Editor and reviewer comments (we provide below the full context of each review) are in *italics*, while our responses are in regular text.

Editor's comments:

As you can see from the comments of the expert reviewers below, both are broadly positive about the manuscript, but both have suggestions for further improvement. Referee 1, in particular, thinks that the submission would benefit from some fairly substantial rewriting in places, and I would tend to agree.

We thank the editor for the opportunity to revise our manuscript. We found the reviewers' comments very helpful, with many overlapping requests for clarity or a more careful message. We provide detailed point-by-point responses below.

Reviewer 1 comments:

The authors suggest that tracking individual species' phenological response to environmental change may be insufficient to meaningfully predict community response to future climate change due to the influence of species' interdependence beyond trophic interactions and other influences at the community level. In brief, whether or not a particular species will survive climate change is complex and depends on a range of external factors that affect fitness including the influence of other species in a dynamic system that cannot be predicted by simply tracking its phenology. The impact of climate change on species' interactions with the physical environment and other organisms is notoriously complex and difficult to disentangle. Overall, this is a very timely and important topic and I would support its publication after tightening up of the language and presenting a sharper, clearer message. As it stands this manuscript appears wordy and sometimes laboured in places which tends to mask the key message being conveyed. It would be useful to highlight how this manuscript contributes to advancing the field of predicting species' response to climate change and how this in turn may impact community assemblages.

Thanks for this very helpful review. We completely agree that we could make the message clearer and make the writing easier on readers. We made have now made changes from the abstract to the conclusions. In particular we have tried to reduce the wordiness, including the asides, caveats and examples. We have a lot of these! Partly they come from the struggle of trying to unite the empirical and theoretical framework: the theorist on the author team aims for precision, which led to some of these additions, while the empiricist wanted to clarify what we meant for other empiricists, which led to more of these additions. We have worked to remove many of them, leaving the ones we think are most important for comprehension (but do let us know if you see more that could be removed).

Abstract
Consider replacing 'of' with 'for' in the first sentence. Consider removing 'us' in second last sentence.

We changed first sentence a little (but stuck with 'of' as we found 'for' read odd given other changes), and we removed 'us.' Also, we re-wrote the abstract a good bit for clarity and to better state the aim of our review.

Briefly define 'stationary and non-stationary systems'. Consider conveying in this sentence that the proposed method will 'help' predict rather than definitively 'predict'. It would be useful to emphasize the key findings of the review.

We have changed to 'help understand.' Given our definition of stationary and non-stationary we were not able to briefly define these terms in the abstract, however, we have added a glossary to the paper as the first table.

Introduction
In the second sentence, consider mentioning phenology before space to be consistent with 'time and space' at the end of the first sentence.

Good catch! We switched the ordering of 'space and time' in this first sentence, and altered the second sentence to make the connection to 'time' in the first sentence more clear.

P4L12 Consider adding 'and locations' after '..across species'.

Done.

P4L38 It would be useful to clarify what is meant by '...the underlying distribution of the environment.....' is this related physical climatic parameters?

Good point, we have done this in the new glossary that we added.

P4L53/54 Consider replacing 'measuring tracking in current environments and evaluating the fitness outcomes of tracking' with 'measuring tracking and evaluating its fitness outcomes, in current environments'.

Done. We also just removed 'in current environments' to make the sentence less wordy.

In Figure 1 it would be useful to know how many years before and after 1980 were included in the analysis.

We hinted at this at this at the end of the caption, but can see now it was not very clear. We have 40 years before 1980 and 40 years after 1980 for all the datasets, which we now say early in the caption.

What is the specific aim of this review the authors mention what they are going to review but why is it necessary and what is this synthesis expected to reveal?

Great point, we have added this here (paragraph starting on line 36) and in the abstract. We also adjusted our conclusions some to address this.

Defining and measuring tracking
Consider sticking to vegetation or animal examples not a combination of both or explain the rationale for using both groups.

We have restructured this section slightly and now address this question in line 53-line 56.

L42 consider replacing 'these' with 'the timing and/or intensity'.

We have removed this sentence (we believe, we're not sure of page number, let us know if not).

Overall, I found section 2 very wordy and difficult to follow. P5L47 Consider the following sentence 'Tracking is commonly used to describe how phenology responds to climate change, yet it is rarely defined'. This section could be clearer and 'tighter'.

Agreed! We made lots of changes here, including the one suggested.

P6L25/26 consider replacing '....to be something that can be accurately modeled..' with 'to permit/allow accurate modeling'.

Done.

P6L33/34 remove 'a' before 'interaction'.

We believe this text has been removed in the revision process.

Does environmental tracking include parameters other than climatic variables and photoperiod, such as, nutrient availability etc.? It might be useful to state this.

Environmental tracking is a product of an organism's cue system, and we have worked to clarify this in the text, especially in the glossary. To date major cue systems have found evidence of photoperiod, temperature, moisture cues mainly. Based on this we expect that environmental tracking is not directly related to nutrients in many systems. However, it may be correlated with tracking, as the cue system is tuned to maximizing nutrient availability (but it's not responding directly to it).

