Version dated: June 30, 2021

Running head: Tracking & climate change

How phenological tracking shapes species and communities in non-stationary environments

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] *Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kān'eohe, 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, produced the figures, performed modeling work and edited the paper, EMW wrote the paper and did the literature review, while MJD wrote the supplementary information on the model.

Keywords: community assembly, global change, climate change, phenology, environmental variability

Article information: Abstract: 237 words; Main text: 7880; Tables: 1; Figures: 6; 128 references

Abstract

Climate change alters the environments of all species. Predicting species responses requires understanding how species track environmental change, and how such tracking shapes communities. Growing empirical evidence suggests that how species track phenologically—how an organism shifts the timing of major biological events in response to the environment—is linked to species performance and community structure. Such research tantalizingly suggests a potential framework to predict the winners and losers of climate change, and the future communities we can expect. But developing this framework requires far greater efforts to ground empirical studies of phenological tracking in relevant ecological theory.

Here we review the concept of phenological tracking in empirical studies and through the lens of coexistence theory to show why a community-level perspective is critical to accurate predictions with climate change. While much current theory for tracking ignores the importance of a multi-species context, basic community assembly theory predicts competition will drive variation in tracking and trade-offs with other traits. We highlight how existing community assembly theory can help understand tracking in stationary and non-stationary systems. But major advances in predicting the species- and community-level consequences of climate change will require advances in theoretical and empirical studies. We outline a path forward built on greater efforts to integrate priority effects into modern coexistence theory, improved empirical estimates of multivariate environmental change, and clearly defined estimates of phenological tracking and its underlying environmental cues.

Contents

1 Introduction

Anthropogenic climate change is causing widespread changes in species distributions, with many species shifting in both space and time (??). Species are moving to higher elevations and poleward (?), shifting the timing of recurring life history events (phenology) earlier (??), or both as climate warms (??). These general trends, however, hide high variability across species and locations. A large proportion of species are not shifting (??), 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 by shifting their phenology to changes in biomass, growth and other metrics related to performance (?). Tracking may then be a major component to understanding and predicting the fitness consequences of climate change, with cascading effects on community and ecosystem structure (??).

The hypothesis that tracking improves fitness outcomes with climate change has gained significant traction in the ecological literature (e.g., ?). Simple conceptual models suggest that a warming climate should open up new temporal niche space and favor species that can exploit that space (???). Beyond this, however, there has been little work connecting tracking to community assembly theory. Yet theory shows temporal sequencing and environmental variability can alter the relative fitness and niche differences between species that determine coexistence—suggesting important ecological constraints to tracking.

This disconnect could be because most ecological theory was constructed for stationary systems (e.g., ?). While major arenas of research such as 'modern coexistence theory' or population ecology now embrace environmental stochasticity, they generally still assume stationarity, where the underlying distribution of the environment is unchanged across time (i.e., constant mean and variance, ?).

Climate change upends the assumption of stationarity. By causing increases in temperature, larger pulses of precipitation, shifts in snowpack, increased drought, and more storms (?), climate change has fundamentally shifted major attributes of the environment from stationary to non-stationary regimes (Fig. 1). This transition is reshaping ecological systems. New work has aimed to adapt coexistence theory to non-stationary environments (??). Yet there is still little theoretical work on what such a transition may mean for communities and the species within them, including how processes that shape communities, such as competition and priority effects, might feedback to modify species responses.

Here, we provide a pathway to unify empirical studies of phenological tracking with community ecology theory. We begin by providing the necessary definitions to link empirical estimates to theory: specifically we distinguish between measuring tracking and evaluating its fitness outcome. Next, we provide a brief review of current estimates of tracking, and basic theory that predicts variation in tracking across species and environments in stationary systems. We then examine how well community assembly theory—especially priority effects and modern coexistence theory—can be extended to predict the community consequences of climate change. Our review highlights that we are unlikely to fully understand, and thus predict, phenology without a greater integration of community assembly theory. To this end, we close by reviewing the major hurdles to linking empirical estimates of phenological tracking and new ecological theory in the future.

2 Defining & measuring phenological tracking

Understanding phenological tracking requires defining both phenological events and tracking itself. For our review, this means defining them precisely enough to model using empirical data, and in analytical and simulation studies of community assembly. Below we provide a review of key concepts from empirical phenology studies and life history theory, and provide definitions to bridge this existing literature on tracking to community assembly theory (see Table 1 for a glossary). For generality, we provide examples from a range of organisms and habitats, with a focus on birds and plants that reflects their greater representation in climate change research in phenology (**?**).

2.1 Phenological events

In empirical studies of climate change, phenological events are often coded as on/off switches (e.g., a seed does/does not germinate; a coral does/does not spawn), yet these events are almost always defined by investment decisions that are part of a continuous developmental process (**??**). This is a critical distinction to bridge from empirical work to understanding the life-history and community-level (e.g., niche differences) forces that may shape phenological tracking, and, in turn, how it may structure communities with climate change.

We consider phenological events as the outcome of a two-part process that is repeatedly observed over time. At each temporal unit, an event can either happen or not (when—part 1) and, if it happens, the event can vary in size or degree of investment (how much—part 2). This process is generally applied at the level of the individual (but it could apply at lower levels, for example buds on a branch, or higher levels, such as a recruitment event for a population). Across time, it produces an event's distribution (**??**). After starting, many events are entrained to continue: 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. These individual-level distributions scale up to the population-level estimates of these events generally used by researchers (see **?**, for discussion of the outcomes of this scaling).

2.2 Defining phenological tracking

Tracking is commonly used to describe how phenology responds to climate change, yet it is rarely defined (e.g., **????**). Conceptual and theoretical studies often conceptualize tracking as 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.' In contrast, empirical studies of tracking often focus on the change in the timing of an event relative to a measured environmental variable, with the aim of measuring what we refer to as 'environmental tracking' (Fig. 2-3)—the phenological change due to an organism's cue system given change in the environment (though most studies lack the required knowledge of the underlying cue system, ?).

Fundamental tracking rests on an assumption that there is an ideal timing that yields maximum fitness, with fitness declining as event timings move away from this ideal (a foundational concept of the trophic mismatch literature, ?). This 'ideal timing,' however, is generally only clear in simplified models or in retrospect; thus, most species use environmental cues to try to predict ideal phenological timings over time and space (Fig. 2-3). Each organism's set of cues forms the biological basis for how a species tracks the environment.

An organism's cues combined with the environment's variability determine what we refer to as 'environmental tracking' (Table 1, Fig. 3). While fundamental tracking forms the focus of most conceptual and theoretical treatments of phenological tracking, the majority of empirical studies focus on estimating environmental tracking.

