Version dated: March 17, 2020

Running title: Environmental tracking

How environmental tracking shapes species and communities in stationary and non-stationary systems

E. M. Wolkovich[1] & M. J. Donahue[2]

[1] *Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4 (e.wolkovich@ubc.ca)*

[2] *Hawaii Institute of Marine Biology, University of Hawaiʻi at Mānoa, Kānʻeohe, HI 96744 (donahuem@hawaii.edu)*

Corresponding author: see[1] above; Ph: 604.827.5246 (no fax).

Authorship statement: EMW and MJD both conceived of the paper, performed modeling work and edited the paper, EMW additionally wrote the paper and did the literature review, while MJD additionally wrote the supplementary information on the model.
Data statement: Review, so no new primary data, but data from a comprehensive literature review will be archived in an appropriate public repository and the data DOI will be included at the end of the article.
Keywords: community assembly, global change, climate change, phenology, environmental variability
Article type: Reviews and Syntheses
Article information: Abstract: XXX words; Main text: 6,490; Figures: 5; Boxes: 4 (text in Box 1: 348; Box 2: 500; Box 3: 264, Box 4: 776); 113 references

Abstract

Climate change is reshaping the environments of all species. Predicting responses requires understanding the costs, benefits and constraints of how well species can track environmental change, and how such tracking shapes communities. Growing empirical evidence suggests that environmental tracking—how an organism shifts the timing of key life history events in response to the environment—is linked to species performance and is a structuring force of species and communities today. Here, we review current knowledge on tracking both in empirical data and through the lens of ecological theory. We provide a definition of environmental tracking that highlights both why it must be fundamentally related to fitness, and the challenges of defining it empirically. We then show how life history theory makes predictions for variation in tracking, and trade-offs with other traits, across species and environments. Finally, we examine how well basic community ecology theory can be extended to test the current paradigm that climate change should favor species with environmental tracking. We to aim provide a framework based on existing ecological theory to understand how tracking in stationary and non-stationary systems may shape species and communities and, thus, help predict the species- and community-level consequences of climate change.

1 Main text

Anthropogenic climate change is causing widespread changes in species distributions, with many species shifting in both time and space (IPCC, 2014). Reports often focus on species shifting to higher elevations and poleward (Chen *et al.*, 2011) and/or shifting their recurring life history events (phenology) earlier as climate warms (Wolkovich *et al.*, 2012; Cohen *et al.*, 2018). These general trends, however, hide high variability across species. A large proportion of species are not shifting at all (Cook *et al.*, 2012), which has raised concerns about whether these species may be more vulnerable to population declines with continued warming. Such concerns come in part from increasing research that links how well species track climate change—especially through temporal shifts—to shifts in biomass, growth and other metrics related to performance (Cleland *et al.*, 2012). Tracking climate change may then be a major component to understanding and predicting the indirect effects of climate change, including population declines, with cascading effects on community and ecosystem structure.

The hypothesis that tracking predicts fitness outcomes with climate change has gained significant traction in the ecological literature focused on global change (e.g., Cleland *et al.*, 2012) and several areas of ecological theory support it. Considering tracking as a form of plasticity, evolutionary models predict species that track will be favored in novel environmental conditions. Similarly, some models of community assembly suggest that a warming climate should open up new temporal niche space and favor species that can exploit that space (Gotelli & Graves, 1996; Wolkovich & Cleland, 2011; Zettlemoyer *et al.*, 2019). Yet not all studies find the purported link (e.g., Block *et al.*, 2019), and there has been comparatively little work to improve predictions by formally connecting tracking to foundational ecological theory.

This disconnect could be because most ecological theory today is for stationary systems (e.g., Sale, 1977; Chesson & Huntly, 1997). While major arenas of research such as 'modern coexistence theory' or population ecology are now built on assumptions of a stochastic environment, they generally still assume stationarity, where the underlying distribution of the environment is unchanged across time (i.e., constant mean and variance, Barabas *et al.*, 2018). This assumption is common to much of the theory that underlies ecology, evolution, and myriad other research fields (e.g., Milly *et al.*, 2008; Nosenko *et al.*, 2013).

Climate change upends the assumption of stationarity. By causing increases in temperature, larger pulses of precipitation, increased drought, and more storms (Stocker *et al.*, 2013), climate change has fundamentally shifted major attributes of the environment from stationary to non-stationary regimes (see Fig. 2 and Box: Environmental variability & change). This transition is reshaping ecological systems, and, while new work has aimed to adapt coexistence theory to non-stationary environments (Chesson, 2017; Rudolf, 2019), there is still very little theoretical work on what such a transition may mean for communities and the species within them, despite growing empirical studies.

Here, we review current knowledge on tracking both in empirical data from the climate change impacts literature and through its related ecological theory. We provide a definition of environmental tracking that highlights why it must be fundamentally related to fitness and the complexity of defining it in empirical systems. We show how life history theory—specifically drawing on optimal control, bet-hedging and plasticity—make predictions for variation in tracking across species and environments in stationary and non-stationary systems. We then examine how well basic community ecology theory can be extended to test the current paradigm that climate change should favor species with environmental tracking.

1.1 Defining environmental tracking

While tracking is a commonly used word in the phenology and climate change literature (e.g., Menzel *et al.*, 2006; Cleland *et al.*, 2012; Deacy *et al.*, 2018), there are few, if any, definitions of it. Most interpretations of tracking relate to how well an organism matches the timing of a life history event to the ideal timing for that event, what we refer to as 'fundamental tracking'. Fundamental tracking thus rests on an assumption that there is a timing (an 'ideal timing') that yields maximum fitness, and event timings moving away from this ideal result in reduced fitness (a foundational concept of the trophic mismatch literature, Visser & Gienapp, 2019). Yet this ideal timing is generally only clear in simplified models or in retrospect, thus species must use environmental cues to attempt to predict and match their event timing to the ideal timing across environments in both space and time. Each organism's set of cues forms the biological basis for how a species tracks, but measuring environmental tracking requires two more components.

The first component is the environment's variability—which aspects of the environment vary, how (e.g., temporally each year, spatially at x scale) and how much. If the varying components of the environment do not correspond closely to the organism's cues, then the species may not track this variability. Further, under this definition, identical genotypes will have different tracking across environments, depending on the interaction of the cues and environmental variability.

Second, which aspect(s) of the environment researchers measure will determine measured environmental tracking. If researchers know the exact cue (e.g., a thermal threshold) or suite of cues (e.g., a interaction of thermal sums and daylength) and can perfectly measure these in an environment where the cue(s) varies then an organism will track the environment perfectly. If researchers measure some related attribute (e.g., mean spring temperature in place of thermal sums) or only some of the organism's cues, then the organism will appear to track poorly (i.e., a noisier statistical relationship). If researchers measure an environmental variable that is not directly related to the cue(s) that the species actually uses, but one correlated with it (e.g., an insect tracks daylength but researchers measure temperature) then they have not measured tracking per our definition.

Accurately measuring environmental tracking thus requires a complete knowledge of an organism's cue(s), the environment's variability and the relationship between the actual cues and measured environmental metrics. Knowing an organism's cues is inherently difficult, generally requiring a suite of experiments, process-based models and in-situ data to show that the model of cues is accurate. Not surprisingly then we lack this for almost all species, coming closest for some model species (e.g., *Arabidopsis thaliana*, Kingsolver, 2007; Wilczek *et al.*, 2009), or species with very simple cues (e.g., coral *Acropora millepora*, Levy *et al.*, 2007) and have some basic information for some other species (e.g., the Great Tit, *Parus major*, Charmantier *et al.*, 2008).

1.2 Measuring environmental tracking

Attempting to measure environmental tracking and compare variation in it across species, space and time is a rapidly growing area of ecological research (e.g., Cook *et al.*, 2012; Fu *et al.*, 2015; Thackeray *et al.*, 2016; Cohen *et al.*, 2018). Multiple meta-analyses now show plants' spring phenology shifts with spring or annual temperatures 4-6 days/°C on average across species (Richardson *et al.*, 2006; Wolkovich *et al.*, 2012; Thackeray *et al.*, 2016), but also highlight high variation across species (Cook *et al.*, 2012), even after examining multiple major climate variables (Thackeray *et al.*, 2016). Variability across species appears similar when examining consumers tracking their prey (across diverse species tracking over time is 6.1 days/decade but ranges from zero to 15 days/decade, see Kharouba *et al.*, 2018).

All species-rich studies of phenology-climate relationships find high variation, including some species that do not track or track poorly (i.e., high noise surrounding observed statistical relationships). Researchers have worked to link such variation at times to the underlying cues (e.g., Cook *et al.*, 2012), species traits (e.g., Cohen *et al.*, 2018) or trophic level (e.g., Thackeray *et al.*, 2016). These approaches hint at the three majors classes of reasons that underlie species that do not appear to track climate (or appear to be poor trackers): (1) species do not track, as perfect environmental tracking may either not be possible or optimal for all species, (2) lack of firm biological understanding of the cues that underlie tracking, and (3) statistical artifacts that make it difficult to measure tracking robustly.

Increasingly, research has outlined statistical difficulties in measuring tracking. These issues mostly relate to the challenges of observing phenological distributions (Steer *et al.*, 2019; Carter *et al.*, 2018) through anthropogenically defined temporal units (e.g., day or month) in non-stationary systems, and to the complexity of climate data (more further discussion, see Box 'Statistical challenges in measuring tracking').

Even without statistical issues, translating phenological and climate data into estimates of tracking requires a firm biological understanding of an organism's cues, critical knowledge that researchers rarely have (Chmura *et al.*, 2019). Currently, 'tracking' is often measured simply as the relationship between the dates of the phenological event and a simple abiotic metric, such as mean monthly temperature (with variation in temperature derived from multiple periods of observation or induced through experiments). Simple environmental metrics, however, are almost always proxies for a more complicated underlying physiology where simple cues—such as warm temperatures—can be modified by other cues, such as photoperiod, drought or light spectra (Bagnall, 1993; Stinchcombe *et al.*, 2004). Indeed, multiple studies have shown how simple correlations between phenological events and environmental variables can mask complicated relationships (Cook *et al.*, 2012; Tansey *et al.*, 2017).

Modeling multivariate cues well is inherently difficult (Chuine *et al.*, 2016), especially since one cue may dominate in many conditions. For example, woody plant leafout responds strongly to warm spring temperatures, but also to cool winter temperatures to prevent leafout in midwinter warm snaps that occur long before the last frost. Often this cool-temperature effect may be masked by sufficiently cold conditions. With warming from climate change, however, this additional trigger—which appears to vary by site, species and even inter-annual conditions (Dennis, 2003)—may become critical (and potentially lead many phenological models to fail spectacularly in the future as additional cues come into play, see Chuine *et al.*, 2016). In some semi-arid systems, species time growth to pulses of rain, but only when those rain events occur with cooler temperatures that indicate the start of the rainy season, and not a rare summer rainfall event in the middle of months of drought (Wainwright *et al.*, 2012; Wainwright & Cleland, 2013). Tracking in species with longer generation times may be especially complicated, as species may track low frequency climate signals and make investment choices on far longer timescales than species with shorter lifespans (Morris *et al.*, 2008).

Researchers are increasingly recognizing the need to consider multiple climate variables, though currently most estimates are based on long-term observational data (e.g., Chmielewski *et al.*, 2013; Simmonds *et al.*, 2019), which can lead to spurious correlations without experiments to test hypothesized cues (Chuine & Regniere, 2017). Further, estimates of 'tracking' from long-term data that are not linked to mechanistic experiments may sometimes serve as proxies (i.e., environmental variables correlated with one or more actual cue that a species uses) for an organism's environmental tracking, but may not directly connect to an organism's cue(s).

Limited understanding of organisms' phenological cues combined with statistical issues may make many current estimates of variation in tracking less reliable than they appear, and make robust quantitative analyses across species difficult (Brown *et al.*, 2016; Kharouba *et al.*, 2018). Yet these estimates provide the crucial first step to understand variation. As estimates improve, ecologists will better capture a picture of which species, when and where, do and do not track. Yet given the current difficulty of measuring environmental tracking, clear testable predictions from ecological theory are perhaps most critical to guide research today (Smaldino & McElreath, 2016).

1.3 Understanding variation in environmental tracking

A number of research areas in ecology predict variation across species in how well they track the environment. Applying these areas of research to environmental tracking, however, first requires understanding phenological events. In particular, while phenological events are often coded as on/off switches (e.g., a seed does/does not germinate; a coral does/does not spawn), they are almost always defined by investment decisions that are part of a continuous developmental process (Inouye *et al.*, 2019).

Phenological events are best considered as the outcome of a two-part sequential process that is repeatedly observed over time. At each temporal unit, an event can either happen or not (step 1) and, if it happens, there is a secondary part regarding the size or degree of investment of the event (step 2). This process is generally applied at the level of the individual (but it could potentially apply at lower levels, for example buds on a branch, or potentially higher levels, such as all the offspring from a parent). Across time, it produces an event's distribution. After starting, many events are entrained to continue based on the underlying physiological process:

for example, laying eggs within one clutch (here, the first part of the process is whether to lay eggs or not and the second is whether to continue to invest in that process, which would lead to additional eggs, which researchers then observe as number of eggs per temporal unit) or flowering each growing season. In such cases, first events at the individual-level are somewhat unique from the rest of the event's distribution. In all cases, these individual-distributions scale up to the population-level estimates of these events generally used by researchers (see Inouye *et al.*, 2019, for discussion of the outcomes of this scaling).

Considering the life history events that define part of environmental tracking as a two-part process highlights that tracking is ultimately shaped by resources that species need to grow and reproduce, and circles back to an organism's fundamental tracking. This is perhaps best recognized in the literature on trophic synchrony where there is often focus on how well consumers' environmental tracking matches to the seasonal distributions of their prey (Deacy *et al.*, 2018; Kharouba *et al.*, 2018). For example, decades of work has studied how birds (e.g., *Parus major*) time their peak food demands—during their nesting season—to maximum prey (caterpillar) abundance (e.g., Charmantier *et al.*, 2008). Failure of environmental tracking to match prey year-to-year or over time with long-term warming has been well tied to individual-level fitness consequences in some systems (Charmantier *et al.*, 2008), but not all (Visser *et al.*, 2006), which may be due to the complexity of mechanisms that influence total fitness (Singer & Parmesan, 2010; Johansson & Jonzen, 2012). Environmental tracking in plants and other lower trophic levels is also about resources. Alpine plant species that emerge in step with snowmelt or temperature are likely responding, at least in part, to light resources for photosynthesis. Light equally appears critical to the sequence of phenology in many temperate forests: with lower-canopy species, and younger (shorter) individuals of higher-canopy species, routinely risking frost damage to leaf out before the canopy closes and access to light becomes severely reduced (Vitasse, 2013; Heberling *et al.*, 2019). These ultimate controllers on tracking—which determine fundamental tracking—are then filtered through the abiotic environmental cues species use to time events. From here, predicting tracking relates to predicting which cues an organism should use: an optimal control problem.

