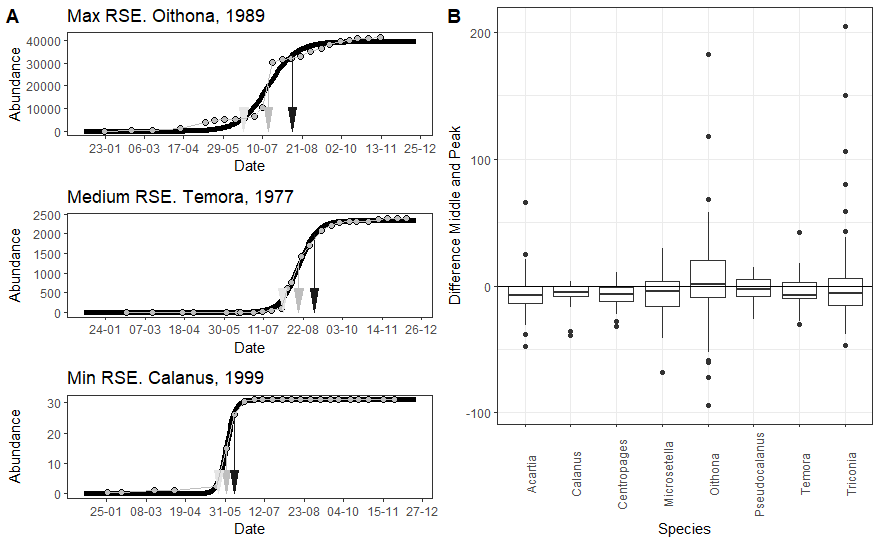
**Materials and methods**

<Reviewer 1>

Line 213-256: this part of the 'Material and Methods' section is quite long and could be easily resumed by means of a figure such as the one used in Mackas et al. 2012 (their Fig. 4 in "Changing zooplankton seasonality in a changing ocean: Comparing time series of zooplankton phenology"; 10.1016/j.pocean.2011.11.005).



Для ответа на первую часть замечания мы, действительно, можем дать вот такой график (часть А). Здесь будут показаны “Begin”, “Middle” и “End”, реконструированные из аппроксимации кумуляты логистической кривой. На этом рисунке приведены три примера логистических кривых, аппроксимирующих кумуляты при максимальном отклонении наблюдаемых значений от предсказанных (max Residual Standard Error, RSE), для среднего уровня RSE, и для минимального уровня RSE. Верхняя картинка - это когда кумулята наименее всего похожа на логистическую кривую, нижняя - это когда кумулята максимально похожа на логистическую кривую. Надо подумать, как эту информацию правильно подать. Предлагаю сделать электронное приложение, в котором будет приведена информация, позволяющая оценить качество логистических моделей, которые были использованы для реконструкции Begin и End.

I would appreciate more information about how was assessed the quality of the logistic models.

На эту часть замечания ответить сложнее. Обычные методы оценки качества модели (типа коэффициента детерминации, R2) в случае с нелинейными моделями не работают (Spiess, Neumeyer, 2010).

Поэтому the quality of the logistic models оценивали косвенными методами. (1 способ) Визуальная инспекция соответствия кумуляты и логистической кривой. Собственно это и приведено на рисунке А. (2 способ) У нас есть референс, который можно напрямую прочитать из данных - это дата пика (Peak, дата, когда наблюдалось максимальное значение обилия). Если модель хорошая, то дата точки перегиба в логистической кривой должна быть близка дате пика. Мы оценили распределение разниц между этими величинами (часть В, на рисунке). Видно, что медианы этих распределений очень близки к нулю. Это свидетельствует о том, что величина, не использованная для построения модели, хорошо предсказывается логистической моделью.

Line 289: "Filling in missing values". Did the authors controlled for possible biases when 'modelled' data were estimated? This step could be considered as a data pre-treatment. When missing data are re-estimated, it could be useful (and more robust) to use sensitivity analysis approaches in combination with the numerical analyses to avoid possible biases.

