Few final comments (line numbers refer to the track changes document):

The authors have made good improvements to the manuscript according to the review. As I point out in my comments below it would be useful to add extra clarity on the standard deviations used for the DOM metrics data (as the other reviewer and I pointed out already in our initial comments). I see the reply of the authors regarding this, however, to me it is not fully clear how they have reduced the standard deviations so dramatically.

1. Line 191: Presumably these filters were also 0.7 µm?
We added the pore size to the text: 'GF/F (Whatman®, 0.7 µm, pre combusted)'.
2. Line 436: These standard deviations (in Table 2) do not seem 'high' as described in this sentence.
This section refers to Table S2 in the supplementary material (not Table 2 in the main text). Please see response for next section for changes made to address this.

3. Table 2. The authors seem to have been able to decrease the standard deviation by an order of magnitude or fully remove any deviation for the DOM metrics by using replicate samples instead of the compounds. I am not sure I fully understand how these differ from each other here. Any extra clarification in the text could be useful for the readers.
We agree that including explanations for both standard deviations in the main text could be confusing and to add to the confusion, the two standard deviation equations use the same symbol, N, to describe two different things. To address this, we have changed replaced the two N symbols to different symbols (F and N). F represents the number of formulas within a sample and N the number of samples within a treatment. Additionally, the text describing Table S2 was moved to the supplementary material in the caption text for Table S2. We also reference the equations for the respective standard deviation in the caption of the table.
The weighted standard deviation, SDw (Equation 3, Table S2) in the supplementary material describes the variability of formulas within a sample, and the standard deviation, SD, (Equation 2, Table 2) in the main text explains the variability between treatment means. The high variability of formulas within a sample (SDw) help explain our rationale for interpreting small changes between means as large changes among hundredths of formulas per sample. We hope that these changes will provide better guidance for the reader.

New text on the supplementary Table S2:
L106: 'The weighted standard deviation (SD_w, Equation 3) and the standard error of the weighted mean (SEM, Equation 4) are computed for each sample using the DOM metrics associated with the molecular formulas within that sample. The variability in metric values among the many identified formulas is reflected in the standard deviations. The benefit of the large number of formulas for each treatment is the high certainty in the mean which is also shown by the low standard error of mean.'

New text for Table 2 in manuscript:
L439: 'The standard deviation (SD) is computed (Equation 2) for the start (t0) and end (t1) of incubation of the 36 hours incubation for the filtered (F) fjord water treatment (N = 3). The standard error of the difference of means ($SEM_{x_1-x_2}$) is computed (Equation 5) for each experiment.'

We also added the line:
L430: 'A table presenting the individual mean values per sample for each DOM metric is provided in the supplementary material (Table S2).'

Additionally, we replace the symbol for N with F in the weighted standard deviation equation in the methods section:
L311: The equation for weighted average for DOM metrics (H/C_{wa}, O/C_{wa}, $AI_{mod\ wa}$, MW_{wa}) is shown below. Here, I_i is the signal normalized intensities for a given formula and A_i represents the DOM metric value for that formula and F is the total number of formulas per sample.

$$wa = \frac{\sum_{i=1}^{F} I_i \cdot A_i}{\sum_{i=1}^{F} I_i} \quad (1)$$

The standard deviation shown in Table 2 is calculated for the mean DOM metric in each treatment. Here, wa_i is the intensity weighted average for each sample and (N) represents the total number of samples per treatment and $\overline{x}$ is the sampling mean for the treatment.

$$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(wa_i - \overline{x})^2} \quad (2)$$

Additionally, the weighted standard deviation (SD_w) for compounds in each sample is estimated by the following equation and shown in Table S2. Normalized intensities for a given formula are represented by I_i. DOM metric value is represented by A_i, and F is the total number of formulas per sample. The weighted average, *wa*, equation is the same as Equation (1).

$$SD_w = \sqrt{\frac{\sum_{i=1}^{F} I_i (A_i - wa)^2}{\sum_{i=1}^{F} I_i}} \quad (3)$$

The standard error of a single mean's true estimated value (and not the distribution of the population, which is estimated by SD) is approximated (as the SEM) by the following equation where the weighted standard deviation, SD_w, is divided by the square root of the number of identified formulas, *F*.

$$SEM = \frac{SD_w}{\sqrt{F}} \quad (4)$$

The following equation is used for calculating standard error of the difference of means ($SEM_{x_1 - x_2}$) between treatments as shown in Table 2. Here, the standard deviations are divided by the number of samples, N, for each treatment.

$$SEM_{x_1 - x_2} = \sqrt{\frac{SD_1^2}{N_1} + \frac{SD_2^2}{N_2}} \quad (5)$$

4. Supplement line 83-84: How is the n only three for the t-test as seemingly there are so many points (n=53-73) shown in the figure? I'd advise also better practise in reporting t test results: include the t value and degrees of freedom in addition to the p value.

Thank you for this point. The individual highlighted points on the van krevelen refer to each individual tpeak formula for that sample. We ran t-test on the total number of tpeaks reported per replicate (n=3) per treatment. A total of four t-test with a 95% confidence level. We have added the t-values and degrees of freedom for each t-test reported and corrected a typo on the months.

L700: 'Additionally, there was a significant decrease of a group of low H/C compounds referred to as 'terrestrial peaks' (t-Peaks) in February (t-test, p = 0.04, t-value: 2.96, df = 4) and October (t-test, p = 0.04, t-value: 3.02, df = 4). T-Peaks are a group of compounds that are commonly present in vastly different rivers as reported by Medeiros et al. (2016; Fig. S7). Removal of these compounds could contribute to the increase in average H/C_{wa} ratios observed in February and October incubations. This suggests a potential degradation of t-Peak compounds during February and October, in contrast to September (t-test, p = 0.08, t-value: -2.31, df = 4) and December (t-test, p = 0.09, t-value: 2.21, df = 4) when t-Peaks did not significantly change during the incubation (Fig. S7).'