CONCEPTS & SYNTHESIS
EMPHASIZING NEW IDEAS TO STIMULATE RESEARCH IN ECOLOGY

Ecological Monographs, 0(0), 2019, e01352
© 2019 by the Ecological Society of America

Phenology as a process rather than an event: from individual reaction norms to community metrics

Brian D. Inouye [1,2,3,5] Johan Ehrlén,[2,4] and Nora Underwood[1,2,3]

[1]*Biological Science, Florida State University, Tallahassee, Florida 32306 USA*
[2]*Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm 106 91 Sweden*
[3]*Rocky Mountain Biological Lab, Gothic, Colorado 81224 USA*
[4]*Bolin Centre for Climate Research, Stockholm University, Stockholm 106 91 Sweden*

Citation: B. D. Inouye, J. Ehrlén, and N. Underwood. 2019. Phenology as a process rather than an event: from individual reaction norms to community metrics. Ecological Monographs 00(00):e01352. 10.1002/ecm.1352

Abstract. Measures of the seasonal timing of biological events are key to addressing questions about how phenology evolves, modifies species interactions, and mediates biological responses to climate change. Phenology is often characterized in terms of discrete events, such as a date of first flowering or arrival of first migrants. We discuss how phenological events that are typically measured at the population or species level arise from distributions of phenological events across seasons, and from norms of reaction of these phenological events across environments. We argue that individual variation in phenological distributions and reaction norms has important implications for how we should collect, analyze, and interpret phenological information. Regarding phenology as a reaction norm rather than one year's specific values implies that selection acts on the phenologies that an individual expresses over its lifetime. To understand how climate change is likely to influence phenology, we need to consider not only plastic responses along the reaction norm but changes in the reaction norm itself. We show that when individuals vary in their reaction norms, correlations between reaction norm elevation and slope make first events particularly poor estimators of population sensitivity to climate change, and variation in slopes can obscure the pattern of selection on phenology. We also show that knowing the shape of the distribution of phenological events across the season is important for predicting biologically important phenological mismatches with climate change. Last, because phenological events are parts of a continuous developmental process, we suggest that the approach of measuring phenology by recording developmental stages of individuals in a population at a single point in time should be used more widely. We conclude that failure to account for phenological distributions and reaction norms may lead to overinterpretation of metrics based on single events, such as commonly recorded first event dates, and may confound meta-analyses that use a range of metrics. Rather than prescribing a single universal approach to studying phenology, we point out limitations of inferences based on single metrics and encourage work that considers the multivariate nature of phenology and more tightly links data collection and analyses with biological hypotheses.

Key words: phenological distribution; phenological mismatch; phenology; reaction norm; response to climate change.

Introduction

Although studies of the seasonal timing of biological phenomena have a very long history (Demaree and Rutishauser 2011), the need for more accurate measures of phenology has become particularly acute with the recognition of anthropogenic climate change. Changes in phenological events such as the timing of bird migration, frog spawning, mammal hibernation, and first flowering are considered to be bellwethers of climate change (e.g., Beebee 1995, Inouye et al. 2000, Miller-Rushing et al. 2008, Charmantier and Gienapp 2014), and there is broad concern about the potential consequences of climate-induced changes in phenology for species persistence (Willis et al. 2008), species interactions (Yang and

Manuscript received 27 June 2018; revised 8 November 2018; accepted 3 December 2018. Corresponding Editor: Jessica R. K. Forrest.
[5] E-mail: binouye@bio.fsu.edu

Rudolf 2009), and ecosystem function (Cleland et al. 2007). Studies of phenology focus on issues ranging from natural selection (e.g., Valdés and Ehrlén, 2017) to species interactions (e.g., Reed et al. 2013), community structure and ecosystem function, and changes in phenology over time and space (e.g., CaraDonna et al. 2014, Hua et al. 2016). There is still, however, considerable uncertainty about how best to measure phenology.

The most common measures of phenology account for only a small amount of the biologically important information about phenology through a season, across environments and among individuals. Phenology is most often characterized in terms of discrete *events* at the population level (Fig. 1A), particularly "firsts" such as the date on which the first individual animal emerges from hibernation, or the first individual plant blooms. But these events are merely discrete landmarks within the *distribution* of events through the growing season that makes up a full phenology (Fig. 1A). Phenological distributions and the landmarks extracted from them can be measured at many levels of biological organization from within individuals to ecosystems, with phenologies at population levels and above being the sums of individual-level phenologies (Fig. 1B, C). Although single-event descriptions of phenology have logistical advantages, in many cases, it is not clear how individual events relate to outcomes we are really interested in, such as fitness, interaction strength, or resource availability across a season. For example, the seasonal flux of floral resources for pollinators and thus effects of floral timing on pollinator fitness cannot be inferred from only the date on which flowers are most abundant in a community or the dates of the first flower.

Phenological traits are typically plastic, changing with abiotic and biotic cues across space and time. For example, plants may bloom earlier in warmer years or locations than in colder ones. This plasticity in phenology, or sensitivity to climate, is important for understanding how climate change may influence phenology and species success, and the evolution of phenology. In fact, we are often specifically interested in how the timing of biological phenomena change relative to climate or other environmental factors, including interactions with conspecifics (Marquis 1988, Ison and Wagenius 2014) or heterospecifics (e.g., Elzinga et al. 2007, Rafferty et al. 2013, Reed et al. 2013, Kharouba et al. 2018). The fact that phenological traits are so often plastic, and particularly that individual organisms, populations and species can *differ* in their plasticity, is critical for understanding how to interpret measures of phenology and for recognizing the limitations of describing phenology in terms of events. For example, if individuals vary in the plasticity of their responses to climate, the date at which the first individual in a population reaches a phenological stage can be a poor representation of the average population response.

Understanding the phenological response of populations to changing environments will likely remain a key goal of ecology for the coming decades. We argue that dates of single events provide limited information about phenology, particularly when individuals vary in their phenological traits and in the plasticity of those traits. We will consider how such a view can help researchers identify the most appropriate measures of phenology for their research questions. This paper is not a manual specifying particular metrics or tests that researchers should use; questions about phenology and biological systems are too varied and complex for any one paper to anticipate the best course of action in every case. The most appropriate metric (summary statistic describing an aspect of a phenological distribution) will depend on both the biological question and the organisms being studied. For some questions, what matters most might indeed be the initiation of biological activity (a first event), whereas for others what matters most might be interactions between organisms across the season. Researchers are often limited by the data available, so that ideal metrics may not be attainable even if they are identifiable. Our goal is to help researchers identify useful metrics for their systems and questions, and appropriately interpret the metrics that are feasible if the ideal is out of reach. We provide some background on characterizing plasticity in phenological traits, suggest how appreciating the many sources of variation in phenology can help identify more powerful approaches to answering key questions about phenology, discuss an under-used approach to gathering phenological data, and conclude with recommendations for future directions. The insights gained by appreciating phenological distributions, plasticity, and variation among individuals are relevant to all types of organisms so we draw on examples from a range of taxa, but given our research backgrounds we use phenological responses of plants in much of our discussion.

Characterizing Phenological Plasticity: Phenological Reaction Norms and Sensitivity

Many phenological traits, like the timing of flowering, migration, or mating, are plastic, so their expression depends on environmental conditions. The effects of environmental conditions on phenology can be both direct and indirect (Chmura et al. 2019). For example, enzyme kinetics are typically faster at warmer temperatures, so that metabolic rates can respond directly to temperature, but temperature could also act indirectly as a cue that influences a developmental program. Developmental timing can also be affected by interactions with other organisms, such as when bird breeding schedules respond to food supply (Martin 1987) or the maturation and senescence of flowers depend on pollinator visitation rates (Herrera 1995). If we were to measure phenological distributions for the same individual in different years with varying environmental conditions (for a long-lived species) or for the same genotype in different environmental conditions within the same year, we would thus expect phenological distributions to shift with the environment (Fig. 2A).

