Caterpillar Patterns in Space and Time: Insights From and Contrasts Between Two Citizen Science Datasets

Grace J. Di Cecco and Allen H. Hurlbert

Introduction

Caterpillars are an important component of forest ecosystems as both herbivores on woody and herbaceous vegetation and as a food resource for higher trophic levels. Trends in the abundance, biomass, or phenology of caterpillars may thus have impacts that propagate up and down food chains. Region-specific combinations of climate and land use change may result in geographically variable responses by caterpillar populations, and understanding this geographic variation would help identify the most important global change drivers and their mechanisms. Geographically widespread data on occurrence has increased tremendously with the digitization of museum records (Nelson and Ellis 2018), the establishment of online repositories such as the Global Biodiversity Information Facility (GBIF; Telenius 2011), and the rise of popular citizen science platforms like iNaturalist (Seltzer 2019), resulting in millions of observations across the globe. Nevertheless, questions have been raised about the types of inferences that can be made from such data when information on the sampling effort underlying those records is biased or unknown (Isaac and Pocock 2015; Mair and Ruete 2016; Ries et al. 2019; Di Cecco et al. 2021).

Standardized monitoring schemes, on the other hand, are explicit about effort and are able to provide more accurate estimates of abundance and phenology, yet tend to yield fewer and more sparsely distributed data (Soroye et al. 2018). Given

G. J. Di Cecco
Department of Biology, University of North Carolina, Chapel Hill, NC, USA

A. H. Hurlbert (✉)
Department of Biology, University of North Carolina, Chapel Hill, NC, USA

Environment, Ecology, and Energy Program, University of North Carolina, Chapel Hill, NC, USA
e-mail: hurlbert@bio.unc.edu

R. J. Marquis, S. Koptur (eds.), *Caterpillars in the Middle*, Fascinating Life Sciences, https://doi.org/10.1007/978-3-030-86688-4_17

this tradeoff between data collection approaches, it would be useful to know how well unstandardized occurrence records can recapitulate patterns of caterpillar abundance and phenology observed by more standardized monitoring schemes. If patterns are similar, this implies that the large volume of occurrence records can be used to infer patterns at greater geographic extents and with greater spatiotemporal resolution than would be possible from standardized monitoring data alone. If not, then more work will be needed to understand and deal with the various sources of bias that hamper opportunistic occurrence records.

Lepidoptera in general, and the caterpillar stage specifically, are well represented on iNaturalist. The "Caterpillars of Eastern North America" project on iNaturalist (https://www.inaturalist.org/projects/caterpillars-of-eastern-north-america) has over 400,000 observations from eastern North America alone at the time of publication. Unfortunately, while a number of geographically widespread monitoring programs exist for adult Lepidoptera (e.g. Fourth of July Butterfly Count, state- or nationwide butterfly atlases) or for single species (e.g., Monarch Larva Monitoring Project; Prysby and Oberhauser 2004), there have historically been no community-wide monitoring efforts for caterpillars as a group that have collected data across many regions. Community-wide monitoring is especially important when trying to understand variation in total resource availability for a higher trophic level like foliage-gleaning birds, rather than attempting to understand the dynamics of a particular taxonomic subset of caterpillars. To address this monitoring gap, the citizen science project *Caterpillars Count!* was created in 2015 to collect data on the phenology and abundance of foliage arthropods on woody vegetation (Hurlbert et al. 2019).

Here, we contrast patterns in the taxonomic representation, occurrence, and phenology of caterpillars derived from these two very different citizen science datasets, iNaturalist and Caterpillars Count!. We describe broadscale caterpillar patterns using each dataset and assess the strengths of each for inferring spatiotemporal variation in caterpillar occurrence across North America.

Datasets

iNaturalist

Participants of the iNaturalist citizen science project typically submit photo observations along with a date and georeferenced location of the observation. The observer can suggest a taxonomic identification, and then other iNaturalist users can agree or suggest alternative identifications. The Caterpillars of Eastern North America project is a collation of all larval Lepidoptera records in iNaturalist from Mexico, the United States, and Canada east of 100° W longitude. The number of records contributed to this iNaturalist project has grown steadily over time, with more than 59,000 unique observers contributing >206,000 photos as of June 2020.

We used all observations of larval lepidopterans including those not identified more finely than order. Taxonomic family was specified for 178,702 observations. iNaturalist records are the result of an unknown observation process that depends both on the number of users in space and time as well as their individually variable reporting behavior (Di Cecco et al. 2021).

Caterpillars Count!

Caterpillars Count! is a monitoring program in which participants conduct sets of standardized foliage arthropod surveys on woody vegetation by either carrying out visual inspection of an area of 50 leaves and their associated petioles and twigs or by striking a branch ten times over a beat sheet (Hurlbert et al. 2019). Each participating site has between 10 and 60 marked survey branches, all of which are ideally surveyed once every week or two from after leaf out at least through July. Survey branches are selected in a quasi-standardized fashion so as to capture representative vegetation within the area deemed of interest by each individual site coordinator (Hurlbert et al. 2019).

Participants identify all arthropods found to order and record their abundance and length in millimeters. They may optionally take photos of their observations, which are then submitted automatically to iNaturalist with the potential to receive finer-level taxonomic identifications. Through June 2020, 1,278 unique users have conducted 48,384 branch surveys at 116 sites from throughout the United States and Canada, mostly east of 100° W. Of the 9,981 total caterpillar observations from Caterpillars Count!, 605 were photographed, submitted to iNaturalist, and subsequently identified to the family level or below. All Caterpillars Count! observations shared with iNaturalist, including those that were not yet identified, were removed from the analyses representing iNaturalist patterns.

Family Composition

While Caterpillars Count! observations are specifically from trees and shrubs, iNaturalist observations are unconstrained by substrate and are more likely to capture caterpillars in gardens, on herbaceous vegetation, and even on the ground during a wandering phase. These differences in "sampling" process result in differential representation of Lepidoptera families within the two datasets.