Our definition of environmental tracking highlights the difficulty of measuring it. If the varying components of the environment are not in the organism's set of cues, then the organism does not 'track' per this definition (although covariation with other environmental components might give the appearance of tracking). Which aspect(s) of the environment are changing and which aspects researchers measure will determine estimates of environmental tracking (Fig. 3). If researchers know the exact cue or suite of cues and can perfectly measure these in an environment where the cue(s) varies, then an organism will track the environment near perfectly (e.g., the photo-thermal model of flowering of *Arabidopsisis thaliana*, ?). 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). Aside from a few model systems (e.g., ??), most studies lack the required knowledge of the underlying cue system (?). This makes it difficult to evaluate the accuracy of most current estimates of tracking.

2.3 Measuring phenological tracking

Measuring 'tracking' and comparing variation in it across species, space, and time is a rapidly growing area of ecological research (e.g., ????). Studies that directly quantify fundamental tracking are uncommon (but see ??), given in part the difficulty of estimating fitness, though many studies in the synchrony literature attempt to link consumer change to resource change, with an assumption that the measured resource determines the ideal timing for the consumer (though this may rarely be true, see ???). Instead, most studies focus on estimates closer to environmental tracking. Some studies estimate simply change in days over time (e.g., ??), though most studies now estimate shifts as responses per unit temperature (???) or precipitation (??).

All species-rich studies of phenology-climate relationships find high variation (??), including some species that do not track or track poorly. Researchers have worked to link such variation to the underlying cues (e.g., ?), species traits (e.g., ?) and trophic level (e.g., ?). These approaches hint at several major explanations for why some species do not appear to track climate or appear to track poorly: environmental tracking is either not possible or optimal (discussed below in 'Tracking in single-species environments' and see ?), researchers have measured an environmental variable that species do not track (?), and statistical artifacts that make it difficult to measure tracking robustly (discussed below in 'Robust comparable measures of phenological tracking').

These confounding factors may make many current estimates of interspecific variation in tracking less reliable than they appear. This in turn makes robust quantitative analyses across species difficult (??), yielding a muddy picture of which species, when, and where, do and do not track. Given this difficulty, clear testable predictions from ecological theory would be especially valuable in guiding the field forward (?).

3 Tracking in single-species environments

Community assembly theory provides a major paradigm to predict and understand variation in phenological tracking. Like most of ecology it builds upon theory from evolutionary biology of when and where tracking should evolve. Thus, before discussing models of community assembly, we briefly review foundational evolutionary theory for single-species systems (where most work has focused, but see ???).

3.1 Predicting variation in environmental tracking in stationary systems

Evolutionary models predict selection for tracking in heterogeneous environments where there are cues for the ideal timing of events (??) and the underlying genetics to develop a heritable cue system (tracking is likely strongly heritable, given that many cue systems are themselves heritable, e.g., ??). The predictability of the environment via relevant cues that an organism can monitor is particularly critical for irreversible plastic traits, which includes many phenological traits, and must exist at an appropriate timescale for an organism to monitor and respond to.

The strength of selection is then determined by the costs and benefits of cues (?). The costs include the machinery an organism uses to monitor its environment (e.g., accumulated temperature or daylength), while the benefits are the increases in fitness gained from better timing (e.g., how much tissue is saved by avoiding a coldsnap). Adaptation, however, can be lower than expected from reaction norms predicted by simple evolutionary models for many reasons, including trade-offs with tracking (??), gene flow from other environments that push a population away from its local optimum (?), limits due to standing genetic variation (??), or deeper evolutionary history that may produce co-evolved traits making it difficult for selection to act solely on tracking (?).

Apparently unreliable cues may occur for organisms in environments where there is both a low cost and low benefit to the cue(s). Whereas expensive cues, such as complex multivariate ones, are possible given a high pay-off. Most in-depth empirical studies of species' cue systems find evidence for complex multivariate systems that appear adapted to handle unusual—though not completely uncommon—years (?). This suggests that multivariate cues may better couple environmental tracking to fundamental tracking, while simple cues are more likely to trigger growth or reproduction at a suboptimal time. Such ideas are supported by models built upon the genetic architecture of phenological events (e.g., ?), which highlight the complexity of cues underlying even apparently simple events. This research has also highlighted how gene pathways may shape, and thus constrain, multiple phenological events. To predict what cues an organism should have, even in simple stationary systems, would require considering a suite of costs, benefits, and constraints (??). Not surprisingly, we lack this understanding for most organisms. General theory has developed, however, to try to predict which stationary environments do, or do not, favor tracking.

Tracking should generally not be favored where early season environment cannot be used to predict later season environment, or where species otherwise face high uncertainty in the timing of investment decisions (?). Instead theory suggests the optimal strategy may often be to bet-hedge (???) via a high diversity of timings or a conservative timing. Because bet-hedging, by definition, maximizes geometric-mean fitness in the long-run, its short-term outcomes can appear maladaptive. How often observed 'maladaptations,' which may easily include species that do not track or appear to track poorly, are actually the outcome of bet-hedging is difficult to estimate, as robustly assessing bet-hedging requires studies of fitness over longer timescales than many current field experiments (?). Environmental variation, however, is rarely simply predictable or not; it more often includes both predictable and less predictable aspects. In such cases theory predicts organisms may evolve tracking that is a mixed strategy between bet-hedging and plasticity (?).

Evolutionary theory, which integrates over the sometimes hidden costs and benefits of particular cue systems and considers environmental predictability, thus provides multiple reasons species may not track or track weakly. This suggests that—even in simple single-species stationary systems—we should expect a number of species that do not track.

3.2 Predicting variation in environmental tracking in non-stationary systems

A major open area of research is whether conclusions derived from evolutionary theory developed for stationary systems extend to non-stationary systems (?). In regards to phenological tracking a major question is whether tracking should be more or less favored in non-stationary environments.

One approach to this focuses on cue systems and makes predictions based on whether cue systems maintain their reliability in a changing environment; i.e., whether they consistently yield high fundamental tracking (?). Consider a simple case in which an organism's cues evolved based on a correlation between peak prey abundance and daylength: in a stationary environment the daylength cue may be fairly reliable, but would become unreliable, and lead to fitness declines, if warming continually advances peak prey abundance. Multivariate cues are often argued to be more reliable because they can capture multiple attributes of the environment (??), but they may be equally vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (?) and thus optimal fitness is no longer associated with the cue system. Under this framework, predicting whether tracking is more or less favored in non-stationary environments requires that researchers know: (1) the full cue system of an organism, (2) how it relates to ideal timing (i.e., fundamental tracking), and (3) how both the cue system and the ideal timing shift with a changing environment. Given this high bar for prediction, researchers have also worked towards more general predictions based on models of trait evolution.

