Global Change Biology

Global Change Biology (2017) **23**, 1783–1791, doi: 10.1111/gcb.13472

OPINION

Confounding effects of spatial variation on shifts in phenology

CHARLOTTE W. DE KEYZER[1,2], NICOLE E. RAFFERTY[1,2,†], DAVID W. INOUYE[2,3] and JAMES D. THOMSON[1,2]

[1]*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3B2,* [2]*Rocky Mountain Biological Laboratory, Crested Butte, CO 81224, USA,* [3]*Department of Biology, University of Maryland, College Park, MD 20742, USA*

Abstract

Shifts in the timing of life history events have become an important source of information about how organisms are responding to climate change. Phenological data have generally been treated as purely temporal, with scant attention to the inherent spatial aspects of such data. However, phenological data are tied to a specific location, and considerations of sampling design, both over space and through time, can critically affect the patterns that emerge. Focusing on flowering phenology, we describe how purely spatial shifts, such as adding new study plots, or the colonization of a study plot by a new species, can masquerade as temporal shifts. Such shifts can look like responses to climate change but are not. Furthermore, the same aggregate phenological curves can be composed of individuals with either very different or very similar phenologies. We conclude with a set of recommendations to avoid ambiguities arising from the spatiotemporal duality of phenological data.

Keywords: *Cardamine cordifolia*, climate change, flowering time, long-term data, phenology, Rocky Mountain Biological Laboratory, spatial ecology, temporal ecology

Received 28 April 2016 and accepted 4 August 2016

Introduction

The phenomenon of climate change has stirred up interest in phenology as a source of response variables (Parmesan & Yohe, 2003; Root *et al.*, 2003; Visser & Both, 2005; Forrest & Miller-Rushing, 2010; Rafferty *et al.*, 2015). Because phenological records concern the timing of natural events, it is possible to think of them as purely temporal, nothing more than calendar dates. But such data are necessarily spatial as much as they are temporal. A plant species may be recorded as coming into bloom on a particular date, but what really happened is that the species came into bloom on a particular date *at the particular place* where the datum was recorded. In the northern temperate zone, for example, we know that spring takes weeks to sweep north across immense latitudinal gradients (Schwartz, 1998). Furthermore, numerous studies have documented latitudinal clines in spring phenology traits, such as bud flush (Aitken *et al.*, 2008) and flowering time (Stinchcombe *et al.*, 2004). At smaller spatial scales, spring also arrives later at higher elevation, on north-facing slopes, in shaded microhabitats, and in places that accumulate deeper snowpack over the winter. Such observations of heterogeneity are commonplace, but their consequences for data analysis are not always appreciated. In particular, variation in space and variation in time may interact in ways that allow one to masquerade as the other. Although this general principle affects all phenological data, here we focus on how spatiotemporal duality can potentially produce misleading interpretations of data on flowering times. We warn that such problems may become more severe with the need to forecast phenological change at greater spatial scales and as datasets are expanded, especially by spatially extensive efforts such as citizen science (Dickinson *et al.*, 2012; Schwartz *et al.*, 2012; Primack & Gallinat, 2016).

The critical roles of sample placement and sample pooling

In many long-term studies, space may essentially be held constant because the data all come from a single place. In such cases, it is not necessary to dwell on spatial variation – because there is none – but it is still important to realize that the data are spatial. They represent events at that single place and no other. It is also important to understand that the characteristics that determine the phenological earliness of a particular

[†]Present address: Department of Entomology, Washington State University, Pullman, WA 99164, USA

Correspondence: Charlotte W. de Keyzer, tel. +1 905 995 3778, fax +1 416 978 8532, e-mail: charlotte.dekeyzer@utoronto.ca

© 2016 John Wiley & Sons Ltd

place may not be inherent to that place per se. The place-specific characteristics that govern blooming may change through time even if the place is held strictly constant. For example, Thomson (2010) described how the earliest *Erythronium grandiflorum* flowers in a subalpine meadow consistently appeared around the bases of large spruce trees where the snow melted earlier than elsewhere. When such a tree dies and falls, the herbs around its base will no longer experience earlier snow melt even though their position in the meadow has not changed. The early-flowering propensity caused by the tree acts only in one special location, but the action is not a permanent property of that location's latitude and longitude. It is a transient property of the ecological community at that location. Similarly, trees deposited by avalanche debris can shade snow, causing it to melt weeks later than it would otherwise and delaying plant growth and flowering for at least 1 year in that location (D. Inouye, personal observation).