Across both projects, Erebidae was one of the two most commonly represented families in eastern North America, but there were clear differences in family representation between projects as well (Fig. 1a; $\chi^2 = 320.6$, df = 10, $p < 2e\text{-}16$). In the iNaturalist dataset, Nymphalidae, Papilionidae, Sphingidae, and Lasiocampidae were relatively overrepresented compared to Caterpillars Count!, and caterpillars in these families are frequently large, conspicuous, and found in gardens and yards.

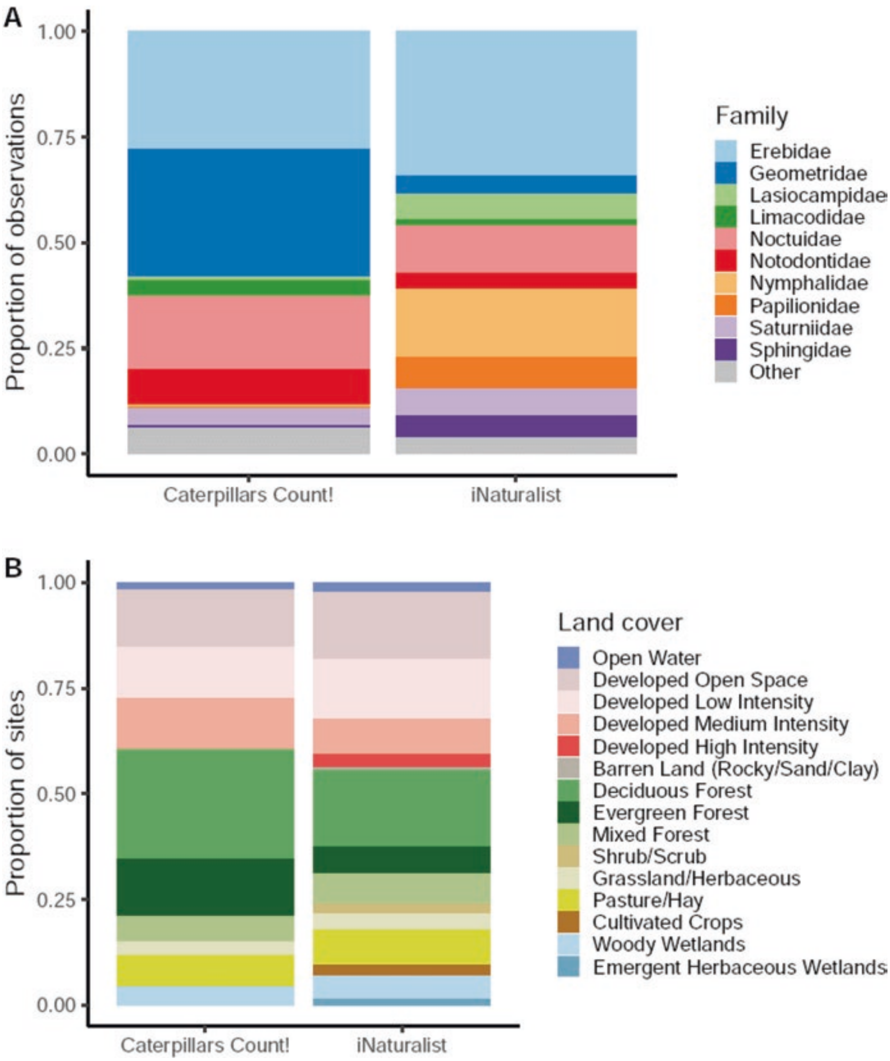

Fig. 1 Relative representation of the (**a**) most common caterpillar families and (**b**) land cover classes represented in the Caterpillars Count! (605 records identified to family, 66 sites) and iNaturalist datasets (178,702 records; 109,939 sites)

Conversely, Geometridae, and to a lesser extent, Notodontidae and Noctuidae were relatively overrepresented in the Caterpillars Count! dataset which was restricted to woody vegetation and which involved people actively searching for caterpillars, many of which are cryptic in form and color (Fig. 1a). To the extent that Caterpillars Count! participants are more likely to submit photos of showy or conspicuous caterpillars, the true proportion of these less conspicuous groups may be even greater.

A comparison of the land cover types within a 75 m radius of each Caterpillars Count! site and at the latitude and longitude of each iNaturalist observation based on data from the 30 m resolution National Land Cover Database from 2016 (Yang et al. 2018) highlights some of the differences in habitats represented between the two datasets (Fig. 1b). Forest habitats are more sampled by Caterpillars Count!, while high-intensity developed areas and open areas including cropland, pasture, grassland, and scrub are more sampled by iNaturalist users (although the small sample size of Caterpillars Count! sites precludes a chi-square test here).

Based on data from Caterpillars Count! as well as a study by Seifert et al. (2020) which examined caterpillar density across 15 common eastern North American tree species, the most commonly encountered families of caterpillars on woody vegetation were Geometridae, Erebidae, Noctuidae, Notodontidae, Depressariidae, and Tortricidae. Even restricting the comparison to these most commonly observed families, Geometridae is still strikingly underrepresented in iNaturalist (8.5% of observations compared to 36% in Caterpillars Count!), highlighting the impact of potential biases in the observation process in opportunistic versus survey-based citizen science datasets.

Geographic Patterns

We examined spatial patterns of caterpillar occurrence during June 2019 in both datasets using a uniform hexagonal grid (distance between cell centers of 285 km; per cell area of approximately 70,000 km^2). During this month-long window, there were 6,380 caterpillar observations submitted to iNaturalist, and we scaled the number of observations per hex cell by the total number of insect observations per hex cell in order to control for spatial variation in iNaturalist activity. During this period there were 3,882 Caterpillars Count! surveys conducted which reported 876 total caterpillars, and we calculated the number of caterpillars observed per survey within each hex cell. Importantly, this snapshot of caterpillar density or occurrence within the month of June may reflect slightly different points in time relative to forest leaf out depending on latitude.