Predicting variation in environmental tracking in stationary systems

An optimal control framing can help predict which cues an organism should have based on a consideration of the costs, benefits and constraints, in any one organism by environment system (Donahue *et al.*, 2015). First, it requires that benefits vary depending on the timing of event; this effect may be stronger in highly seasonal environments. Next, there must be a useful cue—some aspect of the environment that predicts resources or otherwise links to back to the ultimate factors that shape environmental tracking (Gremer *et al.*, 2016). Some environments may inherently lack useful predictors, such as desert systems where few early-season variables seem to predict high or consistent rainfall years.

From here, the exact cue or suite of cue(s) that an organism should have depends on the cost of those cues (e.g., the machinery of monitoring temperature or daylength), the benefits of a cue (for example, how much tissue is saved by avoiding a coldsnap). Ultimately the balance of the costs of cue(s) and their benefits should determine exactly what cue(s) a species uses: apparently poor cues may occur for organisms in environments where there is both a low cost and low benefit to the cue(s). Similarly, expensive cues, such as complex multivariate ones, are possible given a high pay-off. Most in-depth studies of species' phenological cues find evidence for complex multivariate ones (Chuine & Regniere, 2017). These cues almost always appear adapted to handle unusual—though not completely uncommon—years when the simple cue alone would fail (that is, would trigger growth, reproduction or another life history event at a suboptimal time), suggesting that multivariate cues may provide a large benefit in best coupling the timing of cues to the ultimate controls (i.e., cues that couple environmental tracking strongly to fundamental tracking).

Optimal control highlights that not all species should track, but instead that tracking is based on an optimization of costs, benefits and constraints. In environments where there is no clear optimal strategy, species should bet-hedge (de Casas *et al.*, 2015). In general, species in highly variable environments, or which otherwise face high uncertainty in when to time investment decisions, should gain a substantial benefit from bet-hedging or employing other approaches that spread out risk given uncertainty (Venable, 2007; Donaldson-Matasci *et al.*, 2012). Assessing bet-hedging in many systems, however, requires studies of fitness over longer timescales than many current field experiments.

Constraints also shape cues and may limit tracking. Fundamental differences in life history impose constraints—for example, the type and amount of loss an organism can sustain each season is limited by its generation time and other attributes related to long-lived lifestages that yield buffered population growth (Chesson & Huntly, 1997). Additionally, constraints may arise if a species cannot closely measure relevant environmental cues (Arnold, 1992; Singer & Parmesan, 2010), through unavoidable trade-offs with tracking (Singer & Parmesan, 2010; Johansson & Jonzen, 2012), or through evolutionary pathways. Gene flow from other environments may continually push a population away from its local optimum (Lenormand, 2002), standing genetic variation limits phenotypic variation and thus can slow the evolution of optimal cues (Franks *et al.*, 2007; Ghalambor *et al.*, 2015), deeper evolutionary history may produce co-evolved traits making it difficult for selection to act on a single trait axis (Ackerly, 2009), or other fundamental evolutionary limits to the rates of trait change and what traits are possible (Gould & Lewontin, 1979).

Predicting variation in environmental tracking in non-stationary systems

Much of life-history theory, including optimal control and bet-hedging, rests on assumptions of stationarity, thus a major open area of research is adapting life history theory to non-stationary environments. Multivariate cues may be especially robust to a non-stationary environment if they provide a tight coupling of cues to fundamental tracking, and that coupling is maintained in the non-stationary environment (Dore *et al.*, 2018). But multivariate cues may equally be most vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (Bonamour *et al.*, 2019). For example, consider an organism's whose cues evolved based on a correlation between peak prey abundance and daylength—these cues may work well in a stationary environment but fail if warming advances peak prey abundance. Predicting the outcome of non-stationarity thus relies on knowing both the full cue system of an organism, how it relates to fundamental tracking, and how both that cue system and the underlying fundamental model shift with non-stationarity.

Another area of life-history theory, that focused on plasticity, may be primed to provide insights on non-stationarity (or 'sustained environmental change,' see Chevin *et al.*, 2010). Considering phenology as a trait (as we and others do, e.g., Charmantier *et al.*, 2008; Nicotra *et al.*, 2010; Inouye *et al.*, 2019), environmental tracking is one type of plasticity. Researchers could thus more broadly understand environmental tracking through modeling an organism's reaction norms (Pigliucci, 1998; Chmura *et al.*, 2019) and understanding how cues and suites of cues—across environments—determine how fundamentally plastic an organism may be in its tracking. For example, multivariate cues should yield higher plasticity in this framework. From here, models of the role of plasticity in novel environments provide an important bridge to understanding the outcomes of non-stationarity, generally predicting non-stationarity should favor highly plastic species. This outcome, however, assumes there are no costs related to plasticity (Ghalambor *et al.*, 2007; Tufto, 2015). If there are costs associated with plasticity, akin here directly to costs associated with tracking, then species may evolve lower tracking, because it should trade-off with other traits (Auld *et al.*, 2010).

1.4 Tracking in multi-species environments

Plasticity theory—in contrast to much of the life-history theory discussed above (where other species are, at best, filtered into models as an aspect of the environment)—shows how critical a multi-species perspective is to understanding environmental tracking (Metcalf *et al.*, 2015). In this light, tracking cannot be considered as a singular trait, but must be evaluated as part of a trait syndrome (or mosaic of traits, Ghalambor *et al.*, 2007) and selection in multi-species environments should produce communities of species where tracking trades-off with other traits.

As tracking often relates to the timing of a resource pulse, traits related to resource acquisition are likely contenders for a trade-off. Species with traits that make them poor resource competitors may need to track the environment closely to take advantage of transient periods of available resources, but will risk tissue loss to harsh environmental conditions more prevalent early in the season (e.g., frost or snow). In contrast, species with traits that make them superior resource competitors may perform well even if they track environments less closely, because their resource acquisition is not strongly constrained by competitors. Examples include under-canopy species leafing out earlier to gain access to light (Heberling *et al.*, 2019) or species with shallow roots starting growth sooner in an alpine meadow system, while species with deeper roots begin growth later (Zhu *et al.*, 2016). In such cases, tracking is akin to a competition-colonization trade-off (Amarasekare, 2003), where species that track well gain priority access to resources and, thus, may co-exist with superior competitors. Research to date supports this, with several studies linking higher tracking to traits associated with being poor competitors (Dorji *et al.*, 2013; Lasky *et al.*, 2016; Zhu *et al.*, 2016). Further, many studies have found a correlation between higher tracking and 'earlyness' each season, which has been linked to resource acquisition traits associated with lower competitive abilities (Wolkovich & Cleland, 2014, see Box 'Trait trade-offs with tracking').

Understanding these trade-offs is clearly critical, but understanding the short-term dynamics of a changing environment with plastic species is additionally important. Most theory predicts the outcome of a new environment, but non-stationarity in the climate today means understanding the trajectory to that outcome may be most relevant to interpreting current trends. For example, models show how plasticity may limit standing variation and thus reduce fitness in novel environments (Ghalambor *et al.*, 2007; Fox *et al.*, 2019). Whether such findings extend to systems transitioning from stationary to non-stationary will likely depend on how non-stationarity affects the rate of adaptation (Chevin *et al.*, 2010). Efforts to model expected outcomes given climate projections and current understanding of plasticity and genetic variation underlying event timing in some organisms provide the empirical start to this (e.g., Fournier-Level *et al.*, 2016), but more eco-evolutionary models that bridge this gap may prove especially useful.

Including tracking in multi-species community assembly models

Predicting how tracking may determine which species are effectively winners and losers with climate change requires integrating non-stationary environments into models of community assembly. Recent advances in coexistence models, sometimes called 'modern coexistence theory,' recognize that both mechanisms independent of fluctuations in the environment (e.g., R^* and other classical niche differences) and mechanisms dependent on fluctuations in the environment (relative non-linearity and storage effect) can lead to coexistence (Chesson & Huntly, 1997; Chesson, 2000). These models, which underlie much of current community ecology research (Mayfield & Levine, 2010; Barabas *et al.*, 2018; Ellner *et al.*, 2019), provide a framework to begin to model environmental tracking and non-stationarity.

How the environment is defined in most community models falls into two broad categories. In some models the environment is expressed as variation in parameters related to species. For example, in some lottery models the environment appears, effectively, as variation in birth and death rates. Building a changing environment into such models thus requires knowing how environmental shifts filter through to species-level parameters (Tuljapurkar *et al.*, 2009). For example, Rudolf (2019) added the temporal environment to competition models by defining interaction strength as dependent on the temporal distance between species. In other models, the environment is more specifically defined. Many of these models define the environment as a resource (e.g., many seed germination models that begin with a resource pulse each year), and thus generally model something close to fundamental tracking. Building a changing environment into these models requires knowing how the environment is changing (see Box 'Environmental variability & change').

Models that explicitly include the environment provide a major opportunity to predict how environmental tracking and non-stationarity determine future communities (see Box: 'Adding tracking and non-stationarity to a common coexistence model'). Yet most current models generally examine the environment from only one of two relevant angles: they represent the environment as used for species' cues (e.g., many models of plasticity) or they represent the environment as directly affecting fitness (e.g., the storage effect model). Combining these two angles may be especially critical to understanding the costs and benefits of tracking in non-stationary environments.

Layered onto the different angles that different models take on the environment is how species responses to the environment are defined. In general, species responses to the (resource) environment can be broadly grouped into models that explicitly define when species start an event (e.g., spawning or germination) in response to the environment versus those that model the magnitude (e.g., the number of propagules or seeds) of response to the environment. Models that explicitly model when a species starts an event are often focused on situations where order of arrival is critical to predicting coexistence outcomes. Models of priority effects through niche pre-emption highlight the advantage tracking may provide when it allows species to be early (and when there is no cost to being too early): early arrivals receive a head-start advantage, by gaining priority access to resources (the environment) they can draw down the resources available to later arrivals (Fukami, 2015).

Other models canalize species' responses to the environment into production and investment. For example most models of inter-annual competition (much of 'modern coexistence theory') fall into this camp. Species produce (via offspring, tissue etc.) differentially depending on the environment each year and outcomes are mediated through density. While these models superficially may seem disconnected from timing, they critically highlight how phenology relates to production and, thus, investment in future years (in contrast, priority effect models explicitly model timing).

A model where species vary both when they start an event and how much they produce dependent on the environment would capture the important attributes of tracking—combining head-start advantages from being early with production variation based on the fitness of the environment. To our knowledge, however, most models approach these questions separately, though models of bet-hedging come closest (Gourbiere & Menu, 2009; Tufto, 2015). A model that explicitly models the linked decisions of when to time an event and how much to offspring/tissue to produce in the event could provide fundamental insights on the relative importance of each aspect of this process. Such a model could be adapted to address multiple questions of environmental tracking, including how these decisions may trade-off and which other traits tracking may be most strongly linked to, as well as explicitly modeling the costs and benefits of tracking in stationary systems, a critical precursor to extending it to non-stationary systems.

Extending models to non-stationary systems is crucial to fully testing how environmental tracking relates to species persistence with climate change, and research has already begun to tackle this non-trivial challenge (Chesson, 2017; Legault & Melbourne, 2019; Rudolf, 2019). Most work to date, however, focuses on conclusions from stationary or non-stationary systems. Thus, the transition between stationary and non-stationary is often ignored, yet we expect it may be most critical. Communities formed in stationary environment periods (or periods with environments lower non-stationarity) are effectively filtered and assembled by that environmental regime and thus produce the base variation and assembly dynamics for a shifting environment. While analytical solutions for systems transitioning from stationary to non-stationary may take time to develop (Chesson, 2017), simulation work can provide an immediate intuition and framework to address this challenge (see Box: Adding tracking and non-stationarity to a common coexistence model).

1.5 Future directions

Growing empirical research highlights that environmental tracking is linked to species performance and, thus, may be critical to understanding the forces that assemble communities and determine species persistence—especially as anthropogenic climate change is reshaping the environment of all species. Ecological theory, including from areas of optimal control, plasticity and community assembly, is clearly primed for understanding how a variable environment can shape the formation and persistence of species and communities. To understand what advances in theory may be most useful for making predictions in the Anthropocene we need more focus on understanding the attributes of an environment shaped strongly by humans. In turn, to test theory we need more robust estimates of environmental tracking and how it fits within a mosaic of other correlated traits. To this aim, we review several major areas of research that we believe could most rapidly unite empirical and theoretical research in environmental tracking to advance the field.

How is an organism's environment changing?
Currently, much research has focused on one major shift in the climate system (earlier growing seasons), but research on multivariate environmental shifts is growing and will be critical to understanding how climate change affects an organism's whole environment. Ecologists can guide these efforts by identifying environmental shifts that are often linked (e.g., warming temperatures may drive earlier seasons and higher evaporative loss of some resources such as water). Researchers can also aim to more consistently and fully characterize the environmental distributions of their systems that appear to most drive species performance and interactions: the environment of the years of study should be clearly reported and compared against long-term and recent climate for each system.

More interdisciplinary research with climate science could also speed a fuller understanding of what shifts are and are not expected with climate change, and what climate variables are inherently correlated. Such correlations make estimating cues and other biological parameters from long-term data especially precarious (Tansey *et al.*, 2017). But these correlations are equally critical in considering how species may view their environment and whether environmental change will couple or uncouple links between proximate cues and fundamental tracking (Bonamour *et al.*, 2019).

Understanding and measuring 'tracking'
Understanding how the environment is changing represents just one step towards robust measures of environmental tracking. Shifting environmental regimes must then be filtered through species cues. As robust estimates of these cues are currently available for few if any species, we suggest several major improvements on current methods to help interpret current trends and make comparisons more feasible.

