Response to Reviews

Eric Scott

8/9/2021

Dear Editor,

Thank you for the opportunity to submit a revised version of our manuscript “Delayed effects of climate on vital rates lead to demographic divergence in Amazonian forest fragments” to Global Change Biology The feedback we received from the referees was clear, fair, and very helpful, for which we would like to extend our thanks.

Overall the referees were pleased with the submission, but they also had several excellent suggestions for improving the analyses, interpretation, and clarity of our presentation. We have carefully gone through the manuscript to incorporate their suggestions. We identified three main themes in the reviews:

1. A need for a more in-depth explanation of distributed lag non-linear models, both for a technical audience (for replicability) and for a general ecology audience (for interpretabillity). We have expanded and revised the statistical modeling subsection to improve on this area.
2. A need for more discussion of the caveats and limitations of our modeling. For example, since we were not able to test for a statistical interaction between habitat and lagged SPEI, we have revised the results to be more clear about which conclusions are strongly supported by statistical results.
3. Finally, reviewer 3’s comments made us re-evaluate our approach to limiting the number of “knots” (or maximum basis complexity) in our models. This (along with an updated version of the data with some duplicate plants corrected) lead to some cascading changes in our methods, results, and discussion, although the general conclusion that there are lagged effects of SPEI that differ among vital rates and habitats remains unchanged. Increasing , the maximum basis complexity, resulted in changes to the shape of the crossbasis smooths, and have ensured that increasing $k$ further yet does not result in further changes to the shape of the smooths (see model validation supplement). The change in crossbasis shape changed the statistical significance of some lag times for some models. This increase in increased the computational time for models in a way that scales multiplicatively with random effects terms such that it was no longer feasible to include individual-level random effects. We ensured that removing these individual-level random effects did not bias the results using some data subsets (now mentioned in the updated methods).

The suggestions from all reviewers really improved the manuscript, and we are very thankful for such thorough and thoughtful reviewers. We have addressed all suggestions and comments in detail in our response to reviewer comments below.

Sincerely,

Eric Scott

# 1 Reviewer 1

**General comments**

Scott and colleagues present the results of an interesting series of analyses designed to disentangle the lagged effects of climate and forest fragmentation on the demographic performance of a tropical understory plant. As ecologists seek to understand not just how changes in mean climate, but also changes in extremes, will influence population persistence, studies such as this that capitalize on long time series are critical to making progress. The authors find that the effects of swings in precipitation (both drought and particularly wet years) affect survival, growth and flowering, but that these effects are both lagged and differ by whether plants were in continuous forest or the smallest fragment size (1 ha). Overall, it’s a well written paper, and an application of a relatively new modeling approach to address the questions. The results highlight both the challenges with understanding the complex factors driving population persistence and also how critical it will be to grapple with them. Below I provide suggestions for improvement, particularly aimed to make the new modeling approach more accessible.

We are delighted that the reviewer found the manuscript engaging and valuable. Our responses to specific suggestions follow.

## 1.1

By the end of the introduction, I expected that the paper would focus on drought, and was then puzzled by the choice of the words ‘extreme precipitation’ – this of course makes sense later that the results are focused both on the effects of drought and extreme wet. The introduction could set up the reader better to understand both. Since drought is well covered, adding background and hypotheses for why extreme wet would have a negative effect on vital rates would help.

We edited a sentence in the introduction that cites Esteban et al. (2021) to clarify that both extreme dry and wet conditions reduce growth and survival of trees in the Amazon. *“…recent results suggest increases in the frequency and severity of extreme precipitation events* ***(both drought and extreme wet)*** *reduce survival and reproduction…”* (L67–69). Additionally, we added a sentence describing forecasted changes in precipitation in the Amazon (L64).

## 1.2

In the intro and discussion, I think it would also be helpful to describe a little bit more about mechanisms for why extreme precipitation matters more in fragments compared to continuous forest. I wondered if the authors thought this was too obvious (edge effects, etc), but I think it’d be helpful to spell this out more clearly. L466 – 467 offer a new hypothesis, but briefly revisiting the ‘old’ evidence/hypotheses also seems important.

At the reviewer’s suggestion, we have added a short paragraph (starting on L74) to the introduction explaining why the pressures of habitat fragmentation and climate change might interact (i.e. be more than additive).

## 1.3

One challenge of using a new modeling approach is making the methods of the paper understandable and to stand alone. As written in the methods and results, I found the implementation very difficult to follow. A slightly longer summary of the methods would help.

