Four decades of phenology in an alpine amphibian: trends, stasis, and climatic drivers

Omar Lenzi, Kurt Grossenbacher, Silvia Zumbach, Beatrice Lüscher, Sarah Althaus, Daniela Schmocker, Helmut Recher, Marco Thoma, Arpat Ozgul, Benedikt R. Schmidt

General comment

We are grateful to both the editors and the recommenders for their comments. The main concern regarded the terminology and the presentation of the method used to account for uncertainty in the calendar dates and the related results. We agree with this concern and we addressed it as best as possible by clarifying the terminology and improving the description of the method and the presentation of the result. All the other comments were addressed as well. We believe that the revised manuscript is now of higher quality thanks to the comments by the reviewers and the following revision. Below we provide a point-by-point response to the comments. Line numbers refer to the updated version on bioRxiv. We also uploaded a copy of the manuscript where the changes are visible.

Response to the Recommender

Comment*:* *Dear authors, after a first review both reviewers agreed that your manuscript is very clear and without doubts an interesting contribution to the field. However, both (specially reviewer 2) suggested several points that need to be clarified or modified. Please keep in mind these comments in your new version.*
I think that including the previous points and clarification current concerns could be a major improvement of it.

Response R1: We thank the recommender for the encouraging feedback and for the opportunity to resubmit a revised version of the manuscript.

Response to Reviewer 1

Comment: *As general observations, I might suggest to reduce the number of citations, and maybe work around the title to make it more descriptive and representative of your work.*

Response R2: We thank the reviewer for the suggestion, but after careful deliberation we would like to keep the title as it is. We also kept all the references. We believe it is necessary to cite previous work in this field of research even if the number of citations is high. For an online-only journal, we hope that the number of citations does not matter.

Comment: *lines 109-115. These are generalizations that only apply to temperate/cold regions. Tropical regions will not experience any of these events, and hibernation does not occur. So I suggest to adapt the paragraph to make generalizable statements or to indicate its regional limitations. The modified paragraph will still fit perfectly in you discursive line.*

Response R3: We thank the reviewer for the comment, with which we agree. However, we believe that there is some form of seasonality in most climates and therefore eggs could be exposed to unfavourable conditions. Nevertheless, we adapted the paragraph (line 111) to include a comment on the regional limitations. We also added a clarification in the discussion, at line 589.

Comment: *Line 369. Is this some sort of implementation of the Kaiser criterion? I'm familiarized with the criterion of keeping principal components whose eigenvalues are higher than one, but I haven't found anything on using the standard deviation.*

Response R4: This is indeed the Kaiser criterion, essentially. The *prcomp* function of the *stats* R package gives as an output the standard deviation of the principal components, which corresponds to the squared root of the eigenvalues (see RDocumentation of this function:

https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp). Therefore, components with a SD below 1 must also have an eigenvalue below 1, and can be discarded following Kaiser's rule. We changed lines 389-390 and the description of Table S5 to make it clearer.

Response to Reviewer 2

Comment: *The authors use simulated datasets to assess the plausible consequences of the uncertainty in the dates of the first observations on parameter estimates (ie toads may have come a few days earlier as observations are not made every single days and they may have been missed). They do this by "allowing the date of the start of the breeding to be as early as seven days before the initially assigned first capture night". First the authors need to define what distribution they used for the simulated dates (uniform distribution between 0 and 7?).*

Response R5: We thank the reviewer for the comment and we agree that our description was incomplete. We changed the sentences at lines 207-209, 211-212, 269, 298, 330 to make it clear that the dates are drawn from a uniform distribution.

