Supporting information for

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

Authors: 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], 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*

*To whom correspondence should be addressed.

Correspondence to: Jinghao Zhai, Email: zhaijh@sustech.edu.cn

Xin Yang, Email: yangx@sustech.edu.cn

Contents

Text S1. The single particle mass spectrometer used in this study.

This study employed a state-of-the-art single particle aerosol mass spectrometer with high performance (HP-SPAMS, Hexin Instrument Co., Ltd.) to investigate the chemical composition of individual particles.(Li et al., 2011; Du et al., 2022) Particles were directed into an aerodynamic focusing lens and propelled at velocities dictated by their size.(Carson et al., 1995; Clemen et al., 2020) The vacuum aerodynamic diameter of each particle was then determined by calculating the time delay detected by two continuous lasers (Nd: YAG, 532 nm) oriented at right angles to each other. Particles are then sequentially ionized by laser desorption/ionization (LDI) technique. (Du et al., 2022; Peacock et al., 2017) Within the ion source region, particles triggered a pulsed desorption/ionization laser (Nd: YAG, 266 nm), generating ion fragments. The resulting ions fly under the force of accelerating potential in a TOF-MS. (Murray et al., 2013) For each particle, both positive and negative mass spectra were recorded using a bipolar time-of-flight mass spectrometer (Fig. S1). The raw *m/z* are calculated by TOF data with predetermined parameters of the TOF-MS, e.g., length of flight path and voltage of the accelerating potential.

The initial positions of ions before flight in the TOF-MS are deviated resulting from the focusing limit of AFL,(Dienes, 2003) leading to shifts in the flight path and the accelerating potential from the designed parameters (Fig. 1b), and consequently, a drift in the flight time and the corresponding *m/z*.(Wang et al., 2010) Since the deviation is uncertain for each particle, specific coefficients are required in the calibration function for individual particles.(Chen et al., 2020; Clemen et al., 2020) Moreover, studies have shown that the inhomogeneity of ionizing laser could also influence the mass measurement.(Wenzel and Prather, 2004)

The power of the desorption/ionization laser was set to ~0.6 mJ/pulse in this study. To ensure accurate size determination, the aerodynamic diameter measurements were calibrated using monodisperse polystyrene latex spheres (Nanosphere size standards) with known diameters ranging from 0.05 to 5.0 μm. To improve data quality, all acquired single particle mass spectra were subjected to a filtering process that set a minimum signal threshold of 50 mV above the baseline for each *m/z* within the range of ± (1-500) Th.

The relation between *m/z* and mass resolution (R_m) is described in the empirical equation (S1). (Du et al., 2022) The equation is used for the localization of our algorithm.

$$R_m = -0.04549(\mathrm{m}/z)^2 + 22.98\frac{m}{z} + 199.6 \quad \text{(S1)}$$

The $\boldsymbol{m/z}_\mathrm{r}$ was converted from TOF data $\boldsymbol{t}$ using equation (S2), with coefficients provided in the following table, the unit of TOF is ns.

$$\boldsymbol{m/z}_\mathrm{r} = \left(\frac{\boldsymbol{t} - t_0}{m_0}\right)^2 \quad \text{(S2)}$$

polarity	t_0	m_0
positive	-714.829468	1216.546875
negative	-714.993408	1209.078735

Text S2. Detailed description of traits in the prototype.

It is essential to choose traits to establish the prototype to ensure accurate calibration of mass spectrometer spectra. Prototype is a concept borrowed from statistic machine learning, but the mechanism of prototype in our algorithm is more like the reward defined in reinforcement learning.(Hastie et al., 2009)

