PRACTICAL TOOLS

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

iPhenology: Using open-access citizen science photos to track phenology at continental scale

Yves P. Klinger[1] | R. Lutz Eckstein[2] | Till Kleinebecker[1]

[1]Division of Landscape Ecology and Landscape Planning, Justus Liebig University Gießen, Gießen, Germany

[2]Department of Environmental and Life Sciences, Biology, Karlstad University, Karlstad, Sweden

Correspondence
Yves P. Klinger
Email: yves.p.klinger@umwelt.uni-giessen.de

Handling Editor: Hooman Latifi

Abstract

1. Photo observations are a highly valuable but rarely used source of citizen science (CS) data. Recently, the number of publicly available photo observations has increased strongly, for example, due to the use of smartphone applications for species identification. This has enabled the raising of ecological insights in poorly studied subjects. One of the fields with the highest potential to benefit from the use of photo observations is phenology.
2. We propose a workflow for iPhenology, the use of publicly available photo observations to track phenological events at large scales. The workflow comprises data acquisition, cleaning of observations, phenological classification and modelling spatiotemporal patterns of phenology. We explore the suitability of iPhenology to observe key phenological stages in the plant reproductive cycle of a model species and discuss limitations and future prospects of the approach using the example of an invasive species in Europe.
3. We show that iPhenology is suitable to track key phenological events of widespread species. However, the number and quality of available observations may differ among species and phenological stages.
4. Overall, publicly available CS photo observations are suitable to track key phenological events and can thus significantly advance the knowledge on the timing and drivers of plant phenology. In future, integrating the workflow with automated image processing and analysis may enable real-time tracking of plant phenology.

KEYWORDS
iEcology, image analysis, photo observation, plant ecology, workflow

INTRODUCTION

Citizen science (CS) comprises the participation of nonprofessional volunteers in scientific projects or investigations. During the last decades, CS has enabled data collection at unprecedented scales (Dickinson et al., 2012). Among the different types of data collected by citizen scientists, photo observations are an invaluable but underused source of research data (Depauw et al., 2022). Recently, the collection of photos potentially available for ecological research has strongly increased, mainly due to the use of smartphone applications for species identification, such as iNaturalist (inaturalist.org), Pl@ntNet (plantnet.org) or Flora incognita (Mäder et al., 2021). These applications have become widely popular during the last years with more than 12 million downloads for the abovementioned apps alone.

This has drastically increased both the number and range of photo observations. Many photos are (semi-)automatically used to populate databases such as the Global Biodiversity Information Facility (GBIF) (gbif.org). This upsurge enables the raising of ecological insights in relatively poorly studied subjects or processes.

Phenology remains one of the most understudied aspects of plant functional ecology (Garnier et al., 2016). The capability to reach pivotal reproductive stages such as bud burst, flowering and seed production under differing climates is essential to persist across geographical ranges. Consequently, studies on the latitudinal distribution of species can give valuable information on their climatic niches and are crucial to predict future range shifts under a warming climate. Despite their relevance, empirical phenological studies across large geographical scales are limited (but see Ludewig et al., 2022; Nordt et al., 2021). This is mainly because such studies require frequent simultaneous observations across different latitudes. This is very time-intensive and usually requires the cooperation of scientists from many different institutions, which makes observations costly and hard to organize. Accordingly, such projects can only observe a limited number of individuals, species and ecosystems, often under relatively artificial conditions. Due to these limitations, phenological research is one of the fields that can benefit the most from the use of CS photo observations.

Despite their potential in phenology (Depauw et al., 2022), few papers have yet explored the use of CS photo observations to track phenological events across large scales (but see Puchałka et al., 2022; Reeb et al., 2022). As CS data collected without guidance may be biased concerning their spatial and temporal distribution (Pötzelsberger et al., 2021; Tiago et al., 2017), it can be difficult for researchers interested in phenology to deal with challenges arising from the use of CS data. Here, we propose a workflow for iPhenology, the use of publicly available photos to track phenological events (following the iEcology definition of Jarić et al., 2020). In the workflow, we check the data for frequently occurring biases, provide an example of phenological classification and model the spatiotemporal patterns of two key phenological stages. We use the herbaceous species *Lupinus polyphyllus* to demonstrate the workflow of iPhenology. The species is particularly suitable, as it is widespread, has prominent flowers and fruits, and is easy to identify due to the lack of native relatives in Europe (Eckstein et al., 2023). Through the utilization of photo observations, we track the flowering and fruiting stages in the invaded range across Europe. In particular, we address the following questions:

1. Can key phenological stages be tracked at the continental scale using iPhenology?
2. What are the current limitations and future prospects of iPhenology?

METHODS

The workflow for iPhenology comprises the acquisition of photo observations, pre-processing, and phenological classification

(Figure 1). With the resulting phenological observations, a plethora of potential analysis is possible, for example, assessing spatiotemporal patterns or modelling climatic drivers of phenology (see e.g. Puchałka et al., 2022).