Another possible reason that species may not appear to 'track' maybe that some species require a greater amount of change in the environmental cue i.e. a higher threshold, before they respond. It seems likely that there are more than 3 major reasons why tracking is not detectable. A clear rationale for selecting and focusing on these reasons would be useful.

Good point, we actually see that reason as covered by our first reason, but we agree their could be others and have adjusted our text here somewhat.

P7L40/41 consider placing 'briefly' before 'review'.

Done.

P9L18/19 replace 'that' with 'the'.

Done.

P10L54/55 repetition of 'fluctuations in the environment' consider something like 'mechanisms which are dependent on, or independent of, fluctuations in the environment.....

Done.

P11L9/10 Consider 'In community ecological modeling, definitions of the environment generally fall into two broad categories'.

Done.

P13L5/6 consider removing 'including the previous example'. L16-18 is it necessary to be so explicit? Throughout the MS I find much of the information presented within brackets distracting.

Removed and we agree! We have tried to remove these caveats and 'let's be super explicit' additions throughout. We were trying to be exact and thorough, but it makes it hard to read and isn't always needed.

P13L21-25 if the resource is in limited supply this may provide an advantage to early arrivals but later arrivals will benefit from the same resource if it is not depleted by earlier arrivals and other conditions may be better such as less risk of late frost. It is well established that the environmental variables being monitored are not exactly what the organism is responding to as micro-climate varies considerably in even the simplest ecosystem, furthermore, do we even know if budburst for example is triggered to the temperature of the bud surface, some internal temperature or some interaction with root (or some other organ) temperature. Defining the specific environmental variable and the specific threshold each species and each phenophase is responding to appears overly complex there must be some trade-off between a researchers effort and the applicability of the results. I just wonder where we draw (or don't) the line. This is a very challenging and important topic that the authors are addressing.

Thanks? We agree ... [ADD more here].

Conclusions are more of a summary of the topics reviewed rather than a comprehensive synthesis of the literature to draw new and more advanced conclusions based on the collective information from the review. It might be useful to make some recommendations on what is needed and why. Therefore, perhaps point 5 could be expanded.

We have expanded point 5 (now point 6) and adjusted our conclusions to more clearly state the path forward. Our sections 4.5 ('Frontiers of community assembly models') and section 5 ('Linking empirical and theoretical research') are focused on recommending what is needed, so we worked to summarize these better in the conclusions, while also following journal guidelines (which state that the "conclusions section should be in the form of a short list of numbered points summarising the main findings of your article. Avoid the introduction of new material or new references in the conclusions section.").

Reviewer 2 comments:

In this review the authors argue that phenological tracking data needs to be combined with co-existence theory to help make predictions about how climate change will affect communities,

particularly in non-stationary environments. Generally, this was well-written, clear in its logic, and really laid out how phenological tracking vs. coexistence theory are currently divided but could be integrated. The one aspect that fell short was suggesting specific types of experimental data that might help us reach this integration of phenological cues and climate change in a multi-species framework. I also have a couple of clarifying questions throughout. But otherwise, this was a really interesting review proposing a novel combination of two research areas that would really strengthen predictions about the consequences of climate change.

1. The argument for phenological tracking combined with coexistence theory in stationary environments itself is something that needs more work, but it was kind of glossed over in favor of non-stationary environments. It might be worth laying out more about the gaps between tracking and coexistence in stationary environments in the intro (pg.4, between lines 33 and 34?).

We tried to address this as best we could without impacting the flow and coherence of the introduction (see line 16-line 22). As we are focused here on climate change and thus non-stationarity our introduction is focused on these interwoven topics.

2. Pg.4, L49: Can you be more specific about examples of the community-level processes here? At first I thought this meant phenological tracking as a process but that would be individual or population level.

Yes! We have added 'such as competition and priority effects' (see line line 35).

3. Pg.5, L31: Isn't number of offspring still measured at the individual level? What does "higher levels" mean here? Perhaps an example like synchrony at the community level would fit better?

Good point, switched to population.

4. Pg.5, L35: I agree that the first event is not equivalent to the number of flowers (or a continuous phenological metric), but yes/no flowering is not the same as first event either these variables seem conflated in this section.

We have removed the sentence about first events.

5. Pg.6, L34-36: This section made me wonder if there are any examples where researchers know the exact cues and can measure phenological tracking more precisely in systems like Arabidopsis?

Great point! We have added mention of the photothermal model of flowering of *Arabidopsisis thaliana* (line 105) and we also discuss it in a new section 'Building from cue systems to phenological tracking' in our future directions (section 5, 'Linking empirical and theoretical research), see line 511-line 520.

6. Pg.7, L49: This is somewhere where I thought empirical directions could be suggested (or using this, have a section on future research at the end) the heritability of phenological tracking made me think maybe we need more studies linking phenological plasticity to genetic variation and heritability (e.g., do more plastic populations harbor more genetic variation in phenological traits)?

Good point, we added this to line 511-line 520 in a new section 'Building from cue systems to phenological tracking' in our future directions (section 5, 'Linking empirical and theoretical research).

