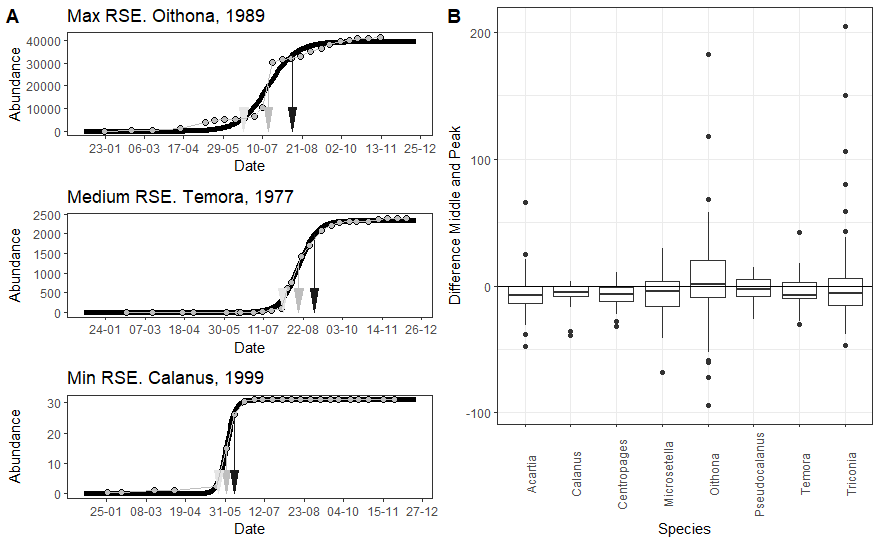
**Materials and methods**

<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.

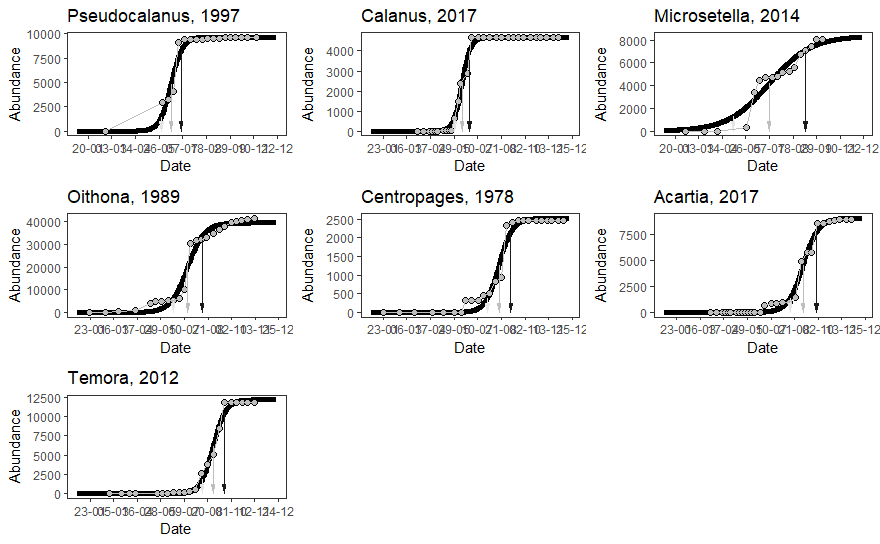
(1 способ) Визуальная инспекция соответствия кумуляты и логистической кривой. Собственно это и приведено на рисунке А. Можно вообще для каждого вида привести пример с максимальным значением RSE (наихудшая аппроксимация) и минимальным значением RSE (наилучшая аппроксимация). Такая картинка приведена ниже. В текст можно вставить следующее.

Since the standard tools of regression quality control such as R2 is not appropriate for nonlinear estimations (Spiess, Neumeyer, 2010) we controlled the goodness of fit for logistic regression by the mean of visual inspection of correspondence between cumulative curve and logistic regression line. Both in the cases of the best fits and in the case of the worst ones congruency between observed and fitted values was rather high (см. Электронное приложение).