In recent years plasticity theory has developed to provide insights on non-stationarity (or 'sustained environmental change,' see **?**). Models of the role of plasticity in novel environments provide an important bridge to understanding the outcomes of non-stationarity, and generally predict that non-stationarity should favor highly plastic species. At the individual level, environmental tracking is a plastic response, and thus this theory would predict greater individual tracking in non-stationary environments. This outcome, however, assumes there are no costs related to plasticity (**??**) or costs that may limit the evolution of tracking (**?**).

These predictions from plasticity theory may not hold, however, if ecological dynamics reshape the environment as systems transition from stationary to non-stationary. At the community level, competitive hierarchies and fitness asymmetries are likely to shift with changes in the environment. The importance of such short-term dynamics of a changing environment with plastic species highlights how much we need ecological theory for tracking in multi-species environments.

4 Tracking in multi-species environments

Life history theory often ignores other (non-focal) species or abstracts them as an aspect of the environment. While the trophic mis-match literature has addressed this gap for trophic interactions (**??**), there is little consideration of competitive coexistence. Yet decades of research show that competition drives the niche differences necessary for species to co-exist (**??**). Considering how selection in multi-species environments is structured by competition highlights that tracking cannot be considered as a singular trait, but must be evaluated as part of a trait syndrome (or mosaic of traits, **?**) and should ultimately produce communities of species where tracking trades-off with other traits.

4.1 Trait trade-offs with tracking

As environmental 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) and thus may need to be more stress tolerant. 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 (**?**) or species with shallow roots starting growth sooner in an alpine meadow system, while species with deeper roots begin growth later (**?**). In such cases, tracking is akin to a competition-colonization trade-off (**?**), where species that track well gain priority access to resources and, thus, may co-exist with superior competitors.

To examine support for a competition-tracking trade-off in the empirical literature, we reviewed research on phenological tracking and other traits (see Supplement 'Literature review of studies examining tracking & other traits' for search terms and additional methods). This research area has increased greatly in recent years, with a major uptick in studies after 2010 (Fig. 4). Most papers examining tracking and other traits across species 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, or start of migration for a species, termed 'earlyness' by some authors), with earlier species tending to track more (studies included both birds and Lepidoptera, **?????**). This correlation between higher tracking and 'earlyness' each season has been linked to resource acquisition traits associated with lower competitive abilities in plants (e.g., they tend to be of lower height, have shallower roots, narrower diameter vessels, thinner leaves, and grow faster, reviewed in **?**), but our review found few studies that directly 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 rooted or lacking a taproot, **???**). These species were often also early (e.g., **??**), supporting the hypothesis that environmental tracking may relate to a syndrome of traits that allows species to be rapid colonizers each season, but poor competitors in lower-resource periods.

4.2 Including tracking in multi-species community assembly models

Predicting how tracking may determine which species are 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 mechanisms that are both dependent on, or independent of, fluctuations in the environment can lead to coexistence (**??**). These models, which underlie much of current community ecology research (**???**), provide a framework to integrate tracking and non-stationarity into community ecology theory.

In community ecology modeling, definitions of the environment generally fall into two broad categories. In some models the environment is expressed as variation in species' parameters. For example, in an early formalization of the lottery model (**?**), the environment appears as interannual variation in birth and death rates. In later generalizations of competitive coexistence in temporally-varying environments, including the storage effect model (**?**), the environment is formalized as the 'species response to the environment' (E_i), which translates environmental variation (potentially complex and multivariate) into the common currency species' low density per capita growth rates. Building a changing environment into such models thus requires knowing how environmental shifts filter through to species-level parameters (**?**) to impact fundamental tracking. For example, storage effect models predict shifts in communities when environmental change alters the long-term covariance between the environment and competition (i.e., decreasing $cov(E_i, C_i)$), leading to a decrease in the storage effect as a means of competitive coexistence. In another example, **?** added the temporal environment to competition models by defining interaction strength as dependent on the temporal distance between species. This is somewhat similar to models that include the environment effectively through different levels of asynchrony (e.g., **??**).

In other models, the environment is more specifically defined as a resource (e.g., seed germination models where an explicit resource pulse each year initiates germination) and models something close to fundamental tracking. Models that explicitly include the environment provide a major opportunity to predict how tracking and non-stationarity determine future communities. As an example, we modeled a shift to earlier growing seasons using a common coexistence model where the environment is defined as a limiting resource that determines the start of growth each year.

4.3 Adding tracking and non-stationarity to a common coexistence model

To show how resource-based coexistence models can be adapted to study tracking in non-stationary environments we used 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 (**?**), we follow a similar terminology for ease here; however the basic structure of the 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 changes from year to year via variable germination, and within-years is explicitly modeled as a resource pulse at the start of the season. The timing of the resource relative to each species' ideal timing determines how much each species germinates each year, allowing us to include fundamental tracking. Specifically, in the model 'tracking' moves a species intrinsic start time (τ_i for species i) closer to the environmental start time (τ_P), resulting in a higher germination fraction—making it, effectively, a superior colonizer (see SI for complete description and equations).

As with all coexistence models, species can co-occur via equalizing mechanisms, but require stabilizing mechanisms to coexist (see Table 1). Thus species cannot coexist given only variation in tracking—coexistence requires variation in another trait axis. Following the theory and empirical work reviewed above, we included a trade-off between species' tracking and R^* (where species with lower R^* are superior competitors). With variation in tracking and in R^*, species can persist together when the species with a temporal niche advantage is also the inferior competitor (Fig. 5). 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 had persisted together for 500 years—to non-stationary, via an earlier start of season (earlier timing of the resource pulse, τ_p, Fig. 5; 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 (15.1%, see Fig. 5). 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. Taken together, these 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 (**?????**), but may be part of a larger trait syndrome. Indeed, this model trivially shows 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 (**??**)—this must be the case for multi-species persistence. Otherwise, the species best matched to the environment would drive the other extinct.

Additionally, 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-occurred via equalizing mechanisms persisted longer—these equalized species were similarly affected by the changing environment. While this longer persistence of equalized species seems obvious once observed, 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.

4.4 Fundamental versus environmental tracking in multi-species models

Most current models examine the environment from only one of two relevant perspectives: they represent the environment through its effects on fitness (e.g., the storage effect model), or they represent the environment as used for species' cues (e.g., many models of plasticity). Combining these two perspectives, which connect to fundamental and environmental tracking, respectively, may be critical to understanding the costs, benefits, and community outcomes of tracking in non-stationary environments.

