

“Distinguishing molecular characteristics of aerosol water soluble organic matter from the 2011 trans-North Atlantic US GEOTRACES cruise” by A. S. Wozniak et al.

Anonymous Referee #1

Received and published: 22 April 2014

Overview: The presented publication reports molecular characteristics of aerosol water-soluble organic matter (WSOM) collected as part of the 2011 trans-North Atlantic US GEOTRACES cruise. Aboard the ship, a high volume filter sampler was used to collect 24 individual samples Nov 7, 2011 - Dec 9, 2011. The samples were analyzed using ultrahigh resolution FT-ICR MS. Using multivariate statistical analysis the identified molecular formulas were used to separate the 24 samples into 5 key groups associated with the air mass histories. For each of the groups, the molecular characteristics are described based on the unique molecular formulas identified using the PCA. Overall, this is an interesting piece of work with very high quality data. The molecular insights to marine organic aerosol from this paper represent a substantial contribution of interest to the readers of ACP.

Specific Comments: Several specific comments are listed here in no particular order.

1) P6428, L26: The following includes a typo, “higher O/C ratios and lower O/C ratios”.

We thank the reviewer for picking up this typo. The revised text will reflect that this has been corrected to read “higher H/C ratios and lower O/C ratios”.

2) P6432, L09: The PCA was done on selected assigned molecular formulas and not on the MS data.

We agree with the reviewer that the precision of this sentence could be better. The PCA was performed on the relative spectral intensities of a subset of the assigned molecular formulas for our samples. Those molecular formulas correspond to peaks at m/z ratios which are derived from the MS data. The revised text will be changed to read “Principal component analysis (PCA) is performed on data derived from the FTICR-MS spectra...” (lines 89-90 in the revised manuscript) as this is a more precise description of what was done.

3) P6434, L14: Please note the addition of HCl may catalyze many oligomerization reactions. Also, methanol has been shown by Bateman et al. ES&T 2008 to contribute to artifacts from acetal reactions.

We are aware of potential artifacts due to the use of methanol as a solvent as documented by McIntire and McRae (2005) and Bateman et al. (2008), and every effort was made to minimize self esterification. McIntire and McRae found self-esterification to be more significant in positive ion mode relative to negative ion mode and to increase with the amount of time the sample remained in methanol. Our samples were run in negative ion mode within 2 h of elution and stored in the dark in a freezer between elution and running. We feel comfortable that this minimized the extent of self-esterification and did not affect the conclusions of the manuscript.

To ensure high WSOM recoveries for PPL extractions, WSOM must be acidified, and the addition of HCl in PPL extractions for FTICR MS studies is commonplace (e.g., Dittmar et al., 2008; Gonsior et al., 2009; Stubbins et al., 2010; Kujawinski et al., 2011). The water extracts

were acidified to a very dilute concentration (~0.005 M) and samples were eluted and run within 2 h of elution as described in the preceding paragraph in an effort to minimize any artifacts. Examination of sample mass spectra do not show evidence for oligomerization reactions, and we are comfortable that any artifacts have been minimized.

4) P6435, L01 – L22: Molecular formula assignments appear to be limited only the elemental ratios, mass accuracy and the number of heteroatoms. Are there additional measures to ensure data quality? The method described by Kujawinski et al., 2009 is more extensive than just the number of heteroatoms. They also advocate using formula extensions because mass accuracy alone is insufficient with the high number of elemental combinations. Furthermore, the limit on the O/C is relatively low for atmospheric aerosol.

Formula assignments were also confirmed using Kendrick mass defect series in a manner similar to Kujawinski et al. (2009). Text has been added to the methods reflecting the use of a formula extension approach (lines 152-154 in the revised manuscript).

In spite of what has been reported in recent aerosol OM publications, we have chosen a maximum O/C ratio of 1.2. In a study testing appropriate limits for molecular formula assignments, Kind and Fiehn (2007) examined more than 68,000 formulas reported in Wiley and DNP databases and found that 99.7% of formulas were assigned correctly using an O/C limit of 1.2. Increasing that limit to 3.0, only increased the percentage of correctly assigned formulas by 0.2%. Using an O/C maximum of 1.2 allows us to assign more than 90% of OM peaks in our mass spectra (excluding salts and ^{13}C peaks). It is likely that the chance for incorrectly assigning a formula with O/C between 1.2 and 3.0 is higher than identifying a correct formula. We do acknowledge that compounds having molecular formulas with O/C greater than 1.2 do exist. These may include short carbon length compounds with nitrate and/or sulfate groups including nitrooxyorganosulfates that have been reported in other atmospheric work but that are below the m/z range in our study ($m/z < 200$). As a result, we are very comfortable using the O/C limit we have chosen and suggest that molecular formulas with higher O/C limits be viewed with caution.

