Electronic Supplementary Material (ESI) for Analyst.
This journal is © The Royal Society of Chemistry 2024

Supplementary information for:

Post column infusion of an internal standard into LC-FT-ICR MS enables semi-quantitative comparison of dissolved organic matter in original samples

Rebecca Rodrigues Matos[a], Elaine K. Jennings[a], Jan Kaesler[a], Thorsten Reemtsma[a,b], Boris P. Koch[c,d], Oliver J. Lechtenfeld[a,e,*]

[a] Department of Environmental Analytical Chemistry, Helmholtz Centre for Environmental Research − UFZ, Permoserstr. 15, 04318 Leipzig, Germany

[b] Institute of Analytical Chemistry, University of Leipzig, 04103 Leipzig, Germany

[c] Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research, Ecological Chemistry, Am Handelshafen 12, 27570 Bremerhaven, Germany

[d] Hochschule Bremerhaven, University of Applied Sciences, An der Karlstadt 8, 27568 Bremerhaven, Germany

[e] ProVIS−Centre for Chemical Microscopy, Helmholtz Centre for Environmental Research − UFZ, Permoserstr. 15, 04318 Leipzig, Germany

* Corresponding author

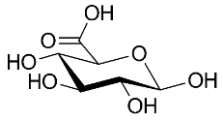
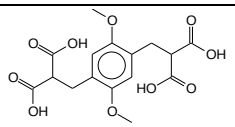
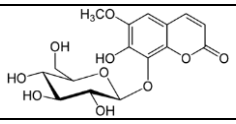
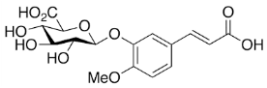
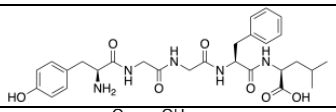
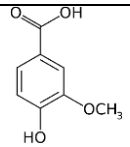
SI Text: 1

SI Tables: 7

SI Figures: 9

SI 1 – Materials and Methods

Table SI 1.1 – Model compounds used in this study.

Model compound	$[M+H]^-$	Molecular formula	Chemical structure	RT (min)	Concentration (ng mL^{-1})
D-glucuronic acid	193.0354	$C_6H_{10}O_7$		5.04	25
2-(4-(2,2-Dicarboxy-ethyl)-2,5-dimethoxy-benzyl)malonic acid	369.0827	$C_{16}H_{18}O_{10}$		14.25	4
Fraxin	369.0827	$C_{16}H_{18}O_{10}$		16.04	8
Isoferulic acid 3-*O*-β-D-glucuronide	369.0827	$C_{16}H_{18}O_{10}$		15.06	4
Leu-enkephalin	554.262	$C_{28}H_{37}N_5O_7$		16.13	5
Vanilic acid	167.035	$C_8H_8O_4$		15.34	50

Text SI 1.1 – Holtemme and Elbe sample site description.

While Holtemme spring (51° 47' N 10° 40' E) and Elbe River (53° 3' N, 11° 32' E) native samples were comparable based on their DOC concentration (6.39 mg C L^{-1} and 5.27 mg C L^{-1} respectively), they represent a large compositional gradient. Holtemme is a small third order stream draining a mountainous forested in the Elbe catchment area.[1] The Holtemme river merges with the rivers Bode, then Saale and ultimately Elbe, which drains a catchment of 148,000 km^2. Elbe, thus, represents the DOM from a mix of sources, including photoautotrophic and anthropogenically impacted DOM (*e.g.* agricultural and wastewater treatment effluent). The native Elbe sample was collected on May 5th of 2022 and Holtemme spring on May 24th of 2022.

Table SI 1.2 – Calibrants and their *m/z* values used in this study.

Calibrant	m/z	Calibrant	m/z
C8H5O4	165.019332	C21H21O9	417.1191
C8H7O4	167.034982	C21H23O9	419.1348
C8H11O4	171.066282	C21H25O9	421.1504
C6H11O6	179.056112	C20H23O10	423.1297
C10H9O4	193.050632	C20H25O10	425.1453
C9H7O5	195.029897	C17H15O13	427.0518
C10H15O4	199.097583	C19H23O11	427.1246
C10H9O5	209.045547	C18H21O12	429.1039
C9H7O6	211.024812	C22H23O9	431.1348
C11H9O5	221.045547	C22H25O9	433.1504
C10H7O6	223.024812	C22H27O9	435.1661
C9H5O7	225.004076	C21H25O10	437.1453
C10H9O6	225.040462	C20H23O11	439.1246
C11H15O5	227.092497	C20H25O11	441.1402
C11H7O6	235.024812	C21H15O11	443.062
C12H11O5	235.061197	C21H17O11	445.0776
C11H9O6	237.040462	C20H15O12	447.0569
C11H11O6	239.056112	C22H25O10	449.1453
C11H15O6	243.087412	C22H27O10	451.161
C12H7O6	247.024812	C21H25O11	453.1402
C12H9O6	249.040462	C21H27O11	455.1559
C12H11O6	251.056112	C20H25O12	457.1352
C13H17O5	253.108147	C23H23O10	459.1297
C16H31O2	255.232954	C23H25O10	461.1453
C12H17O6	257.103062	C23H27O10	463.161
C13H9O6	261.040462	C22H25O11	465.1402
C14H15O5	263.092497	C22H27O11	467.1559
C12H9O7	265.035376	C21H25O12	469.1352
C13H15O6	267.087412	C20H23O13	471.1144
C13H17O6	269.103062	C21H13O13	473.0362
C12H15O7	271.082326	C23H23O11	475.1246
C14H11O6	275.056112	C23H25O11	477.1402
C15H17O5	277.108147	C22H23O12	479.1195
C14H15O6	279.087412	C21H21O13	481.0988
C14H17O6	281.103062	C21H23O13	483.1144
C13H15O7	283.082326	C21H25O13	485.1301
C13H17O7	285.097976	C23H19O12	487.0882
C12H15O8	287.077241	C25H29O10	489.1766
C14H9O7	289.035376	C24H27O11	491.1559
C15H15O6	291.087412	C24H29O11	493.1715
C15H17O6	293.103062	C24H31O11	495.1872
C15H19O6	295.118712	C19H15O16	499.0366
C12H9O9	297.025205	C22H27O13	499.1457
C14H17O7	297.097976	C23H17O13	501.0675
C13H15O8	299.077241	C24H23O12	503.1195
C15H9O7	301.035376	C22H17O14	505.0624
C16H15O6	303.087412	C24H27O12	507.1508
C15H13O7	305.066676	C23H25O13	509.1301
C15H15O7	307.082326	C23H27O13	511.1457
C15H17O7	309.097976	C24H17O13	513.0675
C13H11O9	311.040856	C26H27O11	515.1559
C15H19O7	311.113627	C23H17O14	517.0624
C13H13O9	313.056506	C26H31O11	519.1872