Layered onto these different views of the environment is how species responses are defined. In general, species responses to the environment can be broadly grouped into models that explicitly define when species start an event (e.g., spawning or germination) versus those that model the magnitude of response (e.g., the number of propagules or seeds, as discussed above in 'Adding tracking and non-stationarity to a common coexistence model'). Models that explicitly include when a species starts an event are often focused on situations where order of arrival is critical. For example, models of priority effects through niche pre-emption highlight the advantage tracking may provide when it allows species to be early: early arrivals receive a head-start advantage, by gaining priority access to resources they can draw down, reducing resources available to later arrivals (**?**). Such models predict early-arriving species to out-compete other species, unless there is a cost to being too early or there are trade-offs with other species' traits (Fig. 6).

Other models canalize species' responses to the environment into production and investment. Most models of inter-annual competition (most explicit examples of 'modern coexistence theory,' e.g., **??**) fall into this camp. Species produce (via investment in offspring, tissue, etc.) differentially depending on the environment each year and outcomes are mediated through density. While these models may seem disconnected from timing, they are built on the idea that how well species are 'matched' to the environment varies across both species and years, and determines the density independent component of fitness. Some models explicitly define this 'match' based on phenology (**???**), highlighting how phenology often relates to production and, thus, investment across years. Further, they almost always model the environment as a distribution (**??**), which provides the opportunity for the environment to alter the competitive environment each year and, thus, structure coexistence.

A model where species vary both when they start an event and how much they invest would capture the important attributes of fundamental tracking—combining head-start advantages from being early with production variation based on the resource environment. To our knowledge, however, most models approach these questions separately, though models of bet-hedging come closest (**??**).

4.5 Frontiers of community assembly models

A model that explicitly includes the linked decisions of when to time an event and how much offspring/tissue to produce during the event could provide insights on the relative importance of each aspect of this process. Such a model could be adapted to address multiple questions of tracking, including how these decisions ('when' and 'how much') may trade-off and which other traits may be most strongly linked to tracking, 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 testing how environmental tracking relates to fundamental tracking and species persistence with climate change, and research has already begun to tackle this non-trivial challenge (**???**). Most work to date, however, focuses on conclusions from systems that are initialized as non-stationary, ignoring the transition between stationary and non-stationary environments. Yet we expect this transition may be critical because communities formed in stationary environments (or periods with lower non-stationarity) are effectively filtered and assembled by that environmental regime and thus produce the baseline of variation and assembly dynamics for a shifting environment. While analytical solutions for systems transitioning from stationary to non-stationary may take time to develop (**?**), simulation work can provide an immediate intuition and framework to address this challenge.

Outcomes for such community assembly models also depend on how effectively closed communities are. 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, 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 (**??**). Evolutionary responses could also rescue species with low plasticity. Long-term population (e.g., **?**) and resurrection studies (**??**), and field experiments (**??**), 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., **?**), associated with reduced population viability (e.g., **?**), and that other factors may constrain adaptive responses.

5 Linking empirical and theoretical research

We have outlined above how current community ecology theory could make advances through models that combine effects of variation in timing and production and models that include the environment as impacting species' cues, and species' fitness. Such models would explicitly include the potential costs and benefits of tracking depending on how closely environmental tracking matches fundamental tracking. But to best test and develop such models we need a greater understanding of how the environment is changing alongside more robust estimates of environmental tracking and how it fits within a mosaic of correlated traits that determine individual fitness.

5.1 Defining the change in an organism's environment

Currently, much research has focused on one major shift in the climate system—rising temperatures, but research on multivariate environmental shifts is critical to understanding how climate change affects an organism's whole environment. Research in this area is already increasing (e.g., **?**), and empirical research can guide work on theory by identifying environmental shifts that are often linked (e.g., **?**); for example, warming temperatures may drive earlier seasons and higher evaporative water loss. Empirical studies should also consistently characterize the environmental distributions of study systems that appear linked to 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 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 (**?**). 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 (**?**).

5.2 Robust comparable measures of phenological 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 to impacts on growth and survival. Studies should clarify their definition of tracking, how the environment is defined, how an event relates to fitness, and how well—or not—the underlying cue system is understood. Currently, some studies of trophic asynchrony examine fundamental and environmental tracking simultaneously (e.g., **???**), but most studies are comparatively less clear. The more researchers can clarify when and how they are addressing fundamental versus environmental tracking, the more easily we can compare results across studies.

Even with clearer definitions, progress in documenting and understanding empirical variation requires more robust estimates of phenological tracking. Increasingly, research has outlined statistical difficulties in measuring tracking, which may underlie many observations of species that do not track. These issues mostly relate to the challenges of non-stationarity in units caused by using discrete (calendar) time, low power, and the complexity of climate data.

Non-stationarity in units comes in many forms—estimates of mean days shifted per decade depend strongly on the climate of the decade(s) studied, which is not consistent in many systems (**?**). Estimates based on a relevant climate variable can sometimes ameliorate this problem, but may be equally vulnerable to non-stationarity in units (e.g., **?**). 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 depend on 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).

Even 'long' time-series may be too short for robust analyses of trends (**?**). Authors should be especially cautious if they find only large effects appear significant (e.g., **?**), which is a well-known statistical bias associated with p-values (**?**). 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.

Many statistical issues can be addressed by improved statistical approaches (e.g., **??**), though such approaches may uncomfortably highlight how uncertain many current estimates are (**?**) or reveal lower effect sizes. Impacts of start-years for long-term time-series can be muted by applying change-point or hinge models (e.g., **?**). We suggest mixed models should be used more widely alongside randomization and/or data-simulation approaches (e.g., **?**), and we need models that can discriminate among confounding factors. For example, we reviewed above growing evidence that suggests a potential fundamental trade-off where early species track, grow fast and die young, while later species track less, grow slowly and live longer—this might suggest later species bet-hedge more given their longer investment window. Or it could be an artifact where early species use simpler cues, and, thus, their tracking is measured more accurately given current methods.

5.3 Building from cue systems to phenological tracking

Even without statistical issues, translating event date and climate data into estimates of tracking requires a firm biological understanding of an organism's cues, which we rarely have (**?**). Currently, 'tracking' is often measured as the relationship between the dates of an event and a simple abiotic metric. Such measures, however, are almost always proxies for a more complicated underlying physiology where simple cues—such as warm temperatures or snowpack—can be modified by other cues, such as photoperiod, drought or light spectra (**??**). Modeling multivariate cues, however, is inherently difficult (**?**), especially since one cue may dominate in many conditions (**?**). 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 (**?**).