Patterns of caterpillar occurrence were spatially heterogeneous. As a fraction of the total insect observations, iNaturalist caterpillar occurrence in June was highest in the Upper Midwest (Minnesota, Wisconsin, Michigan), as well as in Louisiana and Mississippi, typically making up fewer than 10% of all insect observations. Caterpillar occurrence was lowest across the Midwest (Missouri, Illinois, Indiana, Ohio), the southeastern states east of Mississippi, and easternmost Canada. Within the subset of hex cells that had sufficient Caterpillars Count! surveys, Michigan, Massachusetts, and North Carolina had the highest caterpillars observed per survey branch, and there was a positive correlation between iNaturalist caterpillar occurrence and Caterpillars Count! density estimates (Fig. 2c; $r = 0.729$, $p = 0.002$, $n = 15$ hex cells).

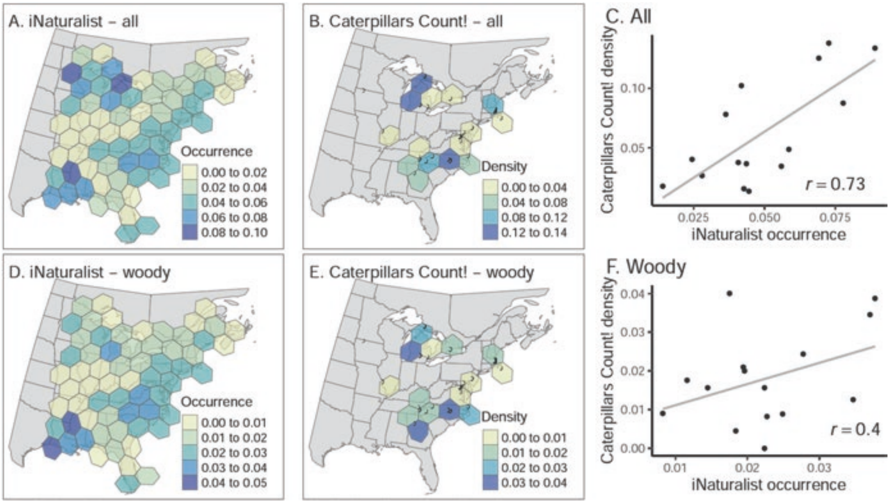

Fig. 2 Geographic variation in June 2019 in (**a**) iNaturalist caterpillar occurrence (proportion of all insect observations), (**b**) Caterpillars Count! caterpillar density (the number of caterpillars observed per branch survey), and (**c**) the relationship between the two at the scale of individual hex cells. (**d**–**f**) Same as (**a**–**c**) but for the subset of caterpillars from families well represented on woody vegetation (Erebidae, Geometridae, Notodontidae, Noctuidae, Depressariidae, and Tortricidae). The locations of Caterpillars Count! sites with a minimum of 20 surveys conducted in June 2019 are shown as open circles in panels (**b**) and (**e**)

Because many iNaturalist observations are presumably from gardens, yards, and nonwoody substrates, we expected that by filtering those observations to the most common families found on woody vegetation that the correlation between the two datasets would become stronger. Limiting analysis to the six most common families named above actually weakened the correlation between iNaturalist and Caterpillars Count! datasets (Fig. 2f; r = 0.396, p = 0.144, n = 15 hex cells). In particular, restricting the analysis to these families reduced iNaturalist caterpillar occurrence more severely than it did Caterpillar Count! density estimates in the Upper Midwest. While still positive, the weakness of this correlation between datasets may be due in part to the reduced number of iNaturalist observations within these focal families, and also highlights the remaining sources of uncertainty in the iNaturalist observation process, including geographic variation in the types of taxa, habitats, and substrates which observers sample.

Caterpillar density and diversity are well known to vary by host plant species (Futuyma and Gould 1979; Tallamy and Shropshire 2009; Singer et al. 2012; Shutt et al. 2019), and so the spatial patterns of caterpillar density in Fig. 2 are necessarily impacted by the plant species on which those observations were made. For the Caterpillars Count! dataset, unlike iNaturalist, we have information on host plant identity, and closer examination of three of the most commonly represented species reveals density patterns that appear to be host plant specific (Fig. 3). For example, on American beech (*Fagus grandifolia*), caterpillar density was highest at lower

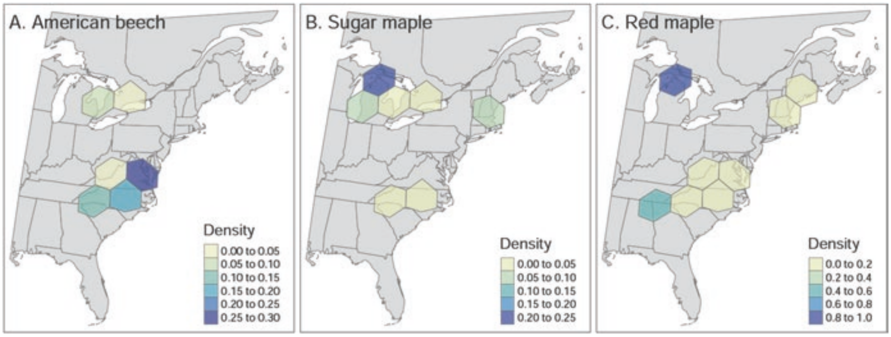

Fig. 3 Host-specific patterns of caterpillar density (number per survey branch) from Caterpillars Count! surveys on (**a**) American beech (*Fagus grandifolia*), (**b**) sugar maple (*Acer saccharum*), and (**c**) red maple (*Acer rubrum*)

latitudes, while on sugar maple (*Acer saccharum*) and red maple (*Acer rubrum*), density was highest at the northernmost site near Sault Ste. Marie, Ontario. Even within regions and on the same host plant, fine-scale gradients in temperature, humidity, and landscape context may lead to variation in caterpillar density (Kendeigh 1979; Jeffries et al. 2006; Reynolds et al. 2007; Smith et al. 2011; Seress et al. 2018), and thus these regional geographic patterns should be interpreted with caution pending more thorough characterization via the addition of more sampling sites within hex cells.