The importance of spatial variation in phenological data may be clarified by explicitly considering different spatial scales or hierarchical levels of organization (e.g., from an individual plant to an entire population). Phenological events are often recorded as the onset of a process, such as the first day of flowering of a species. However, such data are outliers by definition (van Strien *et al*., 2008), so more reliable estimates of flowering times are provided by entire flowering distributions as revealed by serial censuses taken throughout the flowering period (Miller-Rushing *et al*., 2008; CaraDonna *et al*., 2014). Investigators using historical data may be restricted to using dates of onset (Fitter & Fitter, 2002; Miller-Rushing & Primack, 2008; Rafferty & Ives, 2011), but we expect that newly designed studies will include more detail. Here, we focus on the 'flowering curve' of a species as a graph of the number of open flowers each day vs. day of year. Such a distribution has several properties that can describe the temporal distribution of flower openings: start date, end date, mean date of bloom, median date of bloom, breadth of the flowering curve (as a variance, say, or an interquartile range), and shape (e.g., skewness and kurtosis for unimodal curves; number of modes for others). We could also consider the evenness of flowering across days, as represented by the sorts of diversity indices used to measure how individuals are spread across species. Regardless of what we estimate, the values of all of these measures depend on arbitrary sampling decisions. The only species-level flowering curve that would not be affected by the spatial extent of sampling would be a theoretical *universal curve* that includes all flowers produced by all individual plants across the entire range of the species. Such a curve would amount to an envelope that would contain any subset of the total numbers that might emerge from sampling any smaller portion of the whole, all the way down to a single plant.

Any such subset, even a single plant, would amount to a partial summary of the universal curve, just as any statistical sample provides an incomplete estimate of population parameters (see, e.g., Sokal & Rohlf, 1981). Naturally, we would expect any randomly chosen single plant to provide a deeply inadequate representation of the complete curve. Counting flowers on many plants would better represent the whole, especially if those plants were randomly drawn from the set of all plants across the range. As one sampled more and more plants, the sample curve would more closely approach the complete curve. We would expect summary statistics to approach their parametric values gradually and asymptotically, just as estimates of species richness increase toward the true value as more individuals are added to the sample (Gotelli & Colwell, 2001). Attempts to measure flowering curves, as opposed to capturing first flowering dates, are more likely to be based on counts of flowers within study plots (e.g., Thomson, 1980; Kudo & Hirao, 2006; Inouye, 2008; CaraDonna *et al*., 2014; Wheeler *et al*., 2015), rather than single plants. Here again, sample estimates of phenological descriptors will more closely approach the parametric values as investigators sample more plots, larger plots, and plots that span more of the species' range.

Whether plants or plots are added, increasing the spatial extent of the pooled sample will improve its ability to represent the universal curve. Such improvement would be good if the goal of the study is to represent the universal curve, but that will seldom be the goal. Gathering data across the entire range of a species would almost never be practical and would seldom be desirable; investigators are far more likely to restrict their studies to more restricted areas. These could be determined geologically (such as a mountain range or a catchment basin), politically (such as a township or county), expediently (such as 'a study site' or 'the vicinity of a field station'), or collaboratively (such as the areas under observation by citizen scientists). Almost any such restriction is defensible, but the choice will inevitably affect the data that emerge. At the regional level, for example, a collection of plots spaced along a line of longitude will yield earlier start times, later end times, and a broader flowering curve than a similar collection spread along a line of latitude. At smaller scales, plots in a mountainous region will yield different results if they are concentrated on south-facing vs. north-facing slopes or if they are located at different elevations as opposed to being restricted to a narrow elevational range. Such quirks will be relatively

inconsequential if a single set of fixed plots is studied across time, but if plots are added to or subtracted from a study, the change in spatial extent may be misinterpreted as a change in timing.

Example of spatial variation presented as temporal variation

An example comes from a valuable dataset compiled by Inouye and colleagues on flowering times and abundance (CaraDonna *et al.*, 2014). Uniquely extensive, and now available online (https://osf.io/jt4n5/), this dataset continues to generate numerous papers (e.g., Inouye, 2008; Forrest *et al.*, 2010; Diez *et al.*, 2012; McKinney *et al.*, 2012; CaraDonna & Inouye, 2015; Wright *et al.*, 2015). Since 1974, flowers have been counted approximately every other day in a series of permanent 2 × 2 m plots in the vicinity of Gothic, Colorado, USA. The plots are distributed across a small spatial extent (approximately 1 km), small elevation range (approximately 110 m), and occur in different habitat types, from dry to mesic to wet meadows in this area.

The plot-based design of this study imparts particular spatial characteristics. The spatial and temporal duality of the data is an important consideration for analyses of changes in flowering phenology over time and for attributing phenological shifts to temporal processes such as climate change. Indeed, phenological shifts in this subalpine community have been attributed to increased temperatures and earlier snow melt associated with climate change over four decades (Iler *et al.*, 2013; CaraDonna *et al.*, 2014). The plots not only differ in space but also may have different temporal histories of plant representation within plots. For example, two plots were added in 1985, three in 1998, and two in 2004. Furthermore, species may newly colonize or go extinct in particular plots at any time, even if the plots were established and monitored for the same time periods. An analysis that does not explicitly consider spatial effects runs the risk of attributing all variation to temporal effects. The probability of organisms entering or dropping out of plots during a study will be greatest for mobile animals or short-lived plants, but can occur with long-lived perennials, too.

We present a case study of flower count data from the perennial herb, *Cardamine cordifolia* (hereafter *Cardamine*). Based on Fig. 2 of CaraDonna *et al.* (2014), *Cardamine* has exhibited the greatest shift in both first flowering and peak flowering dates of the 60 species analyzed for this plant community. We chose to examine this species because its reported shift was greater than other plants with similar life histories; we wondered whether this could be due to a spatial change masquerading as a temporal change, such as the colonization by *Cardamine* into an existing plot or the addition of a new plot to the study.

