Supplement

A Novel Approach for Simple Statistical Analysis of High-Resolution Mass Spectra

Yanjun Zhang[1], Otso Peräkylä[1], Chao Yan[1], Liine Heikkinen[1], Mikko Äijälä[1], Kaspar R. Daellenbach[1], Qiaozhi Zha[1], Matthieu Riva[1,2], Olga Garmash[1], Heikki Junninen[1,3], Pentti Paatero[1], Douglas Worsnop[1,4], and Mikael Ehn[1]

[1] Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Helsinki, 00140, Finland

[2] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, IRCELYON, F-69626, Villeurbanne, France

[3] Institute of Physics, University of Tartu, Tartu, 50090, Estonia

[4] Aerodyne Research, Inc., Billerica, MA 01821, USA

First author: Yanjun Zhang & Otso Peräkylä

Correspondence to: Yanjun Zhang (yanjun.zhang@helsinki.fi) & Chao Yan (chao.yan@helsinki.fi)

Figures

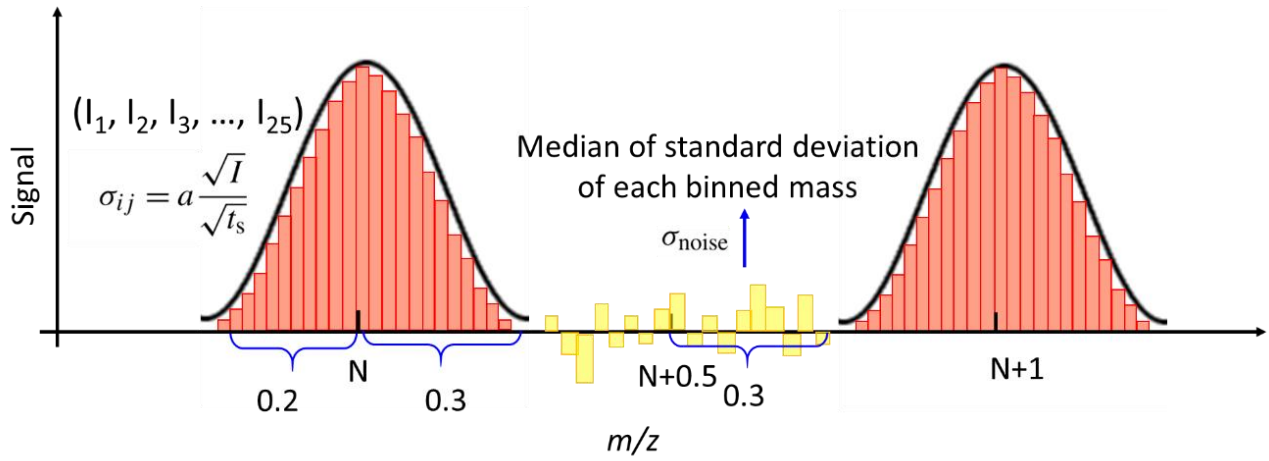


Figure S1. Schematic diagram of data matrix preparation and error estimation for binPMF. binPMF input, data matrix was constructed with the bins in the signal region (the red bars). The error matrix contained two parts and was estimated based on bins in both the signal region (red bars) and noise region (yellow bars).