We have added an additional paragraph to the methods describing how the crossbasis smooth was fit using the mgcv and dlnm packages (L229–239). We are not entirely sure if this is the type of description the reviewer was looking for. If the reviewer(s) or editor see this as too detailed for the main text, we are happy to remove it, move it to a supplement, or simply replace it with a reference to the (now available) archived code for the analysis. See also our response to 3.3.

In the results, I was surprised to see only a few p-values and not much else about the models in terms of how they were assessed for goodness of fit. While I am not at all suggesting that p-values are the solution, I would have liked a little more sense of how well the models fit. One way to do that might be to show the raw data plotted with the model fits so the reader can see, for example, what kind of functions were drawn. Another might be to present something in table format (the comparable version of a table of AICs or of the results of frequentist tests).

Rather than comparing the fit of multiple models with different predictors, we took advantage of the select = TRUE option in the gam() function, which adds an additional penalty to smooth terms that effectively removes them from the model if they are not supported by data. Therefore we cannot show something like a dAIC table. However, we have now included a table with R2 values for each model indicating fit as well as estimated degrees of freedom and p-values for each term (Table 2). We have replaced many of the parenthetical p-values reported in the results section with references to this table. Additionally, we have added supplemental Figure S3 which shows the curves as separate panels for continuous forest and fragments with raw data superimposed.

## 1.4

Related to above, I think it’s important to see more of the raw data than are presented in Fig. 1. (I’m thinking of a version of Fig 3 that shows size in t+1, survival, and flowering rather than effect size on the y-axis, with the data and lines showing the shape of the GAMs). This could be an appendix, but it would help with interpreting Figures 3 – 6.

We changed Fig 3 to be on the response scale as suggested (see our response to 1.9), but showing the raw data made the plots visually cluttered, in our opinion. Another way we approached this comment is by excluding areas of the heatmaps in figures 4-6 that are far from observed data, equivalent to not drawing regression lines beyond the range of data. We implemented this with the exclude.too.far() function in the mgcv package. Displaying raw data on the heatmaps was too messy and not informative in a 2d plot.

## 1.5

I suggest that a little more about the limitations of this modeling approach in the discussion would help set the context for interpreting the results and make the paper more robust.

The main limitations of this method are of data (see also reviewer 2’s comment 2.14). We have added some text to the discussion describing limitations found in previous simulation studies, and new unpublished evidence that spatial replication may be more valuable than long time series for detecting demographic effects of the environment (L542–550).

In some ways it seems like the analysis is adding more precision (by using monthly climate data) than the plants perceive and that would be measured in a yearly census.

Counter-intuitively, averaging weather conditions over an entire year may obscure ways plants perceive environmental variation. For example, a drought in the wet season may have a different effect than a drought in the dry season. This differential response to season- or month-level climate data is observable in yearly plant census. Collapsing finer temporal scale environmental drivers to yearly averages (or a few means of arbitrarily chosen climate windows) is exactly what we avoid with this method. The caveat to this is that without finer scale demographic data, one may not be able to discern the cause of an apparent delayed effect, which is why we expand on several possible mechanisms for delayed effects in detail in the discussion.

While the authors mention this, what about adding something short about when using this type of analysis is critical and when it might not be (should all demographers adopt this approach? Why or why not?)

We have added additional detail to the conclusions recommending that all demographers consider lagged effects when possible in light of recent research (including this study) (L536–541).

## 1.6

I was curious about why the paper did not include analyses of recruitment. Perhaps it is beyond the scope of this paper, but I wondered how new recruits responded to drought and wet also, both in number, as well as how big they were in the first year.

The reviewer is correct that recruitment is an important aspect of demography that could be effected by SPEI and fragmentation. However, we believed the analyses in this paper were complicated enough to not include additional analyses. We intend to follow up this paper with analyses of recruitment and integrating all vital rate models with an integral projection model. We simply felt doing all of this (describing and interpreting DLNMs, modeling recruitment, and integrating all into an IPM) would be too much for a single paper.

## 1.7

I had a hard time linking these statements in the results to the figures.

L272–273—How do you see disparities in plant size here? I thought this was about growth?

Models which predict size in year t+1 as a function of size in year t, such as ours, are often described as growth functions because they can be used to project change in size (i.e. growth). However, we agree that it would be more clear for this section heading to be “size” rather than “growth.”

