ORIGINAL RESEARCH

Unstructured citizen science reduces the perception of butterfly local extinctions: the interplay between species traits and user effort

Elia van Tongeren[1,2] · Ginevra Sistri[2] · Marco Bonifacino[2] · Mattia Menchetti[3] ·
Lorenzo Pasquali[2] · Vania Salvati[2] · Emilio Balletto[4] · Simona Bonelli[4] ·
Alessandro Cini[5] · Mariagrazia Portera[6] · Leonardo Dapporto[1,2]

Abstract
The detection of local extinctions is often hindered by the lack of long-term monitoring schemes, and thus relies on time series of presence data. Recently, citizen science has repeatedly shown its value in documenting species occurrences. We investigated the effectiveness of unstructured citizen science records in reducing the perception of local extinctions in butterfly populations across Italian National Parks. We addressed three research questions: (i) the ability of citizen science data to supplement existing knowledge to complete time series of occurrences, (ii) the impact on data collection of three species features (species size, distribution and length of flight period) determining their appearance, and (iii) the interplay between participant effort and species appearance in the amount of diversity recorded on the iNaturalist platform. Our analysis of 98,922 records of Italian butterflies (39,929 from literature and 58,993 from iNaturalist of which 7427 from National Parks) showed that the addition of iNaturalist data filled many recent gaps in time series, thus reducing the perception of potential local extinctions. Records from more engaged users encompassed a higher fraction of local biodiversity and were more likely to reduce the perception of local extinctions. User effort strongly interacted with species features in determining the frequency of records for individual species. In particular, more engaged users were less affected by species size. We provided updated butterfly checklists for Italian National Parks and a R package to calculate potential extinction upon time series. These results offer guidance for protected areas, conservationists, policymakers, and citizen scientists to optimize monitoring of local populations.

Introduction

The Anthropocene epoch is characterized by a severe crisis in biodiversity, leading to a rapid and unprecedented global loss of species and genetic diversity (Ceballos et al. 2015; Turvey and Crees 2019). Insects are especially affected, with up to 40% of species at risk of extinction (Riley 1986; Habel et al. 2019; Sánchez-Bayo and Wyckhuys 2019). In some protected areas of Central Europe, insects have lost 75% of their biomass (Hallmann et al. 2017) and the species included in the grassland butterfly index declined by 39% in number of individuals in recent decades (Warren et al. 2021). This widespread decline could have significant impacts on ecosystems, as insects play a crucial role in ecological networks (Rosenberg et al. 1986; Yang and Gratton 2014; Noriega et al. 2018; Montgomery et al. 2020).

Despite the urgency for monitoring local extinctions and declines, species trends across space and time are largely unavailable, especially for less visible and unpopular taxa like most insects (Rocha-Ortega et al. 2021; Montgomery et al. 2020; Lobo 2016). Insects are also characterized by a taxonomic complexity that is often solvable only by specialists (Roskov et al. 2019; Marshall 2008; Hochkirch et al. 2022), so monitoring schemes are not available for most insect taxa or, if present, are limited to specific regions and run by professional taxonomists.

Due to the lack of long-term monitoring data, detecting extinctions at local and regional scales in many southern European biodiversity hotspots must be based on time series of occurrence data. Published records follow a typical pattern: there are a few large data-sets published in dedicated faunistic papers, rarely replicated over time, and many single records published over the years. This pattern reduces the completeness of time series. In recent decades, citizen science, the involvement of the public in scientific projects (Heigl et al. 2019; Fontaine et al. 2021; Chowdhury et al. 2023a), has repeatedly demonstrated its contribution to documenting species occurrences and spread of invasive species (Crall et al. 2010; Gallo and Waitt 2011; Maistrello et al. 2016; Mannino and Balistreri 2018) and in monitoring protected species and entire communities, as in the case of the European Butterfly Monitoring Scheme (eBMS) (Zapponi et al. 2017; Campanaro et al. 2017; Oberhauser and Prysby 2008; Warren et al. 2021). Citizen science activities provide both structured and unstructured data. Structured data come from organized initiatives that aim to involve the public in pre-defined objectives, such as monitoring a specific taxon (e.g. Callaghan et al. 2020; Krabbenhoft and Kashian 2020; Carpaneto et al. 2017). Unstructured data are provided randomly, without a predefined and taxon-specific (or area-specific) goal, and usually come from citizens uploading their observations on citizen science platforms (e.g. iNaturalist and eBird (Sullivan et al. 2009)). Records from unstructured citizen science are numerous and more distributed temporally and spatially than those from structured citizen science. However, they are more influenced by operator biases since the type and amount of data depend on the user's personal preferences (Isaac and Pocock 2015; Callaghan et al. 2021; Van Eupen et al. 2021). In this study, we assess the contribution of unstructured citizen science records in discarding local extinctions in the Italian National Parks. Specifically, we addressed three main research questions:

1. To what extent can unstructured citizen science data supplement existing knowledge on butterfly diversity obtained through literature to determine the persistence or local extinction of butterflies' populations? Unlike literature data, citizen science records are

more recent and have seen a significant rise in recent years (Fischer et al. 2021) offering the possibility of resolving concerns regarding local extinctions.

2. To what extent do data collected through citizen science on Italian butterflies depend on species appearance? The number of records uploaded on citizen science platforms may be influenced by species features. Larger species with a wider geographical distribution and longer flight periods might receive more records (Callaghan et al. 2021; Barbato et al. 2021; Stoudt et al. 2022; Van Eupen et al. 2022).