Studies should clarify their definition of tracking, how the environment is defined and how well, or not, the underlying cue system is understood for study species. Currently, many studies examine fundamental and environmental tracking at once (e.g., Yang & Cenzer, 2020), which is clearly helpful in advancing the field. However, the more researchers can clarify when and how they are addressing fundamental tracking versus environmental tracking the more easily we can compare across studies. Next, and relatedly, studies should define their environment: are they considering primarily the abiotic environment or measuring an environment fundamentally shaped by other species? This difference connects to fundamental versus realized niches and whether systems are primarily top-down (resources and the environment may be strongly shaped by other species) or bottom-up controlled. Finally, all researchers working on environmental tracking need to embrace their inner-physiologist, or collaborate with one. For many species there is often a related species (albeit, sometimes distantly) whose cue system has been studied. Thus, researchers should draw on the literature of their study species' close relatives to bracket which environmental variables may represent environmental tracking and which may be proxies, and to highlight uncertainty. We expect progress will come from a balance between measures of fundamental tracking, estimating an organism's system of cues, and measuring environmental tracking. Clear statements of what is and is not known and measured will help.

Embrace the sometimes contradictory pulls of conducting experiments to identify mechanistic cues and the multivariate climate of the real world. Clearly, we need more experiments to identify which specific aspects of the environment different species cue to and how these cues are filtered by their actual environmental regime (as outlined above and see Chmura *et al.*, 2019). Suites of experiments, which build from identifying cues to understanding how they act when correlated, are a major gap for most organisms.

Build a model of your species' cues and interrogate it. As research progresses in trying to estimate environmental tracking, greater progress will come from fuller and more diverse interrogations of current (and future) models. Define the framework under which you expect your cue system developed or works (e.g., bet-hedging) then test how well it works, and where it fails (see Johansson & Bolmgren, 2019, for an example). Model interrogation can also use realistic environmental regimes to provide field predictions (Wilczek *et al.*, 2010, 2009) or predict future species and communities. One example of this comes from in silica resurrection experiments of model organisms where future environmental regimes included a mix of regular climate projections and projections modified to test and advance understanding of environmental tracking for the study species (e.g., warmer winter and altered photoperiod scenarios in Fournier-Level *et al.*, 2016).

What major traits trade-off with tracking?

Basic theory of plasticity and competition suggest that environmental tracking must trade-off with other traits to allow multi-species communities. Yet to date empirical work has mainly documented tracking, linked it to performance, or focused on how it varies between native and non-native species (Willis *et al.*, 2010; Wolkovich *et al.*, 2013; Zettlemoyer *et al.*, 2019). Such work lays the groundwork that environmental tracking is important, but future empirical research should address how this trait co-occurs with other traits. Research has highlighted some traits that co-vary with tracking (e.g., Kharouba *et al.*, 2014; Lasky *et al.*, 2016; Zhu *et al.*, 2016), but to tie this empirical work to models requires more research on traits that link clearly to theory, and a fuller understanding of how tracking and other traits jointly contribute to performance under varying environments.

Traits that link to resource competition, as we focus on here (as others have as well, see Rudolf, 2019), may be especially fruitful for greater research, but should not be the only ones considered. For example, traits related to predator tolerance or avoidance may also play a role, but have been effectively unstudied. As empirical research in this area grows, models can aid progress in understanding the outcomes of these trade-offs for community assembly.

Developing new models for tracking in stationary to non-stationary systems
As outlined above many areas of ecological and evolutionary modeling contribute to our understanding of environmental tracking. But most are limited in various ways. Community ecology models generally bifurcate in modeling differences in timing versus production amounts across species, thus studies of whether these models lead to similar or different conclusions could help predict community outcomes and advance our understanding of trade-offs. As outlined above, understanding tracking likely requires models that combine effects. This includes models that combine effects of variation in timing and production amounts and models that include environment as impacting species' cues, as well as species' fitness. Such models would explicitly allow the potential costs and benefits of tracking depending on how closely environmental tracking matches fundamental tracking.

New models will also need to examine how relaxing assumptions of closed communities (i.e., without dispersal or evolution) alter predictions. In practice, dispersal of species or individuals with traits that make them better matched to the non-stationary environment would lead to new communities that may persist longer or be continually re-assembled as long as the environment remains non-stationary. Indeed, this logic underlies the argument that invasive species may be superior trackers benefiting from how climate change has altered growing seasons (Willis *et al.*, 2010; Wolkovich & Cleland, 2011). Evolution equally could alter predictions by allowing species traits to evolve in step with environmental change. Long-term population (e.g., Colautti *et al.*, 2017) and resurrection studies (Wilczek *et al.*, 2014; Yousey *et al.*, 2018), as well as field experiments (Colautti *et al.*, 2017; Exposito-Alonso *et al.*, 2019), have repeatedly shown species can shift to earlier flowering times, higher thermal tolerances or related genetically-controlled traits that confer higher fitness in warmer climates. Yet these studies also highlight that responses can be lagged (e.g., Wilczek *et al.*, 2014), associated with reduced population viability (e.g., Colautti *et al.*, 2017), or other factors that may constrain adaptive responses.

1.6 Stationarity in the future

While most environments today are climatically non-stationary and have been for decades, the climate will return to stationarity in the future. There are many possible pathways to climatic stabilization, but almost all require first the stabilization of greenhouse gases—the subject of much policy and political debate. Once greenhouse gas emissions stabilize climate will not quickly snap back to a new stationary phase. Instead systems will slowly approach a new climatic stationarity depending on how they are effected by the earth's multiple thermal reservoirs, and, in turn, how quickly those reservoirs stabilize. The timescale of this approach is generally expected to be on the scale of centuries, but could be much longer in certain oceanic systems (Collins *et al.*, 2013). Thus, ecologists are—and will remain for the foreseeable future—in a research area structured by climatic non-stationarity.

As paleobiologists and evolutionary biologists often point out, climatic nonstationarity is a common part of the earth's history (Jansson & Dynesius, 2002)—even if stationary periods—be they cold or warm (glacial and interglacial periods)—are more common. Indeed, while much of this work has examined how species survive for millions of years given large oscillations in climate (Provan & Bennett, 2008), the periods that provide the most dramatic community reshuffling are periods shifting from stationary to non-stationary climate regimes (Vrba, 1980, 1985). Such stories of the past are now fundamentally happening today, and ecology is challenged to understand how transitions between stationary and non-stationary environments are reshaping the species and communities we have today and will in our warmer future.

2 Acknowledgments

We thank I. Breckheimer, D. Buonaiuto, E. Cleland, J. Davies, G. Legault and two anonymous reviewers for helpful comments that improved the manuscript and the NSERC Discovery Award (RGPIN-05038 to EMW) and Canada Research Chair in Temporal Ecology (EMW) programs that provided funding.

3 Boxes

3.1 Box: Environmental variability & change

Decades of ecological research highlight how temporally variable environments shape species and their communities at multiple scales (Sale, 1977; Chesson & Huntly, 1997). In seasonal landscapes, the environment limits periods for growth each year (e.g., by temperature or drought); within-year variability in the environment (e.g., daily, hourly or finer resolution temperatures or rainfall amounts) compounds into inter-annual variability that shapes the distribution of the start and end of growing seasons. For long stretches of history this variability has been effectively stationary; that is, the underlying probability distribution that describes the start (or end) of the season (e.g., the date of the last major frost) does not change, even though the date may be dramatically different from one year to the next.

In other time periods, variability has been non-stationary in one or multiple dimensions. For example, climate in the northern hemisphere includes long warming and then cooling periods (i.e., increasing then decreasing means of the probability distribution) at the start of the Holocene, when the earth was coming out of the last glacial maximum. Anthropogenic climate change is a similar non-stationary process, with warming evident around the globe and knock-on effects for other climate metrics, such as heat extremes and the size of precipitation events.

Understanding non-stationarity in ecological systems requires first identifying which aspects of the environment have shifted—and how they have shifted with respect to one another—as the underlying distributions transition from stationary to non-stationary (Fig. 2). For example, with climate change, warming has increased mean temperatures over time, with minimum temperatures generally increasing more than maximum—this results in an underlying distribution for daily temperature where the mean is increasing through time while the within-day variance is decreasing (Stocker *et al.*, 2013; Screen, 2014). Understanding the impacts of climate change further requires recognizing that many systems can be considered stationary or non-stationary depending on the timescale and period of study. Thus, predicting the consequences of current non-stationarity in ecological systems benefits from identifying the type and scale of non-stationarity, relative to long-term trends.

3.2 Box: Statistical challenges in measuring tracking

A potentially widespread reason for observations of species that do not track is statistical bias and artifacts, including non-stationarity in units and unrecognized low power. All of these can be addressed given improved statistical approaches (e.g., Gienapp *et al.*, 2005; Pearse *et al.*, 2017), though such approaches may (uncomfortably) highlight how uncertain many current estimates are (Brown *et al.*, 2016). Non-stationarity in units comes in many forms—estimates of mean days shifted per decade (i.e., days/decade, a common metric summarizing observed shifts in phenology over time in long-term datasets) depend strongly on the climate of the decade(s) studied, which is not consistent in many systems (Ault *et al.*, 2011; McCabe *et al.*, 2012). Estimates based on a relevant climate variable can sometimes ameliorate this problem, but may be equally vulnerable to non-stationarity in units (e.g., Sagarin, 2001). For example, processes that depend on thermal sums reported as days/°C will generally appear to decline with warming, as the thermal sum of an average day has increased in most regions with climate change. Relatedly, estimates of long-term change using simple linear regression are influenced by the climate at the start of the time-series (with greater changes seen from time-series that started in unusually cold decades, such as the 1950s for much of North America). Impacts of start-years for long-term time-series can be muted by applying change-point or hinge models (e.g., Kharouba *et al.*, 2018), while metrics that move away from calendar units (e.g., day) can help address non-stationarity in units.

Low power is widespread in ecology, where even 'long' time-series may be far too short for robust analyses (Bolmgren *et al.*, 2013; Kharouba *et al.*, 2018). Authors should be especially cautious if they find only large effects appear significant (e.g., CaraDonna *et al.*, 2014), which is a well-known statistical bias associated with p-values (Loken & Gelman, 2017). Additionally, effect sizes that are higher when climate variability is higher (for example, in temperate habitats temperature is highly variable in the spring and autumn compared to summer) may be more related to variation in statistical power than to biology (periods with higher variation yield greater variation in the predictor variable, and thus higher power). Mixed models can help better leverage understanding by pooling information across species, and often better capture uncertainty (Pearse *et al.*, 2017). We suggest mixed models should be used more widely alongside randomization and/or data-simulation approaches (e.g., Bolmgren *et al.*, 2013; Kharouba *et al.*, 2018) to better estimate and communicate uncertainty in studies. And researchers should identify what results bias may produce. For example, growing evidence suggests a potential fundamental trade-off where early species track and possess a suite of traits to related to faster growth and shorter lifespans, while later species track less and possess traits related to slower growth and longer lifespans—these later species may bet-hedge more given their longer investment window. This, however, could equally be an artifact where early species use simpler cues, and, thus, their tracking is measured more accurately given current methods.

3.3 Box: Trait trade-offs with tracking

Research on phenological tracking and traits has increased greatly in recent years, with a major uptick in studies after 2010 (see SI Fig. **??**). Most papers examining tracking and other traits across species have focused on plants (20/30), followed by birds and Lepidoptera (both 4/30), plankton and aphids (both 1/30). The most studied trait was how early or late a phenophase occurred (e.g., date of flowering of start of migration for a species, termed 'earlyness' by some authors), with earlier species tending to track more (studies included both birds and Lepidotera, Diamond *et al.*, 2011; Ishioka *et al.*, 2013; Kharouba *et al.*, 2014; Jing *et al.*, 2016; Du *et al.*, 2017). While this is an important link, it is vulnerable to statistical challenges (see Box 'Statistical challenges in measuring tracking'). Few studies examined whether tracking correlates with resource acquisition traits; those that did generally found species with higher tracking also had traits associated with lower competitive abilities under low resources (e.g., being shallower or lacking a taproot rooted Dorji *et al.*, 2013; Lasky *et al.*, 2016; Zhu *et al.*, 2016). These species were often also early (e.g., Dorji *et al.*, 2013; Zhu *et al.*, 2016), suggesting tracking may relate to a syndrome of traits that allow species to be rapid colonizers each season, but poor competitors for resources. Indeed, previous work has documented that species with earlier phenophases tend to have resource acquisition traits associated with lower competitive abilities (e.g., they tend to be of lower height, have shallower roots, narrower diameter vessels, thinner leaves, and grow faster, reviewed in Wolkovich & Cleland, 2014).

3.4 Box: Adding tracking and non-stationarity to a common coexistence model

To understand the role of environmental tracking by species in variable environments we use a simple model that allows within- and between-year dynamics to contribute to coexistence. As the model is akin to many commonly used seed germination models (Chesson *et al.*, 2004), we follow a similar terminology for ease; however the basic structure of our model could apply to other systems with one dominant (non-renewing) pulse of a limiting resource each season (e.g., water from rain or snowpack), or a multivariate resource pulse that acts effectively as one resource (e.g., nitrogen and light drawn down together over the season). In this model the environment is included between-years via variable germination, and within-years the environment is explicitly included as a resource pulse at the start of the season. We adjust the biological start time of species (τ_i for species i) to also allow species to respond to the environment dynamically through what we refer to as tracking. Here, tracking effectively moves a species intrinsic start time closer to the environmental start time in that year, resulting in a higher germination fraction (see SI for complete description and equations).

Species can co-occur via equalizing mechanisms, but they require stabilizing mechanisms to coexist. Species that is, cannot coexist given only variation in tracking—coexistence requires variation in another trait axis. As theory and empirical work suggest this trade-off may involve traits related closely to resource competition, we varied species' R^*. With variation in tracking and in R^* species can persist together as long as those species with a temporal niche advantage are also the inferior competitors (Fig. 3). That is, species that can draw resources down to a lower level and are thus the superior within-season resource competitors (lower R^*) can persist with species with that are inferior competitors but have realized biological start times closer to the environmental start time—a finding inline with currently observed empirical trade-offs (see Box 'Trait trade-offs with tracking'). These trade-offs, however, are all environmentally dependent. They hold only so long as the environment is stationary.

We examined how trade-offs may be transformed by a non-stationary environment, by transitioning a stationary environment—in which two-species communities had persisted for 500 years—to non-stationary, via an earlier start of season (earlier timing of the resource pulse, τ_p, Fig. 3a; see SI for more details). By changing a fundamental niche axis (the distribution of the environment, an axis along which these communities were structured), we shifted one major part of the trade-off: the new non-stationary environment favored an earlier start time than the previous stationary environment. This, in turn, reshaped our two-species communities, which depended on this trade-off for persistence.

