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

Спасибо за совет. Мы вставили рекомендованные иллюстрации в специальное электронное приложение (ES ++++), в котором мы даем объяснения того, как оценивалось качество логистических регрессий.

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. Using the Residual Standard Error (RSE) we ordered the fitted logistic models from the worst (maximal RSE) to the best (minimal RSE) for each species. Both in the first case and in the second one the correspondences between observed and fitted values was rather high (see ES++). Additionally we compared the observed data of Peak with predicted Middle-of-season date. We suppose that in condition of highly expressed seasonality presented in high latitudes this two events should be close each to other. We found that medians of differences between these two values are close to zero in the case of all species.

+++++++++++++++++++++++++++++

Коля, надо будет следующим образом модифицировать текст.

После строки 247 (пдф. Сабмита) надо вставить фразу.

Since standard methods of goodness of fit assessment are not appropriate for nonlinear estimations (Spiess, Neumeyer, 2010) the quality of fitted logistic curves was evaluated by visual inspection of correspondence between observed cumulative curve and fitted regression line (ES ++++). Additionally we inspected the level of similarity between observed date of Peak and predicted Middle-of-Season date (ES +++), which expected to be close to each other. Both methods gave acceptable results.

Spiess, A. N., & Neumeyer, N. (2010). An evaluation of R 2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC pharmacology*, *10*(1), 6.

+++++++++++++++++++++++++++++

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.

Спасибо за замечание! Действительно, надежного и общепринятого метода восстановления пропущенных значений во временных рядах нет. Мы стояли перед выбором - либо удалять из массива годы, в которые какой-то из параметров не был измерен, либо использовать тот или иной метод восстановления пропусков. Поскольку основная ценность наших данных заключается в их продолжительности мы решили пойти по второму пути. Мы основывались на следующих соображениях. Во-первых, таких пропусков было немного (6 пропусков в средовых параметрах и в 14 случаях мы воспользовались методом восстановления пропусков фенологических показателей видов, когда логистическая модель, описывающая кумуляту обилия видов, либо не была подобрана, либо давала неадекватные оценки событий, см. Материал и методику). Во-вторых, метод SSA позволяет реконструировать пропущенные значения с учетом общей тенденции многолетних изменений, имеющейся во временном ряде. Это приводит к минимальному смещению восстановленных данных. В-третьих, мы провели специальный анализ (см ES+), основанный на симуляции пропусков, который не выявил явных тенденций к смещении реконструированных значений, относительно восстановленных.

++++++++++++++++++

Нужно вставить в текст отсылку к эдетронному приложению ES+.

++++++++++++++++++

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).

Спасибо за ценное указание на нашу неточность. В действительности мы использовали два метода оценки присутствия многолетних трендов.

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.

Использование регрессионного анализа было продиктовано необходимостью оценки того, на сколько дней произошло смещение тех или иных фенологических параметров за период наблюдений. Это можно оценить только зная 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. Поэтому мы воспользовались методом, предложенным для поиска пространственных градиентов (Clarke & Gorley, 2006). Кратко суть этого метода сводится к тому, что временой ряд превращается в матрицу эвклидовых расстояний (фактически это матрица квадратов 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 уровня значимости (adjasted p-level) за счет False Discovery Rate monitoring procedure (Benjamini, Hochberg, 1995).

++++++++

Где-то в тексте (например на строке 294, сразу за подщаголовком). надо вставить фразу.

The presence of long-term trends in parameters dynamics we assessed by two ways. Firstly, by fitting of linear regression model. Secondly, we used model matrix approach (Clarke & Gorley, 2006) comparing the observed time series with model “ideal” gradient represented by numerical series: 1961, 1962…. 2017, 2018.

++++++

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.

Необходимо было оценить связь между двумя матрицами: зависимой матрицей (фенологические события для видов) и матрицей-предикторов (ряды фенологических событий в средовых параметрах и средние показатели обилия видов). Вычислительные особенности Canonical Correspondence Analysis исключительно полно описаны в Legendre, Legendre, 2012.