C14H17O8	313.092891	C25H29O12	521.1665
C13H15O9	315.072156	C24H27O13	523.1457
C15H9O8	317.030291	C24H29O13	525.1614
C15H11O8	319.045941	C24H31O13	527.177
C16H17O7	321.097976	C27H29O11	529.1715
C16H19O7	323.113627	C25H23O13	531.1144
C15H17O8	325.092891	C26H29O12	533.1665
C15H19O8	327.108541	C24H23O14	535.1093
C14H17O9	329.087806	C25H29O13	537.1614
C18H19O6	331.1187	C24H27O14	539.1406
C17H17O7	333.098	C24H29O14	541.1563
C15H11O9	335.0409	C28H31O11	543.1872
C16H17O8	337.0929	C27H29O12	545.1665
C14H11O10	339.0358	C26H27O13	547.1457
C16H19O8	339.1085	C26H29O13	549.1614
C16H21O8	341.1242	C26H31O13	551.177
C15H19O9	343.1035	C25H29O14	553.1563
C14H17O10	345.0827	C24H27O15	555.1355
C18H19O7	347.1136	C26H21O14	557.0937
C17H17O8	349.0929	C25H19O15	559.0729
C17H19O8	351.1085	C26H25O14	561.125
C17H21O8	353.1242	C27H31O13	563.177
C16H19O9	355.1035	C26H29O14	565.1563
C16H21O9	357.1191	C26H31O14	567.1719
C17H11O9	359.0409	C26H17O15	569.0573
C19H21O7	361.1293	C27H23O14	571.1093
C19H23O7	363.1449	C26H21O15	573.0886
C18H21O8	365.1242	C25H19O16	575.0679
C17H19O9	367.1035	C25H21O16	577.0835
C15H13O11	369.0463	C25H23O16	579.0992
C17H21O9	369.1191	C27H33O14	581.1876
C16H19O10	371.0984	C26H17O16	585.0522
C16H21O10	373.114	C26H19O16	587.0679
C19H19O8	375.1085	C29H33O13	589.1927
C19H21O8	377.1242	C27H27O15	591.1355
C19H23O8	379.1398	C25H21O17	593.0784
C18H21O9	381.1191	C27H19O16	599.0679
C17H19O10	383.0984	C27H21O16	601.0835
C16H17O11	385.0776	C27H23O16	603.0992
C19H15O9	387.0722	C26H21O17	605.0784
C19H17O9	389.0878	C26H23O17	607.0941
C20H21O8	389.1242	C29H37O14	609.2189
C20H23O8	391.1398	C28H21O16	613.0835
C19H21O9	393.1191	C27H19O17	615.0628
C19H23O9	395.1348	C27H21O17	617.0784
C16H13O12	397.0412	C27H23O17	619.0941
C18H21O10	397.114	C27H25O17	621.1097
C17H19O11	399.0933	C28H21O17	629.0784
C19H13O10	401.0514	C27H21O18	633.0733
C20H19O9	403.1035	C27H23O18	635.089
C20H21O9	405.1191	C29H21O17	641.0784
C20H23O9	407.1348	C29H23O17	643.0941
C19H21O10	409.114	C29H25O17	645.1097
C19H23O10	411.1297	C29H27O17	647.1254
C19H25O10	413.1453	C29H29O17	649.141
C17H19O12	415.0882	C29H31O17	651.1567