For instance, in the S_r to be calibrated, if a peak is observed around 51 Th, and the calibration target is mass deviation less than 0.025 Th, it remains uncertain whether the exact value is 50.95 Th (V^+) or 51.00 Th ($C_4H_3^+$) before calibration (in atmospheric aerosols, V^+ and $C_4H_3^+$ are only choices for a reasonable peak around 51 Th). However, suppose a calibration attempt aligns it to 51.00 Th, and simultaneously, a minor peak at 52.00 Th adheres to the isotopic distribution of $C_4H_3^+$ (i.e., p(51.00 Th): p(52.05 Th) = 1: 0.43). In that case, this peak is likely corresponding to 51.00 Th ($C_4H_3^+$). Conversely, if it is calibrated to 50.95 Th and satisfies the monoisotopic characteristic of V^+, it could be identified as 50.95 (V^+). Additionally, if prominent peaks are both observed around 51 Th and 67 Th and a series of organic signals (e.g., C_mH_n signals) is absent, these peaks may be identified as V^+ (50.95 Th) and VO^+ (66.95 Th), respectively. The isotopic distributions thus serve as traits of exact spectra. Any calibration attempt that relocates the peak contrary to these distributions would be deemed incorrect. Trait matching is considered specific because a peak matched to more than one trait is in low probability. Although it is easy to determine whether a trait is not possessed/matched by a spectrum, it is hard to judge whether a trait is truly possessed (definitely positive) by a spectrum. Once the trait is matched, it is essential to assess the behavior of S_t across the entire prototype database using the value function before confirming a trait is truly positive. Therefore, the traits that matched are only considered truly positive in S_{opt}.

Consider another example spectrum containing only C^+ (12.00 Th), C_2^+ (24.00 Th), C_3^+ (36.00 Th), C_4^+ (48.00 Th), C_5^+ (60.00 Th), and $C_4H_3^+$, and these peaks are collected in the prototype database. Although this information is unknown before calibration, we can evaluate any calibration attempt by matching these *m/z*. The optimal calibration attempt, S_{opt}, would align all six peaks to their respective theoretical positions by the prototype, resulting in peaks at 12.00 Th, 24.00 Th, 36.00 Th, 48.00 Th, 60.00 Th, and 51.00 T. Any S_t calibrate these peaks to other *m/z* position is not the optimal calibration.

A special type of traits is isolate ions because the trait matching of this type is certain. For instance, if we observed the presence of a peak around 12 Th. Given the mass deviation range (± 0.1 Th at 12 Th), the only valid theoretical choice for its value would be 12.00 Th (C^+) in atmospheric aerosol samples. Consequently, any calibration attempt that relocates this peak to a position other than 12.00 Th is miscalibration. Additionally, if a small peak is observed around 13 Th, along with a strong peak at 12 Th and following the isotope distribution of carbon, the assertion is further confirmed.

Text S3. Simplified pseudocode of the algorithm.

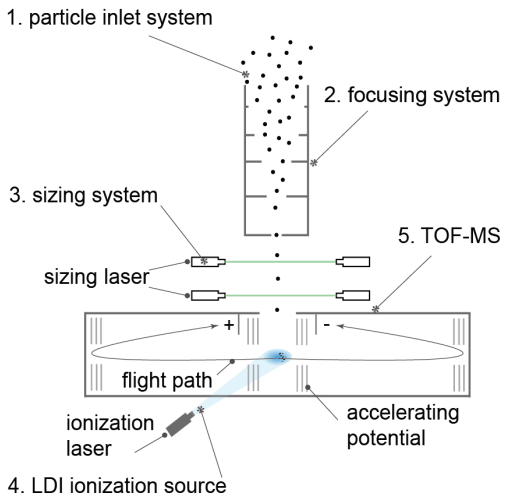
Standard-free calibration Algorithm: calibration core

Input: The raw mass spectrum *raw*, the prototype database *ptt*, the θ space *space*.

