Establishing a Measure of Reproducibility of Ultrahigh-Resolution Mass Spectra for Complex Mixtures of Natural Organic Matter

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2012

Impact Factor: 5.64 · DOI: 10.1021/ac3018026 · Source: PubMed

6 AUTHORS, INCLUDING:

Hongmei Chen
Oak Ridge National Laboratory
22 PUBLICATIONS **267** CITATIONS
SEE PROFILE

Andrew Wozniak
Old Dominion University
26 PUBLICATIONS **407** CITATIONS
SEE PROFILE

Amanda S Willoughby
Old Dominion University
8 PUBLICATIONS **46** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

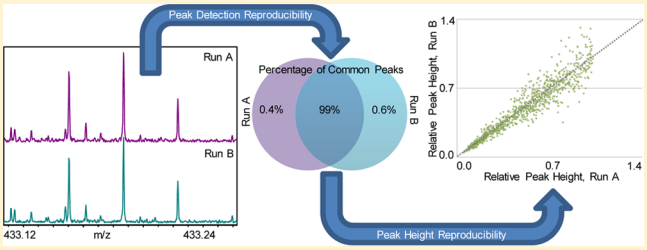
Establishing a Measure of Reproducibility of Ultrahigh-Resolution Mass Spectra for Complex Mixtures of Natural Organic Matter

Rachel L. Sleighter, Hongmei Chen, Andrew S. Wozniak, Amanda S. Willoughby, Paolo Caricasole, and Patrick G. Hatcher*

Department of Chemistry and Biochemistry, Physical Sciences Building, Room 3100, Old Dominion University, 4402 Elkhorn Avenue, Norfolk, Virginia 23529, United States

Supporting Information

ABSTRACT: This study describes a method for evaluating the reproducibility of replicate mass spectra acquired for complex natural organic matter (NOM) samples analyzed by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, with regard to both peak detection and peak magnitude. Because studies of NOM characterization utilize not only the emergence and disappearance of peaks but also changes in relative peak magnitude, it is important to establish that the differences between samples are significantly larger than those between sample replicates. Here, a method is developed for correcting strict signal-to-noise thresholds, along with a new scheme for assessing the reproducibility of peak magnitudes. Furthermore, a systematic approach for discerning when the comparison of samples by the presence or absence of peaks is appropriate and when it is necessary to compare based on the relative magnitude of the peaks is presented. A variety of 10 different types of NOM samples are analyzed in duplicate or triplicate instrumental injections or experimental extractions. A framework for these procedures is provided, and acceptable reproducibility levels are recommended.



Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS) has emerged as a powerful tool for the characterization of natural organic matter (NOM) over the course of the past decade.[1] Its use and application has increased with time, from early papers optimizing spectral acquisition and studies describing how mass defect can be utilized to elucidate differences in NOM[2−4] to the first papers employing molecular formula assignments and visualization diagrams[5−8] and the more recent application of multivariate statistical analysis to examine the large data sets generated from the analysis of many samples.[9−12] However, this tool can still be considered quite immature when compared to more traditional techniques for NOM analysis (e.g., chromatography, bulk elemental analysis, infrared, ultraviolet−visible, and fluorescence spectroscopies, etc.). The utilization of ESI-FTICR-MS has been somewhat restricted, mainly due to the cost and limited instrumental availability and the time/labor intensive data analysis. Additional instruments now exist in laboratories throughout the world, making ESI-FTICR-MS more accessible. Furthermore, with the development of novel software programs and MatLab scripts,[13−15] data analysis has become more standardized and streamlined.

ESI-FTICR-MS has been utilized to characterize nearly all types of NOM samples (i.e., aquatic, aerosol, soil, sediment, etc.) and recent work has focused on correlating this data with other instrumental analyses, such as nuclear magnetic resonance (NMR) spectroscopy[16−19] and excitation emission matrix spectroscopy (EEMS).[20−22] The information obtained from NMR or EEMS can be associated with specific types of peaks (i.e., molecular formulas) in FTICR-MS data, especially those that align in a certain region of the van Krevelen diagram. In such correlations, there is a need to go beyond the traditional reporting of formula presence/absence; the utilization of peak magnitudes for individual samples is necessary. While ESI-FTICR-MS is not currently a quantitative technique (due to ionization efficiency variations among analytes), the peak magnitudes can be compared between samples to observe what types of formulas are detected in the highest relative magnitude. While the presence/absence is appropriate for comparing samples that are compositionally very different, Sleighter et al.[12] highlighted that peak magnitudes were essential for determining differences, based on principal component analysis.

Despite the numerous NOM studies conducted using ESI-FTICR-MS, to our knowledge there is only a single study where the reproducibility of peak magnitudes is assessed.[23] Many studies inherently assume that one FTICR-MS analysis is sufficient for making comparisons and evaluating differences between samples based on the relative magnitudes of peaks. Kido Soule et al.[23] performed an extensive study examining the various instrumental parameters that can affect spectral reproducibility. They found that peak detection and the