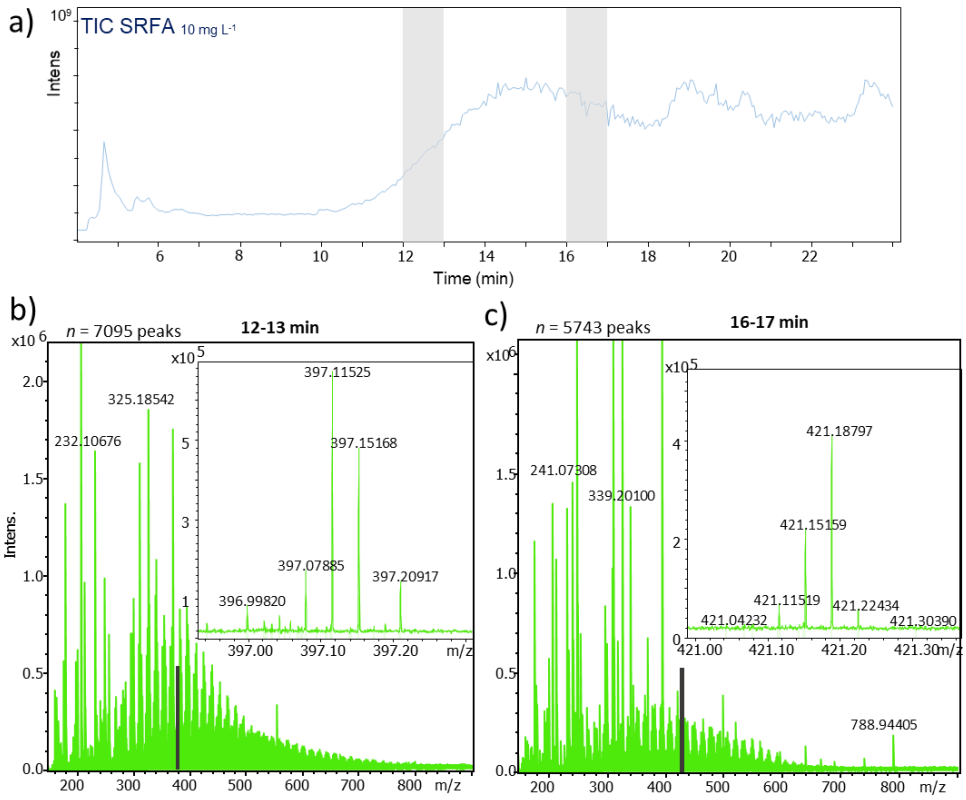


Figure SI 1.1 – a) Total ion chromatogram (TIC) of SRFA injected at 10 mg carbon L^{-1}. The grey areas at 12-13 min and 16-17 min represent the retention times for which averaged mass spectra are displayed below in b and c. The inlet in b and c represents the zoomed mass spectra at b) nominal m/z 397 and c) nominal m/z 421. Note that 2 MW (corresponding to 0.8 s transients with 147 start m/z at 12 T) provides sufficient mass resolution to resolve major DOM species in the considered mass range.