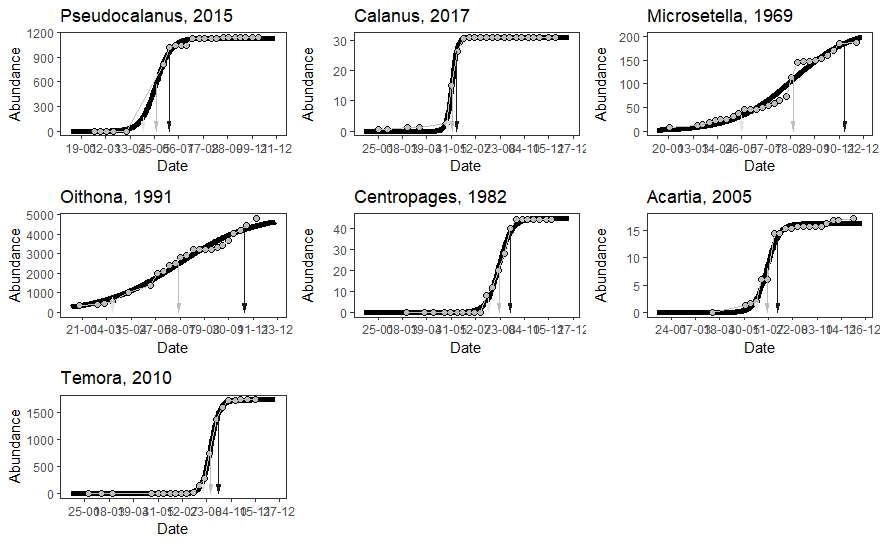
(2 способ) У нас есть референс, который можно напрямую прочитать из данных - это дата пика (Peak, дата, когда наблюдалось максимальное значение обилия). Если модель хорошая, то дата точки перегиба в логистической кривой (“Middle”) должна быть близка дате пика. Здесь можно сослаться на Figure 3, где уже видно, что для каждого вида распределение значений дат Peak очень близко к распределению дат Middle.

Однако можно привести и более детальный анализ, который надо вынести в приложение. Мы оценили распределение разниц между Middle и Peak (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.

Maximal RSE (the worst approximation of cumulative abundance curve by logistic regression)

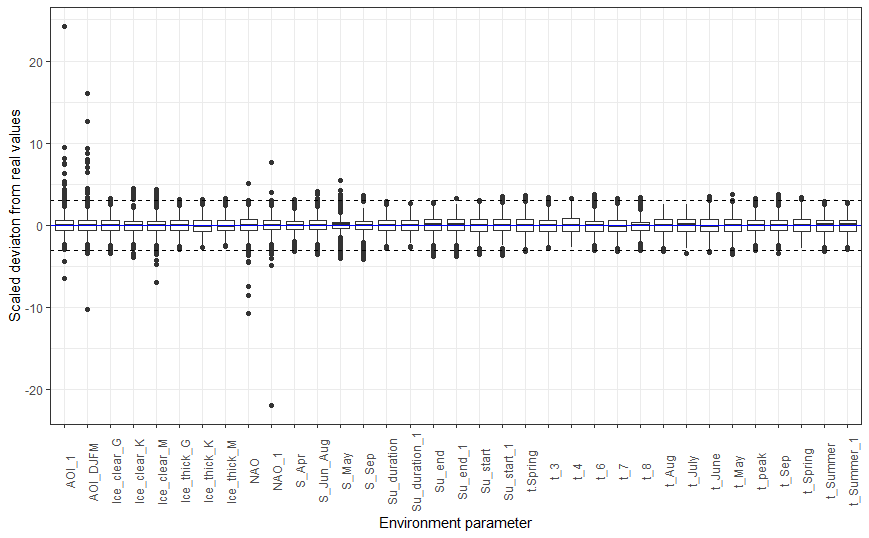


Minimal RSE (the best approximation of cumulative abundance curve by logistic regression)



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 исходных значений. Далее эти пропущенные значения были восстановлены с помощью SSA. 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).



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.

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)

Да, мы, в том числе, использовали регрессионные модели для детерндинга временного ряда. Это еще одна причина, по которой мы применяли регрессионный анализ. Остатки от этой модели дают информацию о межгодовой динамике за вычитом многолетнего тренда. Детерндирование необходимо для получений stationar time series which is needed for assessment of correlations between time series (I.e. between abundance time series and phenologycal parameters time series).

На самом деле у нас было два способа выявления многолетнего тренда. Во-первых, мы вписывали линейные модели. Они были использованы для оценки линейного тренда, угловой коэффициент, и для детерндинга временных рядов (см. Ниже). Во-вторых, мы использовали метод модельных матриц.

**Это надо в явном виде прописать в тексте.**

Использование регрессионного анализа было продиктовано необходимостью оценки того, насколько дней произошло смещение тех или иных фенологических параметров за период наблюдений. Это можно оценить только имея значение углового коэффициента регрессионной модели. Mann-Kendall trend test or the Spearman's rank correlation coefficient дают лишь информацию о наличии связи между временем и той или иной величиной. Мы также решили не использовать упомянутые ревьюером методы, основанные на estimating trends over time by finding the best fit in the long-term dynamic of each parameter. Такой подход, действительно, мог бы дать более тонкую оценку многолетних изменений (в том числе и нелинейных) того или иного фенологического параметра. Однако это потребовало бы тонкой настройки модели для каждой величины, для каждого вида и, как следствие, проведение множественных анализов. Это привело бы к существенному повышению риска Error Type I. Более того, потребовалось бы специальное изучение характера распределения анализируемых величин (например, для количества Julian days, as dependent variable poisson distribution seems to be appropriate). Все эти анализы надо было бы проделать для каждого вида в отдельности. Все это привело бы к порождению множества сущностей, что существенно снизило бы надежность результатов.