In the Inouye dataset, *Cardamine* has been found in four different plots since 1974. *Cardamine* flowers have been fairly consistently counted in the Willow-Meadow Interface #3 (INT3) plot from 1974 to 2011. For 24 of these years, the first date of *Cardamine* flowering has been recorded from INT3, a plot characterized as mesic. However, since the addition of the Stream (STR) plot in 2004, the date of first flowering for *Cardamine* across all plots has always been recorded in this plot (Fig. 1a). The STR plot is characterized as wet habitat and is located at the top of a stream drainage site. We asked whether the addition of this plot has had an effect on the changes in first and peak flowering dates for *Cardamine* over time and whether those changes should be attributed to space or time.

When plots are pooled and the change in first date of flowering for *Cardamine* (1974–2012) is analyzed using a simple linear regression, as done by CaraDonna *et al.* (2014), first flowering is estimated to have significantly advanced over time by 8.3 ± 1.2 days per decade (Fig. 1a; $R^2 = 0.59$, $F_{1,32} = 48.1$, $P < 0.0001$). When the STR plot is excluded from the analyses, the estimated advance drops to 2.6 ± 1.3 days per decade, and the significance falls short of the conventional threshold of 0.05 (Fig. 1b; $R^2 = 0.076$, $F_{1,31} = 3.65$, $P = 0.066$). We further tested for spatial effects by fitting a linear model to unpooled data with plot and year as predictor variables ($R^2 = 0.51$, $F_{4,39} = 12.2$, $P < 0.0001$, slope = 0.26); both year ($P = 0.048$) and STR plot ($P < 0.0001$) were significantly correlated with first flowering. Our findings indicate that the addition of the STR plot has had a significant effect on the perceived change in first flowering for *Cardamine*.

We also asked whether the more robust phenological descriptor, peak flowering, could be similarly affected by a change in sampling extent. We have adopted the same definition for peaking flowering as CaraDonna *et al.* (2014): the day of the year on which 50% of flowers pooled across plots have been counted. When we include the STR plot in the pooled flower counts, the change in peak flowering for *Cardamine* (1974–2012), analyzed by simple linear regression, is estimated to have significantly shifted earlier by 7.6 ± 1.2 days per decade ($R^2 = 0.55$, $F_{1,32} = 40.8$, $P < 0.0001$). When we exclude the STR plot from the same analysis, we find that the change in peak flowering is still significant but of lesser magnitude, advancing by 3.4 ± 1.2 days per decade ($R^2 = 0.18$, $F_{1,31} = 8.093$, $P = 0.0078$). By not accounting for the addition of the STR plot, the change in peak flowering is perceived to have advanced by an additional 4.2 ± 1.7 days per decade. We elaborate

