Technical note: Determining chemical composition of atmospheric single particles by a standard-free mass calibration algorithm

Shao Shi[1,2], Jinghao Zhai[1,2*], Xin Yang[1,2*], Yechun Ruan[3], Yuanlong Huang[4], Xujian Chen[5], Antai Zhang[1,2], Jianhuai Ye[1,2], Guomao Zheng[1,2], Baohua Cai[1,2], Yaling Zeng[1,2], Yixiang Wang[1,2], Chunbo Xing[1,2], Yujie Zhang[1,2], Tzung-May Fu[1,2], Lei Zhu[1,2], Huizhong Shen[1,2] and Chen Wang[1,2]

[1]Shenzhen Key Laboratory of Precision Measurement and Early Warning Technology for Urban Environmental Health Risks, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[2]Guangdong Provincial Observation and Research Station for Coastal Atmosphere and Climate of the Greater Bay Area, Shenzhen 518055, China
[3]Institute of Future Networks, Southern University of Science and Technology, Shenzhen 518055, China
[4]College of Engineering, Eastern Institute for Advanced Study, Ningbo, 315200, China
[5]Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, China

Correspondence to: Jinghao Zhai (zhaijh@sustech.edu.cn) and Xin Yang (yangx@sustech.edu.cn)

Abstract. The chemical composition of individual particles can be revealed by single particle mass spectrometers (SPMS). With higher accuracy in the ratio of mass to charge (m/z), more detailed chemical information could be obtained. In the SPMS, the conventional standard-based calibration methods (internal/external) are constrained by the inhomogeneity of ionization lasers and the finite focusing ability of the inlet system, etc., therefore, the mass accuracy is restricted. In this study, we obtained the detailed and trustable chemical composition of single particles utilizing a standard-free mass calibration algorithm. In the algorithm, the characteristic distributions of hundreds of ions were concluded and collected in a database denoted as prototype. Each single-particle mass spectrum was initially calibrated by a function with specific coefficients. The range of coefficients was constrained by the magnitude of mass deviation to a finite vector space. To find the optimal coefficient vector, the conformity of each initially calibrated spectrum to the prototype dataset was assessed. The optimum calibrated spectrum was obtained with maximum conformity. For more than 98% ambient particles, a twentyfold improvement in mass accuracy, from ~10,000 ppm (integer) to ~500 ppm (2 decimal places) was achieved. The improved mass accuracy validated the determination of adjacent ions with m/z difference ~0.05 Th. Furthermore, atmospheric particulate trace elements that were poorly studied before are specified. The obtained detailed single-particle-level chemical information could help understand the source apportionment, reaction mechanism, and mixing state of atmospheric particles.

1 Introduction

Chemical composition analysis of individual particles is crucial in studying the mechanism of atmospheric multiphase reactions, single-particle mixing states, and aerosol source apportionment (Pratt et al., 2009; Wang et al., 2019; Hatch et al., 2011; Gard et al., 1998; Kirpes et al., 2019, Zhai et al., 2023). The in-situ mass spectra generation of individual particles can

be enabled by single particle mass spectrometers (SPMS) (Prather et al., 1994; Gard et al., 1997; Lu et al., 2022; Pratt and Prather, 2012; Su et al., 2004; Spencer and Prather, 2006; Schade et al., 2019; Zawadowicz et al., 2015; Li et al., 2023). To reveal the chemical composition from single-particle mass spectra, accurate determination of ions is essential. While the resolving ability of SPMS has been improved by the delayed extraction technique (Li et al., 2018), discerning ions according to the single-particle mass spectra is still limited due to the lack of proper calibration methods (Chudinov et al., 2019; Zhu et al., 2020). Thus, the ratio of mass to charge (m/z) in single-particle mass spectra was in low accuracy and were conventionally rounded to the nearest integers (~10,000 ppm). The obtained integral mass spectra were frequently used in previous studies (Anders et al., 2023; Mcnamara et al., 2020; Zhang et al., 2016; Zhang et al., 2020; Lian et al., 2020; Ault et al., 2010; Xiao et al., 2018; Wang et al., 2019).

Mass calibration transforms the biased mass spectrum into the corresponding accurate mass spectrum (Montaudo et al., 1994). The transformation could be described using mathematical functions (Kozhinov et al., 2013; Kolarova et al., 2017; Lou et al., 2010). For instance, the calibration functions in the time-of-flight (TOF) mass spectrometer (mass analyzer used in the SPMS) could be described by polynomial functions (Li et al., 2011; Romson and Emmer, 2021; Green et al., 2006). For mass calibration, the coefficients in the calibration functions are the uncertainties requiring determination (Montaudo et al., 1994). The coefficients are typically determined using internal or external mass calibration methods, where standard chemicals (calibrants) are conventionally measured alongside the analyte as references, either collectively or separately (O'connor and Costello, 2000; Boskamp et al., 2020; Gobom et al., 2002). Calibrants are used to specify the exact m/z of the measured reference peaks (usually at least two), thereby determining the calibration coefficients before operating the raw spectrum with the calibration function. The internal calibration method is not feasible in the SPMS, and may alter the particle composition caused by the addition of calibrants (O'connor and Costello, 2000). Meanwhile, the application of external mass calibration to SPMS is primarily limited by various uncertainties, including but not limited to the focusing capability of the aerosol inlet system (Wang et al., 2010; Dienes, 2003), and the inhomogeneity of the ionization laser (Text S1, Fig. S1) (Wenzel and Prather, 2004), hence, coefficients in the calibration function should be specifically determined for each particle (Wang et al., 2010; Chudinov et al., 2019; Zhu et al., 2020; Chen et al., 2020; Clemen et al., 2020). Generally, neither internal nor external calibration methods could be practical in the SPMS, it is necessary to establish calibration methods that are standard-free.

Previous studies attempted to avoid the addition of standard chemicals for calibrating the SPMS spectra, by assuming the exact m/z of certain peaks in the raw spectrum (Chudinov et al., 2019; Zhu et al., 2020). For instance, if a peak was observed between 23.50 Th and 24.49 Th in the raw spectrum, it was empirically assumed to be Mg^+ (23.98 Th), without considering other possibilities such as C_2^+ (24.00 Th) within the same range. Logically, the exact m/z could only be obtained after the mass calibration, implying that the rationality of the assumption could not be proved during the calibration process. Additionally, chemicals in the assumptions were used to serve as calibrants, although standard chemicals were not physically added. Therefore, the calibration approach in previous studies was empirical and could be subject to unreliability (Chudinov et al., 2019; Zhu et al., 2020).

65   Here, we present a calibration algorithm that eliminates the need for internal, external, or assumed calibrants for single-particle mass spectra. We first developed a calibration theory, in which the traits possessed by accurate mass spectra were concluded. A value function was established to quantitatively assess the quality of calibration. Calibration was successfully applied to more than 98% of the total 12 million SPMS spectra of ambient aerosols with an average mass resolution of 2000, and the bilateral mass deviation range was limited to ~0.05 Th on average. With the improved mass accuracy (~500 ppm), the

70 detailed chemical composition of single particles was revealed. The improved understanding of particle composition was also proved by increased spectra entropy from an information theory perspective.