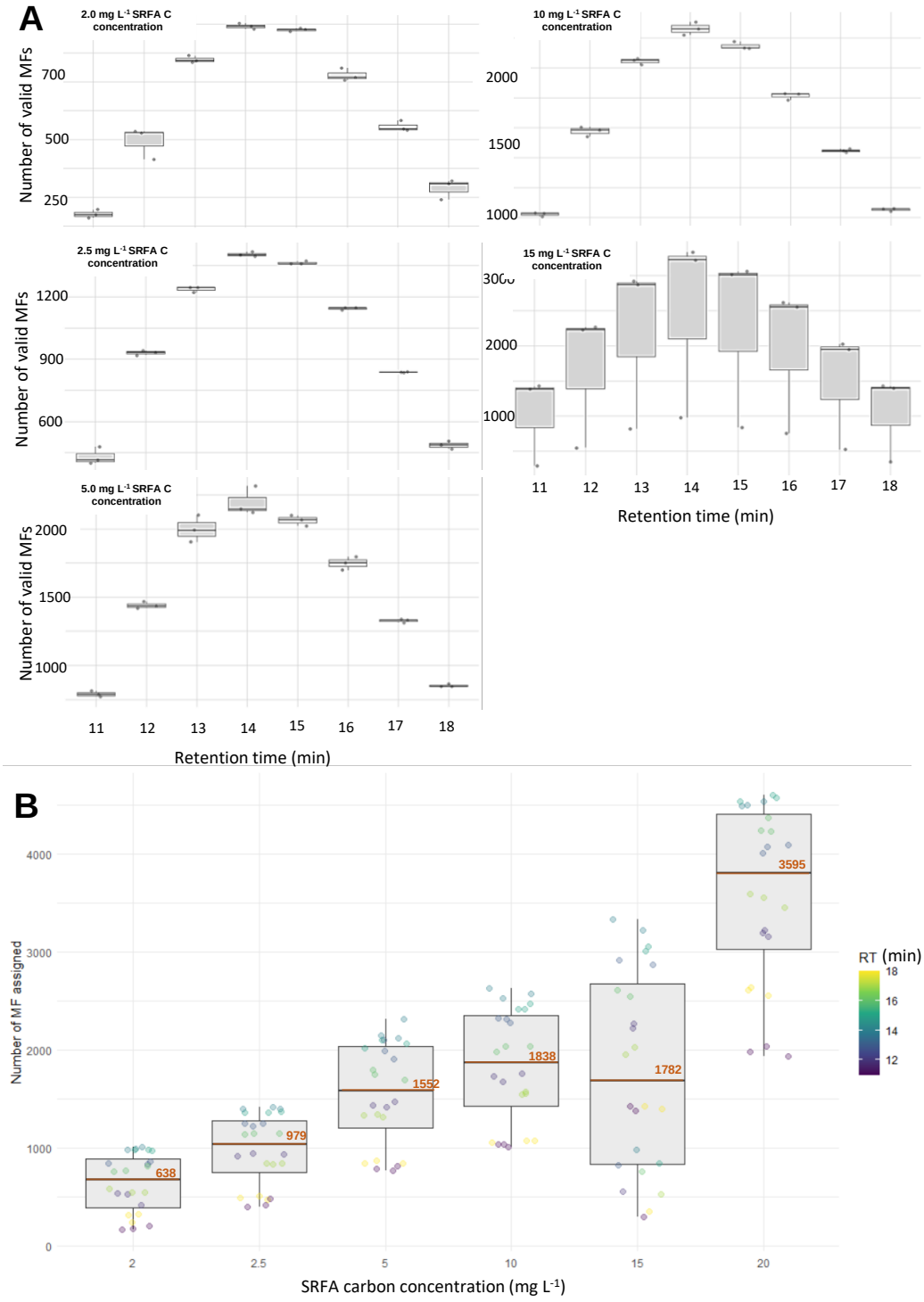


Figure SI 1.2 – Box plot showing the variation on the number of valid molecular formulas (MFs) found across triplicates at a given concentration (2, 2.5, 5, 10, and 15 mg C L^{-1}) split according to a) the processed retention times (RTs), and b) found across triplicates at a given concentration (2, 2.5, 5, 10, 15, and 20 mg C L^{-1}). In a) within a RT (comparing the triplicates), the number of valid MF is highly reproducible. The most prominent variation is observed for 15 mg C L^{-1} probably due to an error during sample management. In b) one dot represents one measurement, therefore, one RT. The variation on the number of valid MFs increased with the concentration (713 MFs @ 2 mg L^{-1}, 1072 MFs @ 2.5 mg L^{-1}, 1690 MFs @ 5 mg L^{-1}, 2007 MFs @ 10 mg L^{-1}, and 1854 MFs @ 15 mg L^{-1}).

Table SI 1.2 – Percentage of the total molecular formulas (MFs) used for coefficient of variance (CV) calculation. Each row is a sum of processed RTs (from 11 to 18 min) of the respective concentration triplicates. For CV calculations, the testable peaks were the MFs shared between the triplicate measurements.

Concentration (mg C L^{-1})	Total number of peaks in the triplicates	Percentage of MFs present in the triplicates (testable peaks) (%)
2	23495	60
2.5	15311	69
5	37238	74
10	44111	75
15	42770	15

Table SI 1.3 – Count of molecular formulas (MFs) used for linear regression model fit for each RT. Each row is the sum of MFs from the SRFA concentrations 2 to 15 mg C L^{-1} in triplicate. For the linear regression assessment, the testable peaks were the MFs found more than 10 times.

Retention time (min)	Number of valid MFs	Number MFs occur > 10 times in all conc (testable peaks)	Percent of MFs (%)
11	18233	3048	17
12	31076	9233	30
13	41971	13213	31
14	47875	15719	33
15	46271	15159	33
16	39972	12495	31
17	30955	8496	27
18	21299	4388	21

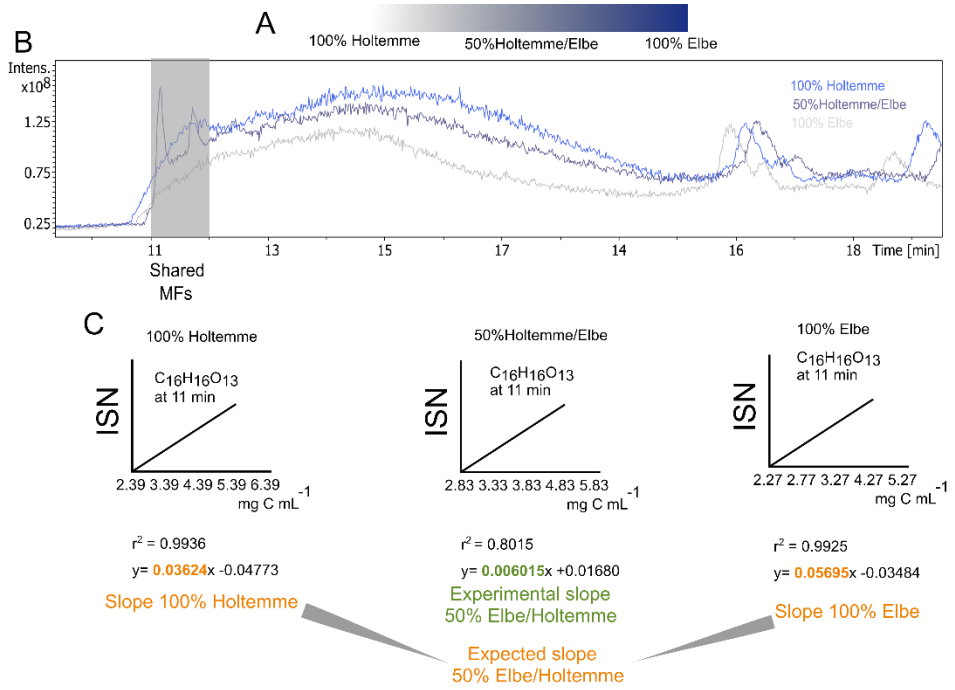


Figure SI 1.3 – A) Environmental gradient between Holtemme spring (100% Holtemme) and Elbe river (100% Elbe) samples. The third sample (50% Elbe/Holtemme) is composed of half volume of each of the two firsts samples. B) Total ion chromatogram of three samples. The grey area in (segment 11-12 min) is an example of how the data was treated: only molecular formulas (MFs) shared between the three samples in a given segment were considered. C) A dilution series of each of the three samples was made and for each of the shared MFs a linear regression analysis was performed. From the linear regression analysis, the expected slope 50% Elbe/Holtemme was calculated from the slopes of the 100% samples.