Received: July 8, 2012
Accepted: October 17, 2012
Published: October 17, 2012

   dx.doi.org/10.1021/ac3018026 | *Anal. Chem.* 2012, 84, 9184−9191

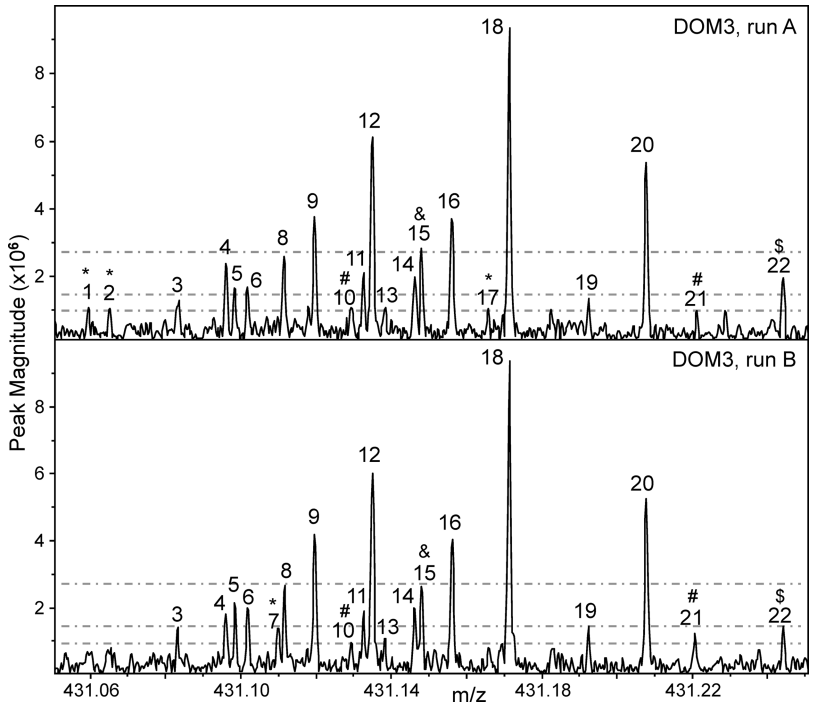


Figure 1. Duplicate mass spectra of DOM3 expanded at a nominal mass of 431. Details of each peak are given in Table 1. Asterisks (∗) indicate peaks that are detected in one spectrum with an S/N ≥ 3 but not detected in the other spectrum. The #, $, and & symbols indicate peaks that are detected in one spectrum with a S/N ≥ 3, ≥ 5, and ≥ 10, respectively, but are detected in the other at S/N ≥ 2.5. Horizontal lines are shown at S/N 3, 5, and 10. The average resolving power at this nominal mass is ≥510 000.

percentage of common peaks were greater in broadband mode than narrow band and that a preliminary assessment of peak magnitudes indicated that these were also more reproducible in broadband mode. Not unexpectedly, they found that more scans provided a greater overlap for peak detection and magnitude reproducibility. This study described the importance of estimating and understanding experimental variability, so that FTICR-MS users can appropriately interpret their data. Most of the emphasis of Kido Soule et al.[23] was on peak detection, with less evaluation of peak magnitude reproducibility. As we advance our understanding of FTICR-MS data for NOM, it is important to not only examine instrumental parameters that impact peak detection reproducibility but also to establish standard limits of reliability when parameters are fully optimized. It is particularly important to have this information for studies that seek to use FTICR-MS to evaluate changes in NOM brought about by environmental processes, many of which currently employ peak magnitude as a proxy for determining differences and changes that occur between samples.

The goal of this manuscript is to establish a standard level of reliability for peak magnitudes using a variety of NOM samples. Prior to assessing peak magnitude reproducibility, it is necessary to investigate the percentage of common peaks in replicate analyses to ensure that sufficient overall spectral reproducibility is achieved. Instrumental replicates and experimental replicates are evaluated. Furthermore, because replicate injections require longer instrument times (as well as higher costs and more time devoted to data analysis), duplicate injections vs triplicate injections are discussed. The problems encountered when using strict signal-to-noise (S/N) thresholds are also described in detail. Because peak magnitude has been shown to be beneficial (if not essential) for statistical analyses of a large number of samples,[12] a new approach to assessing the reproducibility of peak magnitudes is illustrated. To be as comprehensive as possible, 10 NOM sample types, prepared with different isolation/extraction techniques, are investigated. Instrumental parameters are optimized for each sample set and then held constant for replication, so that the reproducibility of peak detection and magnitude can be evaluated, allowing for recommendations to be made for future studies.

EXPERIMENTAL SECTION

Sample Description. Numerous dissolved organic matter (DOM) and NOM samples are utilized in order to evaluate mass spectral reproducibility under optimized spectral conditions. Table S-1 in the Supporting Information shows the abbreviations used throughout the manuscript, describes the preparation of the samples, and explains how they were analyzed by ESI-FTICR-MS. Instrumental replicates are multiple injections of the same sample, while experimental replicates are separate aliquots of the same original sample extracted individually for analysis.

ESI-FTICR-MS Analysis. Samples were analyzed in negative ion mode using broadband acquisition from 200 to 2000 m/z. Samples were infused by a syringe pump at 120 μL/h into an ApolloII ESI ion source of a Bruker Daltonics 12 T ApexQe FTICR-MS. Spray voltages and parameters were optimized for each sample set to maintain stable, consistent ion currents. Ions were accumulated in a hexapole (as specified in Table S-1 in the Supporting Information) before being transferred to the ICR cell, where coadded transients were collected with a 4 MWord time domain. The summed free induction decay signal was zero-filled once and sine-bell apodized prior to fast Fourier transformation and magnitude calculation using Bruker Daltonics Data Analysis software.

dx.doi.org/10.1021/ac3018026 | Anal. Chem. 2012, 84, 9184−9191

Table 1. Details of the Numbered Peaks in Figure 1, Labeled with the Corresponding Symbols[a]