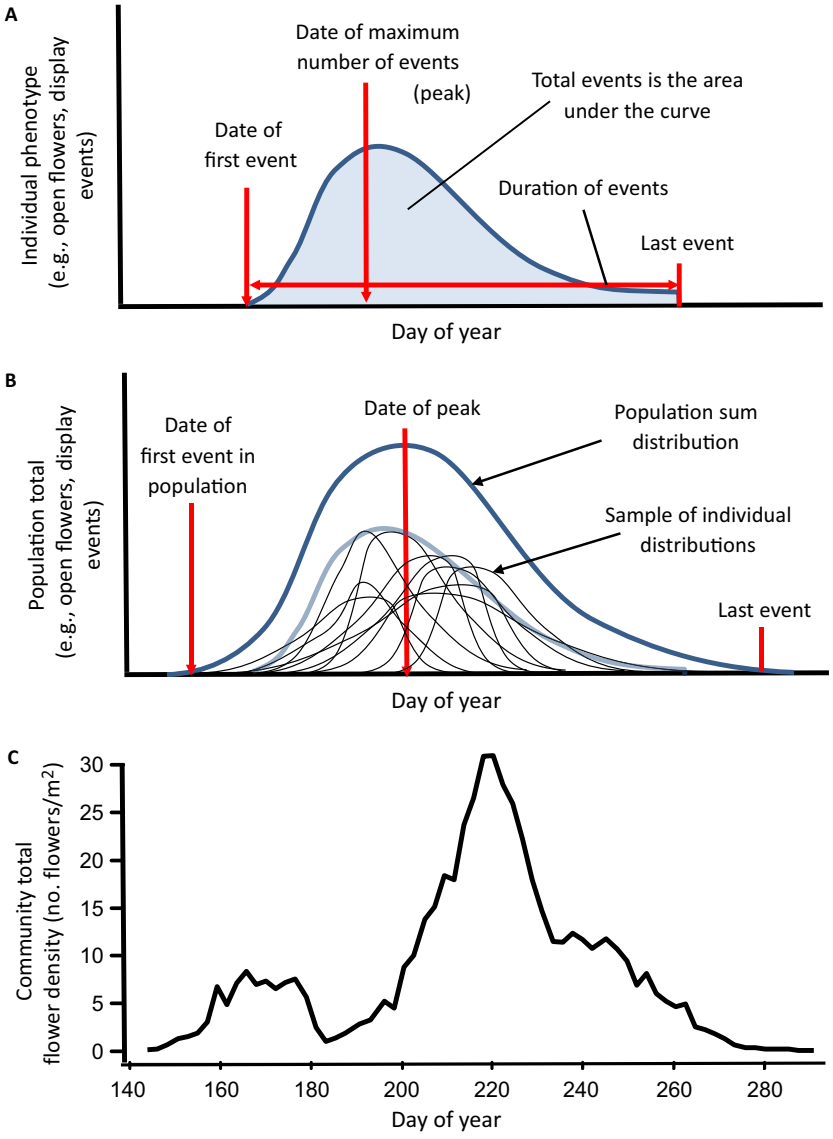


Fig. 1. Phenologies across scales of biological organization. A phenological *event* is any biological occurrence that has a seasonal pattern of expression (Denny et al. 2014), such as flowering, spawning, hatching, metamorphosis, or leaf flush. The most complete characterization of phenology would be the *distribution* of phenological events over the entire growing season; for example, of daily open flowers on an individual plant, eggs laid by a butterfly, or bird nesting activity. Fig. 1A shows an idealized distribution for one individual for a trait whose expression increases and then decreases. Traits such as animal growth and development would instead have monotonic distributions (a one-way trajectory from a beginning to an ending state along the *y*-axis). From full phenological distributions one can derive phenological metrics based on dates of key events such as first flowers, peak flight activity, or hatching time as well as integrative descriptions including duration of non-zero events, total events, or periods of resource scarcity (Ogilvie et al. 2017). Phenological distributions can be measured at a range of scales, from individuals to communities (e.g., Rathcke and Lacey 1985, Stucky et al. 2018), such as flower openings within an individual plant (Bolmgren 1998), the timing of the first flower in a population of plants (e.g., Anderson et al. 2012), the distribution of flower abundance for all plant species used by a pollinator through a season (Ogilvie et al. 2017), or the timing of green-up in an ecosystem (e.g., Cleland et al. 2006). When we study phenology at the level of a population, species, community, or ecosystem, we aggregate the distributions of individuals (e.g., CaraDonna et al. 2014; see Fig. 1B for idealized representation). For example, open flowers on individual plants can be summed to flowers in a population, or the singing activity of individual birds aggregated into the density of singing birds in a population. Thus, population distributions have a duration at least as long as the longest individual duration, and the date of the first event is the date for the earliest individual. Similarly, population or species data can be aggregated to form phenological distributions for communities or ecosystems. Fig. 1C provides an example using real data at the community scale: the total flowers (summed over 32 plant species) available for long-tongued bumblebees (*Bombus* spp.) in 2014 in 23 4-m^2 plots in a subalpine meadow at the Rocky Mountain Biological Laboratory (Colorado, USA). Other examples of community or ecosystem metrics would include the density of plankton (e.g., Thackeray et al. 2016), normalized difference vegetation index (NDVI; as a proxy for productivity), or plant green-up (e.g., Piao et al. 2015, Wu et al. 2017).

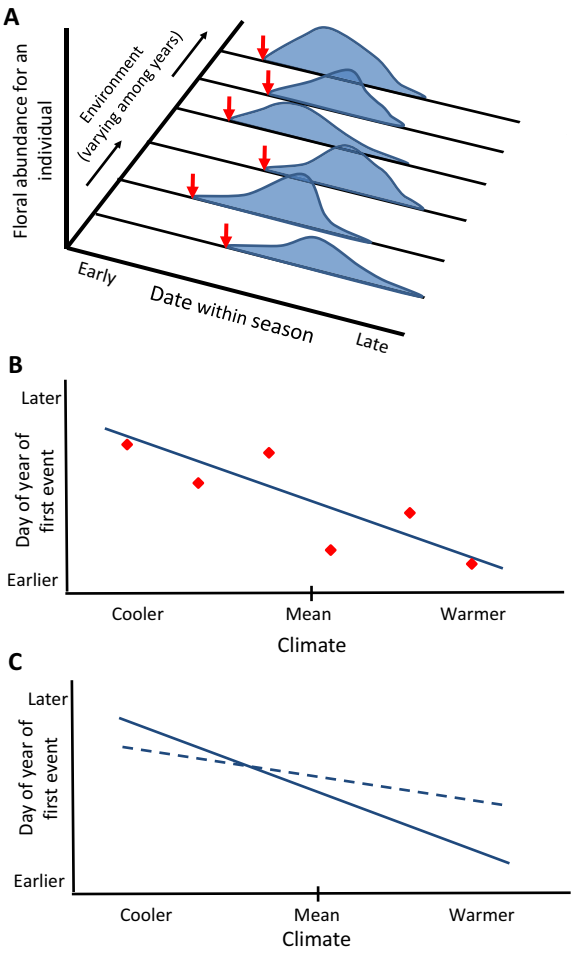


FIG. 2. (A) The date on which phenological events begin (e.g., first flowers, as indicated by the red arrows) and how events are distributed through the season can vary across years. Note that variation in phenological distributions could also be measured for a population or species, in which case, the environment can vary among years or sites. (B) The norm of reaction for a phenological trait is the relationship between the timing of that trait and some environmental variable, in this case "climate." In this idealized example, each red dot is derived from one of the distributions in a year, such as in panel A at the locations of the red arrows. For example, the graph could show the time of first flowering for an individual over six years that span a range of climate conditions. The reaction norm has an intercept at the average climate, and its slope is the sensitivity of phenology to climate. Reaction norms can be measured for an individual (for which trait expression can be measured in different environmental conditions in different years), for genotypes (if replicates are exposed to different environments or persist across years as clonal lineages), and for populations or species. (C) Norms of reaction can vary in slope, in elevation, or both slope and elevation (as shown here) among individuals, populations or species. The dashed line is a reaction norm with a shallower slope, i.e., less responsive to climate variation, and higher central intercept, indicating later phenology onset in mean climate conditions.

The variation in phenological event dates across environmental gradients can be visualized as a *reaction norm* (Fig. 2B). Reaction norms are characterized by their elevation (i.e., central intercept), which describes the trait level at mean environmental conditions, and their slope, which describes how responsive the trait is to environmental conditions (Visser et al. 2010). In the context of phenology, the *x*-axis of the norm of reaction is some environmental variable, such as temperature, and the *y*-axis is the timing of a biological event of interest (Fig. 2B). When individuals respond differently to the environment their norms of reaction differ, either in slope, in elevation, or both (as in Fig. 2C). As a result, the phenologies of individuals or species can converge, diverge, or even change order, depending on the environment (but see Crawley and Akhteruzzaman 1988).

Reaction norms can be quantified in several ways, all of which require a lot of data (Weis and Gorman 1990, Brommer et al. 2012, Morrisey and Liefting 2016). At the individual level, ideally we put the same genotype (or siblings) in multiple environments and record phenologies (e.g., Anderson et al. 2012). Alternatively, for perennial organisms we can estimate reaction norms by measuring phenotypes in years that differ in environmental conditions. This latter approach will work for traits that are expressed multiple times during the lifetime of an organism, such as flowering time in perennial plants, and where phenotypic trait values in each year depend on environmental conditions (i.e., labile plastic traits; Lynch and Walsh 1998). However, this approach requires that we be able to separate the effects of environmental variation from effects of individual age or past experience (Tavecchia et al. 2005). We note that while most studies estimating reaction norms have used linear functions, recent work suggests that the curvature of reaction norms is an important component of the evolution of plasticity (Murren et al. 2014).

Reaction norms can also be defined at the population or species level, where a reaction norm describes changes in the mean expressed trait level within the population or species with changes in the mean environmental conditions. As with individual reaction norms, at the species level, we can estimate the influence of environmental factors by observing variation in phenology over environmental gradients, often in terms of among-year variation in climate (Brommer 2013). In many phenology studies, this relationship is referred to as the environmental *sensitivity* of a population or species (e.g., Willis et al. 2008, Wolkovich et al. 2012). Studies of the effects of climate on species interactions often ask whether species differ in sensitivity to environmental factors such as temperature and imply that differing sensitivities to temperature in combination with climate warming may lead to changes in the relative timing of species' activities and increasing "mismatches" (e.g., Kharouba and Vellend 2015, Thackeray et al. 2016, Kharouba et al. 2018). Although these studies rarely explicitly use the term "reaction norm," the reasoning is implicitly based on the same concept.