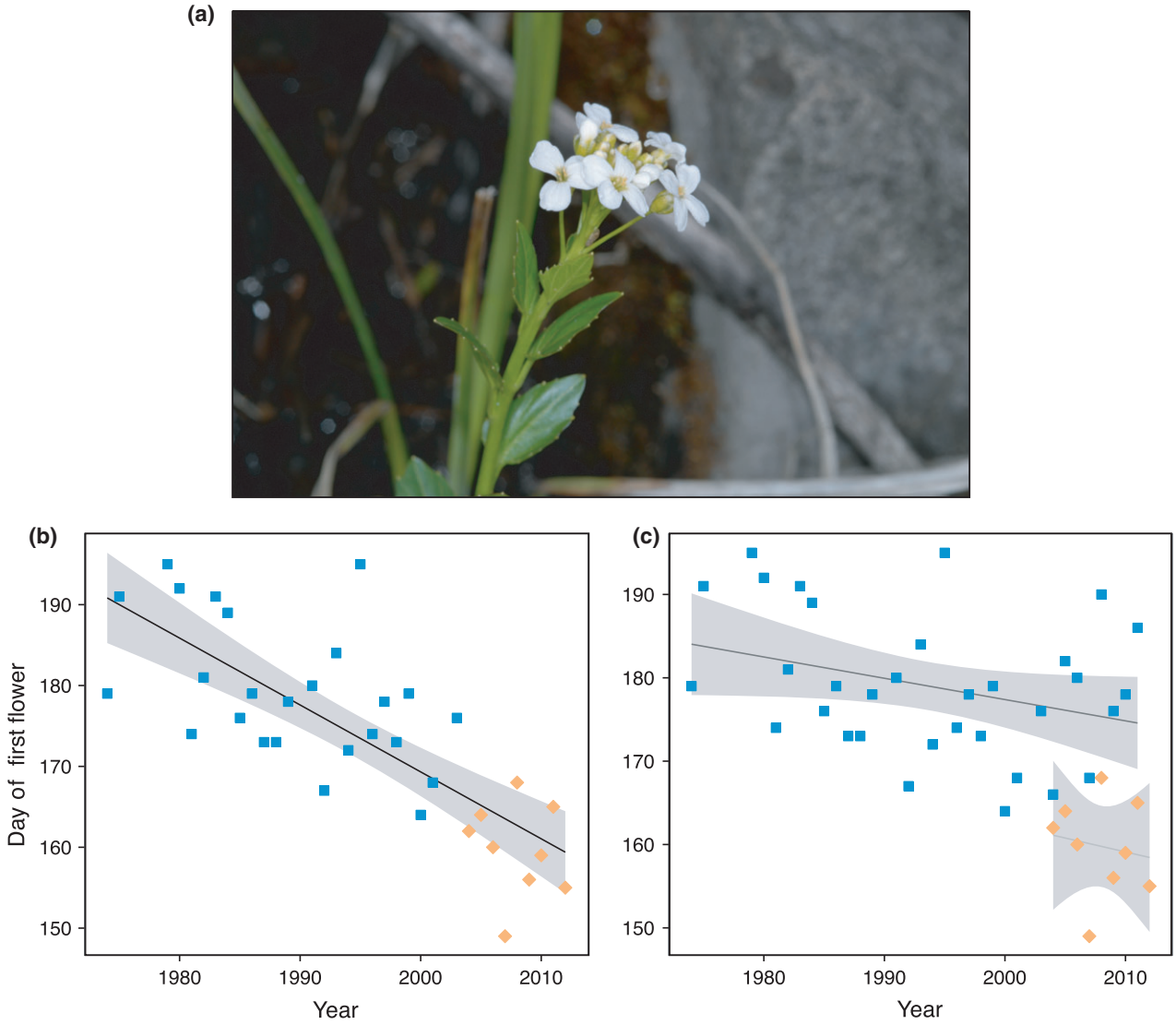

Fig. 1 Change in day of year of first flowering for *Cardamine cordifolia* (a) over a 38-year time period (1974–2012) when (b) data are pooled across plots; each point is the first date of flowering recorded for that year. Plot identity of each point is shown (blue squares = INT3 plot; orange diamonds = STR plot). $R^2 = 0.59$, $F_{1,32} = 48.1$, $P < 0.0001$, slope = −0.83. (c) Data are unpooled; each point is the first date of flowering recorded in each plot that year, only plots with multiple years of data are shown. Within plots, the regression slopes are less and the confidence intervals are so wide that the relationships become insignificant. INT3 (blue squares): $R^2 = 0.076$, $F_{1,31} = 3.65$, $P = 0.066$, slope = −0.26. STR (orange diamonds): $R^2 = 0.12$, $F_{1,7} = 0.18$, $P = 0.69$, slope = −0.33. Confidence bands (95%) of regression slopes are shown.

further on the *Cardamine* example in an online appendix. It is important to know that the issue affecting *Cardamine* is an isolated example that does not affect the general value of the Inouye phenology data.

In the *Cardamine* example, the spatial change that masqueraded as a temporal change was the addition of a new study plot. That might seem to be a simple matter to avoid, but the same effect could arise if a plant species being studied were to migrate into a new plot during the study, and if that plot happened to be in a site that favored early flowering, or if it were to disappear from a plot, especially a late-flowering one. The plant species would appear to begin flowering earlier, but the shift would be attributable to the local-scale dispersal dynamics of the species rather than a change in climate. Of course, spatial range shifts can themselves be driven by climatic changes (Parmesan & Yohe, 2003; Freeman & Class Freeman, 2014; Kuhn *et al.*, 2016), but these will usually be captured at spatial scales larger than individual plots.

The *Cardamine* example demonstrates, first, that pooling of data may introduce problems; second, that

archiving raw data in sufficient detail can allow some of those problems to be detected and fixed. The first point has been made before; for example, Miller-Rushing & Primack (2008, p. 339) discuss the issue with regard to their combining four separate sets of phenological data in the area of Concord, Massachusetts, USA:

> '[W]e, Thoreau, and Hosmer observed flowering times throughout Concord, while Logemann observed flowering times only on her property in Concord. Because Logemann observed a smaller area and fewer plants, the first flowering dates she observed for many species were later than they were for the other observers'.

Although Miller-Rushing & Primack (2008) clearly recognized the dependence of phenological records on the extent of sampling, they did not dwell on it or offer solutions. For example, they did not discuss the possibility of restricting their present-day observations to the area of Logemann's property, or of deriving some correction factor by comparing contemporary data from Logemann's property to data from more extensive samples.

Miller-Rushing *et al.* (2008) discussed another issue of sampling extent. Species with increasing population sizes, such as invasive species, will colonize a wider range of microclimates over time. Some of those sites will tend to promote earlier flowering, so species that expand their ranges over the course of a study will tend to show accelerated first flowering dates. Similarly, ubiquitous species will presumably have occupied more of the early sites than rare species, leading to a positive relationship between population size and early onset. In Concord, shifts in flowering phenology for some species could be explained only by change in population size and not by change in climate; however, the same overriding effect of population size was not found for observations from Gothic (Miller-Rushing *et al.*, 2008). Miller-Rushing *et al.* (2008) suspect that the extent of sampling accounted for this difference between Gothic and Concord. The much smaller area covered by the 2×2 m plots sampled at Gothic would naturally provide a much smaller range of microclimates for expanding populations to colonize.