	DOM3, run A				DOM3, run B			
peak no.	m/z	S/N	magnitude	ratio	m/z	S/N	magnitude	ratio
1[b]	431.05948	3.2	1 017 999		nd	nd	nd	nd
2[b]	431.06511	3.1	990 607		nd	nd	nd	nd
3	431.08348	4.0	1 213 714	0.13	431.08314	4.8	1 424 626	0.15
4	431.09607	8.7	2 359 572	0.25	431.09596	6.4	1 812 980	0.19
5	431.09824	5.7	1 634 836	0.18	431.09844	7.9	2 177 012	0.23
6	431.10171	5.7	1 616 660	0.17	431.10183	7.3	2 025 205	0.22
7[b]	nd	nd	nd	nd	431.10977	4.9	1 433 334	
8	431.11140	9.4	2 532 886	0.27	431.11152	9.9	2 664 950	0.29
9	431.11955	14.2	3 676 951	0.40	431.11958	16.3	4 220 919	0.45
10[c]	431.12921	3.2	1 023 960	0.11	431.12924	2.9	967 544	0.10
11	431.13248	7.3	2 027 288	0.22	431.13247	6.7	1 894 393	0.20
12	431.13484	24.4	6 155 545	0.66	431.13489	23.6	6 016 505	0.64
13	431.13821	3.2	1 013 273	0.11	431.13815	3.6	1 123 322	0.12
14	431.14613	7.0	1 939 610	0.21	431.14608	7.3	2030587	0.22
15[e]	431.14788	10.3	2 744 602	0.29	431.14796	9.9	2673659	0.29
16	431.15594	14.2	3 692 316	0.40	431.15600	15.7	4073468	0.44
17[b]	431.16566	3.0	985 629		nd	nd	nd	nd
18	431.17119	37.4	9 307 421	1.00	431.17121	37.3	9 339 390	1.00
19	431.19222	4.3	1 282 848	0.14	431.19215	4.9	1 447 682	0.16
20	431.20750	20.9	5 310 755	0.57	431.20749	20.5	5 259 780	0.56
21[c]	431.22095	2.8	934 052	0.10	431.22051	4.0	1 223 430	0.13
22[d]	431.24382	6.8	1 893 671	0.20	431.24384	4.9	1 444 617	0.15

[a]Ratio is the relative peak magnitude (ratio of the peak's magnitude to the base peak (largest peak at each nominal m/z, peak 18), as described in Reproducibility of Peak Magnitudes). nd, not detected. [b]*, Peaks detected in only one spectrum with a S/N ≥ 3 but not detected in the other spectrum. [c]#, Peaks detected in only one spectrum with a S/N ≥ 3 but detected in the other spectrum at S/N ≥ 2.5. [d]$, Peaks detected in only one spectrum with a S/N ≥ 5 but detected in the other spectrum at S/N ≥ 2.5. [e]&, Peaks detected in only one spectrum with a S/N ≥ 10 but detected in the other spectrum at S/N ≥ 2.5.

Table 2. Total Number of Peaks in Each Set of Replicate Analyses and the Percentage of Peaks That Are Common to Each Set, At Strict S/N Thresholds of 3, 5, and 10 (White) and after Inclusion of Threshold-Corrected Peaks (Gray)

Sample	Total Number of Peaks in Replicates			Number of Common Peaks (% of total)		
	S/N≥3	S/N≥5	S/N≥10	S/N≥3	S/N≥5	S/N≥10
DOM1, RO/ED[a]	2467	1460	729	1759 (71%)	1216 (83%)	637 (87%)
				+163 (78%)	+226 (99%)	+92 (100%)
DOM2, RO/ED[a]	2768	1692	833	2030 (73%)	1393 (82%)	751 (90%)
				+195 (80%)	+279 (99%)	+82 (100%)
DS-WW, same day[b]	4997	3006	1299	3089 (62%)	2129 (71%)	1044 (80%)
				+291 (68%)	+618 (91%)	+235 (98%)
DS-WW, different days[a]	4677	2802	1218	2685 (57%)	1796 (64%)	890 (73%)
				+267 (63%)	+497 (82%)	+259 (94%)
DOM3, PPL[c]	3255	1927	907	2280 (70%)	1538 (80%)	748 (82%)
				+233 (77%)	+342 (98%)	+158 (99.9%)
DOM4, PPL[c]	3487	2112	1017	2458 (70%)	1691 (80%)	876 (86%)
				+209 (76%)	+346 (96%)	+136 (99.5%)
Water Soluble Aerosol[c]	3464	1735	443	2172 (63%)	1163 (67%)	266 (60%)
				+308 (72%)	+509 (96%)	+172 (99%)
Pyridine Soluble Aerosol[c]	3038	1439	607	1941 (64%)	1066 (74%)	483 (80%)
				+307 (74%)	+337 (97%)	+117 (99%)
Poultry microcosm[c]	2088	1040	436	1286 (62%)	773 (74%)	332 (76%)
				+194 (71%)	+239 (97%)	+104 (100%)
Hairy Vetch microcosm[c]	2828	1474	664	1573 (56%)	1035 (70%)	526 (79%)
				+179 (62%)	+213 (85%)	+87 (92%)

[a]Instrumental duplicates. [b]Instrumental triplicates. [c]Experimental duplicates.

Data Analysis. Mass spectra were externally calibrated using a polyethylene glycol standard and internally calibrated using naturally present homologous series detected within the spectra,[24] ensuring that at least 1 calibrant peak existed every 28 m/z units (i.e., every other CH_2) at 200−600 m/z. The mass lists exported for each mass spectrum were constructed using

dx.doi.org/10.1021/ac3018026 | Anal. Chem. 2012, 84, 9184−9191

S/N ≥ 2.5 for m/z 300−600 (see the Supporting Information for more details). Peaks that were detected in solvent blanks and/or extraction blanks that were also detected in those samples' spectra (within 0.5 mDa) were removed from further consideration. For the purposes of comparing peaks detected in replicates, 3 different S/N thresholds were explored (3, 5, and 10) to test the effect of S/N on peak detection and relative magnitude.