Phenology

Almost all of the work on the phenology of Lepidoptera has focused on modeling individual species and the variance between species (e.g., Diamond et al. 2014; Thorson et al. 2016; Belitz et al. 2020), but from the perspective of bird food, it is caterpillar phenology in aggregate that is of interest, and specifically the phenology during the avian nesting season (e.g., Visser et al. 2006; Lany et al. 2016; Shutt et al. 2019), which varies latitudinally but can broadly be considered to span May through July in North America. Thus, despite the often multimodal nature of aggregate caterpillar phenology from shortly after leaf out through the fall (Fig. 4), we estimate two phenometrics related to the period of peak caterpillar activity that should be most relevant to birds: (1) the peak caterpillar date between May 15 and July 30 and (2) the temporal centroid of caterpillar density or occurrence during this period. The former measure reflects when food resources for birds were ostensibly at their highest point, but may fail to capture shifts in the overall caterpillar distribution if there are multiple peaks of similar magnitude (Fig. 4). The latter measure captures variation in the center of mass of caterpillar observations even when the actual peak date does not vary.

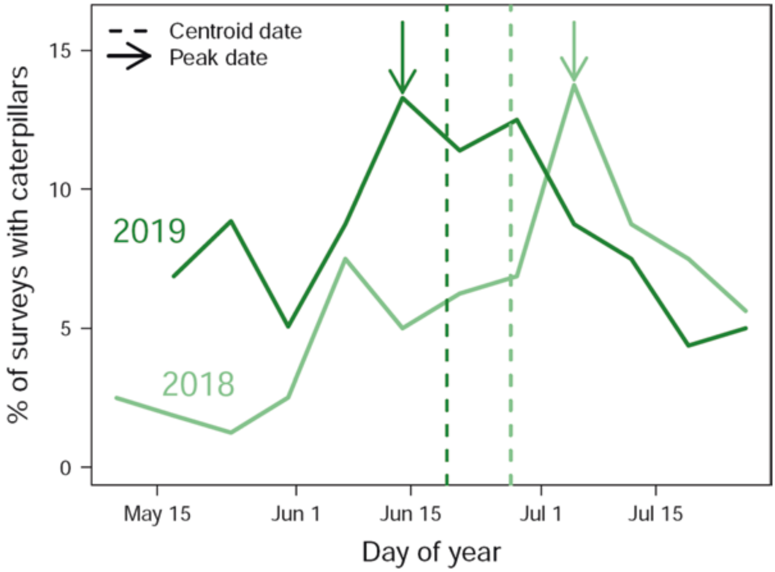

Fig. 4 Caterpillar phenology from Caterpillars Count! surveys at the North Carolina Botanical Garden in 2018 and 2019, illustrating the two phenometrics used, centroid date and peak date

For Caterpillars Count! data in 2019, we used data from all sites that had at least 20 branch surveys over at least 6 distinct sampling weeks spanning May 15 to July 30 and calculated the fraction of surveys detecting a caterpillar on any given date. When there were multiple Caterpillars Count! sites within one hex cell, we calculated the peak or centroid date for each site separately, and then averaged values across sites within a cell. For each hex cell with at least 100 iNaturalist insect observations per week, we calculated caterpillar phenology from iNaturalist data over the 6–10-week time window that matched the period of Caterpillars Count! data collection in that cell. Observations were binned by week to account for a "weekend effect," whereby users are more likely to contribute observations on Saturdays and Sundays (Courter et al. 2013; Di Cecco et al. 2021).

Because caterpillar phenology is tied to leaf out and both are closely related to spring temperatures (van Asch and Visser 2007; Uelmen et al. 2016), we also calculated average spring temperatures in each hex cell from March to June using Daymet climate data at 1 km resolution (Thornton et al. 2019) to investigate how standardized and opportunistic citizen science surveys capture relationships between caterpillar occurrence, density, and phenology and temperature.

Peak caterpillar date was poorly correlated between the datasets (Fig. 5a–c, $r = 0.330$, $p = 0.352$, $n = 10$ hex cells). The correlation was slightly stronger for centroid date (Fig. 5d–f, $r = 0.472$, $p = 0.169$, $n = 10$ hex cells), which also exhibited smoother variation with latitude (later dates at higher latitudes) than peak date. These results suggest that peak date in particular may be an overly sensitive phenometric, at least when estimated from an underlying distribution that is often

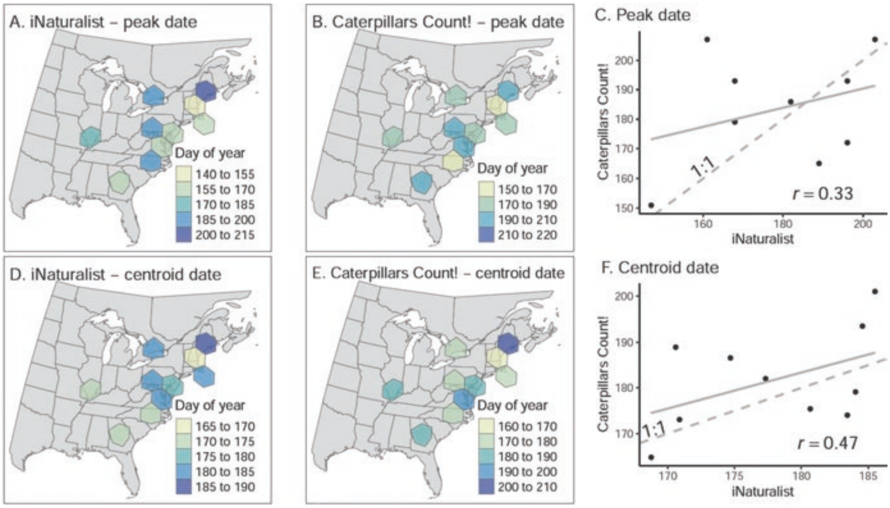

Fig. 5 Geographic variation in caterpillar peak date in 2019 based on data from (**a**) iNaturalist and (**b**) Caterpillars Count!. (**c**) The relationship between peak date estimates from the two datasets. Geographic variation in caterpillar centroid date in 2019 based on data from (**d**) iNaturalist and (**e**) Caterpillars Count!. (**f**) The relationship between centroid date estimates from the two datasets

multimodal and from a time series of only 6–10 weekly data points. Peak date is also expected to be more sensitive to single species outbreaks which may vary from year to year and region to region. Centroid date is a more promising phenometric, however, and one for which opportunistic and heterogeneous iNaturalist data may actually convey useful information about the geographic signal of forest caterpillar phenology.