When there is variation in reaction norms, it is possible for the slope and the intercept to be correlated. At

the individual level, a positive correlation would mean that individuals with earlier phenology in an average year (lower intercept) have greater sensitivity (more negative slope) and a negative correlation would mean that individuals with earlier phenology are less sensitive to the environmental factor (Fig. 2C). For example, there are correlations between slopes and intercepts for individual thermal reaction norms of salmonids (Janhunen et al. 2016), and data on reaction norms for flowering date in Phillimore et al. (2012, Fig. 2H) and Ison and Wagenius (2014) suggest positive, albeit nonsignificant, correlations between slopes and intercepts (B. Inouye, *personal observation* from analyses of data on Dryad). Although several studies have documented individual variation in slopes of phenological reaction norms (e.g., van Asch et al. 2007, Reed et al. 2009), we found no studies where authors have tested correlations of slopes and intercepts for phenological traits at this level. In contrast, patterns that have been reported for species-level reaction norms suggest positive correlations may be common, as species that have earlier mean dates of activity (lower intercept) also tend to be more sensitive to climate (more negative slope, e.g., Brooks et al. [2017] for butterflies, Hülber et al. [2010] and Wolkovich et al. [2012] for plants).

Using Phenological Distributions and Reaction Norms to Answer Questions About Phenology

Three fundamental types of questions are typically asked about phenology: what timing of events in an individual's life maximizes fitness (e.g., Valdés and Ehrlén 2017), how phenology influences biotic interactions (e.g., Reed et al. 2013), and how phenology changes over time or space (e.g., Inouye et al. 2000). Each of these questions uses phenological metrics in somewhat different ways. A growing number of papers specifically address phenological metrics and limits to their interpretation (e.g., Miller-Rushing et al. 2008, Moussus et al. 2010, Strebel et al. 2014, Chapman et al. 2015, Lindén 2018). With new metrics appearing and work on metrics for different types of organisms (principally birds and plants) proceeding independently, it is not always clear which are best to use for a particular question (e.g., Moussus et al. 2010, Thackeray et al. 2012, Pearse et al. 2017). In this section, we discuss how considering full distributions of phenological traits and their underlying reaction norms can help researchers think critically about the ideal metrics for the focal question of their study, and how to interpret less-than-ideal phenological metrics when data are limited.

Questions about natural selection on phenology

Studies of natural selection on phenology are motivated by a desire both to improve our fundamental understanding of life history and to predict evolutionary responses to environmental change; full phenological distributions rather than single dates could provide greater insight into these questions in two ways. First, while studies often focus on phenological events that represent important life-history landmarks (e.g., onset of reproduction), other events (e.g., peak reproduction) or more integrative measures of the phenological distribution, such as duration or skew, could also be key targets of selection. If data are available for a full phenological distribution, one could include multiple facets of the distribution as predictors in a regression model and determine where selection acts most strongly in a given study system; without a full phenological distribution, many interesting hypotheses are untestable. Data on phenological distributions for individuals within a population would also allow testing for frequency-dependent selection on phenological traits, i.e., effects of intraspecific synchrony (Ison and Wagenius 2014, Weis et al. 2014). The timing of reproductive effort relative to conspecifics can have large effects on fitness, but synchrony can't be estimated well without full phenological distributions.

Because phenological traits are plastic (e.g., Anderson et al. 2012, Vedder et al. 2013, Charmantier and Gienapp 2014), selection on long-lived organisms (or genotypes) actually acts on the norm of reaction (Ehrlén 2015). Lifetime selection is the accumulated result of selection acting on the phenotypes expressed in each environment and across years differing in environmental conditions. Although we expect selection to act strongly on the reaction norms of phenological traits (Chevin et al. 2010, but see Oostra et al. 2018), data on this are sparse across all types of organisms, even for the relatively well-studied topic of plant flowering time (Ehrlén 2015). Unfortunately, the relationship between within-season phenotypic selection on phenology (which is relatively easy to measure) and selection on reaction norms (much harder to measure) is not straightforward. Consistent yearly relationships between a phenology metric and fitness can actually cause inconsistent selection on underlying reaction norms. For example, consistent selection for early flowering times could favor individuals with a later intercept and less environmental sensitivity in a cooler year, but earlier intercepts and greater sensitivity in warmer years (as in Fig. 2C). Similarly, different selection gradients in different years could correspond to consistent selection on reaction norms. Only if reaction norms are parallel, an unlikely situation, do we expect observed patterns of selection on phenology to correspond exactly to selection on the underlying reaction norms. To understand better how natural selection acts on phenology it is thus critical to have more data on individual phenological reaction norms. If we have data on the phenology of individuals from multiple years, or experimental data from replicated genotypes or family lines, then we can start to explore how phenological reaction norms vary among individuals and how such variation influences natural selection on phenology.

Questions about effects of phenology on inter- and intraspecific interactions

Many important ecological and evolutionary processes depend on the relative phenologies of interacting species or individuals within a population. For example, consumers must overlap in time with their resources, plants must overlap with pollinators, and individuals must overlap in time with suitable mates to facilitate reproduction. Many studies have thus focused on how phenology influences species interactions or access to mates (e.g., Tikkanen and Julkunen-Tiitto 2003, van Asch et al. 2007, Yang and Rudolf 2009, Rasmussen et al. 2014, Bolmgren and Eriksson 2015, Wadgymar and Weis 2017), and especially how changes in climate might change the degree of synchrony of interacting species. If interacting species use different cues, for example one being sensitive to temperatures and another to photoperiod, mismatches may become increasingly common.

The intensity of an interaction between two species at a given point in time depends on abundances of individuals in developmental phases during which interaction is possible. The interaction intensity across a season will thus be a function of how relative abundances vary over a season. To assess how intensities of interactions are influenced by phenology, we ideally need to quantify the distributions of phenological events over the season in each of the interacting species (Wadgymar and Weis 2017, Lindén 2018). By using full phenological distributions and looking at interaction intensity across the season, we can also begin to recognize critical periods for species interactions, or thresholds for interactions, such as periods of resource scarcity (Fig. 1C), which can be more important for consumers than peak abundances (Ogilvie et al. 2017). Using full distributions can also provide us with information about the phenology of an individual relative to phenological events in the other interacting partner(s) (Forrest and Thomson 2010, Wadgymar and Weis 2017), which will provide a better basis for understanding how natural selection mediated by multiple species interactions acts on traits associated with phenology (Bolmgren and Eriksson 2015).

Using full distributions can also improve our ability to answer the popular question of whether climate change leads to increased phenological mismatches. Previous studies have mostly quantified phenological mismatches by testing for divergence in the first event dates for different species, although some work instead uses estimates of median or peak dates (e.g., Reed et al. 2013, Kharouba and Vellend 2015). However, conclusions about the strength or existence of a phenological mismatch can depend on the choice of metric, particularly if the shapes of phenological distributions are sensitive to climate (Lindén 2018). Although we have few data to say whether distributions change shape with environment, they do seem likely to change. For example, if individual reaction norms have correlated intercepts and slopes then we should expect population-level phenological distributions to have different durations in different years or environments (See Questions about how phenology changes through time or across environments). Furthermore, those studies that have compared phenological responses of both first and last event dates generally find that these change at different rates or have different climate sensitivities, implying that the duration of the phenological distribution must change (e.g., Thackeray et al. 2012, Iler et al. 2013, Austen et al. 2014, CaraDonna et al. 2014, Miles et al. 2016). Fig. 3 illustrates one way in which estimates of a phenological mismatch between two species depends on the choice of metrics, if the duration of the phenological distribution also responds to climate. Even when duration remains constant and reaction norms are linear, overlap metrics based on abundances can show nonlinear responses to climate due to abundances increasing and then decreasing through a season. The fact that estimates of mismatch based on first, last, or peak dates will differ from estimates based on absolute or proportional temporal overlap, even in sign, may help explain why meta-analyses of phenological mismatches (Thackeray et al. 2016, Kharouba et al. 2018) that include divergence in different metrics for different species have found variable results.

Given that different metrics of overlap will give different answers, it is critical for researchers to state biological hypotheses that justify the use of particular phenological mismatch metrics for their particular system and question. For example, for insects needing new leaves to begin development the most appropriate metric might be synchrony of the hatching period and the period when young leaves are available, not only synchrony of first hatching and first budburst dates. When the only data available are first event dates (e.g., Bolmgren et al. 2013), special care should be given to interpretation of results because using only the sensitivities of first event dates to estimate phenological mismatches is unlikely to capture biologically important aspects of many species interactions, limiting the strength of inferences from past studies. Ultimately, understanding the impact of phenological mismatches, regardless of how carefully mismatches are quantified, will require studying the relationship between the degree of mismatch and mean absolute fitness or population growth rate (Visser and Both 2005, Miller-Rushing et al. 2010, Bolmgren and Eriksson 2015, Johansson et al. 2015, McLean et al. 2016). When the degree of overlap affects fitness and the phenologies of interacting species are differently sensitive to changing environmental conditions, or use different types of cues (e.g., van Asch et al. 2007, Kharouba and Vellend 2015), we can also expect changes in natural selection on traits associated with the timing of development in interacting partners. For example, Fogelström et al. (2017) showed that differences in relative phenology of a perennial herb and its seed predator influenced the magnitude and direction of selection on flowering phenology.