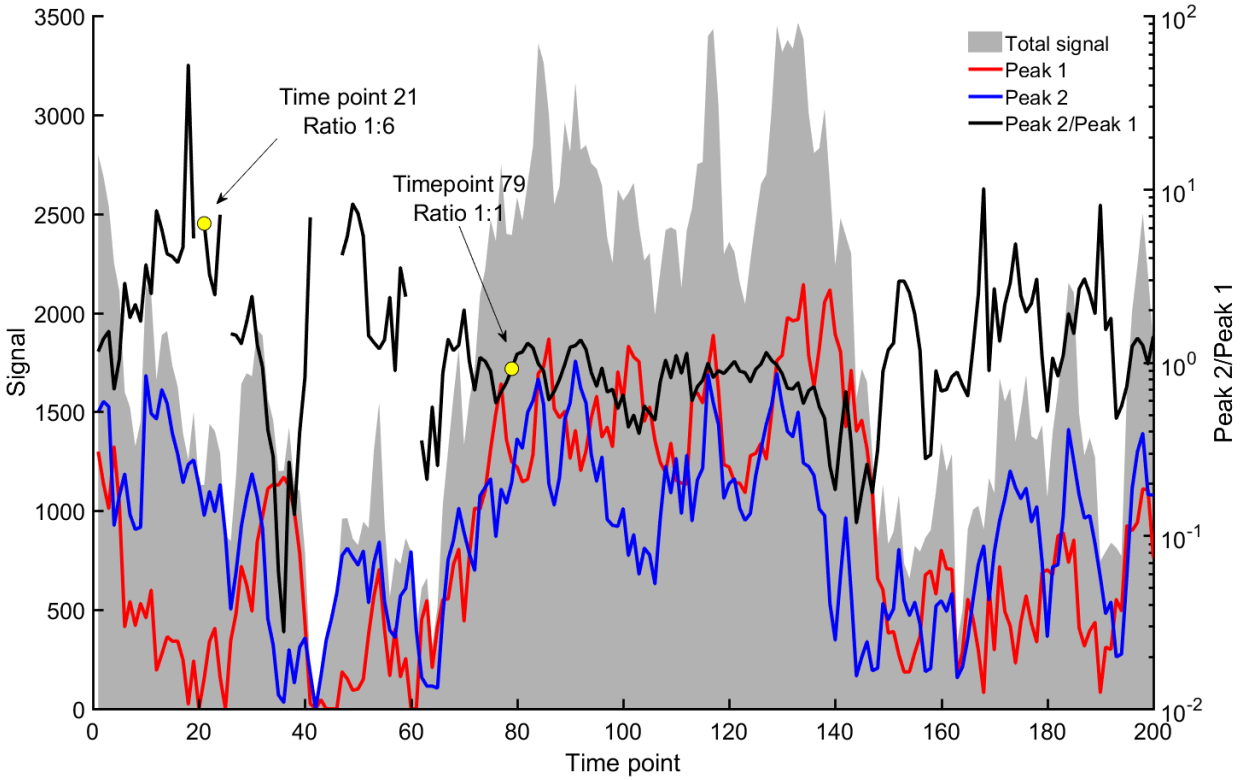


Figure S2. Time series of the two sources in the synthetic datasets. Source signals are shown in the left Y axis, while the source ratio is on the right Y axis. The signal intensity ratios of Source A and

Source B were approximately 1:1 and 1:6, respectively, at the 79th and the 21th time point, which were used for peak the fitting comparison in section 3.1.2.

