Supplementary Material

Molecular signals of heterogeneous terrestrial environments identified in dissolved organic matter: A comparative analysis of Orbitrap and ion cyclotron resonance mass spectrometers

Carsten Simon, Vanessa-Nina Roth, Thorsten Dittmar, Gerd Gleixner*

*** Correspondence:** Gerd Gleixner: gerd.gleixner@bgc-jena.mpg.de

Contents

Supplementary Table 1. Comparative features of both analyzers (compiled from Marshall and Hendrickson, 2008; Perry et al., 2008; Zubarev and Makarov, 2013).

	FT-ICR MS instruments	**Orbitrap instruments**
Size of the instrument	High due to superconducting magnet	Theoretically small, benchtop and portable instruments possible
Flexibility of the analyzer	High due to storage of trapped ions in the ICR cell	Low, no storage of ions; depending mainly on front end
Costs	Very high instrumental costs High maintenance costs	High instrumental costs Low maintenance costs
Instrument used in this study	SolariX XR (introduced 2009; mod. 2013; Bruker Daltonik GmbH, Bremen, Germany)	Orbitrap Elite (introduced 2011; Thermo Fisher Scientific, Bremen, Germany & Waltham, USA)
Resolution	~ 1.000.000, up to 2.000.000 possible with modified XR cell This study: ~ 600.000 @ 400 m/z	~ 150.000; Up to 480.000 possible with beta software; This study: ~ 200.000 @ 400 m/z
Accuracy in mass determination	< 1 ppm, < 0.1 ppm possible with modified XR cell sub-ppm also by internal calibration	1 – 5 ppm < 2 ppm with internal calibration
Stability of mass accuracy	Stable for months	Stable for days to weeks
Field applied for trapping	Magnetic	Electrostatic
Ion motion used for detection	Cyclotron rotational frequency	Axial oscillation frequency
Frequency link to m/z	$f \propto \frac{1}{m/z}$	$f \propto \frac{1}{\sqrt{m/z}}$

Supplemental Table 2. Compilation of Orbitrap/ FT-ICR MS comparison studies for different natural organic matter (NOM) samples taken from [1] Cortés-Francisco et al., 2011; [2] Pomerantz et al., 2011; [3] Remucal et al., 2012; [4] Smith et al., 2012; [5] Zhurov et al., 2013; [6] Mangal et al., 2016; [7] Hawkes et al., 2016; [8] This study; n.s., not stated; * improved analyzer & high transient length.

Ref.	Type of sample(s)	Instrument, ESI mode, m/z range	Evaluation parameters	Sample information
[1]	Standard mix of four ions and IHSS samples	LTQ Orbitrap/ Orbitrap Exactive, n.s., n.s.	Mass accuracy (drift) and precision, resolving power	None
[2]	Petroleum	LTQ Orbitrap XL, negative mode, 150 – 1050	Mass resolution, mass accuracy	Heteroatom class distribution, DBE distribution, cyclic/ acyclic ratio, PM degradation scale
[3]	IHSS samples, SRFA	Orbitrap Exactive, negative mode, (100) 290 – 600	Mass distribution, shared formulae	None
[4]	Fraction of a bio-oil produced by fast pyrolysis	LTQ Orbitrap Discovery, negative mode, 100 – 400	Mass distribution, mass resolution	Heteroatom class distribution, DBE distribution
[5]	Resin and maltene fraction of 2 crude oils	Orbitrap Elite*, positive mode, 200 – 1000	Resolving power, mass accuracy, spectral dynamic range	Kendrick mass defect analysis, heteroatom class subsets, heteroatom class distribution
[6]	IHSS samples, SRFA and PLFA	Orbitrap Q Exactive, negative mode, 200 – 1000	Mass distribution	Kendrick mass defect analysis, average DBE, average elemental composition
[7]	Mixture of a dystrophic lake sample and a marine sample	LTQ Velos Pro Orbitrap, negative mode, 150 – 2000	General performance, mass distribution, differentiation of composition changes	Average H/C and m/z values, critical mass differences, Bray-Curtis dissimilarities
[8]	Set of 17 SPE-DOM samples from a wide range of ecosystems	Orbitrap Elite*, negative mode, 200 – 650	Mass distribution, sample-specific loss of information, causes of information loss, reproducibility/ retrieval of biogeochemical trends	Ion abundance patterns, Van Krevelen patterns, critical mass differences, number of formulae and molecular group contribution, trends of molecular indices (DBE, NOSC, H/C, etc)

Supplemental Table 3. Instrument settings of the two FTMS methods. Orbitrap-specific setting of Automatic Gain Control™ and S-Lens RF level were adjusted at 1xE6 and 70%.

Conditions	**FT-ICR MS**	**Orbitrap**
Ionization mode	ESI Negative	ESI Negative
Flow [µl*min^{-1}]	4	7
Accumulation time [ms]	100	Max. 100 [1]
DOC [ppm]	10	20
Scan range [m/z]	115 – 2000	115 – 2000
Scans [n]	500	300
Source/ Capillary Temp. [°C]	200	275
Source fragmentation [eV]	40	40
Spray voltage [kV]	4.5	2.65
Transient length [s]	2.1	0.8

[1] Due to Automatic Gain Control (AGC).

Supplemental Table 4. Effect of the variation of a single factor (columns) on the different responses (with increasing factor level). Arrows show either a positive or a negative relation, no effect is marked by "o". Colors correspond to the ideal conditions (e.g., high percentage of assigned peaks, small variation coefficient): Green marks *positive*, yellow *negative*, and blue *neutral* evaluation. Source fragmentation, inject time and sheath gas were the most influential factors but not checked for interactions. Variation coefficient of triples was always very low but was increased substantially by use of source fragmentation and high inject times.

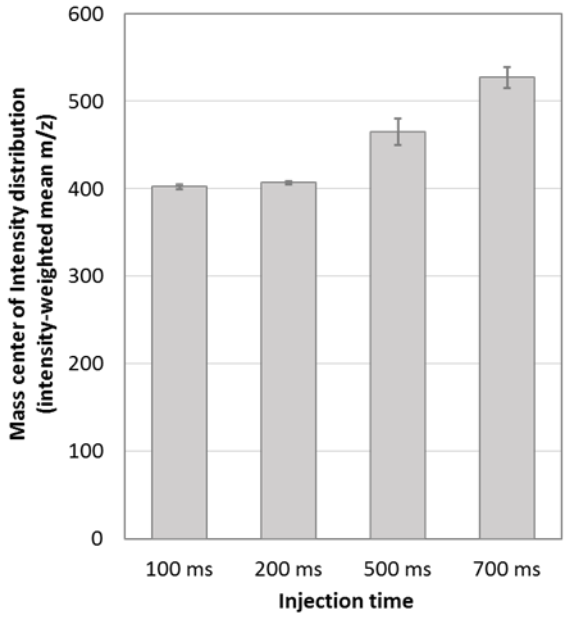
	DOC	**Flow rate**	**Source fragm.**	**S-Lens**	**Spray voltage**	**Sheath gas**	**Capillary temp.**	**Inject time**
Unit	ppm	µl*min^{-1}	eV	%	kV	a.u.	°C	ms
Levels	10, 20, 30	5, 7, 10	0, 25, 50	30, 50, 70	2.5, 3.5, 4.5	0, 10, 20	225, 275, 325	100, 200, 500, 700
Number of peaks	↑	↑	o	o	↓	o	↑	o
% assigned formulae	o	o	↑↑	o	o	↑	o	↓↓
Similarity of data with ICR method (Pearson's r)	o	↓	↑↑	↑	o	↑↑	o	↓↓
% triple detected formulae in triplicates	o	o	↓	o	↑	↑	↓↓	↓↓
Variation coefficient of triples	o	↓	↑↑	↑	↑	o	o	↑↑

Supplemental Table 5. FT-ICR MS metadata of the full dataset. Sample names: refer to Table 3. WA, weighted average. Indices are explained in Figure 5 of the main text and Supplemental Figure 5. CHONSP classes add up to an own scale.

	TAP	H5-3a	H3-2b	SAALE	IHSS	N8R	H2S-5	T-5	W1-5	JE-10	JE-20	JE-30	JE-60	N3B	N8B	BZWA	NELHA
Formulae	4754	3768	4643	5115	3607	4279	4844	4574	3015	5742	6643	6602	5985	4431	3724	6317	4710
m/z_{WA}	382.9	358.5	384.6	345.8	346.5	370.4	364.6	368.7	352.0	344.1	357.7	368.1	370.7	363.0	364.7	372.4	416.1
C_{WA}	19.28	17.36	19.28	17.43	17.31	18.08	17.82	17.88	17.07	17.39	17.96	18.41	18.67	17.74	17.77	17.99	20.11
H_{WA}	24.04	22.02	23.70	21.30	18.16	21.19	19.25	18.58	16.33	19.13	20.25	21.66	22.06	18.76	19.33	19.18	25.76
O_{WA}	7.78	7.85	7.75	6.80	7.38	8.15	7.95	8.31	8.07	6.74	7.09	7.37	7.28	8.09	8.19	8.33	8.91
N_{WA}	0.19	0.18	0.38	0.25	0.11	0.11	0.32	0.18	0.13	0.50	0.49	0.42	0.46	0.11	0.07	0.26	0.37
S_{WA}	0.03	0.02	0.03	0.12	0.05	0.03	0.02	0.03	0.02	0.07	0.07	0.07	0.07	0.04	0.02	0.04	0.04
P_{WA}	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02
DBE_{WA}	8.36	7.44	8.62	7.91	9.29	8.54	9.35	9.68	9.98	9.08	9.08	8.80	8.87	9.41	9.14	9.53	8.43
$AI_{MOD,WA}$	0.28	0.25	0.29	0.30	0.40	0.31	0.38	0.40	0.45	0.39	0.36	0.33	0.33	0.39	0.37	0.38	0.24
$NOSC_{WA}$	-0.40	-0.33	-0.36	-0.37	-0.15	-0.24	-0.10	-0.05	0.03	-0.20	-0.22	-0.28	-0.30	-0.10	-0.13	-0.09	-0.33
H/C	1.25	1.27	1.23	1.22	1.05	1.17	1.08	1.04	0.96	1.10	1.13	1.18	1.18	1.06	1.09	1.07	1.28
O/C	0.40	0.45	0.40	0.39	0.43	0.45	0.45	0.47	0.47	0.39	0.39	0.40	0.39	0.46	0.46	0.46	0.44
DBE/C	0.43	0.43	0.45	0.45	0.54	0.47	0.52	0.54	0.58	0.52	0.51	0.48	0.48	0.53	0.51	0.53	0.42
DBE-O	0.58	-0.41	0.87	1.11	1.91	0.39	1.40	1.36	1.91	2.34	1.99	1.43	1.59	1.33	0.95	1.20	-0.49
$BC_{\%}$	0.8	0.5	0.7	2.2	5.9	3.6	3.6	4.7	7.3	3.8	3.3	2.5	1.9	5.8	5.8	4.9	0.5
$PP_{\%}$	13.1	11.6	13.7	16.7	23.0	17.2	18.8	20.3	23.3	20.2	19.1	17.7	17.2	20.0	18.8	20.0	9.1
$HU_{\%}$	72.6	72.6	73.5	61.3	54.0	62.0	60.2	57.6	52.5	55.5	59.1	62.8	65.5	55.2	57.2	60.1	74.5
$UA_{\%}$	9.2	9.7	7.5	11.3	5.8	10.4	6.4	6.4	6.3	8.1	6.9	7.1	7.2	8.5	9.9	5.4	11.2
$SUG_{\%}$	0.0	0.0	0.0	0.0	0.4	0.1	0.3	0.4	0.6	0.1	0.1	0.0	0.0	0.3	0.5	0.2	0.0
$PEP_{\%}$	2.1	4.3	2.4	2.4	0.2	0.7	0.9	0.3	0.2	1.1	1.2	1.5	1.5	0.4	0.3	0.8	3.1
$CHO_{\%}$	44.2	38.7	40.6	44.3	61.3	58.4	51.0	57.0	68.9	41.2	38.3	36.7	37.3	61.5	69.4	44.4	39.0
$CHNO_{\%}$	39.5	43.2	48.0	35.0	24.3	27.2	41.6	32.8	25.3	43.7	44.5	44.7	45.1	24.8	18.8	40.2	45.1
$CHOS_{\%}$	11.0	9.2	7.1	17.5	11.6	9.9	5.3	7.0	2.1	12.3	13.5	13.8	12.7	9.5	7.3	10.7	8.6
$CHOP_{\%}$	3.1	4.4	2.0	1.0	1.2	2.4	0.4	0.4	0.3	0.7	0.9	1.3	1.5	1.1	1.7	0.6	4.5

