The paper by Croghan and co-authors entitled "Seasonal and interannual Dissolved Organic Carbon transport process dynamics in a subarctic headwater catchment revealed by high-resolution measurements" presents a comprehensive data set of high-frequency flow and DOC data (and associated hydroclimatic variables) over four years to assess the DOC dynamics in an arctic catchment (without permafrost). The authors utilize this data to explore high-frequency metrics and other indices. They use random forest models to assess the drivers of these metrics, and find some interesting and some confounding patterns.

This paper is well structured and clear - with a thorough analysis that doesn't stretch too far beyond the data in hand and seeks to advance our understanding of the system. The data set is valuable and unique, and the process insights strong. There are a number points I would ask the authors to consider, and also explore a newly published paper that has a similar analysis in a subarctic environment with at times similar and at times different results that likely was not spotted before submission. Some of the authors interpretations may (or may not) be informed by this paper:

Shatilla, N. J., Tang, W., & Carey, S. K. (2023). Multi-year high-frequency sampling provides new runoff and biogeochemical insights in a discontinuous permafrost watershed. Hydrological Processes, 37(5), e14898.

>> We thank you for your positive evaluation of our paper and constructive feedback. The suggested paper was indeed not spotted before submission but will be extremely helpful to us, thus we are thankful to you for the suggestion. The insights in the paper are very useful for us, and we will implement some of the observations into our revised paper.

Other comments:

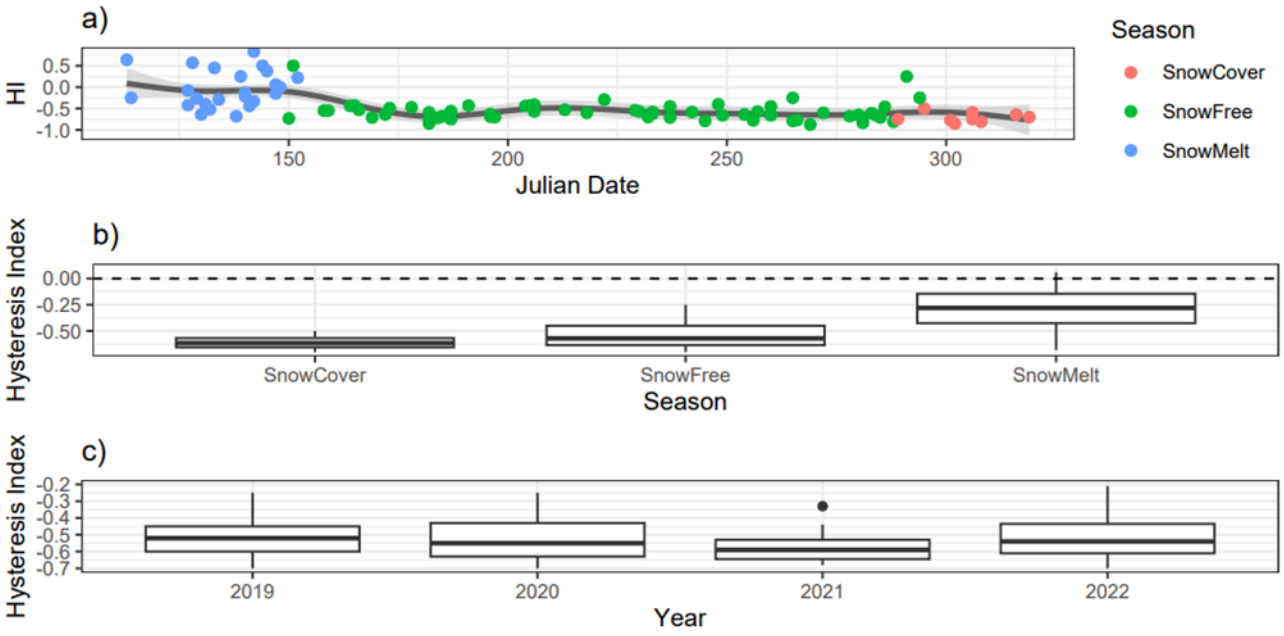
~Equation 1. Is there a reference for this equation?

>> The equation was derived through internal lab calibrations of the instrument thus there is no reference. We will make this clear in the revised manuscript.

~Line 237. Only single peaks were used for analysis. How many events were single vs multiple peaks in the entire data set? Did the authors examine multiple peaks or was this simply 'too messy'?

>> The motivation was to compare between but certainly there would be merit in examining multiple peaks. We've subsequently included an additional 37 peaks (from multi peak events) to see how multiple peaks would change the dataset, but there is no overall difference in results. The snow melt season still has the highest average HI values and in snowmelt and snow cover the values are still consistently very negative in comparison. Thus, the single peaks vs multiple peaks did not make a difference and there

will be no need to change any of the text, though we will include the multiple peaks data (shown in the figure below) in the revised manuscript.



~I would like the authors to consider including summary hydroclimatic information that is easier to 'digest' than the time series. This could be in the SI, but seasonal and yearly temperatures and precipitation (as well as rain v snow) may be informative to the reader as they interpret the data. At several times during the discussion I was trying to evaluate warm vs cold seasons or wet vs dry and it was difficult from the time series alone.