In this case also, Miller-Rushing *et al.* (2008) acutely point out the sensitivity of temporal inferences to the extent of sampling in a special case but do not provide more general guidelines for properly attributing patterns to causes. We argue that further attention is warranted to the spatial extent of phenological data, and we close this paper with specific recommendations for researchers who are either designing new studies or analyzing existing data.

Thus far, we have emphasized how spatial variation can present a difficulty for interpreting long-term phenological data; however, we do not wish to convey a message that calls all previous work on long-term flowering phenology into question. For example, most species in the Inouye dataset have been less affected by changes in sampling extent (P. CaraDonna & A. Iler, personal communication). *Cardamine* is an outlier, likely because only a small number of plants are represented in a small number of plots, making this species more likely to be affected by plot additions and subtractions, especially when plots differ in habitat characteristics. Other studies that have analyzed long-term phenology datasets have accounted for spatial variation by including plot identity as a fixed (Kudo & Hirao, 2006) or random effect (Anderson *et al.*, 2012; but see Wheeler *et al.*, 2015) in mixed-effects models. Furthermore, experimental manipulations (e.g., snow removal) have been used to demonstrate that shifts in phenology are attributable to temporal processes, such as the effect of climate change on the timing of snow melt (Anderson & Gezon, 2015). Indeed, spatial variation in phenology can also be used to a researcher's advantage; a particularly powerful method for studying climate change effects on phenology has been to combine spatial gradients with warming experiments over time (Dunne *et al.*, 2003, 2004).

Decomposing pooled curves

The spatiotemporal nature of phenological data also affects other issues, such as the hierarchical structure of ecological heterogeneity (i.e., the ecological patterns found at various scales) and the nature of interplant variation, which in turn affect the prospects for response to selection. By considering a pooled, universal flowering curve as the sum of smaller-scale processes, we can better understand how a particular property of the pooled distribution might arise from different underlying processes (Elzinga *et al.*, 2007 discuss other examples). This is most easily seen with respect to properties such as flowering curve breadth or symmetry. Fig. 2 illustrates how two very different sets of plot data can give rise to identical pooled curves, but how the underlying differences can be summarized by partitioning flowering-time diversity of the whole collection (γ diversity) into α (within-plot) and β (between-plot) components. The hypothetical populations in Fig. 2a, c are ones in which all plots are homogeneous, and each one reproduces the pooled data. In Fig. 2b, d, β diversity is much higher; plots vary wildly, and any one plot is a very poor representative of the