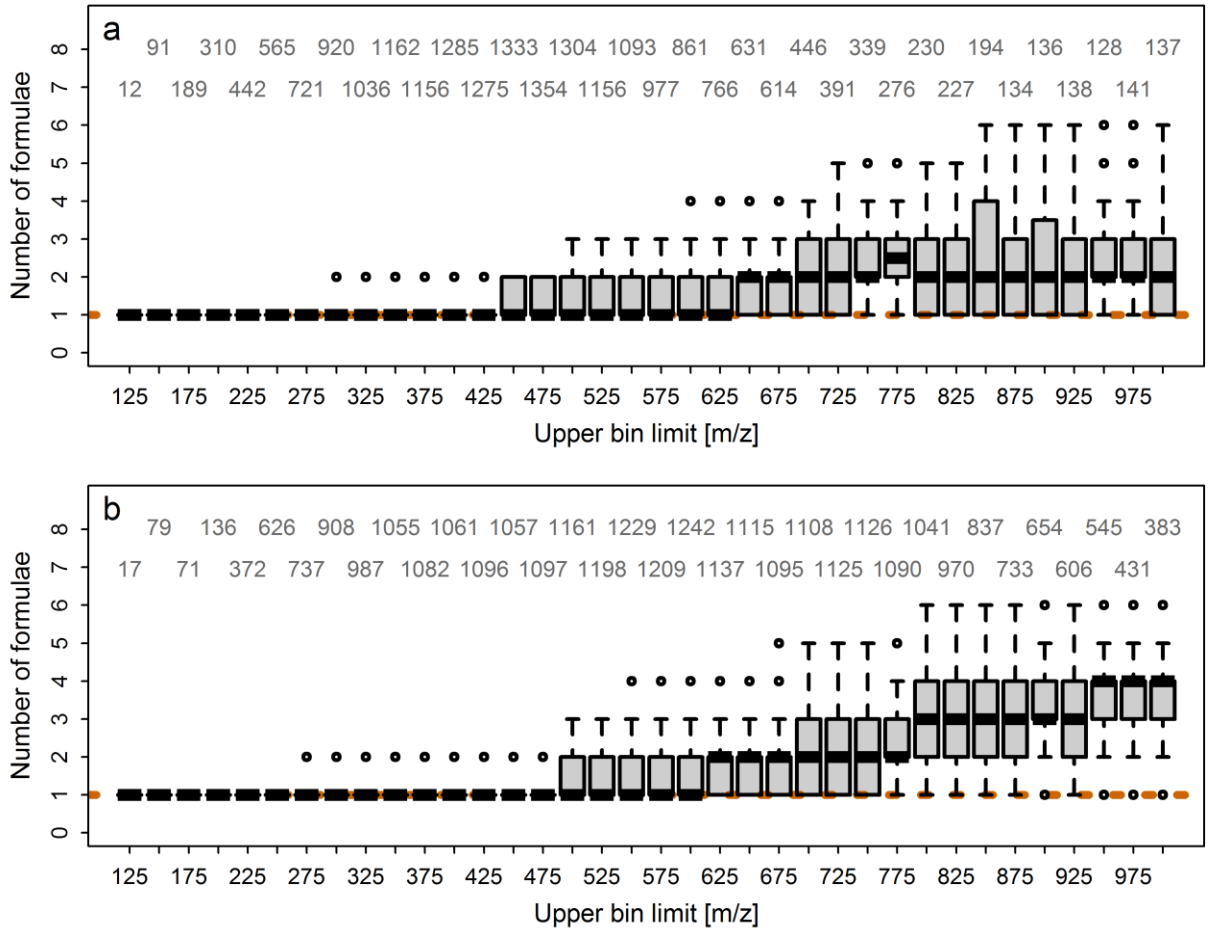
Supplemental Table 6. Orbitrap metadata of the full dataset. Details similar as described in Supplemental Table 6.

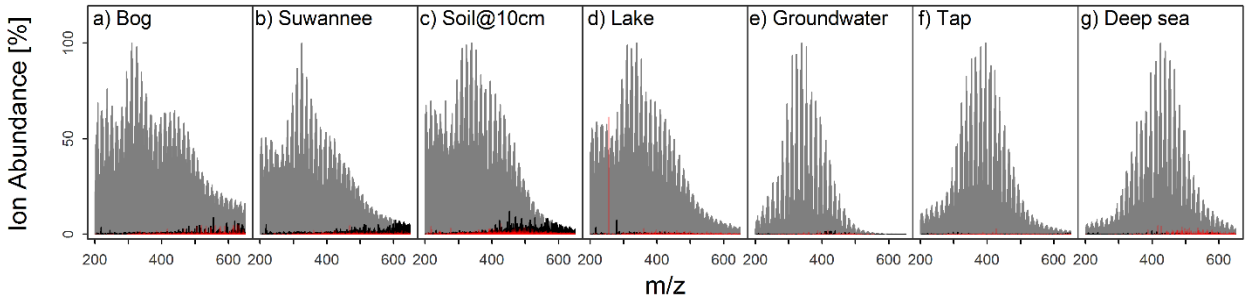
ORBITRAP	TAP	H5-3a	H3-2b	SAALE	IHSS	N8R	H2S-5	T-5	W1-5	JE-10	JE-20	JE-30	JE-60	N3B	N8B	BZWA	NELHA
Formulae	4987	3890	5123	6001	4818	5251	5506	5234	4146	5932	6053	5949	5821	5003	4861	5734	4381
m/z_{WA}	391.1	353.5	398.2	355.3	375.3	386.6	392.4	390.2	385.0	363.5	367.9	386.3	397.7	387.1	390.6	368.0	428.8
C_{WA}	19.38	17.17	19.71	17.21	17.97	18.45	18.66	18.63	18.30	17.70	18.25	18.73	19.22	18.29	18.44	18.22	20.45
H_{WA}	23.90	21.79	24.08	20.32	18.27	21.11	20.16	19.52	17.63	18.81	20.77	21.94	22.51	18.83	19.43	19.98	25.92
O_{WA}	8.35	7.79	8.46	7.72	8.82	8.98	9.15	9.16	9.22	8.00	7.75	8.48	8.77	9.30	9.39	7.99	9.63
N_{WA}	0.09	0.12	0.19	0.15	0.06	0.04	0.18	0.10	0.07	0.28	0.28	0.22	0.24	0.05	0.02	0.14	0.22
S_{WA}	0.02	0.01	0.01	0.12	0.01	0.02	0.01	0.00	0.00	0.04	0.03	0.05	0.05	0.01	0.01	0.01	0.03
P_{WA}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
DBE_{WA}	8.48	7.33	8.76	8.12	9.86	8.92	9.67	9.92	10.52	9.44	9.01	8.87	9.09	9.90	9.74	9.30	8.60
$AI_{MOD,WA}$	0.28	0.25	0.28	0.30	0.40	0.32	0.36	0.38	0.43	0.39	0.35	0.31	0.30	0.39	0.37	0.37	0.23
$NOSC_{WA}$	-0.35	-0.34	-0.33	-0.24	-0.02	-0.16	-0.05	-0.04	0.06	-0.09	-0.22	-0.21	-0.21	0.00	-0.03	-0.18	-0.29
H/C	1.23	1.27	1.22	1.18	1.02	1.14	1.08	1.05	0.96	1.06	1.14	1.17	1.17	1.03	1.05	1.10	1.27
O/C	0.43	0.45	0.43	0.45	0.49	0.49	0.49	0.49	0.50	0.45	0.42	0.45	0.46	0.51	0.51	0.44	0.47
DBE/C	0.44	0.43	0.44	0.47	0.55	0.48	0.52	0.53	0.57	0.53	0.49	0.47	0.47	0.54	0.53	0.51	0.42
DBE-O	0.12	-0.46	0.30	0.41	1.05	-0.07	0.52	0.75	1.30	1.44	1.26	0.39	1.32	0.60	0.35	1.31	-1.03
$BC_{\%}$	2.2	2.0	2.1	4.6	6.5	5.9	5.6	5.9	7.2	5.3	4.8	4.7	4.5	6.3	6.5	5.4	2.2
$PP_{\%}$	15.4	14.1	15.5	18.9	22.6	18.8	19.5	20.4	22.1	20.1	19.3	18.7	18.2	20.2	19.5	19.9	13.3
$HU_{\%}$	68.3	69.1	68.8	59.6	58.6	60.0	60.5	59.9	58.6	59.4	60.7	61.7	63.1	58.1	58.2	59.0	70.0
$UA_{\%}$	7.4	8.5	6.3	7.9	3.8	7.5	5.7	5.5	4.8	5.9	5.7	5.8	5.6	7.3	8.2	6.5	8.2
$SUG_{\%}$	0.0	0.0	0.0	0.1	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.1
$PEP_{\%}$	2.8	3.9	2.9	2.1	0.2	1.2	1.1	0.5	0.1	1.2	1.7	1.6	1.7	0.5	0.6	1.1	3.2
$CHO_{\%}$	48.6	46.6	46.0	48.5	55.9	53.8	51.2	54.6	60.6	48.0	47.3	46.4	45.9	56.1	58.0	51.0	45.6
$CHNO_{\%}$	40.6	41.9	44.7	39.2	33.8	34.7	39.4	36.1	32.6	39.7	40.6	41.6	42.3	32.3	29.8	37.4	43.7
$CHOS_{\%}$	7.9	7.9	6.6	9.7	9.2	9.1	8.0	7.8	6.0	9.7	9.5	9.4	9.2	9.9	10.0	9.3	7.8
$CHOP_{\%}$	0.9	1.4	0.7	0.2	0.2	1.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.6	1.3	0.2	0.9



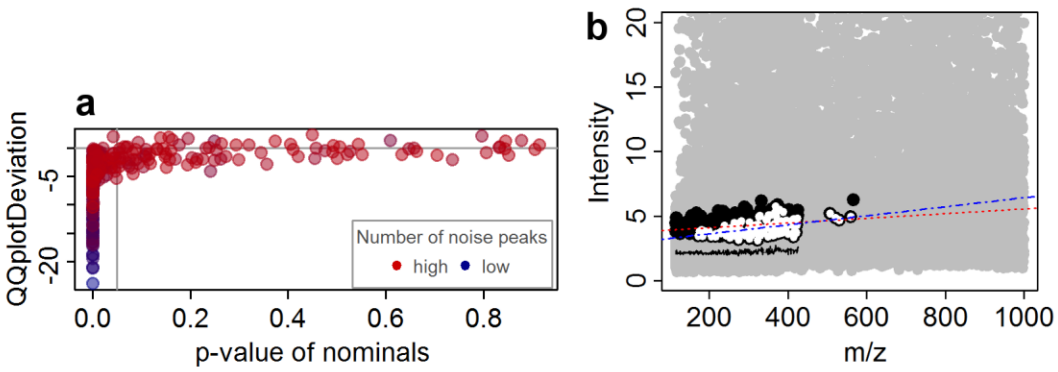
Supplemental Figure 1. Inject time did control the average mass center of the IHSS ion abundance distribution, showing a distinct shift to higher m/z values with increasing inject time (Cao et al., 2016; Hawkes et al., 2016). Error bars are ± standard deviation of triplicate measurements.

Supplemental Figure 2. Initial distributions of the number of calculated formulae within the elemental limits $C_{1\text{-}60}\,H_{4\text{-}210}\,O_{1\text{-}60}\,N_{0\text{-}4}\,S_{0\text{-}2}\,P_{0\text{-}1}$, $O/C_{max} = 1$, $H/C_{min} = 0.3$, $DBE_{min} = -0.5$ and mass error < 0.5 ppm, shown here for bins of 25 Da width spanning from m/z 115 – 1000 of a) FT-ICR MS and b) Orbitrap data. Grey numbers over each boxplot denote the number of peaks (with at least one assigned formula) contributing to each distribution. Boxplots were constructed from the following five values: The minimum (lower whisker), first quartile (lower hinge of the box), median (black line), third quartile (upper hinge), and the maximum (upper whisker). Single dots denote outliers of the distribution and the orange line is indicating unambiguous hits (only one potential formula found).