Image acquisition and pre-processing

We acquired data using the GBIF (gbif.org). In the GBIF query, *L. polyphyllus* was selected based on human observations providing at least one image. The query covered the distribution of the species across Central and Northern Europe and included photo observations from −4.6° to 41.3° Longitude and from 42.3° to 69.7° Latitude for the years 2018–2021. The observations originate from several sources (e.g. iNaturalist Research-grade observations, Plant.net observations, Swedish Species Observation Service records and others). The full original dataset contains 8429 observations and is available under https://doi.org/10.15468/dl.jnxvnn, classified observations are available under https://doi.org/10.5061/dryad.h70rxwdpk.

We used the CoordinateCleaner package (Zizka et al., 2019) to identify and remove observations with problematic coordinates (e.g. locked to centroids of administrative areas or located in the ocean). Furthermore, coordinates with an uncertainty of >20 km were removed from the dataset. Additionally, observations with identical coordinates were removed. After cleaning the data, we used a custom R-script to access links to the images in each GBIF observation (Appendix S1). Images were then manually classified according to their phenology using a modified version of the R-Shiny application found in the appendix of Puchałka et al. (2022). Classification of ~8,000 images took approximately one working day (8 h) each for a researcher and a student assistant experienced in phenological classification.

For phenological classification, we used photo observations showing a full stand, plant or inflorescence. If multiple images were associated to one observation, we checked the first image for suitability. Originally, we distinguished between seven phenological stages, based on flowering and fruiting phenology, which were then aggregated to four main stages (*vegetative*, *flowering*, *fruiting* and *open pods*; Table S1). Classification was based on the majority of flowers, individuals were classified to be flowering and fruiting at the same time if both flowers and fully developed fruits were visible. Observations were assigned to the class *open pods* when any opened pods were visible. Images with low resolution, displaying only single leaves or small parts of inflorescences or with personal information/people visible were excluded from the analysis and classified as either *unsuitable* for phenological classification or *misidentified* if no specimen of *L. polyphyllus* could be found in the image.

Data analysis

For the analysis of spatiotemporal patterns of the flowering phenology, observations were dummy-coded to represent flowering
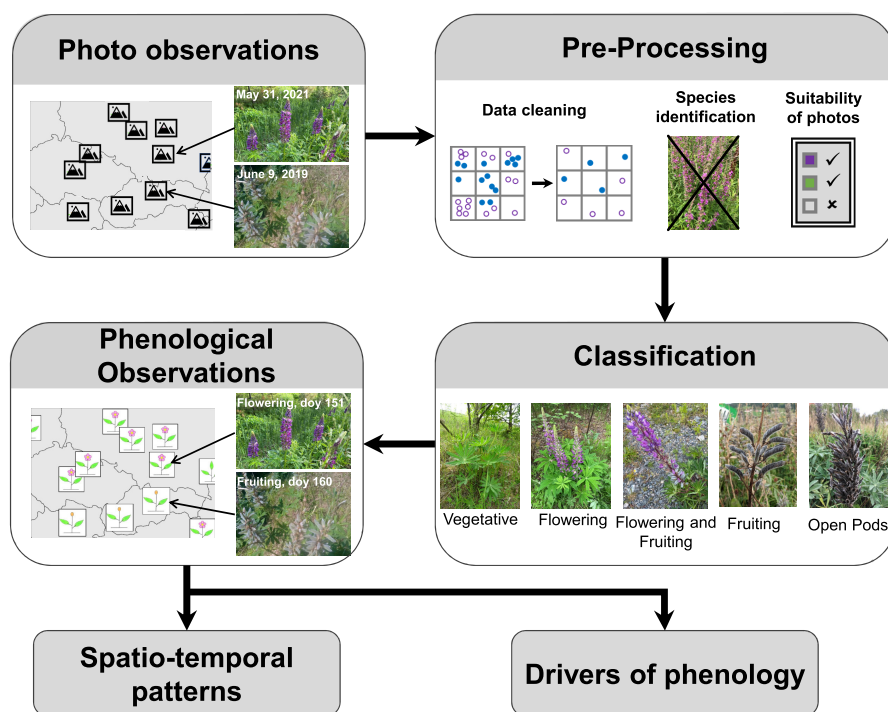
FIGURE 1 Proposed workflow for iPhenology. First, observations are pre-processed by removing problematic observation and, if necessary, reducing spatial aggregation. Second, photos are checked for correct identification and suitability before being classified. Unsuitable or misidentified photos are removed. For the resulting phenological observations, there are many potential uses.