For evaluating peak detection reproducibility, peak lists are compared at S/N thresholds of 3, 5, and 10. The first step is to determine the percentage of peaks common to replicate analyses of each sample at each S/N. This is accomplished using a MatLab script that aligns the replicate peak lists and considers peaks to be the same if the m/z values exist within 0.5 mDa of each other. Once the percentage of common peaks for each sample set is evaluated at S/N 3, 5, and 10, peaks that were detected in 1 replicate but not the other(s) were searched for in the other m/z list(s) using a minimum S/N value of 2.5 (the lowest S/N value where a well-defined signal can be distinguished from the noise). If the peak was found in the peak list of the other spectrum (or spectra) at S/N ≥ 2.5, the peak was considered to be present in all spectra. These peaks are referred to as "threshold-corrected peaks". In this way, the issue of well-defined peaks present below the S/N threshold was handled, and a more representative percentage of reproducibility is calculated for the presence/absence of peaks among replicates.

RESULTS AND DISCUSSION

Reproducibility of Peak Detection. The acquired mass spectra are similar to those previously published, with peaks existing at 200−700 m/z.[1,8,9,14−16,25,26] Complexity is apparent, with clusters of peaks existing at every nominal mass (Figure 1). Establishing levels of acceptable reproducibility of peaks and their magnitudes is the objective, and thus assigning molecular formulas is beyond the scope of this paper. Figure 1 highlights the threshold issues encountered when using strict S/N cutoffs, and the details of each peak are given in Table 1. Peak 10 is detected in DOM3 run A with a S/N of 3.2 and run B with a S/N of 2.9. Thus, if a mass list is constructed using only peaks with S/N ≥ 3, peak 10 would not be included in run B despite the presence of a well-defined peak. The same is observed for peaks 15, 21, and 22 at S/N thresholds of 10, 3, and 5, respectively. As a result, using these arbitrary S/N thresholds is not adequate for establishing peak detection reproducibility. A percentage of common peaks calculated using strict S/N thresholds will not give a realistic view of the consistency between replicates because well-defined peaks go undetected. Therefore, mass lists were exported using S/N ≥ 2.5 to correct the threshold problem when determining the percentage of peaks common to replicate analyses, as described above.

Table 2 shows the total number of peaks for each of the NOM samples analyzed in replicate at S/N ≥ 3, ≥ 5, and ≥ 10. The number and percentage of peaks that are common to each set of replicates using strict S/N thresholds are given, with the additional common peaks after threshold correction (e.g., peak 10 in Figure 1 and Table 1) highlighted in gray. For the 10 NOM sample sets examined here, the average percentage of common peaks for S/N 3, 5, and 10 are 65% (range of 56−73%), 75% (range of 64−83%), and 79% (range of 60−90%), respectively (Table 2). With the inclusion of the threshold-corrected peaks, the average values of common peak detection for S/N 3, 5, and 10 rise to 72% (range of 62−80%), 94% (range of 82−99%), and 98% (range of 92−100%), respectively. The threshold-corrected peaks increased the total percentage of common peaks by 6−10% (7.3% on average) at S/N ≥ 3. At S/N ≥ 5 and ≥ 10, the ranges are 16−23% (18.5% on average) and 10−39% (18.8% on average), respectively. As expected, peaks detected in only one replicate at higher S/N thresholds (S/N 5 or 10) are more likely to be counted as common after threshold correction than when S/N 3 is utilized. Thus, threshold correction results in larger percentage increases for S/N 5 and 10.

The common peak percentages reported here are similar to those in the few studies that have addressed the reproducibility of FTICR-MS peak detection. The percentage of common peaks in negative ion broadband analyses performed by Kido Soule et al.[23] were 68−87% at the reported S/N ≥ 5 threshold. Approximately 83% of peaks, accounting for 97% of total spectral magnitude, were common to effluent DOM replicates desalted using electrodialysis.[27] Percentages in the range of 62−65% and 87−92% at S/N ≥ 3 and S/N ≥ 5, respectively, were reported for replicate analyses of C_{18} extracted waters from the Lake Superior watershed and Suwannee River fulvic acid.[28]

DOM1 and DOM2, instrumental duplicates from reverse osmosis/electrodialysis (RO/ED) isolations of North Pacific and North Atlantic seawater, give reproducibilities (i.e., common peak percentages of 78−80%, 99%, and 100% at S/N 3, 5, and 10, respectively) that are very similar to DOM3 and DOM4 (76−77%, 96−98%, and 99.5−99.8% at S/N 3, 5, and 10, respectively, Table 2), which are experimental duplicates from PPL extracts of coastal Pacific Ocean seawater. Instrumental triplicates of DS-WW analyzed on the same day give reproducibilities (68%, 91%, and 98% at S/N 3, 5, and 10, respectively) that are slightly lower than that of DOM1−4, implying that duplicate analyses give higher percentages of common peaks (as expected). The DS-WW instrumental duplicates analyzed 12 days apart give percentages (63%, 82%, and 94% at S/N 3, 5, and 10, respectively) that are lower than that of the same day analyses, indicating that instrumental analysis of sample sets on different days may be an important factor to consider during data analysis. The replicate analyses of the aerosol (72−74%, 96−97%, and 99% at S/N 3, 5, and 10, respectively) and microcosm (62−71%, 85−97%, and 92−100% at S/N 3, 5, and 10, respectively) samples also give commonalities similar to the other DOM samples. In general, the experimental duplicates yield common peak percentages that are only slightly lower than that of instrumental duplicates, with the instrumental triplicates giving slightly lower reproducibilities. It is clear from the NOM samples examined that correcting for a strict S/N threshold is important for obtaining a more representative percentage of common peaks (here, we correct at S/N ≥ 3, ≥ 5, and ≥ 10 using a S/N minimum of 2.5). Failing to threshold correct spectra may exclude peaks from analysis because their S/N values are only slightly lower than the threshold but in actuality are present.

