Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC3
https://doi.org/10.5194/bg-2021-279-RC3, 2022


# Comment on bg-2021-279

Anonymous Referee #3

Referee comment on "Sensitivity of plankton assemblages to hydroclimate variability in the Barents Sea" by Elliott L. Price et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-279-RC3, 2022

This paper examines phyto- and zooplankton communities together with their environmental drivers on a large transect across the western Barents Sea. The CPR dataset behind the work is a large and impressive dataset, and the region is of great importance. However, the paper has some very serious flaws and I cannot recommend it for publication in its present state. In my opinion it needs to be re-focused, re-structured and re-written before it can be publishable. I do not list the minor comments here, just major ones.

I agree with Reviewer 1 that the region of study is not “the Arctic”, although it is certainly en-route to the Arctic and will influence the Arctic. Unlike reviewer 1 I do not believe that this alone is enough to warrant rejection, but it does mean that the introduction and parts of the discussion would need to be re-written. This is not an Arctic ecosystem, it is the North Atlantic.

I found it strange to pool phyto- and zooplankton communities into a single community analysis since they are operating on different size, spatial and temporal scales and clearly have different factors driving them. Also one may influence the other (in both directions, as you rightly mention top-down control).

After throughly reading the paper several times I am still not exactly sure what the main finding is here. That plankton are inter-annually variable? That plankton communities are influenced by water masses? Both of those statements are well established facts that do not require proof. To be interesting and scientifically valuable you need to show is (a) the extent of this variability (and how it compares - in absolute numbers - not arbitrary placements along ordination axes - with for example seasonal or spatial variability, and also with other similar studies, (b) exactly which water masses/properties influence the communities and in what way.

The presentation of the results is messy and confusing, and I believe that the statistical methods chosen in the paper are not appropriately applied or interpreted. Moreover, the way most of the model results are presented make them to be an "result in themselves", whereas ordination models are merely a mostly visual tool to aid scientific interpretation of complex data. Most of the text in the result section, Figures 3-4,7 and especially Tables 1-6 - correlations of species to axes - carry no scientific value on their own (they can be in the supplementary material for transparency, of course), but all of this information should already be visible on the ordination plots (which are, incidentally, not even all shown). An in-depth discussion of the results is missing, the existing discussion is scattered and very superficial.

Some more specific comments on the stats are provided below, but I am certain a statistician would have more to say. I highly suggest to the editor that the authors consult with one prior to re-submitting the paper.

The main (and really only) tool the author apply is a pCCA - a partial canonical correspondence analysis, but to me it is not clear that the authors fully understand what this method is. Line 145: "In brief, pCCA is a cluster analysis that describes variability in the plankton assemblage by producing an artificial 'axis' that represents a certain plankton community." CCA is not a cluster analysis, it is an ordination method! It does not produce artificial axes (pl.!) that represent communities. Ordination methods take multidimensional data and try to reduce the number of dimensions, ideally to a 2- or 3-dimensional space. Then we can look at where the data points are relative to each other on this reduced space, and which original variables influenced their placement to infer "communities". A CCA compared to other ordination methods additionally places these data points to maximise correlation to a second matrix of variables (environmental drivers), so we can also see which variables are most correlated to those communities. A partial CCA is applied when we have a set of covariates we are not interested in. Here the authors want to look at spatial variability without year, or interannual variability without space - so that part is OK. But still they do not seem to grasp that CCA is primarily a *visual* method, a way to visualise complex data in reduced space - hence the 6 tables, and many pages of text that describe what should already be visible on the ordination diagrams. They also do not even present the diagrams until figure 6, for the "biplots" - which are already part of the CCA, not a separate analysis. The ordination diagram for the spatial data is not presented at all.

Moreover, the way the data was treated is not clear. Line 155: Any taxa that contributed less than 5 % of the total count zooplankton and phytoplankton counts were removed prior to the pCCA analysis. 5% in any one sample or all samples pooled? If the latter, that seems unnecessarily conservative and would result in the removal of a lot of species that could be very important, just in a smaller number of samples. I wonder how many total species were left in the dataset? I would remove singletons, but the rarer species (especially larger ones that are just unlikely to be caught in large numbers) can also be important indicators of communities!

Line 156. “Species counts in both phytoplankton and zooplankton datasets differed by several orders of magnitude and tended towards zero inflation. To mitigate against these effects, species counts were standardized from 0 and 1.” First of all, zero inflation does not necessarily equal “a high percentage of zeros”, we will come back to that later. Species counts were standardised from 0 to 1, what does that mean? That you took the ratio of each species in each sample? Or you took the highest abundance and set it to 1 and adjusted the rest accordingly as a fraction of 1? But how does that mitigate the order of magnitude problem? If you have 10,000 Oithona and 1 Calanus in your sample, then you will have 0.9999 and 0.0001 (or 1 and 0.0001), still same order of magnitude difference. I think what you might mean here is that you were interested in the *relative composition* of the communities, not total species abundances? In this case you should of course use ratios, not absolute abundances. To address the order of magnitude problem I would also do a log- or square/fourth root transformation to de-emphasise the most dominant species and bring up the importance of the rarer ones, but consult with a statistician whether that is appropriate in your case.

You standardised your abundance values (or at least tried to), what about the environmental variables? These should be scaled too.

Line 111 onwards - you said you use SST separate into 3 regions based on water masses, but it looks like you just use latitude. Otherwise, you need to show this somewhere. Moreover, I am not sure you can separate water masses based on SST alone.