(classes 2–5 in Table S1) vs. non-flowering (all others) and fruiting (class 6) vs. non-fruiting (all others) individuals. We used generalized additive models (GAMs) for binomially distributed response data to model the probability of flowering/fruiting of *L. polyphyllus* using restricted maximum likelihood estimation. GAMs provide a highly flexible framework for exploratory data analysis and can be used to model both spatial and temporal patterns in complex datasets (Wood, 2017). We included smoothers for the day of the year (doy), the coordinates (long/lat) and the interaction between both as explanatory variables. Year was included as factor to account for yearly variation in phenology. To address seasonal patterns emerging from phenological data, we included fixed knots for doy at the start and the end of each year. Elevation above sea level (GMTED, 2010) and if the observations were located in urban heat islands, areas with a distinct urban microclimate (CIESIN-Columbia University, 2016), were included first but later removed as they had no effect on flowering or fruiting phenology. Model assumptions were checked using the gam.check function. Analyses were performed using the MGCV package (Wood, 2017) in R 4.1.3 (R Core Team, 2022), plots were created using the GGPLOT2 package (Wickham, 2016).
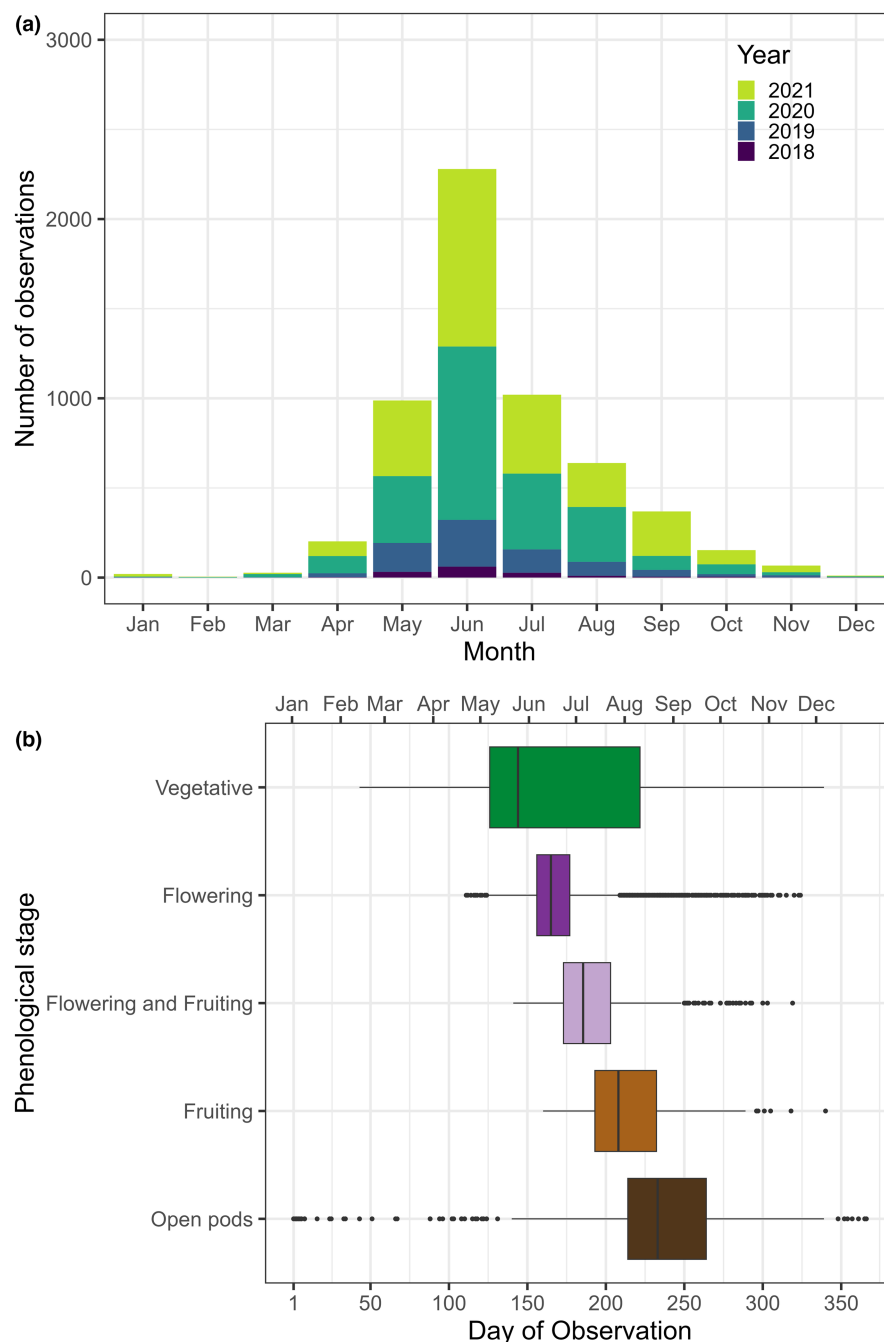
RESULTS

Spatiotemporal patterns

Suitable photo observations of *L. polyphyllus* ($n = 5780$) showed a distinct temporal pattern with a summer peak in June (Figure 2a), following the phenological cycle of the species (Figure 2b). The number of photo observations of *L. polyphyllus* increased nearly 10-fold from 185 observations in 2018 to 1646 observations in 2020 and stayed on a similar level in 2021 (Figure 2a). In the course of the year, most observations were performed in spring and summer (May–July), when the species is flowering. Thus, more than half of photo observations (2958) showed *flowering* or *flowering and fruiting* (640) individuals, whereas *vegetative* (1290) and *fruiting* (484) specimens were observed less frequently (Figures 2b and 3, Table S2). Almost half of suitable observations (48%) were carried out in urban heat islands (CIESIN-Columbia University, 2016). *Unsuitable* observations of *L. polyphyllus* ($n = 1570$) were mostly due to low image quality, images not allowing phenological classification, or the specimen being removed from its habitat. Only very few (23) observations were *misidentified*, that is they did not show an individual of *L. polyphyllus*.