L277–278—How does this show the effect of size on growth is greatest for mid-size plants? More generally, I got a little confused with the flipping back and forth between growth & plant size in the results and discussion. More precise language would help.

We have replaced this sentence. See our response to 2.5.

L331–332—How can you see in this figure that the response is ‘more muted as indicated by the shape of the crossbasis function?’

Our point was actually that SPEI had less of an impact on flowering probability in fragments. We edited this sentence to read: “SPEI had less of an effect on flowering probability in 1-ha fragments as indicated by the narrower range of the evaluated smooth (Figure 6b).”

## 1.8

L442-443 – What is a mechanism for why plants might flower more after drought? (the discussion talks about the opposite) The next paragraph kind of goes there, but the link could be tighter.

We are a little hesitant to speculate too much here, as not much is known about the optimal microclimate conditions for flowering in *H. acuminata*, but we have revised this paragraph (see also our response to 2.16). It now includes the following possible explanation: *"Drought conditions could be favorable for* H. accuminata *flowering due to increased temperatures or decreased cloudiness associated with droughts (Pau et al. 2013),* *and the effects could be delayed due to the development time of inflorescences."*(L470-473).

## 1.9 Figures:

Fig 2 – This was only cited in the discussion as far as I can tell. Bring it back into the results?

We have added a sentence about this figure in the results.

Fig 3A – The results mention something about survival approaching 1 – I wasn’t sure if that was supposed to be visible here.

This was referring to a previous iteration of the figure where the y-axis was on the response scale. We have reverted the figure to this version (see response below) and this should be more obvious now.

Fig 3 – I mentioned this earlier, but I wasn’t really clear on how to interpret effect size. I see the explanation in L 539 in the legend, but I don’t really know what that means. Since this is such an important result of the paper, it would help the reader if they were walked through this idea. What is a negative vs. positive effect size in relation to plant size (does this compare to the median size)? I am likely missing something here. (I think the challenge for me is that I kept wanting the figures to fit into a more standard demographic paper – like the probability of survival based on size, but then I realized something different was being shown here.)

These plots were the evaluated smooths on the link scale. We agree that they were difficult to interpret in that form, so we have added the model intercept and back-transformed them to be on their respective response scales (e.g. probability for flowering and survival). This is equivalent to the marginal effect of log(size\_t)—that is, without considering the additive effect of lagged SPEI or the random intercepts for different plots and for plants that had not flowered in the previous year (the reference level for the flwr\_prev factor we included as a parametric covariate). We have now clarified that in the figure caption.

Figs 4 – 6 – Where the isocline circles are large, does that mean it’s a very large time span? I was a bit confused when describing the results of Fig 4b (L296-297), for example, the lag is described as the preceding 13 months in the 1 ha fragment, rather than 0 – 13 months (although now I see maybe the word preceding was meant to indicate this whole range). A little bit more explanation in the legends would help these figures be more intuitive.

We’ve revised the figure captions to clarify: *“Outlines show combinations of SPEI and lag time that have a significant effect on survival, defined as areas where the 95% confidence interval around the response does not overlap the intercept.”*

# 2 Reviewer 2

**General comments**

The manuscript by Scott et al analyzed an eleven year long demographic data set on the herb, Heliconia acuminata in the Central Amazon to evaluate the effect of drought and fragmentation on survival growth and fecundity. The authors also employ distributed lag non-linear models to explore delayed effects of drought. This work is an important contribution adding to the sparce literature documenting the role of weather variables on plant vital rates. The manuscript is generally well written. I hope the authors will find my comments helpful in improving this already very good manuscript.

We are glad to hear the reviewer saw this work as important and well written. We address specific concerns and suggestions below.

## 2.1

The results are all about the effect of SPEI on plant vital rates, but the abstract implies that the manuscript presents the effect of temperature and precipitation on vital rates separately.

We have edited the abstract. Now the sentence to which the reviewer refers reads: *“We used distributed lag non-linear models to understand the delayed effects of climate (measured as standardized precipitation evapotranspiration index, SPEI) on survival, growth, and flowering.”* (L33).

## 2.2

Line 51-52. I find this sentence confusing. What specific demographic mechanisms are you referring to? Do you mean that high mortality rate lead to extinction and we don’t know what causes high mortality rates?