SI 2 – Performance of PCI-IS for variable concentrations

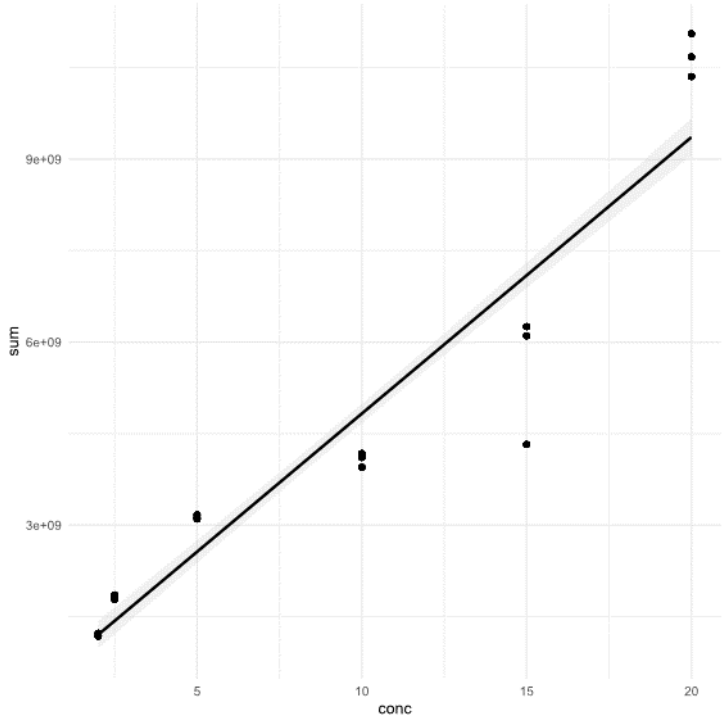


Figure SI 2.1 - Linear regression of sum of all assigned peak intensities versus SRFA carbon concentration (2, 2.5, 5, 10, 15 and, 20 mg C L^{-1})**.**

Table SI 2.1 – Mean, standard deviation (SD) and coefficient of variance (CV) of naproxen-D_3 intensity for the 15 runs shown in Figure 2b in the main text.

Naproxen-D_3 intensity (± SD)	**RT (min)**	**CV (%)**
818384 ± 112532	11	14
1844777 ± 136195	12	7
3176678 ± 124387	13	4
4012133 ± 169294	14	4
4046379 ± 202194	15	5
3550172 ± 171373	16	5
2947751 ± 72978	17	2
2167286 ± 277418	18	13

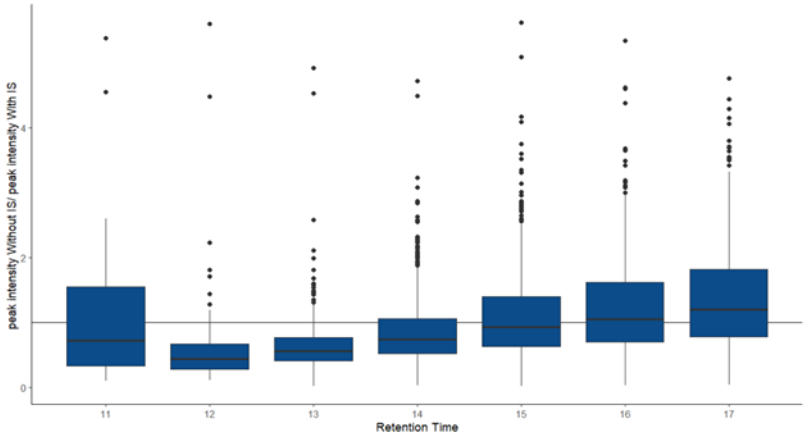


Figure SI 2.2 – Box plot of the ratio between molecular formulas (MFs) absolute peak intensity (RAW) and PCI-IS transformed peak intensities (ISN) for RTs between 11 and 18 min. The shared MFs between the SRFA injected at 2 mg C L^{-1} with and without PCI-IS in the respective RT was used. The y-axis represents the ratio between peak intensity without IS and peak intensity with IS from 1. If the IS does not play an impact in the DOM analytes, the absolute intensity of the sample with and without IS should be similar and the ratio between them close to 1. The deviation from 1 particularly occurred in earlier RTs, indicating that those analytes suffer from enhancement when the PCI-IS is used. As an overall result, the use of PCI-IS slightly impact the ionization of DOC analytes.