The average percentage of post-threshold-corrected common peaks at S/N ≥ 3 for all 10 sample sets analyzed here is 72%. For comparison, when evaluating the peaks detected in the poultry manure microcosm run A to the hairy vetch microcosm run A (2 samples that are compositionally different), the percentage of common peaks at strict thresholds of S/N ≥ 3, ≥ 5, and ≥ 10 is 23%, 21%, and 14%, respectively, and 28%, 39%, and 49%, respectively, after threshold-correction. These values are significantly lower than the common peak percentages for the replicates of each substrate's microcosms (Table 2),

dx.doi.org/10.1021/ac3018026 | Anal. Chem. 2012, 84, 9184−9191

Table 3. Average Percent Relative Standard Deviation (RSD) of the Peak Magnitude Ratios in Each Set of Replicate FTICR-MS Analyses As Well As the Percentage of Those Common Peaks Whose Peak Magnitude Ratios Fall within 10% RSD and 25% RSD, at S/N ≥ 3, S/N ≥ 5, and S/N ≥ 10

	avg % RSD for peak magnitude ratios			percent of peaks within 10% RSD			percent of peaks within 25% RSD		
sample	S/N ≥ 3	S/N ≥ 5	S/N ≥ 10	S/N ≥ 3	S/N ≥ 5	S/N ≥ 10	S/N ≥ 3	S/N ≥ 5	S/N ≥ 10
DOM1, RO/ED[a]	8%	7%	4%	66%	67%	89%	95%	98%	99.8%
DOM2, RO/ED[a]	8%	6%	4%	75%	77%	89%	95%	99%	99.9%
DS-WW, same day[b]	10%	8%	6%	57%	68%	83%	96%	99%	99.9%
DS-WW, different days[a]	14%	12%	10%	50%	57%	64%	81%	86%	89%
DOM3, PPL[c]	10%	7%	4%	61%	71%	85%	92%	97%	99.7%
DOM4, PPL[c]	9%	7%	4%	65%	73%	87%	94%	97%	99%
water-soluble aerosol[c]	12%	9%	5%	53%	63%	82%	87%	93%	99%
pyridine soluble aerosol[c]	11%	8%	5%	57%	70%	81%	89%	96%	98%
poultry microcosm[c]	9%	6%	3%	64%	75%	94%	94%	98%	100%
hairy vetch microcosm[c]	8%	6%	4%	66%	78%	90%	95%	98%	99%

[a]Instrumental duplicates. [b]Instrumental triplicates. [c]Experimental duplicates.

indicating that the replicates are sufficiently reproducible at each level of S/N to compare differences between the two NOM substrates. On the basis of the analyses of these 10 sample sets, we suggest 67% as a minimum for the percentage of post-threshold-correction common peaks, indicating that acceptable reproducibility is obtained when $^2/_3$ of the peaks are common to instrumental or experimental replicates of the same sample. This recommended percentage is slightly lower than the average of these NOM samples given above but is consistent with the few previous studies that report on reproducibility.[23,27,28]

In the case where samples are substantially different from one another, using presence/absence as a mode of comparison is adequate, without considering relative peak magnitudes. However, when comparing DOM1 run A to DOM2 run A (seawaters desalted by RO/ED), the percentage of common peaks at S/N ≥ 3, ≥ 5, and ≥ 10 is 56%, 61%, and 59%, respectively, and 62%, 84%, and 96%, respectively, after threshold-correction. While these reproducibility values are still lower than those for each set of replicates of DOM1 and DOM2, these 2 sample sets are more similar to one another than the example of varying microcosm substrates. When comparing the peaks common to DOM1 runs A and B to the common peaks in DOM2 runs A and B, the percentage of threshold-corrected common peaks at S/N 3, 5, and 10 is 59%, 62%, and 59%, respectively. The similarity of these oceanic DOM samples is not surprising, since the majority of marine DOM is thought to be dominated by a refractory DOM pool that is homogeneously distributed across different oceanic areas and depths.[16,29] In this instance, peak magnitude is essential for discerning the less obvious differences between these two different seawater DOM samples, but only if that peak magnitude is reproducible from one analysis to the next.

Reproducibility of Peak Magnitudes. Relative peak magnitude is typically reported as either the percent of the total spectral magnitude (i.e., the sum of the magnitudes of all peaks) or as the percent of the base peak (i.e., the largest peak detected in the spectrum). Here, we present each peak's relative magnitude as a ratio of its magnitude to the tallest peak at its specific nominal mass. For example, in Figure 1 peak 18 is the largest peak at nominal mass 431 (and represents the base peak at this nominal mass), and thus the relative magnitudes of the other peaks detected at nominal mass 431 are the ratio of their magnitude to the magnitude of peak 18 (Table 1). This calculation was done for every nominal mass in the spectrum. In the case of Figure 1 (and for the vast majority of nominal masses for all sample sets in this study), the tallest peak at each nominal mass is the same in all replicate analyses. In the few instances where the tallest peak at the nominal mass differs between replicates, the base peak for the first spectrum analyzed is chosen as the base peak at that nominal mass for all replicates, ensuring that the ratios are calculated consistently from one replicate to the next. Thus, in a few rare cases, the ratio of one peak to the base peak at that nominal mass can be greater than 1.

In order to compare the calculated ratios between replicates of the same sample, only the peaks common to all replicates can be compared. Consequently, when a peak was detected in one spectrum but not in the replicate, the ratio is not calculated (Table 1). Once the ratios at each nominal mass are calculated (representing the relative peak magnitude of the common peaks), the average, standard deviation, and percent relative standard deviation (RSD, standard deviation divided by the average of the relative peak magnitude) is calculated for the set of replicates. Table 3 displays the average percent RSD for each sample set at S/N ≥ 3, ≥ 5, and ≥ 10 as well as the percentage of peaks within 10% RSD and 25% RSD at each S/N threshold.