Flowering and fruiting phenology

According to the model prediction, flowering of *L. polyphyllus* begins in the Southwest of the Europe and gradually extents further north (Figure 4). Additionally, southern populations of *L. polyphyllus* have either an extended or a second flowering period in July–August. Flowering probability of *L. polyphyllus* changed with doy ($p < 0.001$) and longitude/latitude ($p < 0.001$) and there was a significant interaction between both ($p < 0.001$, adjusted $R^2 = 0.55$). There were significant but minor differences in total flowering probability between the 3 years (2019–2021), with year 2021 having lower flowering probabilities than 2019 and 2020. Fruiting of *L. polyphyllus* was characterized by one peak between doy 180 and 220, depending on the location. As for flowering, the GAM model identified a distinct pattern of fruiting with changing doy ($p < 0.001$) and longitude/latitude ($p < 0.001$; Figure 5). According to the model, the

FIGURE 2 Temporal patterns of photo observations of *L. polyphyllus*. (a) Monthly number of photo observations for the years 2018–2021 (*n* = 5780). While the number of observations has increased between 2018 and 2020, the highest number of observations is carried out in June, when the species is flowering in Europe. (b) Box-Whisker-Plot showing the timing of observations classified to main phenological stages vegetative (*n* = 1290), flowering (*n* = 2958), flowering and fruiting (*n* = 640), fruiting (*n* = 484) and open pods (*n* = 408). Plot shows median ± 1,5-fold IQR.

DISCUSSION

Our results confirm that iPhenology, the use of publicly available CS photo observations, is suitable to track plant phenology across large geographical scales. It can be assumed that prominent and widespread species are most suitable for the approach. However, depending on the ecology of the target species and the behaviour of citizen scientists, there may be large differences between the observed phenological phases and the suitability of photos for phenological classification.

fruiting probability was lower compared to the flowering model, and the model captured less variation in the data (adjusted R^2 = 0.3).

In this regard, Puchalka et al. (2022) found that for the forest herb *Anemone nemorosa*, most observations comprised flowering individuals and available CS data were less suitable to track other phenological stages such as fruiting. In contrast to other CS data, photo observations can be checked for correct identification and suitability before further handling. For our model species, most images were suitable for phenological classification and photo observations covered the distributional range of *L. polyphyllus* across Europe (cf. Eckstein et al., 2023). Thus, they can be considered representative of the species' geographical distribution.

Despite the spatial aggregation of observations in densely populated areas, which is typical for CS observations (Speed et al., 2018), more than half of observations were carried out outside of highly
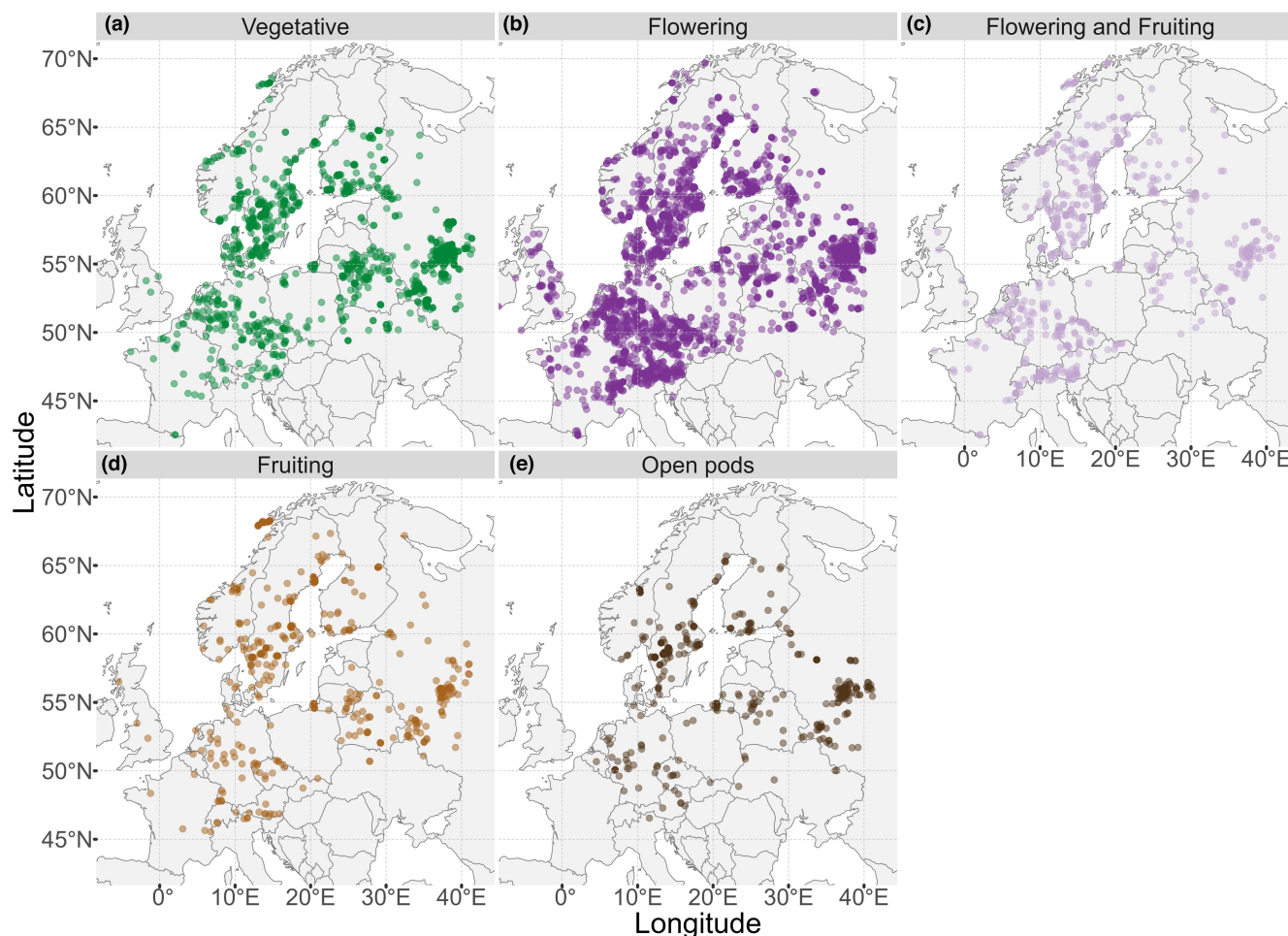
FIGURE 3 Spatial distribution of CS observations of different phenological stages 2018–2021. CS observations mostly comprise (b) flowering or (c) flowering and fruiting individuals, whereas (a) vegetative and (d) fruiting individuals are observed less frequently. Senescent plants with open seed pods (e) are observed the rarest

