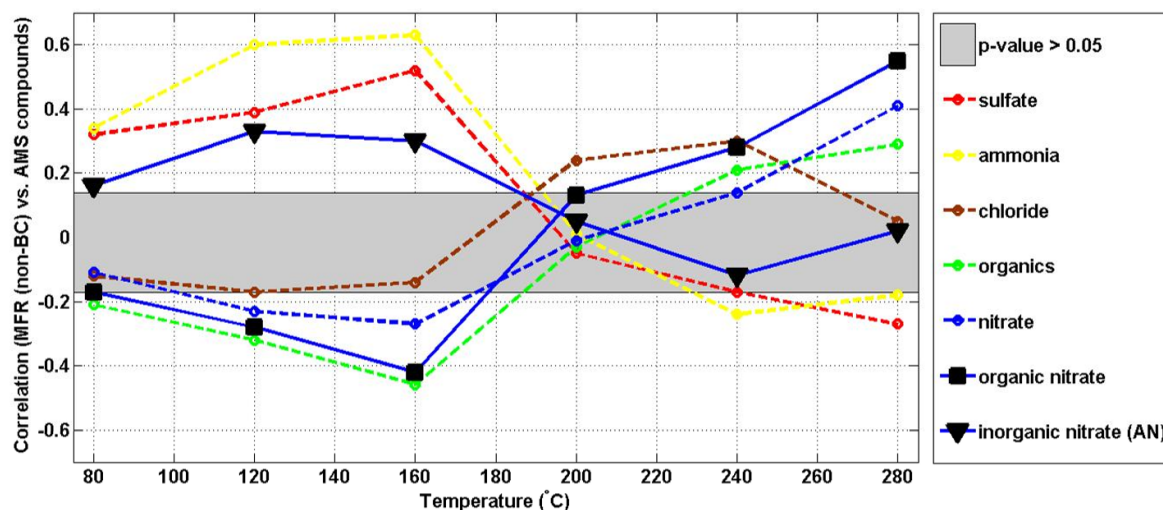
Figure 1: (Paper I). Non-BC mass fraction remaining (MFR) correlations with mass fractions of AMS ion species and ion balance derived nitrate species (inorganic and organic NO_3). Below 200ºC the MFR seems to be explained by inorganic salts, whereas the MFR at higher temperatures implies connection to organics and organic nitrate.

We conclude the origins for the non-BC MFR may indeed be largely biogenic - a notion also supported by the wind direction analysis implying higher observed MFR for air masses arriving from the clean north-west section of SMEAR, where no significant anthropogenic sources exist (Sogacheva et al., 2005). We additionally put forward a hypothesis that the non-BC MFR might be explained by low/non-volatile organics in amorphous, glass-like structures, as described by of Virtanen et al. (2010). Other potential candidates include aminium salts (Smith et al., 2010) or organic polymers (Kalberer et al., 2004). However, the actual volatilities of such compounds and structures remain unknown. Our findings additionally suggest organic nitrate may be connected to the chemical processes resulting in non-volatile (at 280 °C) aerosol

33

species. This notion is also supported by the later study by F. Drewnick et al., (2015), which reports that the fragmentation patterns of organic nitrates inside the AMS may be indicative of very slow thermal decomposition, even at very high temperatures (of around 600 ºC).

In Paper IV, we focused on volatility properties of organic aerosols. A V-TDMA was deployed at the SMEAR II station to measure aerosol volatility distributions in the spring of 2014. An aerosol volatility model (Riipinen et al., 2010; Hong et al., 2017) was then applied to explain the observed aerosol volatility distributions. The model describes organic aerosols using three volatility classes (semi-volatile, low-volatile, and extremely low-volatile) and calculates an estimate for their relative contributions that would best explain the experimentally measured data. Incidentally, there is also evidence that factorisation of AMS organic mass spectra also carries information on volatilities of the resulting organic aerosol subtypes (often termed low-volatile organic aerosol, LV-OOA, and semi-volatile organic aerosol, SV-OOA; Aiken et al., 2008). This has resulted in the current widespread convention of naming the AMS-derived organic aerosol types according to their expected volatilities, even when experimental volatility data is not available. While the connection of chemical composition and volatility certainly exists, it is understood that the operative definitions for volatility classes do not correspond exactly (Paciga et al., 2016; Zhang et al., 2011). In Paper IV, we examined if these differences in classifications could be reconciled by modifying the parameters in the volatility model of Riipinen et al. (2011).

The Riipinen et al. (2011) volatility model is extremely sensitive to values given for enthalpies of vapourisation (ΔH_{vap}), yet little data for ΔH_{vap} of atmospheric organic aerosols exists. We thus tested the agreement between the volatility model and the AMS factorisation, using a variety of enthalpy values for the volatility classes in the model. With optimised enthalpy values the model predictions do correlate with AMS factorisation results (Figure 2). However, the agreement is moderate at best, implying that either (1) the operative definition for the volatility classes (derived from AMS and V-TDMA) may be inherently different, or (2) the volatility model or the AMS speciation (PMF factorisation combined with an ion balance scheme for inorganic classification; Paper IV, Paper VI) may need refinement.
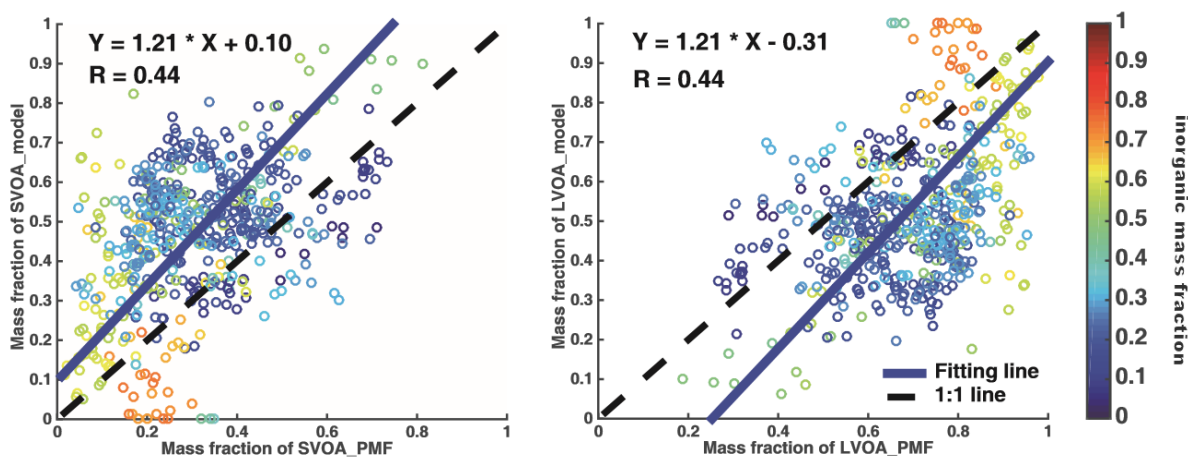
Figure 2: (Paper IV). Modeled mass fraction (V-TDMA) versus PMF-derived mass fraction (AMS low-volatile and semi-volatile oxidised organic aerosols; LV-OOA and SV-OOA). For the V-TDMA derived model, LV-OOA class contains also the extremely low-volatile oxidised organic aerosol type, which was not separately resolvable by the AMS factorisation.

Overall, volatility is a difficult quantity to (1) measure, due to technical reasons (e.g. aerosols not reaching thermal equilibrium inside the thermodenuder; Riipinen et al., 2010), possible thermal decomposition or charring), and (2) model, due to scarcity of knowledge and data on e.g. enthalpies of aerosol components. The technical and modelling developments in volatility measurements may in time resolve these issues. In the meanwhile, improving AMS chemical speciation via new chemometric approaches may also result in improved agreement.

4.1.2 Hygroscopicity

The connection between aerosol hygroscopic properties and chemical speciation was examined in Paper III. While there exists a more established theoretical framework and parameterisations for hygroscopicity than for aerosol volatility, the influence of chemical composition on hygroscopicity is not easy to capture. Multivariate linear parameterisations are traditionally used to model hygroscopic growth factor (HGF) for aerosol consisting of multiple chemical components.

Hygroscopicities for inorganic aerosols (e.g. pure $(NH_4)_2SO_4$) are fairly well known from laboratory data, but less so for atmospheric organic aerosols. Aged, highly oxidised organics are known to be more hygroscopic than fresh organics, and experimental parameterisations have been proposed to describe the HGF dependence on organic aerosol oxidation level, (represented by aerosol oxygen-to-carbon ratio; O:C; Duplissy et al., 2011; Pajunoja et al., 2015; Massoli et al., 2010).

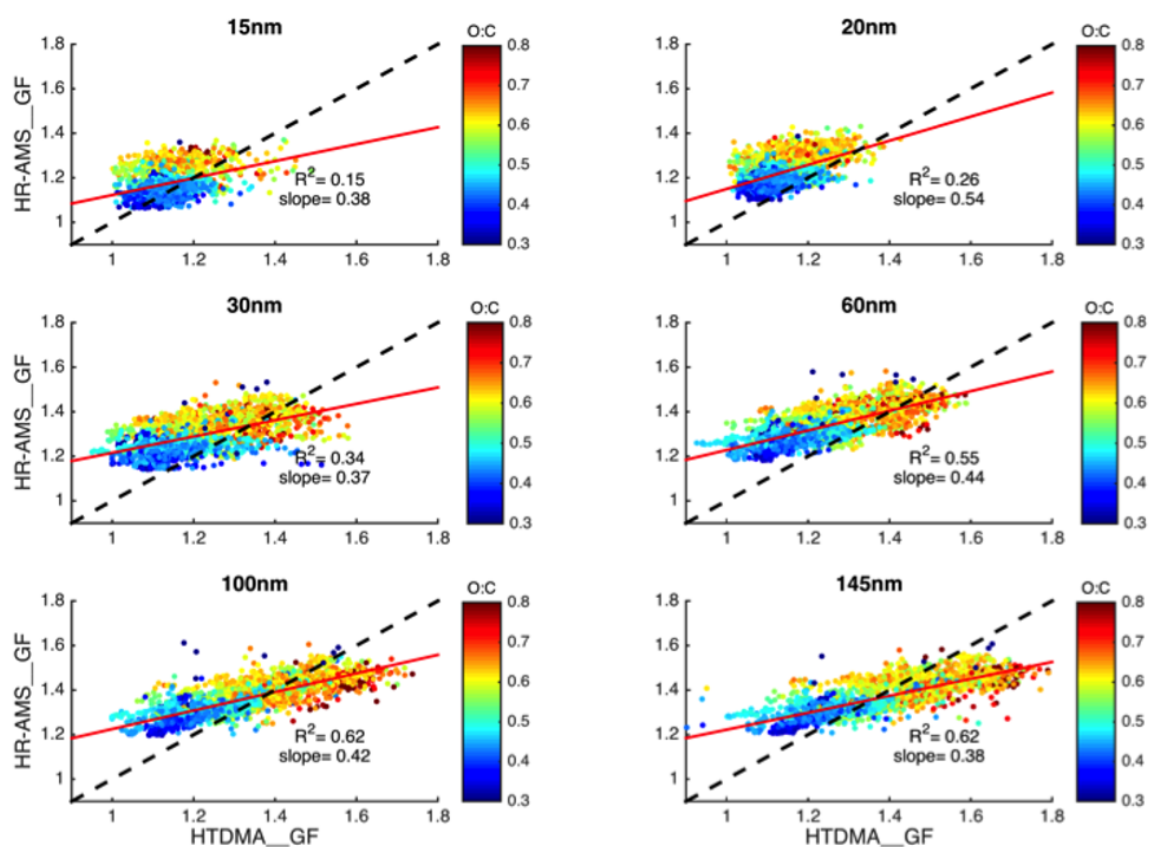Figure 3: (Paper III). AMS-derived hygroscopic growth factor (HGF) versus the directly measured, experimental value from H-TDMA, for aerosol particle sizes of 15 to 145 nm. Organics were considered as single species and inorganics were speciated using an ion-balance scheme (see e.g. Paper VI). Oxygen-to-carbon-ratio estimate (Aiken et al., 2008), representing aerosol oxidation level, is plotted in colour.