Addressing these issues is possible if we embrace our inner physiologists—or collaborate with one—to develop models that explicitly include species' cues. Research on model systems has highlighted the multivariate nature of most cues at the genetic level (**?**)—where expressed differences in phenology are the outcome of one genetic pathway under different environmental regimes (**???**). Such work on the heritability and underlying genetics of phenological plasticity has often found similar genes with similar functions across taxa (**??**). This provides hope for a more general framework where cue systems can more quickly be identified. Such a framework would also allow forecasts that include the shifting genetics of phenology as species shift their ranges with climate change (e.g., **?**).

Models that include species' cues and consider the framework under which we expect cue systems have evolved (e.g., bet-hedging) could further a general framework for what cue systems we expect across species and environments. We then must interrogate these models to understand when they work and where they fail (see **?**, for an example). This approach can help embrace the contradictory pulls of conducting experiments to identify mechanistic cues and understanding how they are filtered through the multivariate climate of the real world (see **??**).

5.4 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 empirical work has mainly documented tracking, linked it to performance, or focused on how it varies between native and non-native species (**???**). Research has highlighted some traits that co-vary with tracking (e.g., **???**), but to tie this 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.

Progress may come from greater efforts to measure and report phenological differences in species-interaction studies. In particular, ecology has a long history of lab and field experiments on competition—which have been critical to our understanding of niche differences and how competition stabilizes and shapes communities (**??**). After decades of research hinting at the role of phenology in determining competitive outcomes, recent research has highlighted the role of phenology through 'seasonal priority effects,' 'within-season niche differences' or 'size-mediated priority effects' (**????**). While these studies have focused on phenology explicitly, we suggest all competition studies should measure and report phenological differences, which could rapidly help elucidate how phenology contributes to per-capita fitness outcomes of competitive interactions.

Finally, while traits that link to resource competition may be especially fruitful for greater research, they should not be the only ones considered. For example, traits related to stress, predator tolerance and 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.

5.5 Embrace non-stationarity

While most environments today are climatically non-stationary and have been for decades, the climate will return to a more stationary form in the future—likely some centuries after the stabilization of greenhouse gases (**?**). As paleobiologists and evolutionary biologists often point out, climatic nonstationarity is a common part of the earth's history (**?**)—even if stationary periods—be they cold or warm (glacial and interglacial periods), or dry or wet (megadroughts or pluvials)—are more common. Indeed, while much of this work has examined how species survive for millions of years given large oscillations in climate (**?**), the periods that provide the most dramatic community reshuffling are periods shifting from stationary to non-stationary climate regimes (**??**). Such stories of the past are now happening today, and have caused ecologists to question their simplifying assumption of stationarity (**?**). We argue that better predictions of climate change impacts—and fundamental insights for ecology—will come from embracing the complexity of non-stationary environments.

6 Conclusions

(1) Growing empirical evidence highlights that phenological tracking may be linked to species performance and critical to understanding the forces that assemble communities and determine species persistence. Anthropogenic climate change has shifted many systems from generally stationary to non-stationary climate dynamics—making how well species can track this change an important topic of research both for empirical studies of climate change and for foundational ecological theory.

(2) Definitions of tracking in conceptual and theoretical studies often diverge from empirical global change studies of tracking, which may hinder efforts to combine theory and empirical data for better predictions. In conceptual and theoretical studies tracking often refers to how well an organism matches the timing of a life history event to the ideal timing for that event and connects to an organism's fitness (**?**). In contrast, in empirical studies tracking often refers to a statistical estimate of a change in the timing of an event relative to a measured environmental variable (**?**).

(3) We outline a suite of confounding factors that may make many current estimates of inter-specific variation in tracking less accurate than they appear, including a weak understanding of organisms' underlying cue systems, simplified estimates of complex multivariate changes in the environment, and issues of statistical power. This in turn means we may have only very rough estimates of which species, when, and where, do and do not track. Given this difficulty, we argue that clear testable predictions from ecological theory would be especially valuable to guide the field forward (**?**).

(4) We show how ecological theory designed on how a variable environment can shape the formation and persistence of species and communities could guide future research on phenological tracking. Basic models of coexistence in stationary environments highlight that tracking must trade-off with other traits for multi-species communities to exist. This suggests the paradigm from empirical studies of invasive species that climate change should favor tracking may need to expand to include more traits. To fully apply these findings to tracking of global change, however, requires new models that examine how communities shift as previously stationary environments become non-stationary.

(5) We outline how uniting several major divides in current modeling approaches could improve predictions and guide empirical studies. These divides include: *(i)* whether the focus is on the timing of an event or the investment in an event (e.g., seeds or other offspring), *(ii)* whether the environment affects fitness or affects species cues that trigger events (that may eventually affect fitness), and *(iii)* whether a changing environment is modeled directly via a resource or similar abiotic component or considered only via species-level parameters.

(6) Areas where empirical research could help guide theory are clear. In particular we need: *(i)* a greater focus on understanding the attributes of a multivariate environment shaped strongly by humans, *(ii)* measures of phenological tracking that are more comparable across species and sites, and statistically robust, which will require *(iii)* efforts to build a framework to identify species' cue systems, *(iv)* more studies of how phenological tracking fits within the complicated mosaic of an organism's traits. Across both empirical and theoretical research a greater focus on non-stationarity, including transitions between stationary and non-stationary systems, could provide fundamental and applied advances.

7 Acknowledgments

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

8 Tables

Term – definition
community assembly – the suite of processes that determine which species are in a community—this includes processes that determine which species arrive and persist, including stabilizing and equalizing mechanisms, constrained by the regional species pool
cue reliability – the correlation between an organism's fitness given the ideal timing of a life history event and its fitness given the actual timing, which is based on its cue system (Fig. 2-3)
environmental tracking – the change in timing of a biological event due to an organism's cue system given change in the environment (Fig. 3, note the shift in timing between sites); for example, for a tree whose budburst cue system is a combination of chilling, forcing, and photoperiod, its environmental tracking would be the shift in timing of budburst in response to changing environmental conditions, such as warmer winters and springs
equalizing mechanism – processes that minimize fitness differences between species in a community
fundamental tracking – the change in the ideal (fitness-maximizing) timing of a phenological event; for example, how the 'best day' for a phenological event changes from year to year. This is typically represented in an idealized way (Fig. 2), but may be more complicated in a multivariate system (Fig. 3).
non-stationary environment – the underlying distribution of abiotic characteristics of a location changes over time (e.g., warming temperatures, larger rainfall events)
phenological events – the outcome of a two-part process that is repeatedly observed over time. At each temporal unit, an event can either happen or not (when—part 1) and, if it happens, the event can vary in size or degree of investment (how much—part 2).
stabilizing mechanisms – processes that cause species in a community to more strongly limit their own fitness than other species' fitness (e.g., the common requirement for coexistence that intraspecific competition must be stronger than interspecific); includes niche differences.
stationary environment – the underlying distribution of a location's abiotic characteristics is unchanged across time (i.e., constant mean and variance); this suite of characteristics varies by habitat type, but generally includes physical climatic factors, such as temperature and precipitation

Table 1: Glossary of major terms related to phenological tracking and community assembly.