The average percent RSD decreases as S/N increases, and the percentage of peaks within 10% or 25% RSD increases as S/N increases, indicating that larger peaks have more reproducible magnitudes than smaller peaks. For the 10 sample sets examined, the average percent RSD at S/N 3, 5, and 10 are 10% (range of 8−14%), 8% (range of 6−12%), and 5% (range of 4−10%), respectively (Table 3). The percentage of peaks within 10% RSD at S/N 3, 5, and 10 are 61% (range of 50−75%), 70% (range of 57−78%), and 84% (range of 64−94%), while the percentage of peaks within 25% RSD at S/N 3, 5, and 10 are 92% (range of 81−96%), 96% (range of 86−99%), and 98% (range of 89−100%), respectively.

At S/N 3, 5, and 10, DOM1 has average percent RSDs of 8%, 7%, and 4%, respectively, while DOM2 has average percent RSDs of 8%, 6%, and 4%, respectively (Table 3). These values are quite similar to DOM3 (average percent RSDs of 10%, 7%, and 4% at S/N 3, 5, and 10, respectively) and DOM4 (average percent RSDs of 9%, 7%, and 4% at S/N 3, 5, and 10, respectively). Instrumental triplicates of DS-WW analyzed on the same day give average percent RSDs (10%, 8%, and 6% at S/N 3, 5, and 10, respectively) that are slightly higher than that of DOM1−4. This, again, implies that duplicates perform better than triplicates and give lower average RSD percentages (as

dx.doi.org/10.1021/ac3018026 | Anal. Chem. 2012, 84, 9184−9191

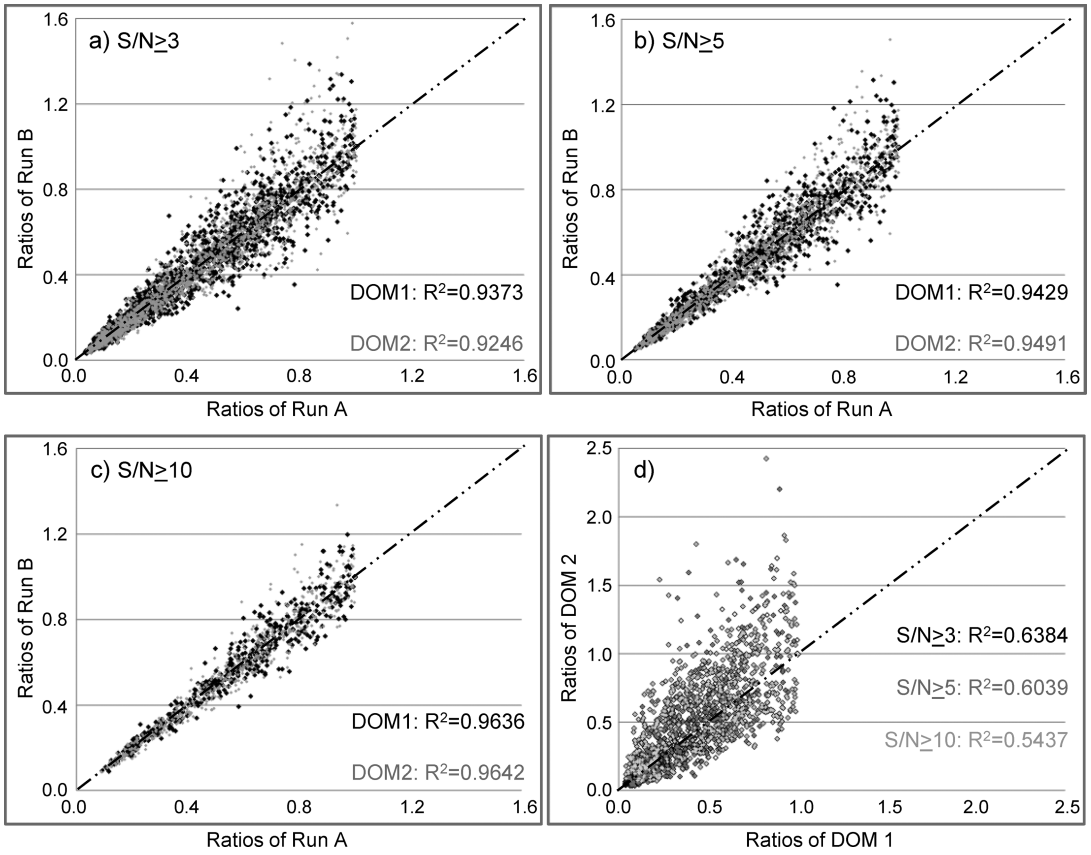


Figure 2. Ratios of peak magnitudes (the ratio of the peak's magnitude to the magnitude of the base peak at each nominal m/z) at (a) S/N ≥ 3, (b) S/N ≥ 5, and (c) S/N ≥ 10 of Run A plotted against Run B for DOM1 (black points) and DOM2 (gray points), highlighting better fits (higher R^2) at higher S/N thresholds. When different DOM samples are compared, peak magnitudes are not reproducible, as in part d where the ratios of the common peaks detected in both DOM1 and DOM2 (each run A) are compared at S/N ≥ 3 (open black points), S/N ≥ 5 (solid dark gray points), and S/N ≥ 10 (light gray × symbols), and show decreasing R^2 fits. The 1:1 line is dashed on each plot.

expected based on the percentage of common peaks). The DS-WW instrumental duplicates analyzed 12 days apart give average percent RSDs of 14%, 12%, and 10% at S/N 3, 5, and 10, respectively, which are higher than that of the same day triplicates. At S/N 3, 5, and 10, the replicates of the aerosol (11−12%, 8−9%, and 5%, respectively) and microcosm (8−9%, 6%, and 3−4%, respectively) samples also give average percent RSDs similar to the other DOM samples. In general, the instrumental and experimental duplicates yield similar average percent RSDs for the common peaks detected during replication.