In Paper III, we compared H-TDMA-measured HGF to a corresponding estimate derived from AMS data. For inorganic speciation, we used an ion balance scheme (see Paper VI) to estimate the mass fractions of inorganic salts. In this study, organics were considered as a single separate class. A simple multilinear model manages to explain measured hygroscopicity quite well (Figure 3), especially for the larger particle sizes. As the AMS data is not size segregated, composition of larger particles dominates the AMS results due to the relatively larger mass and volume, so the AMS chemical speciation mostly represents the composition of large particles ($>$ 100 nm). The linear model squared correlation with measured HGF is around 0.62, which can be interpreted as the model capturing 62 % of variation in measured HGF. While there is still a lot of room for improvement, we find this to be a promising result, considering the approximations used.

From Figure 3 (data coloured by O:C) it is apparent that the treatment of organic aerosols as a single species explains some of the variation not captured by the parameterisation. When organic fraction and O:C was included in the experimental parameterisation (Figure 4), the correspondence between the model and experimental data improved markedly, reflected by the slightly higher correlation values and the slope values of linear fits closer to unity.

The remaining, unexplained variability may be due to (1) non-linear dependences not captured by the multivariate linear model (2) non-optimal AMS speciation (ion balance scheme uncertainty; treating organics as single species), (3) effects of refractory aerosol species not measured by the AMS, such as black carbon, or (4) measurement technical uncertainties. The nonlinearity of the relationship can arise from many sources, e.g. it may be due to interactions between the organic and inorganic species. Again improving the AMS-based speciation may be part of the solution, as well as application of non-linear chemometric models (for example nonlinear regression or a neural network approach).
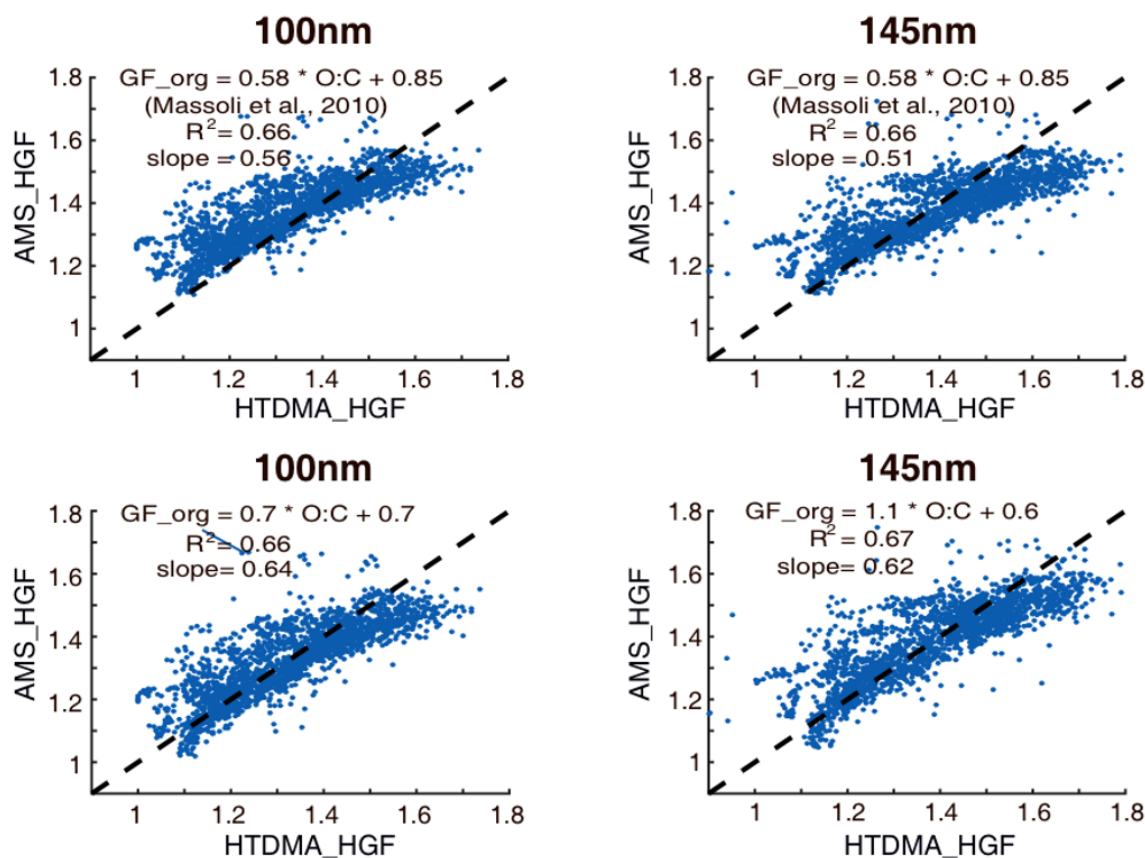
Figure 4: (Paper III) AMS-derived hygroscopic growth factor (HGF) versus the directly measured, experimental value from H-TDMA, with the influence of organic aerosol O:C included (upper panels: Massoli et al., 2010, parameterisation; lower panels: parameterisation from Paper III).

4.1.3 Optical properties

Aerosol light scattering and its enhancement in humid environments, $f(RH)$ were studied in Paper II. Light scattering properties for dry particles depend strongly on particle size and the effect of chemical composition is relatively minor and well understood. However, for atmospherically relevant humidity levels, the situation changes, necessitating consideration of particle chemical composition. As discussed in the hygroscopicity section (2.1.3), particle water uptake properties do significantly depend on composition. Besides growing the particles in size, which itself affects scattering, a layer of condensed water on particle surface or a change of particle phase towards a liquid droplet may change the particle optical properties dramatically.
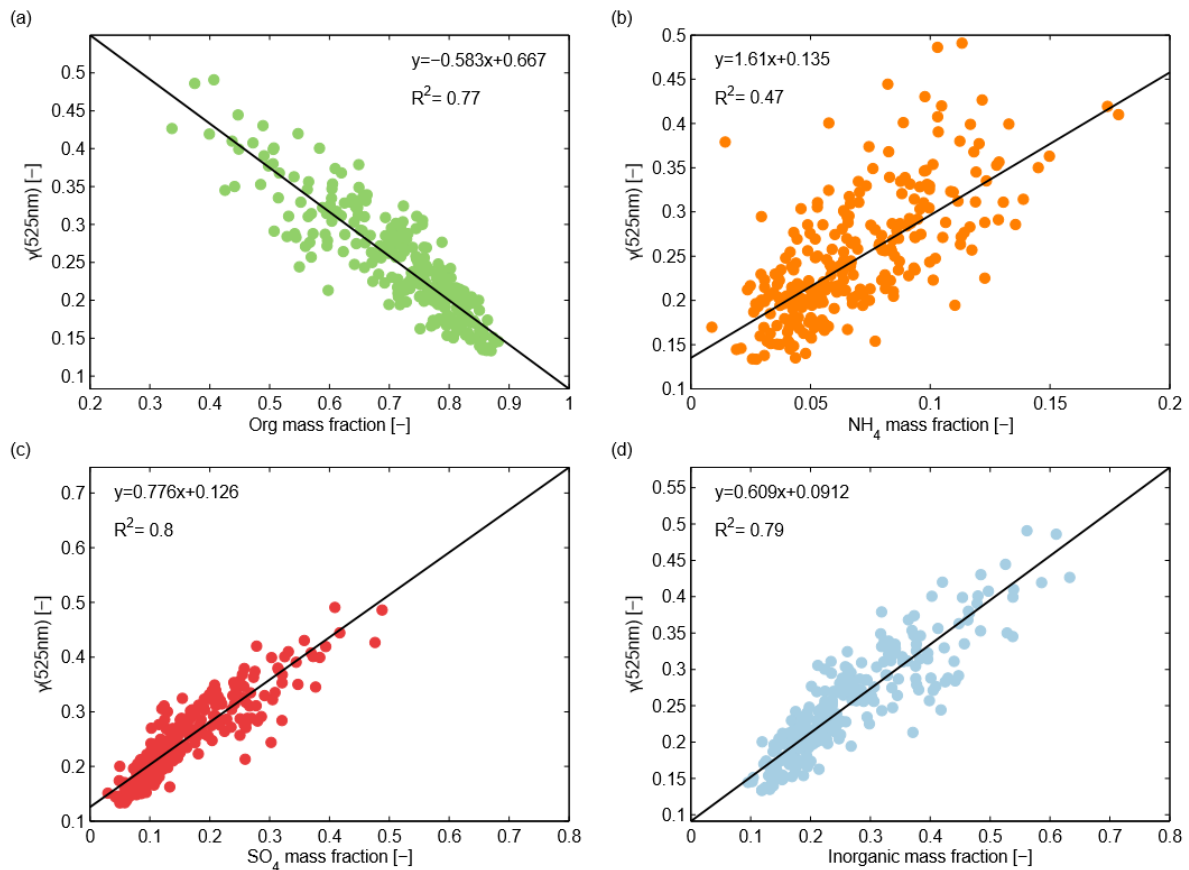
Figure 5: (Paper II unpublished figures, courtesy of Paul Zieger). Magnitude of the hygroscopic scattering enhancement, described by the γ parameter (Clarke et al., 2002; Zieger et al., 2015) as a function of AMS-measured ion species fraction; organics (panel a); NH_4 (b), SO_4 (c) and sum of inorganics (d).