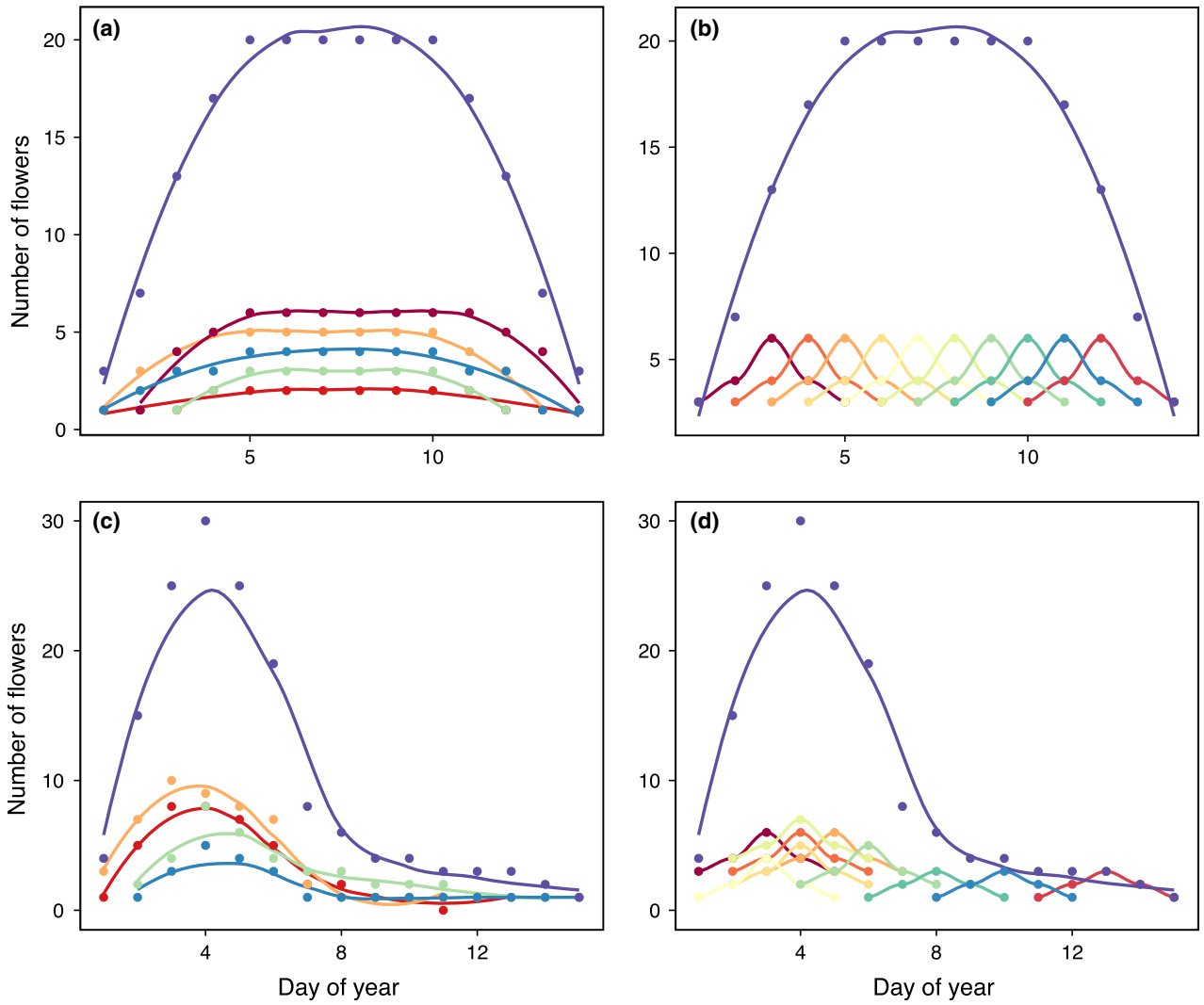

Fig. 2 Hypothetical data to demonstrate effects of pooling; here we focus on pooling across plots; however, the same effects apply to data pooled across individual plants. These data were created by assigning daily counts of open flowers to individual plots. Population-level (pooled) flowering counts were created by summing the plot-level flowering counts across days. For pairs of identical pooled curves (a and b, and c and d; shown in purple), we show possible differences in the flowering curves of the underlying plots. The differences between (a) and (b) are summarized by partitioning flowering-time diversity of the whole collection (γ diversity) into α (within-plot) and β (between-plot) components. For simplicity, we use Simpson's index for α and γ diversity and follow Whittaker's original definition of $\beta = \gamma/\alpha$ (Whittaker, 1972). For these equations, instead of calculating the proportional abundance of each species, we calculate the proportional abundance of each day's open flowers. (a) All plots are a good representation of the pooled curve, with similar symmetry and breadth (γ = 11.66, β = 1.03). This hypothetical scenario could occur when plots sampled for a given species occur in a homogeneous flowering environment. (b) Any one plot is a poor representation of the pooled curve (γ = 11.66, β = 2.51). β diversity is much higher in (b) than in (a). This hypothetical scenario could occur when plots sampled for a given species occur in a heterogeneous flowering environment. (c) All underlying plots have skewed flowering curves that are well represented by the skewness of the pooled curve. (d) None of the underlying plots show skewed flowering curves and instead the skewness of the pooled curve is a result of a greater number of plots flowering earlier. All curves are computed by a locally weighted regression smoother (LOESS).

whole. The two situations would present very different challenges to flower feeders and different opportunities for individual selection on timing (Thomson, 1980; Elzinga *et al.*, 2007), but those differences become invisible when the data are pooled across plots.

The same principle holds if data are recorded for individual plants instead of individual plots (Primack, 1985; Elzinga *et al.*, 2007). For example, Thomson (1980) noted that 53 of 57 plant species in subalpine meadows showed positively skewed flowering curves in pooled

census data from rectangular grids of 108 plots of 4 m^2. Although he speculated about possible adaptive advantages for individual plants with skewed curves, he pointed out that the shapes of individual plants' curves cannot be deduced from the shape of the pooled curve to which they contribute. As shown in Fig. 2c, a pooled curve might be positively skewed because the individual plant curves are positively skewed and start more or less at the same time. Alternatively, the individual curves might be symmetrical (or even negatively skewed) but their locations in time might tend to be aggregated toward the beginning of flowering, producing a skewed pooled curve by a different mechanism (Fig. 2d; see also CaraDonna *et al.*, 2014). Moving from the plot level to the plant level, in a later study of *Diervilla lonicera*, Thomson (1985) decomposed a pooled flowering curve – which was skewed – into its constituent curves from each plant. In that case, individual shoots did tend to be skewed, thereby presenting the same skewed nature as the pooled curve. However, a different result might have emerged. The same argument applies to individual plants and individual plots: once phenological data have been pooled across individuals or sample plots, the characteristics of those subunits are lost. To the extent that phenological sampling collects data at multiple hierarchical levels, such as plants within plots, plots within meadows, and meadows within floodplains, the preceding arguments about plants should apply at each hierarchical level (Primack, 1985).

Concluding recommendations

These considerations suggest some good practices for phenological monitoring. We offer these particularly to investigators who are planning new studies. Although we have focused on flowering phenology, many of the following recommendations could also be applied to phenological data collection for mobile organisms, especially when collected from fixed locales (e.g., permanent bird-banding stations).