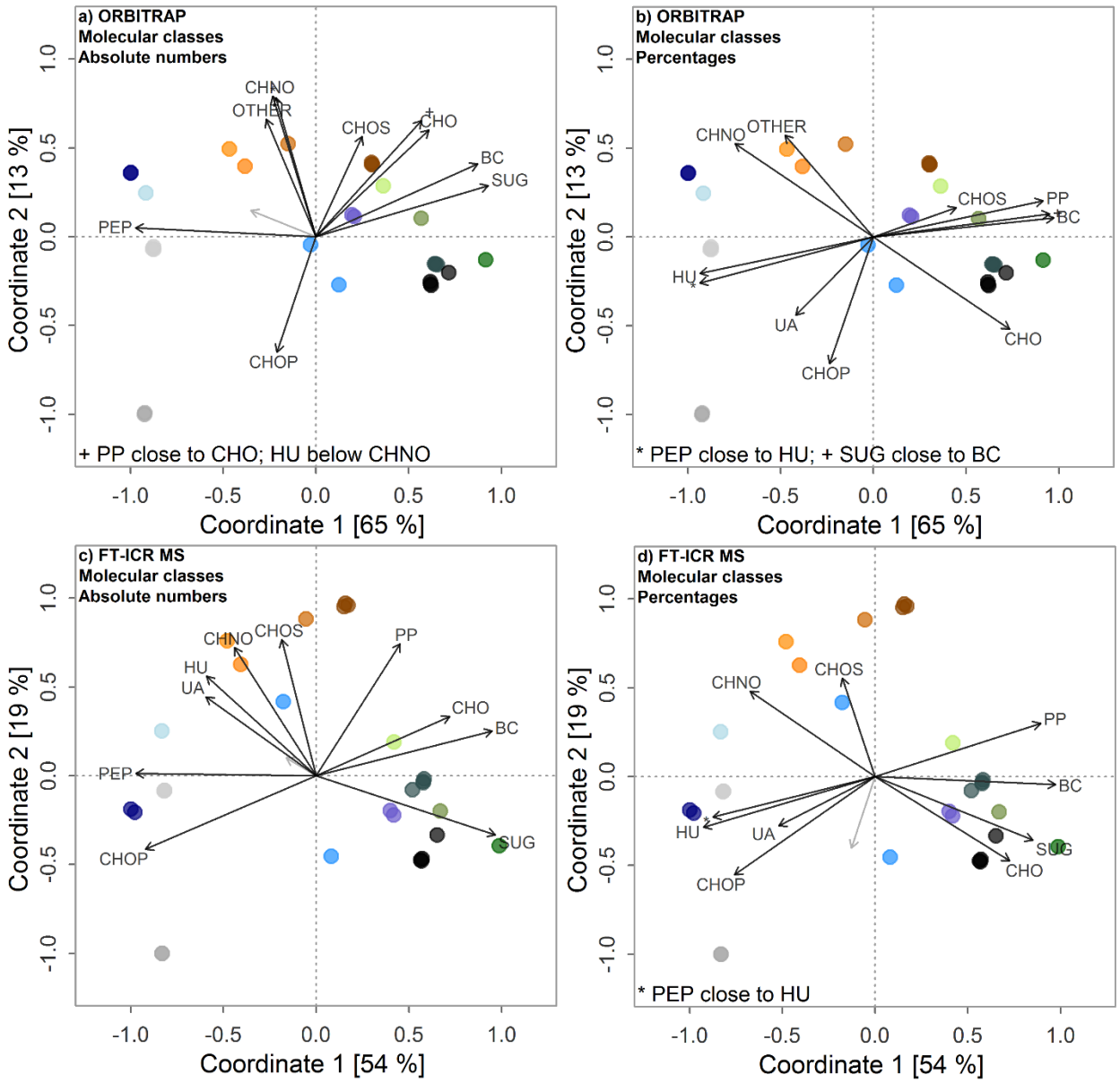




Supplemental Figure 3. Distribution of Orbitrap signals (formulae) for the samples of the subset used for detailed analyses (compare Fig.1 and Fig. 2 in main text). Shown are specific signals (red, specific s.s.), signals also encountered in the FT-ICR MS dataset but not in same samples (black, specific s.l.), and common signals (both datasets, and exact same sample (grey). It is obvious that common signals are by far the most prominent fraction for all sample types. In the higher mass range, unavoidable differences in tuning led to slight deviations in apparent sample composition (black signals with higher intensities, especially in a, b, c). The effect was most pronounced for samples with ion abundance maxima in the lower mass range and influenced multivariate separation of samples, too (e.g., leading to an offset between mean average m/z as derived from FT-ICR MS and Orbitrap data, Supplementary Figure 8; see also main text).

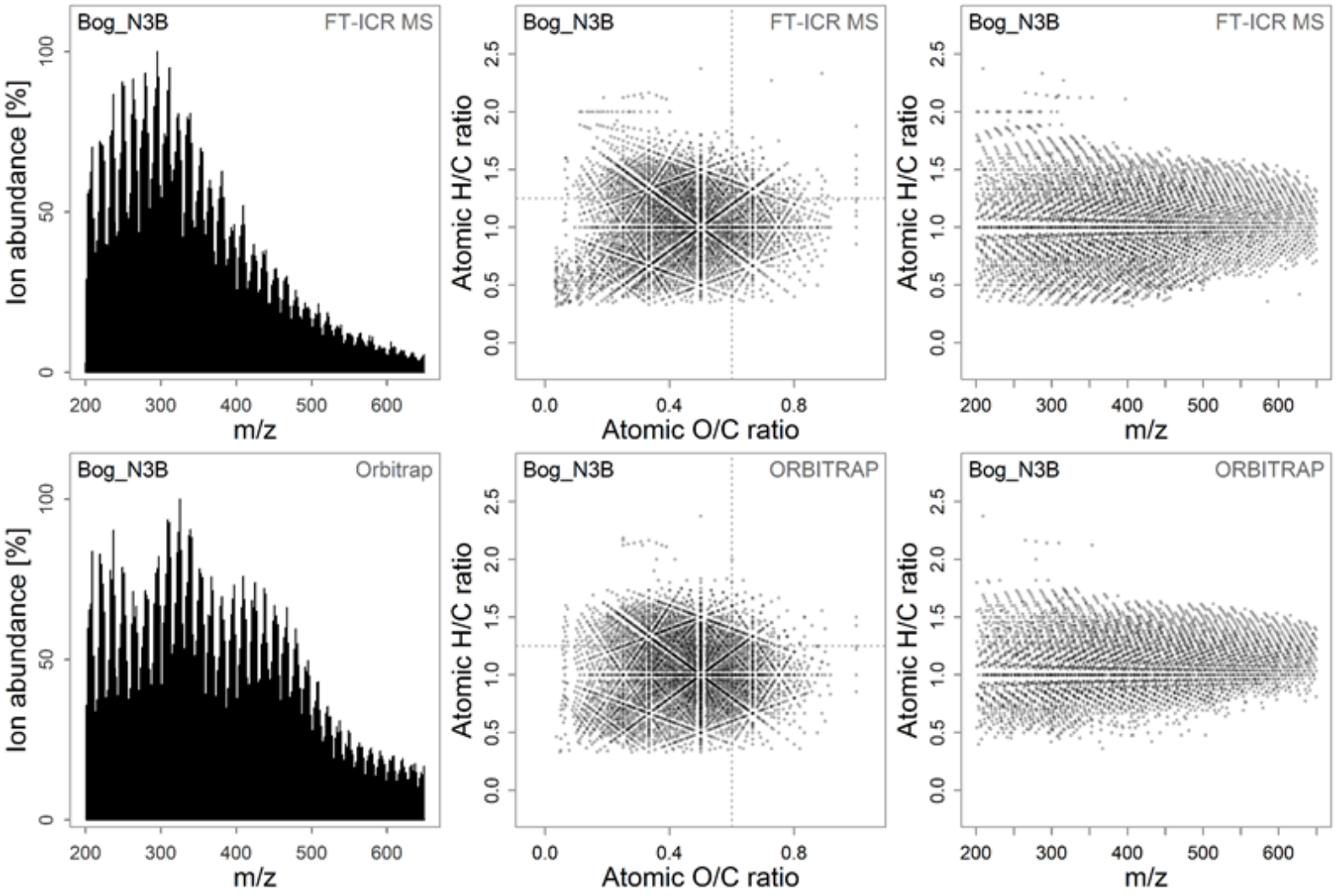


Supplemental Figure 4. Application of a method detection limit to Orbitrap data (see also Supplemental note 1). a) Normality assessment of nominal mass noise peak ensembles (p-values > 0.05 in a t-test, QQ-plot deviation ~ 0) being the basis of MDL calculation, and b) MDL approximation of the respective (NELHA) sample (black dots: all nominal masses with a peak number > 20; white dots: normally distributed peak ensembles from a). White dots were used for approximation of a linear regression model (blue line; red line is based on all nominal mass ensembles), showing a general increase of MDL-values over m/z. The black line is mean noise without added standard deviation, grey background represents the raw peak data.

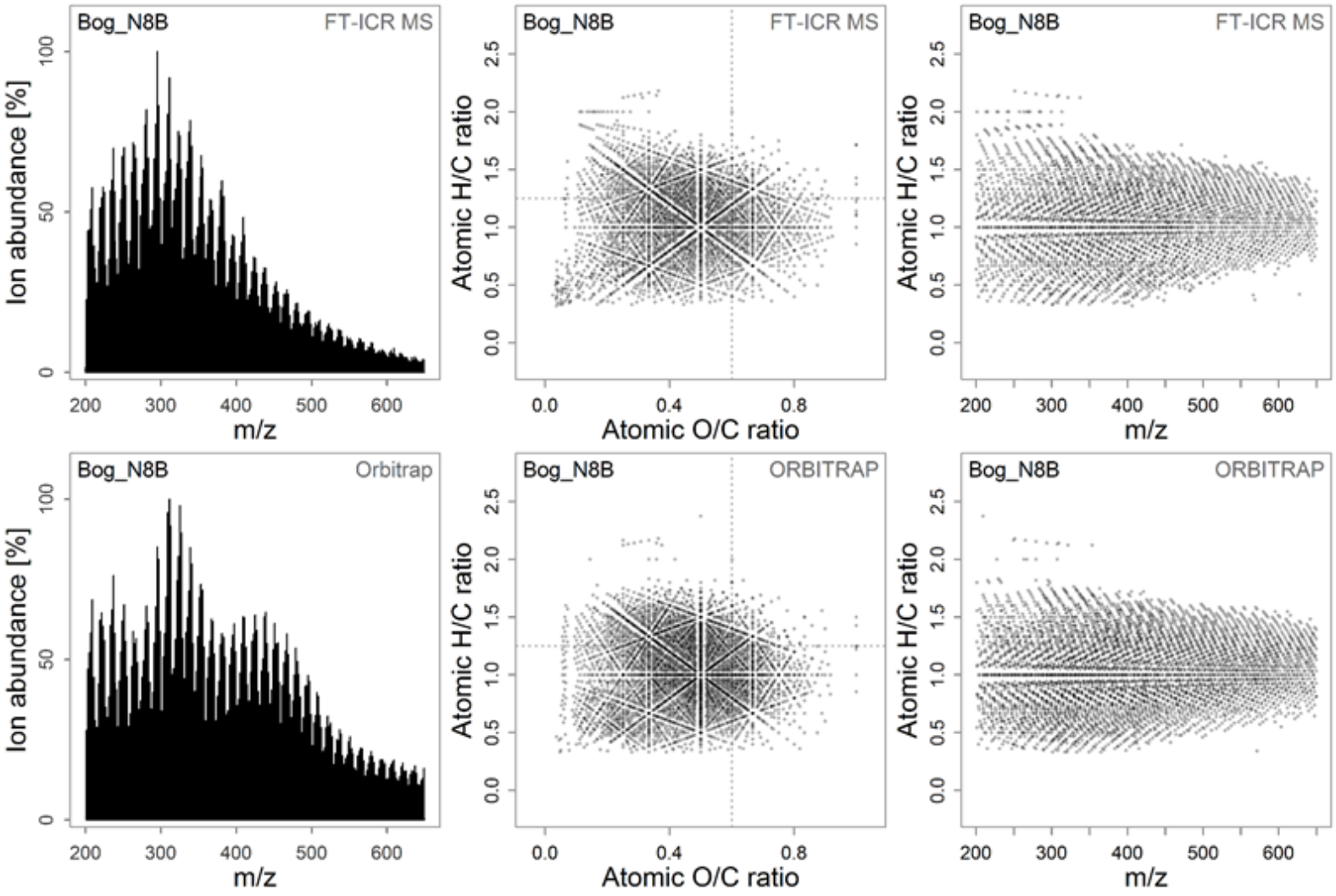


Supplemental Figure 5. Post-gradient fitting on ordination obtained by Orbitrap (upper panels) and FT-ICR MS (lower panels) for the molecular class data (Molecular groups/ Formula classes). Fitted gradients are based on percentage data (b, d; see Supplemental Tables 6 and 7) or absolute numbers (a, c). Molecular classes: BC, black carbon; PP, polyphenol; HU, highly unsaturated; UA, unsaturated aliphatics; SUG, carbohydrate; PEP, peptide. CHNO includes all Nitrogen-containing formulae ($N_{1\text{-}4}$), and CHOS includes all Sulfur-containing formulae ($S_{1\text{-}2}$). OTHER is the sum of CHOSP, CHONS and CHONP formulae.

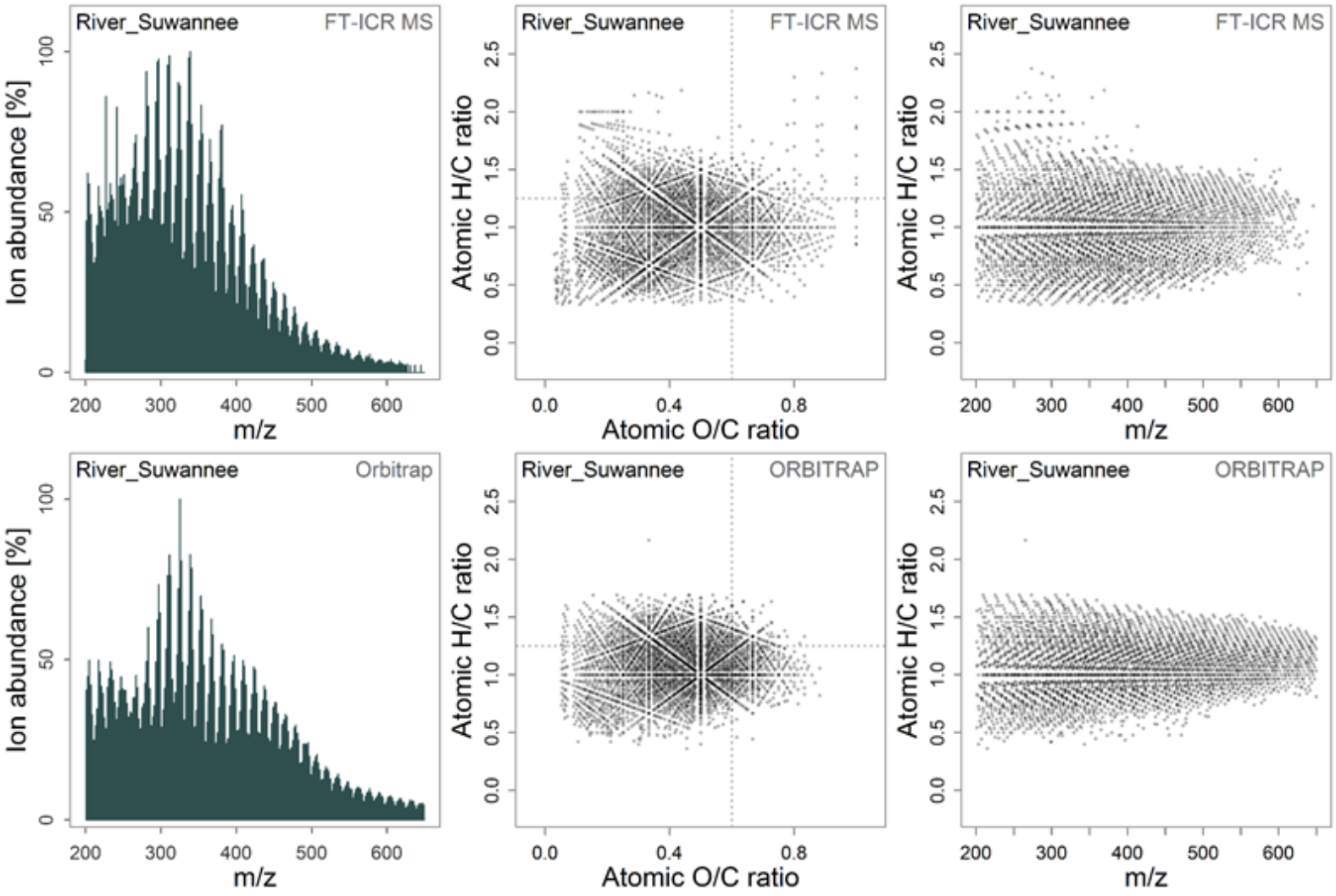
Supplemental Figure 6a. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the bog sample N3B, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