9 Figures

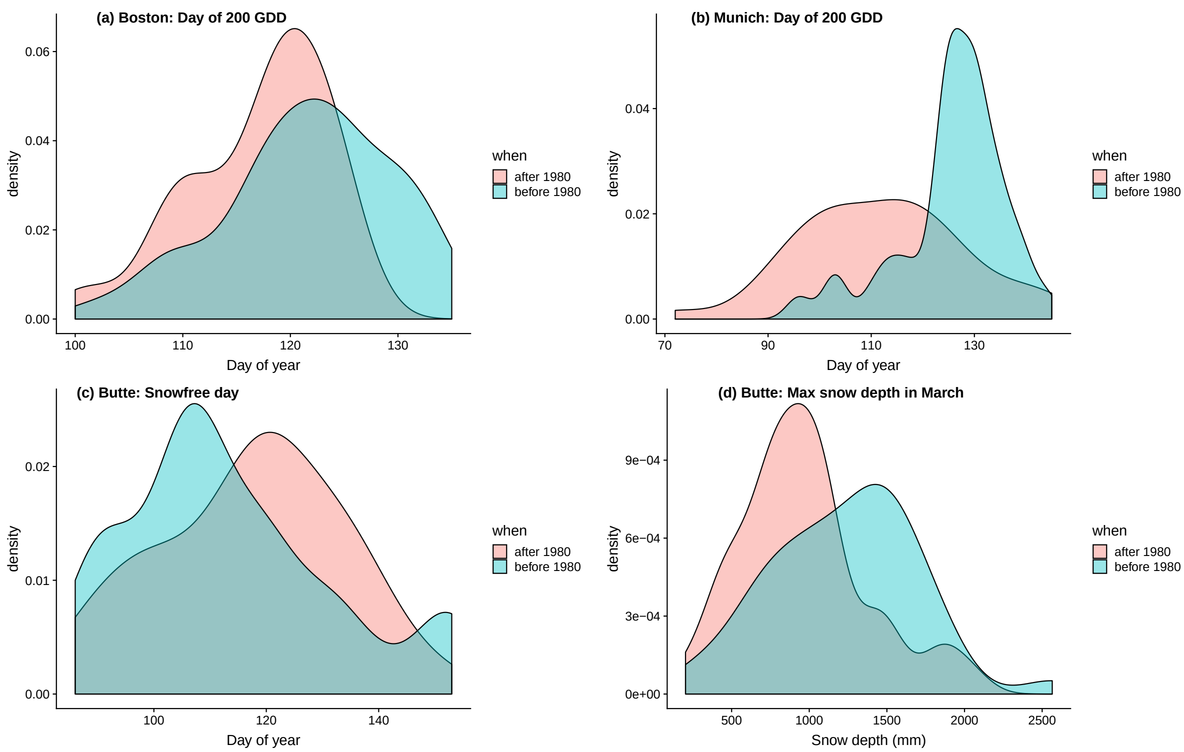

Figure 1: Understanding non-stationarity in ecological systems requires first identifying which aspects of the environment have shifted—and how they have shifted with respect to one another (e.g., **??**)—as the underlying distributions transition from stationary to non-stationary. Here we show examples of non-stationarity in climate variables that affect phenology by comparing 40 years of data before and 40 years of data 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.

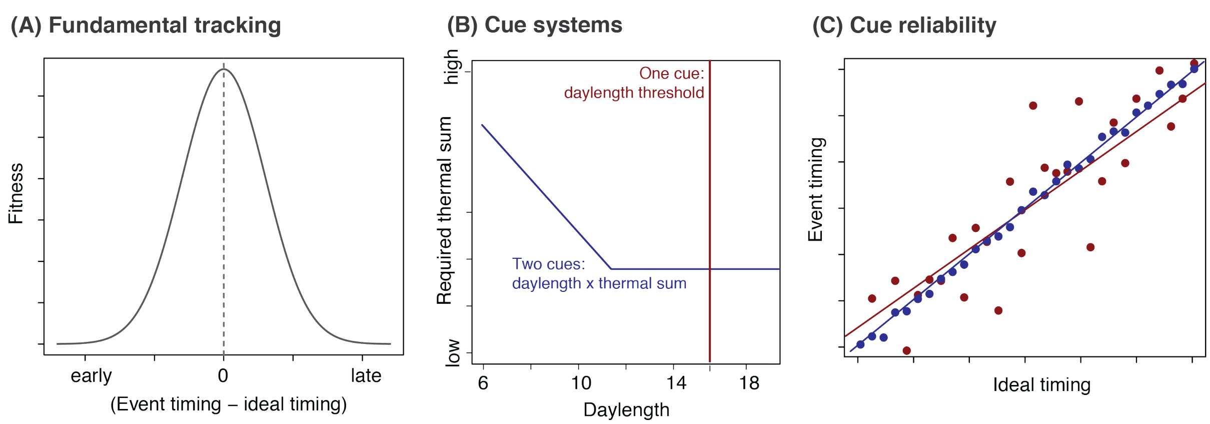

Figure 2: Fundamental tracking (A) represents how an organism matches the actual timing of a life history event (phenology) to the timing that maximizes its fitness (i.e., ideal timing). Here, we show a common conceptualization where fitness declines as event timing moves away from the ideal timing, though realizations in nature may take diverse forms. As the ideal timing is generally only clear in simplified models or in retrospect, species that phenologically track must use a cue system (B) to try to match their phenology to the ideal timing across environments (temporally and/or spatially). Here we show two cue systems: one single cue system dependent only on daylength (red line: the event occurs when the organism's environment exceeds a certain daylength) and one multivariate cue system, which depends on a combination of daylength and thermal sums (navy line: the event occurs when the organism accumulates enough temperature for the current daylength). The match between ideal timing and actual timing represents cue reliability (C).

