Both reviewers, and I, consider this to be an excellent preprint. The analyses that are conducted across four impressive datasets reveals several important results. In particular, the analyses reveal that a lack of good baseline data make it close to impossible to assess long-term abundance trends, and this is particularly the case with short-time series exhibiting non-monotonous dynamics.

One reviewer has very few comments, stating that, for them, this was a rare occassion to read such a well-produced, and important paper that requires little modification. The suggested changes are minor. The second reviewer has a number of suggested edits, all that seem sensible, and none that will impact the results or conclusions. Given this, I am asking the authors to revise their manuscript to address these minor issues. Once that is done, I do not consider it necessary to seek re-review. Instead I will read the revised manuscript, and will recommend.

I would like to finish by congratulating the authors on a very impressive, and important, piece of work.

We thank the Editor for these encouraging comments and the reviewers for their supportive remarks about our work. We also appreciated that the editor and the reviewers devoted some of their time to contribute to a new model of scientific publication. A point by point response to the comments of the reviewer can be found below. In the version of the manuscript sent per email, editor and reviewers will find a track of all modification done on the main text.

Reviews

Reviewed by anonymous reviewer, 22 Mar 2022 22:37

The authors use a series of simulations, some using independent large, published datasets to thoroughly address the issue of how artifacts of methodology can generate different conclusions about long-term arthropod population trends. Surveys taken over too short a period can be especially prone to give results reflecting only a portion of a taxon's nonlinear or non-monotonic population dynamics, predicting an abundance trend that differs from the true long-term trend. The results make intuitive sense, but the simulations bring rigor and statistical demonstration of how this happens. The analyses are particularly relevant given the conflicting conclusions being reported in the literature about long-term arthropod abundance trends: some showing widespread arthropod decline, others that arthropods are increasing, and others that no substantive change is detected. This paper clarifies how and why that can happen. There are no easy fixes, because there is no control over monitoring in the past. But it seems clear that predicting long-term trends from short-term datasets is particularly hazardous. My comments are mostly related to presentation and the need for some better explanation in spots.

We would like to thank the reviewer for his/her comments on the manuscript, which helped us to clarify the writing and the figures. We agreed with all the comments and changed the text accordingly, except two, for which we argue below why we would rather retain our initial choice (comments regarding Figure 4 and results).

L28-29: Change "abundance variations" to "variation in abundance"

Done

L52: Change "but" to "and"

Done

L78-83: The phrase "temporal coverage" is used 4 times in these lines, but it is not obvious what it means. Does it mean the particular span of years (e.g., 1970-1995) over which a taxon or community is sampled? the number of years for which data are available regardless of which particular years those are? the length of time series? The authors should define it so the reader is not in doubt.

We now define temporal coverage at its first use (Line 78).

Fig. 1: Excellent figure for getting the point across. Would it be possible to provide more separation between the red, green, and yellow lines in panels a-d? As it is, there is so much overlap it is hard to distinguish the three lines.

Sure, done.

Also, for all figures relying on color, it is best practice to avoid red and green in the same figure because they cannot be easily distinguished by those with red-green color blindness (~8% of males and 0.4% females).

We apologize for that, we changed the colors.

L106: Please explain what is meant by "annual occupancy estimate". Is this presence/absence per geographic unit? What is being occupied?

Here, an occupancy estimate is the proportion of 1km2 grid cells occupied by a species over a country. We added this definition line 106.

L108: change "form" to "from"

Done

L109: "aggregated" is used twice, once for taxonomic resolution, which is clear. But the first use: "annual estimates of arthropod abundances aggregated by van Klink…" is unclear. What particular estimates did van Klink aggregate, and from where? Who generated the estimates? I could go to the original paper and find out, but this should not be necessary for the reader to have to do to understand the study.

We used aggregated because van Klink et al. is a meta-analysis, so they did not produce abundance estimates themselves but collected and collated them from literature. Perhaps the word "aggregated" is misleading: "annual estimates of arthropod abundances aggregated by van Klink et al." is equivalent of "annual estimates of arthropod abundances from the meta-analysis of van Klink et al.", so we changed the sentence to be clearer.

For the four data sets in general: please give some basic background on the nature of each dataset. Who and how were the data collected, what was the geographic extent of each dataset. It does not have to be super detailed, but give the reader some idea of what the data analyzed in these different studies were and how they were gathered.

We added a Table S1 to fill this gap.