Although the geographic variation in centroid date with latitude may be due, at least in part, to variation in spring temperature, phenometrics might also vary from cell to cell due to regional differences in the habitats, host plants, and Lepidoptera fauna sampled, and the nature of this cell-to-cell variation may differ in magnitude between iNaturalist and Caterpillars Count!. As such, it is perhaps not surprising that these two datasets exhibit different relationships between phenometrics, latitude, and spring temperature (e.g., centroid date-temperature relationship: iNaturalist, $R^2 = 0.223$, $p = 0.168$; Caterpillars Count!, $R^2 = 0.009$, $p = 0.791$, $n = 10$ hex cells). However, a comparison of phenometrics between years within the same hex cell is more likely to hold some of these sources of heterogeneity constant, and thus we might expect iNaturalist and Caterpillars Count! data to agree on the extent to which a year is early or late for a specific hex cell. As such, we identified all of the hex cells with sufficient Caterpillars Count! and iNaturalist data for 2 or more years from 2015 to 2019. For each dataset-cell-year and for each phenometric (peak date and centroid date), we calculated the difference between the phenometric in that year and the mean phenometric across years within that cell. All phenometrics within each dataset were thus represented as deviations from that cell's mean. We

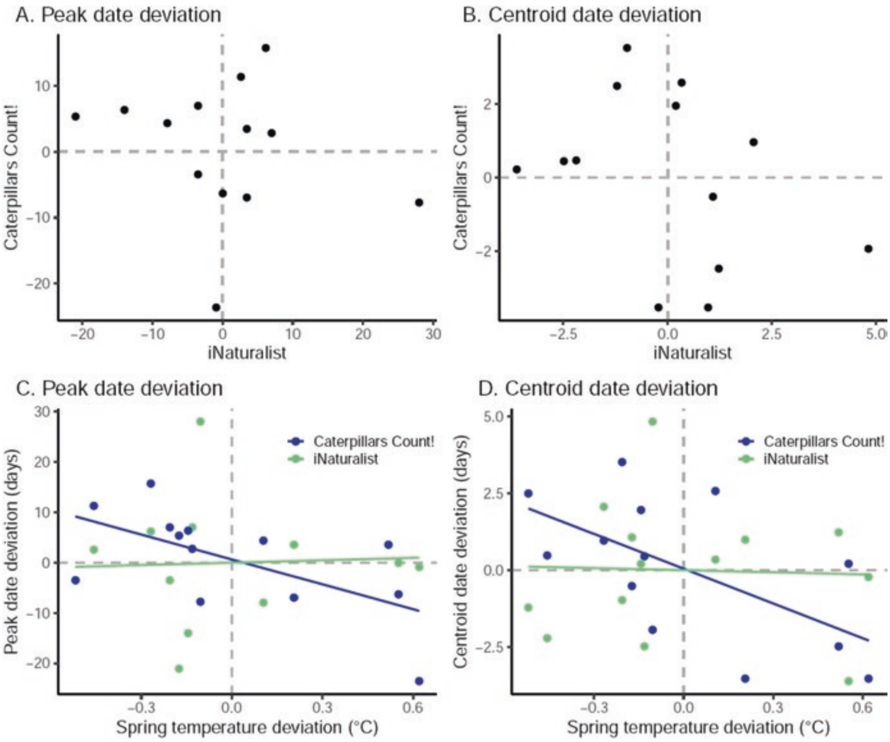

Fig. 6 The relationship between interannual deviations in phenology estimates between iNaturalist and Caterpillars Count! based on (**a**) peak date and (**b**) centroid date. (**c**) Peak date deviations and (**d**) centroid date deviations in caterpillar density as a function of interannual deviation in spring (March through June) temperature for iNaturalist (green) and Caterpillars Count! (blue)

were then able to ask whether there was agreement between iNaturalist and Caterpillars Count! in which years were early or late relative to that mean, as well as how interannual deviations in caterpillar phenology were related to interannual deviations in spring temperature.

Unfortunately, there was effectively no relationship between interannual deviations in either phenometric between the systematic monitoring of Caterpillars Count! and the opportunistic observations reported to iNaturalist (peak date, Fig. 6a, $r = -0.217$, $p = 0.477$; centroid date, Fig. 6b, $r = -0.354$, $p = 0.235$, $n = 13$ hex cell years). Within the Caterpillars Count! dataset, years with relatively early caterpillar phenology were also warmer, while years with later phenology were cooler (peak date, Fig. 6c, $R^2 = 0.367$, $p = 0.028$; centroid date, Fig. 6d, $R^2 = 0.364$, $p = 0.029$; blue lines, $n = 13$ hex cell years). While similar in variance explained, the parameter estimates differed substantially for the two phenometrics. A shift in centroid date by ~4 days per degree warming indicates the impact of temperature on the caterpillar fauna as a whole. In contrast, because peak date may be more sensitive to individual species outbreaks, a shift in peak date of ~17 days per degree suggests that a seasonally distinct caterpillar fauna is achieving higher abundances under warming than the fauna present during the peak period in cooler years (e.g., Fig. 4), either through direct effects of temperature or indirectly through phenological shifts in synchrony with host plants. Conversely, phenometric deviations within the iNaturalist dataset showed no relationship to interannual temperature deviations (peak date, Fig. 6c, $R^2 = 0.003$, $p = 0.868$; centroid date, Fig. 6d, $R^2 = 0.001$, $p = 0.902$; green lines, $n = 13$ hex cell years).