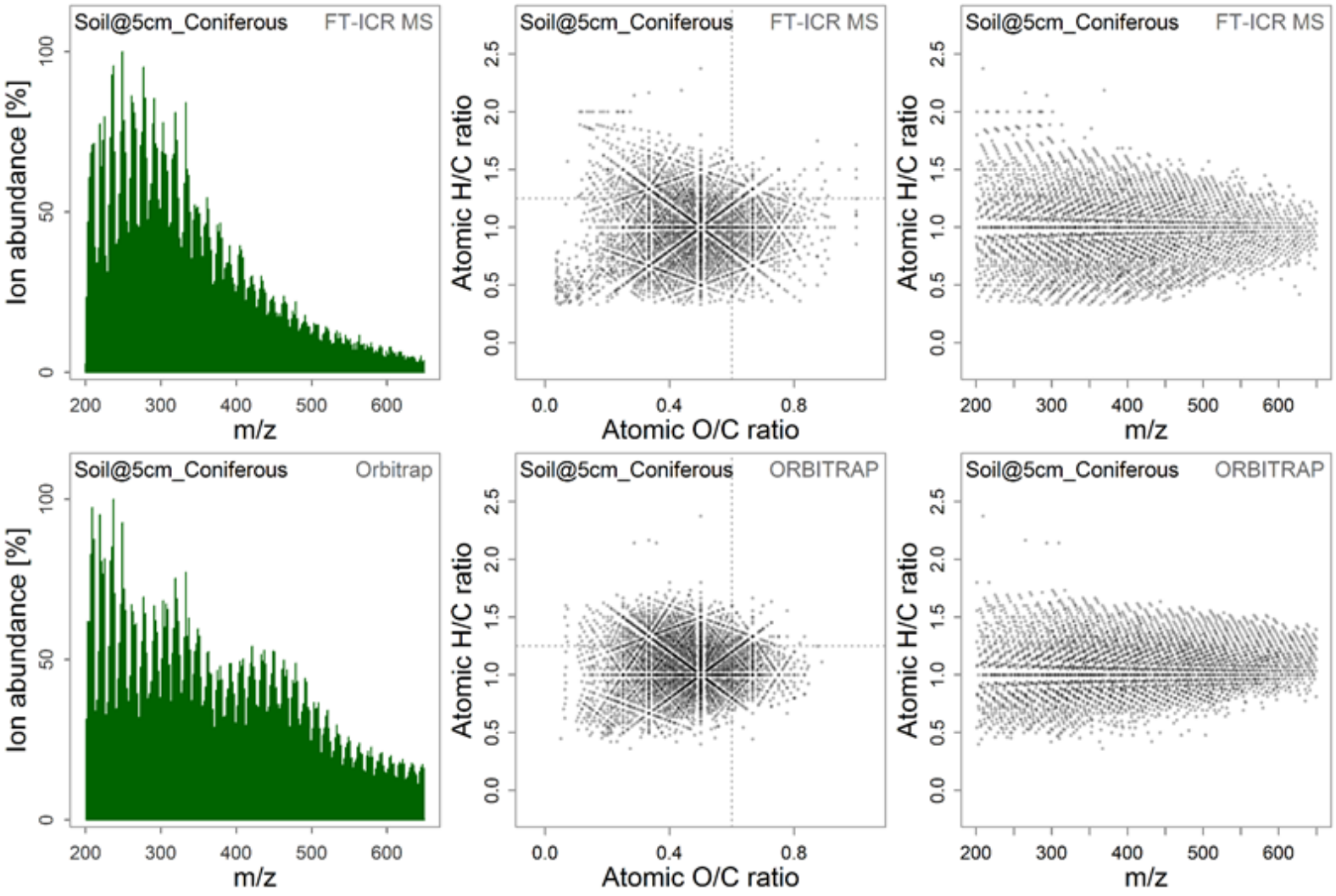
Supplemental Figure 6b. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the bog sample N8B, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



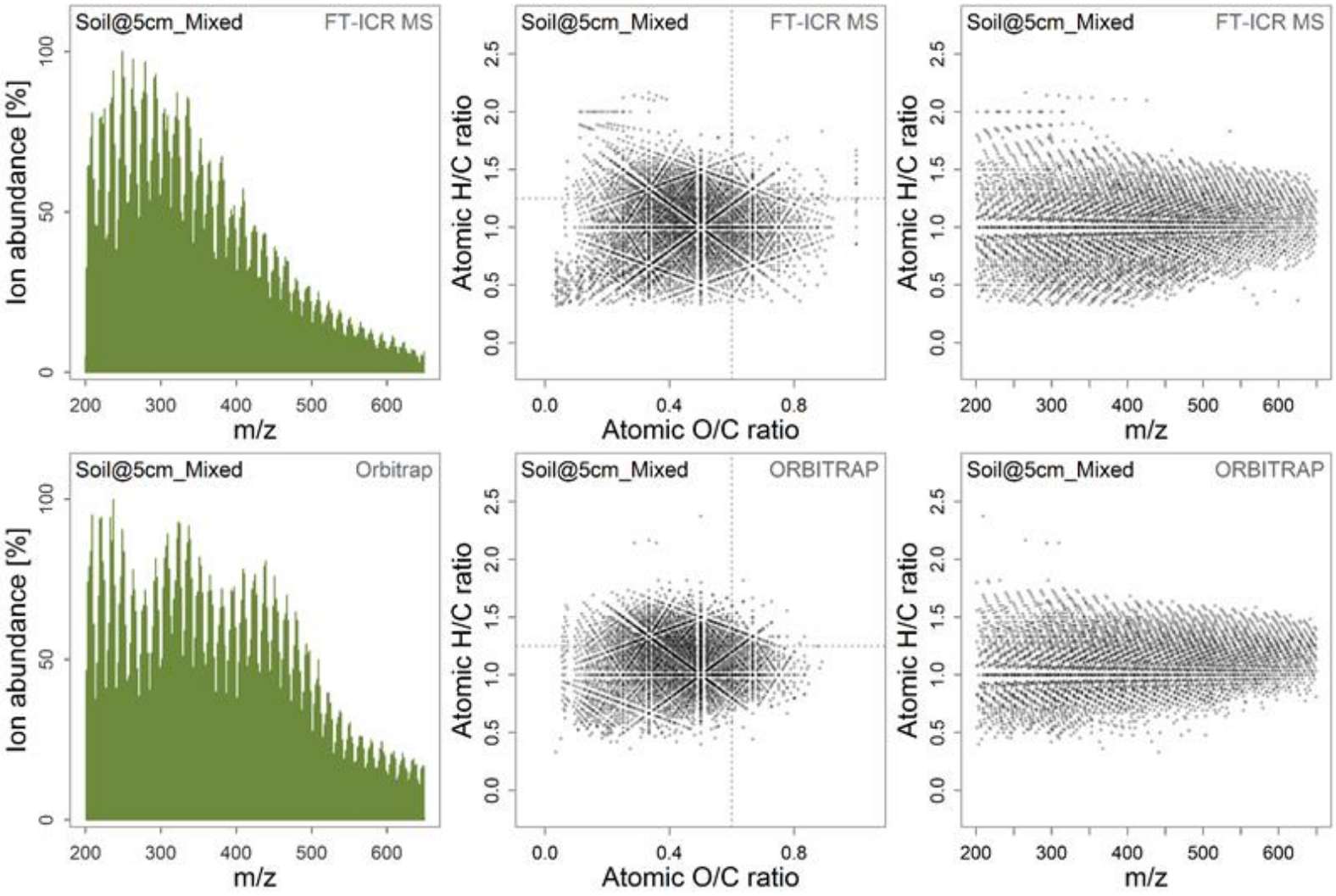
Supplemental Figure 6c. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the black water Suwannee river (IHSS) sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



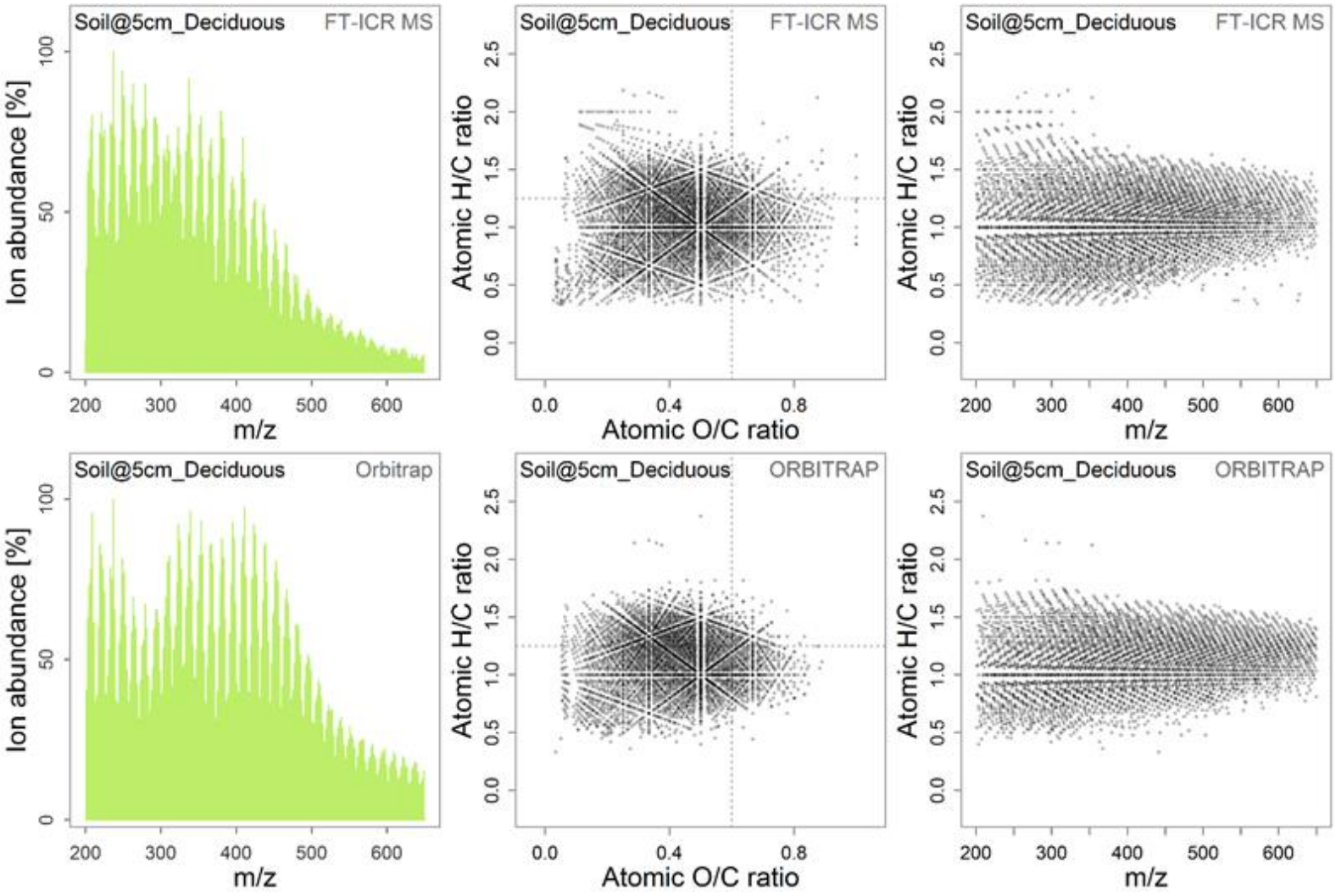
Supplemental Figure 6d. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water W1-5 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