2 Methodology

2.1 Field sampling and data acquisition

A sampling campaign was conducted at Southern University of Science and Technology (SUSTech) located in urban Shenzhen,

75 China (22.604°N, 114.006°E, 100 m above sea level) from April 2, 2021, to April 30, 2021. TOF spectra of 12,371,204 individual particles were continuously collected using an SPMS (HP-SPAMS, Hexin Instrument Co., Ltd.) with detailed instrumental information provided in Text S1 (Zhai et al., 2023). Briefly, aerosols are focused using an aerodynamic focus lens (AFL) to form a particle beam. Particles in the beam are then sequentially ionized by laser desorption/ionization (LDI) technique. The resulting ions fly under the force of accelerating potential in a TOF-MS. Time-of-flight data are obtained by

80 calculating raw m/z with predetermined parameters of the TOF-MS, e.g., length of flight path and voltage of the accelerating potential. In our HP-SPAMS, the delayed extraction technique is implemented, which improves the mass resolution from 500 to 2000. The improved resolution provided the foundation for achieving ~500 ppm mass accuracy by calibration. Peaks in the obtained TOF spectra (in profile form) were automatically extracted using a built-in software provided by the instrument manufacturer. The resulting centroid spectra were converted to raw mass spectra utilizing a quadratic function (Eq. (S2) in

85 Text S1).

2.2 Development of the calibration algorithm

Calibration functions are used to describe the relationship between m/z_r (the raw mass) and m/z_c (the calibrated mass) in MS (Kozhinov et al., 2013; O'connor and Costello, 2000; Kaiser et al., 2005). The coefficients to be determined could be collectively listed and referred to as a coefficient vector $\boldsymbol{\theta}$ (Eq. (1)), therefore, each dimension of $\boldsymbol{\theta}$ is a specific coefficient in

90 the calibration function. The m/z_r and m/z_c are vectors with variable lengths according to the number of peaks in each spectrum. S_r and S_c are the spectra which include peak intensity information (peak height, peak area, or relative peak area) corresponding to m/z_r and m/z_c.

$$m/z_c = f(m/z_r, \boldsymbol{\theta}) \tag{1}$$

The only uncertainty in Eq. (1) is θ, which means mass spectra can be calibrated if θ is determined. Such that, the mass calibration process is in principle determining θ. Since the range of mass deviation is finite, the vector space of all possible θ is also constrained. Therefore, if we obtained the mass deviation range through a large dataset, we could accordingly find the variation range of each dimension of θ.

An initially calibrated spectrum is created by substituting θ into Eq. (1). Among all θ in θ space, there exists only one optimal θ, denote θ_{opt}, that yields the accurate spectrum. An indicator is required to determine whether a specific θ is optimal. In our algorithm, the indicator is a value function. The value function assesses the quality of calibration by analyzing the traits that the calibrated spectrum possesses, and could potentially guide the search for the θ_{opt}. Traits are characteristics that distinguish accurate spectra from inaccurate ones. The *value* function is defined in Eq. (2) and (3), where w_i represents the weight assigned to each trait and is the number of *m/z* involved in the specific trait, n represents the total number of traits, S_t is the initially calibrated spectrum, T_i is the ith trait in a database named prototype, and *match* is a function for determining the presence of T_i in each spectrum. The method for the selection of traits is provided in detail in Text S2. A searching range of *m/z* comparable to the resolution is used in *match*. Generally, there exist 3 types of traits, and examples of the traits are given in Table 1 for atmospheric aerosol research. All the traits are collected in a database named prototype, and the congruence of each trait in the prototype should be tested with S_t as shown in Eq. (3). Specifically, in the traits typed "isotopic distribution (fraction) of ions", both the isotopic ratios and the exact m/z serve as conditions for trait matching in our calibration.

Table 1. Three types of traits in the prototype database for calibration.

Trait type	Criterion [a]	Related ions
	$p(\sim 6.015 \text{ Th}): p(\sim 7.015 \text{ Th}) \approx 0.082: 1$	Possibly $^6Li^+$, $^7Li^+$
	$p(\sim 11.999 \text{ Th}): p(\sim 13.003 \text{ Th}) \approx 1: 0.011$	Possibly $^{12}C^+$, $^{13}C^+$
	$p(\sim 17.026 \text{ Th}): p(\sim 18.023 \text{ Th}) \approx 1: 0.004$	Possibly NH_3^+
	$p(\sim 23.984 \text{ Th}): p(\sim 24.985 \text{ Th}): p(\sim 25.982 \text{ Th})$ $\approx 1: 0.127: 0.140$	Possibly $^{24}Mg^+$, $^{25}Mg^+$, $^{26}Mg^+$
Isotopic distribution (fraction) of ions [b]	$p(\sim 28.030 \text{ Th}): p(\sim 29.034 \text{ Th}) \approx 1: 0.021$	Possibly $C_2H_4^+$
	$p(\sim 29.013 \text{ Th}): p(\sim 30.010 \text{ Th}) \approx 1: 0.007$	Possibly N_2H^+
	$p(\sim 31.018 \text{ Th}): p(\sim 32.021 \text{ Th}) \approx 1: 0.011$	Possibly CH_3O^+
	$p(\sim 35.999 \text{ Th}): p(\sim 37.003 \text{ Th}) \approx 1: 0.032$	Possibly C_3^+
	$p(\sim 38.015 \text{ Th}): p(\sim 39.018 \text{ Th}) \approx 1: 0.032$	Possibly $C_3H_2^+$
	$p(\sim 38.963 \text{ Th}): p(\sim 40.961 \text{ Th}) \approx 1: 0.072$	Possibly $^{39}K^+$, $^{41}K^+$
	$p(\sim 39.962 \text{ Th}): p(\sim 43.955 \text{ Th}) \approx 1: 0.022$	Possibly $^{40}Ca^+$, $^{44}Ca^+$

	$p(\sim57.935$ Th$): p(\sim59.930$ Th$): p(\sim61.928$ Th$) \approx 1: 0.385: 0.05$	Possibly $^{58}Ni^+$, $^{60}Ni^+$, $^{62}Ni^+$
	$p(\sim125.859$ Th$): p(\sim127.857$ Th$): p(\sim129.855$ Th$) \approx 1: 0.889: 0.120$	Possibly $^{126}Cu_2^+$, $^{128}Cu_2^+$, $^{130}Cu_2^+$
	$p(\sim197.967$ Th$): p(\sim198.968$ Th$): p(\sim199.968$ Th$) : p(\sim200.970$ Th$): p(\sim201.971$ Th$): p(\sim203.973$ Th$) \approx 0.334: 0.565: 0.774: 0.441: 1: 0.23007$	Possibly $^{198}Hg^+$, $^{199}Hg^+$, $^{220}Hg^+$, $^{201}Hg^+$, $^{202}Hg^+$, $^{204}Hg^+$
	$p(\sim205.974$ Th$): p(\sim206.976$ Th$) : p(\sim207.977$ Th$) \approx 0.460: 0.422: 1$	Possibly $^{206}Pb^+$, $^{207}Pb^+$, $^{208}Pb^+$
Chemical distribution patterns [c]	~50.95 Th & ~66.95 Th	Possibly V^+ / VO^+
	~47.948 Th & ~63.943 Th	Possibly Ti^+ / TiO^+
	$\sim n*11.999$ Th, $n = 2, 3, 4, 5, \ldots$	Possibly C_n, $n = 2, 3, 4, 5, \ldots$
	W^+ distribution & WO^+ distribution	Possibly W^+, WO^+
	Pb^+ distribution & PbO^+ distribution	Possibly Pb^+ / PbO^+
	$\sim(n*1.007 + m*11.999)$ Th, $n, m = 1, 2, 3, \ldots$	Possibly C_nH_m, $n, m = 1, 2, 3, \ldots$
Isolate ions [d]	~7.015 Th exists	Certainly $^7Li^+$
	~11.999 Th exists	Certainly ^{12}C
	~-1.008 Th exists	Certainly $^1H^-$
	~-12.001 Th exists	Certainly $^{12}C^-$

a. The peak area at specific m/z is denoted $p(m/z)$.

b. Once the isotopic pattern is matched, the criterion of trait corresponding to it could be satisfied. However, the behavior of the initially calibrated spectrum on the entire prototype database should be checked using the value function before determining whether one trait is truly positive. Thus, a weight was assigned to each trait for quantification of the matching process. In addition, isotopic fraction may vary in different aerosol particles of different sources, such that an error range (5% in our calibration) was considered in practice.

c. The listed pairs of species may or may not coexist in a spectrum, with the weight assigned to each trait corresponding to the number of ions involved.

d. Isolate ions are far from adjacent ions ($\Delta m/z > \sim1$ Th) with no other ions confusing their determination.