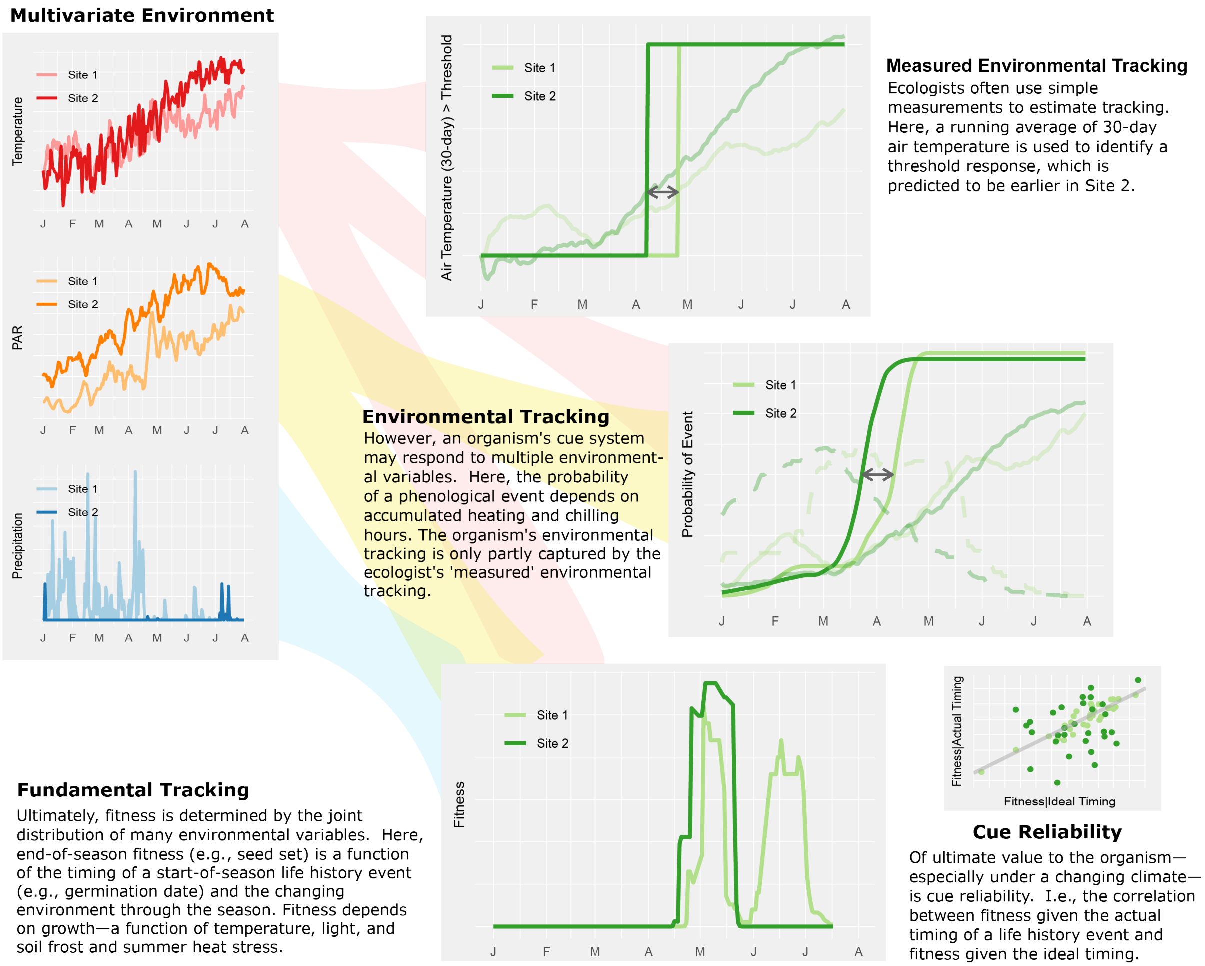

Figure 3: Different components of a multivariate environment influence phenological tracking. Ecologists may use a simple seasonal metric to *measure environmental tracking*, but an organism's *environmental tracking* may reflect more complex cues. Ultimately, fitness is determined by the joint distribution of many environmental variables through time after the start-of-season life history event. Cue reliability is the relationship between the timing that results from an organism's cue system and the fitness of the organism. See SI 'Fig. 3 methods' for further methods and details.

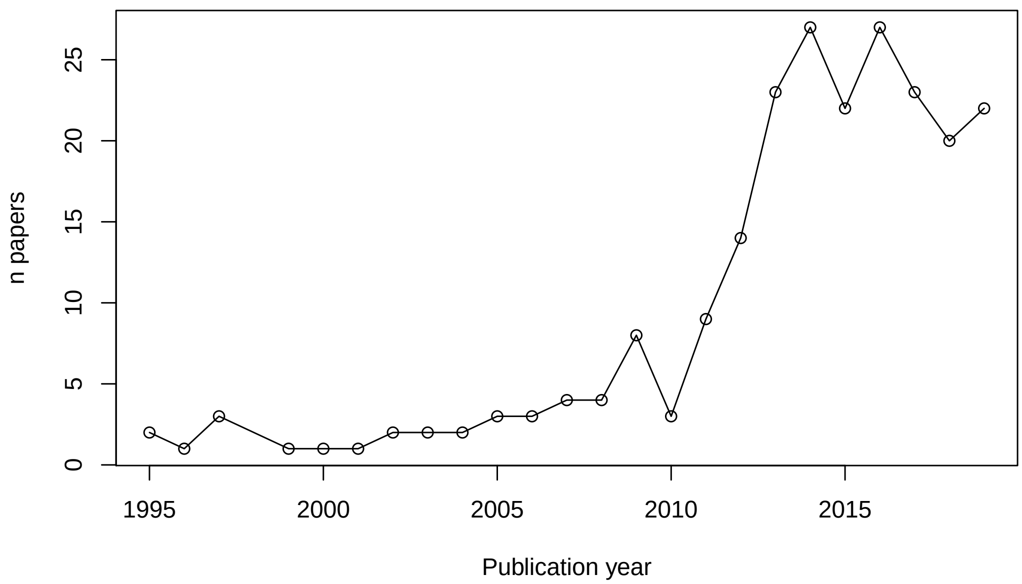

Figure 4: Trends in all papers examining tracking & other traits from a systematic literature review. We searched ISI (four searches: (1) Topic: 'phenolog* chang*' and Title: phenolog* AND trait*, (2) Topic: 'warming shift*' AND trait* and Title: phenolog*, (3) Topic: 'phenolog* track*' AND trait* and Title: phenolog*, (4) Topic: 'phenolog* sensitiv*' AND trait* and Title: phenolog*), which resulted in 231 papers, 83% of which were published in 2011 or onward. Of papers from which we could extract data 25 of 30 were published in the same period. See SI, 'Literature review of studies examining tracking & other traits,' for detailed methods.