On the basis of the FTICR-MS analyses of these 10 NOM sample sets, we suggest 10% as a maximum average percent RSD and that a minimum of 60% and 90% of peaks fall within 10% and 25% RSD, respectively, at S/N ≥ 3. These recommended values are the average or slightly better than the average for the NOM samples analyzed here at S/N ≥ 3. For comparison, when evaluating the common peaks detected in the poultry manure microcosm run A to the hairy vetch microcosm run A (two samples that were determined to be compositionally different in Reproducibility of Peak Detection), the average percent RSD values at S/N ≥ 3, ≥ 5, and ≥ 10 are 49%, 47% and 40%, respectively. These values are significantly higher than the average percent RSDs of the replicate analyses shown in Table 3. Furthermore, when comparing peak ratios for common peaks detected in the poultry and hairy vetch microcosms (each run A) only 35% and 42%, 39% and 46%, and 54% and 58% fall within 10% and 25% RSD at S/N thresholds of 3, 5, and 10, respectively. The values given here for the two compositionally different microcosm samples are much lower than the suggested acceptable values of reproducibility given above, further proving that these samples can be compared sufficiently on a presence/absence basis.

When comparing DOM1 run A to DOM2 run A (seawaters desalted by RO/ED that gave common peak percentages of 62%, 84%, and 96% at S/N ≥ 3, ≥ 5, and ≥ 10, respectively, Table 2), the average percent RSD values at S/N ≥ 3, ≥ 5, and ≥ 10 are 24%, 24%, and 25%, with 32% and 59%, 33%, and 57%, and 34% and 57% of peaks falling within 10% and 25% RSD, respectively (Table 3). Figure 2a−c shows a clear depiction of the enhanced reproducibility of the peak magnitude ratios for the replicates of DOM1 and the replicates of DOM2 (at S/N ≥ 3, ≥ 5, and ≥ 10), differing from that of the peak magnitude ratios of DOM1 and DOM2 compared to each other (Figure 2d). As portrayed in Table 3 and Figure 2, peak magnitudes are more reproducible at higher S/N, as the distribution of points becomes narrower and more closely follows the 1:1 line as one progresses from part a to part b to part c of Figure 2. The fit and distribution of DOM1 run A to DOM2 run A are visibly less reproducible than those of actual sample replicates (Figure 2d). While these two samples appear to be fairly similar based on their common peak percentages, when this new method of comparing relative peak magnitudes is utilized, it is clear that these samples are compositionally different based on their peak magnitude ratios. In Figure 2d, the R^2 value decreases as S/N increases, which is opposite from the

dx.doi.org/10.1021/ac3018026 | Anal. Chem. 2012, 84, 9184−9191

trend observed in Figure 2a–c. This is further evidence that DOM1 differs significantly from DOM2, because the largest peaks in each spectrum are more different than those with lower S/N. Thus relative peak magnitude can be exploited in order to discern the subtle differences that may not be observed otherwise.

CONCLUSIONS

Reproducibility of peaks and their magnitude is an important consideration, as many studies of complex NOM samples rely upon the appearance and disappearance of peaks for characterization. This study illustrates the need for threshold-correcting S/N when calculating the percentage of peaks common to replicates and evaluating the overall reproducibility of replicate mass spectra. For replicates to be considered reproducible, we suggest that a minimum of 67% of threshold-corrected peaks be common to replicates. If peak detection reproducibility is significantly higher for replicates than different samples, then using presence/absence as a mode of comparison is sufficient. However, if different samples closely resemble one another, then peak magnitude is necessary to evaluate the variation or changes that occur between samples. In this case, the ratio of each peaks' magnitude to the base peak at each nominal mass can be utilized to evaluate sample differences. When comparing the calculated ratios in replicates, a maximum percent RSD of 10% is suggested, with at least 60% and 90% of peaks being within 10% and 25% RSD, respectively.

Subsequent data analysis involves the assignment of molecular formulas to the detected peaks. Analyzing samples in duplicate and establishing that the replicates are reproducible will allow a more reliable assessment of similarities/differences among samples. Assigning formulas only to those peaks that fall within the reproducibility measurements outlined here will ensure that variations between samples are, in fact, real and environmentally significant. As the NOM community moves toward advancing the quantitative (or even semiquantitative) capabilities of ESI-FTICR-MS, reproducibility becomes even more vital, especially with regard to peak magnitude. While the goals of each NOM study can vary substantially, the routines conducted here provides an outline for how to evaluate reproducibility for replicate analyses.

ASSOCIATED CONTENT

Supporting Information

Expanded experimental section and Table S-1 as referenced in the manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 757-683-6537. Fax: 757-683-4628. E-mail: phatcher@odu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Susan Hatcher and Jared Callan at the COSMIC facility at ODU for their assistance with the ESI-FTICR-MS analyses, and we thank Aron Stubbins and Thorsten Dittmar for useful discussions regarding our approach to assessing reproducibility. We also thank two anonymous reviewers for their comments that improved the quality of this manuscript. This work was funded in part by the Batten Endowment Fund at ODU. We thank Aron Stubbins, E. Michael Perdue, and Nelson W. Green for collecting water and processing the RO/ED DOM, the researchers and crew on the *R/V Oceanus* during Cruise OC449-3 in September 2008, and the crew at NELHA for sampling during September 2009. The RO/ED work was funded by the NSF Chemical Oceanography Program (Grant Number OCE-0728634). We also thank John Pohlman at the U.S.G.S. for the marine near seep samples. A. S. Willoughby was funded by the U.S. Environmental Protection Agency (EPA) Science to Achieve Results (STAR) graduate fellowship program, Grant FP 91736001-0. The EPA has not officially endorsed this publication and the views expressed herein may not reflect the views of the EPA. The microcosm biodegradation work was supported by the USDA-National Research Initiative Competitive Grant Number 2008-35107-04480.