$$value = \sum_{i=1}^{n} w_i \, match(T_i, S_{\mathrm{t}}) \tag{2}$$

$$match(T_i, S_{\mathrm{t}}) = \begin{cases} 1, & S_{\mathrm{t}} \; possesses \; T_i \\ 0, & S_{\mathrm{t}} \; does \; not \; possesses \; T_i \end{cases} \tag{3}$$

As the value function is defined, the quality of an initial calibration is quantifiable, because the optimal calibration attempt generates the maximum *value*. If none of the $\boldsymbol{\theta}$ allows a spectrum possessing a *value* > 0, the spectrum cannot be calibrated until the prototype is expanded to include additional traits. However, how the *value* is distributed in the $\boldsymbol{\theta}$ space is unknown, which means an advanced optimization algorithm is not currently suitable for optimizing $\boldsymbol{\theta}$. Thus, the obtainment of $\boldsymbol{\theta}_{\mathrm{opt}}$ requires a thorough search in the entire $\boldsymbol{\theta}$ space. Therefore, every possible $\boldsymbol{\theta}$ in $\boldsymbol{\theta}$ space should be traversed and tested to generate a S_{t} (Fig. 1). In our iterative search for $\boldsymbol{\theta}_{\mathrm{opt}}$ in $\boldsymbol{\theta}$ space, the step size of iteration is set to be sufficiently small (at least comparable to the resolving ability of the MS) to ensure thorough exploration of the $\boldsymbol{\theta}$ space for reaching the global optimum. In other words, the step size ensures the calibration converges toward the optimal solution and was determined by the resolution of the mass spectrometer used.

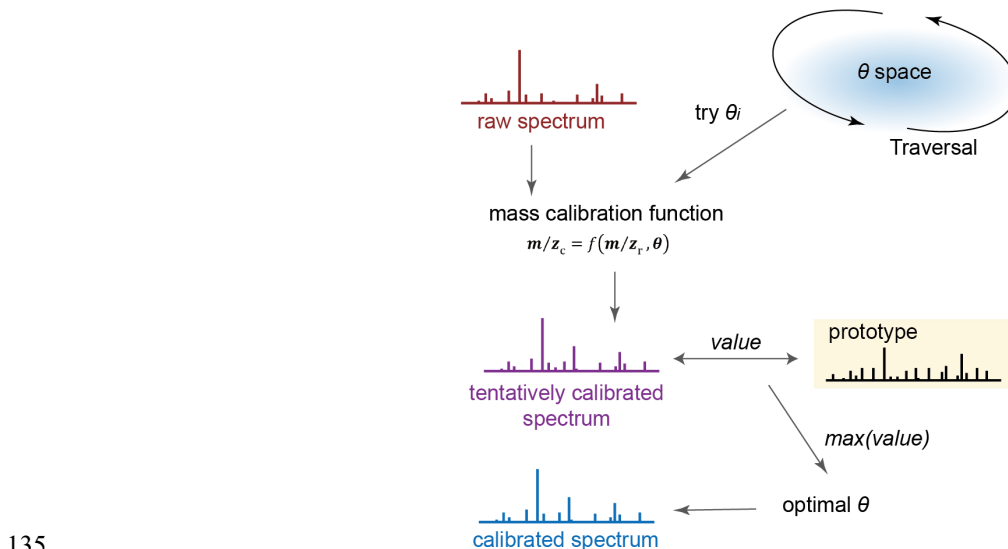



Figure 1. Simplified procedure of the calibration algorithm. Raw spectrum is calibrated using mass deviation function with candidate coefficient pairs θ_i to generate initially calibrated spectra. A *value* is calculated using a value function based on a prototype dataset. A calibrated spectrum is obtained when the *value* is maximal.

2.3 Initialization and localization of the algorithm

The calibration function for TOF-MS is polynomial (Eq. (4)) (Chudinov et al., 2019; Zhu et al., 2020; Gobom et al., 2002), where a_j is the coefficient to be determined, and v refers to the order of approximation.

$$m/z_c = \sum_{j=0}^{v} a_j (m/z_r)^j \qquad (4)$$

The coefficients of Eq. (4) are restricted within a specific range, partially because of the constrained initial position deviation of particles and the size of the laser spot (Text S1, Fig. S1). Nevertheless, the actual deviation range of the measured m/z could be obtained by analyzing the m/z deviation of ions that have no adjacent ions interfering within the deviation range in the average spectrum. For instance, $^{12}C^+$ was selected as such an ion. With the maximum mass deviation, we constrained the variation range of each dimension in θ in the calibration function and, therefore constrained the coefficient space.

The calibration algorithm was implemented using MATLAB (ver. 2022b), and a data processing and calibration software named SSSDA ver. 1.0 was developed (accessible at https://github.com/s129136908794904/SSSDA-ver.1.0). For the current resolution of the SPMS, a linear function is sufficient for calibration (let $v = 1$) (Chudinov et al., 2019; Zhu et al., 2020; Gobom et al., 2002). Prototype was specifically tailored for ambient atmospheric aerosol research, and includes a large number of chemical species and distribution patterns commonly found in ambient aerosols (Table1, Text S2) (Steinfeld, 1998; Zhang et al., 2020; Zhang et al., 2016; Chen et al., 2020; Hatch et al., 2011; Zhang et al., 2019a; Zhai et al., 2023; Anders et al., 2023; Pratt et al., 2009; Pratt and Prather, 2012; Mcnamara et al., 2020). More than 1000 traits are used in this study for calibration, not limited to the examples in Table 1. The step size for iteration ensures the m/z variation of 0.025 Th at $m/z = 50$ Th, corresponding to the resolving ability of the SPMS used in this study. An empirical function (Eq. S1) was employed to describe the relationship between m/z and the resolution of SPMS (Du et al., 2022). Out of the total dataset of 12,371,204 mass spectra, 12,195,009 were successfully calibrated.

2.4 Efficiency, universality, and validity of the algorithm

The algorithm exhibits a linear complexity to the number of spectra (Fig. 1, Text S3), as the interference between the calibration of different spectra was absent. The linear complexity theoretically makes the algorithm efficient in large datasets. From another perspective, the complexity of calibrating individual particles is $O(nlogn)$ related to the number of peaks in each spectrum (Text S3), which ensures that the algorithm is efficient in calibrating individual spectra. Furthermore, parallel computing is suitable for implementing the algorithm because there is no interference in the spontaneous calibration of different mass spectra. We performed the calibration of 12 million mass spectra on a private computer with 12th Gen Intel (R) Core (TM) i9-12900K CPU. Exceptional high time efficiency is observed with only ~12 hours consumed.

The calibration failure rate was less than 2% of the total spectra ($value = 0$), primarily due to the absence of corresponding traits in the prototype database for calibration. The high success rate (> 98%) reflects the universality of the algorithm. In addition, since a searching range of m/z is used for matching traits (Text S3), relationships between m/z_c and m/z_r may deviate from the calibration function. In our calibration, the correlation between m/z_c and m/z_r values was quantified by the absolute value of the correlation coefficient $|r|$. All the spectra possess $|r|$ approximating 1 (Fig. S2), indicating high accordance with the calibration function (Zhu et al., 2020).

3 Results and discussions

3.1 Determination of isotopic ions in single particles

The determination of ions becomes accurate after calibration since the calibrated spectra demonstrate a noticeable improvement in mass accuracy (Fig. S3, Table S1). To evaluate the enhancement in mass accuracy, the mass deviations of several isotopic ions has been examined (Fig. 2). Based on the calibrated spectra, multiple markers were employed in searching for each isotopic ion to ensure accuracy. For instance, the presence of both Pb^+ and PbO^+ signals were required to ensure the inclusion of Pb^+ in the spectrum. The raw spectra exhibit deviations larger than 1.5 Th at ~208 Th, indicating challenges in achieving integer-level identification of ions. Instead, the calibrated spectra exhibit a variation less than ± 0.05 Th (500 ppm) around the same position, enabling the accurate identification of ions such as Pb isotopes (Fig. 2e, Fig. 3a).