Output: The optimal calibrated spectrum *opcalied.*

```
maxValue = 0
optCoef = 0
for θ in space
    S.mz = f(raw.mz, θ)
    thisValue = 0
    for Ti in ptt
        thisValue = thisValue + Ti.wi * match(Ti, S)
    end for
    if thisValue > maxValue then
        maxValue = thisValue
        optCoef = θ
        opcalied = S
    end if
end for
return opcalied
```

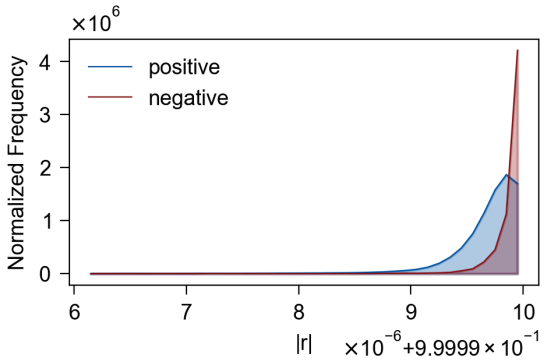
The core algorithm of the match function (equation (3)) could be a modification of binary search, or positive matrix factorization. In this study, the match function adopted a variation of binary search with a complexity of *O(nlogn)* to the number of peaks in the mass spectrum. Tolerance of mass difference was set to the resolving ability for the fuzzy matching of *m/z*. In addition, *f(raw, θ)* is the mass calibration function mentioned in equation (1), and *value* is the function in equation (2).



115

116 **Figure S1.** Simplified operational mechanism of the HP-SPAMS. Particle sequentially
117 goes through stage 1-5 to generate a raw TOF spectrum.