The reference, Bruna et al. (2009), discusses the lack of knowledge of what demographic vital rates (e.g. survival, reproduction, recruitment) in which life stages (e.g. small or large individuals) contribute most to population declines in fragments. We have re-ordered this sentence (now L50) to emphasize the hypothesis that fragmentation alters survival, growth, or reproduction with the caveat that few studies have determined which of these vital rate changes are driving extinctions.

## 2.3

Line 146. Can you explain what you mean with “the height of each plant to the tallest leaf?” If the tallest leaf is growing more horizontally the height of the plant would be shorter than the length of the largest leaf?

This is just the height of the plant from the ground to the tallest leaf leaf tip, not the tip of the youngest leaf. The reviewer is correct that the height of a plant growing more horizontally (for example if it was knocked over by a branchfall) could be shorter than the length from the base to the end of the tallest leaf. We edited the description to read: *“…the height of each plant (i.e. distance from the ground to the tallest leaf tip)”* (L152)

## 2.4

Lines 251ff: I don’t see a reference of Fig 2 in the result section.

We have added a sentence to the results that refers to this figure (L284).

## 2.5

Line 277-278: I don’t see in Figure 3a that the impact of size on growth was greatest for mid-sized plants in continuous forest. It looks to me that the 1-ha and CF lines are extremely close together and not statistically different. Perhaps you can say that there is a tendency of small plants having a larger growth in fragments while large plants grow slightly more in CF.

We thank the reviewer for pointing out this mistake. We have edited this line to read “*Although plants were significantly larger in continuous forest compared to fragments, the effect of size in year t on size in year t+1 is nearly identical in the two habitats.*” (L303). See also our response to 1.9.

## 2.6

Lines 287-289. While this statement is true, there are also years where flowering is lower in CF (e.g., 2002 similar values as following the drought in 2003 in fragments).

We thank the reviewer for pointing this out. In fact, with updated data, 2006 actually shows the largest disparity in proportion of large plants flowering with a greater proportion flowering in 1 ha fragments. However, the total number of flowering plants in fragments was always lower due to 1) lower plant density in fragments and 2) fewer large reproductive-sized plants in fragments. We have clarified this in the revisions.

I feel that the reporting of the results is a bit biased towards concluding that plants do worse in fragments compared to CF, especially since none of the result reported so far are significant.

We have added Table 1 to clarify what statistically significant results we found between habitats. The table shows the model intercepts ± 84% confidence intervals, which is comparable to testing a main effect of habitat. The difference in flowering probability between continuous forests and fragments is in fact significant, although the difference is slight.

## 2.7

Delayed effects: Since the authors fit separate models for plants in fragments and in continuous forest to allow for the crossbasis function to differ between habitats we don’t know if any of the small differences reported in Fig 4c, 5c and 6c are statistically significant. Hence, I would recommend interpreting the results concerning the difference between fragments and CF with caution.

Because we were not able to test for a significant interaction between habitat and lagged SPEI, we did not intend to suggest that our results implied statistically significant differences in the shape of the crossbasis functions between fragments and continuous forest. We have made changes throughout the manuscript that clarify this and remove some of the stronger declarative language about differences in the effects of SPEI between continuous forest and forest fragments.

I would like to see at least a justification for why representing the difference figures are meaningful.

We believe the difference figures (panel c in figure 4,5, and 6) are helpful in showing combinations of SPEI and lag time where the marginal effects on vital rates are most divergent between continuous forest and fragments. For example, at a lag of 0, wet conditions have a positive effect on survival in continuous forest and a negative effect on survival in fragments, leading to a large predicted difference in survival between habitats under those conditions.

Line 301-301: This is somewhat misleading because as you describe later SPEI at time lags 15-20 and 32-36 are significant also. I wonder if it might be better to describe the entire Fig 4b and then the entire Fig 4a. What springs to my mind is that the area that is significant is much larger in in fragments compared to CF. It seems that in CF only a few months are significant. Surprisingly they fall not consistently in the dry or wet season.

We see how this could be confusing and have taken the reviewers suggestion to include the 32–36 month lag when describing lags that are significant for each habitat.

## 2.8

Line 323-325: I don’t know why the growth of plants in fragments would affect the size of plant in CF.

This sentence meant to describe the difference between CF and fragments, not the absolute size in fragments. It has been re-worded to clarify (L348).

## 2.9

Lines 355-359: The analysis did not consider testing a statistical difference between habitat type (see comment 7), so I think this sentence is misleading (overinterpreting the results).