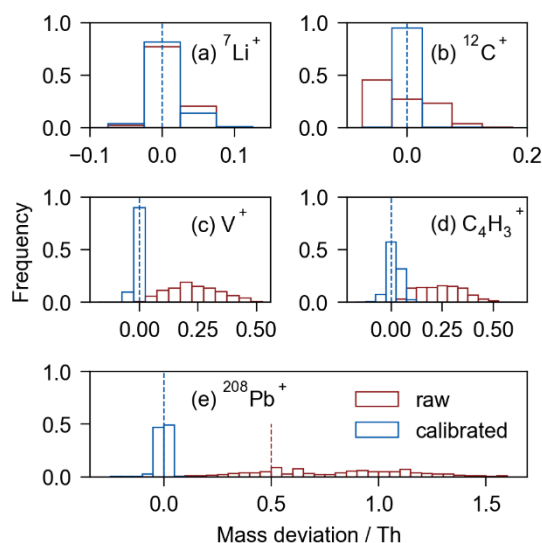



Figure 2. Mass deviation distribution for the raw (in red) and calibrated (in blue) spectra of 5 selected ions. The blue dashed line indicates the location of the exact mass. (a) Signals of Li^+ (42,858 Li^+-containing spectra) exhibit mass deviation mostly less than 0.025 Th in both raw and calibrated spectra. (b) Most $^{12}C^+$ signals (7,739,415 spectra out of 8,168,248 spectra) exhibit mass deviations larger than 0.025 Th in the raw spectra, meanwhile, they are 100% calibrated within 0.025 Th deviation. (c) - (d) More than 95% of V^+ (50.95 Th, 296,259 spectra in total), and $C_4H_3^+$ (51.00 Th, 105,477 spectra in total) signals exhibit mass deviations larger than 0.025 Th in raw spectra. Nevertheless, most deviations in the calibrated spectra are less than 0.025 Th. (e) Most of the $^{208}Pb^+$ signals (127,138 spectra in total) exhibit mass deviation larger than 0.5 Th (out of the red dash line) in raw spectra. Overall, ~100% of the signals in the calibrated spectra are controlled within ± 0.05 Th deviation interval.

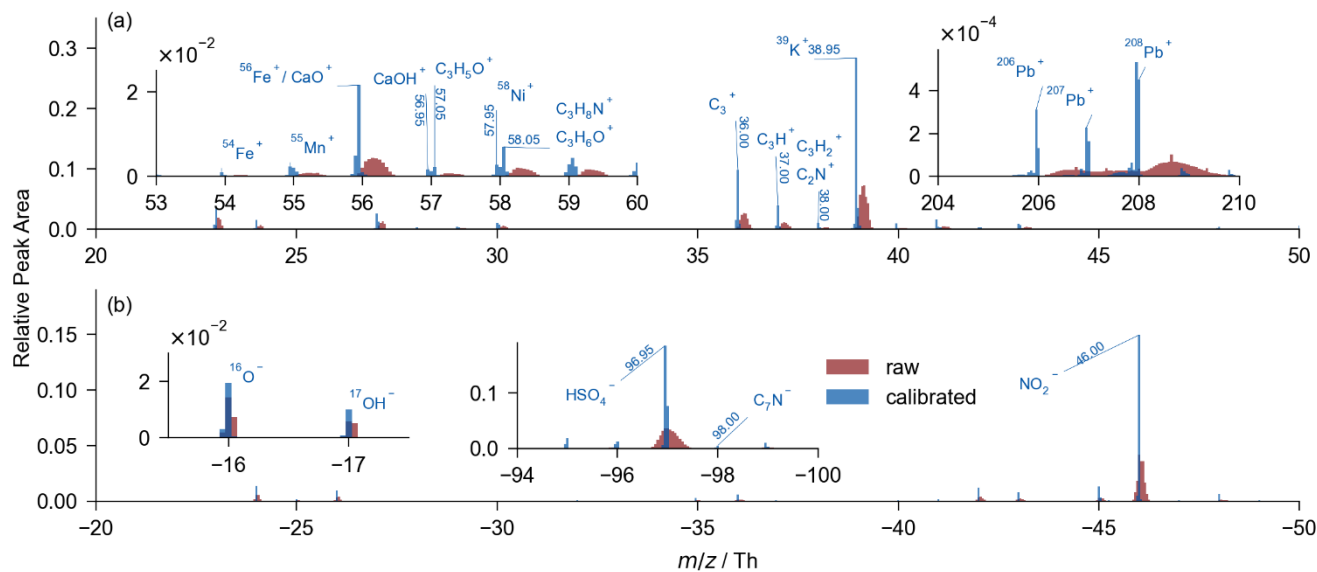



Figure 3. Averaged spectrum of 12 million raw (in red) and calibrated (in blue) spectra. Several intervals of *m/z* are examined for detailed spectrum comparison. (a) The positive spectrum. (b) The negative spectrum. The summation of relative peak area is 1 in each positive or negative spectrum.

In the averaged spectrum of all the 12-million individually calibrated spectra, there is a substantial reduction of mass deviation (Fig. 3). The identification of ions, such as ^{206}Pb, ^{207}Pb, ^{208}Pb, $^{55}Mn^+$ (54.95 Th), and C_7N^- (-98.00 Th), is reliable due to the reduced mass deviation. For studies involving large amounts of mass spectra, accurate mass value is crucial for data analysis such as searching with marker ions or clustering with similarity algorithms. In addition, a substantial increase in the average signal intensity was demonstrated in the calibrated spectrum due to the contraction of the mass deviation range in the averaged spectrum. Therefore, the ions with low signal intensity can be examined in the averaged spectrum with our algorithm. During the sampling campaign, particles containing trace elements (44,407 Ti-containing, 124,293 Ga-containing, 144,385 Rb-containing, 153 Zr-containing, 4,715 Mo-containing, 12,014 Ag-containing, 351 Cd-containing, 4,775 Sn-containing, 5,905 Sb-containing, 3,613 Ba-containing, 1,176 W-containing, and 493 Hg-containing particles) were successfully identified with a precision of ~500 ppm (Fig. 4), although their presence is difficult to detect in the raw spectra and has therefore been little studied in the literature. Generally, the analysis of the trace-element-containing particles is enabled after the calibration. In addition, the isotopic distributions and the accurate *m/z* values in Fig. 4 jointly imply the validity of our calibration.