SI 3 – Compensation of matrix effects by PCI-IS

Table 3.1 – Linearity data of model compound (MCs)

Compound name	Matrix	RAW			PCI-IS		
		Linear equation	Pearson`s correlation coef. (RAW)	p-value (RAW)	Linear equation	Pearson`s correlation coef. (MC/PCI-IS)	p-value (MC/PCI-IS)
D-glucuronic acid RT 5.05 min	MQW	y = 386215x -2129943	0.9632	8.64E-09	y = 0.2337x - 0.7187	0.9947	3.02E-14
	SRFA 10 mg L^{-1}	y = 385448x + 42075	0.9882	1.22E-08	y = 0.2321x - 0.0266	0.9968	3.51E-11
	Elbe	y = 8112x - 98457	0.9446	3.85E-05	y = 0.0104x + 0.0507	0.9549	1.72E-05
	Strob.	y = 10112x - 91327	0.9816	1.53E-05	y = 0.0139x - 0.0733	0.9874	4.96E-06
2-(4-(2,2-Dicarboxy-ethyl)-2,5-dimethoxy-benzyl)malonic acid RT 14.26 min	MQW	y = 1322192x - 260479	0.9487	7.20E-08	y = 0.0828x - 0.1339	0.9817	9.63E-11
	SRFA 10 mg L^{-1}	y = 902728x-1435520	0.9964	6.12E-11	y = 0.0604x - 0.0002	0.9988	3.48E-13
	Elbe	y = 1266567x-2433241	0.9953	1.55E-13	y = 0.0793x - 0.0838	0.9927	2.11E-12
	Strob.	y = 1210244x-1945558	0.9993	1.16E-18	y = 0.0790x - 0.0990	0.9992	5.27E-18
Fraxin RT 16.03 min	MQW	y = 701776x – 1626112	0.9294	5.45E-07	y = 0.0618x - 0.1609	0.9773	6.89E-12
	SRFA 10 mg L^{-1}	y = 373620x - 226099	0.9982	2.67E-12	y = 0.0489x - 0.0234	0.9978	6.56E-15
	Elbe	y = 541921x - 842505	0.7690	8.06E-04	y = 0.0584x - 0.1421	0.9958	9.64E-17
	Strob.	y = 594537x - 1563531	0.9990	1.37E-17	y = 0.0549x - 0.0874	0.9986	1.29E-11
Isoferulic acid 3-O-β-D-glucuronide RT 15.01 min	MQW	y = 1485234x-2186603	0.8683	2.69E-05	y = 0.1006x - 0.1094	0.9804	1.48E-10
	SRFA 10 mg L^{-1}	y = 1041394x - 631967	0.9995	1.01E-14	y = 0.0748x - 0.0267	0.9990	1.82E-13
	Elbe	y = 1378398x - 1140110	0.9972	6.38E-15	y = 0.0927x - 0.0428	0.9882	3.89E-11
	Strob.	y = 1351206x - 1251206	0.9954	1.37E-13	y = 0.0959x - 0.0899	0.9943	4.90E-13
Leu-enkephalin RT 16.20 min	MQW	y = 470395x - 419901	0.9345	3.38E-07	y = 0.0404x - 0.0505	0.9866	7.05E-08
	SRFA 10 mg L^{-1}	y = 172487x - 81563	0.9866	2.17E-08	y = 0.0215x - 0.0113	0.9825	5.25E-04
	Elbe	y = 81393x - 166258	0.8548	4.91E-05	y = 0.0391x + 0.0320	0.7851	1.69E-07
	Strob.	y = 336556x - 467857	0.9522	1.55E-07	y = 0.0370x - 0.0035	0.9515	1.42E-10
Vanilic acid RT 15.30 min	MQW	y = 106929x - 629983	0.9673	4.03E-09	y = 0.0076x - 0.0227	0.9915	6.46E-13
	SRFA 10 mg L^{-1}	y = 147777x - 188842	0.9969	2.86E-11	y = 0.0078x - 0.0395	0.9995	6.27E-15
	Elbe	y = 157110x - 863815	0.9931	1.72E-13	y = 0.0060x - 0.0349	0.9951	1.85E-14
	Strob.	y = 109103x -1640825	0.9828	3.60E-10	y = 0.0074x + 0.0229	0.9970	3.86E-10

SI 4 – Repeatability and effect of PCI-IS on DOM MFs

Table SI 4.1 - Coefficient of variance (CV) mean of all testable molecular formulas (MFs) at different SRFA carbon concentrations (2.0, 2.5, 5.0, 10, and 15 mg C L^{-1}) and retention time (RT, min) for ISN and RAW.

SRFA concentration (mg C L^{-1})	RT (min)	Mean of CV for all testable MFs	
		ISN	RAW
2.0	**11**	17.5	15.7
	12	15.0	14.2
	13	14.9	13.6
	14	15.0	13.3
	15	15.1	14.1
	16	14.9	14.0
	17	15.5	14.6
	18	17.5	15.7
Mean CV of RTs		15	15
2.5	**11**	14.9	15.2
	12	14.7	15.9
	13	13.7	14.9
	14	13.3	15.0
	15	14.4	14.9
	16	14.0	14.9
	17	14.5	15.4
	18	15.3	17.2
Mean CV of RTs		14	15
5.0	**11**	13.7	14.8
	12	13.2	12.4
	13	12.7	12.2
	14	12.0	11.8
	15	12.2	12.0
	16	14.1	13.0
	17	13.8	13.4
	18	14.9	14.4
Mean CV of RTs		13	13
10	**11**	13.7	13.7
	12	12.0	12.0
	13	12.5	12.0
	14	11.8	12.0
	15	11.8	11.7
	16	12.7	12.3
	17	13.6	13.4
	18	13.5	13.4
Mean CV of RTs		13	13
15	**11**	10.4	12.4
	12	12.3	11.9
	13	13.9	13.4
	14	13.0	12.9
	15	13.1	12.8
	16	11.2	10.8
	17	11.7	11.5
	18	12.7	12.7
Mean CV of RTs		12	12