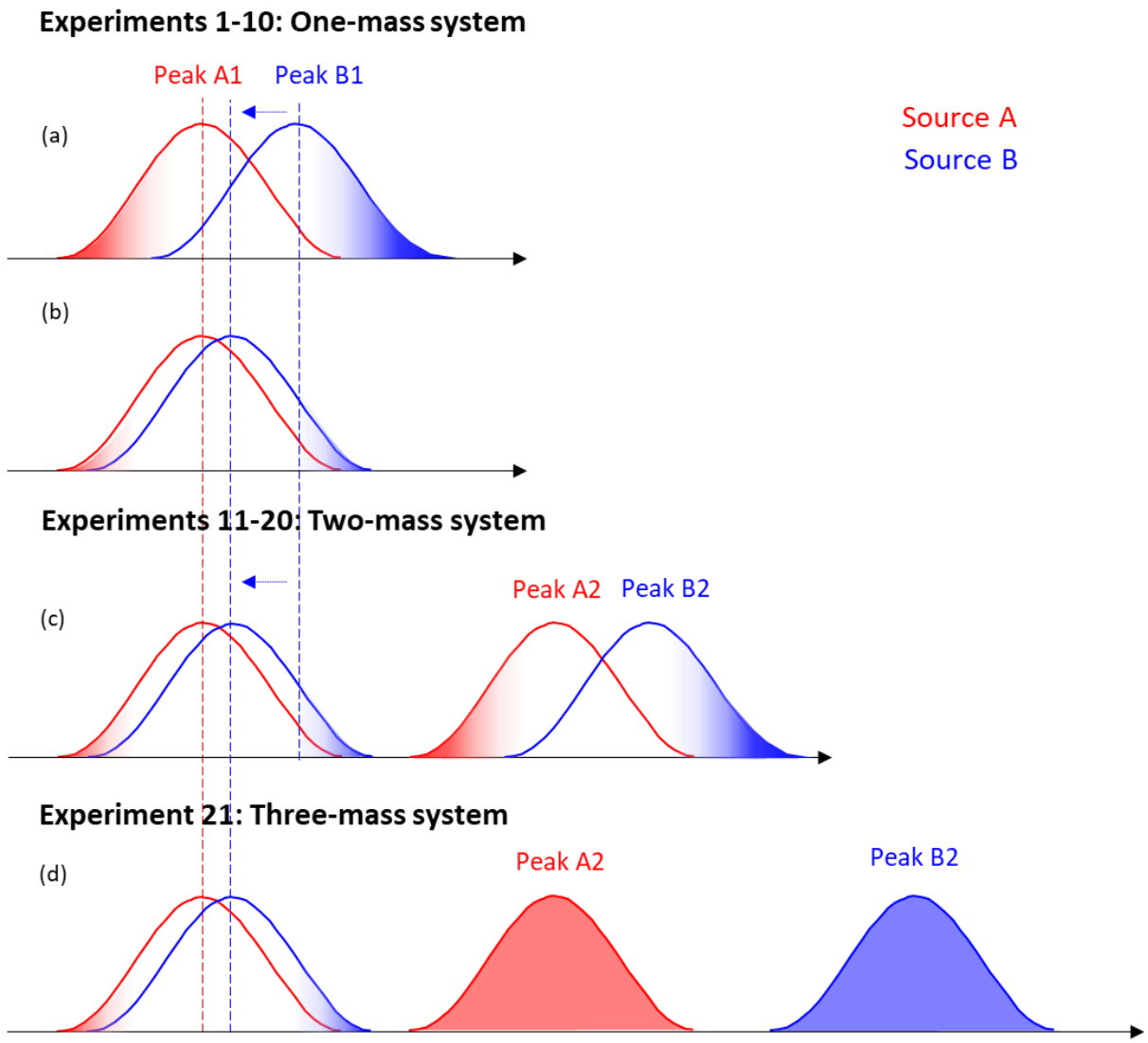
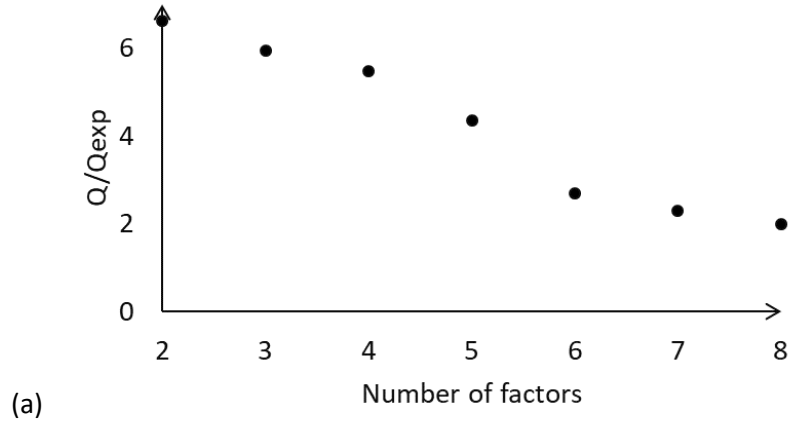
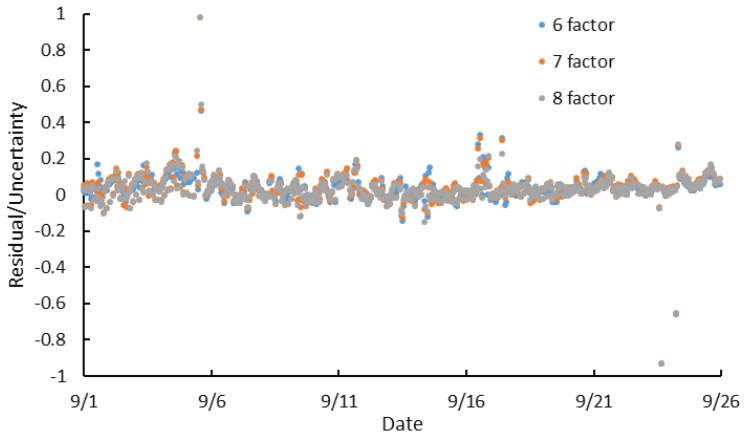