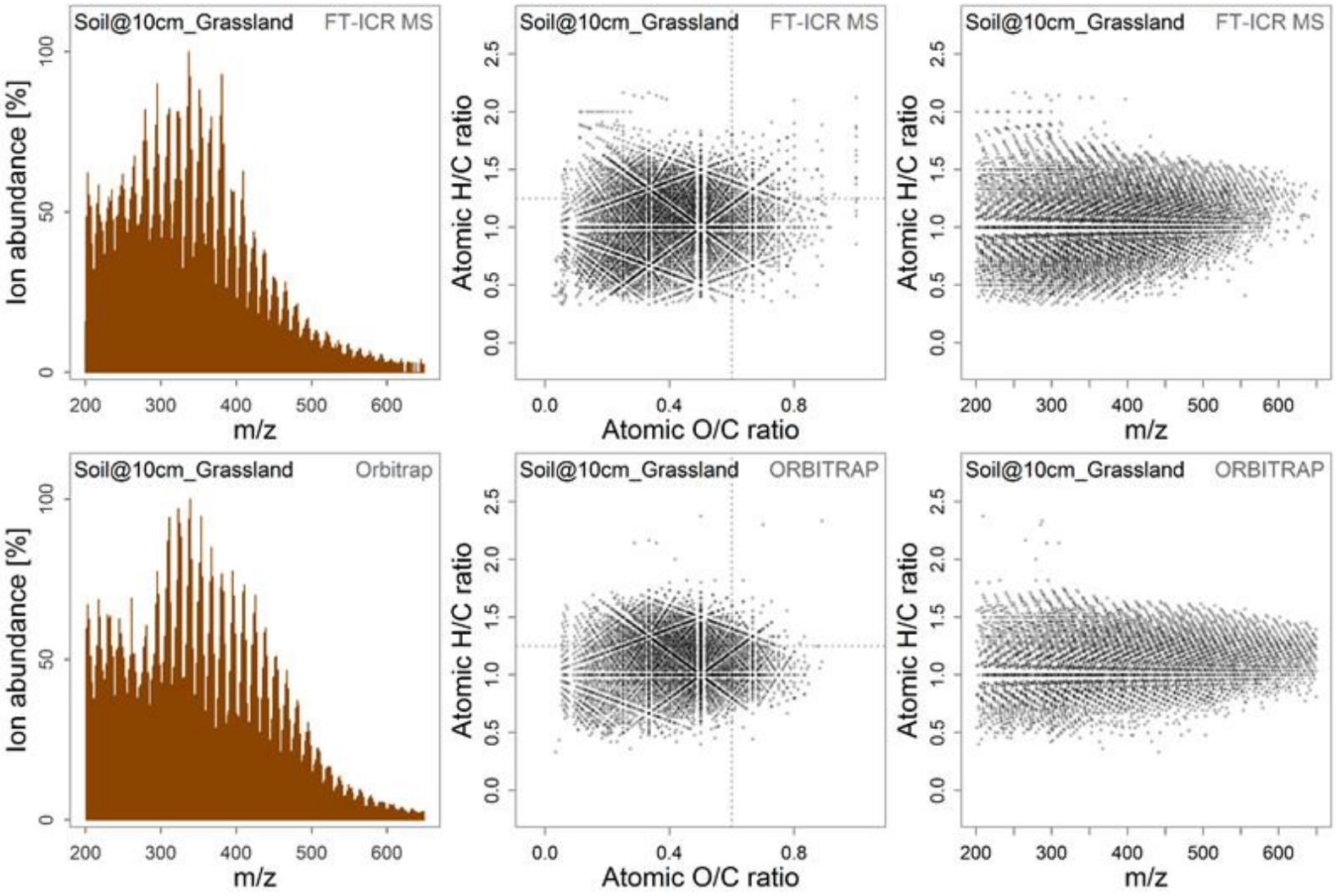
Supplemental Figure 6e. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water T-5 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



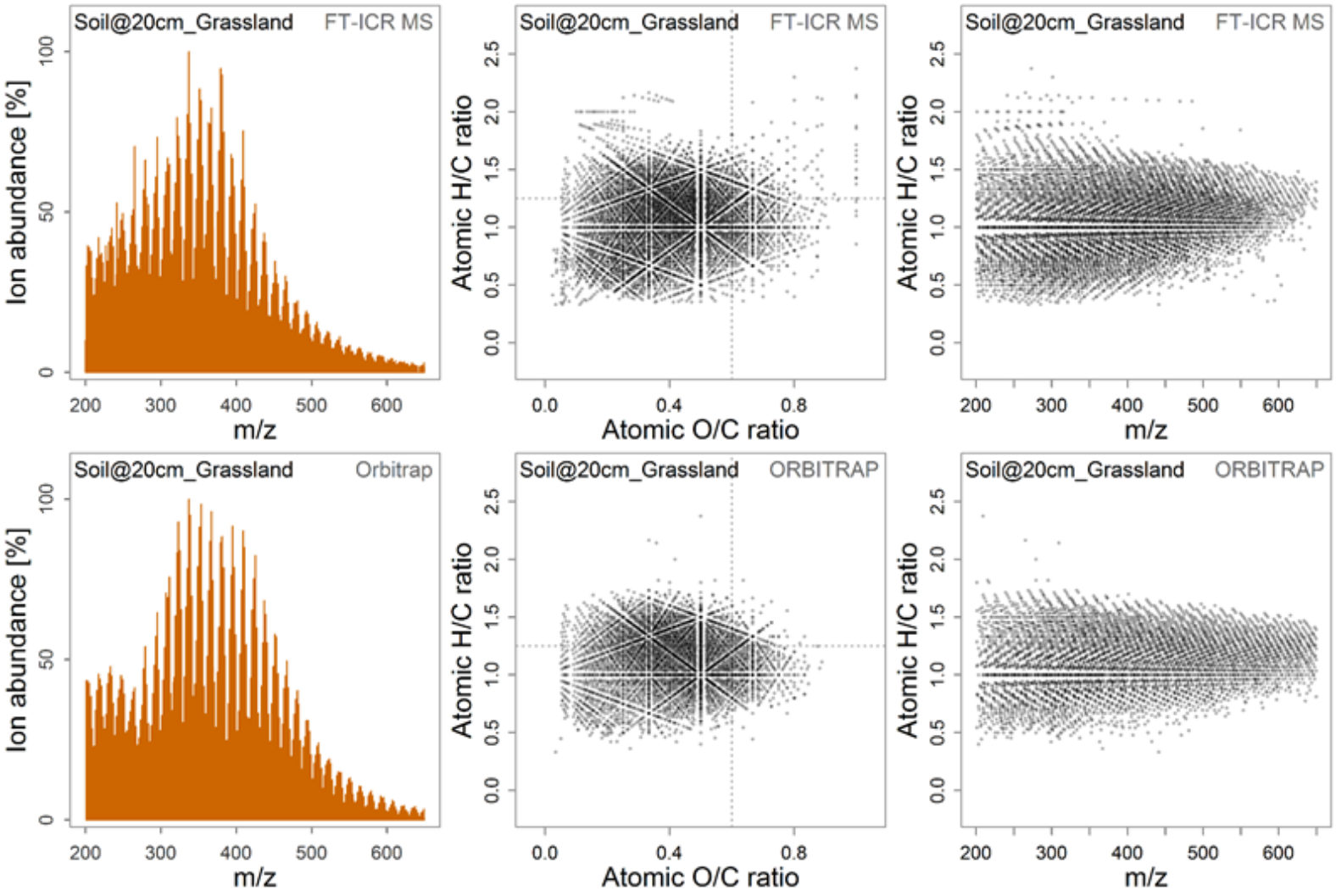
Supplemental Figure 6f. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water H2S-5 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



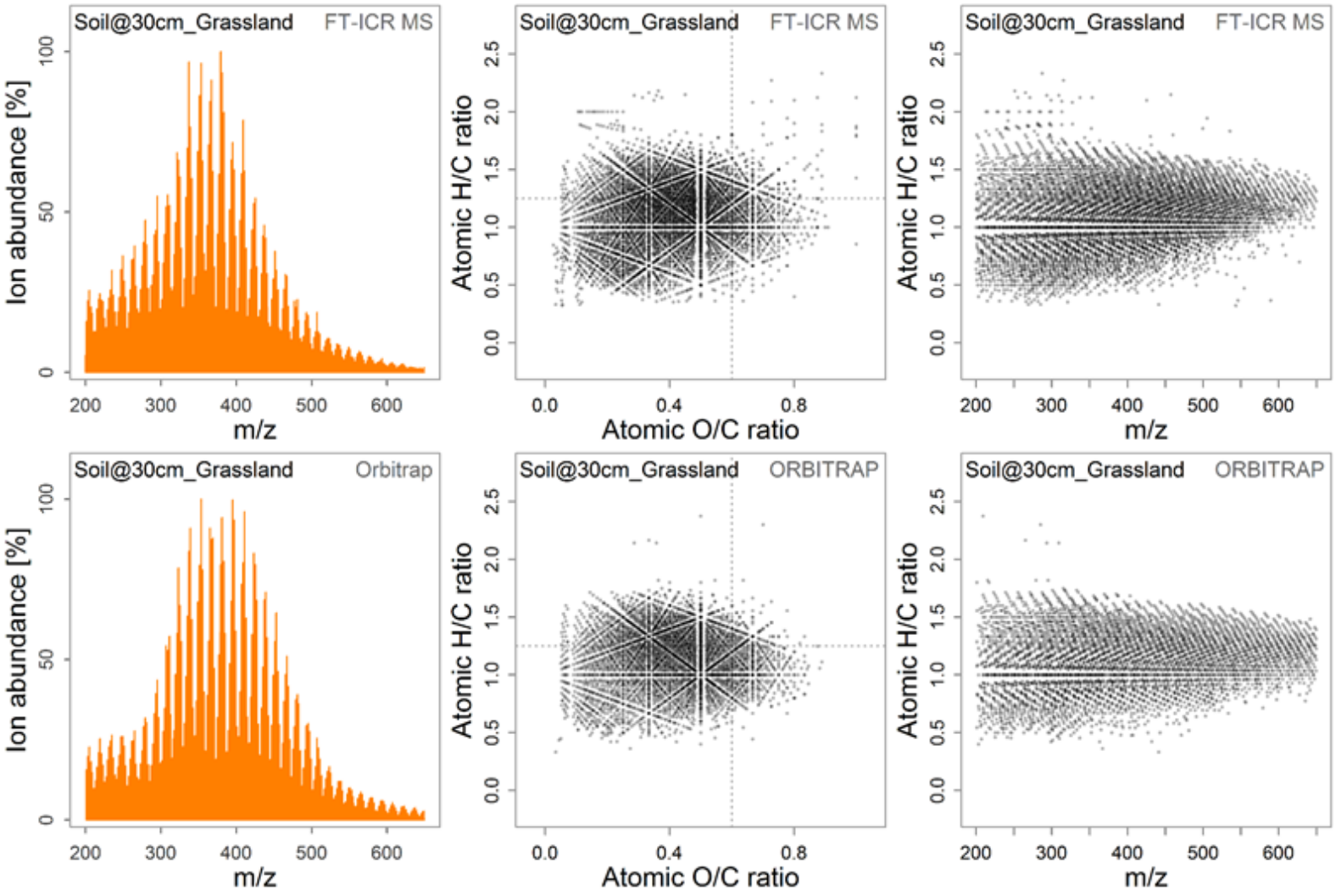
Supplemental Figure 6g. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water JE-2-5-10 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



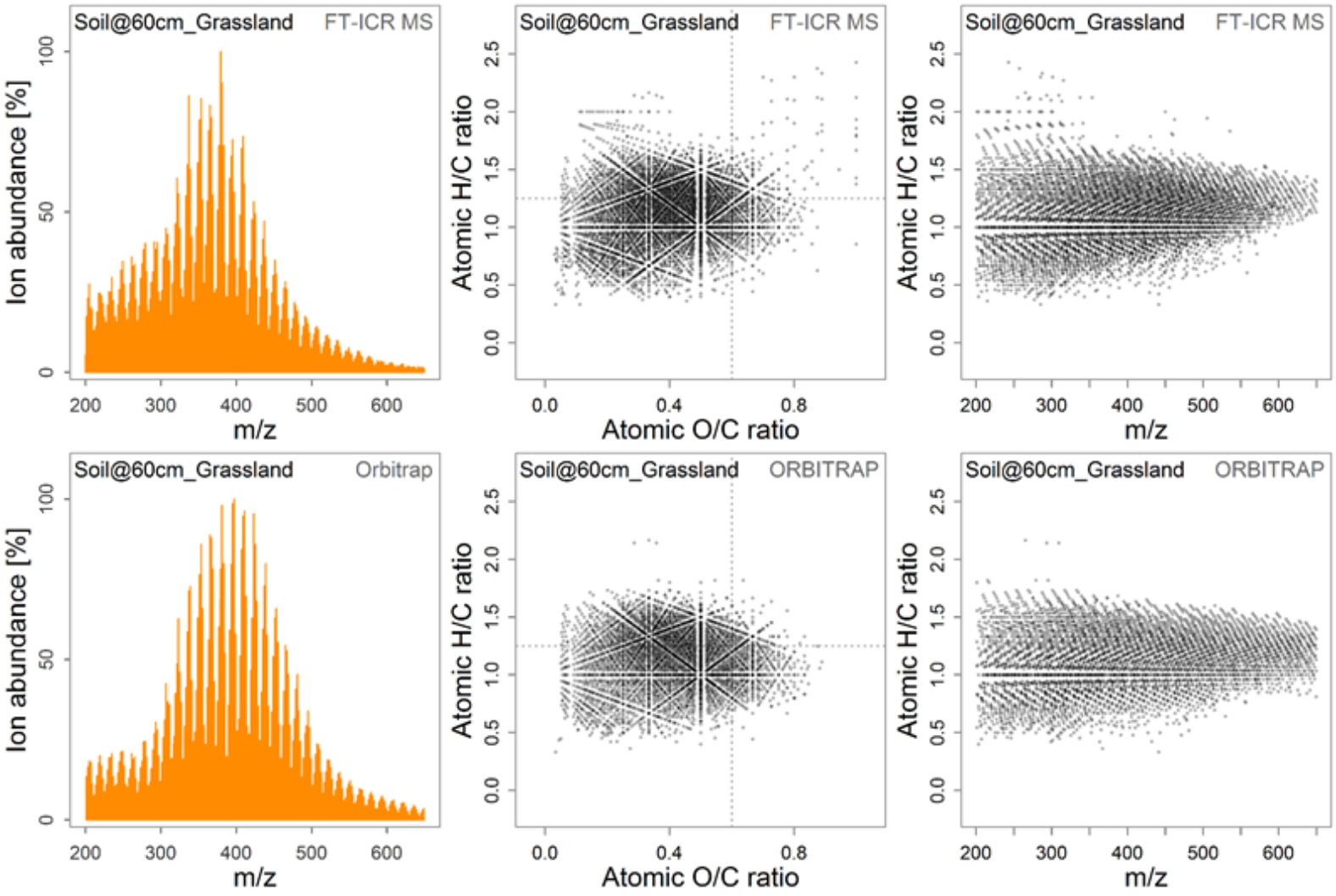
Supplemental Figure 6h. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water JE-2-5-20 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