5) P6440, L03 – L18: Why are the compounds considered ubiquitous? Perhaps they are strongly associated with biogenic hydrocarbons? A similar statement about “terpene-like” molecular formulas that has been presented in Schmitt-Kopplin et al., 2010 and Mazzoleni et al., 2012. This underlines the significance of biogenic hydrocarbons as sources of atmospheric OA components.

As described in the text, the compounds are considered to be ubiquitous because they are present in the majority of the samples from this study and additional samples collected in terrestrial environments in Virginia and New York, USA. Of course, biogenic sources account for major portions of all types of organic matter including that in the atmosphere (minus that contributed by fossil fuel combustion), but we assume that the reviewer refers specifically to terpene-like biogenic compounds such as secondary organic aerosols from isoprene, alpha-pinene, and other terpene-like precursors. While it is possible, these ‘ubiquitous’ molecular formulas could be derived from terpene-like biogenic compounds – which are undoubtedly important atmospheric OA components – we do not have confirmation that they are indeed terpene-like biogenic

compounds. Specific confirmation of a terpene-like source for these 'ubiquitous' formulas requires more detailed structural analysis and is beyond the scope of this paper.

6) P6443, L03 -06: How do you explain the amino functional groups in negative ion mode? Typically the H-affinity of amines is so high, that the multifunctional compounds (amino acids) will be zwitterions in the negative ion mode unless there are multiple deprotonations. However, most of the studied OM components have been singly charged. Several of the tentative structures drawn in Figure 7 do not seem likely to be observed in the negative ion mode as indicated on P6433.

Examination of peptides by our group using peptide standards in the negative ion mode has not observed this issue. Peptide standards are observed as singly charged ions. The structures presented in Figure 7 are intended to demonstrate that the formulas are consistent with peptide-like structures and not meant to identify the specific compounds in our samples. Identifying compounds from molecular formula information alone is not possible. Text has been added to the revised manuscript to clarify that these structures are tentative: lines 301-303: "It is noted that the structures in Figure 7 are tentative and represent only some of many potential isomers that correspond to the assigned molecular formulas. LC/MS or a comparable technique is needed to verify the structures of the compounds that correspond to these formulas, but this is beyond the scope of this particular paper."

7) P6449, L08– 09: How was the WSOM defined as combustion-influenced? Or, how was the contribution of biogenic SOA removed from the complex mixture to better understand the composition of the combustion influence? How does the fossil fuel combustion composition differ from biomass burning?

The WSOM is defined as combustion-influenced based on it having air mass trajectories from the North American continent which is widely known to have higher combustion-influence (specifically, anthropogenic fossil fuel combustion influence) than the north African continent (the other continental influence in this study). Because combustion processes impact WSOM composition via increased SOA formation, increased inclusion of combustion byproducts (NO_x, SO_x) into SOA and aged POA (e.g., de Gouw et al., 2008), and the direct emission of WSOM, the terminology is accurate. Biogenic SOA were not removed from the complex mixture and likely comprise a portion of the North American influenced WSOM. Text has been added to the revised manuscript that clarifies our definition of WSOM coming from the North American continent as combustion-influenced. lines 409-413: "The data presented here demonstrate that samples with clear Saharan influence show a predominance of CHO aerosol WSOM compounds with less aliphatic character than typically found for aerosol WSOM influenced by the North American continent (e.g., Figure 6c and d; Wozniak et al., 2008) which experiences considerably more combustion influence known to both emit fossil fuel OM and stimulate secondary organic aerosol formation (e.g., de Gouw et al., 2008; Carlton et al., 2010)."

8) P6451, L23 - 27: A lack of condensed aromatic compounds could be a result of other factors. For example, aerosol-aging processes may promote transformation of the compounds over the longer distances. Have condensed aromatic compounds been observed in the remote atmospheres previously?

The reviewer is correct that our explanation for the lack of condensed aromatic compounds as important PCA loadings to the North American samples is insufficient. Yes, condensed aromatic compounds measured as soot or black carbon have been observed in the remote marine atmosphere previously (e.g., Posfai et al., 1999). Drawing on the data from the full FTICR MS dataset tens to hundreds of peaks in the North American samples are assigned formulas that correspond to condensed aromatic or aromatic compounds using Koch and Dittmar's (2006) aromaticity index. It is our contention that these compounds do not ionize efficiently enough to account for an important component of the spectral intensity relative to the more aliphatic sulfated WSOM compounds in these spectra which are known to ionize efficiently. We do recognize that these aromatic and condensed aromatic WSOM compounds could be transformed via photochemical aging, but these processes would impact the North African influenced samples to which the North American influenced are being compared. In the revised manuscript, we have added text that explains our argument more clearly while acknowledging alternate hypotheses. Lines 459-463 "Combustion processes are well known to emit condensed aromatic compounds including black carbon, and tens to hundreds of condensed aromatic formulas were found in each North American-influenced sample (data not shown). It is likely, however, that the aliphatic, highly oxygenated compounds in the North American-Influenced samples dominated the signal in the electrospray source keeping these more condensed compounds at low abundance and therefore not considered in the PCA."