While the non-stationary environment favored higher trackers (who in turn drove the extinction of species with lower tracking values from many two-species communities) some two-species communities persisted (257 out of 1698 two-species communities persisting after end of stationary, or 15.1%, Fig. 3). These two-species communities persisted because the same fundamental trade-off between biological start time and within-season competitive ability, while narrowed, was not fully lost (Fig. 3). Taken together, these simple simulations show how non-stationarity can drive local species extinction and reshape the underlying assembly mechanisms of communities.

Our simulations support growing work that tracking should not be considered alone (Diamond *et al.*, 2011; Dorji *et al.*, 2013; Ishioka *et al.*, 2013; Kharouba *et al.*, 2014; Du *et al.*, 2017), but may be part of a larger trait syndrome. Indeed, this model trivially show that multi-species communities cannot form given only variation in the temporal niche—a trade-off is required. Our results thus support empirical work showing a trade-off where trackers are also inferior resource competitors (Lasky *et al.*, 2016; Zhu *et al.*, 2016)—we show this must be the case for multi-species persistence; otherwise, the species best matched to the environment would drive the other extinct.

Finally, our results highlight that non-stationarity may reshape the balance of equalizing versus stabilizing mechanisms. As environments shifted from stationarity to non-stationarity, species that co-occured via equalizing mechanisms persisted longer. While the outcome that equalized species will be more similarly affected by environmental shifts is rather obvious, it has several important implications. First, it may make identifying which traits climate change promotes through stabilizing mechanisms more difficult. Second, it suggests climate change—or other factors that cause an environment to shift from stationary to non-stationary—may cause a fundamental shift away from assembly via stabilizing mechanisms. Thus understanding the prevalence of stabilizing versus equalizing mechanisms (which ecology has worked on for many decades, Caswell, 1976; Chesson, 2000) becomes critical for understanding the implications of transitions to non-stationary environments.

4 Figures

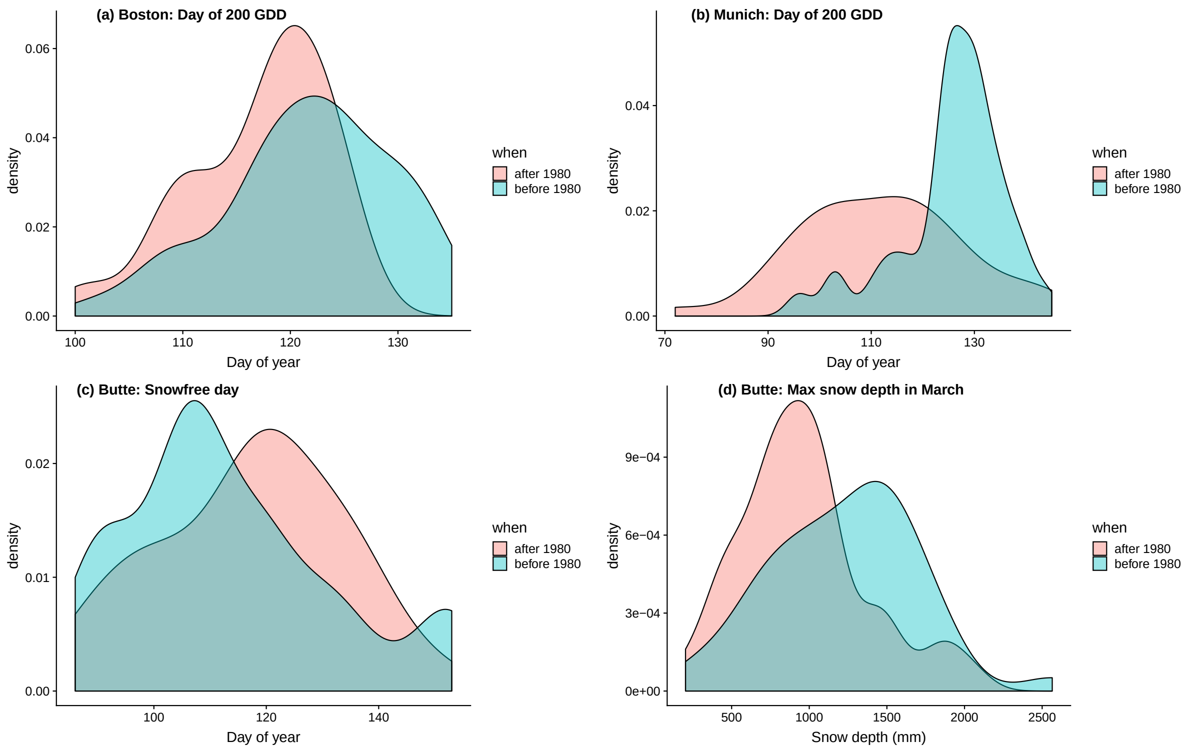

Figure 1: Boom, we define ET.

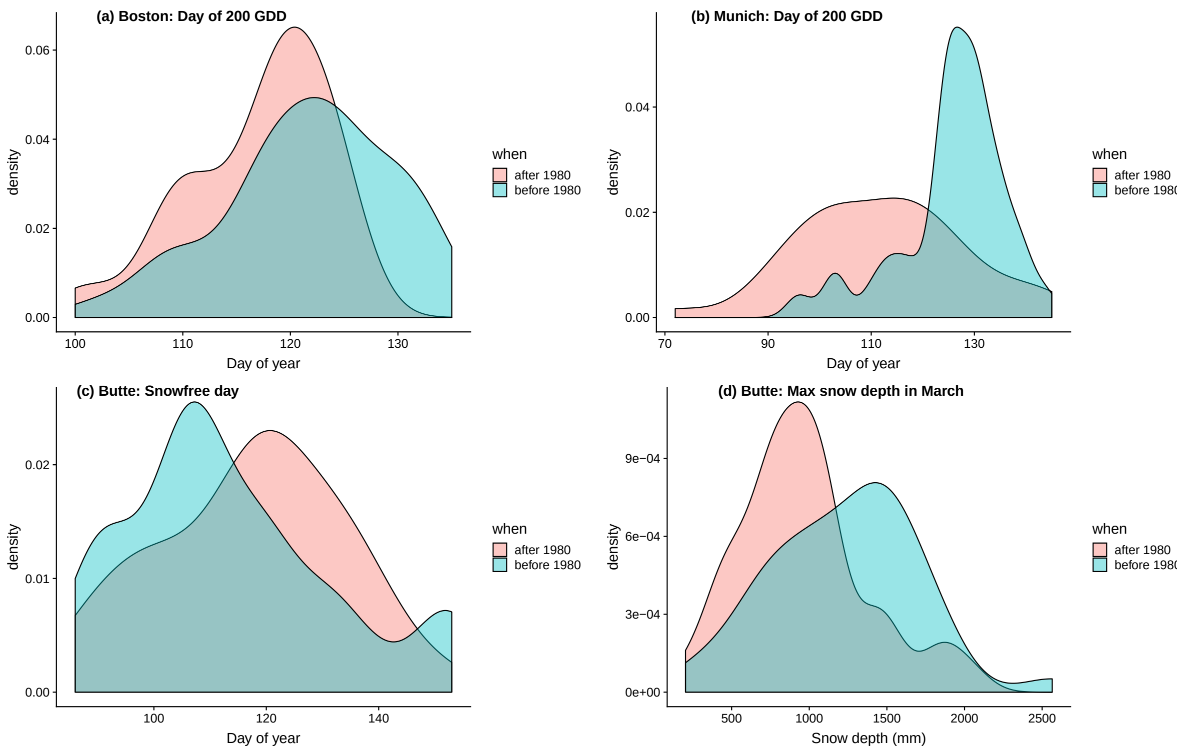

Figure 2: Examples of non-stationarity in climate variables linked to environmental tracking: shifts before and after 1980 (a major change-point in climate for many regions) in several metrics related to the start of growing seasons (a-c) or resource pulse connected to growing season length (d). Density plots of day of 200 growing degree day units (a metric of thermal sum, here based on 0 degree base temperature using daily minima in °C) in Boston, MA, USA (a), and Munich, Germany (b), first snowfree day (followed by at least 9 snowfree days) in Crested Butte, CO, USA (c) and maximum snowdepth (mm) in March (often the month before the first snowfree day) in Crested Butte, CO, USA (d). Note that (c) and (d) are likely related, with lower snowpacks leading to an earlier first snowfree day. We selected sites that have been studied for plant phenological data and included at least 80 years of daily climate data from a Global Historical Climatology Network site; we subsetted data so that there were 40 years before and after 1980 for all sites.

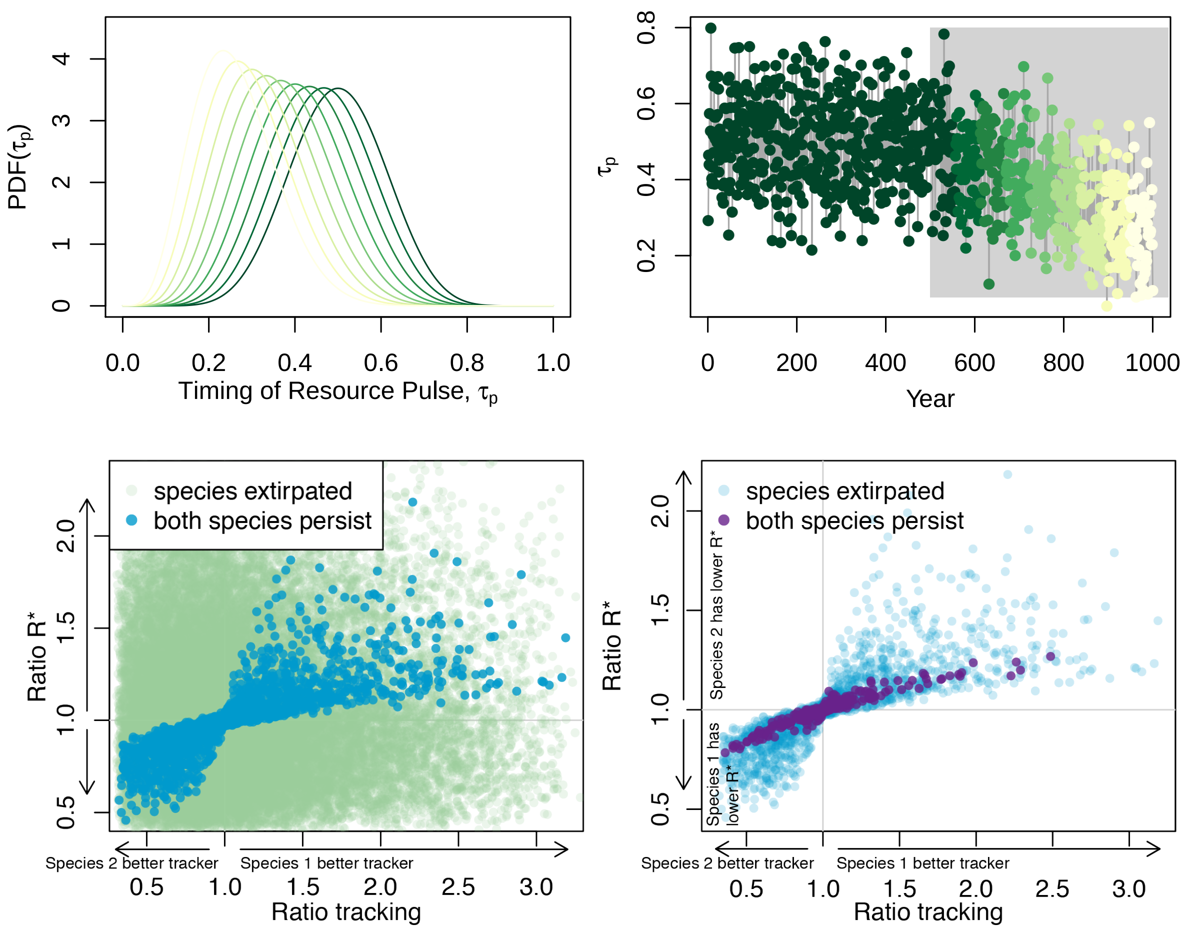

Figure 3: Example of how non-stationarity can reshape communities in a simple model. We shifted the environment (top panels) by changing the timing of the resource pulse from a stationary period ($\tau_p \sim \beta(10, 10)$ for the 500 years) to a nonstationary period ($\tau_p \sim \beta(5, 15)$ over the 500 years), then examined outcomes for two-species communities (bottom panels) where tracking (X axis: species 1/species 2) trades off with R^* (Y axis: species 1/species 2): each point represents one two-species community color-coded by whether both species persisted or one or more species was extirpated through 500 years of a stationary environment (bottom-left), followed by an additional 500 years of non-stationary environment (bottom-right, only two-species communities that persisted through the stationary period are shown).

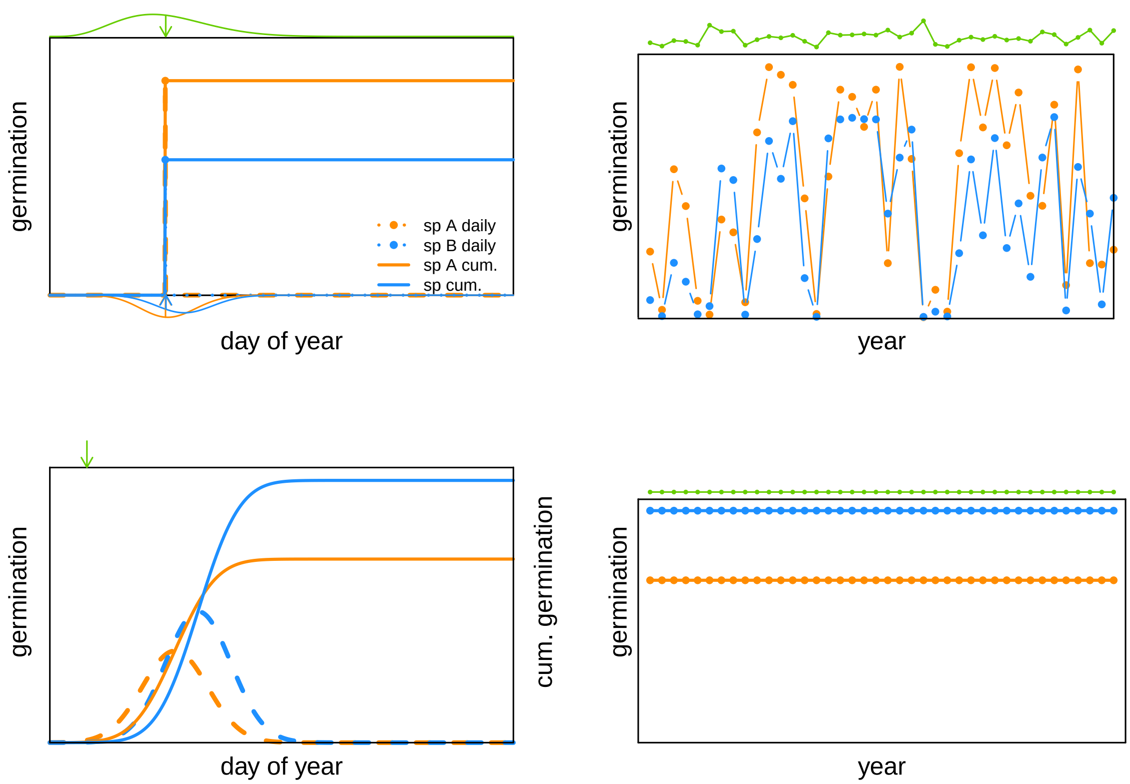

Figure 4: Goober.