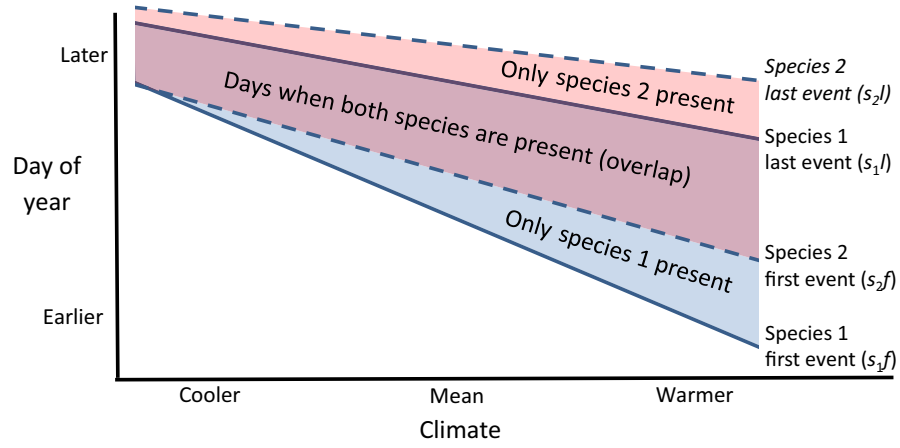


FIG. 3. An illustration that estimates of the phenological mismatch between two species can depend on the choice of mismatch metric. In this figure, the climate sensitivities of first and last event dates are not equal and can be different for two interacting species. The two species have similar phenologies in cooler climates but divergent phenologies in warmer climates, and the duration of phenological distributions increases with warmer temperatures in both species. In other words, the two species differ in climate sensitivities, i.e., the change in day of year for first and last dates (s_if and s_il, respectively) per change in climate. Phenological divergence (i.e., mismatch) could be defined using differences in first event dates ($s_1f - s_2f$), last event dates ($s_1l - s_2l$), peak event dates ($[(s_1f + s_1l) - (s_2f + s_2l)]/2$ if abundance distributions are symmetric), absolute amount of overlap, or proportional overlap. In this figure, each of these divergence metrics has a different value. Considering, for example, an activity that occurs during the shaded periods between first and last event dates, it is clear that the strength and even presence of a phenological mismatch will depend on how the mismatch is quantified. Mismatches based on first, last, or peak dates will differ from mismatch metrics based on absolute or proportional temporal overlap and can differ even in sign. If $s_1l > s_2f$, as in the example shown here, then absolute overlap and proportional overlap can increase while first event dates diverge, so metrics of phenological mismatch based on overlap vs. first event dates differ in sign as well as magnitude. If $s_1l = s_2f$ then the absolute amount of overlap will be constant while other metrics diverge.

Questions about how phenology changes through time or across environments

There is much current interest in understanding how phenologies have changed over time, presumably due to shifts in climate, and in predicting how species' phenologies may change in the future. In this section, we first consider how this understanding can be improved by considering full phenological distributions and norms of reaction, and then describe results of a simulation model including individual variation in reaction norms, which suggests caution in using first event dates to estimate climate sensitivities.

Most studies examining the effects of climate warming have focused on changes in single events like first or peak dates, often due to data constraints, although we know that phenological distributions can respond to environmental variation in multiple ways. For example, the onset and duration of a period of activity can change in different ways (e.g., CaraDonna et al. 2014, Miles et al. 2016). If the shapes of phenological distributions change from year to year (Forrest and Thomson 2010), then relying on single event metrics can give an incorrect picture of how changing environments will affect phenologies, species interactions, and selection on phenology (Kopp and Cleland 2015).

Because most researchers understand the importance of considering full phenological distributions, several papers have proposed methods for estimating population- or species-level phenological distributions from individual observations, including those from herbaria or citizen science projects (e.g., Malo 2002, Lindén 2011, Austen et al. 2014, Strebel et al. 2014, Cayton et al. 2015, Chapman et al. 2015, Pearse et al. 2017). For example, Bayesian methods make it possible to include information on temporal variation in observer effort and detection probabilities (i.e., Lindén 2011, Chapman et al. 2015). These methods generally assume a simple symmetric parametric shape for phenological distributions, which is a reasonable starting point, but more data on actual phenological distributions are needed to ensure that these methods are robust. The empirical phenological distributions that have been published are rarely symmetric (e.g., Forrest and Thomson 2010, Weis et al. 2014, Bolmgren and Eriksson 2015) and very few studies have tested for changes in the shape of phenological distributions (but see Austen et al. 2014). Estimating phenological distribution shapes in multiple environments would improve our ability to make robust predictions about the consequences of climate change.

Interpreting relationships between phenology and climate can also be improved by regarding phenology as a reaction norm rather than one year's specific trait values. Long-term trends in event dates, e.g., changes in first flowering dates, or dates for migration, hatching, or spawning, are the result of both genetic changes and phenotypic plasticity. Relatively few studies have

separated contributions of plasticity and evolution (but see, e.g., Réale et al. 2003, Anderson et al. 2012, Vedder et al. 2013, Charmantier and Gienapp 2014), but distinguishing these components will be critical for making better predictions about phenological response to climate change (Oostra et al. 2018). For traits related to phenology in temperate, alpine, or Arctic regions where the length of the growing season is a limiting factor, we might often expect a counter-gradient response to climate warming, i.e., that phenotypic and genotypic responses are in opposite directions (Levins 1969, Conover and Schultz 1995). To illustrate this, imagine a system where the optimal day of first flowering is a function of the benefits of being early, e.g., in terms of being able to mature seeds before autumn, and costs, e.g., in terms of an increased risk of frost damage if flowering early. Plastic changes in floral development phenotype will occur with a warming climate along the species' reaction norm, because development is faster and flowering is earlier with increasing temperatures. However, with a warming climate we also expect that growing seasons will become longer. Thus, the cost associated with not maturing seeds before the autumn will be smaller, resulting in selection to decrease development rates and flower later at a given temperature, i.e., to shift the intercept of the thermal reaction norm of development downwards. Due to the potential for evolution of reaction norms and the frequency-dependent nature of selection for synchrony, the results from short-term studies of selection on phenologies (i.e., finding directional selection for first event dates in a single field season) are much less predictive of future responses to a warming climate than one might wish. Measuring reaction norms of individuals will help us understand how phenotypes may change in response to climate change and interpret observed patterns of change.

Phenological sensitivity to climate is often estimated using first events, and although there are known problems with using first events for this purpose we argue that considering the full distributions of phenological events and viewing phenology as a reaction norm suggest additional concerns for how best to estimate climate sensitivity. Previous authors generally advocate the use of population mean, median, or peak dates instead, or a quantile (i.e., 5% or 10% along an estimated phenological distribution, e.g., Brooks et al. 2017), emphasizing the statistical sampling issues associated with minimum and maximum values drawn from any probability distribution (e.g., Moussus et al. 2010, Strebel et al. 2014). They have also pointed out that first dates are likely to be more sensitive than means or peak metrics due to differences in population size (Miller-Rushing et al. 2008) as well as more sensitive to differences among years in observer effort or detection biases (Chapman et al. 2015). In addition, studies have shown that for plants (e.g., CaraDonna et al. 2014), birds (Goodenough et al. 2015, Miles et al. 2016), insects (Iler et al. 2013), and a *Daphnia* species (Thackeray et al. 2012), a population's temporal trends in first event dates can be different from the temporal trends in other phenological metrics due to changes in phenological distributions among seasons (c.f. Fig. 3). This may, in turn, be the result of individual variation in phenology reaction norms and correlation between slopes and intercepts of reaction norms.

To explore how using first event dates can provide biased estimates of a species' climate sensitivity, we used a simulation model that estimates climate sensitivity using both first dates and mean dates for systems with different degrees of correlation between slopes and intercepts of individual reaction norms. The simulation (detailed description and annotated R code in Appendix S1) illustrates two important differences between estimates of phenological sensitivity based on first event dates and those based on other metrics. First, and most importantly, if the slope and intercept of individual reaction norms are correlated, the estimated population sensitivities can be very different depending on whether they are estimated from the first date in the population or a population average. When positively correlated (as in Fig. 4A), the estimated sensitivity from first-in-population data will be more negative (generally indicating greater sensitivity) than the estimate from the mean of all individuals, and vice versa. In general, a sensitivity estimated from the mean of all individuals is close to the value used to generate the simulation data. However, when individuals differ in reaction norm slopes, different individuals can claim the title of first in different years (Fig. 4B). The same principle applies when estimating community sensitivities when there is variation in slopes for the component species, such as for green-up rates in remotely sensed data; fortunately, satellite data can provide full phenological distributions (Helman 2018). When the slope and intercept are strongly correlated, the estimated sensitivities using first dates can deviate quite far from the population mean estimates. In the particular run of the simulation shown in Fig. 4B, the upper 95% confidence interval even includes results where first dates suggest later phenology with a warming climate, even though the average individual in the population has the opposite response, i.e., a potential "Type S" error where not only the magnitude of the estimated response is wrong but also the sign of the response (Gelman and Carlin 2014).