In the study described in Paper II, nephelometer-measured scattering coefficients were measured for dry and humidified particles at the SMEAR II site. The humidity enhancement in scattering efficiency, was examined as a function of AMS-derived particle composition. The magnitude of the scattering enhancement effect was estimated from nephelometer data using an empirical parameterisation (Clarke et al., 2002; Zieger et al., 2015). The parameterisation provides an RH-independent metric (γ) for estimating the magnitude of the scattering enhancement, with $\gamma = 0$ signifying no enhancement and higher values corresponding to higher magnitude of the effect. In Figure 5, γ is shown as a function of mass fraction of various AMS ion species. From the results it is evident that the magnitude of scattering enhancement strongly and clearly depends on particle chemical composition. For fully organic particles the scattering enhance-

ment is close to zero, which indicates that predominantly organic particles are nearly hydrophobic, and accumulate very little condensed water even in high-humidity conditions. Contrarily, the scattering enhancement is proportional to fraction of AMS ion species associated with inorganics (SO_4, NO_3, and NH_4). This is to be understood in the framework of inorganic salt formation ((NH_4)$_4SO_2$ and NH_4NO_3) – the common inorganic salts in aerosol are typically hydrophilic, so particles with high fraction of these species eagerly take up water from the ambient air. The extra water in these particles results in enhanced light scattering.
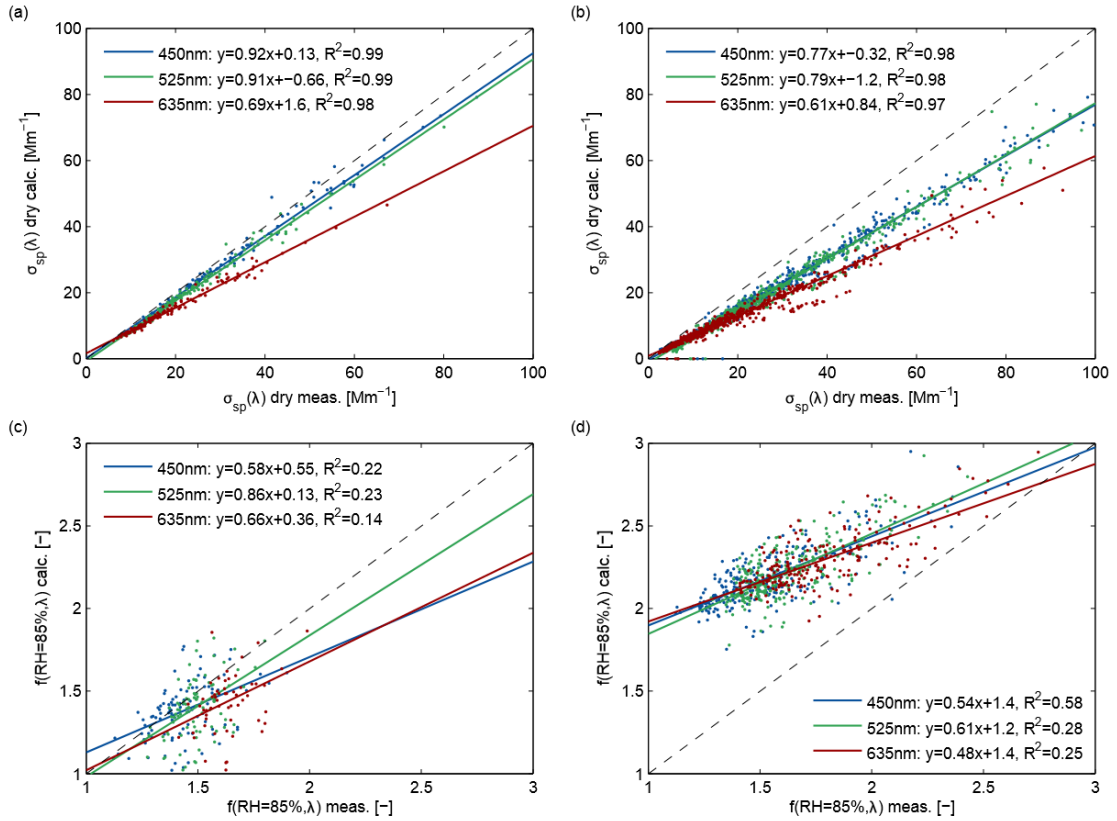
Figure 6: (Paper II unpublished figures, courtesy of Paul Zieger). Panels (a) and (b): the particle light scattering coefficient σ_{sp}, calculated from chemical composition (y-axis) versus measured (x-axis) values. In panels (c) and (d): the calculated (y-axis) versus measured $f(RH)$ (x-axis) is described similarly. In panels (a) and (c), refractive index of ammonium sulphate was used, whereas in panels (b) and (d) the refractive index was estimated based on aerosol chemical composition.

A test was also run for estimating if using composition-dependent refractive index for aerosol (instead of refractive index for ammonium sulphate) would improve estimation of σ_{sp} and $f(RH)$ from aerosol size distributions (Figure 6). From the results it seems that using the composition-dependent refractive index does improve the correlation coefficient between measured and predicted $f(RH)$, but decreases the agreement given by the slope of a linear fit. This was taken to hint that the current method for refractive index estimate may be inadequate, especially regarding the influence of organics.

4.2   Improved speciations via chemometric methods

Papers V and VI deal with the prospect of improving AMS chemical speciation using combined chemometric methods. In summary, the problematic issues with the traditional AMS speciations relate to (i) treating organics as one bulk species and (ii) the encoded separation of inrganic and organic sub-species. Incorporating various O:C related parameterisations for oxidation level (Aiken et al., 2007; Aiken et al., 2008) somewhat improves the handling of biogenic, oxidised organics, but it does not account for minor organic species. As the inorganic-organic division is "hard-coded" into most speciation schemes, inorganic-organic mixtures of components are not allowed in the classifications. This poses problems for e.g. organic nitrates. We know a large fraction of nitrate in Hyytiälä is organic, but this is not accounted for in many models where nitrate is assumed to be inorganic by default. The simplified ion balance models used for inorganic speciation are inadequate to describe organic nitrates or sulphates.

To answer to the problem (i), factor analysis is often presented as a solution. Indeed, it has been applied to successfully resolve organic sub-species. However, PMF models suffer from two major problems common to factorisation methods: (a) selection of number of factors is ambiguous, and the mathematical diagnostics for this problem are limited. (b) Rotational ambiguity of factor analytical models results in problems especially for the extraction of minor components, which are not extracted reliably. Especially the components with smaller loadings (typically anthropogenic aerosol types, (Crippa et al., 2014)) are affected. This has prompted for the use of constraints in PMF. However, the problem then becomes (c) the proper selection of these constraints. Paper V touches the aforementioned organic speciation problem (i), and in Paper VI we address both of the problems (i) and (ii). Papers V and VI also attempt to remedy the factor analytical issues surrounding problems (a), (b), and (c) for AMS data analysis.

4.2.1 Classifying anthropogenic aerosol pollution plumes

While chemometric factorisation methods (specifically PMF) have proven invaluable for de-convolving organic aerosol sub-species, analyst-selected constraints are often necessary for resolving subspecies with low total mass contribution (Canonaco et al., 2013; Crippa et al., 2014; Ulbrich et al., 2009). The selection of these constraints has traditionally been up to the analyst. Lacking easy-to-use diagnostics and sensitivity tests for different selections, the selection process is thus prone to subjective bias by the analyst. As any additional constraints set for the factorisation algorithm strongly affect the outcome of factorisation, this problem needs addressing.

Paper V shows how reliable constraints and reference spectra for anthropogenic aerosols can be derived from data. As the anthropogenic aerosols often manifest in the form of distinct pollution plumes (at non-urban sites), directly observable as increased mass loadings, I factorised 81 such events with PMF. Crucially, the factorisation of short-lived pollution events circumvents many of the problems related to rotational ambiguity, as minimisation of factor correlation can be used as a criterion for solution quality. An example of such a factorisation is described in Figure 7.

Figure 7: (Paper V). Example of a factorisation solution for a pollution event, with a biomass burning pollution plume separated from background aerosol. PMF was tasked to separate the data into two components, the mass spectra and loadings of which are shown in different colour for the background (green) and pollution (turquoise) factors.

Analysis of pollution events from three measurement campaigns produced a total of 81 pollution spectra. While many of the aerosol types could be identified based on previous literature, many were too ambiguous. Also, a manual classification would nullify the effort of avoiding subjective bias. Thus, the pollution spectra were classified algorithmically. A clustering algorithm was optimised (with respect to spectral similarity metric and data weighting), to maximise classification solution quality (indicated by silhouette value; Figure 8).
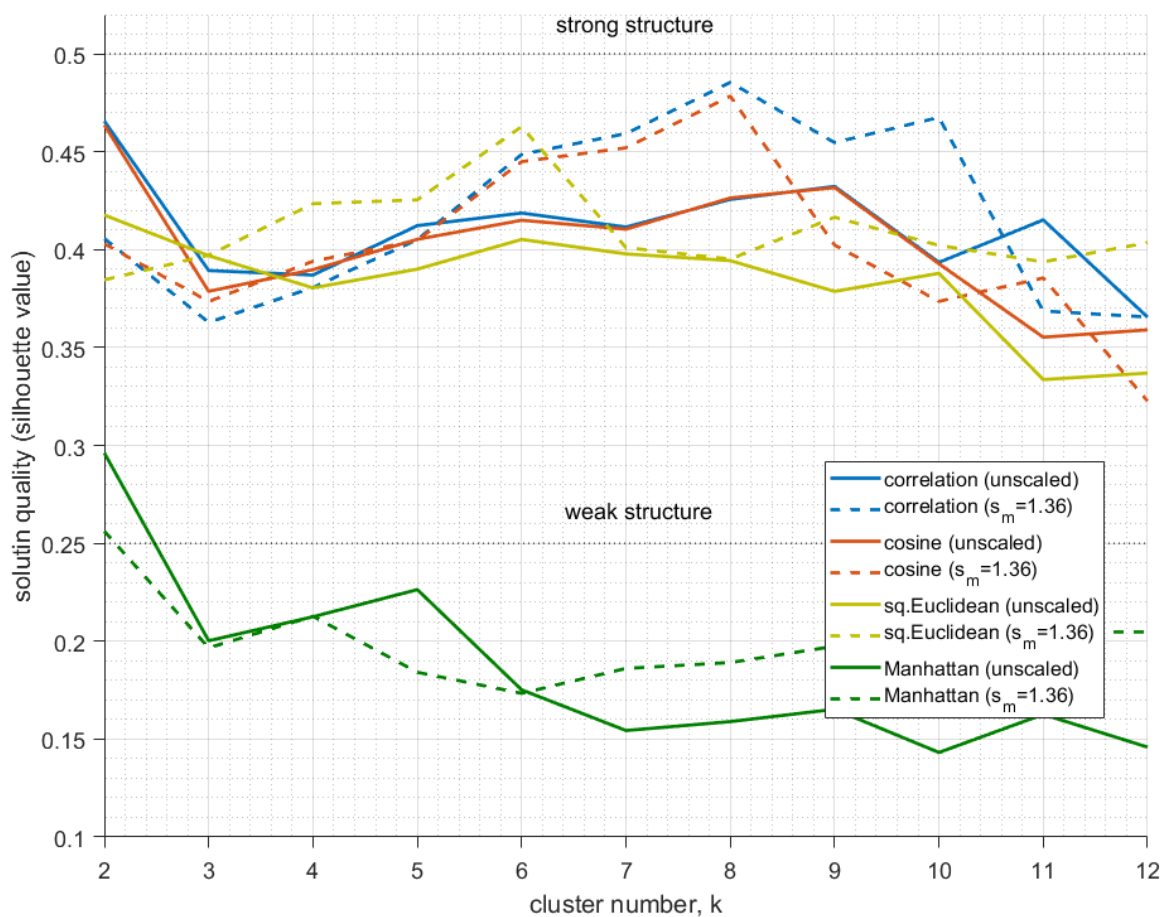
Figure 8: (Paper V). Optimisation of spectra classification procedure with respect to spectral similarity metric (coloured lines) and data weighting (by a mass scaling factor s_m). Solution quality is described by the silhouette metric (Rousseeuw, 1987).