References

Ackerly, D. (2009). Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19699–19706.

Amarasekare, P. (2003). Competitive coexistence in spatially structured environments: a synthesis. *Ecology Letters*, 6.

Arnold, S.J. (1992). Constraints on phenotypic evolution. *American Naturalist*, 140, S85–S107.

Auld, J.R., Agrawal, A.A. & Relyea, R.A. (2010). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B-Biological Sciences*, 277, 503–511.

Ault, T.R., Macalady, A.K., Pederson, G.T., Betancourt, J.L. & Schwartz, M.D. (2011). Northern hemisphere modes of variability and the timing of spring in western north america. *Journal of Climate*, 24, 4003–4014.

Bagnall, D.J. (1993). Light quality and vernalization interact in controlling late flowering *Arabidopsis* ecotypes and mutants. *Annals of Botany*, 71, 75–83.

Barabas, G., D'Andrea, R. & Stump, S.M. (2018). Chesson's coexistence theory. *Ecological Monographs*, 88, 277–303.

Block, S., Alexander, J. & Levine, J.M. (2019). Phenological plasticity is a poor predictor of subalpine plant population performance following experimental climate change. *Oikos*, p. 10.1111/oik.06667.

Bolmgren, K., Vanhoenacker, D. & Miller-Rushing, A.J. (2013). One man, 73 years, and 25 species. evaluating phenological responses using a lifelong study of first flowering dates. *International Journal of Biometeorology*, 57, 367–375.

Bonamour, S., Chevin, L.M., Charmantier, A. & Teplitsky, C. (2019). Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 374.

Brown, C.J., O'Connor, M.I., Poloczanska, E.S., Schoeman, D.S., Buckley, L.B., Burrows, M.T., Duarte, C.M., Halpern, B.S., Pandolfi, J.M., Parmesan, C. & Richardson, A.J. (2016). Ecological and methodological drivers of species' distribution and phenology responses to climate change. *Global Change Biology*, 22, 1548–1560.

CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111, 4916–4921.

Carter, S.K., Saenz, D. & Rudolf, V.H.W. (2018). Shifts in phenological distributions reshape interaction potential in natural communities. *Ecology Letters*, 21, 1143–1151.

de Casas, R.R., Donohue, K., Venable, D.L. & Cheptou, P.O. (2015). Gene-flow through space and time: dispersal, dormancy and adaptation to changing environments. *Evolutionary Ecology*, 29, 813–831.

Caswell, H. (1976). Community structure - neutral model analysis. *Ecological Monographs*, 46, 327–354.

Charmantier, A., McCleery, R.H., Cole, L.R., Perrins, C., Kruuk, L.E.B. & Sheldon, B.C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, 320, 800–803.

Chen, I.C., Hill, J.K., Ohlemuller, R., Roy, D.B. & Thomas, C.D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, 333, 1024–1026.

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

Chesson, P. (2017). Aedt: A new concept for ecological dynamics in the ever-changing world. *Plos Biology*, 15.

Chesson, P., Gebauer, R.L.E., Schwinning, S., Huntly, N., Wiegand, K., Ernest, M.S.K., Sher, A., Novoplansky, A. & Weltzin, J.F. (2004). Resource pulses, species interactions, and diversity maintenance in arid and semi-arid environments. *Oecologia*, 141, 236–253.

Chesson, P. & Huntly, N. (1997). The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *American Naturalist*, 150, 519–553.

Chevin, L.M., Lande, R. & Mace, G.M. (2010). Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *Plos Biology*, 8.

Chmielewski, F.M., Blumel, K., Scherbaum-Heberer, C., Koppmann-Rumpf, B. & Schmidt, K.H. (2013). A model approach to project the start of egg laying of great tit (parus major l.) in response to climate change. *International Journal of Biometeorology*, 57, 287–297.

Chmura, H.E., Kharouba, H.M., Ashander, J., Ehlman, S.M., Rivest, E.B. & Yang, L.H. (2019). The mechanisms of phenology: the patterns and processes of phenological shifts. *Ecological Monographs*, 89.

Chuine, I., Bonhomme, M., Legave, J.M., de Cortazar-Atauri, I.G., Charrier, G., Lacointe, A. & Ameglio, T. (2016). Can phenological models predict tree phenology accurately in the future? the unrevealed hurdle of endodormancy break. *Global Change Biology*, 22, 3444–3460.

Chuine, I. & Regniere, J. (2017). Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics*, 48, 159–182.

Cleland, E.E., Allen, J.M., Crimmins, T.M., Dunne, J.A., Pau, S., Travers, S.E., Zavaleta, E.S. & Wolkovich, E.M. (2012). Phenological tracking enables positive species responses to climate change. *Ecology*, 93, 1765–1771.

Cohen, J.M., Lajeunesse, M.J. & Rohr, J.R. (2018). A global synthesis of animal phenological responses to climate change. *Nature Climate Change*, 8, 224–+.

Colautti, R.I., Aring;gren, J. & Anderson, J.T. (2017). Phenological shifts of native and invasive species under climate change: insights from the boechera - lythrum model. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 372.

Collins, M., Knutti, R., Arblaster, J.L., Dufresne, T., Fichefet, P., Friedlingstein, X., Gao, W., Gutowski, T., Johns, G., Krinner, M., Shongwe, C., Tebaldi, A., Weaver & Wehner, M. (2013). *Climate Change 2013: The Physical Science Basis – Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, chap. Long-term Climate Change: Projections, Commitments and Irreversibility, pp. 1029–1119.

Cook, B.I., Wolkovich, E.M. & Parmesan, C. (2012). Divergent responses to spring and winter warming drive community level flowering trends. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 9000–9005.

Deacy, W.W., Erlenbach, J.A., Leacock, W.B., Stanford, J.A., Robbins, C.T. & Armstrong, J.B. (2018). Phenological tracking associated with increased salmon consumption by brown bears. *Scientific Reports*, 8.

Dennis, F. (2003). Problems in standardizing methods for evaluating the chilling requirements for the breaking of dormancy in buds of woody plants. *HortScience*, 38, 347–350.

Diamond, S.E., Frame, A.M., Martin, R.A. & Buckley, L.B. (2011). Species' traits predict phenological responses to climate change in butterflies. *Ecology*, 92, 1005–1012.

Donahue, M.J., Karnauskas, M., Toews, C. & Paris, C.B. (2015). Location isn't everything: Timing of spawning aggregations optimizes larval replenishment. *Plos One*, 10.

Donaldson-Matasci, M.C., Bergstrom, C.T. & Lachmann, M. (2012). When unreliable cues are good enough. *American Naturalist*, 182, 313–327.

Dore, A.A., McDowall, L., Rouse, J., Bretman, A., Gage, M.J.G. & Chapman, T. (2018). The role of complex cues in social and reproductive plasticity. *Behavioral Ecology and Sociobiology*, 72.

Dorji, T., Totland, O., Moe, S.R., Hopping, K.A., Pan, J. & Klein, J.A. (2013). Plant functional traits mediate reproductive phenology and success in response to experimental warming and snow addition in Tibet. *Global Change Biology*, 19, 459–472.

Du, Y.J., Chen, J.R., Willis, C.G., Zhou, Z.Q., Liu, T., Dai, W.J., Zhao, Y. & Ma, K.P. (2017). Phylogenetic conservatism and trait correlates of spring phenological responses to climate change in northeast china. *Ecology and Evolution*, 7, 6747–6757.

Ellner, S.P., Snyder, R.E., Adler, P.B. & Hooker, G. (2019). An expanded modern coexistence theory for empirical applications. *Ecology Letters*, 22, 3–18.

Exposito-Alonso, M., Burbano, H.A., Bossdorf, O., Nielsen, R., Weigel, D., Exposito-Alonso, M., Rodriguez, R.G., Barragan, C., Capovilla, G., Chae, E., Devos, J., Dogan, E.S., Friedemann, C., Gross, C., Lang, P., Lundberg, D., Middendorf, V., Kageyama, J., Karasov, T., Kersten, S., Petersen, S., Rabbani, L., Regalado, J., Reinelt, L., Rowan, B., Seymour, D.K., Symeonidi, E., Schwab, R., Tran, D.T.N., Venkataramani, K., Van de Weyer, A.L., Vasseur, F., Wang, G., Wedegartner, R., Weiss, F., Wu, R., Xi, W.Y., Zaidem, M., Zhu, W.S., Garcia-Arenal, F., Burbano, H.A., Bossdorf, O., Weigel, D. & Genomes Field Expt, T. (2019). Natural selection on the arabidopsis thaliana genome in present and future climates. *Nature*, 573, 126–+.

Fournier-Level, A., Perry, E.O., Wang, J.A., Braun, P.T., Migneault, A., Cooper, M.D., Metcalf, C.J.E. & Schmitt, J. (2016). Predicting the evolutionary dynamics of seasonal adaptation to novel climates in arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E2812–E2821.

Fox, R.J., Donelson, J.M., Schunter, C., Ravasi, T. & Gaitan-Espitia, J.D. (2019). Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 374.

Franks, S.J., Sim, S. & Weis, A.E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 1278–1282.

Fu, Y.S.H., Zhao, H.F., Piao, S.L., Peaucelle, M., Peng, S.S., Zhou, G.Y., Ciais, P., Huang, M.T., Menzel, A., Uelas, J.P., Song, Y., Vitasse, Y., Zeng, Z.Z. & Janssens, I.A. (2015). Declining global warming effects on the phenology of spring leaf unfolding. *Nature*, 526, 104–107.

Fukami, T. (2015). *Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects*, vol. 46 of *Annual Review of Ecology Evolution and Systematics*, pp. 1–23.

Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N. & Hughes, K.A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, 525, 372–+.

Ghalambor, C.K., McKay, J.K., Carroll, S.P. & Reznick, D.N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21, 394–407.

Gienapp, P., Hemerik, L. & Visser, M.E. (2005). A new statistical tool to predict phenology under climate change scenarios. *Global Change Biology*, 11, 600–606.

Gotelli, N.J. & Graves, G.R. (1996). The temporal niche. In: *Null Models in Ecology* (eds. Gotelli, N.J. & Graves, G.R.). Smithsonian Institution, Washington, DC, pp. 95–111.

Gould, S.J. & Lewontin, R.C. (1979). Spandrels of San-Marco and the Panglossian paradigm - a critique of the adaptationist program. *Proceedings of the Royal Society Series B-Biological Sciences*, 205, 581–598.

Gourbiere, S. & Menu, F. (2009). Adaptive dynamics of dormancy duration variability: evolutionary trade-off and priority effect lead to suboptimal adaptation. *Evolution*, 63, 1879–1892.

Gremer, J.R., Kimball, S. & Venable, D.L. (2016). Within-and among-year germination in sonoran desert winter annuals: bet hedging and predictive germination in a variable environment. *Ecology Letters*, 19, 1209–1218.

Hausfather, Z., Drake, H.F., Abbott, T. & Schmid, G.A. (2020). Evaluating the performance of past climate model projections. *Geophysical Research Letters*, 47, e2019GL085378.

Heberling, J.M., Cassidy, S.T., Fridley, J.D. & Kalisz, S. (2019). Carbon gain phenologies of spring-flowering perennials in a deciduous forest indicate a novel niche for a widespread invader. *New Phytologist*, 221, 778–788.

Inouye, B.D., Ehrlen, J. & Underwood, N. (2019). Phenology as a process rather than an event: from individual reaction norms to community metrics. *Ecological Monographs*, 89.

IPCC (2014). *Climate Change 2014: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ishioka, R., Muller, O., Hiura, T. & Kudo, G. (2013). Responses of leafing phenology and photosynthesis to soil warming in forest-floor plants. *Acta Oecologica-International Journal of Ecology*, 51, 34–41.

Jansson, R. & Dynesius, M. (2002). The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, 33, 741–777.

Jing, J., Li, K. & Liu, Z.G. (2016). Effects of varying temperature on leaf phenology and herbivory of dominant tree species in subtropical evergreen broad-leaves forest in eastern china. *Polish Journal of Ecology*, 64, 53–69.

Johansson, J. & Bolmgren, K. (2019). Is timing of reproduction according to temperature sums an optimal strategy? *Ecology and Evolution*, 9, 11598–11605.

Johansson, J. & Jonzen, N. (2012). Game theory sheds new light on ecological responses to current climate change when phenology is historically mismatched. *Ecol Lett*, 15, 881–8.

Kharouba, H.M., Ehrlen, J., Gelman, A., Bolmgren, K., Allen, J.M., Travers, S.E. & Wolkovich, E.M. (2018). Global shifts in the phenological synchrony of species interactions over recent decades. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 5211–5216.

Kharouba, H.M., Paquette, S.R., Kerr, J.T. & Vellend, M. (2014). Predicting the sensitivity of butterfly phenology to temperature over the past century. *Global Change Biology*, 20, 504–514.

Kingsolver, J.G. (2007). Variation in growth and instar number in field and laboratory manduca sexta. *Proceedings of the Royal Society B-Biological Sciences*, 274, 977–981.

Knutti, R., Rugenstein, M.A.A. & Hegerl, G.C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10, 727–+.

Lasky, J.R., Uriarte, M. & Muscarella, R. (2016). Synchrony, compensatory dynamics, and the functional trait basis of phenological diversity in a tropical dry forest tree community: effects of rainfall seasonality. *Environmental Research Letters*, 11.

Legault, G. & Melbourne, B.A. (2019). Accounting for environmental change in continuous-time stochastic population models. *Theoretical Ecology*, 12, 31–48.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology Evolution*, 17, 183–189.

Levy, O., Appelbaum, L., Leggat, W., Gothlif, Y., Hayward, D.C., Miller, D.J. & Hoegh-Guldberg, O. (2007). Light-responsive cryptochromes from a simple multicellular animal, the coral acropora millepora. *Science*, 318, 467–470.