1. In estimating phenological properties, one should consider and anticipate how those estimates will depend on the spatial extent of the sampling and the heterogeneity of the area sampled. In analogy to the role of species–area curves in the estimation of species diversity, one should examine how estimates of phenological descriptors such as flowering duration respond to the successive addition of sample plots (Miller-Rushing *et al.*, 2008). Ideally, a study would include enough plots for those estimates to stabilize – as a species–area curve levels off – but this may not be practical. Rarefaction procedures, analogous to those used for estimating species diversity, could help gauge the adequacy of sampling. The analogy to species diversity could extend to the point of breaking down phenological variation into hierarchical components akin to α-, β-, and γ-diversity.
2. In designing a study that will establish permanent sample plots, the criteria for selecting plots should be considered and recorded as part of the metadata. To consider questions of functional ecology, it would be important for the sample of plots to include a broad range of local variation. But if the goal is to document change over time, it may be better to select plots using some criterion of homogeneity. Ecological homogeneity is a difficult criterion that might require subjectivity (see Curtis, 1959), but it is better to consider it explicitly rather than to ignore it. For studies wishing to isolate temporal trends in flowering phenology, it may be appropriate to include a plot-level characteristic (e.g., elevation, soil water content, light level, *etc.*) as a covariate in analyses.
3. Once a study has begun, researchers should strongly consider and anticipate the implications of losing plots (or individuals) and adding plots (or individuals). Data and metadata must be archived at the plot (or individual) level so that additions and deletions can be detected and treated as necessary. Data analysis should also be able to detect spontaneous gains or losses of species within plots (see Roth *et al.*, 2014). In most cases, conclusions about changing phenology should focus on temporal changes within plots rather than pooled across plots. When plots are pooled, statistical analyses of time series should account for potential effects of plot identity [e.g., by including individual plot as a fixed (Kudo & Hirao, 2006) or random effect (Anderson *et al.*, 2012)]. Another approach would be to analyze trends within plots only and to use meta-analytic techniques for assessing overall effects and significance. Such an approach would allow one to disentangle the confounding effects of population increase noted by Miller-Rushing *et al.* (2008), especially if the smallest sampling units are small enough to be microclimatically homogeneous.
4. Recording data at the individual-plant level could be performed instead of or in addition to recording data at the plot level. Gathering plant-level data, even if only for a subset of labeled plants, could provide insight into the relationship between individual and populational patterns. In some cases, it may require only slightly more effort to record data at the individual-plant level, which, if combined with data such as seed set, could give important insights into the relationship between phenology and fitness components. Long-term studies that link phenology and

fitness will improve our understanding of how plants are succeeding or failing to adapt to the novel conditions presented by climate change. As well, individual-level data are likely to provide a more precise and accurate dataset that could be summarized with averages instead of by pooling. For example, instead of reporting the date of onset at the plot level, one could calculate the mean date of onset across individual plants within a plot.

5. In considering study designs, a potentially powerful approach would be to recruit a growing group of citizen scientists to record dates of first flowering, leaf budbreak, etc. (as is being done by the USA National Phenology Network, and Project BudBurst). Such datasets would be particularly vulnerable to space-for-time effects, however. If such data were pooled across respondents, the dates of onset would become earlier through time as a purely statistical consequence of adding more observers and thereby increasing the spatial extent of the sample. Ending dates, such as leaf drop, would become later. If observers dropped out of the study, onset would appear to be retarded. It would be critical to analyze such time series within observers or to account for sampling effort, and any measure of central tendency would be better than recording dates of onset or termination (CaraDonna *et al*., 2014). Extension of classic site-occupancy models has been suggested as a possible method for analyzing such temporally and spatially variable phenological data (Roth *et al*., 2014).

Acknowledgements

We thank P. CaraDonna, A. Iler, J. Ogilvie, J. Anderson, and three anonymous reviewers for comments on the manuscript. This research was supported by an NSERC CGS-M to CWdK, a Postdoctoral Fellowship from the Department of Ecology and Evolutionary Biology at the University of Toronto to NER, NSF grants (DEB 75-15422, DEB 78-07784, BSR 81-08387, DEB 94-08382, IBN 98-14509, DEB 02-38331, DEB 09-22080) to DWI, and NSERC Discovery Grants to JDT.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Anderson JT, Gezon ZJ (2015) Plasticity in functional traits in the context of climate change: a case study of the subalpine forb *Boechera stricta* (Brassicaceae). *Global Change Biology*, **21**, 1689–1703.

Anderson JT, Inouye DW, McKinney AM, Colautti RI, Mitchell-Olds T (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3843–3852.

CaraDonna PJ, Inouye DW (2015) Phenological responses to climate change do not exhibit phylogenetic signal in a subalpine plant community. *Ecology*, **96**, 355–361.

CaraDonna PJ, Iler AM, Inouye DW (2014) Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 4916–4921.

Curtis JT (1959) *The Vegetation of Wisconsin: An Ordination of Plant Communities*, pp. 70–71. University of Wisconsin Press, Madison, WI, USA.

Dickinson JL, Shirk J, Bonter D *et al*. (2012) The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, **10**, 291–297.

Diez JM, Ibanez I, Miller-Rushing AJ *et al*. (2012) Forecasting phenology: from species variability to community patterns. *Ecology Letters*, **15**, 545–553.

Dunne JA, Harte J, Taylor K (2003) Subalpine meadow flowering phenology responses to climate change: integrating experimental and gradient methods. *Ecological Monographs*, **73**, 69–86.