The resulting classification algorithm was shown to reproduce the aerosol classes of anthropogenic aerosols known from previous studies, as well as to identify minor aerosol species, which would be neglected in a conventional factorisation analysis. In this study, the anthropogenic organic classes could be divided into classes such as: traffic, cooking, biomass burning, and pollution from a nearby sawmill. The outlier groups notably included amine-dominated aerosols from an unknown source. Although adequately visualising a n-dimensional ($n > 3$) clustering is difficult, it is possible to project the clustering solution to a plane (defined by two axes). For this, two important axes which also have meaning in an aerosol chemical interpretation were identified. A projection of the resulting classification to these axes - one describing aerosol age (x-axis) and one connected to aerosol origin (y-axis) - is shown in Figure 9.
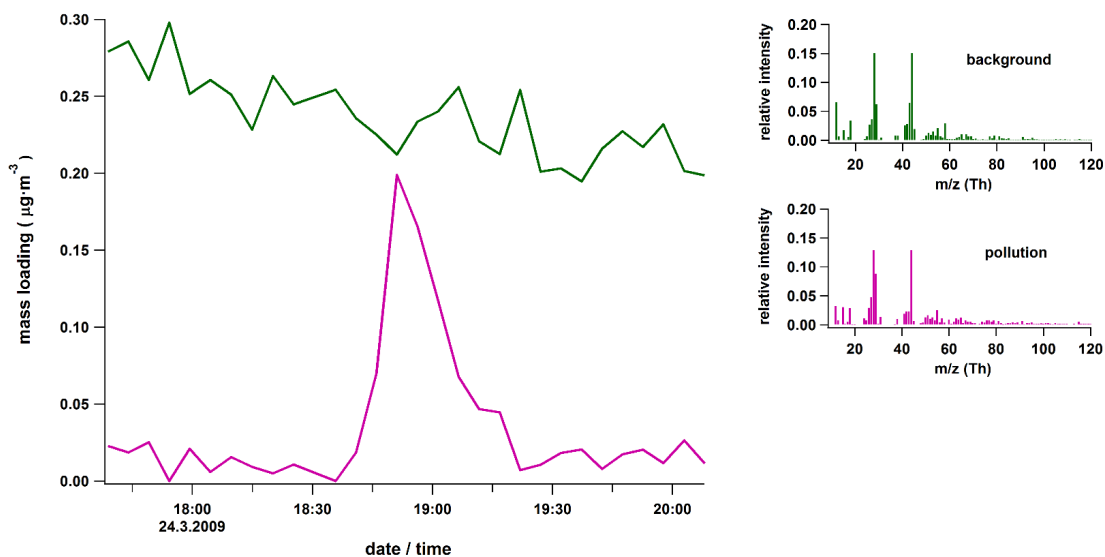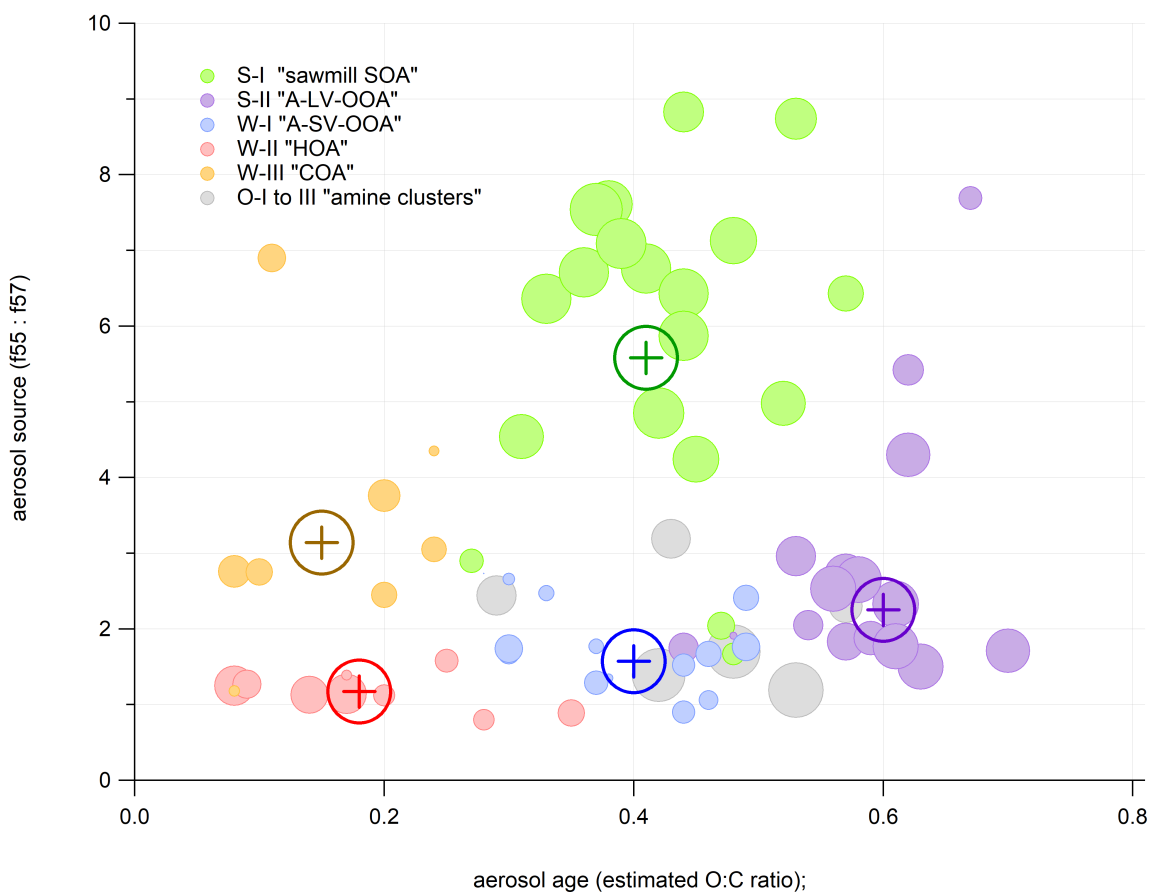

Figure 9: (Paper V). Example of a factorisation solution for a pollution event, with a biomass burning pollution plume separated from background aerosol. PMF was tasked to separate the data into two components, the mass spectra and loadings of which are shown in different colour for the background (green) and pollution (turquoise) factors.

4.2.2 Improved chemometric chemical speciation

Finally, in Paper VI the traditional division into organic and inorganic components is re-evaluated. This fundamental classification has traditionally served as the starting point, and sub-classes, such as calculation of inorganic salts from ion balance models, or organic aerosol factorisation, have been done within the confines of either the inorganic or the organic dichotomous class. However, it is well known that e.g. aerosol organonitrates or organosulphates cannot easily be described within the traditional ion species-based, organic/inorganic classification.

In this study, we combined the traditionally separate organic and inorganic mass spectral results and subsequently factorised a total of eight data sets from SMEAR II. To avoid the pitfalls of manual factor number selection, I used a data driven approach with affixed number of PMF factors, and (similarly to Paper V) used clustering to find the most likely number of aerosol types. The method resulted in a statistically robust classification of aerosol into seven species (Figure 10).
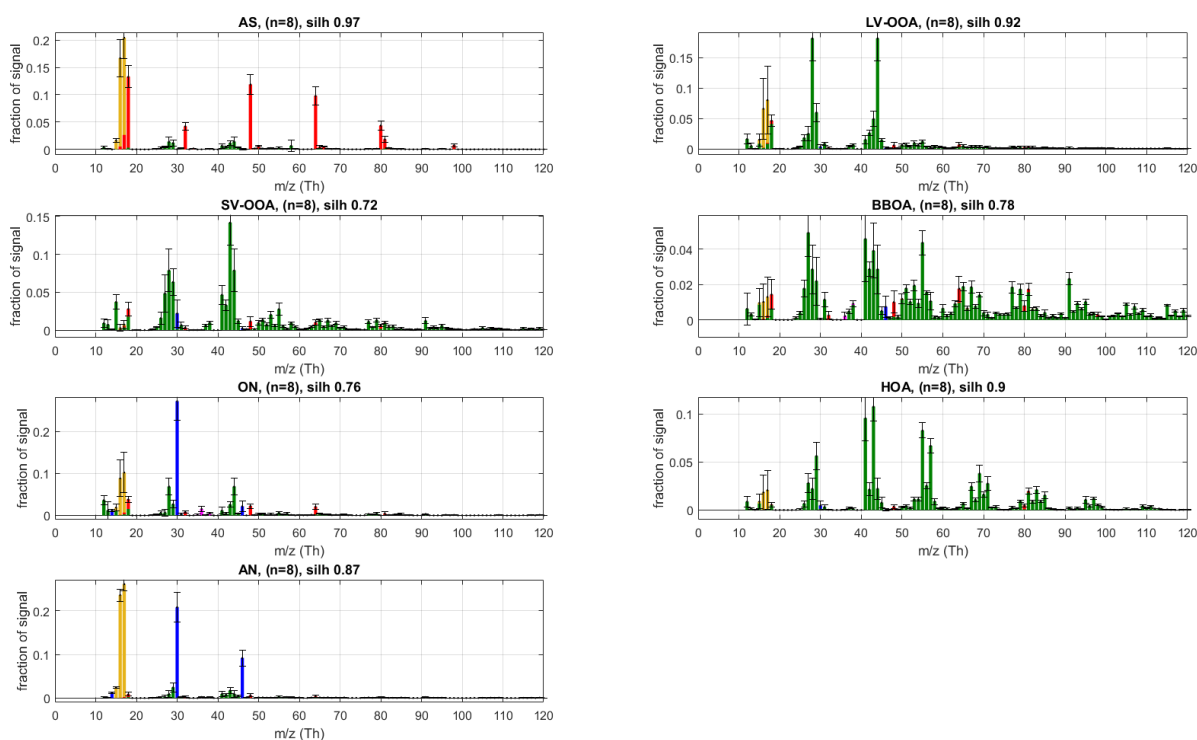
Figure 10: (Paper VI) Results for the data-driven, comprehensive aerosol classification mass spectra for the main aerosol species at SMEAR II.