Единственной модификацией анализа, примененой в нашей работе, был подбор оптимальной модели, то есть формирование такого набора переменных из матрицы-пердиктора, которые наиболее сильно коррелируют с зависимой матрицей. Подбор оптимальной модели описан в Borcard et al (2018). Этот результат изображен в виде биплота (Fig.6).

+++++++

На строке 335 вставить ссылку

CCA comprised several steps. Firstly, collinear predictors with variance inflation factor exceeding 5 were rejected and were not included into the model (Oksanen, 2011).

Oksanen, J. (2011). Multivariate analysis of ecological communities in R: vegan tutorial. *R package version*, *1*(7), 1-43.

+++++

<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).

Спасибо за очень правильное предложение. Мы всячески поддерживаем идею reproducable research. К сожалению, за время, отведенное на ревизию, мы не успеваем полностью подготовить R-код, включающий все необходимые комментарии. Кроме того, пока не решены юридические вопросы о выкладывании в открытый доступ первичной информации по многолетним наблюдениям, а без публикации этих первичных данных код, работающий с ними, теряет смысл.

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.

Сезонные тренды имеет смысл удалять только для временных рядов обилий видов. Остальные временные ряды (даты фенологических событий) включают единственное для каждого года значение и не имеют сезонного тренда. Мы рассматривали вариант анализа, в котором обилие видов было бы разложено на составляющие (main trend, seasonal dynamics and some residuals) с помощью SSA. Несомненно это дало бы более адекватную оценку многолетних изменений обилия. Однако мы отказались от этого анализа в пользу более грубого, но более универсального анализа, когда все изучаемые временные ряды обрабатываются однотипно. Выбирая более грубый анализ, да еще и с массой поправок на множественные сравнения (см. Выше), мы старались уменьшить вероятность ложноположительных сигналов многолетних изменений фенологических показателей.

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. Однако он потребовал бы тонкой настройки моделей, подобранных GLS. В частности потребовалось бы настраивать параметр “correlation” в функции gls() из пакета “nlme”, подбирая разные варианты моделирования автокорреляции. В этой ситуации пришлось бы подбирать для каждого вида и для каждой зависимой переменной особенную, оптимальную, форму автокорреляционной функции. Это привело бы к тому, что для каждого вида будет подобрана своя собственная модель с большим количеством своих собственных параметров. То есть мы значительно усложнили бы анализ. Единственное на что повлиял бы такой подход - это на оценку статистической значимости оценок параметров линейных моделей. К тому же, идя по этому пути, нам пришлось бы решать вопрос о типе статистического распределения, соответствующего тем или иным переменным. В частности не факт, что правильно было бы моделировать поведение фенологических параметров (количества дней от начала года, то есть счетных величин, которые скорее подчиняются poisson distribution), с помощью подходов, основанных на gaussian distribution. Поэтому в нашем, предельно упрощенном анализе мы специально отказались от параметрических методов оценки статистической значимости в пользу непараметрических методов. 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 (see Literature section)

**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.

Спасибо за совет! Мы создали специальное электронное приложение ES ++, в котором приведены все фенологические характеристики, использованные в работе.

+++++++++

Надо вставить в текст отсылку к электронному приложению ES ++.

+++++++++

Коля!

По ходу дела обнаружил, что предложение на строке 252 надо дописать

In 1963, 1972 and 1990, the observations did not adequately describe the cumulative for *C. glacialis* (a very short period of the species presence in the plankton fell on the intervals between observations).

Надо так

In 1963, 1972 and 1990, the observations did not adequately describe the cumulative for *C. glacialis* (a very short period of the species presence in the plankton fell on the intervals between observations). The same was in the case of *Pseudocalanus* spp. in 1961, 1962, 1994, 1998.

+++++++

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 ES +).
2. Only few percent of missed values 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.

!!!!!!!!!!

У нас было два способа оценить тренды.

1. Это угловой коэффициент линейной регрессии.
2. Мантеловская корреяция с модельным линейным градиентом. Так что мы имплицитно использовали два способа оценки.