Loken, E. & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355, 584–585.

Mayfield, M.M. & Levine, J.M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, 13, 1085–1093.

McCabe, G.J., Ault, T.R., Cook, B.I., Betancourt, J.L. & Schwartz, M.D. (2012). Influences of the el nino southern oscillation and the pacific decadal oscillation on the timing of the north american spring. *International Journal of Climatology*, 32, 2301–2310.

Menzel, A., Von Vopelius, J., Estrella, N., Schleip, C. & Dose, V. (2006). Farmers' annual activities are not tracking the speed of climate change. *Climate Research*, 32, 201–207.

Metcalf, C.J.E., Burghardt, L.T. & Koons, D.N. (2015). Avoiding the crowds: the evolution of plastic responses to seasonal cues in a density-dependent world. *Journal of Ecology*, 103, 819–828.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. & Stouffer, R.J. (2008). Climate change - stationarity is dead: Whither water management? *Science*, 319, 573–574.

Morris, W.F., Pfister, C.A., Tuljapurkar, S., Haridas, C.V., Boggs, C.L., Boyce, M.S., Bruna, E.M., Church, D.R., Coulson, T., Doak, D.F., Forsyth, S., Gaillard, J.M., Horvitz, C.C., Kalisz, S., Kendall, B.E., Knight, T.M., Lee, C.T. & Menges, E.S. (2008). Longevity can buffer plant and animal populations against changing climatic variability. *Ecology*, 89, 19–25.

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F. & van Kleunen, M. (2010). Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, 15, 684–692.

Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Mueller, W.E.G., Nickel, M., Schierwater, B., Vacelet, J., Wiens, M. & Woerheide, G. (2013). Deep metazoan phylogeny: When different genes tell different stories. *Molecular Phylogenetics and Evolution*, 67, 223–233.

Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B. & Davies, T.J. (2017). A statistical estimator for determining the limits of contemporary and historic phenology. *Nature Ecology Evolution*, 1, 1876–+.

Pigliucci, M. (1998). Developmental phenotypic plasticity: where internal programming meets the external environment. *Current Opinion in Plant Biology*, 1, 87–91.

Provan, J. & Bennett, K.D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology Evolution*, 23, 564–571.

Richardson, A.D., Bailey, A.S., Denny, E.G., Martin, C.W. & O'Keefe, J. (2006). Phenology of a northern hardwood forest canopy. *Global Change Biology*, 12, 1174–1188.

Rudolf, V.H.W. (2019). The role of seasonal timing and phenological shifts for species coexistence. *Ecology Letters*.

Sagarin, R. (2001). Phenology - false estimates of the advance of spring. *Nature*, 414, 600–600.

Sale, P.F. (1977). Maintenance of high diversity in coral reef fish communities. *American Naturalist*, 111, 337–359.

Screen, J.A. (2014). Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nature Climate Change*, 4, 577–582.

Simmonds, E.G., Cole, E.F. & Sheldon, B.C. (2019). Cue identification in phenology: A case study of the predictive performance of current statistical tools. *Journal of Animal Ecology*, 88, 1428–1440.

Singer, M.C. & Parmesan, C. (2010). Phenological asynchrony between herbivorous insects and their hosts: signal of climate change or pre-existing adaptive strategy? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 3161–3176.

Smaldino, P.E. & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3.

Steer, N.C., Ramsay, P.M. & Franco, M. (2019). nlstimedist: An r package for the biologically meaningful quantification of unimodal phenology distributions. *Methods in Ecology and Evolution*, 10, 1934–1940.

Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., Purugganan, M.D. & Schmitt, J. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 4712–4717.

Stocker, T., Qin, D. & Platner, G. (2013). Climate change 2013: The physical science basis. *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013).*

Tansey, C.J., Hadfield, J.D. & Phillimore, A.B. (2017). Estimating the ability of plants to plastically track temperature-mediated shifts in the spring phenological optimum. *Global Change Biology*, 23, 3321–3334.

Thackeray, S.J., Henrys, P.A., Hemming, D., Bell, J.R., Botham, M.S., Burthe, S., Helaouet, P., Johns, D.G., Jones, I.D., Leech, D.I., Mackay, E.B., Massimino, D., Atkinson, S., Bacon, P.J., Brereton, T.M., Carvalho, L., Clutton-Brock, T.H., Duck, C., Edwards, M., Elliott, J.M., Hall, S.J.G., Harrington, R., Pearce-Higgins, J.W., Høye, T.T., Kruuk, L.E.B., Pemberton, J.M., Sparks, T.H., Thompson, P.M., White, I., Winfield, I.J. & Wanless, S. (2016). Phenological sensitivity to climate across taxa and trophic levels. *Nature*, 535, 241.

Tufto, J. (2015). Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution*, 69, 2034–2049.

Tuljapurkar, S., Gaillard, J.M. & Coulson, T. (2009). From stochastic environments to life histories and back. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 364, 1499–1509.

Venable, D.L. (2007). Bet hedging in a guild of desert annuals. *Ecology*, 88, 1086–1090.

Visser, M.E. & Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. *Nature Ecology Evolution*, 3, 879–885.

Visser, M.E., Holleman, L.J.M. & Gienapp, P. (2006). Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia*, 147, 164–172.

Vitasse, Y. (2013). Ontogenic changes rather than difference in temperature cause understory trees to leaf out earlier. *New Phytologist*, 198, 149–155.

Vrba, E.S. (1980). Evolution, species and fossils - how does life evolve. *South African Journal of Science*, 76, 61–84.

Vrba, E.S. (1985). Environment and evolution - alternative causes of the temporal distribution of evolutionary events. *South African Journal of Science*, 81, 229–236.

Wainwright, C.E. & Cleland, E.E. (2013). Exotic species display greater germination plasticity and higher germination rates than native species across multiple cues. *Biological Invasions*, 15, 2253–2264.

Wainwright, C.E., Wolkovich, E.M. & Cleland, E.E. (2012). Seasonal priority effects: implications for invasion and restoration in a semi-arid system. *Journal of Applied Ecology*, 49, 234–241.

Wilczek, A.M., Burghardt, L.T., Cobb, A.R., Cooper, M.D., Welch, S.M. & Schmitt, J. (2010). Genetic and physiological bases for phenological responses to current and predicted climates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 3129–3147.

Wilczek, A.M., Cooper, M.D., Korves, T.M. & Schmitt, J. (2014). Lagging adaptation to warming climate in arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 7906–7913.

Wilczek, A.M., Roe, J.L., Knapp, M.C., Cooper, M.D., Lopez-Gallego, C., Martin, L.J., Muir, C.D., Sim, S., Walker, A., Anderson, J., Egan, J.F., Moyers, B.T., Petipas, R., Giakountis, A., Charbit, E., Coupland, G., Welch, S.M. & Schmitt, J. (2009). Effects of genetic perturbation on seasonal life history plasticity. *Science*, 323, 930–934.

Willis, C.G., Ruhfel, B.R., Primack, R.B., Miller-Rushing, A.J., Losos, J.B. & Davis, C.C. (2010). Favorable climate change response explains non-native species' success in Thoreau's woods. *PLoS ONE*, 5, e8878.

Wolkovich, E.M. & Cleland, E.E. (2011). The phenology of plant invasions: A community ecology perspective. *Frontiers in Ecology and the Environment*, 9, 287–294.

Wolkovich, E.M. & Cleland, E.E. (2014). Phenological niches and the future of invaded ecosystems with climate change. *AoB Plants*, 6, plu013.

Wolkovich, E.M., Cook, B.I., Allen, J.M., Crimmins, T.M., Betancourt, J.L., Travers, S.E., Pau, S., Regetz, J., Davies, T.J., Kraft, N.J.B., Ault, T.R., Bolmgren, K., Mazer, S.J., McCabe, G.J., McGill, B.J., Parmesan, C., Salamin, N., Schwartz, M.D. & Cleland, E.E. (2012). Warming experiments underpredict plant phenological responses to climate change. *Nature*, 485, 494–497.

Wolkovich, E.M., Davies, T.J., Schaefer, H., Cleland, E.E., Cook, B.I., Travers, S.E., Willis, C.G. & Davis, C. (2013). Temperature-dependent shifts in phenology contribute to the success of exotic species with climate change. *American Journal of Botany*, 100, 1407– 1421.

Yang, L.H. & Cenzer, M.L. (2020). Seasonal windows of opportunity in milkweed-monarch interactions. *Ecology*, 101.

Yousey, A.M., Chowdhury, P.R., Biddinger, N., Shaw, J.H., Jeyasingh, P.D. & Weider, L.J. (2018). Resurrected 'ancient' daphnia genotypes show reduced thermal stress tolerance compared to modern descendants. *Royal Society Open Science*, 5.

Zettlemoyer, M.A., Schultheis, E.H. & Lau, J.A. (2019). Phenology in a warming world: differences between native and nonnative plant species. *Ecology Letters*.

Zhu, J.T., Zhang, Y.J. & Wang, W.F. (2016). Interactions between warming and soil moisture increase overlap in reproductive phenology among species in an alpine meadow. *Biology Letters*, 12, 4.

Editor's comments:

Both reviewers and I really like the 'premise' of this article. Unfortunately, both reviewers were also quite critical of a number of aspects of the paper, including the background, the definitions,

caveats, and model itself. Given these concerns, I am sorry to say that I cannot support publication of this paper in Ecology Letters. I see two options:
line 28.

First, the authors could revise their manuscript as best as they can and seek publication elsewhere. Especially if they were to take on several of the reviewer comments, this might be a relatively straightforward task.

Second, if the authors feel that they can rather fundamentally alter the shape and structure of their manuscript, we might be willing to consider a reworked version. I should say, however, that given the nature of the reviews, and the detailed advice about the concerns and possible ways forward, that this would be a rather significant reworking bordering on a new submission. Such a revision would need to rework the model section, broaden the scope, and really tackle many of the caveats and issues brought up by the reviewers.

Of course, I completely understand if the authors choose the first pathway, as the second pathway would be a lot of work and there is no guarantee that it would satisfy the reviewers. Nevertheless, there is important potential for the authors and topic, and I wanted to leave the 'door open' should the authors be willing to take on this task.

We appreciate the editor's honest assessment of the state of manuscript and the task of a revision for *Ecology Letters.* We agree this is an important topic and the editor and referee's comments have led us to completely redraft the manuscript (we estimate that only 10-20% of the originally submitted text remains in the revised manuscript) with a broader focus. We believe the revised submission better serves the current state of this field and could help rapidly advance progress in research on evironmental tracking.

Referee 1 comments:

The authors present an interesting (Id call it perspective not review) manuscript that is focused on what they call "environmental tracking". With that, they really mean the ability of species to shift their phenologies in response to changing environmental conditions. This is clearly an important and timely topic, and I was excited to see someone tackling this. This said, the title & abstract did not prepare me for the content, and I felt a bit let down. The focus is much narrower than both suggest, and does not provide clear insights into a community context. The main content of the paper is focused on why some species track long term changes (e.g. in temperature) and others dont, with some speculation of how this might affect competition. The literature is not reviewed comprehensively, and largely focused on a few systems and big reviews, without really digging into the available literature. I know there isnt as much out there on this topic, but there is more than is given here. For a review, thats not enough, for a perspective, its still a bit short.

We understand the reviewer's concerns and have worked to completely re-draft the manuscript to provide a more substantial and useful review of the field, while still aiming to be forward-looking. We provide more details on these changes below.

The main part of the paper is a model, and that is where I had the most issues. Ill explain the details below, but it really appears to be a not well developed toy model (albeit I may have missed how exactly it worked, see below), and highly system, condition specific. As a consequence, much of what it shows we already know, or its unclear how we can generalize it to other systems and scenarios. There are no analytical solutions or comprehensive simulations & exploration of the parameter space. The authors also ignore much of the theory we already have on this type of model (it may not be called "phenology", but otherwise very similar). In the end, there are so many restrictions on the model that the outcome is known without any need of simulating the dynamics. There is even little to now discussion of existing models specifically on consequences on phenological shifts, which almost feels like intentional omission, but its not clear why. While highly relevant (some could predict similar outcomes) none of them are discussed or but in context to current models, so we dont really know whats new or different. As a consequence, I dont think that this is a good fit for Ecology Letters. I would suggest the authors either focus on a more comprehensive literature review and drop the model component, or really dig into developing the model and focusing on the exciting new questions that could be addressed with it, but that deserves its own paper. I know this is not an encouraging review, but I like the inherent idea and there is clearly a need for this topic to be emphasized, so Id like to see more of this.

We thank the reviewer for his candid assessment and agree the mix of the model with the overview of the field did not work. As such, we have removed the model from the main text and now review its results briefly in a Box ('Adding tracking and non-stationarity to a common coexistence model') with its description in the supplement. We have also tried to highlight where this model fits in a suite of models that couple provide inference on tracking in stationary and non-stationary systems in a new section, *Tracking in multi-species environments* (see line 257-line 357), which covers theory from plasticity, priority effects and coexistence models similar to the one we present.

Overall, I faced some major confusion with the model and really got stuck on many aspects of it. So let me go more in detail:
- Resource in this system is specified to not be renewed and only gets depleted. This is a reasonable assumption for some systems, but not for others, so it should be clarified and emphasized to avoid confusion. Importantly, this sets the system up for positive priority effect, i.e. resources are always at maximum at start of the reason, so early arriver will always have a benefit over later arrivers. Again, this is reasonable for certain systems, but not for others so it requires some more explanation and justification. It also prevents consumers from overshooting, i.e. there is not punishment for arriving too early (before the resource). Later on the authors confirm this expectation on early arrival advantage. It would be good to cite some literature on this (this is a common optimality problem and has been used in wide range of models). However,

there should also be some detailed discussion on what systems match these specific conditions, and which dont (e.g. systems where resources dont start at max but build up over time, systems where later arrivers have an advantage etc.)

The reviewer is correct that the resource does not renew each season, we now write (see line 578) "one dominant (non-renewing) pulse of a limiting resource each season," to help clarify this. As we have grossly cut the text devoted to this model we do not go into great details over this now, but we have added discussion about the need for more models that place a cost on early arrivals throughout the manuscript (see line 317, line 345, line 454).