SI 5 – Linearity of DOM MF peak intensity

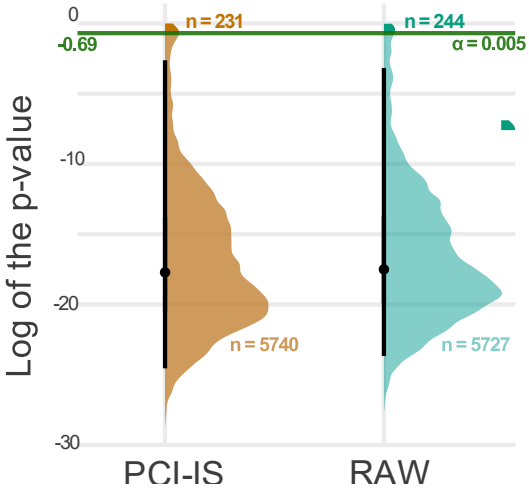


Figure SI 5.1 – Density plot of the *p*-values of all testable molecular formulas (MFs, found more than 10 times) in SRFA (from 2 to 15 mg C L^{-1}) for the tested normalization methods. Log *p*-value of -0.69 represents the 5% confidence interval threshold (orange line).

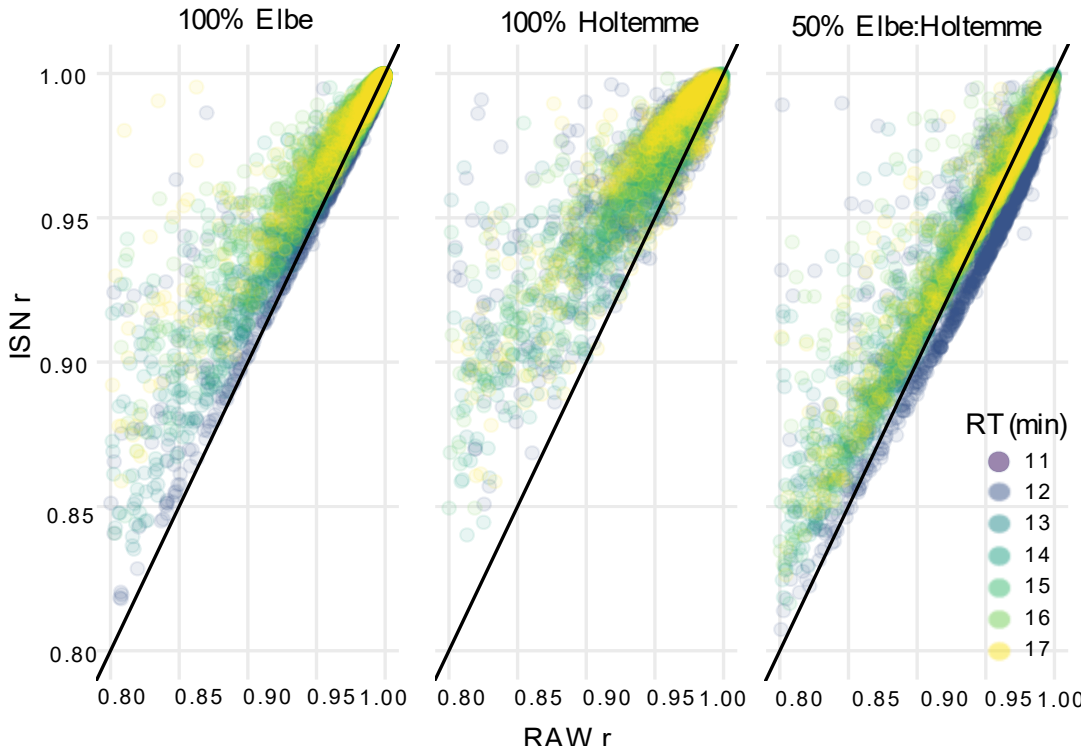


Figure SI 5.2. Pearson's correlation coefficients (r) for individual molecular formulas (MFs) using RAW and ISN for 100% Elbe, 50% Elbe:Holtemme and 100% Holtemme. The black line represents the identity. MFs are colored by the retention time (RT). The negative or low-correlated values correspond to 12% of all MFs. These MFs may correspond to DOM which are structurally more susceptible to suppression or represent contaminants.