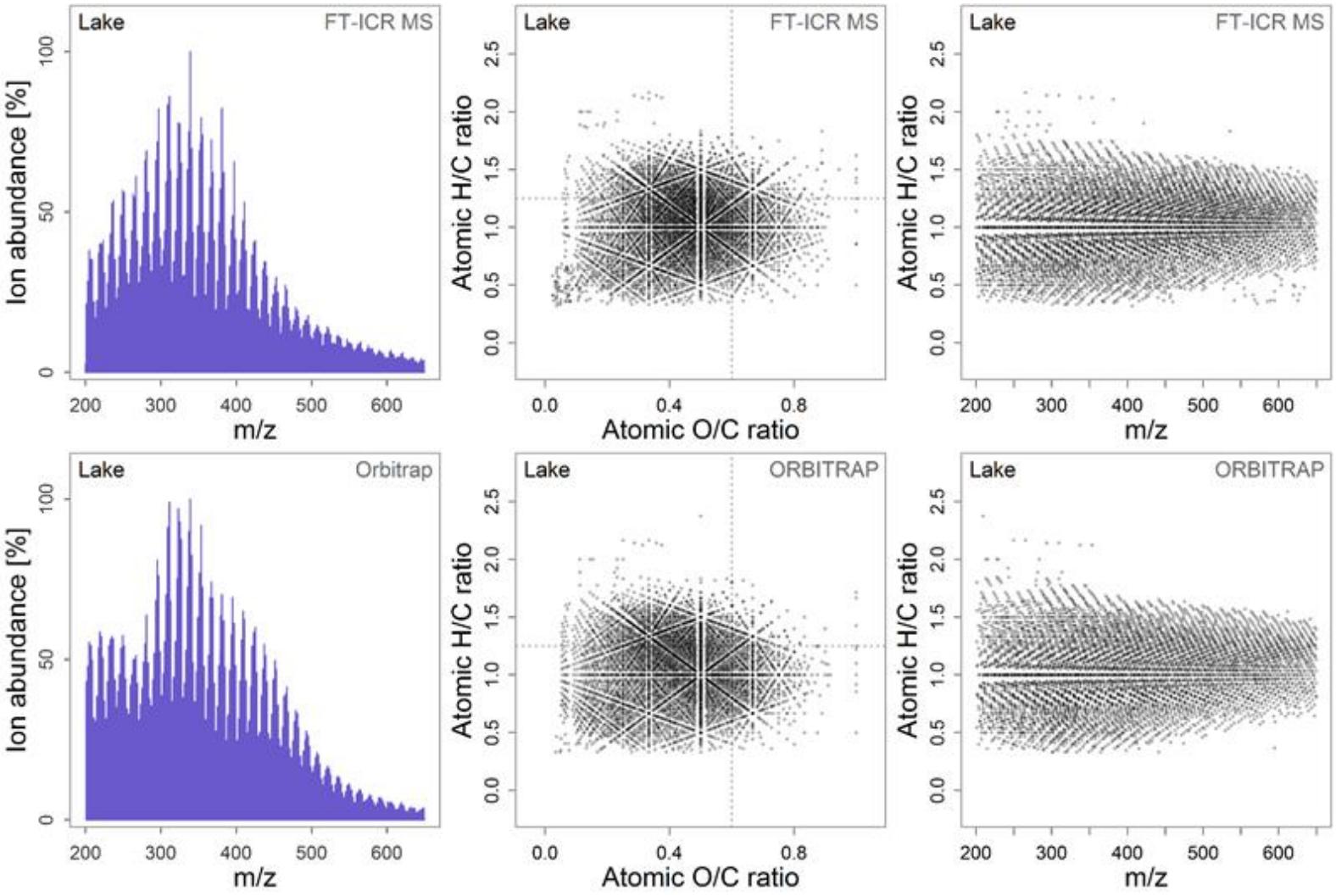
Supplemental Figure 6i. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water JE-2-5-30 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



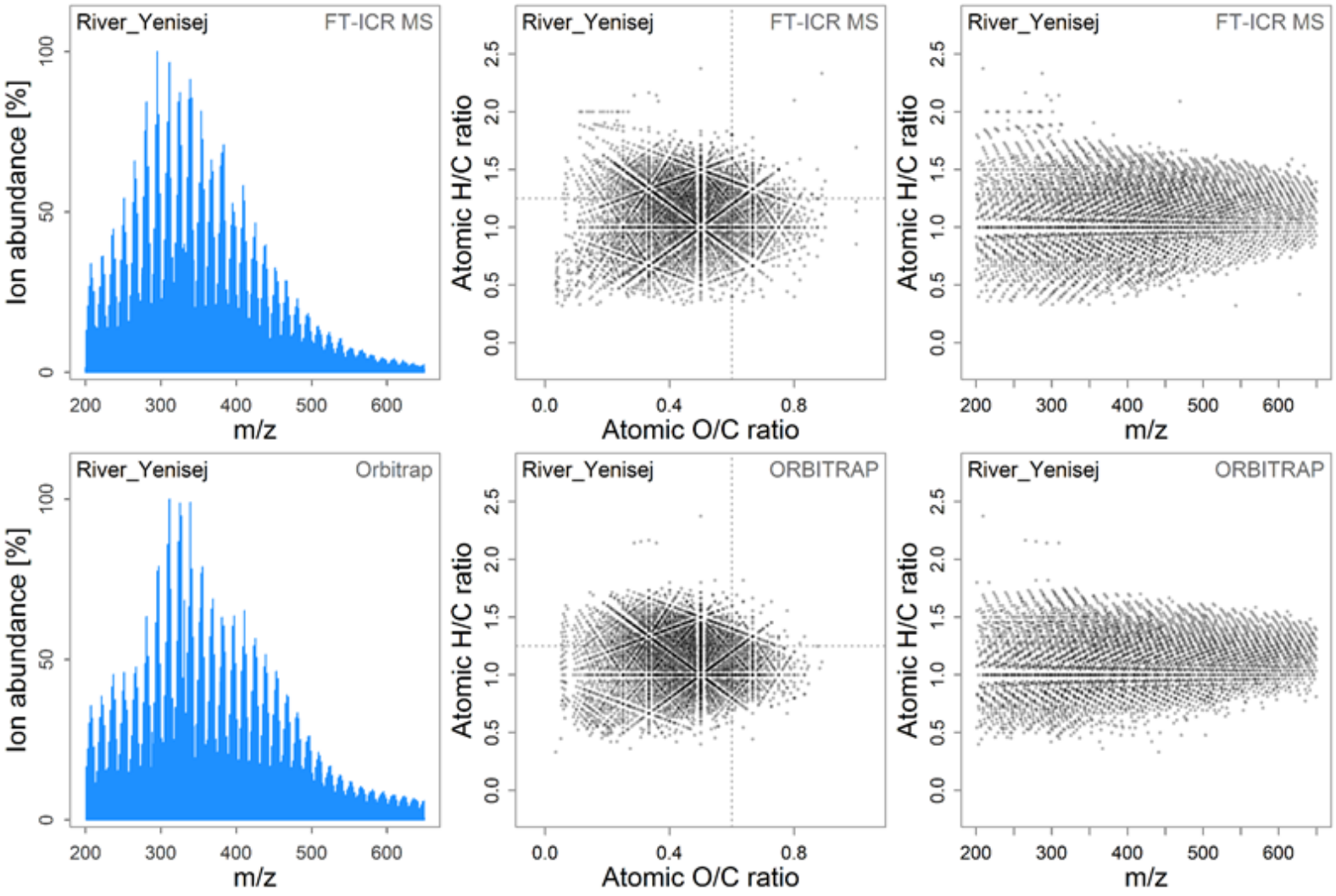
Supplemental Figure 6j. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the soil water JE-2-5-60 sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



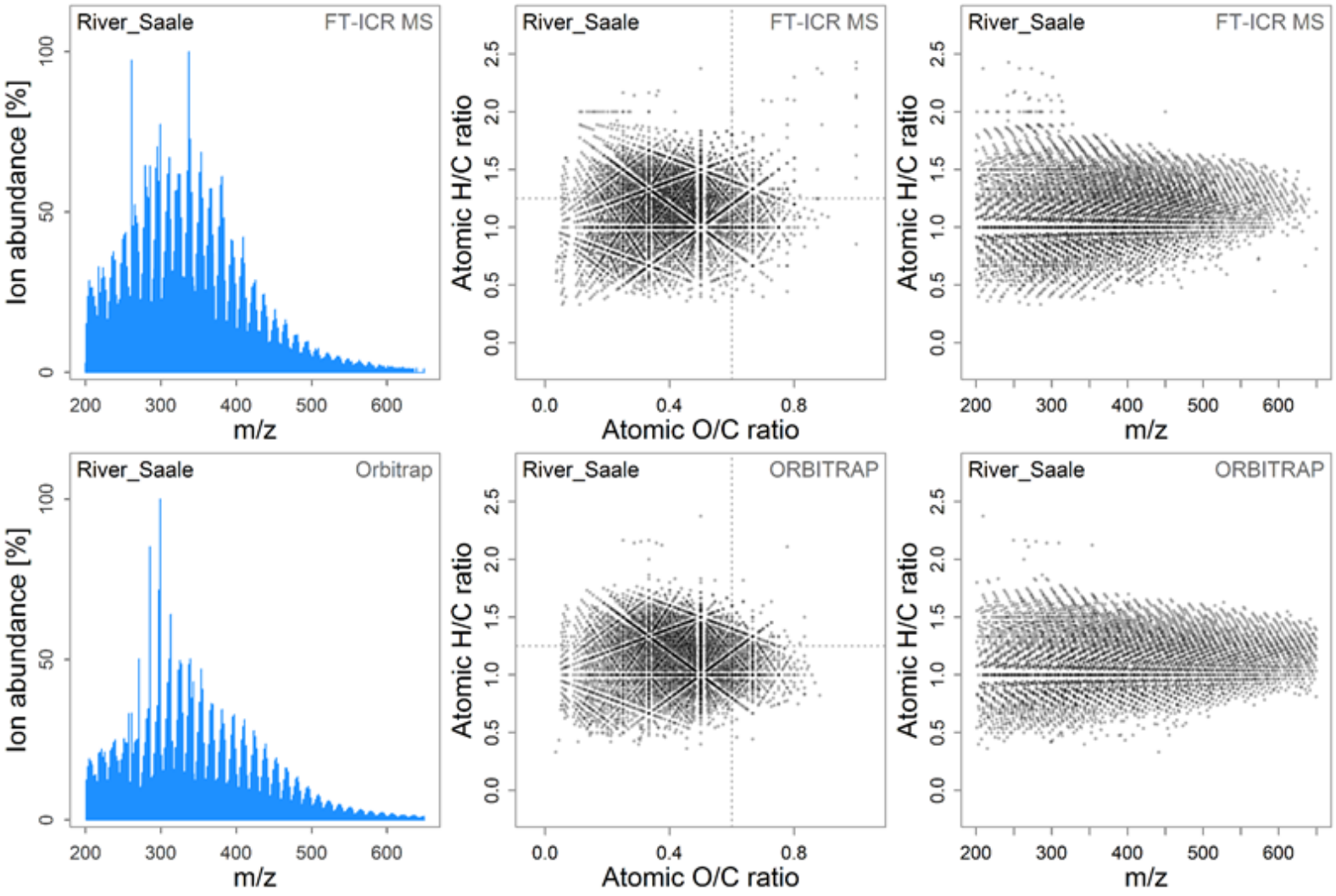
Supplemental Figure 6k. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the lake (BZWA) sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