REFERENCES

(1) Sleighter, R. L.; Hatcher, P. G. In *Fourier Transforms- Approach to Scientific Principles*; Nikolic, G., Ed.; InTech: New York, 2011; pp 295–320. Available from http://www.intechopen.com/articles/show/title/fourier-transform-mass-spectrometry-for-the-molecular-level-characterization-of-natural-organic-matt

(2) Brown, T. L.; Rice, J. A. *Anal. Chem.* **2000**, *72*, 384–390.

(3) Kujawinski, E. B.; Freitas, M. A.; Zang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. *Org. Geochem.* **2002**, *33*, 171–180.

(4) Stenson, A. C.; Landinge, W. M.; Marshall, A. G.; Cooper, W. T. *Anal. Chem.* **2002**, *74*, 4397–4409.

(5) Kim, S.; Kramer, R. W.; Hatcher, P. G. *Anal. Chem.* **2003**, *75*, 5336–5344.

(6) Stenson, A. C.; Marshall, A. G.; Cooper, W. T. *Anal. Chem.* **2003**, *75*, 1275–1284.

(7) Koch, B. P.; Dittmar, T.; Witt, M.; Kattner, G. *Anal. Chem.* **2007**, *79*, 1758–1763.

(8) Reemtsma, T. *J. Mass Spectrom.* **2010**, *45*, 382–390.

(9) Kujawinski, E. B.; Longnecker, K.; Blough, N. V.; Del Vecchio, R.; Finlay, L.; Kitner, J. B.; Giovannoni, S. J. *Geochim. Cosmochim. Acta* **2009**, *73*, 4384–4399.

(10) Hur, M.; Yeo, I.; Park, E.; Kim, Y.; Yoo, J.; Kim, E.; No, M.; Koh, J.; Kim, S. *Anal. Chem.* **2010**, *82*, 211–218.

(11) Ohno, T.; He, Z.; Sleighter, R. L.; Honeycutt, C.; Hatcher, P. G. *Environ. Sci. Technol.* **2010**, *44*, 8594–8600.

(12) Sleighter, R. L.; Liu, Z.; Xue, J.; Hatcher, P. G. *Environ. Sci. Technol.* **2010**, *44*, 7576–7582.

(13) Kujawinski, E. B.; Behn, M. D. *Anal. Chem.* **2006**, *78*, 4363–4373.

(14) Wozniak, A. S.; Bauer, J. E.; Sleighter, R. L.; Dickhut, R. M.; Hatcher, P. G. *Atmos. Chem. Phys.* **2008**, *8*, 5099–5111.

(15) Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. *Limnol. Oceanogr.* **2010**, *55*, 1467–1477.

(16) Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. *Geochim. Cosmochim. Acta* **2006**, *70*, 2990–3010.

(17) Schmitt-Kopplin, P.; Gelencsér, A.; Dabek-Zlotorzynska, E.; Kiss, G.; Hertkorn, N.; Harir, M.; Hong, Y.; Gebefügi, I. *Anal. Chem.* **2010**, *82*, 8017–8026.

(18) Salmon, E.; Behar, F.; Hatcher, P. G. *Org. Geochem.* **2011**, *42*, 301–315.

(19) Zhong, J.; Sleighter, R. L.; Salmon, E.; McKee, G. A.; Hatcher, P. G. *Org. Geochem.* **2011**, *42*, 903–916.

(20) Gonsior, M.; Peake, B. M.; Cooper, W. T.; Podgorski, D.; D'Andrilli, J.; Cooper, W. J. *Environ. Sci. Technol.* **2009**, *43*, 698–703.

(21) Gonsior, M.; Peake, B. M.; Cooper, W. T.; Podgorski, D. C.; D'Andrilli, J.; Dittmar, T.; Cooper, W. J. *Mar. Chem.* **2011**, *123*, 99–110.

(22) Stubbins, A.; Hood, E.; Raymond, P. A.; Aiken, G. R.; Sleighter, R. L.; Hernes, P. J.; Butman, D.; Hatcher, P. G.; Striegl, R. G.; Schuster, P.; Abdulla, H. A. N.; Vermilyea, A. W.; Scott, D. T.; Spencer, R. G. M. *Nat. Geosci.* **2012**, *5*, 198−201.
(23) Kido Soule, M. C.; Longnecker, K.; Giovannoni, S. J.; Kujawinski, E. B. *Org. Geochem.* **2010**, *41*, 725−733.
(24) Sleighter, R. L.; McKee, G. A.; Liu, Z.; Hatcher, P. G. *Limnol. Oceanogr.: Methods* **2008**, *6*, 246−253.
(25) Koch, B. P.; Witt, M.; Engbrodt, R.; Dittmar, T.; Kattner, G. *Geochim. Cosmochim. Acta* **2005**, *69*, 3299−3308.
(26) Sleighter, R. L.; Hatcher, P. G. *J. Mass Spectrom.* **2007**, *42*, 559−574.
(27) Mesfioui, R.; Love, N. G.; Bronk, D. A.; Mulholland, M. R.; Hatcher, P. G. *Water Res.* **2012**, *46*, 622−634.
(28) Minor, E. C.; Steinbring, C. J.; Longnecker, K.; Kujawinski, E. B. *Org. Geochem.* **2012**, *43*, 1−11.
(29) Dittmar, T.; Kattner, G. *Mar. Chem.* **2003**, *83*, 103−120.