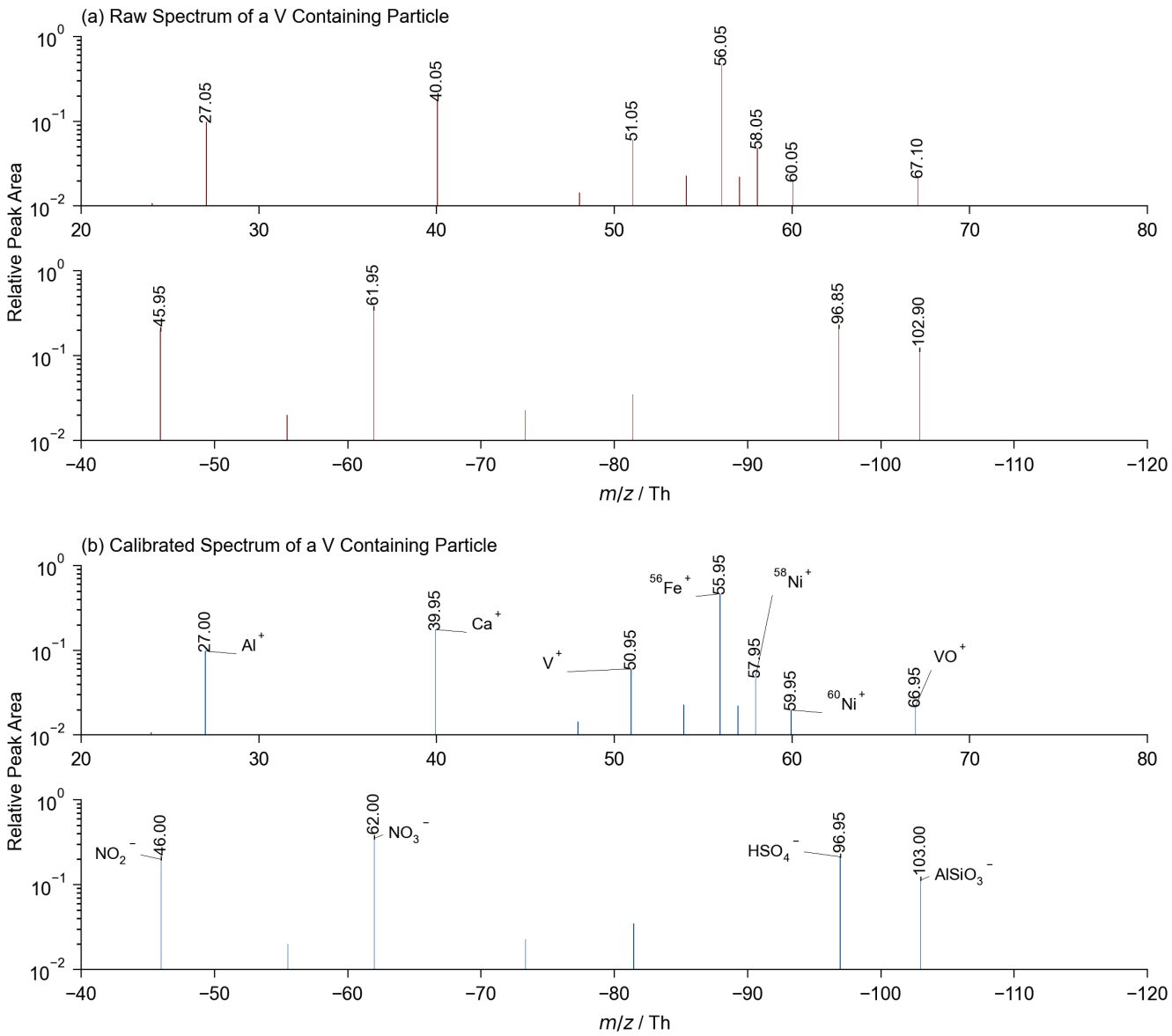
118



119

120 **Figure S2.** Histogram of $|r|$ for positive and negative spectra. The histogram of $|r|$ values
121 for positive and negative spectra demonstrates that all the calibrations result in $|r|$ values
122 greater than 0.999996. Shaded areas are summed to 1.

123



124 **Figure S3.** Example of raw (a) and calibrated (b) single-particle mass spectra of a typical

125 vanadium (V) containing particle.

126

Table S1. Accuracy and resolution required for separating easily confused ions typically encountered in aerosol research (50% separation).

Ion pair	Isotopic *m/z* (Th)	*⊿m/z* (Th)	Accuracy (ppm)	Resolution (FWHM)
Mg^+/C_2^+	23.9845/23.9995	0.0150	625.0	2400
$Al^+/C_2H_3^+$	26.9810/27.0229	0.0419	1551	967.5
$Ca^+/NaOH^+$	39.9620/39.9920	0.0300	750.7	1998
K^+/C_2NH^+	38.9631/39.0104	0.0473	1212	1237
Ti^+/ C_4^+	47.9474/ 47.9995	0.0521	1085	1382
$V^+/C_4H_3^+$	50.9434/ 51.0229	0.0795	1558	962.7
$Mn^+/C_3H_3O^+$	54.9375/55.0178	0.0803	1460	1028
$Cu^+/C5H3^+$	62.9290/63.0229	0.0939	1490	1007
Sn^+/C_{10}^+	119.9016/119.9995	0.0979	815.8	1838
SO^-/C_4^-	47.9675/48.0005	0.0330	687.5	2183
Br^-/PO_3^-	78.9189/78.9591	0.0402	509.1	2946

References

Carson, P. G., Neubauer, K. R., Johnston, M. V., and Wexler, A. S.: On-line chemical analysis of aerosols by rapid single-particle mass spectrometry, J. Aerosol Sci., 26, 535-545, https://doi.org/10.1016/0021-8502(94)00133-J, 1995.

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.

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), EGUsphere [preprint], 2022, 1-23, https://doi.org/10.5194/egusphere-2022-872, 2022.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H.: The elements of statistical learning: data mining, inference, and prediction; Springer, 2009.

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, https://doi.org/10.1016/j.ijms.2011.01.017, 2011.

Murray, K. K., Boyd, R. K., Eberlin, M. N., Langley, G. J., Li, L., and Naito, Y.: Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013), Pure Appl. Chem., 85, 1515-1609, https://doi.org/10.1351/pac-rec-06-04-06, 2013.

Peacock, P. M., Zhang, W. J., and Trimpin, S.: Advances in ionization for mass spectrometry, Anal. Chem., 89, 372-388, https://doi.org/10.1021/acs.analchem.6b04348, 2017.

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.

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.