These results suggest that the observation process by which iNaturalist records are obtained has too many sources of variation to reliably characterize interannual variation in aggregate caterpillar phenology at the scale of these hex cells. Sources of variation include the numbers, spatial distribution, temporal distribution, and habitat representation of observations, phenological variation in the component caterpillar species among which there are early, late, and multivoltine species, as well as other aspects regarding the intent and behavior of individual iNaturalist users (e.g., taxonomic biases, activity of "superusers," etc., Di Cecco et al. 2021). Analyzing the phenology of individual Lepidoptera species with these types of data is more promising (Belitz et al. 2020), and as the iNaturalist dataset continues to grow exponentially, joint dynamic species distribution models and related approaches may improve the ability to detect signal from noise in multispecies assemblages (Thorson et al. 2016). At present, however, modeling caterpillar phenology in aggregate may require more standardized monitoring approaches like that used by the Caterpillars Count! project (Hurlbert et al. 2019).

Climate Change and Phenological Mismatch

Many organisms that depend directly or indirectly on temperature have been observed to shift phenology earlier in recent decades (Parmesan and Yohe 2003; Hurlbert and Liang 2012; Cook et al. 2012), including adult Lepidoptera (Kharouba et al. 2014; Diamond et al. 2014). The dependence of aggregate caterpillar phenology (as opposed to the phenology of any one species) on spring temperature (sometimes indirectly via the timing of green-up) has previously been shown at individual study sites with just one or two dominant caterpillar species (e.g., Visser et al. 2006; Burgess et al. 2018), but we show here that this applies across a much greater geographical extent spanning a much more diverse lepidopteran fauna. Current global climate change scenarios predict continued warming of anywhere from 1.7 to 4.8 °C globally by 2100 (Collins et al. 2013), suggesting that caterpillar phenology will continue to shift earlier. Such phenological shifts are expected both from the direct control of temperature on developmental rates (Knapp and Casey 1986; Gillooly et al. 2002) and from selection imposed by the earlier leaf out of vegetation and hence the earlier incorporation of secondary compounds that reduce leaf quality for herbivores later in the season (Feeny 1970; Martel and Kause 2002). The actual strength of selection for earlier phenology may depend on the relative degree of plasticity in responses by caterpillars and their host plants to changing temperature.

Continued shifts in caterpillar phenology, as well as potentially increased variability in phenology due to an increased frequency in extreme climatic events (Collins et al. 2013), may have negative consequences for organisms from higher trophic levels that depend heavily on caterpillars as a food resource. In particular, many species of foliage-gleaning birds have been found to rely heavily on caterpillars for raising their young (Holmes et al. 1979; Holmes and Schultz 1988; Sillett et al. 2000; Jones et al. 2003) and could face potentially reduced reproductive success if the reproductive period of high nestling resource demand does not shift in parallel with caterpillars. Negative fitness consequences have been demonstrated as a result of such phenological mismatch for resident great tits (*Parus major*) in the Netherlands (Visser et al. 2006; Reed et al. 2013), which have not advanced the timing of breeding as much as the shift in peak caterpillar date. Phenological mismatch is of even greater concern for long-distance migrants which have shown less sensitivity to interannual variation in breeding ground conditions compared to residents and short-distance migrants (Saino et al. 2011; Hurlbert and Liang 2012; Youngflesh et al. 2021) and which are presumably less able to accurately assess those conditions in distant breeding areas.

Several challenges confront research exploring the consequences of phenological mismatch between caterpillars and birds. First, the sensitivity of phenological responses appears to vary between species for both Lepidoptera (Kharouba et al. 2014; Diamond et al. 2014) and birds (Saino et al. 2011; Mayor et al. 2017), and also geographically within species (Hurlbert and Liang 2012; Youngflesh et al. 2021). This means that caterpillar phenology in aggregate may be difficult to predictively model without taking into account the individual responses of common species and species prone to outbreaks, which will vary regionally. Coupled with regional variation in avian sensitivity, an accurate understanding of phenological mismatch across a bird's geographic range will require the integration of avian and lepidopteran datasets spanning large extents and fine temporal resolution to fully unravel.

Examining the consequences of phenological mismatch also requires being able to quantify mismatch in a meaningful way. While the majority of studies that have compared caterpillar phenology with avian nesting phenology have focused on comparing shifts and differences between peak dates (e.g., Visser et al. 2006; Hinks et al. 2015), the height and width of caterpillar phenology curves may be equally or more important to birds than the timing alone (Shutt et al. 2019). Visser et al. (2015) found that an interaction between the height and timing of the caterpillar peak determined the number of nestlings fledged, with stronger seasonal selection during years with lower caterpillar peaks. Integrating the availability of high-value caterpillar prey over the entire nestling period may be the most relevant metric for predicting avian reproductive success, and comparing the integrated value of caterpillar availability during that period to the amount of nestlings would have experienced had the parents shifted reproduction earlier or later could provide a useful measure of phenological match. Both of these measures merit further research, especially with regard to whether and how the large volumes of existing opportunistic data might be integrated with information from structured surveys.

Conclusions

Caterpillars play a central role in forest ecosystems as both herbivores and as a food source for consumers, and citizen science datasets present an increasingly useful resource for understanding spatial and temporal patterns of caterpillar occurrence, abundance, and diversity. Projects like iNaturalist that consist of opportunistic photo observations have proven useful for mapping species distributions (Chardon et al. 2015; Fourcade 2016; Feldman et al. 2020) and for modeling the phenology of individual species (Taylor and Guralnick 2019; Barve et al. 2020). Estimates of caterpillar density, however, require knowledge of the total survey effort expended, and therefore are best obtained through standardized sampling protocols like those of the Caterpillars Count! project that report absences as well as presences. Nevertheless, we found that by scaling caterpillar observations by the total number of insect observations, iNaturalist was still able to recapitulate some of the geographic variation in density observed in the Caterpillars Count! dataset. As these citizen science data collection efforts continue into the future, they will provide a critical means of assessing abundance trends and the impacts of climate and land use change.