A second insight from this simulation is that the range of sensitivity estimates is much larger when using a single first event date for the entire population (Fig. 4B). This reflects both the fact that fewer data are used, only one data point per year instead of population means from many individuals, and the fact that maxima and minima have sampling distributions with much larger variance than does the mean (Forbes et al. 2010). In fact, the bias and the range of sensitivity estimates from first event dates are *smaller* in a very small population (results not shown), because the likelihood of observing an extremely early individual is reduced. The counterintuitive pattern that statistical performance decreases as

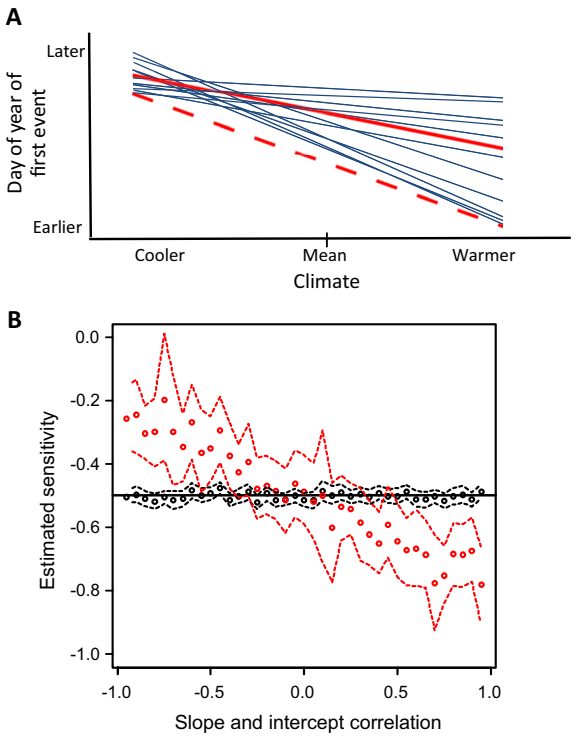


FIG. 4. (A) A simplified example where the slope and intercept of idealized individual reaction norms (thin black lines) vary among individuals and there is a positive correlation between slope and intercept, such that individuals with high mean trait values are also less sensitive to climate. Estimates of phenological sensitivity based on a population mean or peak (solid red line) will be less sensitive to climate than estimates based on the first phenological event in the population (dashed red line). Note that, in this example, the identity of the earliest individual in the population would be different in years with different climate conditions. (B) The results from the simulation model show that estimates of climate sensitivity from first event dates for a population or species depend on the correlation between slopes and intercepts of reaction norms of individuals (red points and dashed 95% confidence intervals), while estimates of sensitivity based on the mean date of an event in the population (black dots and dashed 95% confidence intervals) are not sensitive to this correlation. See Questions about how phenology changes through time or across environments for further explanation. Parameter values for Fig. 4B and code to recreate it are in Appendix S1. These simulation results are qualitatively robust to variation in the number of individuals in the population, the number of years, and the type of probability distribution and variance used to create variation among years or realized first event dates (results not shown).

more individuals are sampled provides another indication that when possible one should avoid estimating population sensitivities from first-in-population data.

The idealized example in Fig 4A also illustrates a new prediction: if individual slopes and intercepts are positively correlated, then the variance in individual phenologies will be much lower in cooler years than in warmer years (Fig. 4A). In cooler years, all individuals in the population achieve a phenological state on a later date than in warmer years, and all the individuals in the population also display similar first event dates. This pattern is consistent with previous work finding longer phenological durations in years with earlier first event dates (e.g., first event dates are more sensitive than peak or last event dates; Miles et al. 2016). Variance in phenology has important implications for outcomes such as reproduction, where synchrony with conspecifics might be key, and resource acquisition, where higher variance in phenology of resource species might provide more consistent resources for consumers.

The Developmental Perspective on Measuring Phenology: Phenology in Snapshots

Phenology is most often recorded by following individuals or populations over time and recording the dates at which pre-defined events occur. Because variation in the date of an event is closely linked to variation in developmental rates, we could instead record the developmental stages of individuals in a population at a single date and use data on the distribution of developmental stages to infer differences in timing among individuals (e.g., Etterson and Shaw 2001), or differences in average timing for populations or species. Many species have predictable developmental progressions, for example from buds to flowers to fruits, from one larval insect instar to the next, or from one anuran Gosner stage to the next. These developmental stages can most simply be described as ordered categories ("phenophases"; Denny et al. 2014), for example by recording insect larval instar, or recording individual plants as being "in bud" vs. "in flower" (e.g., Hülber et al. 2010). Other than the endpoints (i.e., eggs and senescence), the proportion of a population in a single stage will first increase and then decrease, thus the proportion of a population in a single stage on a single date can't unambiguously inform us about phenological progression. However, one could quantify development on a continuous scale for individuals by measuring a phenological score that takes into account multiple aspects of development (e.g., buds, open flowers, and senesced flowers; Dambreville et al. 2015), or sizes of the largest buds in a pre-flowering population. For populations, one could use measurements such as the proportion of individuals that have achieved a developmental milestone (which only increases through time), or the ratios of multiple developmental stages such as larval instars (Murtaugh et al. 2012).

Continuous phenological scores are preferable, in part because they permit a greater range of statistical analyses. Discrete phenophases are on an *ordinal* scale, and thus should only be analyzed with methods designed for ordinal data such as ordinal regression, time-to-event models (e.g., Hülber et al. 2010) or hidden Markov chain models (e.g., Aurdal et al. 2005). Previous analyses that have recorded numbered phenophases on a scale yet used analyses designed for continuous data are logically flawed; most familiar statistical tests such as linear models require data on an interval or ratio scale (an explanation of measurement theory is

beyond the scope of this paper, but see, e.g., Wolman 2006, Houle et al. 2011). For animals that can only express a single developmental stage at a time (e.g., insect larval instars), one can estimate a continuous population-level phenological metric by recording the proportion of a population in each developmental stage on a particular date (e.g., Murtaugh et al. 2012). Note that one could even express the migration state of birds on a given date by looking at the distribution of geographic locations of individuals on that day; typical studies of the phenology of bird migrations invert this relationship and instead focus on the date upon which individuals arrive at (or pass by) a fixed geographic location. With the advent of geographically dispersed citizen science data, it might be possible to map population distributions in a brief time window, adding to what we have learned from data collected at fixed locations such as bird banding stations and observatories (e.g., Ovaskainen et al. 2013, Miles et al. 2016).

Characterizing phenology with developmental scores has both disadvantages and potential advantages. The main potential disadvantages are the lack of information on abundance or intensity of events, and the limited appropriate statistical analyses when data are on an ordinal scale. One potential advantage is that one could assess differences in phenology among individuals or populations in a given year with just a single sampling date, although multiple samples might still be preferable. In addition to comparing individuals or populations at a single date, one could make comparisons across years by sampling at a consistent date each year. Second, estimates of underlying phenology reaction norms based on continuous developmental phenological scores at a given date may be less sensitive to variation in temperature within seasons than those based on dates when individuals pass a threshold. If growing degree days accumulate in fits and starts then some individuals could pass a threshold (i.e., change developmental stage) during a hot spell, but then have stasis during a cold spell. For example, picture two plant individuals with multiple buds that differ only slightly in development, one plant being slightly more advanced than the other. During a warm day the more advanced plant may open its first flower, followed by cooler days during which neither plant opens additional flowers. The two individuals would remain similar in their overall developmental progression, but one will have passed a common phenological landmark. If we then use the date of flowering as our measure of phenology, the small developmental difference among individuals will appear as a larger difference in dates caused by the cold spell. A continuous measure of development recorded for all individuals on the same date will instead provide consistent differences among individuals, assuming that phenological scores have a roughly linear relationship with the environmental variable of interest (e.g., development increases with cumulative growing degree days), and that individuals have similar developmental rates (i.e., slopes of this relationship).

A third potential advantage arises from the fact that while natural selection on phenology is often examined by considering the effects of selective agents acting after a given phenological event has taken place, agents acting before the event might also strongly influence selection. For example, selection on flowering phenology is usually studied by examining the effects of first flowering date on a fitness proxy such as seed number (equivalently, effects of laying date on fledging success). However, selection on flowering phenology might also be mediated by factors acting before flowers have opened, such as late frosts or grazing (Inouye 2008). Omitting the effects of these factors might result in a strongly biased estimate of selection on phenology and development schedules. This is an example of the problem of "the invisible fraction," i.e., that episodes of selection on a trait that occur before the trait is expressed are invisible (e.g., Bennington and McGraw 1995). If instead of events through time we record phenological scores early in a season, then one could alleviate problems associated with assessing this "invisible" selection on phenology. To do this, we need to assume that differences expressed early in development correspond to similar differences in later traits, for example that earlier egg-laying corresponds to earlier hatching dates. This assumption is not valid if the rate of development differs much among genotypes and environments. However, if differences among individuals in later development rates are small relative to developmental differences observed at an early recording date, or if the fraction of individuals that fail to express later phenological stages is high, then measuring early phenological scores will result in a less biased estimate of selection than estimates based only on the individuals that do express later stages. These assumptions may seem restrictive, but we think they are likely to hold true for many organisms. Moreover, responses of phenological scores on a fixed date to shifts in climate are aspects of a multivariate reaction norm. To the extent that the slope of this reaction norm remains the same through a season, evaluating selection on developmental scores early in a season will provide a better estimate of overall selection on phenology than using event dates of later stages with a substantial invisible fraction.