Dunne JA, Saleska SR, Fischer ML, Harte J (2004) Integrating experimental and gradient methods in ecological climate change research. *Ecology*, **85**, 904–916.

Elzinga JA, Atlan A, Biere A, Gigord L, Weis AE, Bernasconi G (2007) Time after time: flowering phenology and biotic interactions. *Trends in Ecology and Evolution*, **22**, 432–439.

Fitter AH, Fitter RSR (2002) Rapid changes in flowering time in British plants. *Science*, **296**, 1689–1691.

Forrest J, Miller-Rushing AJ (2010) Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 3101–3112.

Forrest J, Inouye DW, Thomson JD (2010) Flowering phenology in subalpine meadows: does climate variation influence community co-flowering patterns? *Ecology*, **91**, 431–440.

Freeman BG, Class Freeman AM (2014) Rapid upslope shifts in New Guinean birds illustrate strong distributional responses of tropical montane species to global warming. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 4490–4494.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Iler AM, Høye TT, Inouye DW, Schmidt NM (2013) Nonlinear flowering responses to climate: are species approaching their limits of phenological change? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **368**, 20120489.

Inouye DW (2008) Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology*, **89**, 353–362.

Kudo G, Hirao AS (2006) Habitat-specific responses in the flowering phenology and seed set of alpine plants to climate variation: implications for global-change impacts. *Population Ecology*, **48**, 49–58.

Kuhn E, Lenoir J, Piedallu C, Gégout J-C (2016) Early signs of range disjunction of submountainous plant species: an unexplored consequence of future and contemporary climate changes. *Global Change Biology*, **22**, 2094–2105.

McKinney AM, Caradonna PJ, Inouye DW, Barr B, Bertelsen CD, Waser NM, Irwin RE (2012) Asynchronous changes in phenology of migrating Broad-tailed Hummingbirds and their early-season nectar resources. *Ecology*, **93**, 1987–1993.

Miller-Rushing AJ, Primack RB (2008) Global warming and flowering times in Thoreau's Concord: a community perspective. *Ecology*, **89**, 332–341.

Miller-Rushing AJ, Inouye DW, Primack RB (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.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Primack RB (1985) Patterns of flowering phenology in communities, populations, individuals, and single flowers. In: *The Population Structure of Vegetation* (ed. White J), pp. 571–593. Springer, Dordrecht, the Netherlands.

Primack RB, Gallinat A (2016) Spring budburst in a changing climate. *American Scientist*, **104**, 102–109.

Rafferty NE, Ives AR (2011) Effects of experimental shifts in flowering phenology on plant-pollinator interactions. *Ecology Letters*, **14**, 69–74.

Rafferty NE, CaraDonna PJ, Bronstein JL (2015) Phenological shifts and the fate of mutualisms. *Oikos*, **124**, 14–21.

Root TL, Price JT, Hall KR, Schneider SH, Rosenzweig C, Pounds JA (2003) Fingerprints of global warming on wild animals and plants. *Nature*, **421**, 57–60.

Roth T, Strebel N, Amrhein V (2014) Estimating unbiased phenological trends by adapting site-occupancy models. *Ecology*, **95**, 2144–2154.

Schwartz MD (1998) Green-wave phenology. *Nature*, **394**, 839–840.

Schwartz MD, Betancourt JL, Weltzin JF (2012) From Caprio's lilacs to the USA National Phenology Network. *Frontiers in Ecology and the Environment*, **10**, 324–327.

Sokal RR, Rohlf FJ (1981) *Biometry 2nd edn: The Principles and Practice of Statistics in Biological Research*, pp. 52–53. W.H. Freeman and Company, New York, NY.

Stinchcombe JR, Weinig C, Ungerer M *et al*. (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA.

Proceedings of the National Academy of Sciences of the United States of America, **101**, 4712–4717.

van Strien AJ, Plantenga WF, Soldaat LL, van Swaay CAM, WallisDeVries MF (2008) Bias in phenology assessments based on first appearance data of butterflies. *Oecologia*, **156**, 227–235.

Thomson JD (1980) Skewed flowering distributions and pollinator attraction. *Ecology*, **61**, 572–579.

Thomson JD (1985) Pollination and seed set in *Diervilla lonicera* (Caprifoliaceae): temporal patterns of flower and ovule deployment. *American Journal of Botany*, **72**, 737–740.

Thomson JD (2010) Flowering phenology, fruiting success and progressive deterioration of pollination in an early-flowering geophyte. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **365**, 3187–3199.

Visser ME, Both C (2005) Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2561–2569.

Wheeler HC, Høye TT, Schmidt NM, Svenning J-C, Forchhammer MC (2015) Phenological mismatch with abiotic conditions – implications for flowering in Arctic plants. *Ecology*, **96**, 775–787.

Whittaker R (1972) Evolution and measurement of species diversity. *Taxon*, **21**, 213–251.

Wright KW, Vanderbilt KL, Inouye DW, Bertelsen CD, Crimmins TM (2015) Turnover and reliability of flower communities in extreme environments: insights from long-term phenology data sets. *Journal of Arid Environments*, **115**, 27–34.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Re-analysis of *Cardamine* data controlling for time period and inclusion of data from 1974 to 2015.