9) Figure 6: A few figure labels are missing or use different fonts.

Figure panels that appear to be missing labels do not contain molecular formulas of a given combination that define these groups of samples. For example, panel C does not have CHONP or CHOSP listed because none of the PCA loadings found to define the North American-influenced sample group contained CHONP or CHOSP. In the revised manuscript, one figure label was made to represent all of the figure panels and was placed in the top right panel (Figure 6b) for convenience.

10) Figure 1: Only 3 of the 5 groups are shown here. Why?

As is described in detail in the text, the air mass trajectories displayed on Figure 1 describe the three types of air mass trajectories observed for this set of samples. The five PCA groups discussed in the text are defined from the PCA and are named taking the definitive molecular characteristics and environmental characteristics (e.g., air mass trajectories) into account.

Anonymous Referee #2

The authors use electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for characterisation of water soluble OM molecular fraction of 24 aerosol samples collected as part of the 2011 trans-North Atlantic US GEOTRACES cruise campaign. The authors successfully applied PCA for a very large high resolution MS dataset and identified molecular characteristics of aerosols influenced by primary/secondary marine, continental combustion (North America), and continental dust (North Africa/Saharan dust) sources. The experiments of this work were carefully designed and executed. The paper is clearly written and the subject matter is appropriate for publication in ACP. I recommend the paper be accepted subject to technical corrections.

Technical remarks:

- 1) The authors used solid phase extraction PPL cartridges to remove salt content from the samples. This step generally results in the loss of compounds containing carboxylic groups. Most of the carboxylic acids are important constituents of the atmospheric aerosols and are characteristic for certain emission sources. Have the authors checked the recoveries of the representative organic compounds that are expected to be abundant in their samples (e.g., fatty acids that are characteristic for the marine emissions)? I realise that this may not be in the scope of the study; however, I would suggest making a short statement addressing the possible limitations associated with the use of this step.
- 2) The authors externally calibrated their instrument using fatty acids which is an abundant class of marine aerosol; however, I missed any discussions whether they were able to detect these molecules in the analysed samples. I am wondering whether the homologous series corresponding to fatty acids falls in to the cluster of ions corresponding to primary marine sources identified by PCA. The even to odd carbon ratio of these acids can be used to support their findings.

The reviewer brings up an interesting point. We have not attempted to assess carboxylic acid recoveries. However, the WSOM extract is acidified prior to PPL extraction thereby protonating carboxylic groups and enhancing their retention. Unfortunately, we are unable to use fatty acids as valuable source markers for other reasons: 1) Fatty acids respond especially well to electrospray and are unfortunately present in solvent blanks run by ESI FTICR-MS at the COSMIC facility. Peaks found in solvent blanks – including these fatty acids - are eliminated from our peak lists for formula assignment post-calibration. 2) Further, electrospray ionization is a competitive process, and as a result, the instrument response is not quantitative. Relative magnitudes may well still be useful as source determinants for fatty acids (if relative responses are verified as reproducible among the similar class of compounds), but contributions from solvent blanks (and resultant peak elimination) make this comparison rather challenging. As a result, we are unable to explore fatty acids as source determinants with confidence in this manuscript.

The benefits and limitations of PPL cartridges (and other solid phase extraction techniques) has been discussed elsewhere (Dittmar et al. 2008). In a revised manuscript, the following text has been added (lines 132-133): “Previous work has estimated that ~60% of dissolved OM is retained using this technique (Dittmar et al., 2008).”

- 3) Fig 7 shows structures of several amino acids; the direct infusion analysis does not

allow obtaining the structural (isomeric) information of the molecules. Unless these structures were confirmed by LC/MS analysis or other appropriate techniques I would suggest removing them from the figure. Moreover, I would add a few lines in the text clarifying that these compounds were identified tentatively.

Yes, the author is correct. FTICR MS does not allow for a structural determination of compound identity. The structures drawn in Fig 7 are amino acid containing compounds that have the same molecular formulas as some of the formulas identified in our samples. As noted in the text (p6443, lines 10-11), these structures are simply presented as examples of “potential amino acid containing compounds that correspond to these formulas”, consistent with other lines of evidence that these CHON compounds may be biologically derived. Another reviewer also commented on these structures so text has been added that clarifies that these structures are merely some of many possible structures that can be drawn from the assigned molecular formulas. See lines 301-303 of the revised manuscript: “It is noted that the structures in Figure 7 are tentative and represent only some of many potential isomers that correspond to the assigned molecular formulas. LC/MS or a comparable technique is needed to verify the structures of the compounds that correspond to these formulas, but this is beyond the scope of this particular paper.”