Future Directions and Conclusions

In this paper, we have found that acknowledging full phenological distributions and norms of reaction can inform how we interpret phenological metrics to answer key research questions. For example, inferences about phenological mismatches can be strongly affected by the choice of metrics (i.e., divergence in first event dates vs temporal overlap in abundances; Lindén 2018), when individual and population phenological distributions change shape in different climates. When individuals vary in their phenological reaction norms with respect to

climate, and particularly when the slopes and intercepts of reaction norms covary (Fig. 4A), we should expect population sensitivities of first event dates to differ from the mean population sensitivity (Fig. 4B). In addition to helping us think about existing work on phenology, this perspective of focusing on the underlying nature of phenology as a process, rather than focusing on measuring phenology in terms of dates, suggests important avenues for future work.

First, we need data on phenological variation through seasons and across environments (Denny et al. 2014, Elmendorf et al. 2016). Given the key role of the shape of phenological distributions in determining the potential for species interactions, studies are needed to document full distributions of phenological events across seasons in multiple years, with sufficient temporal resolution for robust tests of distribution shapes. Additional data on distributions for diverse species will allow us to evaluate the common assumptions that phenological distributions are approximately Gaussian and have constant variance, since studies to date suggest that phenological distributions are commonly skewed and their variances change as climates shift. Full distributions can also provide more meaningful measures of mismatch or overlap, although ultimately it is also necessary to link overlap to measures of fitness to determine ecological and evolutionary consequences (McLean et al. 2016). It is also critical to collect data on phenological reaction norms to have a better understanding of potential evolutionary responses, and to know whether slopes and intercepts are commonly correlated. Knowledge of the frequency and type of slope–intercept correlation would allow us to estimate biases in historical data sets that contain only first event dates.

Second, we need more phenological data at the level of individuals. Despite the challenge of recording individual-level data, they would be helpful both for estimating reaction norms, and for quantifying within-population variance in phenology and measures of synchrony (Forrest and Thomson 2010, Ison and Wagenius 2014, Gerst et al. 2017, Wadgymar and Weis 2017), giving us insight into how individual traits scale up to affect first event dates, peak dates, duration and other metrics at the population level. As we demonstrate, metrics based on first event dates can be strongly biased. Without the context provided by estimates of the magnitude of variation in individual phenologies, we can't be confident that previous work on phenological trends based on first-event-in-population data accurately represents true population trends.

Third, in acquiring new data on phenological distributions and reaction norms we need to consider how to include organisms that skip phenological events in a given year, such as perennial plants that don't flower in a given year or birds that skip a year of breeding. This will most likely be a problem for data on individuals because, at the population or species level, it is unlikely that every individual will skip a year (although we have observed this for some plants; B.D. Inouye and N. Underwood, *unpublished data*). A correlation between individuals' reaction norm slopes or intercepts and the probability of skipping a year would affect the relationship between individual and population climate sensitivities, and complicate expectations for how populations may evolve in response to climate change. For example, if late-flowering individuals are more likely to skip flowering in warm years, a population phenology will show climate sensitivity even if individual flowering times are not plastic. If skipped years are common in a population one could model phenology as a two-stage decision process, such as a hurdle model for whether an individual reproduces, and if so, when.

Fourth, as more complete data on phenology become available (and for the few cases where they already exist), researchers will need new statistical methods to make full use of them. While most previous analyses of phenology are univariate analyses of event dates (i.e., regressions of dates against climate or time), phenological distributions are actually complex objects that could be described by a vector of metrics (i.e., first, mean, and last dates, peak abundance, skew, etc.). To analyze these multivariate objects, one could apply methods that quantify the shape of phenological distributions (Malo 2002, Austen et al. 2014), but more promising approaches may be to adapt methods from function-valued trait analysis (Kingsolver et al. 2001, Stinchcombe and Kirkpatrick 2012), to apply methods using response vectors (Collyer et al. 2014, De Caceres et al. 2019), or to develop novel approaches based on tools such as Hellinger distances or multivariate analyses of parametric distributions. These tools would allow one to determine, for example, how duration or skew changes with shifts in peak event dates. We also note that when applying methods that estimate distributions from limited data (e.g., Lindén 2011, Strebel et al. 2014, Chambert et al. 2015, Pearse et al. 2017), it is important to account for parameter uncertainty, as reconstructed phenology metrics can lead to problems with "stats-on-stats" that can greatly inflate error rates in downstream analyses (Brommer 2013, Houslay and Wilson 2017). For any type of analysis, natural history knowledge and non-phenological data are essential for interpreting phenological distributions. For example, basic ecological theory tells us that, with a saturating functional response, consumers will become overwhelmed by the abundance of a resource—birds can't eat all the caterpillars at peak abundance, bees can only visit so many flowers in a day—so a singular focus on the timing of peak abundance may be misplaced. When phenology is visualized as a multidimensional trait that changes in response to the environment, varying across years and environments, the complexity is intimidating, yet acknowledging this complexity can lead to more robust insights.

A final recommendation for future work is that when working with existing studies for reviews and meta-analysis we should avoid grouping different phenological

metrics into single analyses. Reviews often combine studies that have used different measures of phenology, such as first events and peak events (Ovaskainen et al. 2013, Brown et al. 2016, Thackeray et al. 2016, Kharouba et al. 2018). We find that different phenological metrics can give very different estimates of responses to climate even when calculated from a single population (Figs. 3, 4B), which perhaps explains some of the heterogeneity in phenological responses to climate observed in reviews. For some kinds of events, it is impossible to identify a first event date, making firsts less broadly useful; only migratory species have first arrival dates at a field site, whereas all species could have a temporal peak in activity patterns such as singing, mating, or nest building (Strebel et al. 2014). First event dates are also a poor metric for phenophases that do not have a single uninterrupted period, but stop and restart, such as plants that flower sporadically (Gerst et al. 2017). Peak event times (e.g., mating displays for mammals, calling times for frogs, peak flowering times for plants) should be more comparable than disparate metrics such as first flowering for plants vs peak spawning activity for fish, although even peaks will not always be comparable across phenologies at different levels of biological organization. Many phenological studies focus on developmental events, but some focus on the timing of change in numbers of individuals in a population (e.g., the phenology of zooplankton density; Thackeray et al. 2012) and still others use metrics that are a mixture of population growth and individual development (e.g., normalized difference vegetation index [NDVI] or enhanced vegetation index; Helman 2018). As multiplicative processes, the biological mechanisms affecting the timing of peak population sizes are likely to be different from those affecting peaks in individual phenological activity. Reviews and meta-analyses that mix data from different biological processes should at minimum include the type of phenological metric as a covariate in analyses.

Collecting data for full phenological distributions and reaction norms is logistically challenging, but there are reasons for optimism. Keeping track of the individuals from which phenological data are collected in multi-year studies could be relatively straightforward in some systems and would enable us to assess not only among-year variation in phenology but also the reaction norms of individuals. Recording phenological data on ratios of developmental stages could allow more detailed analyses with less of a time investment, in systems where this approach is appropriate. Studies that have partial information on phenological distributions or individual variation can acknowledge phenology as a process, to make better choices and interpretations of metrics. For example, studies of museum collections (e.g., Chapman et al. 2015) and citizen science projects often have valuable data on some aspects of phenological distributions, but either rely on estimates of first events or make unnecessarily simplistic assumptions about shapes of phenological distributions and reaction norms. These pitfalls could be avoided by relying less on first event dates, and allowing for phenology distributions to vary among years (Phillimore et al. 2012). Studies with data sets that provide only records of first events, such as the personal notes of naturalists (e.g., Miller-Rushing et al. 2008, Bolmgren et al. 2013) can continue to provide powerful indications of biotic responses to climate change, while acknowledging their limited power to predict how phenologies will change in the future or to infer potential changes in species interactions.

Studies of phenology address varied questions about evolution, ecological interactions and responses to climate change, and span levels of organization from parts within individuals to ecosystems (Forrest and Miller-Rushing 2010). The development of metrics for phenological analyses has proceeded somewhat differently among studies of different organisms, particularly birds and plants, hindering a more general consideration of the fundamental nature of phenologies. Although the appropriate metrics and analyses for any particular study will vary widely, the field of phenology as a whole could benefit from a greater appreciation of how phenological distributions at each level of biological organization are ultimately a summation of the phenologies of individuals, and how phenological events are shaped by underlying reaction norms. Taking a broader and yet more mechanistic view can improve our basic understanding of phenology and our predictions for responses to climate change in all types of biological systems.

Acknowledgments

The development of these ideas was supported in part by NSF DEB-1354104 to N. Underwood and B. D. Inouye and a Wenner-Gren Stiftelserna Gästforskarstipendier to N. Underwood. All authors contributed equally to the development of ideas in this paper. We thank O. Eriksson, M. Lessard-Therrien, A. Weis, and an anonymous reviewer for helpful comments, and K. Bolmgren for discussion.

Literature Cited

Anderson, J. T., D. W. Inouye, A. M. McKinney, R. I. Colautti, and T. Mitchell-Olds. 2012. Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. Proceedings of the Royal Society B 279:3843–3852.