urbanized areas. However, it has been shown that opportunistic CS activities are prone to further spatial bias such as being more frequent in areas with higher accessibility (e.g. higher density of roads and footpaths, Tiago et al., 2017), or for some species groups in protected areas (Girardello et al., 2019). For plants, Tiago et al. (2017) also found a highly significant positive relationship between number of CS observations and the cover of forests and other (semi-)natural habitats. Since the habitat-specific species pool of these habitats is large (e.g. Jiménez-Alfaro et al., 2018; Pärtel et al., 2005) and many widespread non-native species reside in anthropogenic habitats such as roadsides (Meyer et al., 2021), we suggest that opportunistic CS observations may provide robust representations of the distributions of forest and grassland plants as well as many invasive species. To address potential issues, methods to identify spatial bias in the data, such as cleaning dubious coordinates (Zizka et al., 2019) may be necessary, although suitable tests and thresholds strongly depend on the data at hand (Zizka et al., 2020). In our workflow, we use GAMs to model the timing of flowering and fruiting across Europe. GAMs provide a flexible framework to model complex relationships and have been widely used for ecological data (Wood, 2017). Alternatively, tools to

reduce spatial aggregation such as spatial thinning (Aiello-Lammens et al., 2015) and spatial regression models (Kühn, 2007) may help address oversampling in densely populated areas, but require informed choices, as they can also negatively affect the performance of ecological models (Steen et al., 2021). Despite the potential of iPhenology, our results clearly show that not all phenological phases are observed to the same extent. Furthermore, less accessible areas and ecosystems may be under-represented in the dataset and for species not typically found in anthropogenic habitats, spatial bias may be more problematic compared to our model species.