Fig. 2 caption: Instead of relying only on a narrative of "First we did this, then we did that…", please reference the various Steps and Goals to help the reader know exactly what part of the figure is being described.

Done

L118-119: change "but also" to "as well as"

Done

L121: "keeping"? Do the authors mean "using only"?

Yes we do, thanks. We modified the text.

L122: spell out what GBIF stands for.

Done

L132: what is meant by "time period gathering most of the data"? Do the authors mean the time period when most of the data were gathered?

Yes we do, thanks. We modified the text.

L136: delete "First of all, since" and begin sentence with "Because"

Done

L175: This should be a separate sentence: "Classic standardization…" Something is missing and I suddenly got lost; are you saying classic standardization is what you did? Or is classic standardization something you did NOT do (in contrast to the log/logit transformations) because it gives inappropriate weighting to species with lower variability in abundance? I assume that classic standardization is a good thing based on the corresponding lines in Fig. S4b, but please state explicitly to ensure the reader stays with you.

Sorry for this confusion. Actually, we meant that our transformation (logit or log transformation) gave the same weight to rare and common species, in contrast to classic standardization, which just gives more weight to species with low inter annual variability in abundance. We changed the sentence (see Line 177-180) and the legend of figure S4.

L186-189: "We removed abundance trends…" This sentence is a bit hard to follow because of its length; please consider breaking into two sentences.

We split the sentence in two.

L195: change "three-ways" to "three-way"

Done

L206: change "calculated" to "calculate"

Done

L213: change "depend" to "depends"

Done

Fig. 3 caption: For box plots, please describe the parts of the plots; e.g., what does each box encompass (67% of observations? 95% of observations? upper and lower CI?), what is the thick black horizontal line (mean? median?), what are the thin vertical lines (range? SE? SD?), what is the dot above the box at turning points = 9?

Done. The boxplots represent minimum values (end of the bar), 25% (Q1, lower bound of the box), 50% (thick black line) and 75% (Q3, top bound of the box) quantiles and maximum values (end of the bar) of the distribution. Points with values outside of the range [Q1-1.5(Q3-Q1), Q3+1.5(Q3-Q1)] are considered as outliers and represented as points.

Also, "Proportion" is misspelled on the y-axis.

Corrected.

Fig. 4 caption: Describe symbols comprising box plots as described above for Fig. 3.

Done.

Fig. 4d: It might be more helpful for the reader to shrink the scale on the y-axis so that details of variation around the true growth rate are visible, and find a different way to indicate the outlier values that cannot be shown (e.g., break y-axis // to show extreme values).

We added a Figure S7, which shows the left panel of Figure 4d without the outliers. We actually had this very debate among authors before the initial submission, and we decided on purpose to keep the figure as it was for two reasons: (1) the outliers are at the core of what we want to show: when time series become short, uncertainty around the estimation increases a lot, which justifies not using a broken axis or transformation and rather show that uncertainty can increase substantially ; (2) the details of the variation around the true growth rate is visible for 3 values of decline out of 4, allowing the reader to judge potential bias in case of non-decline to moderate decline, which, we agree, is also a key message of the figure.

L271-273: This is an important conclusion and a main take-home message of the study. It seems misplaced in the Results, and should be in the Discussion instead.

Same is true for L300-302.

Indeed, these results are key and this is why we wanted to briefly present their general meaning in one sentence, as early as in the results, and not wait until the discussion. However, these two results are discussed more deeply in the first part of the discussion as well.

Fig. 6c caption (L333), (and 6c x-axis title): change "relatively" to "relative"

Done.

L323: Collembola is misspelled (only one l in "-bola")

Corrected

L325: Coleoptera is misspelled

Corrected

L361-363: I'm not understanding this sentence. I wonder if the problem is the phrase "posterior to 1990". Does this mean before or after 1990?

Posterior to 1990 means after 1990. We modified the sentence to make it clearer.

L381: "anterior to the rise" – here and elsewhere, instead of "anterior" and "posterior", please use "before" and "after" when talking about time. Their meanings are immediately and unambiguously understood in this context. Anterior and posterior are spatial descriptors for "in front of" and "behind", and their use in a time context is disorienting and ambiguous.

Done.

L383: change "would" to "will"

Done

L385-387: change to "Whether or not scientists can manage to obtain…" "…to effectively turn back the clock, our…"

Done

L389: change "such that our" to "so our" ; change "

Done

L139-140: change to "does not introduce new elements that may affect the reliability…"

We did modify the sentence but did not retain the phrase "may affect the reliability", as we think it would change the meaning

L393-394: not clear what is meant by "joining criticisms". Do the authors mean "inviting" or "provoking" criticism? or something else?