Aurdal, L., H. Ragnar Bang, L. Eikvil, R. Solberg, D. Vikhamar, and A. Solberg. 2005. Use of hidden Markov models and phenology for multitemporal satellite image classification: applications to mountain vegetation classification. Pages 220–224 *in* International Workshop on the Analysis of Multi-Temporal Remote Sensing Images. IEEE Computer Society, Washington, DC, USA.

Austen, E. J., D. A. Jackson, and A. E. Weis. 2014. Describing flowering schedule shape through multivariate ordination. International Journal of Plant Sciences 175:70–79.

Beebee, T. J. C. 1995. Amphibian breeding and climate. Nature 374:219.

Bennington, C. C., and J. B. McGraw. 1995. Phenotypic selection in an artificial population of *Impatiens pallida*: the importance of the invisible fraction. Evolution 49:317–324.

Bolmgren, K. 1998. The use of synchronization measures in studies of plant reproductive phenology. Oikos 82:411–415.

Bolmgren, K., and O. Eriksson. 2015. Are mismatches the norm? Timing of flowering, fruiting, dispersal and germination and their fitness effects in *Frangula alnus* (Rhamnaceae). Oikos 124:639–648.

Bolmgren, K., D. Vanhoenacker, and A. J. Miller-Rushing. 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.

Brommer, J. E. 2013. Phenotypic plasticity of labile traits in the wild. Current Zoology 59:485–505.

Brommer, J. E., P. Kontiainen, and H. Pietiäinen. 2012. Selection on plasticity of seasonal life-history traits using random regression mixed model analysis. Ecology and Evolution 2:695–704.

Brooks, S. J., A. Self, G. D. Powney, W. D. Pearse, M. Penn, and G. L. J. Paterson. 2017. The influence of life history traits on the phenological response of British butterflies to climate variability since the late-19th century. Ecography 40:1152–1165.

Brown, C. J., et al. 2016. Ecological and methodological drivers of species' distribution and phenology responses to climate change. Global Change Biology 22:1548–1560.

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

Cayton, H. L., N. M. Haddad, K. Gross, S. E. Diamond, and L. Ries. 2015. Do growing degree days predict phenology across butterfly species? Ecology 96:1473–1479.

Chambert, T., W. L. Kendall, J. E. Hines, J. D. Nichols, P. Pedrini, J. H. Waddle, G. Tavecchia, S. C. Walls, and S. Tenan. 2015. Testing hypotheses on distribution shifts and changes in phenology of imperfectly detectable species. Methods in Ecology and Evolution 6:638–647.

Chapman, D. S., S. Bell, S. Helfer, and D. B. Roy. 2015. Unbiased inference of plant flowering phenology from biological recording data. Biological Journal of the Linnean Society 115:543–554.

Charmantier, A., and P. Gienapp. 2014. Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. Evolutionary Applications 7:15–28.

Chevin, L.-M., R. Lande, and G. Mace. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biology 8:e1000357.

Chmura, H., H. Kharouba, J. Ashander, S. Ehlman, E. Rivest, and L. Yang. 2019. The mechanisms of phenology: the patterns and processes of phenological shifts. Ecological Monographs. https://doi.org/10.1002/ecm.1337.

Cleland, E. E., N. R. Chiariello, S. R. Loarie, H. A. Mooney, and C. B. Field. 2006. Diverse responses of phenology to global changes in a grassland ecosystem. Proceedings of the National Academy of Sciences USA 103:13740–13744.

Cleland, E. E., I. Chuine, A. Menzel, H. A. Mooney, and M. D. Schwartz. 2007. Shifting plant phenology in response to global change. Trends in Ecology & Evolution 22:357–365.

Collyer, M. L., D. J. Sekora, and D. C. Adams. 2014. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. Heredity 115:357.

Conover, D. O., and E. T. Schultz. 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. Trends in Ecology & Evolution 10:248–252.

Crawley, M. J., and M. Akhteruzzaman. 1988. Individual variation in the phenology of oak trees and its consequences for herbivorous insects. Functional Ecology 2:409–415.

Dambreville, A., P.-É. Lauri, F. Normand, and Y. Guédon. 2015. Analysing growth and development of plants jointly using developmental growth stages. Annals of Botany 115:93–105.

De Caceres, M., L. Coll, P. Legendre, R. Allen, S. Wiser, M.-J. Fortin, R. Condit, and S. Hubbell. 2019. Trajectory analysis in community ecology. Ecological Monographs. https://doi.org/10.1002/ecm.1350

Demaree, G. R., and T. Rutishauser. 2011. From "Periodical Observations" to "Anthochronology" and "Phenology"—the scientific debate between Adolphe Quetelet and Charles Morren on the origin of the word "Phenology". International Journal of Biometeorology 55:753–761.

Denny, E. G., et al. 2014. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. International Journal of Biometeorology 58:591–601.

Ehrlén, J. 2015. Selection on flowering time in a life-cycle context. Oikos 124:92–101.

Elmendorf, S. C., et al. 2016. The plant phenology monitoring design for The National Ecological Observatory Network. Ecosphere 7:e01303.

Elzinga, J. A., A. Atlan, A. Biere, L. Gigord, A. E. Weis, and G. Bernasconi. 2007. Time after time: flowering phenology and biotic interactions. Trends in Ecology & Evolution 22:432–439.

Etterson, J. R., and R. G. Shaw. 2001. Constraint to adaptive evolution in response to global warming. Science 294:151–154.

Fogelström, E., M. Olofsson, D. Posledovich, C. Wiklund, J. P. Dahlgren, and J. Ehrlén. 2017. Plant-herbivore synchrony and selection on plant flowering phenology. Ecology 98:703–711.

Forbes, C., M. Evans, N. Hastings, and B. Peacock. 2010. Extreme value (Gumbel) distribution (Chapter 19). Pages 98–101 *in* C. Forbes, M. Evans, N. Hastings, and B. Peacock, editors. Statistical distributions. Fourth edition. John Wiley and Sons, Hoboken, NJ.

Forrest, J., and A. Miller-Rushing. 2010. Toward a synthetic understanding of the role of phenology in ecology and evolution. Philosophical Transactions of the Royal Society B 365:3101–3112.

Forrest, J., and J. D. Thomson. 2010. Consequences of variation in flowering time within and among individuals of *Mertensia fusiformis* (Boraginaceae), an early spring wildflower. American Journal of Botany 97:38–48.

Gelman, A., and J. Carlin. 2014. Beyond power calculations: assessing Type S (sign) and Type M (magnitude) errors. Perspectives on Psychological Science 9:641–651.

Gerst, K. L., N. L. Rossington, and S. J. Mazer. 2017. Phenological responsiveness to climate differs among four species of *Quercus* in North America. Journal of Ecology 105:1610–1622.

Goodenough, A. E., S. M. Fairhurst, J. B. Morrison, M. Cade, P. J. Morgan, and M. J. Wood. 2015. Quantifying the robustness of first arrival dates as a measure of avian migratory phenology. Ibis 157:384–390.

Helman, D. 2018. Land surface phenology: What do we really 'see' from space? Science of the Total Environment 618:665–673.

Herrera, C. M. 1995. Floral biology, microclimate, and pollination by ectothermic bees in an early-blooming herb. Ecology 76:218–228.

Houle, D., C. Pélabon, G. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. Quarterly Review of Biology 86:3–34.

Houslay, T. M., and A. J. Wilson. 2017. Avoiding the misuse of BLUP in behavioural ecology. Behavioral Ecology 28:948–952.

Hua, F., et al. 2016. Community-wide changes in intertaxonomic temporal co-occurrence resulting from phenological shifts. Global Change Biology 22:1746–1754.

Hülber, K., M. Winkler, and G. Grabherr. 2010. Intraseasonal climate and habitat-specific variability controls the flowering phenology of high alpine plant species. Functional Ecology 24:245–252.

Iler, A. M., D. W. Inouye, T. T. Høye, A. J. Miller-Rushing, L. A. Burkle, and E. B. Johnston. 2013. Maintenance of temporal synchrony between syrphid flies and floral resources despite differential phenological responses to climate. Global Change Biology 19:2348–2359.

Inouye, D. W. 2008. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology 89:353–362.

Inouye, D. W., B. Barr, K. B. Armitage, and B. D. Inouye. 2000. Climate change is affecting altitudinal migrants and hibernating species. Proceedings of the National Academy of Sciences USA 97:1630–1633.

Ison, J. L., and S. Wagenius. 2014. Both flowering time and distance to conspecific plants affect reproduction in *Echinacea angustifolia*, a common prairie perennial. Journal of Ecology 102:920–929.

Janhunen, M., J. Koskela, N. H. Ninh, H. Vehviläinen, H. Koskinen, A. Nousiainen, and N. P. Thỏa. 2016. Thermal sensitivity of growth indicates heritable variation in 1-year-old rainbow trout (*Oncorhynchus mykiss*). Genetics Selection Evolution 48:94.

Johansson, J., N. P. Kristensen, J.-Å. Nilsson, and N. Jonzén. 2015. The eco-evolutionary consequences of interspecific phenological asynchrony—a theoretical perspective. Oikos 124:102–112.