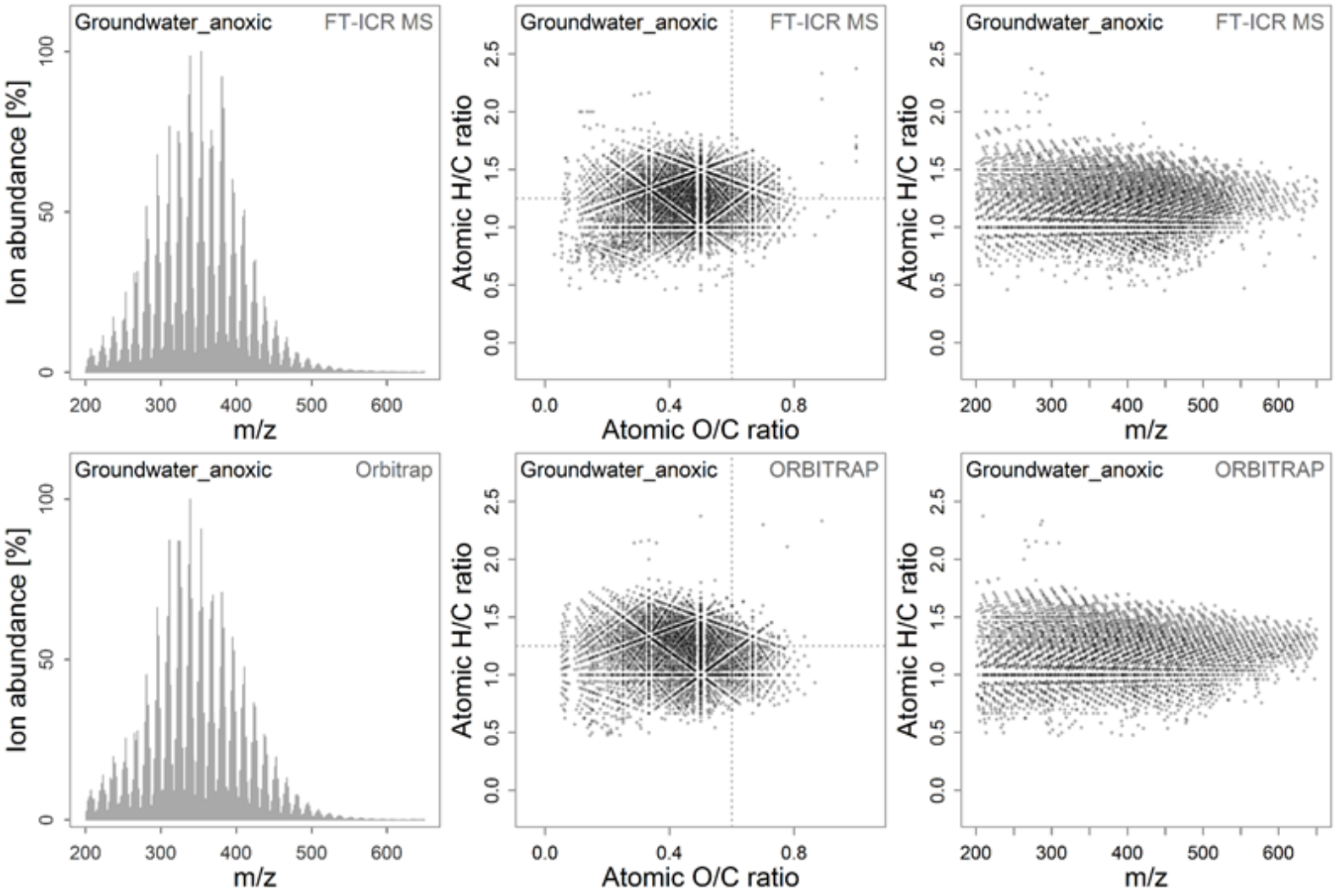
Supplemental Figure 6l. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the Yenisej river sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



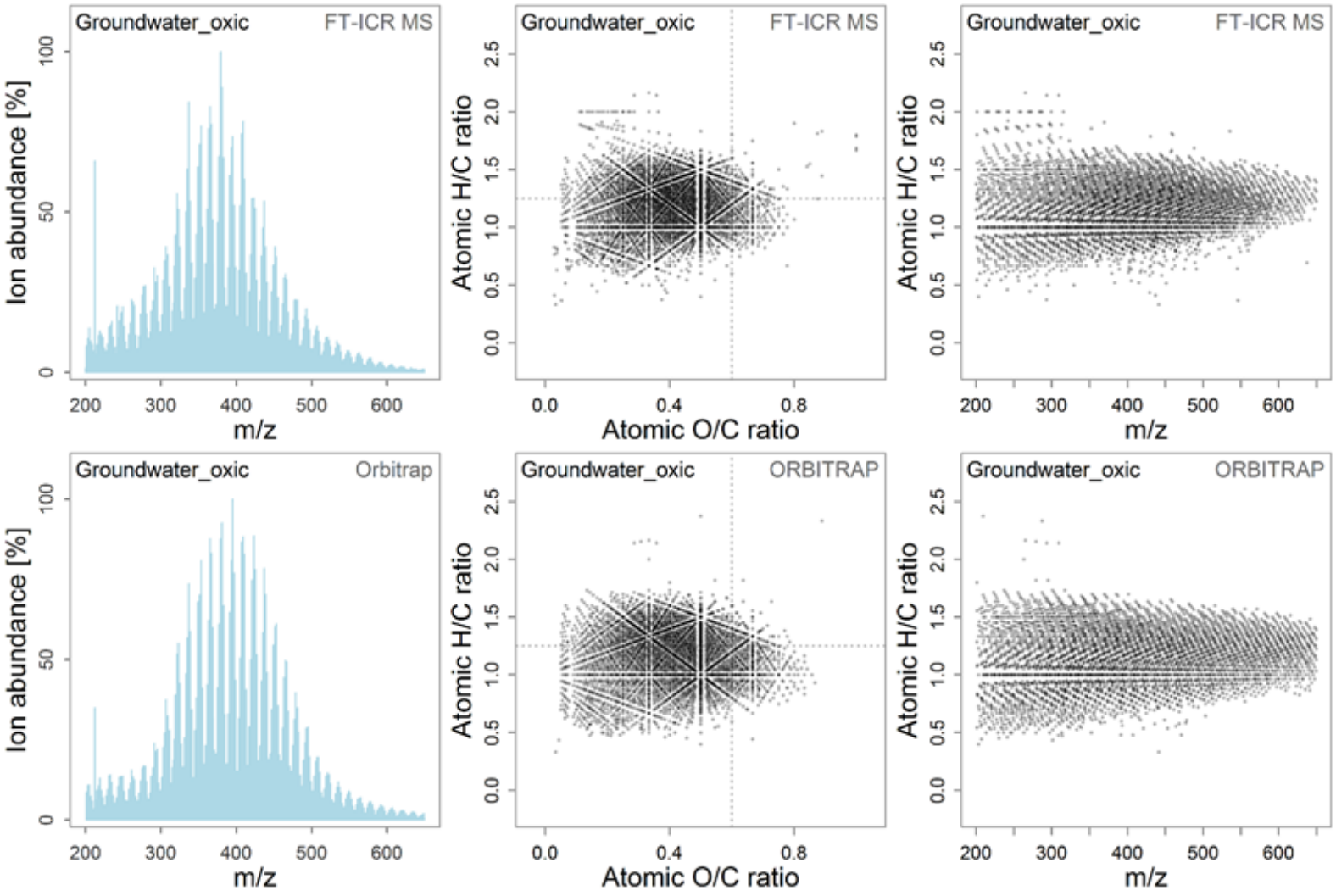
Supplemental Figure 6m. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the Saale river sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



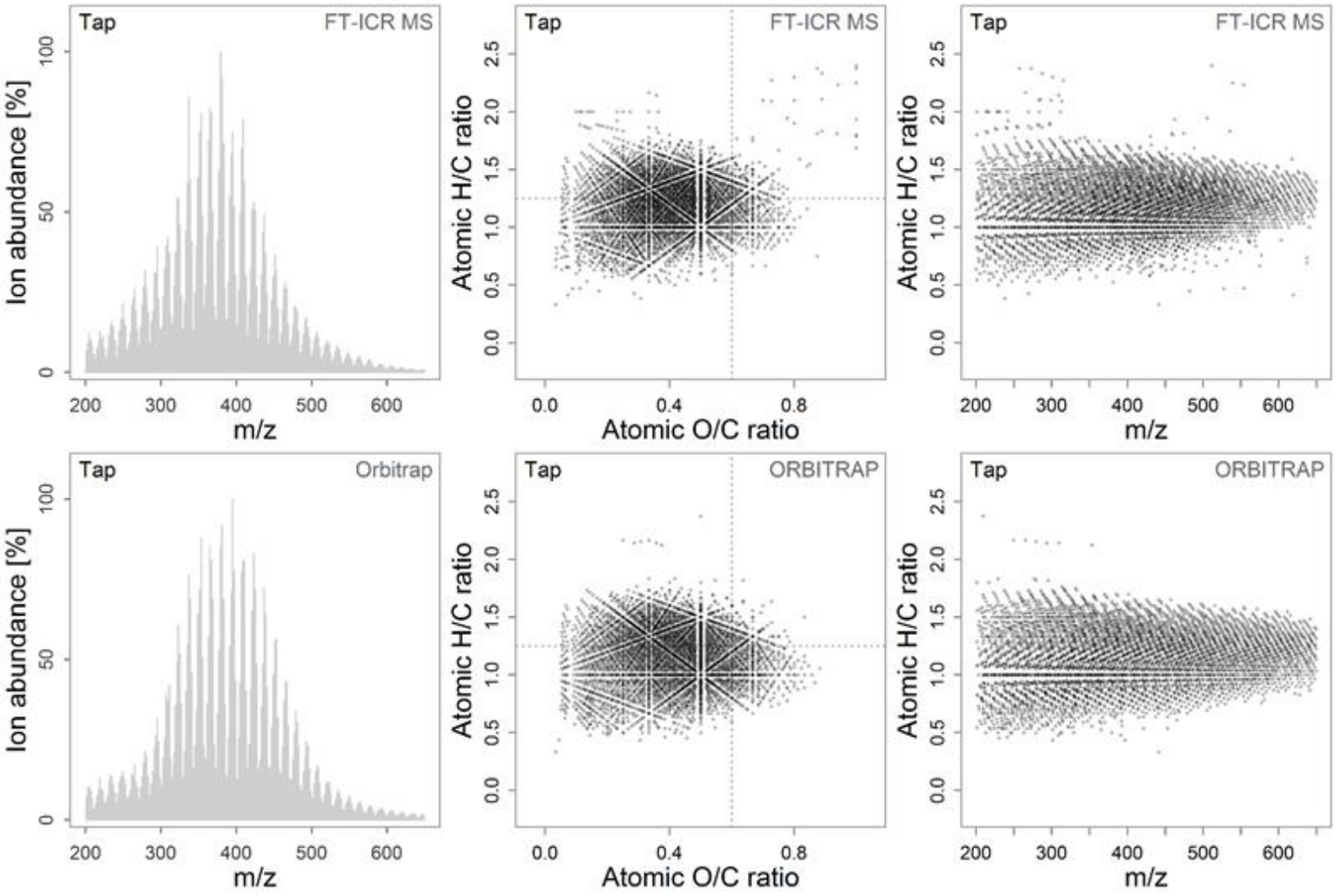
Supplemental Figure 6n. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the H5-3a anoxic aquifer sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



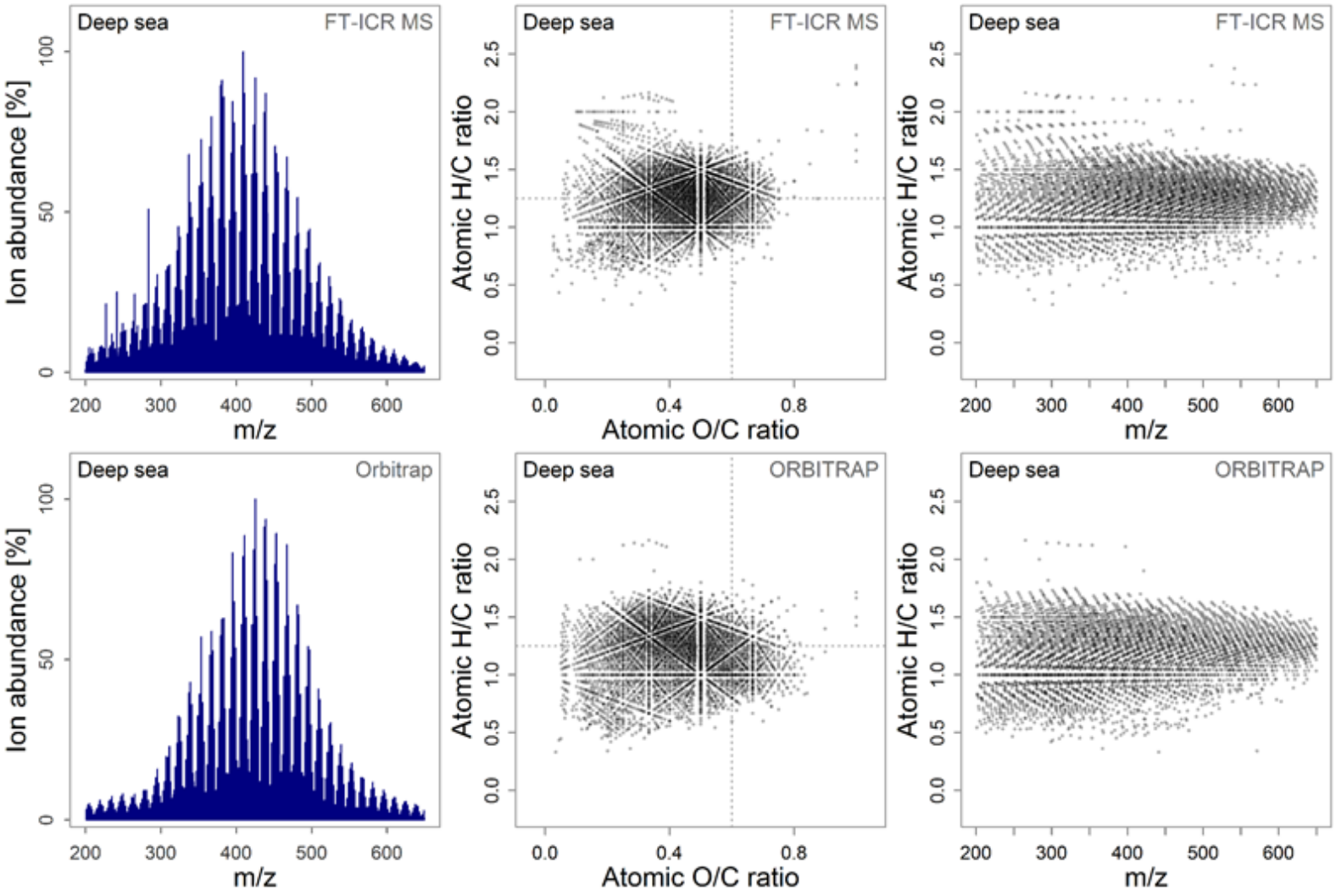
Supplemental Figure 6o. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the H3-2b oxic aquifer sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