В связи с этим мы решили предельно упростить ход анализа, который по сути сводится к тому, что для каждой зависимой переменной мы оценивали лишь линейный тренд, измеренный как величина углового коэффициента в линейной модели.

Однако, как справедливо отметил ревьюер у регрессионных моделей есть много assumptions. Соблюдение этих ограничений гарантирует правильную оценку статистической значимости полученных оценок. В связи с этим, мы опять же решили предельно упростить процедуру оценки статистической значимости выявленного тренда. Она сводится к пермутационной оценке значимости величины Мантеловской корреляции между наблюдаемым временным рядом и модельным линейным градиентом. Кратко суть этого метода сводится к тому, что временой ряд превращается в матрицу эвклидовых расстояний (фактически это матрица модулей разности значений в каждый из годов). Аналогично в матрицу эвклидовых расстояний превращается ряд 1961, 1962, 1963....2017, 2018. Этот ряд моделирует некий абсолютный градиент от минимума к максимуму. Далее между двумя матрицами вычисляется Мантеловская корреляция. Мы не приводим в явном виде значения этих мантеловских корреляций, так как их величина коррелирует с модулем углового коэффициента линейных моделей. Мы используем из этого анализа лишь оценку значимости Мантеловской корреляции, которая проводитится пермутационным методом.

Поскольку и в этом случае мы тестируем множество гипотез, то нам пришлось вводить поправки в уровень значимости (adjasted p-level) за счет False Discovery Rate monitoring procedure (Benjamini, Hochberg, 1995).   
  
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.

Я не понял, ему что тут надо описать всю суть CCA? Она прекрасно описана в куче книжек. Но если надо, то я могу и написать.

<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).

Можно это подготовить, но потребуется время на проверку моего кривого английского в комментариях к коду. Код можно выложить на 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.

Блин! Какой seasonal effect? У нас каждый год характеризует только один параметр, включенный в анализ. Единственное, где это может быть актуально - это динамика обилия. Да, тут можно вытаскивать тренд, удаляя сезонные компоненты. Можем и сделать. Придется, правда, переделать рисунок 3 и переделать раздел Relationship of the abundance of the species and its phenological indicators. Можно отбрехаться тем, что нас интересовали только самые общие закономерности которые мы пытались выявить наиболее грубыми методами (типа если есть закономерность, то она будет видна и так).

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.

Да, такой подход возможен. Но он потребовал бы тонкой настройки моделей, подобранных GLS, так как у каждого вида может быть разная форма автокорреляционной функции. Это приведет к тому, что для каждого вида будет подобрана своя собственная модель с большим количеством своих собственных параметров. Чтобы не плодить множество сущностей, мы решили предельно упростить процедуру для выявления наиболее общих закономерностей.

The applicability of a linear model also needs to be verified (normality and homoscedasticity).

Возможные нарушения условий применимости регрессионного анализа в классическом виде (independency, normality and homoscedasticity) могут привести к overestimation of statistical significance of regression parameters estimation. В связи с этим, мы отказались от классической схемы анализа, которая, действительно, потребовала бы анализа остатков для каждой конкретной зависимой переменной для каждого отдельного вида. Вместо этого мы использовали альтернативный подход. Мы использовали пермутационную оценку значимости мантеловских корреляций между модельным линейным градиентом и наблюдаемым временным рядом (см. Ответы первому рецензенту).

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

We used vegan package version 2.5-6

Надо вставить текст.

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

We did not evaluate statistical significance of the Royama's correlation coefficients because statistical tests for this value are not developed. In fact we did try to assess the significance of these values by permutations. However some special research is needed for assessment of such test validity. It is out of scope of the present work.

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.

Может, действительно, сделать таблицу со всеми оценками параметров, с указанием какая именно величина была реконструирована за счет SSA. Эту таблицу можно поместить в приложение.

Because of the pre-treatment procedure applied to fill missing values, could these values be biased?

1. We did not find any strong evidences of biasment provided by the SSA (see above).
2. Only few percent of missed values 11 from 522 in the case of environmental parameters and ++ from ++ species phenological events were reconstructed.

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.

Еще раз, у нас по сути было два способа оценить тренды!

1. Это угловой коэффициент линейной регрессии (тат же коэффициент корреляции, только вид сбоку).
2. Мантеловская корреяция с модельным линейным градиентом. Так что мы имплицитно использовали два способа оценки.

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…).

Пожалуй, действительно надо что-то такое привести.

Line 426: "[…] since they displayed high variance inflation factor". Please, clarify how this was assessed.

Отсутствие колленеарности предикторов - базовое требование мнжественного регрессионного анализа. Оно оценивается с помощью величины, называемой variance infaltion factor. Можно просто сослаться на то, как это оценивается.

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?

Все эти числа - это просто информация об ординации, ничего более.   
  
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.

Выбор именно этого параметра продиктован тем, что единственое, что может как-то повлиять....

<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.  
Пояснить   
- 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.

Убрать ?

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

См выше.

<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

Дать ссылку?

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.