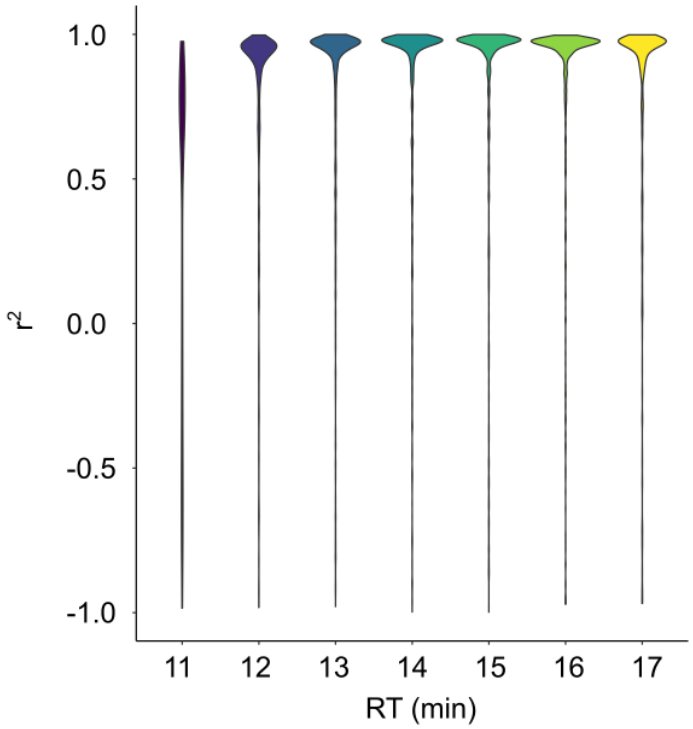


Figure SI 5.3. Squared linear correlation coefficient (r^2) for individual molecular formulas (MFs) using ISN for 100% Elbe, 50% Elbe:Holtemme and 100% Holtemme combined separated by the retention time (RT).

SI 6 – Matrix effects in original, non-extracted freshwater samples

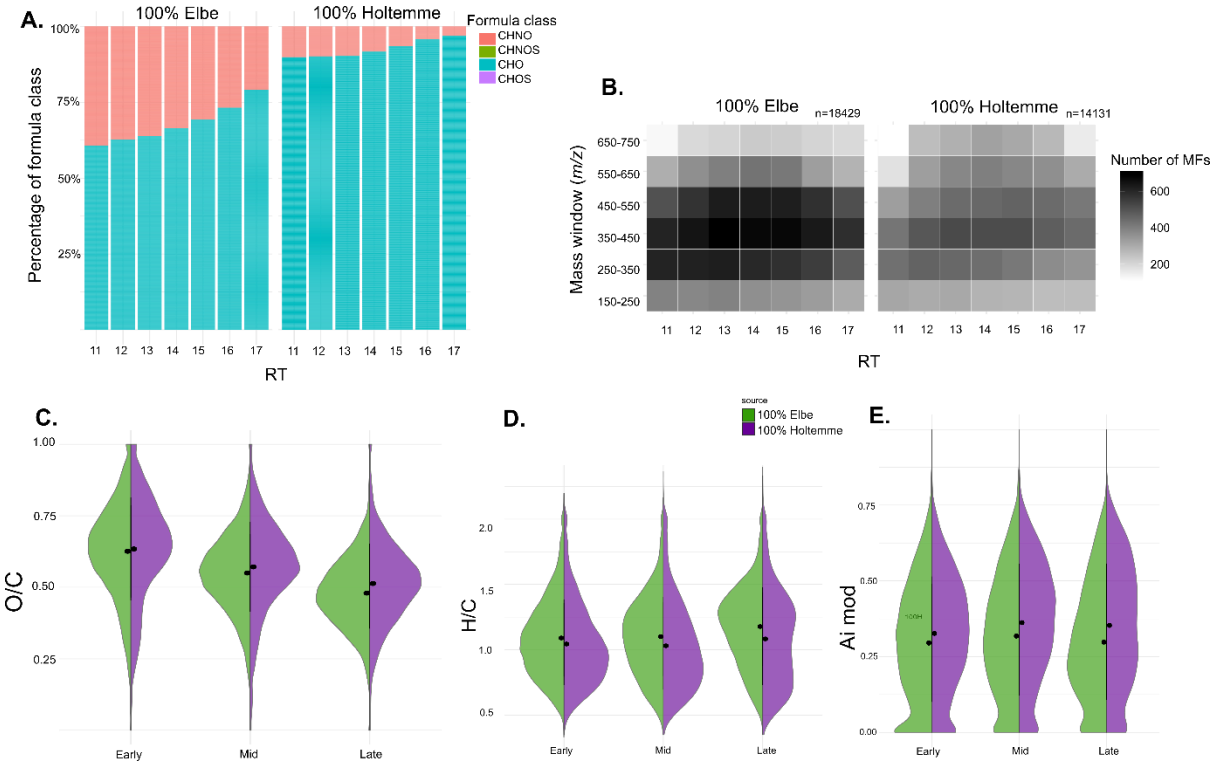


Figure SI 6.1 – Original samples (Holtemme and Elbe chemical composition A) Percentage of molecular formula (MF) classes in each retention time RT segment of both samples: 100% Elbe and 100% Holtemme. B) Assigned MFs in different segments and mass windows. C) O/C, D) H/C and E) Modified aromaticity index $(AI_{mod})^2$ of 100% Elbe (green) and 100% Holtemme (purple) for early (10-11.5 min), mid (10.5-13 min) and late (13-17 min) eluting compounds. The dots in C)-E) represent the mean.

Additional References

1 N. Kamjunke, N. Hertkorn, M. Harir, P. Schmitt-Kopplin, C. Griebler, M. Brauns, W. von Tümpling, M. Weitere and P. Herzsprung, *Water Res.*, 2019, **164**, 114919.

2 A. Zherebker, O. J. Lechtenfeld, A. Sarycheva, Y. Kostyukevich, O. Kharybin, E. I. Fedoros, E. N. Nikolaev and S. I. of K. Paper, *Anal. Chem.*, 2020, **92**, 9032–9038.