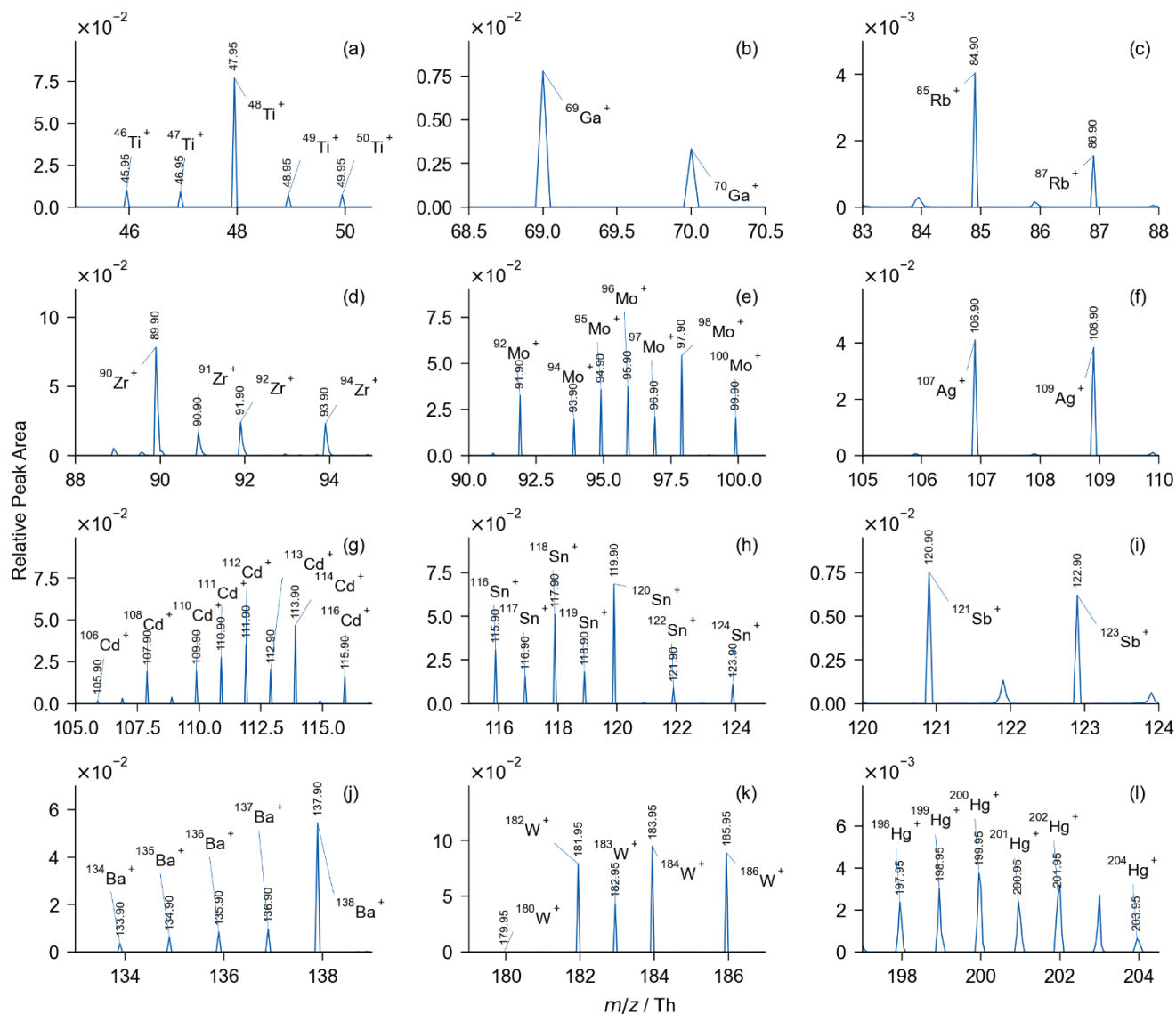



Figure 4. Local averaged spectra of (a) 44,407 Ti-containing, (b) 124,293 Ga-containing, (c) 144,385 Rb-containing, (d) 153 Zr-containing, (e) 4,715 Mo-containing, (f) 12,014 Ag-containing, (g) 351 Cd-containing, (h) 4,775 Sn-containing, (i) 5,905 Sb-containing, (j) 3,613 Ba-containing, (k) 1,176 W-containing, and (l) 493 Hg-containing particles. Particles are grouped using isotopic distribution as markers. Each group of isotopes is identified with the range of fluctuation less than 15%.

3.2 Determination of ions with close *m/z* value

Close ions referring to ions with <0.5 Th difference in exact *m/z* between each other. As shown in Fig. 3, the calibrated peak locations of adjacent ions, such as $^{58}Ni^+$ (57.95 Th), and $C_3H_6O^+$ / $C_3H_8N^+$ (58.05 Th), become precise and therefore could be

determined. Specifically, two pairs of close ions in particles were examined in detail, and the search for these peaks in the spectra was conducted for individual particles.

The first pair is $CaOH^+$ (56.95 Th) / $C_3H_5O^+$ (57.05 Th). We mainly focused on their co-existence in each spectrum here. The marker criterion for spectrum search was *56.95 Th & 57.05 Th coexist in individual spectrum*. We obtained ~6,000 such spectra from our 12-million data (Fig. 5a). The co-existence of $CaOH^+$ / $C_3H_5O^+$ further proved that adjacent peaks could be accurately determined using our calibration algorithm. Overall, the information content of each spectrum is improved by our algorithm.

The second pair is V^+ (50.95 Th) and $C_4H_3^+$ (51.00 Th). V^+ is a typical marker of ship-emitted particles (Zhang et al., 2019a; Zhang et al., 2019b; Wang et al., 2019). Previously the presence of $C_4H_3^+$ often confuses the identification of V^+ signals, because the mass difference (~0.05 Th) is much less than the integer-level accuracy of the SPMS. Therefore, distinguishing V^+ and $C_4H_3^+$ in spectra is crucial in shipping aerosol research. By employing the standard-free calibration algorithm, the achieved high mass accuracy ensures the determination of V^+ and $C_4H_3^+$. Two distinct groups of mass spectra were isolated from our 10-million database of single-particle mass spectra, each containing 296,259 and 105,477 mass spectra with signals at 50.95 Th (V^+) or 51.00 Th ($C_4H_3^+$), respectively. The co-presence of V^+ and $C_4H_3^+$ signal in a single particle is theoretically permitted, though such a scenario was not observed in our database. In the mass spectrum of V^+-containing particles, the signal of vanadium oxide ion (VO^+, 66.95 Th), which is also a marker for shipping emissions, is present (Fig. 5b). Other ions of typical shipping emissions such as Al^+ (27.00 Th), K^+ (38.95 Th), Ca^+ (39.95 Th), Fe^+ (55.95 Th), and $^{58}Ni^+$ (57.95 Th) are also detectable (Zhang et al., 2019a). For the C_4H_3-containing group, signals of $C_mH_n^+$ that correspond to organic aerosols could be observed (Fig. 5c). Additionally, the time series of the two particle groups in terms of number fraction demonstrates distinct temporal behavior (Fig. 5d). The correlation coefficient between the two groups is ~0.15, confirming our differentiation of previously confusing V^+-containing or $C_4H_3^+$-containing particles. Moreover, the successful determination of adjacent ions also validates our calibration algorithm.