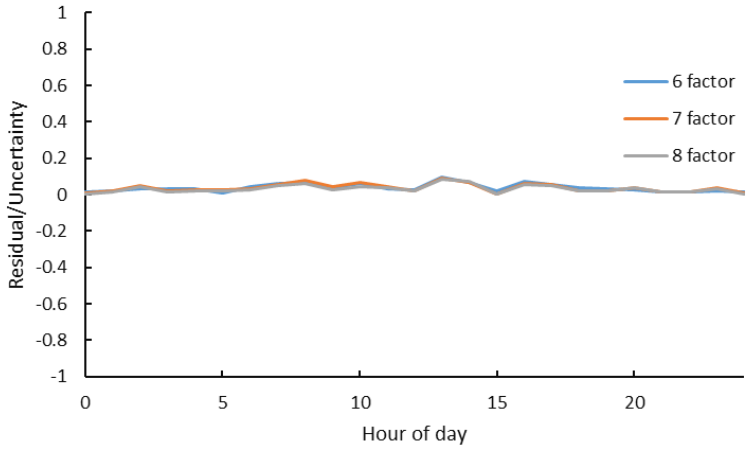
Figure S3. Conceptual illustration of binPMF. Red and blue shaded areas represent signals which are dominated by one factor only, which binPMF can efficiently utilize to identify and separate the two factors.



(a)



(b)



(c)

Figure S4. PMF diagnostic results for the ambient dataset. Q/Q_{exp} (panel a) as a function of the number of factors, time series of median scaled residual (b), and diurnal trend of median scaled residual (c).

Table S1. Detailed settings of the synthetic datasets. Experiments 1-10 are a one-mass system, experiments 11-20 a two-mass system, and experiment 21 a three-mass system.

Experiment	Centroid of peak A1	Centroid of peak B1	*m/z* difference		No. of masses	Q/Q_{exp}
	Th	Th	Th	ppm		
1	310.012	310.062	0.050	161	1	2.49
2	310.012	310.052	0.040	129	1	1.69
3	310.012	310.042	0.030	97	1	1.35
4	310.012	310.032	0.020	65	1	1.04
5	310.012	310.022	0.010	32	1	0.89
6	310.012	310.017	0.005	16	1	0.83
7	310.012	310.016	0.004	13	1	0.82
8	310.012	310.015	0.003	10	1	0.81
9	310.012	310.014	0.002	6	1	0.81
10	310.012	310.013	0.001	3	1	0.81
11	310.012	310.062	0.050	161	2	2.61
12	310.012	310.052	0.040	129	2	2.44
13	310.012	310.042	0.030	97	2	2.54
14	310.012	310.032	0.020	65	2	3.04
15	310.012	310.022	0.010	32	2	3.66
16	310.012	310.017	0.005	16	2	3.92
17	310.012	310.016	0.004	13	2	3.97
18	310.012	310.015	0.003	10	2	4.02
19	310.012	310.014	0.002	6	2	4.06
20	310.012	310.013	0.001	3	2	4.11
21	310.012	310.013	0.001	3	3	4.85