7. Pg.8, L23: This is the other spot where I wanted more background/discussion of gaps in the links between phenological tracking and stationary environments before moving on to non-stationary environments.

Good point, we have expanded on this, see line 171-line 179.

8. Pg.8, L24: Is "unpredictable" environments here different from non-stationary environments? In pg.9, L3, it sounds like non-stationary also means unreliable in this context, so I was a bit unclear on why unpredictability was introduced in 3.1 instead of 3.2 (I might have missed some key difference). A glossary box for some of these terms could also be helpful.

Good point, this was rather inexact, we removed this word and are more specific now (see line 180).

9. Pg.9, L18: What is the fundamental "model" here? Is this underlying fitness?

Good point, we meant the model of fundamental tracking which we now say explicitly (see line 214).

10. Pg. 9, L33: The idea that phenological tracking might vary not only in stationary vs. non-stationary environments but also in environments transitioning from stationary/predictable to non-stationary/non-predictable is cool (and brought up again later in the review), but it's introduced so late in this section and without any context! Definitely worth expanding on.

Good point, we introduced it late as it is setup for our next section, but it does seem a big point that we're making rather too easy to miss. We have expanded to a full paragraph (see line 225-line 231).

11. Pg.9, L47: This statement felt short perhaps include some reasons why competition is critical. Additionally, this is another spot with some future directions implications because phenological shifts aren't often studied in the context of their effects on species interactions (although we often make assumptions based on biology, e.g. invasive-native comparisons). Perhaps a call to include phenological shifts across environments as part of per-capita competition experiments?

Good point, we changed to 'Yet decades of research show that competition drives the niche differences necessary for species to co-exist (Hutchinson, 1959; Chesson, 2000)' (line 236). Also, added this to our future directions (section 5, 'Linking empirical and theoretical research), see line 537-line 547.

12. Pg.10, first paragraph: I found this argument really interesting because although it made sense, it read as counterintuitive to invasive-native comparisons of phenology. The argument here is that more plastic species are probably inferior resource competitors. But studies that find that invasive species are more plastic also assume invasive species are the superior resource competitors. I wonder if this section needs some caveat in our assumptions for invasive species

in these models? (Similarly in 4.2, but I think that the argument that "best-matched" species could drive the other extinct (Pg.12, L45) starts to help with this.)

This is a super interesting point, which we think depends on how you define superior and inferior resource competition. The superior competitor is often defined as the species that can draw down the resource to a lower level, and in the systems we have worked in this is often the native species (for example, invasive annual grasses versus native shrubs in coastal sage scrub in California or invasive shrub glossy buckthorn versus native trees in the US midwest). However, by being earlier the invasive species may outcompete the native through early access to resources, but it would not generally be defined as the superior competitor. Testing this, however, requires measuring more traits across native and invasive species (especially the difficult ones related to resource competition), which we now state in our conclusions (line 593).

13. Pg.12, L55: What if phenological tracking itself could be a stabilizing mechanism by affecting resource partitioning?

This is a great question. Based on our current understanding of community assembly, tracking could only be stabilizing if it trades off with another trait or in some other way creates a niche difference. If tracking somehow partitions resources this could work, but it would have to partition across multiple resources or in some other way limit competition for the same resource across multiple species.

14. Pg.13, L5-12: The review draws on plasticity literature, but there are definitely studies linking plasticity (including phenological plasticity) to fitness that seem overlooked in this first paragraph. E.g., Lustenhouwer, N., et al. (2018). Global Change Biology 24(2): e534-e544.

Agreed, we now address this in line 511-line 520 in a new section 'Building from cue systems to phenological tracking.' This point in general is a very good one and something we have definitely struggled with in writing this manuscript. While we aim to focus on multi-species model, the literature on plasticity is critical to phenology. We have tried to include the major background from this very large literature as much as possible without greatly extending the length and focus of the review.

15. Pg.13, L35: I'm unclear on whether phenology is explicitly in these models. Otherwise, how is phenology connected to studies of production that don't measure phenology?

We did not make this clear at all and have revised the text (line 386 - hr2thisrelatesend) to clarify.

16. Pg.13, L38: Figure 6 is your model, it doesn't show whether other studies model environmental distributions. Why referenced here?

Good point! We have now switched to referencing several major papers on this (line 392).

17. Pg.14, L25: Could you expand on why the new communities would exist only as long as the environment remained non-stationary? Could plastic species that arrived in a non- stationary environment also stay if the environment was just consistently warmer, for example (i.e., stationary)?

18. Pg. 14, L48: I love this call to action great place to put the types of data we still need.

Thanks.

19. The figures are all great I found 2 and 5 particularly helpful for some of the complicated concepts in here. Figure 5a also made me think about whether we need more studies changing the timing of resource (or stress) pulses and how that affects tracking.

It's great to hear these figures are useful to others–as they have been useful to us. Studies how changing the timing of resource pulses affects tracking would be very interesting, especially if they could be done manipulatively. We hope work inspired by our paper might provide the theory to best design such experiments.

References

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–366.

Hutchinson, G.E. (1959). Homage to Santa-Rosalia or why are there so many kinds of animals. *American Naturalist*, 93, 145–159.