>> This is a good idea thanks, and we agree it will be useful to feature in the main manuscript to better inform the data. We will adapt table one in the revised manuscript so it summarises by season and year.

~If I am interpreting this right, DOC is strongly influenced by water temperature (and all the other factors as well) in the winter (Table 2). I'm struggling to interpret the very large values of node purity (compared with the summer months) and the % variance explained (which is less in the winter). This likely is due to my lack of knowledge with the random forest analysis although it is quite commonly applied. I'm more used to seeing the %Var ttributed to each predictor variable.

>> The node purity tells us the relative contribution of each variable to a models performance but should not be compared between models. Thus, though water temperature has a higher value for node purity in the winter months, this does not mean it is a better predictor than water temperature in the summer months.

~Perhaps a small thing, but air and water temperature are highly related, does this affect the random forest models at all (co-linearity issues)

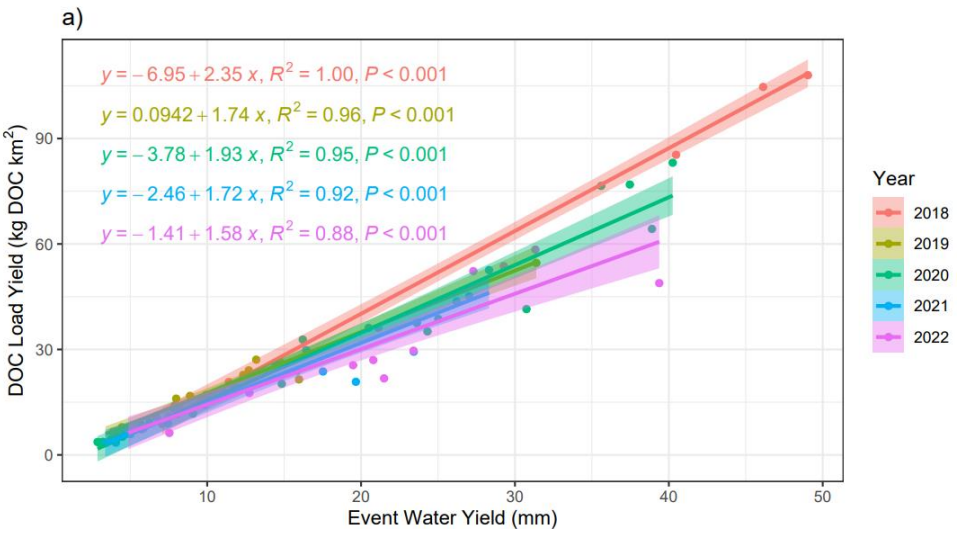
>> By using RandomForests instead of linear regressions, the models shouldn't be impacted by the co-linearity issues. At any rate the motivation for including both air temperature and water temperature in the model was that for the snow cover and snowmelt models, the data wasn't strongly correlated thus they potentially represented different processes.

~ Line 426. How as the event water determined? Via isotopes? Please simply state the method.

>> Indeed, event water was determined using isotope separation, we will clarify this in the revised document.

~For Figure 6, how much of the inter-annual differences is driven by the differences in range of event water yield? Some years have very high flows, whereas others are much more modest. I'm wondering if the 'differences' are simply driven by the end-members. Can a test/consideration be done for values within the same range (say < 50 mm). Largely I'm wondering if doing this constrains the process drivers to more common events, etc., and whether the inter-annual variability in load vs event water yield holds. I'm struggling a bit here and in the interpretation to figure out 'why' this would be. Obviously different climate conditions would be the first place to look, but it remains somewhat unresolved.

>> This is a helpful observation. We re-ran the analysis (figure shown below) and you were correct that the large events accounted for the variation between years and by removing these the differences between years were non-significant. Thus, we will edit the discussion in the revised version, and feature this revised analysis in the supplementary version to show that it was indeed the case that the large events drove the differences between years. Note that 2018 still remained an outlier, although this was likely driven by the fact that monitoring only starting in September that year, and only a much smaller amount of events were monitored that year with none falling in the ranges of 12-40mm.



~Line 469-470. While the forest is mostly conifer, is there much of an understory or mixed forest input of leaf litter? Do you think this affects your interpretation of the DOC sources?

>> The catchment, particularly in the upstream area has an extensive mixture of shrubbery and other non-coniferous vegetation thus this certainly a seasonally variable source of DOC in the catchment. Furthermore, in the peatlands there is extensive dying back of vegetation in the Autumn months which is an ample potential supply of organic matter. With the present analysis we can only speculate exactly to how it contributes to the DOC we observe, but we note that in Figure One for example DOC concentration often seems to increase slightly from September onwards which is perhaps supportive of our interpretation of the presence of decaying organic matter as a source of DOC.

~At the end, you mention stable isotopes of water as an area of future focus (line 537). Do you think that looking at DOM quality would be helpful in advancing our interpretation of DOC data?

>> Certainly, and this is something we are planning to do in the future. Determining the seasonal variability in composition and quality would be extremely beneficial to obtaining better understanding of the system. Using a combination of isotope analysis with compositional analysis is something we are particularly interested in a pursuing going forward to better understand how DOM composition varies between different water sources in the catchments and how subsequently this reflects on the DOC data. We will add in the revised manuscript to reflect these thoughts.