Comment: *Second the intervals they get from these simulated datasets are not confidence intervals in the sense that they only include one source of variability – i.e. proper confidence intervals are those based on all sources of variability and are closer to those considering the standard errors given in Table 1. I would to avoid any confusion put the results associated to the 1000 simulated datasets in the SM (your analyses are robust with respect to errors in records), and provide estimates with proper 95% CI in Table 1 (with P-values if you want, but I am not very fond of P-values). And don't call it bootstrapping (eg as in table 2 and 3) since you are not resampling data (or simulating residuals from the model as you would do with parametric bootstrapping, even if you do the simulations parametrically for one component associated to the error).*

Response R6: We thank the reviewer for pointing this out. We agree that the terms we used were incorrect and we changed them. We do not use the term bootstrapping anymore, and instead of calculating 95% CI we now calculate the 2.5^{th} and the 97.5^{th} percentiles for each estimated parameter, to give an idea of the spread that the parameters values can have. We corrected the text where necessary (e.g., lines 204-207, 212-215, 267-270, 280-282, 295-301, 328-333). We also followed the suggestion of changing the presentation of the results, giving the priority to the results of the analysis conducted on the originally assigned dates, and put in the SM (or just a sentence in the main text) the results of the analysis on the 1000 simulated datasets (see Table S1, S2 and S3 and lines 352-360, 405-409, 420-427, 451-453, 459-463). We also updated Figure 1, as it was not showing the piecewise regression on the originally assigned dates, but it was showing the piecewise regression on the mean dates out of the 1000 simulated datasets.

Comment: *Finally, please remove P-values for nonsensical hypotheses, i.e. for the intercepts (you are testing if toads were breeding at day=0 in year 0…), and rescale your predictor year so that the intercept has a value which is relevant (eg year = 0 means first year of the study, 1982), not an intercept = 6342 (Table 1) which is day number again for year=0.*

Response R7: We rescaled the predictor year as suggested (year 1982 becomes year 0) and added a sentence about it in the text (lines 201-202). We also removed p-values when not appropriate.

Comment: *I don't understand why you include the product (what you call an interaction l. 381) of PC1 and PC2 as a predictor since it is far from being significant and may lead to unstable parameters for the main effects. I fully support to include "non-significant" parameter estimates when they represent the main focus of the study (eg part of the design, or in your case the lunar cycle), but in this case it is not obvious why they should interact multiplicatively.*

Response R8: We removed the product of PC1 and PC2 from both the population-level and individual-level models. The new results essentially do not differ.

Comment: *I would be perhaps a bit more careful when writing that assessing repeatability will help you "understand the role of the genetic component" (abstract, or after l. 481). As you write it could represent an upper limit, but the genetic component could be anything less than that, so one cannot really infer much about genetics, except perhaps as you write that it might be low. Just focus on assessing the individual heterogeneity (the complement of repeatability) which is interesting enough. That this heterogeneity seems to be low is important to know, for example when studies do not have access to individually-based data.*

Response R9: We thank the reviewer for this comment. We toned down this message in various sections of the manuscript (abstract, line 139, 468-470, and paragraph 508-527)

Comment: *The increasing variability in recent years is also something interesting and the discussion after l. 548 is thorough. I understand that one wants to relate it to possible extreme events (which could be due to an increase in environmental variability without having "truly" improbable events – ie not just the 5% tail but events with much lower probability, such as the summer 2003), but it could be nice to show the distribution of residuals on the figure for different time periods to show this explicitly. Is it just a change in the variance, or a change in "extremes" – seems the former is a better description.*
It could be relevant to cite papers by R. Prodon (most recent is Prodon et al. 2020) which have shown nonlinear changes, but that have been related to nonlinear changes in weather patterns.

Response R10: We created a figure (Figure S3) where we show the residuals of the 1000 piecewise regression on start and peak breeding dates, divided in 4 decades (1982-1991, 1992-2001, 2002-2011, 2012-2021). We notice that for the start of the breeding decades 3 and 4 show an increased density around values further from 0, compared to decades 1 and 2 (i.e., they show more pronounced tails). Regarding peak breeding we see less of a difference, even though decade 4 shows a peak of the distribution further away from 0 than the other three decades. We added text concerning Figure S3 on lines 363-364 (Results), 584 (Discussion)

We also cited Prodon et al 2020 in the introduction (lines 101-103) and discussion (lines 561-564), as an example of an environmental phenomenon, different from climate change, that can influence the phenology.