We modified the sentence to explain clearly what we meant.

L403: change "diffuse over" to "apply widely over"

Done

L415: Change "Consistently" to "Consistent".

Done

L415, change to …"we show that transforming data with (log(x+1)) before statistical modeling, as done by Crossley…" I am not sure my suggested change reflects the authors' intended meaning.

It does, thanks, we modified the sentence.

L416: unclear what is meant by "instead of adapting model structure": The log transformation does not adapt model structure as intended? or the authors should have tried adapting model structure instead of a log transformation? or something else? What does "adapting model structure" mean?

Here, adapting model structure means using a GLM (Generalized Linear Model) to model abundance trend, instead of transforming the abundance data using log(x+1) and further applying a LM (Linear Model). We added this explanation to the initial sentence for clarification.

L414-417: This whole sentence is hard to follow and should be restructured, perhaps as two sentences.

We clarified this sentence and hope it is clearer now.

L422: change "questions" to "brings into question"

Done.

L423-426: Hard to follow sentence because of length and too many thoughts at once. Break into two sentences: "…taxonomic group. Losses and gains…"

Done

L427: change "are" to "is" ; change "sometime" to "sometimes"

Done

L426-429: another sentence that is hard to follow because it is trying to accomplish too much at once. Break into two sentences: "…sometimes non-monotonous. This suggests that…"

Done, thanks.

L428: change "although" to "despite"

Done

L429: not sure what the authors are trying to say with "should at least always be associated to temporal coverage" ; associated in what way? associated how?

We mean that any measure of temporal trend should be associated with the temporal coverage of the data, for example "a decline of XX% on 1970-2018", which is rarely the case. We modified the sentence.

L437: change "from a non-successful submission" to "of a previous version

Done

Supplementary: The needed material is all in this pdf. However, the Supplemental material will need to be better organized. It should start with a title page with authors followed by a coherent table of contents that includes everything in the document. As presented, the sub-section for Code is something of a document-within-a-document. This includes its own Table of Contents referencing page numbers (even though the pages of the Code sub-section are not numbered); also the supplementary figures and table are indicated to be at the end of the Code section, but actually come before the Code section in pages that are numbered differently. Providing a holistic Table of Contents at the beginning of everything will also help the reader find the relevant materials referenced in the main paper's text.

Done. To be clearer we separated the code from the other supplementary materials.

Reviewed by Gabor L Lovei, 22 Mar 2022 20:13

This is one of those extremely rare manuscripts where I do not suggest substanial changes - and in fact, very few changes. One is almost ridiculous - I do not think that you should capitalize the word 'arthropods'.

The text is well written, clear and understandable, and even the English is very good. Even if I were to write a little differently, it'd be nit-picking to suggest linguistic changes.

We thank the reviewer for this very nice comment.

The only suggestion I can think of is about defining baselines. One possibility is to establish a mean abundance over a period of X years, and consider that as a baseline. As I understand it, the MS considers single years as baselines - and then the choice of years will very clearly influence the trend/slope. Another possibility is not to force to find linear trends but to employ loess regression.

We agree with the reviewer that a baseline based on an average over an historical period is a good option to define a clear baseline. Debate can follow about the span of this historical period, but it is not the point here. What we wanted to stress is that any measure of abundance trend should be regarded with the temporal coverage of the data. For example, "a decline of X% between YYYY and YYYY". This simple precaution is rarely taken, at least in titles and abstracts. To find this information, readers have to go in the methods, which may lead to dissociate abundance trend from the temporal coverage of the data (time frame, duration, etc) and generates difficulties for comparing results. First, scientific studies need to bring forward the baseline considered in the data, and more generally, specify the temporal coverage (starting and ending date) of their study more explicitly. When defining baseline, historical average would indeed make more sense than taking one year as reference. Actually, in practice, the risk of referring to a single year only is still overall limited. Since most of the abundance trends are estimated using statistical models modelling linear or polynomial temporal trend, the intercept of the model (baseline if the first year is set as year = 0) does not depend only on the first year but also on following years, buffering the effect of the first year of the dataset and getting closer to the idea of averaging value on an historical period. Nevertheless, we agree that more standardized practices in calculating abundance trends should be adopted to insure reliable comparisons.

Other than these minor comments, I was glad to reada a clearly argued, well written MS.