While we could not test for an interaction between habitat and lagged SPEI (i.e. whether the shape of the crossbasis smooth was *significantly* different between habitats), we did test for a main effect of habitat by looking for overlap in 84% confidence intervals of model intercepts (see also response to 2.6). However, to reduce confusion here, we have removed “significantly” so as not to imply statistical significance for all of “magnitude, directions and lag times” (L. 377). See also our response to 2.7.

## 2.10

Line 370: There are no confidence intervals in Fig 2, so we don’t know if the difference is significant.

Figure 2 shows the survivorship curve of plants tagged in 1998. These are exact counts and therefore do not have confidence intervals. They do not require statistics to interpret if they are different. We think the confusion may have come from the axis scale and label (“P(survived)”) which we intended to be interpreted as “proportion of plants surviving” but may have been interpreted as “probability of survival.” We have changed the scale to a percentage and the axis label to “% surviving.”

## 2.11

Lines 383 ff: see comment 7

See our response to 2.7.

## 2.12

Line 412: I think it would be very informative if you were to expand on this and explain how exactly your method differs from what others have used to detect delayed effects.

As reviewer 3 pointed out (3.11), our extension to previous approaches is the use of a 2-dimensional spline to allow for fitting non-linear responses to the environment. We have edited this sentence accordingly (L436).

## 2.13

424-427: I think this is exaggerating since the differences between habitat types is not significant.

See our response to 2.7.

## 2.14

You may also consider arguing that an 11-year long time series is not sufficient to detect delayed effects (see Tenhumberg et al. 2018)

We have added additional discussion on the limitations of DLNMs for detecting delayed effects to the discussion. See also our responses to 1.5.

## 2.15

Lines 437-438 change “starting expand” with “starting to expand”

We have made this correction.

## 2.16

Line 440: Can you explain a bit more in detail why drought increasing flowering probability two years in the future would be a bet hedging response?

We intended this paragraph to focus on one possible explanation for delayed effects—delays due to development times. We believe that the bet-hedging argument was confusing for two reasons: 1) it assumed that drought (as we measured it with SPEI) is detrimental to *H. accuminata* survival, which is not consistently reflected in our results, and 2) it did not address longer lags than the development time of a flowering shoot. We have therefore revised the example in this paragraph (starting on L462) to focus on literature that shows positive effects of temperature and negative effects of cloudiness on flowering in tropical plants—weather conditions both associated with droughts. See also our response to 1.8.

## 2.17

Line 448: Alternatively, drought may reduce nutrient availability such that after covering maintenance costs there is nothing left to allocate.

We thank the reviewer for the suggestion—in fact, effects of drought on maintenance costs is precisely what we were trying to communicate in this line. After shedding shoots and leaves during drought, maintenance costs may be reduced, and plants might invest more in rhizomes, possibly allowing plants to recover after drought to a larger size. We are unaware of literature investigating the belowground effects of drought on *Heliconia*, but this is consistent with results in Bruna et al. 2002 (see response to 2.18).

## 2.18

Lines 451-453: I think herbivory is this is very different from drought. Drought reduced nutrient availability while herbivory removes some plant biomass triggering compensatory growth.

We agree with the reviewer and have replaced this sentence with a more in-depth explanation of Bruna et al. 2002 (L480).

## 2.19

Lines 468 ff: I suggest deleting this paragraph. It doesn’t really matter when exactly a plant dies during a growing season, all that matters for population dynamics is how many plants die from one growing season to the next. Distributed Lag Non-Linear Models account for the exact timing of the weather events.

We agree with the reviewer that the exact timing of mortality within a season doesn’t matter for population dynamics. This is likely obvious to demographers who are used to working with yearly census data and finer scale climate data. However, we think this paragraph is appropriate to follow up with the three before it—a delayed effect indicated by a DLNM plot like those in Figures 4, 5 and 6 could be due to 1) developmental delays, 2) buffering by below-ground storage tissues, 3) indirect effects, or 4) **none of the above.** We want to be sure to point out that in the case of survival especially, a “delayed effect” of SPEI 5 months in the past could simply mean that a plant died *during* a drought 5 months in the past (no actual delay between adverse conditions and death). Importantly, this does not apply to flowering in our system, since flowering happens at a specific time of year (during the census), unlike mortality. We have revised this paragraph (starting on L498) to emphasize the above points.