Finally, estimates of phenological timing of caterpillars in aggregate were less well correlated between the datasets across both space and time. Until methods are developed to better understand and deal with the sources of uncertainty and bias in the sampling process underlying opportunistic datasets like iNaturalist, geographically broadscale attempts to estimate phenological mismatch between caterpillars, and their avian predators will need to rely heavily on more systematic monitoring efforts.

Acknowledgments We thank the thousands of volunteers and amateur naturalists whose observations shared with citizen science projects like iNaturalist and Caterpillars Count! have enhanced scientific understanding of caterpillar diversity, abundance, and distribution. We also thank M. Singer, J. Forrest, and one anonymous reviewer for comments on an earlier draft of this chapter, and R. Marquis and S. Koptur for the invitation to contribute to this volume.

Literature Cited

Barve VV, Brenskelle L, Li D et al (2020) Methods for broad-scale plant phenology assessments using citizen scientists' photographs. Appl Plant Sci 8:e11315. https://doi.org/10.1002/aps3.11315

Belitz MW, Larsen EA, Ries L, Guralnick RP (2020) The accuracy of phenology estimators for use with sparsely sampled presence-only observations. Methods Ecol Evol 11:1273–1285. https://doi.org/10.1111/2041-210X.13448

Burgess MD, Smith KW, Evans KL et al (2018) Tritrophic phenological match–mismatch in space and time. Nat Ecol Evol 2:970–975. https://doi.org/10.1038/s41559-018-0543-1

Chardon NI, Cornwell WK, Flint LE et al (2015) Topographic, latitudinal and climatic distribution of *Pinus coulteri*: geographic range limits are not at the edge of the climate envelope. Ecography 38:590–601. https://doi.org/10.1111/ecog.00780

Collins M, Knutti R, Arblaster J, et al (2013) Long-term climate change: projections, commitments and irreversibility. In: Climate change 2013-the physical science basis: contribution of working group I to the Fifth Assessment report of the intergovernmental panel on climate change. Cambridge University Press, pp 1029–1136

Cook BI, Wolkovich EM, Davies TJ et al (2012) Sensitivity of spring phenology to warming across temporal and spatial climate gradients in two independent databases. Ecosystems 15:1283–1294. https://doi.org/10.1007/s10021-012-9584-5

Courter JR, Johnson RJ, Stuyck CM et al (2013) Weekend bias in citizen science data reporting: implications for phenology studies. Int J Biometeorol 57:715–720. https://doi.org/10.1007/s00484-012-0598-7

Diamond SE, Cayton H, Wepprich T et al (2014) Unexpected phenological responses of butterflies to the interaction of urbanization and geographic temperature. Ecology. https://doi.org/10.1890/13-1848.1

Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH (2021) Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. BioScience biab093. https://doi.org/10.1093/biosci/biab093

Feeny P (1970) Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars. Ecology 51:565–581. https://doi.org/10.2307/1934037

Feldman MJ, Imbeau L, Marchand P, et al (2020) Trends and gaps in the use of citizen science derived data as input for species distribution models: a quantitative review. bioRxiv 2020.06.01.127415. https://doi.org/10.1101/2020.06.01.127415

Fourcade Y (2016) Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. Ecol Inform 36:8–14. https://doi.org/10.1016/j.ecoinf.2016.09.002

Futuyma DJ, Gould F (1979) Associations of plants and insects in deciduous forest. Ecol Monog 49:33–50. https://doi.org/10.2307/1942571

Gillooly JF, Charnov EL, West GB et al (2002) Effects of size and temperature on developmental time. Nature 417:70–73

Hinks AE, Cole EF, Daniels KJ et al (2015) Scale-dependent phenological synchrony between songbirds and their caterpillar food source. Am Nat 186:84–97. https://doi.org/10.1086/681572

Holmes RT, Schultz JC (1988) Food availability for forest birds: effects of prey distribution and abundance on bird foraging. Can J Zool 66:720–728

Holmes RT, Schultz JC, Nothnagle P (1979) Bird predation on forest insects: an exclosure experiment. Science 206:462–463. https://doi.org/10.2307/1749326

Hurlbert AH, Liang Z (2012) Spatiotemporal variation in avian migration phenology: citizen science reveals effects of climate change. PLoS ONE 7:e31662. https://doi.org/10.1371/journal.pone.0031662

Hurlbert A, Hayes T, McKinnon T, Goforth C (2019) Caterpillars Count! A citizen science project for monitoring foliage arthropod abundance and phenology. Citizen Sci Theory Practice 4:1. https://doi.org/10.5334/cstp.148

Isaac NJB, Pocock MJO (2015) Bias and information in biological records. Biol J Linn Soc 115:522–531. https://doi.org/10.1111/bij.12532

Jeffries JM, Marquis RJ, Forkner RE (2006) Forest age influences oak insect herbivore community structure, richness, and density. Ecol Appl 16:901–912

Jones J, Doran PJ, Holmes RT (2003) Climate and food synchronize regional forest bird abundances. Ecology 84:3024–3032. https://doi.org/10.2307/3449971

Kendeigh SC (1979) Invertebrate populations of the deciduous forest: fluctuations and relations to weather. University of Illinois Press

Kharouba HM, Paquette SR, Kerr JT, Vellend M (2014) Predicting the sensitivity of butterfly phenology to temperature over the past century. Glob Change Biol 20:504–514. https://doi.org/10.1111/gcb.12429

Knapp R, Casey TM (1986) Thermal ecology, behavior, and growth of gypsy moth and eastern tent caterpillars. Ecology 67:598–608. https://doi.org/10.2307/1937683

Lany NK, Ayres MP, Stange EE et al (2016) Breeding timed to maximize reproductive success for a migratory songbird: the importance of phenological asynchrony. Oikos 125:656–666. https://doi.org/10.1111/oik.02412