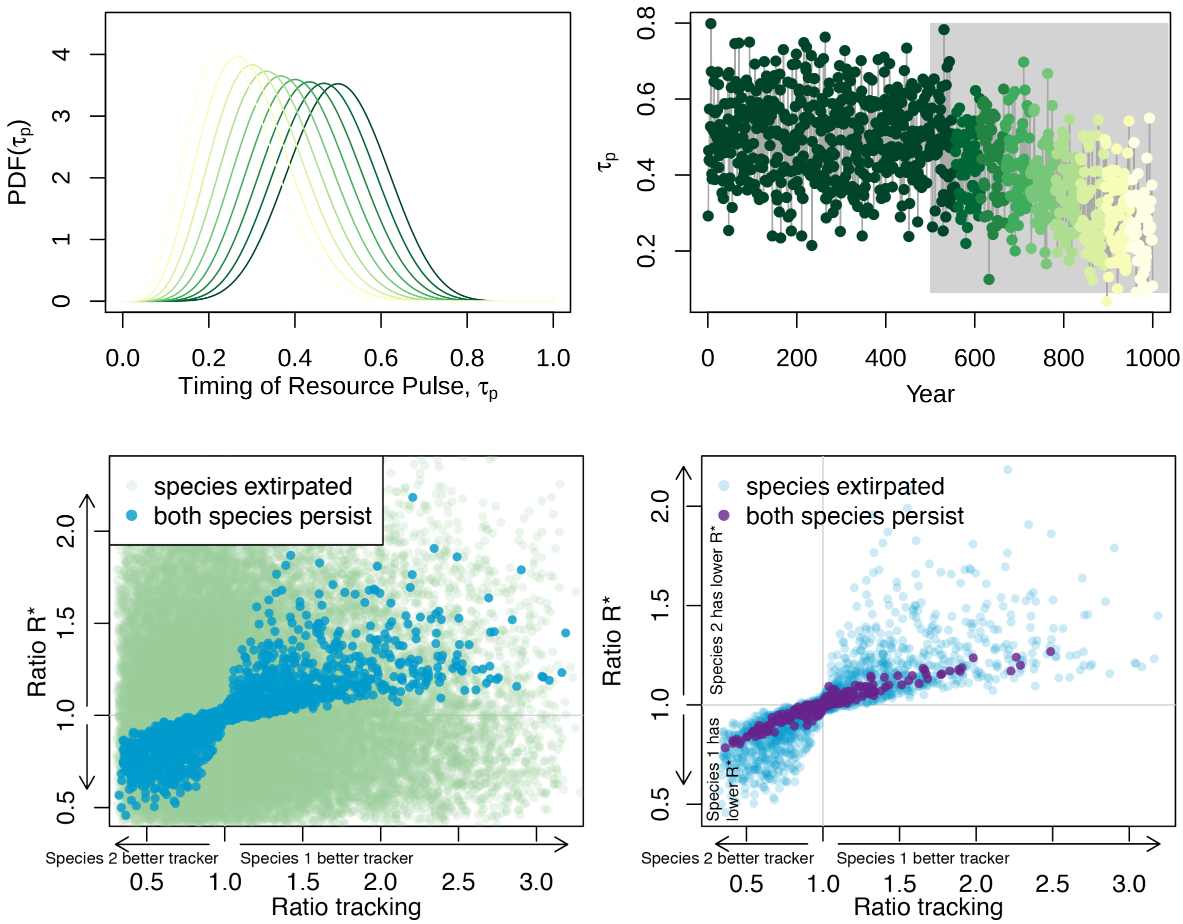

Figure 5: Example of how non-stationarity can reshape communities in a simple coexistence model. We shifted the environment (top panels) by changing the timing of the resource pulse from a stationary period ($\tau_p \sim \beta(10, 10)$ for the 500 years) to a nonstationary period ($\tau_p \sim \beta(5, 15)$ 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).

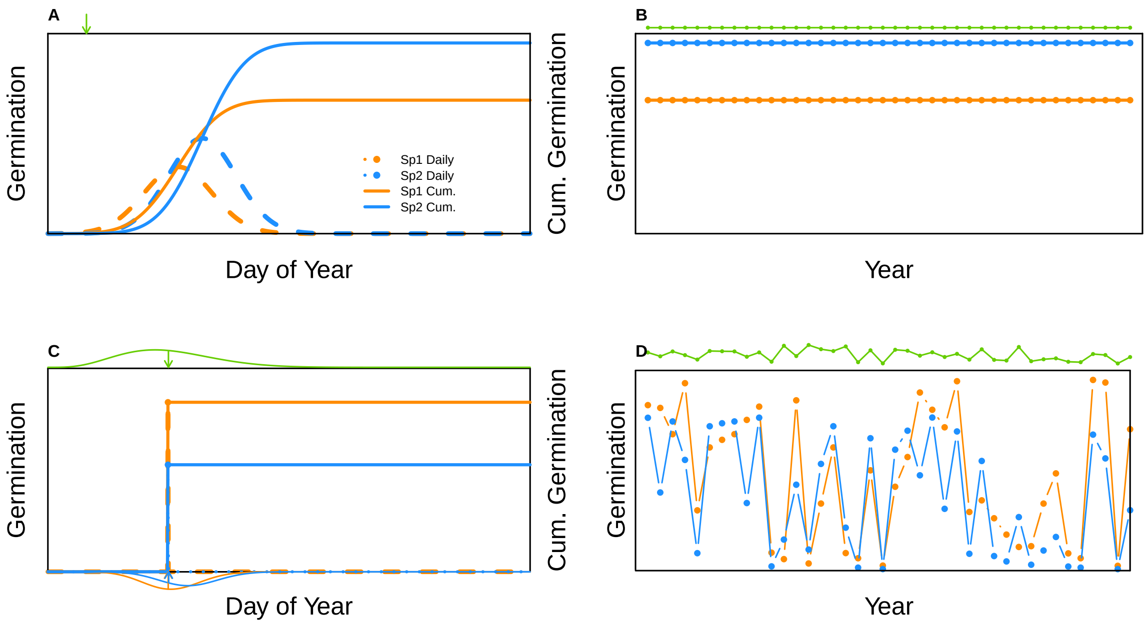

Figure 6: Tracking can be conceptualized as changes in priority effects or changes in storage effects. In a priority effect model (A-B), the coexistence mechanism is a within-year tradeoff between, an early-germinating species that pre-empts resources (sp 1) and late-germinating species that is a superior resource competitor (sp 2) (A, where green arrow indicates the start of season); no between-year variation is required to maintain coexistence. In a storage effect model (C-D), variation in the timing of the start of season (indicated by the distribution in green, top of C) results results in differential species-response to the environment (illustrated by species-specific germination curves, bottom of C); this interannual variation in species-response to the environment (D)—along with a seedbank or other interannual storage mechanism—can maintain coexistence through reduced interspecific competition.