- It took me a bit to think through the model formulation to understand how "timing" is incorporated here, and Im not sure Im still totally clear on it. Part of it stems from confusion about the two time scales, within vs between yeas and the notation was confusing to me which one is which. For instance, is g(t) the germination for year t, or for time t within a year? The latter would suggest that there is some sort of distribution of germination events within a given year, while the first would indicate a single event. From the wording, I assumed that it is indeed a single event per year. Furthermore, it appears that there is not difference in relative timing per se (say relative to the resource), but instead timing effects with a given year are solely driven by how many germinate in a given year, out of the total. So its not a question of "when", but "how many"

The reviewer is again generally correct here and highlights an important point we did not discuss in our original submission: the reality that most phenological events are a mix of both 'when' and 'how much.' Our model asbtracts the '’when' to focus more on the 'how much' and thus may be confusing to some readers. To address this we now discuss these intertwined issues (line 160-line 192) and return to them in our community modeling section (line 318-line 346) where we argue many of the current models to address these questions focus on only one or the other issue ('when' versus 'how much') and we highlight the need for more work combining these aspects, along with costs to mis-timed events.

- Overall, this confusion makes it hard to evaluate what the model does. If we stick with the one germination event a year, lets assume both species are identical for sake of argument. In that case, both species appear at the same time, but at different initial abundances, creating solely numerical priority effects.

- It also assumes that per-capita effects are unchanged, which is a specific assumption that is reasonable for some systems, but not many others. In addition, it would ignore the temporal dynamics, i.e. temporal overlap of competitors should be different, but without an explicit start time, its not. Again, all this is based on not having enough information to determine how the model really works, but based on the supplemental information I assume all populations start at same time within a given year just at different abundances.

All populations do start at the same time, which we have now clarified in our manuscript both in respect to the model through a new figure (see Fig. 4 and in general in respect to the diverse ways the literature currently studies environmental tracking (line 318-line 346).

This confusion is further increased by not providing information on how "non-stationary" is modeled in this system. If there is no real timing, does this mean its modeled as move from environmental to biological timing? So "shift" results in decrease in number of individuals if biological timing doesnt shift but environmental timing shifts earlier?

We agree this was not sufficiently clear in the original draft; in our current version we have moved a supplemental figure into the main text to show how environmental non-stationarity was modeled (see Fig. 3).

So to summarize I took away these following assumptions:
(1) Simulations of within season population dynamics start at the same time, there is no temporal offset of population dynamics, and no "escape" from competition in time. So there is no explicit temporal niche modelled

There is no explicit intra-annual temporal niche, but there is a temporal niche inter-annually, which is critical to our aim to model tracking. We have worked to clarify this in the current version. A model with both intra-annual and inter-annual temporal niche dynamics would be an excellent area for future work, but we believe it could be difficult to tackle this at the same time as adding tracking and non-stationarity to a single model. We now clarify starting on line 336 that this is an important area for future work.

S(2) temporal differences only affect starting densities not temporal dynamics or per capita effects. In other words, this is a model where phenological shift only affect reproduction (and thus numerical), not interaction effects.

Temporal dynamics affect species interactions through density. Interactions terms are not explicitly altered. We have attempted to clarify this by better contrasting models of the type we presented with those where species parameters are directly tied to the environment Rudolf (2019), see line 302-line 305, which read:

> Building a changing environment into such models thus requires knowing how environmental shifts filter through to species-level parameters (Tuljapurkar *et al.*, 2009). For example, Rudolf (2019) added the temporal environment to competition models by defining interaction strength as dependent on the temporal distance between species. In other models, the environment is more specifically defined.

S(3) There is always an early arriver advantage: which ever is closest to environment timing, has higher proportion of seeds emerging and will win (assuming all else equal)

Yes, and we now stress that this is a need across many models related to environmental tracking (line 318-line 346, and Fig. 4).

S(4) Season ends when resource is depleted to lowest R^*. So season is not ended by environmental conditions but resource availability, and just a function of competitor densities.

Each season ends when the R^* value of the better resource competitor is met, which is generally driven by species but can also occur due to abiotic loss in some seasons. As this model is no longer a focus of the paper, we have not addressed these details in the main text.

S(5) Germination function with difference in environmental vs. biological timing is non- linear.

Yes. We have aimed to clarify this in the supplement, but as this model is no longer a focus of the paper, we have not addressed these details in the main text.

S(6) Tracking parameter is difference between fixed vs. moving biological timing. Given this set of assumption, the generality of the model is strongly limited to a few systems/scenarios (very specific plant system), limiting the general inference that can be obtained from it.

The addition of the tracking parameter (which can vary from 0 to 1, which is from low to 'perfect' tracking) shifts how well a species matches to its environment each year (where a high match yields to more offspring). This is a modification of a common coexistence model, used by Chesson widely (Chesson, 2000; Chesson *et al.*, 2004), which is commonly applied to plant systems but extends in the simple form we use here to coral reef fish, forest trees and many other organisms (see Chesson & Huntly, 1997) so we do not believe its inference is limited.

We have worked to further highlight where this model fits within the current literature, which measures a mix of tracking of climate data to fundamental tracking (often studied in the trophic mismatch literature). We hope the changes throughout the manuscript relating to this will help clarify the extent to which this, and many other models, may apply.

Overall, I gained very little from the model, and as far as I can tell nothing new emerged from the model that we did not already know from other systems (e.g. much of this reminds me of stage-specific multi-parasitoid competition systems where the life stage of the host resembles the environment here, or a simple inter-annual model where reproduction varies across years, and this may or may not be correlated across species). In addition, given the specific conditions, the conclusion that non-stationary environments will change coexistence outcome has to be true given the model formulation, and could be easily inferred from recent Rudolf 2019 model (which shows how phenological shifts alter coexistence conditions). As far as I can tell, the only novel aspect here is the tracking aspect, which I quite liked. But, the way its implemented, its not dynamic but forced on the system and simply shifts the initial relative numerical abundance of *species, so again the outcome could be inferred from a simple L-V type competition model. So Im still struggling to understand why this model is necessary and what new insights we gained. Otherwise it just adds confusion, so maybe it would be best shown in a simple verbal or graphical model. In fact, I would strongly favor the graphical option, since that would be clearer, and outcome can easily be predicted without simulation the system from the many existing models we already have. I still think the tracking approach is very interesting, but hasnt been fully developed to ask more detailed question on how tracking will affect long-term dynamics, and rigorously explores when and how it influences long-term dynamics. I think this deserve its own fully developed manuscript, and sticking it in here is really selling it short of its potential. This would also allow the authors to examine how many of the unresolved questions they list later on influence the outcome, e.g. what are consequences of tracking if changes in environmental conditions alter multiple aspects (e.g germination & per-capita effect) etc.*

We appreciate the reviewer's concerns about how useful the model is in this paper and have thus moved it to a box where we focus mainly on its outcomes via figures (after much discussion we have decided to keep the full model description in the supplement as we cannot derive its findings from a verbal or graphical model).

Outside of the model, I generally liked the idea of getting a much better understanding of what species track environments, and which ones dont. I completely agree with the authors that we know way to little about this, and more research needs to be done. This said, the manuscript here did not feel like a review, but a "food for thought" short opinion paper. If this is truly supposed to be a review, I would expect a more thorough and quantitative analyses of the literature, since much literature was missed, and largely restricted to plant systems. So my main complaint here would be that it felt like it was just touching the surface and did not provide enough depth (i.e. go into exiting studies).

We agree that in focusing on the model, we had little room for a more thorough review of the literature. Our revised manuscript draws on literature from vertebrates, corals, plants, arthropods and more to provide a fuller sweep of the literature. In our focus on what may drive variation in tracking, we provide a literature review focused on understanding tracking within a syndrome or traits (see line 263-line 278 and Box 'Trait trade-offs with tracking').

While we understand the desire for a more quantitative review, many have recently tried this and ended up focusing more on methodological issues than ecological predictors (Brown *et al.*, 2016; Kharouba *et al.*, 2018). Indeed, the first author of this manuscript designed the statistical approaches in Kharouba *et al.* (2018) and knows first-hand how difficult it is to accuratel measure 'tracking' across studies currently. We feel the critical needs for this field now are (1) a greater use and development of theory to provide testable predictions and (2) better definitions and guidance on how to define and measure tracking so quantitative reviews will be possible in the future, and have written a manuscript that we see as most useful to the field now, given these needs.

Finally, there was very little coverage over theory on phenological shift. The authors mention Rudolf 2019 in passing, without discussing any similarities, differences that are clearly there. Similarly, they never mention other phenology models, like Nakazawa & Doi, 2012, Revilla et al 2014 etc. Even the simple graphical temporal niche approach that the authors introduced themselves (Wolkowich & Cleland 2011) is not discussed (but brings up interesting question about "single" vs multiple resources approaches).

This is an excellent point. We now have worked to better frame were Rudolf (2019) fits within many ways of introducing phenology into current multi-species models (see line 302-lrS2end) and we have worked to build in more references to the trophic mismatch literature, where is what Nakazawa & Doi, 2012, Revilla et al. 2014 focus on (e.g., line 65, line 178).

Specific comments:
P 3 L30ff: the notion that earlier spring should favor earlier phenologies relies on the assumption that the "niche" is empty i.e. no other species are earlier. So this is applies to very specific systems (i.e. resources are not available before that time point, so temporary resources) and should be clarified.

We have adjusted this text to now read (starting on line 31):

> Considering tracking as a form of plasticity, evolutionary models predict species that track will be favored in novel environmental conditions. Similarly, some models of community assembly suggest that a warming climate should open up new temporal niche space and favor species that can exploit that space (Gotelli & Graves, 1996; Wolkovich & Cleland, 2011; Zettlemoyer *et al.*, 2019).

And we provide a longer discussion of alternative models throughout the manuscript now, especially in the new section, *Tracking in multi-species environments.*

L34-35, there are some studies (and should be cited here), e.g. Block et al 2019 Oikos. Showing that phenological plasticity is a poor predictor of performance.

In the current version of the manuscript this sentence no longer exists, but we have added a reference to the study in the same paragraph (starting on line 34), "Yet not all studies find the purported link (e.g., Block *et al.*, 2019), and there has been comparatively little work to improve predictions by formally connecting tracking to foundational ecological theory."

L 50ff: there has been progress, e.g. Rudolf 2019 specifically incorporates non-stationary systems and variability to examine how it influences coexistence and communities (since its focused on phenology it seems like a highly relevant citation here). In fact this citation would be great to support the claim that it matters, instead of simply stating that nobody looked at it (which is *incorrect).*

We have added this citation where requested.

Equaton 9: "n" is undefined. Along the same line, what determines the end of a growing season?

We apologize for these omissions, n is the number of species, and each season ends when the R^* value of the better resource competitor is met. As this model is no longer a focus of the paper, we have not addressed these details in the main text but they continue to be provided in the supplement.

P13 L 9: this prediction hinges on the assumption of early arriver advantage and single resource competition etc. So as it stands, this is one of the predictions, not the only one.

Agreed, these predictions fail when there are costs to tracking too closely, we now discuss theses costs in detail in multiple places in the modeling section of the manuscript (see line 317, line 345, line 454).

P15L10ff: what about species that are just very plastic, i.e. can adjust to cope with various environmental conditions, and thus take an alternative strategy to shifting. There has been increasing discussion of phenotypic plasticity vs lets call it "environmental" plasticity, i.e. species that can perform equally well at cold and warm temperatures. So environmental generalists.

We now discuss plasticity in depth on line 257-line 289.

Same page, next paragraph (sorry, having not continuous line numbers across pages makes this a bit frustrating). Good examples here would be species where phenologies are correlated across season/life stages. In some cases, phenologies in spring are determined by what happens in fall, or what happens later in summer may depend on how individuals perform during earlier life stages in spring (e.g. changed developmental rates alter later phenologies etc.) In same context, Yang & Cencer 2019 Ecology examine "seasonal windows of opportunity", which fits nicely in the context here. They took rigorous approach in finding what constraints those windows, which would also determine how shifts in them would change the optimal window.

We thank the reviewer for these examples, we now Yang & Cenzer (2020) on line 397 and here and throughout the manuscript have worked to discuss in-depth the contraints on tracking (e.g., line 194-line 231 and line 467).

Our apologies about the line numbers, these are the line numbers provided by *Ecology Letters*; we now provide continuous line numbers and refer to those here (though this may mean we end up with two contrasting sets of line numbers, which we apologize for).

P16 L 34-35: very cool, Ill have to look up change-point and hinge models, never heard of them!

P 16 L38ff: some recent approaches suggest using whole phenological distributions can strongly increase power as well (e.g. single species: Steer et al 2019 Methods E&E, or for species interactions Carter et al 2018 Ecol Letters)

Good point! We have added these references to lines line 112 and we now also discuss the issue of these events as a distribution on line 160-line 173.

Referee 2 comments:

1. Need clearer motivation in the introduction (section 1, "main text").
a. What is the specific definition of tracking applied here? Tracking a set of abiotic conditions? Does it extend to tracking biotic conditions? Is there a way to quantify the relevant set of conditions, and therefore an organisms ability to track them?

This is an excellent point and we now provide an extended definition of environmental tracking (see section, 'Defining environmental tracking,' line 59-lrdefine1end) and a new figure (Fig 1) to help highlight the complexity of defining this. We believe this is an important addition to the literature, where 'tracking' is often used but rarely defined.

b. The second paragraph of the introduction suggests that "a shift toward earlier spring should favor earlier species, especially those that can environmentally track ever-earlier seasons" Im not sure I follow the logic here; it seems like earlier spring conditions could just as easily limit the success of early spring species in particular. Its not that the proposed hypothesis is never true, but it also doesnt seem that it is likely to be necessarily or generally true, at least based on the argument presented. Is this intended as a straw hypothesis?

We have restructed the introduction to try to address this concern. We do not mean for it to be a straw hypothesis, but instead one that has gained traction in the literature and has some basic support from simple models, as we now state.

c. It seems like the assumption of stationarity has never been true, and ecological theory has always been a bit uneasy about this. Though maybe because so much of ecological theory is generally explanatory rather than specifically predictive, these deviations havent been too troublesome. Perhaps the question then is more about how much worse the situation is with rapid climate change.

We agree with the author's point. Most systems can appear either stationary or non-stationary depending on the scale and temporal period. We have tried to clarify this on line 516-line 520 , where we state:

> Understanding the impacts of climate change further requires recognizing that many systems can be considered stationary or non-stationary depending on the timescale and period of study. Thus, predicting the consequences of current non-stationarity in ecological systems benefits from identifying the type and scale of non-stationarity, relative to long-term trends.

2. Environmental variability and change (1.1)
a. L21-24: Is there good evidence of historical stationarity? The distinction between stationary vs. non-stationary environments seems to be scale-dependent, and thus somewhat subjective. Is that a problem?