Mair L, Ruete A (2016) Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLOS ONE 11:e0147796. https://doi.org/10.1371/journal.pone.0147796

Martel J, Kause A (2002) The phenological window of opportunity for early-season birch sawflies. Ecol Entomol 27:302–307. https://doi.org/10.1046/j.1365-2311.2002.00418.x

Mayor SJ, Guralnick RP, Tingley MW et al (2017) Increasing phenological asynchrony between spring green-up and arrival of migratory birds. Sci Rep 7:1–10. https://doi.org/10.1038/s41598-017-02045-z

Nelson G, Ellis S (2018) The history and impact of digitization and digital data mobilization on biodiversity research. Philos Trans R Soc Lond B Biol Sci 374. https://doi.org/10.1098/rstb.2017.0391

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42. https://doi.org/10.1038/nature01286

Prysby M, Oberhauser KS (2004) Temporal and geographical variation in monarch densities: citizen scientists document monarch population patterns. In: Oberhauser KS, Solensky MJ (eds) Monarch butterfly biology and conservation. Cornell University Press, Ithaca, pp 9–20

Reed TE, Jenouvrier S, Visser ME (2013) Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. J Anim Ecol 82:131–144. https://doi.org/10.1111/j.1365-2656.2012.02020.x

Reynolds LV, Ayres MP, Siccama TG, Holmes RT (2007) Climatic effects on caterpillar fluctuations in northern hardwood forests. Can J For Res 37:481–491. https://doi.org/10.1139/X06-211

Ries L, Zipkin EF, Guralnick RP (2019) Tracking trends in monarch abundance over the 20th century is currently impossible using museum records. PNAS 116:13745–13748. https://doi.org/10.1073/pnas.1904807116

Saino N, Ambrosini R, Rubolini D et al (2011) Climate warming, ecological mismatch at arrival and population decline in migratory birds. Proc R Soc B Biol Sci 278:835–842. https://doi.org/10.1098/rspb.2010.1778

Seifert CL, Lamarre GPA, Volf M et al (2020) Vertical stratification of a temperate forest caterpillar community in eastern North America. Oecologia 192:501–514. https://doi.org/10.1007/s00442-019-04584-w

Seltzer C (2019) Making biodiversity data social, shareable, and scalable: reflections on iNaturalist and citizen science. Biodivers Inf Sci Stand 3:e46670. https://doi.org/10.3897/biss.3.46670

Seress G, Hammer T, Bokony V et al (2018) Impact of urbanization on abundance and phenology of caterpillars and consequences for breeding in an insectivorous bird. Ecol Appl 28. https://doi.org/10.1002/eap.1730

Shutt JD, Burgess MD, Phillimore AB (2019) A spatial perspective on the phenological distribution of the spring woodland caterpillar peak. Am Nat 194:E109–E121. https://doi.org/10.1086/705241

Sillett TS, Holmes RT, Sherry TW (2000) Impacts of a global climate cycle on population dynamics of a migratory songbird. Science 288:2040–2042. https://doi.org/10.1126/science.288.5473.2040

Singer MS, Farkas TE, Skorik CM, Mooney KA (2012) Tritrophic interactions at a community level: effects of host plant species quality on bird predation of caterpillars. Am Nat 179:363–374. https://doi.org/10.1086/664080

Smith KW, Smith L, Charman E et al (2011) Large-scale variation in the temporal patterns of the frass fall of defoliating caterpillars in oak woodlands in Britain: implications for nesting woodland birds. Bird Study 58:506–511. https://doi.org/10.1080/00063657.2011.616186

Soroye P, Ahmed N, Kerr JT (2018) Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. Glob Change Biol 24:5281–5291. https://doi.org/10.1111/gcb.14358

Tallamy DW, Shropshire KJ (2009) Ranking lepidopteran use of native versus introduced plants. Conserv Biol 23:941–947. https://doi.org/10.1111/j.1523-1739.2009.01202.x

Taylor SD, Guralnick RP (2019) Opportunistically collected photographs can be used to estimate large-scale phenological trends. bioRxiv 794396. https://doi.org/10.1101/794396

Telenius A (2011) Biodiversity information goes public: GBIF at your service. Nord J Bot 29:378–381. https://doi.org/10.1111/j.1756-1051.2011.01167.x

Thornton PE, Thornton MM, Mayer BW, et al (2019) Daymet: daily surface weather data on a 1-km grid for North America, Version 3. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1328

Thorson JT, Ianelli JN, Larsen EA et al (2016) Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. Global Ecology and Biogeography 25:1144–1158. https://doi.org/10.1111/geb.12464

Uelmen JA, Lindroth RL, Tobin PC et al (2016) Effects of winter temperatures, spring degree-day accumulation, and insect population source on phenological synchrony between forest tent caterpillar and host trees. For Ecol Manag 362:241–250. https://doi.org/10.1016/j.foreco.2015.11.045

van Asch M, Visser ME (2007) Phenology of forest caterpillars and their host trees: the importance of synchrony. Annu Rev Entomol 52:37–55. https://doi.org/10.1146/annurev.ento.52.110405.091418

Visser M, Holleman L, Gienapp P (2006) Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. Oecologia 147:164–172. https://doi.org/10.1007/s00442-005-0299-6

Visser ME, Gienapp P, Husby A et al (2015) Effects of spring temperatures on the strength of selection on timing of reproduction in a long-distance migratory bird. PLOS Biol 13:e1002120. https://doi.org/10.1371/journal.pbio.1002120

Yang L, Jin S, Danielson P et al (2018) A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. ISPRS J Photogramm Remote Sens 146:108–123. https://doi.org/10.1016/j.isprsjprs.2018.09.006

Youngflesh C, Socolar J, Amaral BR, Arab A, Guralnick RP, Hurlbert AH, LaFrance R, Mayor SJ, Miller DAW, Tingley MW (2021) Migratory strategy drives species-level variation in bird sensitivity to vegetation green-up. Nat Ecol Evol 5:987–994. https://doi.org/10.1038/s41559-021-01442-y