Concerning temporal patterns, observations of *L. polyphyllus* clearly followed its phenological cycle (Ludewig et al., 2022). It has been shown that temporal patterns in iEcology data follow human interactions with species, for example, for Google searches on ticks in Denmark (Jensen et al., 2022). These patterns are probably amplified for species observations using mobile applications, where the observation process is part of the interaction. Furthermore, plant species are mostly observed during summer, as was shown for example, for Norway (Speed et al., 2018), when many plants in temperate regions are flowering and more people are outdoors carrying
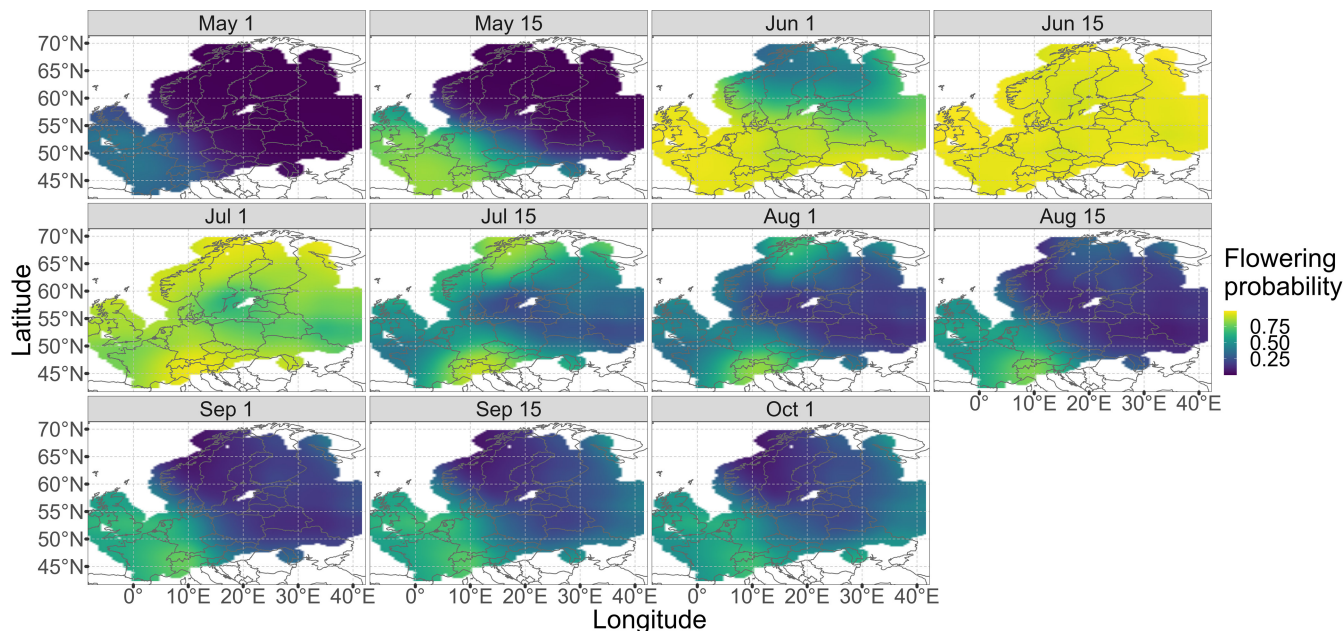
FIGURE 4 Predicted flowering probabilities of *L. polyphyllus* across Europe. Flowering happens between May and October. Northern populations flower later and shorter compared to Southern populations, that can have a second flowering period in August. Model predictions shown for year 2021.
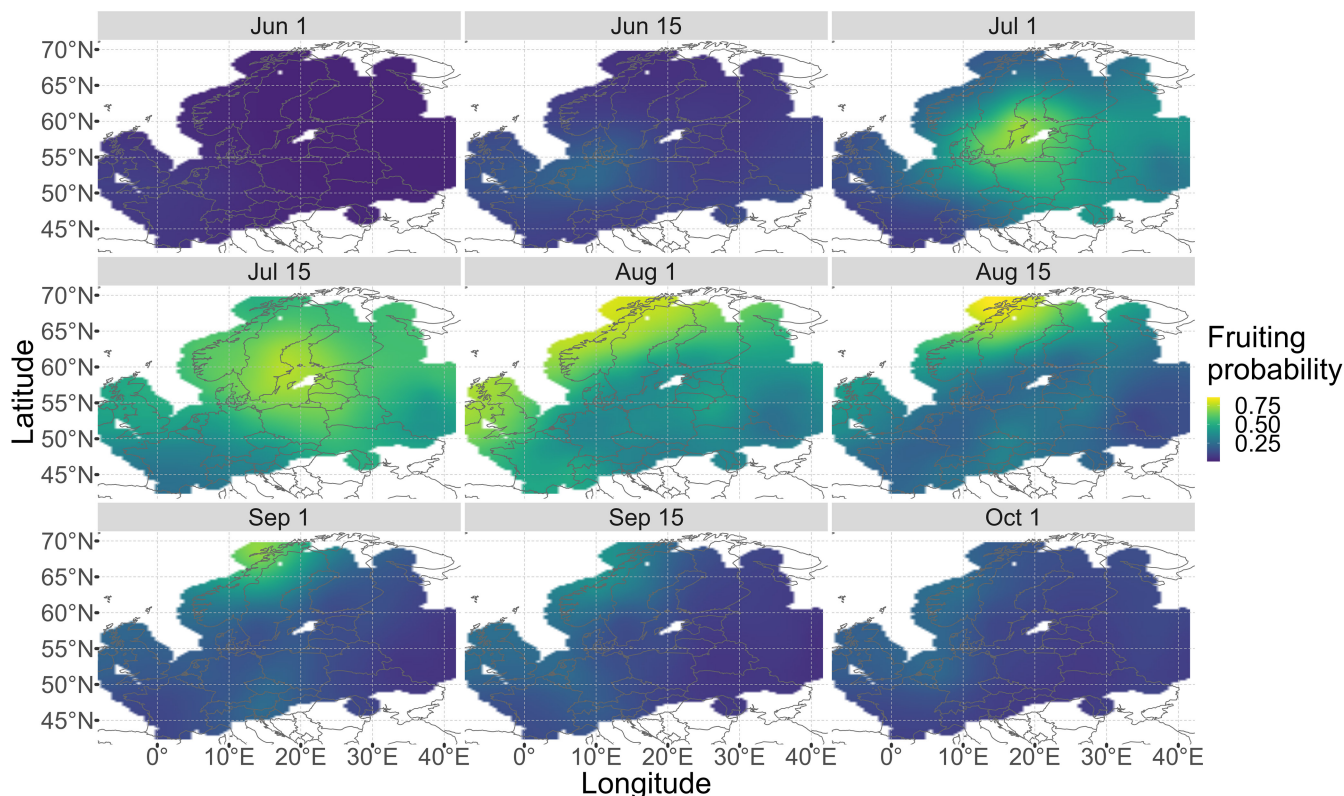

FIGURE 5 Predicted fruiting probabilities of *L. polyphyllus* across Europe. Fruiting happens between June and October. Northern populations fruit later and shorter compared to Southern populations. Compared to flowering, fruiting probability reaches a lower maximum (~0.8), as less fruiting specimens are observed. Model predictions for year 2021.

out recreational activities. This finding is confirmed in our dataset, where more than half of observations comprise flowering individuals. By identifying spatiotemporal patterns and potential bias in

the available observations, targeted CS assessments of underrepresented species, areas or phenological phases may help address such issues (Aavik et al., 2020).