The resolved aerosol classes were ammonium sulphate (AS), ammonium nitrate (AN), organic nitrate (ON), SV-OOA, LV-OOA, biomass burning organic aerosol (BBOA) and hydrocarbon-like organic aerosol (HOA). This classification offers an alternative perspective for the traditional AMS aerosol species classification (Figure 11). The speciation is solidly statistics based, and contains much less assumptions and is more resistant to analyst subjective bias. This classification can be used as a future basis for predicting aerosol physicochemical properties.
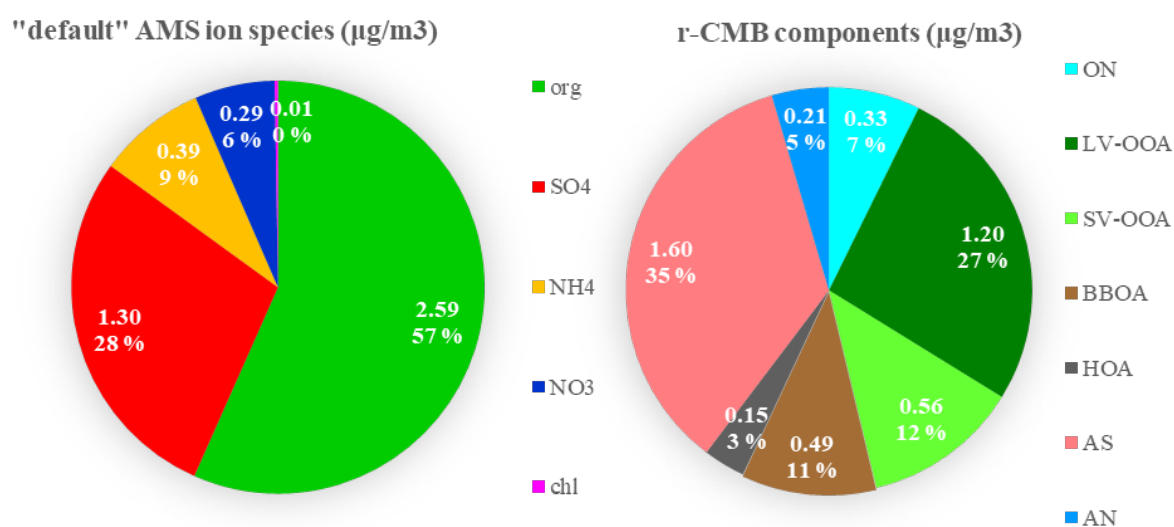
Figure 11: (Paper VI). Traditional AMS speciation, based on division into organic and ion species (left), and the respective statistics-based 7-component speciation derived using the combined methodology of Paper VI.

5 Review of papers and the author's contribution

Paper I investigates atmospheric aerosol volatility characteristics when heated. Fraction of aerosol mass remaining after heating is examined in light of the aerosol chemical data on black carbon and the aerosol mass spectrometer (AMS) chemical classification results. I measured, processed and analysed the AMS data, participated in the data analysis, and wrote parts of the Methods section.

Paper II studies aerosol light scattering properties, and connects humidity-induced scattering enhancement to aerosol chemical composition. I measured, processed, and analysed the AMS data, took part in the wet-nephelometer measurements at SMEAR II and the chemical composition related analysis of the results, and participated in commenting and editing of the manuscript.

Paper III examines aerosol hygroscopic growth, and relates it to particle and aerosol phase chemical composition. I measured, processed, and analysed the AMS data, participated in the data interpretation as well as commenting and editing of the manuscript.

Paper IV reports a comparison between an aerosol volatility model and AMS derived volatility characteristics. I measured, processed, and analysed the AMS data, participated in the data interpretation as well as commenting and editing of the manuscript.

Paper V describes a data-analytical method of deriving anthropogenic aerosol types from mass spectrometric measurements. I was responsible for most of the measurements, processing, and analysis of data, interpreting and visualising the results, and wrote the manuscript with comments from co-authors.

Paper VI reports a study on how to construct a data-driven, statistical aerosol chemical speciation model. I was responsible for most of the measurements, the processing, and analysis of data, interpreting and visualising the results, and wrote the manuscript with comments from co-authors.

6   Conclusions

As outlined in the introduction (Section 1.5), the main aims of this thesis were related to (1) connecting particle chemical composition to aerosol physicochemical properties, (2) applying chemometric and machine learning methods to aerosol mass spectral data, and (3) using them to develop improved chemical speciations for aerosol mass spectral data.

In conclusion, I find that aim (1) was partially achieved: the existing chemical speciations used in Papers I through IV can mostly explain aerosol hygroscopicity properties. For volatility and optical properties, the predictions are not as robust, but they are still clearly useful approximations. The work on aim (2) was certainly successful, as the results of especially Papers V and VI acutely demonstrate the value of chemometric and machine learning tools for aerosol mass spectrometric data analysis. As for aim (3), the study described in Paper VI yielded a promising chemometrics-based speciation that successfully combines the organic and inorganic chemical species. While it seems potentially advantageous over the more traditional schemes, its usefulness will be ultimately be decided by its future applications.

Below, the main results and findings, arranged thematically to new aerosol physical and chemical information (Section 6.1) and chemometric and data analytical insights (Section 6.2), are summarised.

6.1   Linking chemical composition and aerosol physicochemical properties

Different levels of aerosol chemical speciation were evaluated in their usability to explain aerosol physicochemical properties, such as volatility, hygroscopicity, and light scattering observed at the SMEAR II research site.

In Paper I, aerosol volatility in temperatures between 80 and 280 °C was studied. Mass fraction remaining (MFR) after heating (with BC subtracted) was correlated with AMS-derived chemical species. Ammonium (NH_4) and sulphate (SO_4) ion species correlated with MFR temperatures below 200 °C, corresponding to ammonium sulphate vapourisation temperature around 160 °C. Organics correlated with MFR above 200 °C, potentially indicating oligomerisation or amorphous solid phases of organics with

high vapourisation temperatures. Bulk nitrates (NO_3) correlated only marginally, but when an ion-balance scheme was applied to classify nitrate to organic nitrate and NH_4NO_3, the correlations sharpened: NH_4NO_3 behaves more similarly to $(NH_4)_2SO_4$ while NO_3 presumed to be organic correlates markedly with MFR at high temperatures (> 200 °C). The suggestion of organic NO_3 being low-volatile is supported by later observations on their slow thermal decomposition leading to increased fragmentation in the AMS (Drewnick et al., 2015).

In Paper IV, it was found that the AMS-derived speciations suggestive of organic aerosol volatility classes (SV-OOA and LV-OOA) likely do not exactly match the volatility classifications used in V-TDMA studies. Reconciling discrepancies between the model parameterisations is a topic for further study.

Hygroscopicity, on the other hand, was found to be better predicted by AMS-derived chemical speciation, where both organic factorisation and ion balance models were used, resulting in a good match between AMS-derived hygroscopic growth factor (HGF) and that derived from VH-TDMA measurements (Paper III). Specifically, oxygen-to-carbon ratio-dependent classification seems to improve the correspondence between the instruments.

In Paper II, optical properties were examined. Directly measured aerosol (dry) particle scattering coefficient σ_{sp} was well predicted by an estimation derived from DMPS and AMS. However, the hygroscopic scattering enhancement $f(RH)$, was poorly predicted by the existing models, due to difficulties of the linear model in handling the high fraction of organics often observed at the rural measurement site. It seems evident that the hygroscopicity predicting model and/or AMS chemical speciation need to be improved.

After deriving a comprehensive AMS chemical speciation both for organic and inorganic compounds (Paper VI), the statistics-derived speciation was compared with the traditional ion-species based classification. The main outlook of traditional versus statistical speciation is similar, but differences also exist. We find the statistical speciation to be more flexible and less prone to analyst subjective bias.

6.2 Summary of chemometric results

As summarised above, theory-derived, traditional speciations for aerosols can to be prove helpful, but current models are often insufficient for making accurate predictions – whether this is due to bad speciation, insufficient models or measurement uncertainty, is largely unknown. However, in Papers I to IV, there are ample examples where more detailed speciations yield better correspondence with experiments, lending support to the notion that improving chemical speciation schemes is indeed important and serve an important mediating function, in explaining the role of chemical composition on aerosol physical and chemical properties.

In Papers V and VI, new, chemometrics-oriented approaches to chemical speciation were examined. Incorporating typical machine-learning tools for feature extraction and unsupervised classification proved fruitful for aerosol chemical data (Paper VI), but require some initial optimisation to capture the mass spectral features that carry relevant chemical information. The statistical diagnostics associated with chemometric methods increase the objectivity of data analysis and help an analyst in making better-informed decisions and interpretations.

In Papers V and VI, I equally show how combining several chemometric methods can overcome some of the weaknesses of individual methods. For example, hard classification requires discrete or de-convolved samples, an assumption which is rarely fulfilled for mass spectra of atmospheric aerosol particles. This hinders straightforward classification of mass spectral samples and use of typical classification tools such as clustering in aerosol chemical speciation. The prerequisite can, however, be fulfilled by prior factorisation of mass spectral data. The above mentioned combination also helps in mitigating a typical factor analytical problem of rotational ambiguity, which results in ambiguity in the evaluation of factorisation solutions. Posteriori classification of extracted factors reveals which of the data structures are likely stable and contain relevant chemical or physical information, and which outputs are e.g. outliers or data processing artefacts.