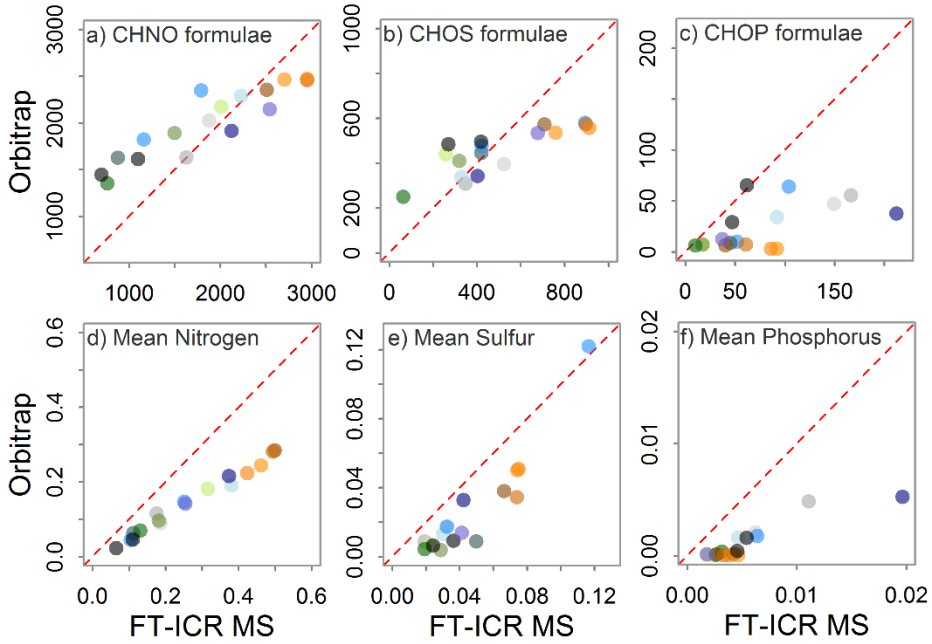
Supplemental Figure 6p. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the TAP water sample, as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



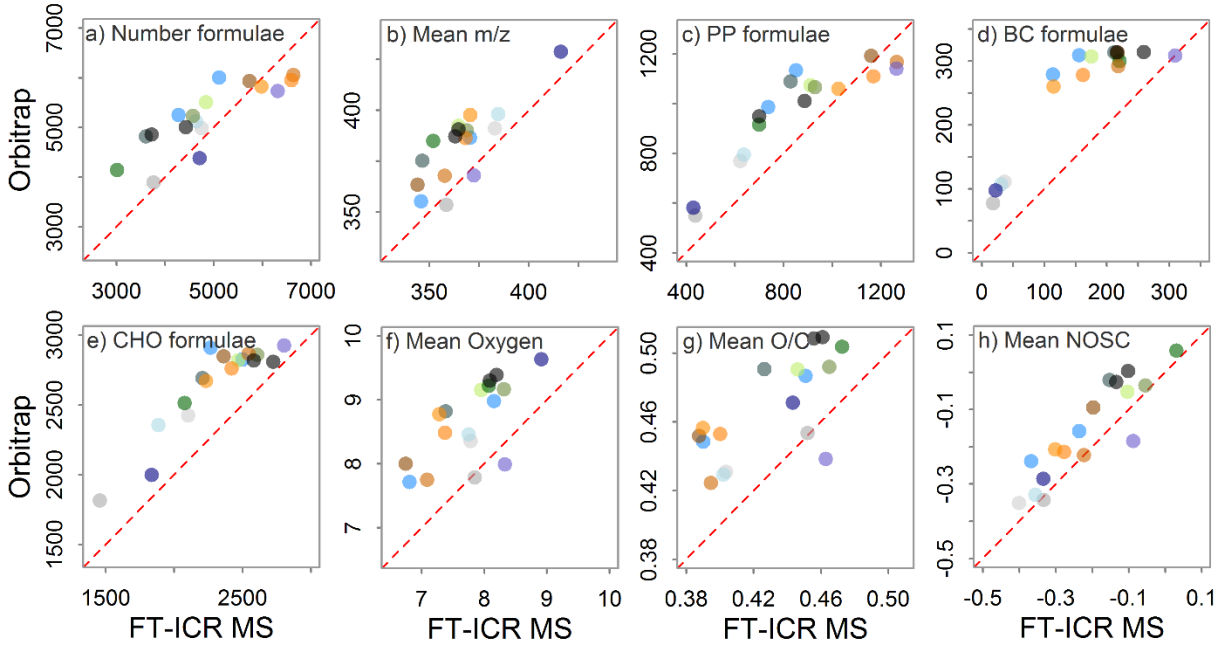
Supplemental Figure 6q. Mass spectra, van Krevelen plots and H/C vs. m/z plots for the deep sea marine sample (NELHA), as used for full set analyses. Upper panels are showing FT-ICR MS data, lower panels contain Orbitrap data.



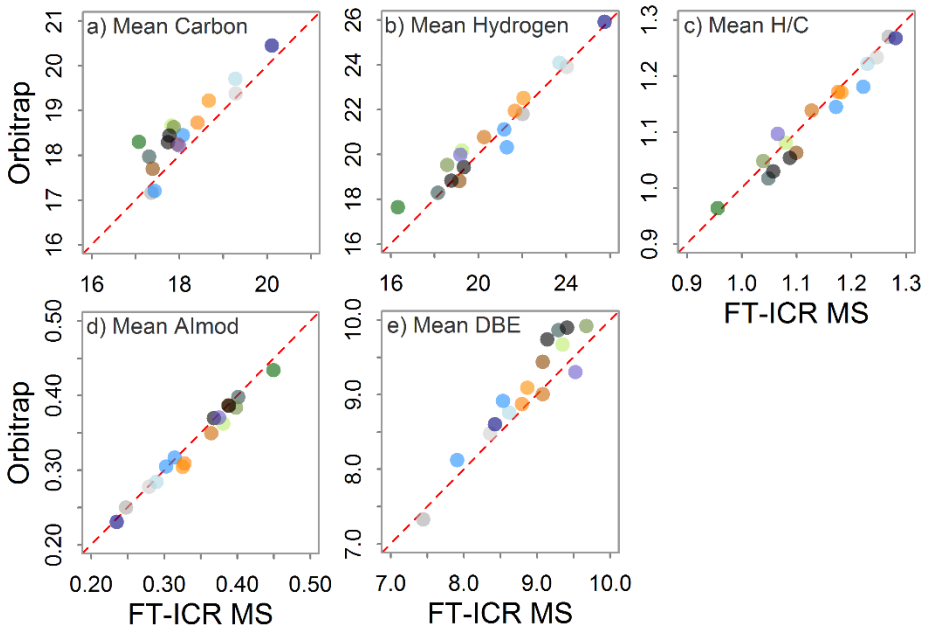
Supplemental Figure 7. Instrumental offsets due to detection of non-common signals by FT-ICR MS (due to better resolution), as shown by molecular class indices linked to N, S and P formulae, and based on the full dataset. The shown indices exerted a trend towards lower values in the Orbitrap due to limitations in resolving power (panels d, e and f) as described in the main text. The differences in tuning lead to a slight mass shift in the Orbitrap (see Supplemental Figure 8b, Supplemental Figure 3) which overlays this general resolution effect (in panels a and b) and accounts for slightly higher numbers of simple CHNO and CHOS formulae detected by the Orbitrap in samples characterized by recent inputs of fresh and degrading organic matter.



Supplemental Figure 8. Instrumental offsets due to detection of signals by Orbitrap in the higher mass range (due to differential architecture, tuning and method setup), shown by selected chemical and molecular class indices, and based on the full dataset. The shown indices exerted a trend towards higher values in the Orbitrap with a tendency to be more pronounced for samples connected to recent inputs of fresh and degrading organic matter. The observed higher number of compounds are related to unavoidable differences in instrumental response due to tuning (Orbitrap was tuned with IHSS sample, FT-ICR MS with NELHA). The Orbitrap-detected signals of mainly aromatic CHO and simple CHNO/ CHOS formulae in the higher mass range were also found by FT-ICR MS but not within the exact same samples; this difference being mainly an effect of tuning and not FT-ICR MS capabilities.



Supplemental Figure 9. Examples of chemical indices that showed no strong offsets between instruments, as based on the full dataset.



Supplemental note 1. Application of a method detection limit (MDL)

The used MDL approach (Riedel and Dittmar, 2014) approximates noise height by analyzing the intensity (i.e., ion abundance) distribution in the signal region between 0.3 and 0.9 Da for each nominal mass. DOM signal "islands" are usually stretched over a mass defect range from -0.1 Da to +0.3 Da. Noise detection can be implemented on the sample data itself that way, by this also saving instrument time. A problem encountered with the Orbitrap data were peaks around 0.5 – 0.65 Da, being indicative of double-charged species. These signals were thus excluded for noise assessment. We used only those nominal masses that contained more than 20 noise signals, and only those showing a normal distribution (of intensity data, assessed by Student's t-test; Riedel & Dittmar, 2014). Noise peak ensembles showing p-values > 0.05 in the t-Test (rejecting the assumption of non-normality) also showed small deviation in a QQ-plot assessment (deviation from the normality line, Supplemental Figure 4a). By applying this MDL approach to the Orbitrap data, we found that the "20 peaks"- limit and the normality criterion were not fulfilled by all nominal masses, especially at higher m/z (Supplemental Figure 4b). This is partly due to decreasing resolution with m/z, but was also negatively influenced by the data acquisition mode. The default mode of data acquisition on the Orbitrap Elite ("reduced profile mode") influences mainly the noise peak yield per nominal mass. This mode of data acquisition reduces file size and scan time and is thus pivotal for high throughput data analysis with regular computing power (standard run with 100 scans, 22 kB file size in reduced vs. 800 kB in full mode; one standard scan, 1 sec in reduced vs. 2.2 sec in full mode). However, this form of acquisition by intention rules out the majority of the noise to improve signal detection, and thus hinders application of the MDL approach. Future studies need to take this into account and should determine the MDL based on full-profile data or at least compare reduced-profile data to it. Despite the named drawbacks, the MDL levels of nominal masses fulfilling the normality criterion were used here for an estimation of a linear regression over m/z and were extrapolated up to m/z 1000 for each sample individually (Supplemental Figure 4b). Each sample was corrected with its own method detection limit (MDL) approximation, set at a conservative confidence level of 99.8 %. For the samples under study, MDL levels increased with m/z, while they were rather constant and generally lower for blank samples with only ~ hundreds of signals.

References

Cao, D., Lv, J., Geng, F., Rao, Z., Niu, H., Shi, Y., Cai, Y., Kang, Y., 2016. Ion accumulation time dependent molecular characterization of natural organic matter using electrospray ionization-Fourier transform ion cyclotron resonance mass spectrometry. Anal. Chem. 88, 12210–12218.

Cortés-Francisco, N., Flores, C., Moyano, E., Caixach, J., 2011. Accurate mass measurements and ultrahigh-resolution: Evaluation of different mass spectrometers for daily routine analysis of small molecules in negative electrospray ionization mode. Anal. Bioanal. Chem. 400, 3595–3606.

Hawkes, J.A., Dittmar, T., Patriarca, C., Tranvik, L., Bergquist, J., 2016. Evaluation of the Orbitrap mass spectrometer for the molecular fingerprinting analysis of natural dissolved organic matter. Anal. Chem. 88, 7698–7704.

Mangal, V., Stock, N.L., Guéguen, C., 2016. Molecular characterization of phytoplankton dissolved organic matter (DOM) and sulfur components using high resolution Orbitrap mass spectrometry. Anal. Bioanal. Chem. 408, 1891–1900.

Marshall, A.G., Hendrickson, C.L., 2008. High-resolution mass spectrometers. Annu. Rev. Anal. Chem. 1, 579–599.

Perry, R.H., Cooks, G., Noll, R.J., 2008. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. Mass Spectrom. Rev. 27, 661–699.

Pomerantz, A.E., Mullins, O.C., Paul, G., Ruzicka, J., Sanders, M., 2011. Orbitrap mass spectrometry: A proposal for routine analysis of nonvolatile components of petroleum. Energy and Fuels 25, 3077–3082.

Remucal, C.K., Cory, R.M., Sander, M., McNeill, K., 2012. Low molecular weight components in an aquatic humic substance as characterized by membrane dialysis and Orbitrap mass spectrometry. Environ. Sci. Technol. 46, 9350–9.

Riedel, T., Dittmar, T., 2014. A method detection limit for the analysis of natural organic matter via Fourier transform ion cyclotron resonance mass spectrometry. Anal. Chem. 86, 8376–82.

Smith, E.A., Park, S., Klein, A.T., Lee, Y.J., 2012. Bio-oil analysis using negative electrospray ionization: Comparative study of high resolution mass spectrometers and phenolic vs. sugaric components. Energy & Fuels 26, 3796–3802.

Zhurov, K.O., Kozhinov, A.N., Tsybin, Y.O., 2013. Evaluation of high-field Orbitrap Fourier transform mass spectrometer for petroleomics. Energy & Fuels 27, 2974–2983.

Zubarev, R.A., Makarov, A., 2013. Orbitrap mass spectrometry. Anal. Chem. 85, 5288–5296.