Метод заполнения пропусков во временных рядах с помощью SSA хорошо разработан (Golyandina, Osipov, 2007, <https://www.gistatgroup.com/cat/mvssa1en.pdf).> Для проверки работоспособности метода на наших данных мы провели специальный анализ. Мы использовали матрицу оценок средовых параметров (датасет с уже заполненными пропусками, который был использован в дальнейшем анализе). В этой матрице мы симулировали 183 пропущенных значения (это количество пропусков было в нашем исходном датасете), которые случайным образом распределяли по матрице. Далее эти пропущенные значения были восстановлены с помощью SSAпо той же методике, которая была применена для основного анализа. После вычислений мы находили разность между реальным значением и значением, восстановленным с помощью SSA. Было произведено 1000 симуляций. Прмененный метод давал outliers (values out of plus/minus 3 standard deviations) только в 0.7% симуляций. Медиана распределения отклонений восстановленных значений от реальных данных для всех средовых переменных была близка к нулю (см. Рсунок).

Var Median\_Delta

1 AOI\_1 0.0

2 AOI\_DJFM 0.0

3 Ice\_clear\_G -1.3

4 Ice\_clear\_K -0.4

5 Ice\_clear\_M -1.0

6 Ice\_thick\_G -0.5

7 Ice\_thick\_K -0.8

8 Ice\_thick\_M -1.2

9 NAO 0.1

10 NAO\_1 0.1

11 S\_Apr 0.0

12 S\_Jun\_Aug 0.0

13 S\_May 0.1

14 S\_Sep 0.0

15 Su\_duration 0.0

16 Su\_duration\_1 0.7

17 Su\_end 2.0

18 Su\_end\_1 2.8

19 Su\_start -0.1

20 Su\_start\_1 -0.1

21 t.Spring 0.1

22 t\_3 -0.2

23 t\_4 -0.6

24 t\_6 -0.7

25 t\_7 -2.3

26 t\_8 0.4

27 t\_Aug 0.1

28 t\_July 0.2

29 t\_June -0.1

30 t\_May 0.0

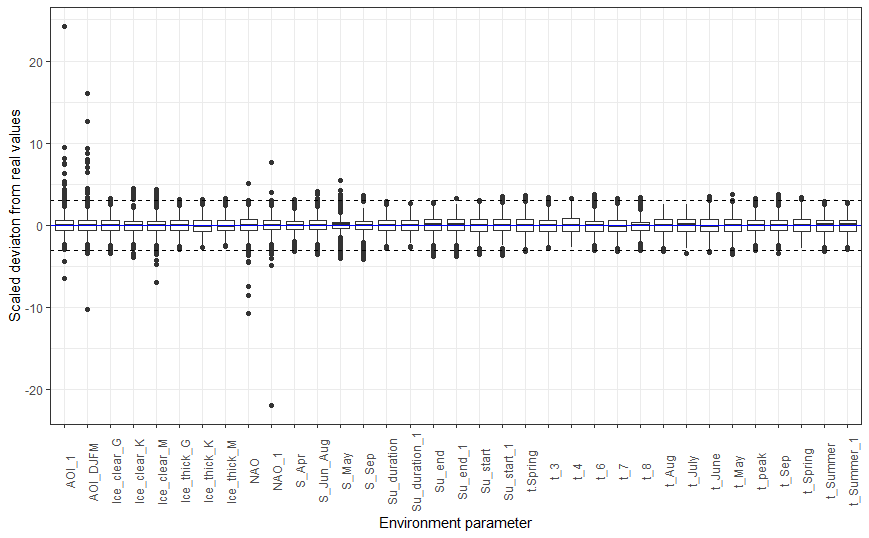
31 t\_peak 0.4

32 t\_Sep 0.0

33 t\_Spring 0.0

34 t\_Summer 0.3

35 t\_Summer\_1 0.2



Line 294: "Analysis of the long-term dynamics of the studied parameters". I do not understand why the authors want to estimate linear trends instead of estimating trends over time by finding the best fit in the long-term dynamic of each parameter, or by using alternative methods such as the modified Mann-Kendall trend test or the Spearman's rank correlation coefficient. If the main reason is to use the residuals from this analysis for the part devoted to "Relationship of the abundance of the species and its phenological indicators" (Line 344-368), it is required to check whether the assumptions associated with each linear regression model are met (normality, no auto-correlation, homoscedasticity) by examining the residuals.

1. Mann-Kendall trend test or the Spearman's rank correlation coefficient дают лишь информацию о наличии связи между временем и тем или иным значением.   
     
   Line 322: "Factors, influencing phenology". I would suggest to present how the CCA works and the different steps of this analysis before explanations about the 'phenological' and 'predictor' matrices.

<Reviewer 2>

Similarly the R code used for statistical analysis does not seem to be openly accessible, as it is more and more the case (through an online platform such as GitHub).