## 2.20

The discussion includes reasons for why we expect delayed effect in plants in general, but it does not relate the specifics of the significant delays to the life history of H. acuminata. For instance, why does too much rain reduce survival? Why is there an effect of SPEI during the dry season (18-20 months in the past but during the wet season 32-36 months in the past? Why do plants in fragments have a decrease survival when precipitation is high during dry and wet periods (1-18 months time lag)? Etc.

While we believe we did discuss *Heliconia*-specific mechanisms for delayed responses of flowering (L 434–438, now L465–469) and delayed effects on growth (L 445–450, now L475–480), the reviewer is correct that we had not previously discussed mechanisms behind mortality or why wet conditions might be detrimental. We have added a paragraph (L542) to the discussion regarding effects of wet conditions on tropical plants.

## 2.21

Lines 484-486: I don’t think this study strongly support that plants do worse in fragments than CF.

Although the differences in plant size and flowering are relatively small, they are statistically significant (see Table 1 in revised manuscript), and we believe they are demographically significant over long time periods. See our response to 2.7.

## 2.22

Line 499: This is confusing. If you measure size every year than you calculate growth as the change in size from one year to the next, don’t you?

That’s correct. This sentence meant to point out that size is usually measured in yearly censuses while weather data can be more easily collected at monthly or daily timescales. We’ve edited this sentence to read: “*The ability to identify size-specific lagged responses may be especially complicated given size is rarely measured at the same time scale (e.g. monthly) as climate drivers.*” (L533).

# 3 Reviewer 3

**General comments**

This is a really nice addition to the growing literature on lagged climate effects on plant demography, and also expands the tool kit for these studies. The combination of experimentally generated fragmentation with long term demographic data is also unique and very valuable. The paper is very clear and concise, with a few minor exceptions discussed in detail below. The one surprising omission was that the authors didn’t use an Integral Projection Model to put the vital rates together into a metric of population growth potential. I enjoyed reading this.

We thought it would be too much to describe and interpret DLNMs for the vital rates, model recruitment, *and* integrate all those models into an integral projection model. We though the description and interpretation of the DLNMs for survival, size, and flowering were complex and interesting enough to stand on their own. We do intend to follow up with an IPM in a future publication. See also our response to 1.6.

## 3.1

Line 152 – “Plants not found for 3 consecutive seasons …” when were they considered to have died? first year they were missed or last? How many plants were not found in an average survey, and why? Does Heliconia show whole plant dormancy?

We have edited to clarify: *“After the initial census year, between 80% and 97% of marked plants were found. Of plants that had missing values for some years, but were found again in a subsequent year, 95% had 2 or fewer years of missing values. Therefore, plants that were not found for three consecutive surveys, and no subsequent survey, were considered to have died in the transition year after their last observation.”* (L159–164).

Some of the missing values in the data do represent *Heliconia* plants that lost all above-ground tissue but re-sprouted in successive years, but other missing values may simply be imperfect “recapture.”

## 3.2

Line 166 - given the shallow roots, why not use 1 month SPEI and let the model do the smoothing over longer time frames? You describe why not using longer smooths, but not shorter.

The choice of scale for the SPEI calculation is indeed difficult, and is related to the operational definition of drought (i.e. is drought defined as dry conditions over the previous 1, 2, 3 or more months?). We are not certain that the reviewer’s suggestion to allow the GAM to “do the smoothing” would have the same effect as changing the scale parameter when calculating SPEI. SPEI is calculated using a sliding window, while GAMs use penalization to achieve smoothing. We believe a thorough comparison of the ability for different SPEI scales to explain variation in survival, size, and flowering is beyond the scope of this paper. Additionally, the paper originally describing SPI recommends 3-month scale for droughts defined as depletion of soil moisture (McKee et al. 1993).

## 3.3

Line 191 - 194 (equation and description) for clarity you should describe what the index variables i and j are indexing, and I think clearer to have y\_ij for ith plant in jth plot. Indexes on the random effect description suggest there is a separate variance estimated for each plot? d\_it suggests that each plant has a separate SPEI estimate? I think it is probably aggregated at a higher level than individual?

We thank the reviewer for the very helpful comments on our model equation. We have changed to and added an explanation of the indexing as suggested. We have also corrected the error in the random effect equation so that only one variance is included. is intended to indicate that each plant observation is associated with a measure of SPEI for the past 36 months. The SPEI history varies among plants spatially (there are three grid cells in the weather data within BDFFP, so different plots experience different climate) and temporally (observations from all years are pooled).