Due to the complexity and large size of data produced in aerosol mass spectrometry, machine-learning methods may be the way to go also in predictive modelling, instead of the current, theory-derived multilinear models. The wide variety of chemometric and machine learning tools available should be considered when aiming to improve or automate data analysis capabilities in aerosol chemical measurements.

References

Aalto, P., Hameri, K., Becker, E., Weber, R., Salm, J., Makela, J. M., Hoell, C., O'Dowd, C. D., Karlsson, H., Hansson, H. C., Vakeva, M., Koponen, I. K., Buzorius, G., and Kulmala, M.: Physical characterization of aerosol particles during nucleation events, Tellus Series B-Chemical and Physical Meteorology, 53, 344-358, DOI 10.1034/j.1600-0889.2001.530403.x, 2001.

Aiken, A. C., DeCarlo, P. F., and Jimenez, J. L.: Elemental analysis of organic species with electron ionization high-resolution mass spectrometry, Anal Chem, 79, 8350-8358, 10.1021/ac071150w, 2007.

Aiken, A. C., Decarlo, P. F., Kroll, J. H., Worsnop, D. R., Huffman, J. A., Docherty, K. S., Ulbrich, I. M., Mohr, C., Kimmel, J. R., Sueper, D., Sun, Y., Zhang, Q., Trimborn, A., Northway, M., Ziemann, P. J., Canagaratna, M. R., Onasch, T. B., Alfarra, M. R., Prevot, A. S., Dommen, J., Duplissy, J., Metzger, A., Baltensperger, U., and Jimenez, J. L.: O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high-resolution time-of-flight aerosol mass spectrometry, Environmental Science and Technology, 42, 4478-4485, 2008.

Aitken, J.: I.—On the number of dust particles in the atmosphere, Earth and Environmental Science Transactions of The Royal Society of Edinburgh, 35, 1-19, 1889.

Allan, J. D., Alfarra, M. R., Bower, K. N., Williams, P. I., Gallagher, M. W., Jimenez, J. L., McDonald, A. G., Nemitz, E., Canagaratna, M. R., and Jayne, J. T.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 2. Measurements of fine particulate chemical composition in two UK cities, Journal of Geophysical Research: Atmospheres, 108, 2003.

Allan, J. D., Delia, A. E., Coe, H., Bower, K. N., Alfarra, M. R., Jimenez, J. L., Middlebrook, A. M., Drewnick, F., Onasch, T. B., and Canagaratna, M. R.: A generalised method for the extraction of chemically resolved mass spectra from Aerodyne aerosol mass spectrometer data, Journal of Aerosol Science, 35, 909-922, 2004.

Anderson, T., Covert, D., Marshall, S., Laucks, M., Charlson, R., Waggoner, A., Ogren, J., Caldow, R., Holm, R., and Quant, F.: Performance characteristics of a high-sensitivity, three-wavelength, total scatter/backscatter nephelometer, Journal of Atmospheric and Oceanic Technology, 13, 967-986, 1996.

Arrhenius, S.: XXXI. On the influence of carbonic acid in the air upon the temperature of the ground, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 41, 237-276, 1896.

Aston, F. W.: LXXIV. A positive ray spectrograph, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 38, 707-714, 1919.

Atkinson, R. W., Mills, I. C., Walton, H. A., and Anderson, H. R.: Fine particle components and health—a systematic review and meta-analysis of epidemiological time series studies of daily mortality and hospital admissions, Journal of Exposure Science and Environmental Epidemiology, 25, 208, 2015.

Blashfield, R. K., and Aldenderfer, M. S.: The methods and problems of cluster analysis, in: Handbook of multivariate experimental psychology, Springer, 447-473, 1988.

Bleakney, W.: A new method of positive ray analysis and its application to the measurement of ionization potentials in mercury vapor, Physical Review, 34, 157, 1929.

Bond, T. C., Anderson, T. L., and Campbell, D.: Calibration and intercomparison of filter-based measurements of visible light absorption by aerosols, Aerosol Science & Technology, 30, 582-600, 1999.

Bond, T. C., Doherty, S. J., Fahey, D., Forster, P., Berntsen, T., DeAngelo, B., Flanner, M., Ghan, S., Kärcher, B., and Koch, D.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research: Atmospheres, 118, 5380-5552, 2013.

Buajarern, J., Mitchem, L., and Reid, J. P.: Characterizing the formation of organic layers on the surface of inorganic/aqueous aerosols by Raman spectroscopy, The Journal of Physical Chemistry A, 111, 11852-11859, 2007.

Buettel, R.G., and Brewer, A.W.: Instruments for the Measurement of Visual Range," Journal of Scientific Instruments, 26, p. 357, 1949.

Burkart, J., Steiner, G., Reischl, G., and Hitzenberger, R.: Long-term study of cloud condensation nuclei (CCN) activation of the atmospheric aerosol in Vienna, Atmospheric environment, 45, 5751-5759, 2011.

Burt, C.: The distribution and relations of educational abilities, P. S. King & Son, London, 1917.

Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L. R., Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the Aerodyne aerosol mass spectrometer, Mass Spectrometry Reviews, 26, 185-222, 10.1002/mas.20115, 2007.

Canonaco, F., Crippa, M., Slowik, J. G., Baltensperger, U., and Prevot, A. S. H.: SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data, Atmospheric Measurement Techniques, 6, 3649-3661, 10.5194/amt-6-3649-2013, 2013.

Cappa, C., and Jimenez, J.: Quantitative estimates of the volatility of ambient organic aerosol, Atmospheric Chemistry and Physics, 10, 5409-5424, 2010.

Carey, N.: Factors in the mental processes of schoolchildren 1: I. Visual and auditory imagery, British Journal of Psychology, 1904-1920, 7, 453-490, 1915.

Chang, D., Lelieveld, J., Tost, H., Steil, B., Pozzer, A., and Yoon, J. J. A. E.: Aerosol physicochemical effects on CCN activation simulated with the chemistry-climate model EMAC, Atmospheric Environment, 162, 127-140, 2017.

Clarke, A., Howell, S., Quinn, P., Bates, T., Ogren, J., Andrews, E., Jefferson, A., Massling, A., Mayol-Bracero, O., and Maring, H.: INDOEX aerosol: A comparison and summary of chemical, microphysical, and optical properties observed from land, ship, and aircraft, Journal of Geophysical Research: Atmospheres, 107, INX2 32-31-INX32 32-32, 2002.

Corrigan, A., Russell, L., Takahama, S., Äijälä, M., Ehn, M., Junninen, H., Rinne, J., Petäjä, T., Kulmala, M., and Vogel, A.: Biogenic and biomass burning organic aerosol in a boreal forest at Hyytiälä, Finland, during HUMPPA-COPEC 2010, Atmospheric Chemistry and Physics, 13, 12233-12256, 2013.

Crippa, M., Canonaco, F., Lanz, V. A., Äijala, M., Allan, J. D., Carbone, S., Capes, G., Ceburnis, D., Dall'Osto, M., Day, D. A., DeCarlo, P. F., Ehn, M., Eriksson, A., Freney, E., Hildebrandt Ruiz, L., Hillamo, R., Jimenez, J. L., Junninen, H., Kiendler-Scharr, A., Kortelainen, A. M., Kulmala, M., Laaksonen, A., Mensah, A., Mohr, C., Nemitz, E., O'Dowd, C., Ovadnevaite, J., Pandis, S. N., Petäjä, T., Poulain, L., Saarikoski, S., Sellegri, K., Swietlicki, E., Tiitta, P., Worsnop, D. R., Baltensperger, U., and Prevot, A. S. H.: Organic aerosol components derived from 25 AMS data sets across Europe using a consistent ME-2 based source apportionment approach, Atmospheric Chemistry and Physics, 14, 6159-6176, 10.5194/acp-14-6159-2014, 2014.

DeCarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K. S., Worsnop, D. R., and Jimenez, J. L.: Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer, Analytical Chemistry, 78, 8281-8289, 10.1021/ac061249n, 2006.

Després, V., Huffman, J. A., Burrows, S. M., Hoose, C., Safatov, A., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M., and Pöschl, U.: Primary biological aerosol particles in the atmosphere: a review, Tellus B: Chemical and Physical Meteorology, 64, 15598, 2012.

Diniz, W., and Canduri, F.: Bioinformatics: an overview and its applications, Genetics and Molecular Research, 16, 2017.

Donahue, N., Robinson, A., Stanier, C., and Pandis, S.: Coupled partitioning, dilution, and chemical aging of semivolatile organics, Environmental Science & Technology, 40, 2635-2643, 2006.

Donahue, N. M., Epstein, S. A., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics, Atmospheric Chemistry and Physics, 11, 3303-3318, 10.5194/acp-11-3303-2011, 2011.

Donahue, N. M., Kroll, J. H., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set - Part 2: Diagnostics of organic-aerosol evolution, Atmospheric Chemistry and Physics, 12, 615-634, 10.5194/acp-12-615-2012, 2012.

Drewnick, F., Hings, S. S., DeCarlo, P., Jayne, J. T., Gonin, M., Fuhrer, K., Weimer, S., Jimenez, J. L., Demerjian, K. L., Borrmann, S., and Worsnop, D. R.: A new time-of-flight aerosol mass spectrometer (TOF-AMS) - Instrument description and first field deployment, Aerosol Science and Technology, 39, 637-658, 10.1080/02786820500182040, 2005.

Drewnick, F., Diesch, J.-M., Faber, P., and Borrmann, S.: Aerosol mass spectrometry: particle–vaporizer interactions and their consequences for the measurements, Atmospheric Measurement Techniques, 8, 3811-3830, 2015.

Driver, H. E., and Kroeber, A. L.: Quantitative expression of cultural relationships, 4, University of California Press, 1932.

Duplissy, J., DeCarlo, P. F., Dommen, J., Alfarra, M. R., Metzger, A., Barmpadimos, I., Prevot, A. S. H., Weingartner, E., Tritscher, T., Gysel, M., Aiken, A. C., Jimenez, J. L., Canagaratna, M. R., Worsnop, D. R., Collins, D. R., Tomlinson, J., and Baltensperger, U.: Relating hygroscopicity and composition of organic aerosol particulate matter, Atmospheric Chemistry and Physics, 11, 1155-1165, 10.5194/acp-11-1155-2011, 2011.

Dusek, U., Frank, G., Hildebrandt, L., Curtius, J., Schneider, J., Walter, S., Chand, D., Drewnick, F., Hings, S., and Jung, D.: Size matters more than chemistry for cloud-nucleating ability of aerosol particles, Science, 312, 1375-1378, 2006.

Ehn, M., Petäjä, T., Birmili, W., Junninen, H., Aalto, P., and Kulmala, M.: Non-volatile residuals of newly formed atmospheric particles in the boreal forest, Atmospheric Chemistry and Physics, 7, 677-684, 2007.