Potential for future applications

iPhenology, the observation of phenological events using publicly available CS photo observations, is highly promising approach to advance phenological research for many widespread species. Among the many potential fields of application are comparing expert-based phenology data with CS data, modelling climatic drivers of phenology using CS observations or determining the right timing for the management of invasive alien species based on their phenology. In future, phenological classification of CS photos using deep learning may allow automated real-time assessments of phenological events for a vast number of species.

CONFLICT OF INTEREST STATEMENT
None.

DATA AVAILABILITY STATEMENT
All data used for this study are publicly available in the Global Biodiversity Information Facility (gbif.org; https://doi.org/10.15468/dl.jnxvnn), classified observations are available under https://doi.org/10.5061/dryad.h70rxwdpk.

ORCID
Yves P. Klinger https://orcid.org/0000-0001-6182-2217
R. Lutz Eckstein https://orcid.org/0000-0002-6953-3855
Till Kleinebecker https://orcid.org/0000-0003-1121-2861

REFERENCES
Aavik, T., Carmona, C. P., Träger, S., Kaldra, M., Reinula, I., Conti, E., Keller, B., Helm, A., Hiiesalu, I., Hool, K., Kaisel, M., Oja, T., Lotman, S., & Pärtel, M. (2020). Landscape context and plant population size affect morph frequencies in heterostylous Primula veris—Results of a nationwide citizen-science campaign. *Journal of Ecology*, *108*, 2169–2183.

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*, 541–545.

CIESIN-Columbia University. (2016). *Global Urban Heat Island (UHI) data set, 2013*. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4H70CRF

Depauw, L., Blondeel, H., De Lombaerde, E., De Pauw, K., Landuyt, D., Lorer, E., Vangansbeke, P., Vanneste, T., Verheyen, K., & De Frenne, P. (2022). The use of photos to investigate ecological change. *Journal of Ecology*, *110*, 1220–1236.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Eckstein, R. L., Welk, E., Klinger, Y. P., Lennartsson, T., Wissman, J., Ludewig, K., Hansen, W., & Ramula, S. (2023). Biological flora of Central Europe—Lupinus polyphyllus Lindley. *Perspectives in Plant Ecology, Evolution and Systematics*, *58*, 125715.

Garnier, E., Navas, M.-L., & Grigulis, K. (2016). *Plant functional diversity: Organism traits, community structure, and ecosystem properties*. Oxford University Press.

Girardello, M., Chapman, A., Dennis, R., Kaila, L., Borges, P. A. V., & Santangeli, A. (2019). Gaps in butterfly inventory data: A global analysis. *Biological Conservation*, *236*, 289–295.

GMTED. (2010). Global Multi-resolution Terrain Elevation Data 2010. https://doi.org/10.5066/F7J38R2N

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., Firth, J. A., Gaston, K. J., Jepson, P., Kalinkat, G., Ladle, R., Soriano-Redondo, A., Souza, A. T., & Roll, U. (2020). iEcology: Harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*, 630–639.

Jensen, P. M., Danielsen, F., & Skarphedinsson, S. (2022). Monitoring temporal trends in internet searches for "ticks" across Europe by Google trends: Tick–human interaction or general interest? *Insects*, *13*, 176.

Jiménez-Alfaro, B., Girardello, M., Chytrý, M., Svenning, J.-C., Willner, W., Gégout, J.-C., Agrillo, E., Campos, J. A., Jandt, U., Kącki, Z., Šilc, U., Slezák, M., Tichý, L., Tsiripidis, I., Turtureanu, P. D., Ujházyová, M., & Wohlgemuth, T. (2018). History and environment shape species pools and community diversity in European beech forests. *Nature Ecology & Evolution*, *2*, 483–490.

Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity & Distributions*, *13*, 66–69. https://doi.org/10.1111/j.1472-4642.2006.00293.x

Ludewig, K., Klinger, Y. P., Donath, T. W., Bärmann, L., Eichberg, C., Thomsen, J. G., Görzen, E., Hansen, W., Hasselquist, E. M., Helminger, T., Kaiskog, F., Karlsson, E., Kirchner, T., Knudsen, C., Lenzewski, N., Lindmo, S., Milberg, P., Pruchniewicz, D., Richter, E., … Eckstein, R. L. (2022). Phenology and morphology of the invasive legume Lupinus polyphyllus along a latitudinal gradient in Europe. *NeoBiota*, *78*, 185–206.

Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H. C., Deggelmann, A., & Wäldchen, J. (2021). The Flora incognita app—Interactive plant species identification. *Methods in Ecology and Evolution*, *12*, 1335–1342.