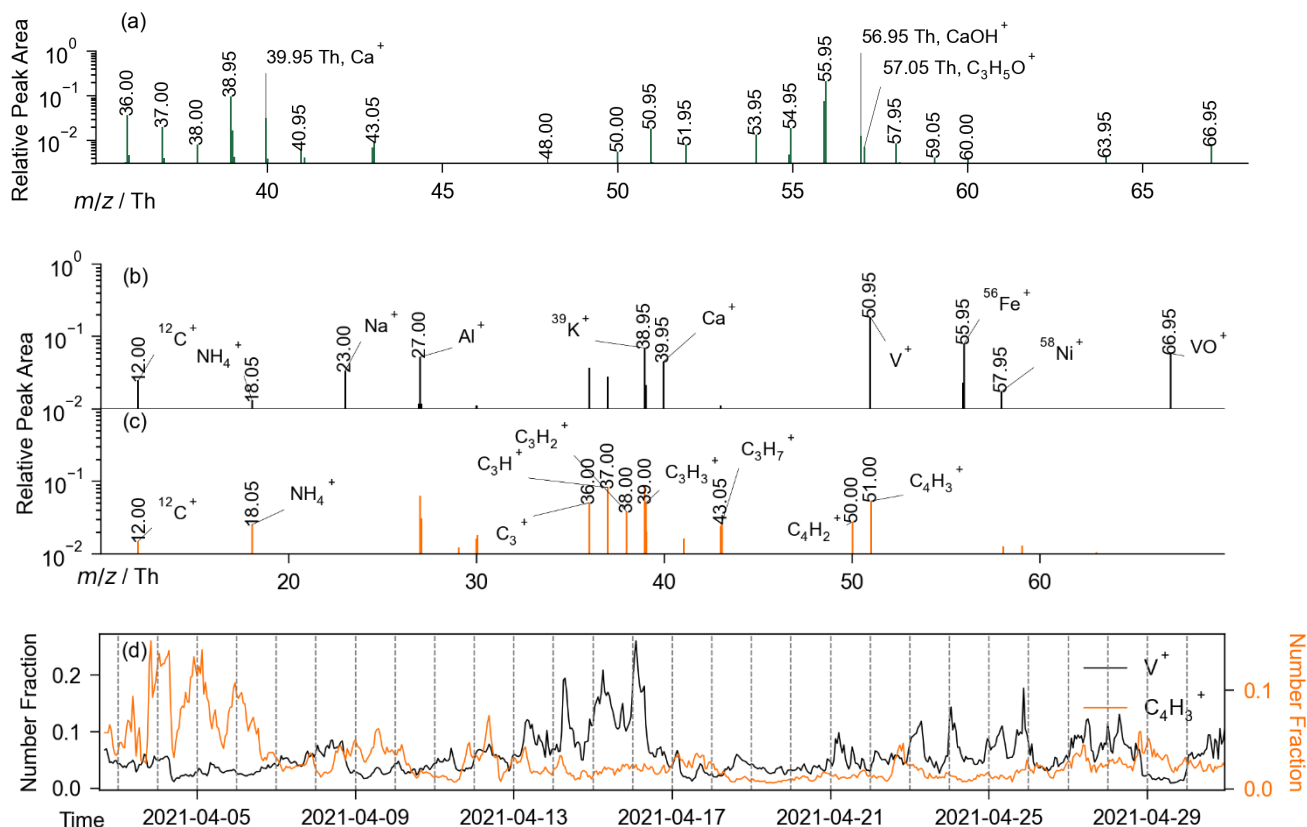



Figure 5. (a) Averaged spectrum (locally illustrated) of ~6,000 spectra that simultaneously contain $CaOH^+$ and $C_3H_5O^+$. Averaged positive spectrum of (b) V^+-containing (in gray) and (c) $C_4H_3^+$-containing (in orange) particles were obtained by averaging millions of particles of the two particle groups separately. (d) Time series of V^+-containing and $C_4H_3^+$-containing particles in number fraction.

3.3 Improved understanding on single-particle composition

Generally, small peaks in mass spectra could emerge if the mass resolution is increased. However, these peaks are not utilizable without calibration. For instance, the coexistence of $CaOH^+ / C_3H_5O^+$ is only determinable with mass accuracy up to ~500 ppm. Therefore, raw m/z is usually rounded to its nearest integer if mass accuracy is limited. The rounded m/z constitutes the integral spectrum, denoted as S_i, and has been frequently used in studies involving SPMS (Anders et al., 2023; Mcnamara et al., 2020; Zhang et al., 2016; Zhang et al., 2020; Lian et al., 2020; Ault et al., 2010; Xiao et al., 2018; Wang et al., 2019). Although the rounding brings convenience to data processing, peaks within the same integer are merged resulting in the elimination of adjacent peaks.

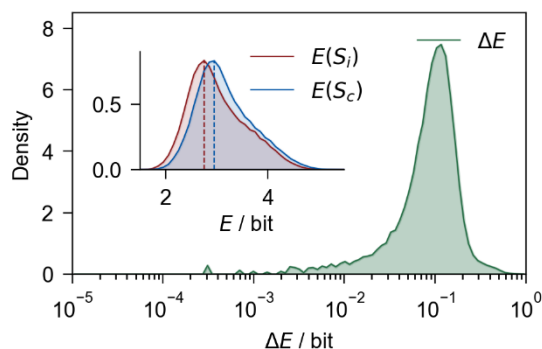
Principally, mass spectra are information that is made up of peaks. According to information theory, the information content in any mass spectrum could be quantified by Shannon entropy (Shannon, 1948). Our algorithm achieves calibration of adjacent m/z, thereby preserving critical peak information. Here, we compared the information content between the calibrated spectrum S_c and its corresponding integral spectrum S_i from an information theory perspective. The amount of information in each spectrum was quantified using Shannon entropy (Eq. (5), $E(S)$ is the Shannon entropy of a spectrum S, I_k is the relative area of the kth peak in S, and q represents the total number of peaks in the spectrum).

$$E(S) = \sum_{k=1}^{q} -I_k \ln (I_k) \tag{5}$$

Figure 6. Increment of Shannon entropy ($\Delta E = E(S_c) - E(S_i)$) of individual spectra. The embedding graph shows the distribution of $E(S_c)$ and $E(S_i)$. The shaded area represents the distribution density.

An obvious increment of E could be observed in Fig. 6, indicating an increment of information content in the S_c compared to the S_i. The peak value of E shifted by ~0.15 bit on average, regardless E is dominantly influenced by major peaks such as K^+, nitrate, and sulfate. In general, our calibration algorithm increased the understanding of single-particle mass spectra, which could be used to determine the chemical composition of individual atmospheric particles.

4 Conclusions and implications

The understanding of single-particle compositions of aerosols is improved by our standard-free mass calibration algorithm. The algorithm efficiently improves the mass accuracy of SPMS spectra from ~10,000 ppm (integer) to ~500 ppm (2 decimal places). The improved mass accuracy enables determining ions in mass spectra within ± 0.025 Th deviation, facilitating detailed investigations of chemical composition at single-particle level. Furthermore, the determination of ions from adjacent ion pairs such as $CaOH^+$ (56.95 Th) / $C_3H_5O^+$ (57.05 Th), and V^+ (50.95 Th) / $C_4H_3^+$ (51.00 Th), which requires a resolution of ~1000 for their separation and an accuracy of ~500 ppm for their determination, is achieved by our algorithm. Particles

280 containing V^+ or $C_4H_3^+$, two confusing ions in previous studies that tracing ship emissions, are successfully determined from each other. In addition, the algorithm enables the analysis of trace species at single-particle level, i.e., Ga, Rb, Zr, Mo, Ag, Cd, Sn, Sb, Ba, W, and Hg isotopes.

Since the detailed information about single particles is revealed at a 500-ppm accuracy level, many more atmospheric species could be determined and involved in the source apportionment of aerosols, the study of chemical reaction mechanisms, and 285 the analysis of single-particle mixing states. Furthermore, the calibration function may appear in different forms for other research interests, depending on the mass analyzer (orbitrap, TOF, etc.) used in the mass spectrometer. Nevertheless, the prototype could be tailored specifically and the algorithm could be implemented in various mass spectrometers. Generally, our standard-free calibration algorithm opens new possibilities for accurate measurement in atmospheric research.

Code and data availability

290 The full data will be made available on request. The SSSDA ver. 1.0 is accessible on GitHub (https://github.com/s129136908794904/SSSDA-ver.1.0). A portion of the prototype was released for testing.

Supplementary information

The single particle mass spectrometer used in this study (Text S1, Fig. S1). Detailed description of traits in the prototype (Text S2). Simplified pseudocode of the algorithm (Text S3). Model fitness of the calibration (Fig. S2). An example of a calibrated 295 single-particle spectrum and the corresponding raw spectrum (Fig. S3). Accuracy and resolution required for separating easily-confused ions typically encountered in aerosol research (Table S1).

Author contribution

SS developed and implemented the algorithm. SS, JZ, XY designed the study. YR, YH, XC, AZ, JY, GZ, BC provided advice on the data analysis process. SS, JZ, and XY prepared the manuscript with comments from all co-authors.

300 **Competing interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41827804, 42305108), Guangdong Basic and Applied Basic Research Foundation (2023A1515011037), Shenzhen Science and Technology Program (KQTD20210811090048025, RCBS20221008093123058), Key-Area Research and Development Plan of Guangdong Province (2020B1111360001), Guangdong Provincial Observation and Research Station for Coastal Atmosphere and Climate of the Greater Bay Area (2021B1212050024), and Shenzhen Key Laboratory of Precision Measurement and Early Warning Technology for Urban Environmental Health Risks (ZDSYS20220606100604008).

References

Anders, L., Schade, J., Rosewig, E. I., Kröger-Badge, T., Irsig, R., Jeong, S., Bendl, J., Saraji-Bozorgzad, M. R., Huang, J.-H., Zhang, F.-Y., Wang, C. C., Adam, T., Sklorz, M., Etzien, U., Buchholz, B., Czech, H., Streibel, T., Passig, J., and Zimmermann, R.: Detection of ship emissions from distillate fuel operation via single-particle profiling of polycyclic aromatic hydrocarbons, Environ. Sci.: Atmos., 3, 1134-1144, https://doi.org/10.1039/d3ea00056g, 2023.