Epstein, S. A., Riipinen, I., and Donahue, N. M.: A semiempirical correlation between enthalpy of vaporization and saturation concentration for organic aerosol, Environmental Science & Technology, 44, 743-748, 2009.

Fierz-Schmidhauser, R., Zieger, P., Gysel, M., Kammermann, L., DeCarlo, P., Baltensperger, U., and Weingartner, E.: Measured and predicted aerosol light scattering enhancement factors at the high alpine site Jungfraujoch, Atmospheric Chemistry and Physics, 10, 2319-2333, 2010.

Fuchs, N.: On the stationary charge distribution on aerosol particles in a bipolar ionic atmosphere, Geofisica pura e applicata, 56, 185-193, 1963.

Griffiths, J.: A brief history of mass spectrometry, Analytical Chemistry, 80, 5678-5683, 10.1021/ac8013065, 2008.

Griffiths, P. R., and De Haseth, J. A.: Fourier transform infrared spectrometry, John Wiley & Sons, 2007.

Gross, J. H.: Mass spectrometry: a textbook, Springer Science & Business Media, 2006.

Hakala, J., Mikkilä, J., Hong, J., Ehn, M., and Petäjä, T.: VH-TDMA: A description and verification of an instrument to measure aerosol particle hygroscopicity and volatility, Aerosol Science and Technology, 51, 97-107, 2017.

Hansen, A.: The Aethalometer Book, Magee Scientific Company, Berkeley, California, USA, 2005.

Hansen, A. D. A., Rosen, H., and Novakov, T.: The Aethalometer - an Instrument for the Real-Time Measurement of Optical-Absorption by Aerosol-Particles, Science of the Total Environment, 36, 191-196, Doi 10.1016/0048-9697(84)90265-1, 1984.

Hari, P., and Kulmala, M.: Station for measuring ecosystem-atmosphere relations (SMEAR II), Boreal Environment Research, 10, 315-322, 2005.

Hinds, W. C.: Aerosol technology: properties, behavior, and measurement of airborne particles, John Wiley & Sons, 2012.

Hudson, J. G.: Variability of the relationship between particle size and cloud-nucleating ability, Geophysical Research Letters, 34, 2007.

Huffman, J., Docherty, K., Mohr, C., Cubison, M., Ulbrich, I., Ziemann, P., Onasch, T., and Jimenez, J.: Chemically-resolved volatility measurements of organic aerosol from different sources, Environmental Science & Technology, 43, 5351-5357, 2009a.

Huffman, J. A., Docherty, K. S., Aiken, A. C., Cubison, M. J., Ulbrich, I. M., DeCarlo, P. F., Sueper, D., Jayne, J. T., Worsnop, D. R., Ziemann, P. J., and Jimenez, J. L.: Chemically-resolved aerosol volatility measurements from two megacity field studies, Atmospheric Chemistry and Physics, 9, 7161-7182, 2009b.

IPCC (editors: Griggs, D. J., and Noguer, M.): Climate change 2001: the scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change, Weather, 57, 267-269, 2002.

IPCC (editors: Stocker, T. F., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.): Climate change 2013: the physical science basis. Intergovernmental panel on climate change, working group I contribution to the IPCC fifth assessment report (AR5), New York, 2013.

IPCC (editors: Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K., and Engelbrecht, F).: Impacts of 1.5 °C global warming on natural and human systems, 2018.

Jacobson, M. Z.: Air pollution and global warming: History, science, and solutions, Cambridge University Press, 2012.

Jain, A. K.: Data clustering: 50 years beyond K-means, Pattern Recognition Letters, 31, 651-666, 2010.

Jayne, J. T., Leard, D. C., Zhang, X., Davidovits, P., Smith, K. A., Kolb, C. E., and Worsnop, D. R.: Development of an aerosol mass spectrometer for size and composition analysis of submicron particles, Aerosol Science & Technology, 33, 49-70, 2000.

Jimenez, J. L., Jayne, J. T., Shi, Q., Kolb, C. E., Worsnop, D. R., Yourshaw, I., Seinfeld, J. H., Flagan, R. C., Zhang, X. F., Smith, K. A., Morris, J. W., and Davidovits, P.: Ambient aerosol sampling using the Aerodyne Aerosol Mass Spectrometer, Journal of Geophysical Research-Atmospheres, 108, Artn 842510.1029/2001jd001213, 2003.

Junge, C., and McLaren, E.: Relationship of cloud nuclei spectra to aerosol size distribution and composition, Journal of the Atmospheric Sciences, 28, 382-390, 1971.

Kalberer, M., Paulsen, D., Sax, M., Steinbacher, M., Dommen, J., Prevot, A., Fisseha, R., Weingartner, E., Frankevich, V., and Zenobi, R. J. S.: Identification of polymers as major components of atmospheric organic aerosols, Science, 303, 1659-1662, 2004.

Kaufman, L., and Rousseeuw, P. J.: Finding groups in data: an introduction to cluster analysis, John Wiley & Sons, 2009.

Knutson, E., and Whitby, K.: Aerosol classification by electric mobility: apparatus, theory, and applications, Journal of Aerosol Science, 6, 443-451, 1975.

Kowalski, B. R.: Measurement analysis by pattern recognition, Analytical Chemistry, 47, 1152A-1162a, 1975.

Kowalski, B. R.: Chemometrics, Analytical Chemistry, 52, 112-122, 1980.

Kowalski, B. R.: Analytical chemistry as an information science, TrAC Trends in Analytical Chemistry, 1, 71-74, 1981.

Kristensen, K., Bilde, M., Aalto, P. P., Petäjä, T., and Glasius, M.: Denuder/filter sampling of organic acids and organosulfates at urban and boreal forest sites: Gas/particle distribution and possible sampling artifacts, Atmospheric Environment, 130, 36-53, 2016.

Baron, P. A., and Willeke, K., (ed. Kulkarni, P.,): Aerosol measurement: principles, techniques, and applications, John Wiley & Sons, 2011.

Kulmala, M., Petäjä, T., Mönkkönen, P., Koponen, I. K., Dal Maso, M., Aalto, P. P., Lehtinen, K. E. J., and Kerminen, V. M.: On the growth of nucleation mode particles: source rates of condensable vapor in polluted and clean environments, Atmospheric Chemistry and Physics, 5, 409-416, 2005.

Kulmala, M., Lappalainen, H. K., Petäjä, T., Kerminen, V.-M., Viisanen, Y., Matvienko, G., Melnikov, V., Baklanov, A., Bondur, V., and Kasimov, N.: Pan-Eurasian EXperiment (PEEX) Program: Grand Challenges in the Arctic-Boreal Context, Geography, Environment, Sustainability, 9, 5-18, 2016.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, nature, 521, 436, 2015.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367, 2015.

Liu, P. S., Deng, R., Smith, K. A., Williams, L. R., Jayne, J. T., Canagaratna, M. R., Moore, K., Onasch, T. B., Worsnop, D. R., and Deshler, T.: Transmission

efficiency of an aerodynamic focusing lens system: Comparison of model calculations and laboratory measurements for the Aerodyne Aerosol Mass Spectrometer, Aerosol Science and Technology, 41, 721-733, 2007.

Malm, W. C., and Day, D. E.: Estimates of aerosol species scattering characteristics as a function of relative humidity, Atmospheric Environment, 35, 2845-2860, 2001.

Massoli, P., Lambe, A., Ahern, A., Williams, L., Ehn, M., Mikkilä, J., Canagaratna, M., Brune, W., Onasch, T., and Jayne, J.: Relationship between aerosol oxidation level and hygroscopic properties of laboratory generated secondary organic aerosol (SOA) particles, Geophysical Research Letters, 37, 2010.

Merriam-Webster.com. 2011. https://www.merriam-webster.com (2 October 2018)

McFiggans, G., Artaxo, P., Baltensperger, U., Coe, H., Facchini, M., Feingold, G., Fuzzi, S., Gysel, M., Laaksonen, A., and Lohmann, U.: The effect of physical and chemical aerosol properties on warm cloud droplet activation, Atmospheric Chemistry and Physics, 6, 2593-2649, 2006.

McMurry, P. H.: The history of condensation nucleus counters, Aerosol Science & Technology, 33, 297-322, 2000a.

McMurry, P. H.: A review of atmospheric aerosol measurements, Atmospheric Environment, 34, 1959-1999, 2000b.

McNeill, V. F.: Atmospheric aerosols: clouds, chemistry, and climate, Annual Review of Chemical and Biomolecular Engineering, 8, 427-444, 2017.

Meadows, D. H.: The limits to growth, Green Planet Blues: Critical Perspectives on Global Environmental Politics, 25, 2014.

Merikanto, J., Spracklen, D., Mann, G., Pickering, S., and Carslaw, K.: Impact of nucleation on global CCN, Atmospheric Chemistry and Physics, 9, 8601-8616, 2009.

Mesarovic, M., and Pestel, E.: Mankind at the turning point. The second report to the Club of Rome, 1974.

Mitchell, T. M.: The discipline of machine learning, Carnegie Mellon University, School of Computer Science, Machine Learning Department Pittsburgh, PA, 2006.

Moosmüller, H., Chakrabarty, R., and Arnott, W.: Aerosol light absorption and its measurement: A review, Journal of Quantitative Spectroscopy and Radiative Transfer, 110, 844-878, 2009.

Murphy, W., and Sears, G.: Production of particulate beams, Journal of Applied Physics, 35, 1986-1987, 1964.

Ng, N., Herndon, S., Trimborn, A., Canagaratna, M., Croteau, P.,Onasch, T., Sueper, D., Worsnop, D., Zhang, Q., Sun, Y., and Jayne, J.: An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol, Aerosol Science and Technology, 45, 780–794, 2011.

Niemi, J. V., Saarikoski, S., Aurela, M., Tervahattu, H., Hillamo, R., Westphal, D. L., Aarnio, P., Koskentalo, T., Makkonen, U., Vehkamäki, H., and Kulmala, M.: Long-range transport episodes of fine particles in southern Finland during 1999-2007, Atmospheric Environment, 43, 1255-1264, 10.1016/j.atmosenv.2008.11.022, 2009.

Oberdörster, G., Sharp, Z., Atudorei, V., Elder, A., Gelein, R., Kreyling, W., and Cox, C.: Translocation of inhaled ultrafine particles to the brain, Inhalation Toxicology, 16, 437-445, 2004.

World Health Organsation: Ambient air pollution: A global assessment of exposure and burden of disease, 2016.

Paatero, P., and Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, Environmetrics, 5, 111-126, 1994.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometrics and Intelligent Laboratory Systems, 37, 23-35, Doi 10.1016/S0169-7439(96)00044-5, 1997.