Meyer, S. E., Callaham, M. A., Stewart, J. E., & Warren, S. D. (2021). Invasive species response to natural and anthropogenic disturbance. In T. M. Poland, T. Patel-Weynand, D. M. Finch, C. F. Miniat, D. C. Hayes, & V. M. Lopez (Eds.), *Invasive species in forests and rangelands of the United States: A comprehensive science synthesis for the United States Forest sector* (pp. 85–110). Springer International Publishing.

Nordt, B., Hensen, I., Bucher, S. F., Freiberg, M., Primack, R. B., Stevens, A.-D., Bonn, A., Wirth, C., Jakubka, D., Plos, C., Sporbert, M., & Römermann, C. (2021). The PhenObs initiative: A standardised protocol for monitoring phenological responses to climate change using herbaceous plant species in botanical gardens. *Functional Ecology*, *35*, 821–834.

Pärtel, M., Bruun, H. H., & Sammul, M. (2005). *Biodiversity in temperate European grasslands: Origin and conservation*. Grassland Science in Europe.

Pötzelsberger, E., Gossner, M. M., Beenken, L., Gazda, A., Petr, M., Ylioja, T., La Porta, N., Avtzis, D. N., Bay, E., De Groot, M., Drenkhan, R., Duduman, M.-L., Enderle, R., Georgieva, M., Hietala, A. M., Hoppe, B., Jactel, H., Jarni, K., Keren, S., … Zlatkovic, M. (2021). Biotic threats for 23 major non-native tree species in Europe. *Scientific Data*, 8, 210.

Puchałka, R., Klisz, M., Koniakin, S., Czortek, P., Dylewski, Ł., Paź-Dyderska, S., Vítková, M., Sádlo, J., Rašomavičius, V., Čarni, A., De Sanctis, M., & Dyderski, M. K. (2022). Citizen science helps predictions of climate change impact on flowering phenology: A study on Anemone nemorosa. *Agricultural and Forest Meteorology*, 325, 109133.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reeb, R. A., Aziz, N., Lapp, S. M., Kitzes, J., Heberling, J. M., & Kuebbing, S. E. (2022). Using convolutional neural networks to efficiently extract immense phenological data from community science images. *Frontiers in Plant Science*, 12, 3148.

Speed, J. D. M., Bendiksby, M., Finstad, A. G., Hassel, K., Kolstad, A. L., & Prestø, T. (2018). Contrasting spatial, temporal and environmental patterns in observation and specimen based species occurrence data. *PLoS ONE*, 13, e0196417.

Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12, 216–226.

Tiago, P., Ceia-Hasse, A., Marques, T. A., Capinha, C., & Pereira, H. M. (2017). Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Scientific Reports*, 7, 12832.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

Zizka, A., Carvalho, F. A., Calvente, A., Baez-Lizarazo, M. R., Cabral, A., Coelho, J. F. R., Colli-Silva, M., Fantinati, M. R., Fernandes, M. F., Ferreira-Araújo, T., Moreira, F. G. L., Santos, N. M. C., Santos, T. A. B., dos Santos-Costa, R. C., Serrano, F. C., da Silva, A. P. A., de Souza Soares, A., de Souza, P. G. C., Tomaz, E. C., … Antonelli, A. (2020). No one-size-fits-all solution to clean GBIF. *PeerJ*, 8, e9916.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10, 744–751.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. R script to access image links.

Appendix S2. R script for iPhenology.

Appendix S3. Table S1. Classification of *L. polyphyllus* images to seven phenological classes and aggregation to four main phenological stages: vegetative, flowering, fruiting and open pods. All example images are Creative Commons Zero (CC0) images derived from iNaturalist (inaturalist.com).

Table S2. Descriptive statistics of phenological observations of *L. polyphyllus* between 2018 and 2021.

Figure S1. Spatial Distribution of CS observations of different phenological stages for the year 2018.

Figure S2. Spatial Distribution of CS observations of different phenological stages for the year 2019.

Figure S3. Spatial Distribution of CS observations of different phenological stages for the year 2020.

Figure S4. Spatial Distribution of CS observations of different phenological stages for the year 2021.

Figure S5. Visreg plots for the flowering model. Flowering probability changes with day of the year (doy), longitude and latitude. According to the model, there are two peaks of flowering for *L. polyphyllus*. Additionally, there are minor differences 2019-2021, where 2021 had the lowest flowering probability. Plots show model predictions for each variable using the median for numeric variables and most common category for factors for variables not shown in the respective panels. Assumptions: doy = 170, longitude = 17.7, latitude = 55.8, year = 2021.

Figure S6. Visreg plots for the fruiting model. Fruiting probability changed with day of the year (doy), longitude and latitude. Fruiting of *L. polyphyllus* in Europe is characterized by one fruiting period between doy 180 and doy 220. Accordin to the CS data, 2019 had higher fruiting probabilities compared to 2020 and 2021. Plots show model predictions for each variable using the median for numeric variables and most common category for factors for variables not shown in the respective panels. Assumptions: doy = 170, longitude = 17.7, latitude = 55.8, year = 2021.