Kharouba, H. M., and M. Vellend. 2015. Flowering time of butterfly nectar food plants is more sensitive to temperature than the timing of butterfly adult flight. Journal of Animal Ecology 84:1311–1321.

Kharouba, H. M., J. Ehrlén, A. Gelman, K. Bolmgren, J. M. Allen, S. E. Travers, and E. M. Wolkovich. 2018. Global shifts in the phenological synchrony of species interactions over recent decades. Proceedings of the National Academy of Sciences USA 115:5211–5216.

Kingsolver, J. G., R. Gomulkiewicz, and P. A. Carter. 2001. Variation, selection and evolution of function-valued traits. Genetica 112-113:87–104.

Kopp, C., and E. Cleland. 2015. A range-expanding shrub species alters plant phenological response to experimental warming. PLoS ONE 10:e0139029.

Levins, R. 1969. Thermal acclimation and heat resistance in *Drosophila* species. American Naturalist 103:483–499.

Lindén, A. 2011. Using first arrival dates to infer bird migration phenology. Boreal Environmental Research 16:49–60.

Lindén, A. 2018. Adaptive and nonadaptive changes in phenological synchrony. Proceedings of the National Academy of Sciences USA 115:5057–5059.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates Inc, Sunderland, Massachusetts, USA.

Malo, J. E. 2002. Modelling unimodal flowering phenology with exponential sine equations. Functional Ecology 16:413–418.

Marquis, R. J. 1988. Phenological variation in the Neotropical understory shrub *Piper arielanum*: Causes and consequences. Ecology 69:1552–1565.

Martin, T. E. 1987. Food as a limit on breeding birds: a life-history perspective. Annual Review of Ecology Evolution and Systematics 18:453–487.

McLean, N., C. R. Lawson, D. I. Leech, and M. Pol. 2016. Predicting when climate-driven phenotypic change affects population dynamics. Ecology Letters 19:595–608.

Miles, W. T. S., et al. 2016. Quantifying full phenological event distributions reveals simultaneous advances, temporal stability and delays in spring and autumn migration timing in long-distance migratory birds. Global Change Biology 23:1400–1414.

Miller-Rushing, A., D. W. Inouye, and R. B. Primack. 2008. How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency. Journal of Ecology 96:1289–1296.

Miller-Rushing, A., T. T. Høye, D. W. Inouye, and E. Post. 2010. The effects of phenological mismatches on demography. Philosophical Transactions of the Royal Society B 365:3177–3186.

Morrisey, M. B., and M. Liefting. 2016. Variation in reaction norms: Statistical considerations and biological interpretation. Evolution 70:1944–1959.

Moussus, J.-P., R. Julliard, and F. Jiguet. 2010. Featuring 10 phenological estimators using simulated data. Methods in Ecology and Evolution 1:140–150.

Murren, C. J., et al. 2014. Evolutionary change in continuous reaction norms. American Naturalist 183:453–467.

Murtaugh, P. A., S. C. Emerson, P. B. McEvoy, and K. M. Higgs. 2012. The statistical analysis of insect phenology. Environmental Entomology 41:355–361.

Ogilvie, J. E., S. R. Griffin, Z. J. Gezon, B. D. Inouye, N. Underwood, D. W. Inouye, and R. E. Irwin. 2017. Interannual bumble bee abundance is driven by indirect climate effects on floral resource phenology. Ecology Letters 20:1507–1515.

Oostra, V., M. Saastamoinen, B. J. Zwaan, and C. W. Wheat. 2018. Strong phenotypic plasticity limits potential for evolutionary responses to climate change. Nature Communications 9:1005.

Ovaskainen, O., S. Skorokhodova, M. Yakovleva, A. Sukhov, A. Kutenkov, N. Kutenkova, A. Shcherbakov, E. Meyke, and M. D. Delgado. 2013. Community-level phenological response to climate change. Proceedings of the National Academy of Sciences USA 110:13434–13439.

Pearse, W. D., C. C. Davis, D. W. Inouye, R. B. Primack, and T. J. Davies. 2017. A statistical estimator for determining the limits of contemporary and historic phenology. Nature Ecology & Evolution 1:1876–1882.

Phillimore, A. B., S. Sandra, R. J. Smithers, and R. Bernard. 2012. Dissecting the contributions of plasticity and local adaptation to the phenology of a butterfly and its host plants. The American Naturalist 180:655–670.

Piao, S., et al. 2015. Leaf onset in the northern hemisphere triggered by daytime temperature. Nature Communications 6:6911.

Rafferty, N. E., P. J. CaraDonna, L. A. Burkle, A. M. Iler, and J. L. Bronstein. 2013. Phenological overlap of interacting species in a changing climate: an assessment of available approaches. Ecology and Evolution 3:3183–3193.

Rasmussen, N. L., B. G. Van Allen, and V. H. W. Rudolf. 2014. Linking phenological shifts to species interactions through size-mediated priority effects. Journal of Animal Ecology 83:1206–1215.

Rathcke, B., and E. P. Lacey. 1985. Phenological patterns of terrestrial plants. Annual Review of Ecology and Systematics 16:179–214.

Réale, D., A. G. McAdam, S. Boutin, and D. Berteaux. 2003. Genetic and plastic responses of a northern mammal to climate change. Proceedings of the Royal Society B 270:591–596.

Reed, T. E., P. Warzybok, A. J. Wilson, R. W. Bradley, S. Wanless, and W. J. Sydeman. 2009. Timing is everything: flexible phenology and shifting selection in a colonial seabird. Journal of Animal Ecology 78:376–387.

Reed, T. E., S. Jenouvrier, and M. E. Visser. 2013. Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. Journal of Animal Ecology 82:131–144.

Stinchcombe, J. R., and M. Kirkpatrick. 2012. Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. Trends in Ecology & Evolution 27:637–647.

Strebel, N., M. Kéry, M. Schaub, and H. Schmid. 2014. Studying phenology by flexible modelling of seasonal detectability peaks. Methods in Ecology and Evolution 5:483–490.

Stucky, B. J., R. Guralnick, J. Deck, E. G. Denny, K. Bolmgren, and R. Walls. 2018. The plant phenology ontology: A new informatics resource for large-scale integration of plant phenology data. Frontiers in Plant Science 9:517.

Tavecchia, G., T. Coulson, B. J. T. Morgan, J. M. Pemberton, J. C. Pilkington, F. M. D. Gulland, and T. H. Clutton-Brock. 2005. Predictors of reproductive cost in female Soay sheep. Journal of Animal Ecology 74:201–213.

Thackeray, S. J., P. A. Henrys, C. G. Jones, and H. Feuchtmayr. 2012. Eight decades of phenological change for a freshwater cladoceran: what are the consequences of our definition of seasonal timing? Freshwater Biology 57:345–359.

Thackeray, S. J., et al. 2016. Phenological sensitivity to climate across taxa and trophic levels. Nature 535:241–U294.

Tikkanen, O.-P., and R. Julkunen-Tiitto. 2003. Phenological variation as protection against defoliating insects: the case of *Quercus robur* and *Operophtera brumata*. Oecologia 136:244–251.

Valdés, A., and J. Ehrlén. 2017. Caterpillar seed predators mediate shifts in selection on flowering phenology in their host plant. Ecology 98:228–238.

van Asch, M., P. H. van Tienderen, L. J. M. Holleman, and M. E. Visser. 2007. Predicting adaptation of phenology in response to climate change, an insect herbivore example. Global Change Biology 13:1596–1604.

Vedder, O., S. Bouwhuis, and B. C. Sheldon. 2013. Quantitative assessment of the importance of phenotypic plasticity in adaptation to climate change in wild bird populations. PLoS Biology 11:e1001605.

Visser, M. E., and C. Both. 2005. Shifts in phenology due to global climate change: the need for a yardstick. Proceedings of the Royal Society B 272:2561–2569.

Visser, M. E., S. P. Caro, K. van Oers, S. V. Schaper, and B. Helm. 2010. Phenology, seasonal timing and circannual rhythms: towards a unified framework. Philosophical Transactions of the Royal Society B 365:3113–3127.

Wadgymar, S. M., and A. E. Weis. 2017. Phenological mismatch and the effectiveness of assisted gene flow. Conservation Biology 31:547–558.

Weis, A. E., and W. L. Gorman. 1990. Measuring selection on reaction norms: an exploration of the *Eurosta-Solidago* system. Evolution 44:820–831.

Weis, A. E., E. Nadone, and G. A. Fox. 2014. The strength of assortative mating for flowering date and its basis in individual variation in flowering schedule. Journal of Evolutionary Biology 27:2138–2151.

Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. Proceedings of the National Academy of Sciences USA 105:17029–17033.

Wolkovich, E. M., et al. 2012. Warming experiments underpredict plant phenological responses to climate change. Nature 485:494.

Wolman, A. G. 2006. Measurement and meaningfulness in conservation science. Conservation Biology 20:1626–1634.

Wu, C., et al. 2017. Land surface phenology derived from normalized difference vegetation index (NDVI) at global FLUXNET sites. Agricultural and Forest Meteorology 233:171–182.

Yang, L. H., and V. H. W. Rudolf. 2009. Phenology, ontogeny and the effects of climate change on the timing of species interactions. Ecology Letters 12:1–10.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecm.1352/full