Line 132 We selected a suite of environmental variables that have previously been known to influence arctic plankton community assemblages.

I think you need to justify this statement with some specific citations. Again, here I have to question your decision to pool the phyto- and zooplankton community together for all your analyses, since they will clearly have different parameters influencing them.

Line 135 we included the average seasonal SST for the spring and summer prior to sampling to capture any lagged effects of SST

Plankton are not stationary, they do not hang out in one spot for one year, by their very definition they are flowing with the currents, and this is a highly advective area. SST *in that location* will certainly not influence the plankton community a year later. It is possible that it may be proxy of what is going on upstream, but why not just use data from upstream if you are using remote sensed data anyway? I would also be very cautious adding such data, for precisely the reason that it might generate "false positives" - you rightfully conclude that last year's summer temperature is unlikely to be driving Calanus abundance, but yet it showed up as a prominent "predictor" in your analysis.

Line 158 "For each axis produced, each species, each hydroclimate variable and each sample was scored to describe how each component relates to that axis. A species and hydroclimate variable with a similar score for an axis were positively correlated, and a sample with a similar axis score for the corresponding species and hydroclimate axes indicated that species was in higher abundance in that sample."

Yes, that is how an ordination works, and this information should be well visible on the plot, not presented in 6 tables.

Line 161-166 What is the purpose of all these additional tests? What do you gain that you cannot see just by looking at the ordination? In my opinion, it adds very little and adds a lot of clutter

Line 180. To determine the impact of phenology on the results of the pCCA, we estimated seasonal abundances by comparing the monthly mean abundance of those taxa. To account for the heavily skewed, zero-inflated distribution of the data, as is common for spatially heterogenous plankton communities, we conducted a zero-inflated negative binomial regression using the 'pscl' package in R. The percentage of zeros in the count data was > 50 % for all species, and so zero inflated models were chosen.

I don't understand this. What did you use a zero-inflated negative binomial regression for? Abundance against … ?

Additionally: like I said before, zero inflated data does not equal data with a lot of zeroes. If you apply a zero-inflated model, you need to have a reasoning behind it. For example, if you are looking at euphausiid abundance, they are known to escape nets, so a lot of zeros in such a dataset will be false zeros. Or if your data includes stations outside of the geographic range/season of the species occurrence. If they are "real" zeros, the model should not ignore them, and it is likely that the negative binomial model alone can account for them!

Results

Section 3.1

For all the plethora of tables and figures and text, some basic information is missing (or hidden). How much of the total inertia was captured by the CCA vs the unconstrained ordination? How much by each axis? Were all the explanatory variables significant - you have a lot, and probably lots of them are junk and some others that may be correlated among each other, some preliminary analysis and step-wise model selection might be very useful here! Why did you chose Axes 1 and 3 for ST1 when usually the first 2 axes describe most of the variability? I see on Fig. 6 that it says that Axis 3 accounts for 15.1% but I honestly find that a bit hard to believe. But if that is indeed true, it is a very large %, but it means that Axis 2 is even higher, and then you can't just choose to ignore it!! In that case you need to show the model as a 3-d box with 3 axes. Like I said, would be nice to see the stats tables for these models. Figure 6 is actually your main result, it should be much larger and come first. Statements like line 194 “consistently negative values for axis 1 across all regions for C. finmarchicus and the low 2014 sample scores” are unnecessary - you can see this in the plot! The fact that they are negative/low is meaningless - this is an arbitrary axis, its values don’t carry meaning on its own. You can actually see a lot of nice information in the plots that you don’t articulate well. For example, in ST1 there is very clear interannual variability (most ellipses are non overlapping), in ST2 a bit less so, and in ST3 I would argue that there is almost no (or very little) separation. That has some real implications that should be discussed!

I don’t really understand how you infer “variability” of species from axes scores - for example, Calanus is the species that is most significantly driving the ordination (i.e. separation of communities), but that doesn’t say much about the variability in its abundance, especially since you standardized(??) the abundance and just looked at relative community composition. Why not just show some box plots/run an ANOVA of the abundances of the different species across the years? Isn’t that what variability is??

Section 3.2

Here instead of Table 4 and Figure 7 you should present (and discuss) an ordination diagram with all appropriate biplots, plus a stats table of the CCA (or describe them in words).

Section 3.3

The word “phenology” is inappropriately used here and throughout the text. You do not look at phenology, just at seasonal variations in abundance (which can be influenced by a hundred different things). These terms do not mean the same thing and cannot be used interchangeably.

This section also feels a bit out of place, since the rest of the paper just focuses on one season.

Discussion

I am missing here a discussion - and comparison - with the large number of zooplankton/ecosystem papers from the Barents Sea. This is the best studied region in the Arctic/sub-Arctic!

Line 261. I would hardly qualify a “high interannual variation of C. finmarchicus abundance” a standout finding. Especially since you don’t really quantify this variability anywhere.

Line 265-268. This is an overly simplistic view. See for example paper by Paul Renaud et al., 2018 from your reference list

Line 285- . SST the year before at this location would not be influencing the community, since it would be a different community

For all the talk of Atlantification in the introduction, one very important aspect is not touched upon at all - advection! This is a highly advective region yet you treat your regions like they are self-contained boxes. The whole transect exists because it is an "Arctic gateway"- it is influenced by several different currents from upstream, and is very important for the Arctic downstream. How are your 3 regions impacted by advection, and by each other?