Ault, A. P., Gaston, C. J., Wang, Y., Dominguez, G., Thiemens, M. H., and Prather, K. A.: Characterization of the single particle mixing state of individual ship plume events measured at the port of Los Angeles, Environ. Sci. Technol., 44, 1954-1961, https://doi.org/10.1021/es902985h, 2010.

Boskamp, T., Lachmund, D., Casadonte, R., Hauberg-Lotte, L., Kobarg, J. H., Kriegsmann, J., and Maass, P.: Using the chemical noise background in MALDI mass spectrometry imaging for mass alignment and calibration, Anal. Chem., 92, 1301-1308, https://doi.org/10.1021/acs.analchem.9b04473, 2020.

Chen, Y., Kozlovskiy, V., Du, X., Lv, J., Nikiforov, S., Yu, J., Kolosov, A., Gao, W., Zhou, Z., Huang, Z., and Li, L.: Increase of the particle hit rate in a laser single-particle mass spectrometer by pulse delayed extraction technology, Atmos. Meas. Tech., 13, 941-949, https://doi.org/10.5194/amt-13-941-2020, 2020.

Chudinov, A., Li, L., Zhou, Z., Huang, Z., Gao, W., Yu, J., Nikiforov, S., Pikhtelev, A., Bukharina, A., and Kozlovskiy, V.: Improvement of peaks identification and dynamic range for bi-polar Single Particle Mass Spectrometer, Int. J. Mass Spectrom., 436, 7-17, https://doi.org/10.1016/j.ijms.2018.11.013, 2019.

Clemen, H.-C., Schneider, J., Klimach, T., Helleis, F., Köllner, F., Hünig, A., Rubach, F., Mertes, S., Wex, H., Stratmann, F., Welti, A., Kohl, R., Frank, F., and Borrmann, S.: Optimizing the detection, ablation, and ion extraction efficiency of a single-particle laser ablation mass spectrometer for application in environments with low aerosol particle concentrations, Atmos. Meas. Tech., 13, 5923-5953, https://doi.org/10.5194/amt-13-5923-2020, 2020.

Dienes, T.: Development, characterization, and refinement of a transportable aerosol time-of-flight mass spectrometer, University of California, Riverside, 2003.

Du, X., Xie, Q., Huang, Q., Li, X., Yang, J., Hou, Z., Wang, J., Li, X., Zhou, Z., Huang, Z., Gao, W., and Li, L.: The design and characterization of a High-Performance Single-Particle Aerosol Mass Spectrometer (HP-SPAMS), Atmos. Meas. Tech., 17, 1037–1050, https://doi.org/10.5194/amt-17-1037-2024, 2024.

Gard, E., Mayer, J. E., Morrical, B. D., Dienes, T., Fergenson, D. P., and Prather, K. A.: Real-time analysis of individual atmospheric aerosol particles: Design and performance of a portable ATOFMS, Anal. Chem., 69, 4083-4091, https://doi.org/10.1021/ac970540n, 1997.

Gard, E. E., Kleeman, M. J., Gross, D. S., Hughes, L. S., Allen, J. O., Morrical, B. D., Fergenson, D. P., Dienes, T., E. Gälli, M., Johnson, R. J., Cass, G. R., and Prather, K. A.: Direct observation of heterogeneous chemistry in the atmosphere, Science, 279, 1184-1187, https://doi.org/10.1126/science.279.5354.1184, 1998.

Gobom, J., Mueller, M., Egelhofer, V., Theiss, D., Lehrach, H., and Nordhoff, E.: A Calibration method that simplifies and improves accurate determination of peptide molecular masses by MALDI-TOF MS, Anal. Chem. 74, 3915-3923, https://doi.org/10.1021/ac011203o, 2002.

345   Green, F. M., Gilmore, I. S., and Seah, M. P.: TOF-SIMS: accurate mass scale calibration, J. Am. Soc. Mass Spectrom., 17, 514-523, https://doi.org/10.1016/j.jasms.2005.12.005, 2006.

Hatch, L. E., Creamean, J. M., Ault, A. P., Surratt, J. D., Chan, M. N., Seinfeld, J. H., Edgerton, E. S., Su, Y., and Prather, K. A.: Measurements of isoprene-derived organosulfates in ambient aerosols by aerosol time-of-flight mass spectrometry-part 2: temporal variability and formation mechanisms, Environ. Sci. Technol., 45, 8648-8655, 350   https://doi.org/10.1021/es2011836, 2011.

Kaiser, N. K., Anderson, G. A., and Bruce, J. E.: Improved mass accuracy for tandem mass spectrometry, J. Am. Soc. Mass Spectrom., 16, 463-470, https://doi.org/10.1016/j.jasms.2004.12.005, 2005.

Kirpes, R. M., Bonanno, D., May, N. W., Fraund, M., Barget, A. J., Moffet, R. C., Ault, A. P., and Pratt, K. A.: Wintertime arctic sea spray aerosol composition controlled by sea ice lead microbiology, ACS Cent. Sci., 5, 1760-1767, 355   https://doi.org/10.1021/acscentsci.9b00541, 2019.

Kolarova, L., Prokes, L., Kucera, L., Hampl, A., Pena-Mendez, E., Vanhara, P., and Havel, J.: Clusters of monoisotopic elements for calibration in (TOF) mass spectrometry, J. Am. Soc. Mass Spectrom., 28, 419-427, https://doi.org/10.1007/s13361-016-1567-x, 2017.

Kozhinov, A. N., Zhurov, K. O., and Tsybin, Y. O.: Iterative method for mass spectra recalibration via empirical estimation 360   of the mass calibration function for Fourier transform mass spectrometry-based petroleomics, Anal. Chem., 85, 6437-6445, https://doi.org/10.1021/ac400972y, 2013.

Li, L., Wexler, A. S., Li, X., Hu, L., and Jiang, G.: In situ characterization of bioaerosols at the single-particle level using single-particle mass spectrometry: a promising tool for defending human health against bioaerosol transmission, Anal. Chem., https://doi.org/10.1021/acs.analchem.2c05324, 2023.

365   Li, L., Liu, L., Xu, L., Li, M., Li, X., Gao, W., Huang, Z., and Cheng, P.: Improvement in the mass resolution of single particle mass spectrometry using delayed ion extraction, J. Am. Soc. Mass Spectrom., 29, 2105-2109, https://doi.org/10.1007/s13361-018-2037-4, 2018.

Li, L., Huang, Z., Dong, J., Li, M., Gao, W., Nian, H., Fu, Z., Zhang, G., Bi, X., Cheng, P., and Zhou, Z.: Real time bipolar time-of-flight mass spectrometer for analyzing single aerosol particles, Int. J. Mass Spectrom., 303, 118-124, 370   https://doi.org/10.1016/j.ijms.2011.01.017, 2011.

Lian, X., Zhang, G., Yang, Y., Lin, Q., Fu, Y., Jiang, F., Peng, L., Hu, X., Chen, D., Wang, X., Peng, P. a., Sheng, G., and Bi, X.: Evidence for the formation of imidazole from carbonyls and reduced nitrogen species at the individual particle level in the ambient atmosphere, Environ. Sci. Technol. Lett. 8, 9-15, https://doi.org/10.1021/acs.estlett.0c00722, 2020.

Lou, X., van Dongen, J. L., and Meijer, E. W.: Generation of CsI cluster ions for mass calibration in matrix-assisted laser 375   desorption/ionization mass spectrometry, J. Am. Soc. Mass Spectrom., 21, 1223-1226, https://doi.org/10.1016/j.jasms.2010.02.029, 2010.