Anonymous Referee #3

The chemical composition of WSOM is essential for tracking their sources and transformations in the atmosphere as well as assessing their environmental outcomes. In this manuscript, the authors employed a state-of-art ultrahigh resolution mass spectrometer for molecular level characterization of WSOM samples collected during the research cruise from North America to North Africa. Both the samples and data are valuable and rare, as considering the fact that recently there are only few studies on the detailed chemical composition of WSOM on marine aerosols. The authors also successfully applied the statistic method (PCA) on classifying and distinguishing the sources of WSOM based on the identified molecular formulas. The results and interpretations are reasonable. I recommend the publication of this manuscript after the authors addressing the following questions:

1. Page 10, line 1: the authors used naturally occurring fatty acids as the internal standards for calibrating the mass accuracy, which can dramatically increase the number of unambiguously identified formulas within WSOM samples. However, in the paper they cited (S and H, 2008), there is no detailed information about how to perform this calibration. Moreover, the studies of S and H are mainly focused on aquatic NOM samples, which might possess different chemical natures with the aerosol samples. I would encourage the authors providing more detailed information about this internal calibration in the supplementary material. For example, what kind of fatty acids were chosen as the internal calibrator? What are the criteria regulating their naturally occurring? How to re-correct the mass errors of other compounds by using the mass errors of the internal standards? I believe that the detailed description on this re-calibration method is very helpful in improving UHRMS data processing in atmospheric chemistry society and will make this paper highly cited in future.

The detailed calibration procedure is beyond the scope of this particular paper, and in fact, would be redundant to the Sleighter and Hatcher (2008) paper. While there are differences in the organic matter used in that paper and atmospheric organic matter, there is quite a large amount of overlap in the types of organic matter compounds making up these organic matter classes. Our calibration utilizes the internal calibration procedure provided by the Bruker software using the internal linear calibration mode. We have accumulated a list of CHO molecular formulas that frequently occur in natural and atmospheric organic matter operated in the negative ion mode. The list of formulas we use correspond to fully saturated mono and di acids of many carbon chain lengths (C_{14-48}) and homologous series (differing by CH_2 groups) of other frequently occurring formulas. The calibration peak list is chosen to cover the mass range examined by the FTICR MS (in this instance, $m/z = 200-800$), and care is taken to ensure that at least one formula is found every 28 mass units ($2 CH_2$ units). The Bruker software looks for these pre-selected molecular formulas within an operator assigned acceptable error range ($0.002 m/z$) and creates a linear calibration based on the measured m/z and exact m/z values. Peaks that show high errors ($>0.5ppm$, measured vs. modeled m/z) are eliminated from the calibration as they may represent incorrectly assigned calibrant peaks and affect the calibration. The remaining measured m/z values in a spectrum are subsequently adjusted to the linear calibration. This procedure gives the m/z peak list that is used for formula assignment and is repeated for each sample.

2. Page 10, line 11: $O/C \leq 1.2$ as one of the constraints of formula assignments. This may work well in aquatic NOM. However, in recent years, atmospheric scientists focusing on chamber studies of secondary organic aerosols do find some organic compounds with O/C greater than 1.2. Will the authors' dataset be changed if this constraint is extended to a larger value (e.g. $O/C \leq 3.0$)?

Reviewer #1 made a similar comment regarding the O/C ratios. Our reply is the same: In spite of what has been reported in recent aerosol OM publications, we have chosen a maximum O/C ratio of 1.2 and believe this to be the best O/C limit for our FTICR MS data. In a study testing appropriate limits for molecular formula assignments, Kind and Fiehn (2007) examined more than 68,000 formulas reported in Wiley and DNP databases and found that 99.7% of formulas were assigned correctly using an O/C limit of 1.2. Increasing that limit to 3.0, only increased the percentage of correctly assigned formulas by 0.2%. Using an O/C maximum of 1.2 allows us to assign >90% of OM peaks in our mass spectra (excluding salts and ^{13}C peaks). It is likely that the chance for incorrectly assigning a formula with O/C between 1.2 and 3.0 is higher than identifying a correct formula. We do acknowledge that compounds having molecular formulas with O/C greater than 1.2 do exist. These may include short carbon length compounds with nitrate and/or sulfate groups including nitrooxyorganosulfates that have been reported in other atmospheric work but that are below the m/z range in our study ($m/z < 200$). As a result, we are very comfortable using the O/C limit we have chosen and suggest that molecular formulas with higher O/C limits be viewed with caution.