We believe there is good evidence for climatic stationarity, at least on ecologically relevant (and certainly researcher-relevant) timescales, and have adjusted text on lines line 516-line 520 and line 469-lrnsfutend to clarify this.

3. Environmental tracking in time (1.2) a. Chmura et al (2019) suggest that relatively little is actually known about the mechanistic/cueing bases of differences in phenological shifts, either because most studies dont consider cues per se, or because they very rarely assess alternative mechanisms. If this is true, how does this affect the framework described in this section?

This is a great point and we believe an area where this paper can offer some guidance. In our overhaul of the paper we now address this two new sections *Defining environmental tracking* and *Measuring environmental tracking* on line 59-line 152.

b. The trade-off between plasticity ("tracking") and bet-hedging has been examined in studies by Chevin, Lande, Ghalambor and others. Do those studies provide a useful perspective here?

Yes, we now review the plasicity literature on lines line 257-line 289, including references to these studies. For example, lrplasticquotestart-line 255, we write:

> Another area of life-history theory, that focused on plasticity, may be primed to provide insights on non-stationarity (or 'sustained environmental change,' see Chevin *et al.*, 2010). Considering phenology as a trait (as we and others do, e.g., Charmantier *et al.*, 2008; Nicotra *et al.*, 2010; Inouye *et al.*, 2019), environmental tracking is one type of plasticity. Researchers could thus more broadly understand environmental tracking through modeling an organism's reaction norms (Pigliucci, 1998; Chmura *et al.*, 2019) and understanding how cues and suites of cues—across environments—determine how fundamentally plastic an organism may be in its

tracking. For example, multivariate cues should yield higher plasticity in this framework. From here, models of the role of plasticity in novel environments provide an important bridge to understanding the outcomes of non-stationarity, generally predicting non-stationarity should favor highly plastic species. This outcome, however, assumes there are no costs related to plasticity (Ghalambor *et al.*, 2007; Tufto, 2015).

c. The long-term value of plasticity vs. bet-hedging may not be apparent in relatively short field studies, since the relevant measure of fitness could require more time to assess.

Agreed, we now discuss bet-hedging in more detail on lines line 214-lrbh1end, and state, "Assessing bet-hedging in many systems, however, requires tsudies of fitness over longer timescales than many current field experiments." See also line 552.

4. Interspecific variation in tracking (1.3)
a. Im a little concerned about the slant of this first sentence, which seems to suggest that tracking is both universally important and positive. Modeling studies seem to suggest that under some circumstances, more plastic responses could be maladaptive. The section goes on to identify some very interesting potential trade-offs with competitive ability, but the broader point is that it doesnt seem to be entirely clear that "tracking" per se is universally favored even absent a competition trade-off. Perhaps this goes back to our limited ability to quantify "tracking" ability, and the implicit assumption that we can assess an organism's ability to find optimal conditions. In most systems, it seems like we dont have enough data to quantify tracking ability. In the absence of this, we can assess plasticity to specific cues, but whereas tracking may implicitly imply adaptive plasticity, plasticity is not always adaptive.

Agreed, the opening sentence now reads (starting on line 95), "Attempting to measure environmental tracking and compare variation in it across species, space and time is a rapidly growing area of ecological research (Cook *et al.*, 2012; Fu *et al.*, 2015; Thackeray *et al.*, 2016; Cohen *et al.*, 2018, e.g.,)." Further, we have overhauled the manuscript and now address many of these concerns throughout the new sections *Defining environmental tracking*, *Measuring environmental tracking* and *Understanding variation in environmental tracking* from line 59 to line 255.

b. L55-56. Because many climatic cues are correlated, and also correlated with other cues (photoperiod, biotic, etc), the observation that temperature models can explain more than 90% of variation in phenology probably shouldnt be assumed as evidence of causation. Temperature in particular can be a very complex cue, and the determination of mechanistic causation is difficult, as described by Chmura et al (2019).

Agreed, we have deleted this note and have worked to stress the complexity of potential cues throughout the manuscript, including mutiple references to Chmura *et al.* (2019) and related work.

5. Model description and simulations (1.4.1)
a. This model conceptualizes "tracking" ability as a variable between 0 and 1 which describes an organisms ability to adjust its biological start time to the (optimal?) environmental start time in a given year. This leaves aside some messy but potentially interesting issues of mechanism and constraint, including any explicit consideration of cues or environmental conditions. Im not sure how I feel about this approach. This could be an effective way to focus on the issue of "tracking" per se, but also risks being too far abstracted from reality to provide a meaningfully realistic model. For example, how should we conceptualize "tracking" ability if the optimal start time becomes worse over time? Or if there is a disconnect between cues and conditions (i.e., an optimal tracking of cues leads to a poor tracking of conditions)?

This was a concern of both reviewers and it highlighted for us the complexity in modeling phenology. This model definitely is a step removed from costs of tracking and we have worked to highlight this. Our new section on *Tracking in multi-species environments* reviews relevant modeling from the plasticity literature, priority effects and our modelling approach, among others. We hope it provides a much more useful and broader view of the challenges in modeling 'tracking' and the broad relevant literature for this issue.

6. Tracking in stationary environments (1.4.2) a. If Im understanding this model correctly, there is the assumption of some kind of intrinsic circannual rhythm (represented by the fixed biological start time) which is then modified by cues (abstractly represented as "tracking") to yield an effective or realized start time. This seems different than my understanding of circannual rhythms and zietgebers in a potentially important way, where the current model would assume that even in the absence of any cues (or with a tracking ability of 0), an organism would consistently start on the same calendar day each year. This seems like a modeling decision that should be explained and justified. Are there studies to indicate that this is a reasonable model?

This model is a form of one commonly applied to plant systems but extends in the simple form we use here to coral reef fish, forest trees and many other organisms (see Chesson & Huntly, 1997). Part of why this model can be applied broadly is in its abstraction, which is also why it may be difficult to link neatly to phenological events. The addition of the tracking parameter (which can vary from 0 to 1, which is from low to 'perfect' tracking) shifts how well a species matches to its environment each year (where a high match yields to more offspring).

Given the reviewers' concerns regarding the model we have now moved it to a box ('Adding tracking and non-stationarity to a common coexistence model'), which precludes an in-depth discussion of its exact potential conceptualizations. We have, however, worked to stress the varied interpretations of 'tracking' throughout the text (e.g., line 59-line 93), including across different modeling approaches (line 291-line 357).

7. Tracking in non-stationary environments (1.4.3) a. I get that this is not intended to be a realistic climate change scenario, but wasnt able to understand the details of how the non-

stationary environment was created without the SI. The key thing that seems clear is that the non-stationary environment favored earlier start times. It wasnt clear if the optimal start time was actually advancing gradually over time, or if it was just changed in a single step. If I understand it correctly, this model doesnt allow for any evolutionary responses.

This was also a concern of both reviewers and, given that part of our aim was to show a model transitioning from stationarity to non-stationarity, we are sorry we failed at this. We have now moved the relevant Figure from the supplement to the text.

8. Model conclusions (1.4.4) a. The observation that tracking is favored seems to be almost an assumption of the model, rather than a conclusion. Could it be otherwise in this model?

Tracking in this model is favored in the same way that a lower R^* is favored in the model. We have worked to clarify this as much as possible, while still keeping the text related to this model within the limits for a Box. We have worked to focus more of the main text on contrasting models and additional approaches (line 291-line 357).

b. The idea that "tracking" should be considered as a part of larger "trait syndrome" seems appealing, though Im not entirely sure what it means. What are the other parts of this syndrome? My concern is that the idea of "tracking" ability per se is not sufficiently defined or justified to develop in this way, abstracted from cues and physiological mechanism.

This is a great point and we have re-drafted the manuscript to dig in deeper on defining and measuring tracking and we return to that definition (and its often multivariate scope) in the new section on *Tracking in multi-species environments* (see line 257-line 357).

c. Despite this, I actually like the idea of a trade-off between tracking ability and competitive ability; it seems intuitively appealing, if not clearly defined. I think Id like some additional justification that the idea of "tracking ability" is a meaningful one in nature, and that there are empirical reasons (not just based on theory) to think that it trades off with competitive ability. As a counterpoint, it seems like phenological traits are just as likely to be used as a tool in competition, where an organism may benefit by showing an earlier phenology in the presence of competitors (due to pre-emption, or asymmetric competition, e.g. for light), even when it would do better to have a later start in the absence of competitors. This requires a more careful definition of "tracking" is an organisms that deviates from its optimal timing in an abiotic-only context showing good tracking or poor tracking? What if a deviation from the abiotic optimum is favored under competition? What if competitive ability depends on the relatively phenological/ontogenetic stages of the competitors? Instead of thinking of ways in which phenological tracking and competitive ability trade-off, Im left wondering more about the complex ways in which they could interact.

We agree and appreciate the reviewer pushing us to better define tracking. Our manuscript is in many ways re-written around this aim with new section on defining and measuring tracking. We provide some examples of empirical trade-offs (line 263-line 278, which read:

> As tracking often relates to the timing of a resource pulse, traits related to resource acquisition are likely contenders for a trade-off. Species with traits that make them poor resource competitors may need to track the environment closely to take advantage of transient periods of available resources, but will risk tissue loss to harsh environmental conditions more prevalent early in the season (e.g., frost or snow). In contrast, species with traits that make them superior resource competitors may perform well even if they track environments less closely, because their resource acquisition is not strongly constrained by competitors. Examples include under-canopy species leafing out earlier to gain access to light (Heberling *et al.*, 2019) or species with shallow roots starting growth sooner in an alpine meadow system, while species with deeper roots begin growth later (Zhu *et al.*, 2016). In such cases, tracking is akin to a competition-colonization trade-off (Amarasekare, 2003), where species that track well gain priority access to resources and, thus, may co-exist with superior competitors. Research to date supports this, with several studies linking higher tracking to traits associated with being poor competitors (Dorji *et al.*, 2013; Lasky *et al.*, 2016; Zhu *et al.*, 2016). Further, many studies have found a correlation between higher tracking and 'earlyness' each season, which has been linked to resource acquisition traits associated with lower competitive abilities (Wolkovich & Cleland, 2014, see Box 'Trait trade-offs with tracking').

As mentioned, we also provide further detail in the Box 'Trait trade-offs with tracking,' however, we feel much more is needed here and thus focus on it further in our future directions section (line 430-line 443).

9. Future research (1.4.5) a. While I agree that improved predictions of climate change would be valuable, it isnt clear how these improved (i.e., more complex) climate predictions would benefit this model in particular. This model already seems quite far abstracted from cues and mechanism. More generally, I actually get the feeling that climatic projections are constantly improving though improved climatological models (especially better local or regional scale models), but our ability to predict ecological outcomes (coexistence or otherwise) is not typically limited by the detail, complexity or resolution of these climatological projections.

We agree. We have placed our modeling results in a box to focus on these bigger issues and we now address the climate projections as needing to focus more on how climate change will impact how we measure tracking, see line 390-line 417 of the sub-section *Understanding and measuring 'tracking'*, which includes:

> Understanding how the environment is changing represents just one step along the towards robust measures of environmental tracking. Shifting environmental regimes

must then be filtered through species cues. As robust estimates of these cues are currently available for few if any species, we suggest several major improvements on current methods to help interpret current trends and make comparisons more feasible.

Studies should clarify their definition of tracking, how the environment is defined and how well, or not, the underlying cue system is understood for study species. Currently, many studies examine fundamental and environmental tracking at once (e.g., Yang & Cenzer, 2020), which is clearly helpful in advancing the field. However, the more researchers can clarify when and how they are addressing fundamental tracking versus environmental tracking the more easily we can compare across studies. Next, and relatedly, studies should define their environment: are they considering primarily the abiotic environment or measuring an environment fundamentally shaped by other species? This difference connects to fundamental versus realized niches and whether systems are primarily top-down (resources and the environment may be strongly shaped by other species) or bottom-up controlled. Finally, all researchers working on environmental tracking need to embrace their inner-physiologist, or collaborate with one. For many organisms there is often a related (perhaps sometimes distantly) species that has been studied for which cues underlie the timing of the life history event. Researchers should draw on this literature to bracket which environmental variables may represent true tracking and which may be proxies, and to highlight uncertainty. We expect progress will come from balance between measures of fundamental tracking, estimating an organism's system of cues and measuring environmental tracking. Clear statements of what is and is not known and measured will help.

Embrace the sometimes contradictory pulls of conducting experiments to identify mechanistic cues and the multivariate climate of the real world. Clearly, we need more experiments to identify which specific aspects of the environment different species cue to and how these cues are filtered by their actual environmental regime (as outlined above). Suites of experiments that build from identifying cues, to understanding how they act when correlated are a major gap for most organisms.

As an aside, while climate models have accurately predicted general trends and anomalies (Hausfather *et al.*, 2020), they have had limited success in predicting many extremes. Further, their reliance on model 'tuning' has made it difficult to understand mechanistically what underlies important divergences between models (Knutti *et al.*, 2017). So they are amazing, but not all would agree they are always improving.

b. I would also be interested to know more about potential trade-offs between "tracking" and other traits, but would want to know first whether "tracking" ability is a meaningful construct. In this model, tracking is mathematically defined as inversely correlated with the difference between the intrinsic timing and the (optimal?) "environmental (abiotic)" start time. My sense *is that there are very few systems where either the intrinsic start time is a realistic concept, or where the (optimal?) environmental (abiotic) start time has been well- characterized. If there are good examples of systems that support these concepts, they should be described. If this model is intended to provide more of any abstract framework, I would suggest that these caveats of definition and characterization should be much more prominent, and assessing these issues would probably be valuable future directions.*

We agree this model is an abstraction (as all are) and we can see it was perhaps not the most useful one here, thus we have moved the model to a box and focused more of the text on defining, measuring and building depths across multiple areas of community ecology theory to better understand tracking.

c. Despite my concerns about the framework of this paper, I do think the question of how climate change will shape coexistence mechanisms is an interesting one. Im not entirely convinced that this model sheds much light on this issue, but would be glad to be convinced otherwise.

We thank the reviewer for their comments, which helped us re-envision this paper. We hope the new version addresses some of the concerns and will provide a path forward for this field of research.

10. Boxes
a. The three boxes in this manuscript touch upon some of the issues that concern me about this manuscript, albeit too briefly. It seems clear that the authors have thought about some of these issues. Why not examine some of these complexities more centrally in this manuscript?

Agreed. We have moved some of the text from the Box 'What underlies variability in species tracking?' into the main text in the new sections *Defining environmental tracking*, *Measuring environmental tracking* and *Understanding variation in environmental tracking.*