Regarding the time series analysis, the seasonal effect should be removed before testing the long-term trend, but it seems that it is not the case here. The seasonal effect could for instance be removed using a Seasonal and Trend decomposition using Loess (or STL; Cleveland et al. 1990). This can be performed in R using the stlplus package (Hafen, 2016) for instance. Then, I would have rather used the Generalized Least Squares (GLS) method to test the significant of the trends, rather than a simple linear regression, in order to take into account the auto-correlation of the residuals, that usual prevent the use of a linear regression on a time series. The GLS can for instance be performed using the nlme package in R (Pinheiro et al., 2018). To do this, the level of auto-correlation can be determined using a partial auto-correlogram of the residuals of a simple linear model. The applicability of a linear model also needs to be verified (normality and homoscedasticity).

- Line 344: which version of the vegan package did you use?

- Line 367: "We did not evaluate statistical significance of the Royama's correlation coefficients": why?

Cited references:  
Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I., 1990. STL: A seasonal-trend decomposition. Journal of official statistics, 6(1), pp.3-73.  
Hafen, R. (2016) stlplus: Enhanced Seasonal Decomposition of Time Series by Loess. R package version 0.1.5. [https://CRAN.R-project.org/package=stlplus](https://CRAN.R-project.org/package=stlplus" \t "_blank).  
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2018) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131.1. [https://CRAN.R-](https://CRAN.R-" \t "_blank) [project.org/package=nlme](http://project.org/package=nlme" \t "_blank).

**Results**

<Reviewer 1>

Line 371: "Seasonal dynamics: temperature and zooplankton species abundance". Following my previous suggestion about species, it could be easier and informative to resume all information in a Table with, for each species, information about both their abundance peak and the dates of phenological events. Because of the pre-treatment procedure applied to fill missing values, could these values be biased?

Because the significance of the trends might be related to the selection of the method (linear regressions) and because this method could be strongly impacted not only by autocorrelation (a potential biased related to temporal autocorrelation has been considered by the authors well) but also by outliers (see, for example, Fig. 4A and B with the exceptional event in the early 1970s), I would suggest to strengthen the results by using another way of calculating trends.

Line 426: "Long-term dynamics: factors influencing phenology timing". In line with my comment about species, a table that resumes changes in the timing of environmental and climate events would be helpful for the readers. As mentioned above, I wonder in what extent these results could be impacted by the data pre-treatment. I would suggest, if the authors decide to not perform sensitivity analysis, to provide more information about missing values in the time series (i.e. at the end of the time series, at the beginning, all along the time period, etc…).  
Line 426: "[…] since they displayed high variance inflation factor". Please, clarify how this was assessed.

Line 445: "The whole model was statistically significant (Table 1a) and explained 40.9% […] It accounted for 14.6% of total inertia". I would suggest to comment more about the meaning of these percentages. The authors should explain better if the percentage of explained variance is high enough to be confident about the robustness of the results. Is the CCA influenced by the high variations in both Acartia and Microsetella abundances that we can observed early 2010s we compared with the previous years?  
  
Line 466: "Relationship of the abundance of the species and its phenological indicators". I would suggest to the authors to explain here why they focused only on the "start-of-season date" without consideration for the other phenological events. Note that I do not understand why the authors did not evaluate the statistical significance of the Royama's correlation coefficients.

<Reviewer 2>

- Fig. 3: I suggest to either transpose the x and y axes to have the Julian days along the x axis as in the Fig. 2, and/or to add the mean values of the 5 phenological events of 3A on 3B with lines.

- Fig. 4 and 5: Find a way to better highlight the significant trends (p<0.01), for instance using red lines instead of blue lines?

- Fig. 6: what are the small dots? Indicate it in the figure legend. You could also number them and indicate for each number/dot to what it correspond in the legend.  
  
- Table 1: I am not sure that Table 1 is necessary. Personally, I do not show these values when I perform a CCA in a manuscript. I just indicate the global significance of the analysis (p-value and % of constrained variances) in the text if the results.

- Line 469: would it be possible to estimate a p-value here (cf. corresponding comment in the Methods section).

<Reviewer 3>

Line 443 onwards: do you mean 40.0% of the variability was explained by the model? Inertia is a very unusual word, please use the common expressions when referring to the CCA model and axis

CCA: when analysing the factors influencing phenology, why did you remove all the factors from the CCA? I would suggest including the different events, start spring, end spring…. Into the CCA as it is usually done to display the correlations, even if no correlation between events and abundances can be established instead of omitting the data.