Lu, H. L., Su, Z. M., Li, L., and Li, X.: Airborne microbial aerosol detection by combining single particle mass spectrometry and a fluorescent aerosol particle sizer, Anal. Chem., 94, 17861-17867, https://doi.org/10.1021/acs.analchem.2c03636, 2022.

380   McNamara, S. M., Kolesar, K. R., Wang, S., Kirpes, R. M., May, N. W., Gunsch, M. J., Cook, R. D., Fuentes, J. D., Hornbrook, R. S., Apel, E. C., China, S., Laskin, A., and Pratt, K. A.: Observation of road salt aerosol driving inland wintertime atmospheric chlorine chemistry, ACS Cent. Sci., 6, 684-694, https://doi.org/10.1021/acscentsci.9b00994, 2020.

Montaudo, G., Montaudo, M. S., Puglisi, C., and Samperi, F.: Determination of absolute mass values in MALDI-TOF of polymeric materials by a method of self-calibration of the spectra, Anal. Chem., 66, 4366-4369, 385   https://doi.org/10.1021/ac00095a038, 1994.

O'Connor, P. B. and Costello, C. E.: Internal calibration on adjacent samples (InCAS) with Fourier transform mass spectrometry, Anal. Chem., 72, 5881-5885, https://doi.org/10.1021/ac000770t, 2000.

Prather, K. A., Nordmeyer, T., and Salt, K.: Real-time characterization of individual aerosol particles using time-of-flight mass spectrometry, Anal. Chem., 6, 1403-1407, https://doi.org/10.1021/ac00081a007,, 1994.

390   Pratt, K. A. and Prather, K. A.: Mass spectrometry of atmospheric aerosols--recent developments and applications. Part II: On-line mass spectrometry techniques, Mass Spectrom. Rev., 31, 17-48, https://doi.org/10.1002/mas.20330, 2012.

Pratt, K. A., DeMott, P. J., French, J. R., Wang, Z., Westphal, D. L., Heymsfield, A. J., Twohy, C. H., Prenni, A. J., and Prather, K. A.: In situ detection of biological particles in cloud ice-crystals, Nature Geoscience, 2, 398-401, https://doi.org/10.1038/ngeo521, 2009.

395   Romson, J. and Emmer, Å.: Mass calibration options for accurate electrospray ionization mass spectrometry, Int. J. Mass Spectrom., 467, https://doi.org/10.1016/j.ijms.2021.116619, 2021.

Schade, J., Passig, J., Irsig, R., Ehlert, S., Sklorz, M., Adam, T., Li, C., Rudich, Y., and Zimmermann, R.: Spatially shaped laser pulses for the simultaneous detection of polycyclic aromatic hydrocarbons as well as positive and negative inorganic ions in single particle mass spectrometry, Anal. Chem., 91, 10282-10288, https://doi.org/10.1021/acs.analchem.9b02477, 2019.

Shannon, C. E.: A mathematical theory of communication, The Bell System Technical Journal, 27, 379-423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x, 1948.

Spencer, M. T. and Prather, K. A.: Using ATOFMS to determine OC/EC mass fractions in particles, Aerosol Sci. Technol., 40, 585-594, https://doi.org/10.1080/02786820600729138, 2006.

405   Steinfeld, J. I.: Atmospheric chemistry and physics: from air pollution to climate change, 3nd ed.; Wiley: Hoboken, N.J., 2016.

Su, Y., Sipin, M. F., Furutani, H., and Prather, K. A.: Development and characterization of an aerosol time-of-flight mass spectrometer with increased detection efficiency, Anal. Chem., 76, 712-719, https://doi.org/10.1021/ac034797z, 2004.

Wang, X., Shen, Y., Lin, Y., Pan, J., Zhang, Y., Louie, P. K. K., Li, M., and Fu, Q.: Atmospheric pollution from ships and its impact on local air quality at a port site in Shanghai, Atmos. Chem. Phys., 19, 6315-6330, https://doi.org/10.5194/acp-19-6315-2019, 2019.

410

Wenzel, R. J. and Prather, K. A.: Improvements in ion signal reproducibility obtained using a homogeneous laser beam for on-line laser desorption/ionization of single particles, Rapid Commun. Mass Spectrom., 18, 1525-1533, https://doi.org/10.1002/rcm.1509, 2004.

Xiao, Q., Li, M., Liu, H., Fu, M., Deng, F., Lv, Z., Man, H., Jin, X., Liu, S., and He, K.: Characteristics of marine shipping emissions at berth: profiles for particulate matter and volatile organic compounds, Atmos. Chem. Phys., 18, 9527-9545, https://doi.org/10.5194/acp-18-9527-2018, 2018.

415

Wang, X., Chen, H.,Yang, F, and Yang, X.: A new algorithm to correct the particle-to-particle shift in single-particle mass spectrometry analysis, J. Chinese Mass Spectrom. Soc., 31, 179, 2010.

Zawadowicz, M. A., Abdelmonem, A., Mohr, C., Saathoff, H., Froyd, K. D., Murphy, D. M., Leisner, T., and Cziczo, D. J.: Single-particle time-of-flight mass spectrometry utilizing a femtosecond desorption and ionization laser, Anal. Chem., 87, 12221-12229, https://doi.org/10.1021/acs.analchem.5b03158, 2015.

420

Zhai, J., Yu, G., Zhang, J., Shi, S., Yuan, Y., Jiang, S., Xing, C., Cai, B., Zeng, Y., Wang, Y., Zhang, A., Zhang, Y., Fu, T.-M., Zhu, L., Shen, H., Ye, J., Wang, C., Tao, S., Li, M., Zhang, Y., and Yang, X.: Impact of Ship Emissions on Air Quality in the Greater Bay Area in China under the Latest Global Marine Fuel Regulation, Environ. Sci. Technol., 57, 12341-12350, https://doi.org/10.1021/acs.est.3c03950, 2023.

425

Zhang, G., Bi, X., Qiu, N., Han, B., Lin, Q., Peng, L., Chen, D., Wang, X., Peng, P. a., Sheng, G., and Zhou, Z.: The real part of the refractive indices and effective densities for chemically segregated ambient aerosols in Guangzhou measured by a single-particle aerosol mass spectrometer, Atmos. Chem. Phys., 16, 2631-2640, https://doi.org/10.5194/acp-16-2631-2016, 2016.

430   Zhang, G., Lian, X., Fu, Y., Lin, Q., Li, L., Song, W., Wang, Z., Tang, M., Chen, D., Bi, X., Wang, X., and Sheng, G.: High secondary formation of nitrogen-containing organics (NOCs) and its possible link to oxidized organics and ammonium, Atmos. Chem. Phys., 20, 1469-1481, https://doi.org/10.5194/acp-20-1469-2020, 2020.

Zhang, X., Zhang, Y., Liu, Y., Zhao, J., Zhou, Y., Wang, X., Yang, X., Zou, Z., Zhang, C., Fu, Q., Xu, J., Gao, W., Li, N., and Chen, J.: Changes in the SO2 level and PM2.5 components in Shanghai driven by implementing the ship emission control policy, Environ. Sci. Technol., 53, 11580-11587, https://doi.org/10.1021/acs.est.9b03315, 2019a.

435

Zhang, Y., Deng, F., Man, H., Fu, M., Lv, Z., Xiao, Q., Jin, X., Liu, S., He, K., and Liu, H.: Compliance and port air quality features with respect to ship fuel switching regulation: a field observation campaign, SEISO-Bohai, Atmos. Chem. Phys., 19, 4899-4916, https://doi.org/10.5194/acp-19-4899-2019, 2019b.

Zhu, S., Li, L., Wang, S., Li, M., Liu, Y., Lu, X., Chen, H., Wang, L., Chen, J., Zhou, Z., Yang, X., and Wang, X.: Development

440   of an automatic linear calibration method for high-resolution single-particle mass spectrometry: improved chemical species identification for atmospheric aerosols, Atmos. Meas. Tech., 13, 4111-4121, https://doi.org/10.5194/amt-13-4111-2020, 2020.

GRAPHICAL ABSTRACT