I would reorganize the following paragraph so that all the descriptive text for the components of (1) occurs before the digression on the details of the crossbasis function.

We moved the explanation of covariates and random effects earlier as suggested.

## 3.4

Line 205 - I’m not sure “bimodal” is correct here – a cubic function can have 2 local extrema, is that what you mean?

This is what we meant, and is no longer relevant as we have increased the number of knots. See our response to 3.5.

## 3.5

Line 205-207 – “number of knots not important …” yes, but package mgcv includes tools for checking those assumptions, would be helpful to include residual analysis and the results of gam.check() in an appendix.

This comment was especially helpful and made us re-think limiting the number of “knots” to 3 in the SPEI dimension. The term *k* is probably more accurately described as the “maximum basis complexity” or “maximum basis dimension.” We had thought that by limiting k to 3, we were simply limiting the complexity of the smooth in the SPEI dimension. However, after taking the reviewer’s suggestions we realized that there were other unexpected consequences. We increased *k* for the SPEI dimension for all models and for other smooths for some models. This resulted in some changes in the the shape of the crossbasis smooth and which areas of it were statistically significant. We have updated the methods and results sections accordingly.  
  
We have also now included a supplemental document with model validation, including checks that *k* is adequately large .

## 3.6

Line 211 - random effect – how did you deal with this in the predictions? set to zero in design matrix, Choose a particular value, sample from the posterior? The code would tell me but not available.

What we show in the figures are not predictions (i.e. results of predict() in R), but partial effects plots generated with the gratia package (equivalent to mgcv::plot.gam()). For figures 4, 5, and 6 we have added the model intercepts to evaluated smooths and back-transformed (when applicable) the estimates. These plots, therefore, do not include the additive effects of other smooths, parametric covariates, or random effects. We think that this is an intuitive way to display the results. We added a paragraph to the methods (L260) explaining our strategy for visualizing the GAM results.

We apologize for not including a link to the code for the project. All analysis code can be found here: [<https://github.com/BrunaLab/HeliconiaDemography>](https://github.com/BrunaLab/HeliconiaDemography). It is so rare for a reviewer to make suggestions based on analysis code, and we are very grateful for these comments!

## 3.7

Line 253 - fig 1a shows mean change in size, not mean size?

We clarified the reference to the figure: “…was preceded by a drop in average plant size in the 2002-2003 transition year (i.e. negative growth in Figure 1a).” Also see our response to 1.7.

## 3.8

Figure 2 isn’t referenced in this section, not used until discussion. Maybe move to supplement or add a sentence in results?

See our response to 1.9.

## 3.9

Line 260 – you mention the CI of the model estimated coefficients, but these aren’t provided anywhere that I could find. I think a table based on the summary of the two fitted models discussed here (and elsewhere too) should be added. And report the variance of the random effect.

We have now included these in Table 1. See also our response to 1.3.

## 3.10

Figure 4 and 5 – very nice way to display the confidence limit. I did have to spend a while looking at them and doing a bit of head scratching to figure out which side the significant values were on, especially for the ones that are along the middle where the difference in color is not large. I wonder if a diverging color scheme with values close to the mean set to white would make this easier to see? Or if you could add some hatching or change the alpha transparency on one side to highlight those regions more clearly.

In the updated version of figure 4 (see response to 3.5), a smaller area of the plot is significantly different from the intercept, and heatmap area far from observed data is grayed out. We believe these two things make it easier to see the outlines around significant values, but we are open to trying other modifications of the plot if the reviewers think it is still difficult to interpret in this way.

## 3.11

Line 413 – “our expansion of this approach …” I’m not sure that the expansion is the penalty that prevents over fitting – both Teller et al and Tenhumberg et al used the penalized spline function in mgcv. What IS an expansion is the non-linear cross basis function. Both Teller et al and Tenhumberg et al used (essentially) linear cross basis functions, so using dlnm is super-cool and interesting.

We thank the reviewer for the positive comment. This is correct—our expansion was not the fitting of lagged effects with a spline, but the use of a 2 dimensional spline to *also* fit non-linear reaction norms. We have edited this sentence accordingly (L436–438).

## 3.12

Line 471 – this demographic delay left me a bit confused for a while until I stopped looking at figure 4b and started looking at 4a. Then it made a lot of sense. However, wouldn’t it be that the demographic effect takes place AFTER the census for an echo to appear at previous lags?

We We have revised this paragraph to clarify (L498). See our response to 2.19.