Paatero, P.: The multilinear engine - A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis

model, Journal of Computational and Graphical Statistics, 8, 854-888, Doi 10.2307/1390831, 1999.

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, Chemometrics and Intelligent Laboratory Systems, 60, 253-264, Doi 10.1016/S0169-7439(01)00200-3, 2002.

Paatero, P., Eberly, S., Brown, S. G., and Norris, G. A.: Methods for estimating uncertainty in factor analytic solutions, Atmospheric Measurement Techniques, 7, 781-797, 10.5194/amt-7-781-2014, 2014.

Paciga, A., Karnezi, E., Kostenidou, E., Hildebrandt, L., Psichoudaki, M., Engelhart, G. J., Lee, B.-H., Crippa, M., Prévôt, A. S., and Baltensperger, U.: Volatility of organic aerosol and its components in the megacity of Paris, Atmospheric Chemistry and Physics, 16, 2013-2023, 2016.

Pajunoja, A., Lambe, A. T., Hakala, J., Rastak, N., Cummings, M. J., Brogan, J. F., Hao, L., Paramonov, M., Hong, J., and Prisle, N. L.: Adsorptive uptake of water by semisolid secondary organic aerosols, Geophysical Research Letters, 42, 3063-3068, 2015.

Paul, W., and Steinwedel, H.: Quadrupole mass filter, Z. Naturforsch. A, 8, 488, 1953.

Raatikainen, T., Vaattovaara, P., Tiitta, P., Miettinen, P., Rautiainen, J., Ehn, M., Kulmala, M., Laaksonen, A., and Worsnop, D. R.: Physicochemical properties and origin of organic groups detected in boreal forest using an aerosol mass spectrometer, Atmospheric Chemistry and Physics, 10, 2063-2077, 2010.

Rastak, N., Pajunoja, A., Acosta Navarro, J. C., Ma, J., Song, M., Partridge, D. G., Kirkevåg, A., Leong, Y., Hu, W., and Taylor, N.: Microphysical explanation of the RH-dependent water affinity of biogenic organic aerosol and its importance for climate, Geophysical Research Letters, 44, 5167-5177, 2017.

Reischl, G. P.: Measurement of ambient aerosols by the differential mobility analyzer method: Concepts and realization criteria for the size range between 2 and 500 nm, Aerosol Science and Technology, 14, 5-24, 1991.

Remiarz, R., Agarwal, J., Quant, F., and Sem, G.: Real-time aerodynamic particle size analyzer, Aerosols in the mining and industrial work environments, 3, 879-895, 1983.

Riipinen, I., Pierce, J. R., Donahue, N. M., and Pandis, S. N.: Equilibration time scales of organic aerosol inside thermodenuders: Evaporation kinetics versus thermodynamics, Atmospheric Environment, 44, 597-607, 2010.

Riuttanen, L., Hulkkonen, M., Maso, M. D., Junninen, H., and Kulmala, M.: Trajectory analysis of atmospheric transport of fine particles, SO 2, NO x and O 3 to the SMEAR II station in Finland in 1996–2008, Atmospheric Chemistry and Physics, 13, 2153-2164, 2013.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., and Schellnhuber, H. J.: A safe operating space for humanity, Nature, 461, 472, 2009.

Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain, Psychological Review, 65, 386, 1958.

Rousseeuw, P. J.: Silhouettes - a Graphical Aid to the Interpretation and Validation of Cluster-Analysis, Journal of Computational and Applied Mathematics, 20, 53-65, Doi 10.1016/0377-0427(87)90125-7, 1987.

Rumelhart, D. E., and McClelland, J. L.: Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations, MIT Press, Cambridge, Massachusetts 1986.

Rönkkö, T., Kuuluvainen, H., Karjalainen, P., Keskinen, J., Hillamo, R., Niemi, J. V., Pirjola, L., Timonen, H. J., Saarikoski, S., and Saukko, E.: Traffic is a major source of atmospheric nanocluster aerosol, Proceedings of the National Academy of Sciences, 114, 7549-7554, 2017.

Saarikoski, S., Mäkelä, T., Hillamo, R., Aalto, P. P., Kerminen, V., and Kulmala, M.: Physico-chemical characterization and mass closure of size-segregated atmospheric aerosols in Hyytiälä, Finland, Boreal environment research, 10, 385, 2005.

Samuel, A. L.: Some studies in machine learning using the game of checkers, IBM Journal of research and development, 3, 210-229, 1959.

Sawyer, J.: Man-made carbon dioxide and the "greenhouse" effect, Nature, 239, 23, 1972.

Seinfeld, J. H., and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, 2012.

Semmler, M., Seitz, J., Erbe, F., Mayer, P., Heyder, J., Oberdörster, G., and Kreyling, W.: Long-term clearance kinetics of inhaled ultrafine insoluble iridium particles from the rat lung, including transient translocation into secondary organs, Inhalation toxicology, 16, 453-459, 2004.

Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J., Fushimi, A., Enami, S., Arangio, A. M., Fröhlich-Nowoisky, J., and Fujitani, Y.: Aerosol health effects from molecular to global scales, Environmental Science & Technology, 51, 13545-13567, 2017.

Shoenfeld, P., and DeVoe, J.: Statistical and mathematical methods in analytical chemistry, Analytical Chemistry, 48, 403-411, 1976.

Smith, J. N., Barsanti, K. C., Friedli, H. R., Ehn, M., Kulmala, M., Collins, D. R., Scheckman, J. H., Williams, B. J., and McMurry, P. H. J. P. o. t. N. A. o. S.: Observations of aminium salts in atmospheric nanoparticles and possible climatic implications, 107, 6634-6639, 2010.

Sobanska, S., Falgayrac, G., Rimetz-Planchon, J., Perdrix, E., Brémard, C., and Barbillat, J.: Resolving the internal structure of individual atmospheric aerosol particle by the combination of Atomic Force Microscopy, ESEM–EDX, Raman and ToF–SIMS imaging, Microchemical Journal, 114, 89-98, 2014.

Sogacheva, L., Dal Maso, M., Kerminen, V.-M., and Kulmala, M.: Probability of nucleation events and aerosol particle concentration in different air mass types arriving at Hyytiälä, southern Finland, based on back trajectories analysis, Boreal Environment Research, 10, 479-491, 2005.

Spath, H.: Cluster analysis algorithms: for data reduction and clasification of objects, John Wiley & Sons, 1980.

Spearman, C.: " General Intelligence," objectively determined and measured, The American Journal of Psychology, 15, 201-292, 1904.

Stein, S. E., and Scott, D. R.: Optimization and testing of mass spectral library search algorithms for compound identification, Journal of the American Society for Mass Spectrometry, 5, 859-866, 10.1016/1044-0305(94)87009-8, 1994.

Stephens, W.: A pulsed mass spectrometer with time dispersion, Physical Review, 69, 691-691, 1946.

Thomson, J. J.: Rays of positive electricity, Journal of the Röntgen Society, 18, 150-151, 1922.

Uin, J.: Integrating Nephelometer Instrument Handbook, DOE ARM Climate Research Facility, Washington, DC (United States), 2016.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, Atmospheric Chemistry and Physics, 9, 2891-2918, 10.5194/acp-9-2891-2009, 2009.

Vandeginste, B.: Optimization of analytical information, TrAC Trends in Analytical Chemistry, 1, 210-215, 1982.

Vincent, D. F.: The orgin and development of factor analysis, Applied Statistics, 107-117, 1953.

Virkkula, A., Mäkelä, T., Hillamo, R., Yli-Tuomi, T., Hirsikko, A., Hämeri, K., and Koponen, I. K.: A simple procedure for correcting loading effects of aethalometer data, Journal of the Air & Waste Management Association, 57, 1214-1222, 2007.

Virtanen, A., Joutsensaari, J., Koop, T., Kannosto, J., Yli-Pirilä, P., Leskinen, J., Mäkelä, J. M., Holopainen, J. K., Pöschl, U., and Kulmala, M.: An amorphous solid state of biogenic secondary organic aerosol particles, Nature, 467, 824, 2010.

Went, F. W.: Blue hazes in the atmosphere, Nature, 187, 641, 1960.

White, W. H.: Reduction of visibility by sulphates in photochemical smog, Nature, 264, 735, 1976.

Wien, W.: Ueber die Fragen, welche die translatorische Bewegung des Lichtäthers betreffen, 1898.

Williams, J., Crowley, J., Fischer, H., Harder, H., Martinez, M., Petäjä, T., Rinne, J., Back, J., Boy, M., Dal Maso, M., Hakala, J., Kajos, M., Keronen, P., Rantala, P., Aalto, J., Aaltonen, H., Paatero, J., Vesala, T., Hakola, H., Levula, J., Pohja, T., Herrmann, F., Auld, J., Mesarchaki, E., Song, W., Yassaa, N., Nölscher, A., Johnson, A. M., Custer, T., Sinha, V., Thieser, J., Pouvesle, N., Taraborrelli, D., Tang, M. J., Bozem, H., Hosaynali-Beygi, Z., Axinte, R., Oswald, R., Novelli, A., Kubistin, D., Hens, K., Javed, U., Trawny, K., Breitenberger, C., Hidalgo, P. J., Ebben, C. J., Geiger, F. M., Corrigan, A. L., Russell, L. M., Ouwersloot, H. G., de Arellano, J. V. G., Ganzeveld, L., Vogel, A., Beck, M., Bayerle, A., Kampf, C. J., Bertelmann, M., Kollner, F., Hoffmann, T., Valverde, J., Gonzalez, D., Riekkola, M. L., Kulmala, M., and Lelieveld, J.: The summertime Boreal forest field measurement intensive (HUMPPA-COPEC-2010): an overview of meteorological and chemical influences, Atmospheric Chemistry and Physics, 11, 10599-10618, 10.5194/acp-11-10599-2011, 2011.

Wiza, J. L.: Microchannel plate detectors, Nuclear Instruments and Methods, 162, 587-601, 1979.

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Ulbrich, I. M., Ng, N. L., Worsnop, D. R., and Sun, Y.: Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review, Analytical and Bioanalytical Chemistry, 401, 3045-3067, 10.1007/s00216-011-5355-y, 2011.

Zieger, P., Aalto, P., Aaltonen, V., Äijälä, M., Backman, J., Hong, J., Komppula, M., Krejci, R., Laborde, M., and Lampilahti, J.: Low hygroscopic scattering enhancement of boreal aerosol and the implications for a columnar optical closure study, Atmospheric Chemistry and Physics, 15, 7247-7267, 2015.

Zubin, J.: A technique for measuring like-mindedness, The Journal of Abnormal and Social Psychology, 33, 508, 1938.