3. Does species appearance affect participants differently depending on their level of engagement in citizen science? A more dedicated involvement in citizen science activities may lead to the observation of less common and less noticeable species (Callaghan et al. 2021). We examined whether users demonstrating greater effort are influenced differently by species appearance and whether they document a larger portion of butterfly diversity compared to those with lower effort, with a similar number of observations.

To address these questions, we analyzed 39,929 butterfly records from literature and 58,993 verified records of Italian butterflies on iNaturalist, one of the best-known and most used citizen science platforms (Aristeidou et al. 2021; Echeverria et al. 2021; Cambria et al. 2021; Nugent 2018; Sanderson et al. 2021). We selected butterflies as a model taxon due to their ecological relevance (Cruden and Hermann-Parker 1979; Courtney et al. 1982; Jennersten 1984), their concerning conservation status (Franzén and Johannesson 2007; Dirzo et al. 2014; McDermott Long et al. 2017; Schultz et al. 2019) and their ease of monitoring by citizen scientists (van Swaay et al. 2008; Wei et al. 2016; Prudic et al. 2017).

Within Italy, National Parks are particularly suitable for our study. They are protected areas that host a high diversity of fauna and flora (Capotorti et al. 2012). Additionally, these are the most extensively studied areas, ensuring a large quantity and quality of literature data on species occurrence that can serve as a basis for evaluating how unstructured citizen science data can complement traditional scientific research. Finally, National Parks are also where citizen science activities occur most frequently, both as independent and organized activities such as bioblitzes (Lundmark 2003).

Two novel and significant resources are also provided: (a) updated butterfly checklists for all Italian National Parks (with time series of occurrence for each species starting from the year 1806—see *Danaus chrysippus* presence in Vesuvio National Park) and (b) a new R package (https://github.com/leondap/pets) that includes a suite of functions designed to calculate the potential extinction upon time series index (PETS) introduced by Labadessa et al. (2021). This is used to evaluate the importance of unstructured citizen science records in reducing the perception of local butterfly extinctions.

Materials and methods

Data collection

We collected butterfly occurrences from two sources (see Table 1 for definitions of the main terms used in the study):

1. Data collected in the Italian CkMap http://faunaitalia.it/documents/CKmap_ITA.pdf (Balletto et al. 2007). This resource contains literature data and butterfly specimen records from the main national collections. The database was published in 2007 and it

Table 1 Definition of the main terms used in the study

Term	Definition
CkMap	The Italian database mapping the distribution of Italian fauna at 10×10 km^2 resolution.
iNaturalist	A citizen science platform built on the concept of mapping and sharing observations of biodiversity across the globe.
PETS index	Potential Extinction upon Time Series index. Evaluate the potential that a community has been eroded in time based on time series.
PETS0	PETS0 index was obtained by considering data from Italian National Parks using dated records in CkMap.
PETS1	As PETS0 but obtained using both data from Ckmap and iNaturalist.
ΔPETS	The difference between PETS0 and PETS1, i.e. the contribution of iNaturalist data in reducing the perception of butterfly extinction within each National Park.
PETSc	The contribution of an individual record (ir) in lowering the PETS0 value, obtained by removing that record for the National Park then recalculating a PETS1ir value without it and obtaining the difference between PETS1ir and PETS1.
User effort	The number of records uploaded by each iNaturalist user on the platform.
Species features	Composed for each species of two functional traits (wingspan, WI and phenology Ph) and by species distribution in Italy obtained as the number of 10×10 km^2 cells in CkMap.
Diversity indices	Diversity encompassed by records from different groups of users has been scored based on number of species and their relative frequency by using Hill's number. For $q = 0$ Hill's formula returns species richness, for $q = 1$ the Shannon index and for $q = 2$ the Simpson index.

is continuously updated by EB. As of December 2021, it contains 335,499 records of Italian butterflies. The spatial resolution of the checklist is represented by 10×10 km^2. We only included in our analysis the occurrences whose square center is not more than 5 km away from a National Park perimeter. Out of the total of 39,929 records present in CkMap within the perimeter of Italian National Parks, 20,191 do not have a precise observation date, so they are unusable for our analyses and marked as NA data. Therefore the CkMap has a total of 19,738 usable data.

2. iNaturalist observations. We identified 58,993 records of Italian butterflies uploaded until December 2021 with a location error lower than 1000 m in order to select for high quality data. Among these data, 7427 records were included in National Parks perimeters.

We did not collect data from the Global Biodiversity Information Facility (GBIF) for the following reasons: (i) it is impossible to verify the species identification of citizen science data as pictures are not uploaded, (ii) it contains a mixture of data collected by professional recorders and citizen scientists, making it difficult to separate the two, (iii) in many cases, the observer is not recorded or cannot be precisely identified, which would impede our ability to reliably assess user effort.

All the collected data were organized in the Darwin Core format, a widely used standard format in biodiversity research applications (Wieczorek et al. 2009, 2012; Groom et al. 2019). The main fields for each record are "occurrenceID", which contains the specimen reference; "Scientific name" in genus species format; "Locality", where the occurrence is located; "decimalLatitude" and "decimalLongitude" that are the locality coordinates in decimal degrees; "basisOfRecord" that specifies the source of the records, "literature" for

CkMap records and "iNaturalist" for records obtained from that platform; "recordedby" is the source of bibliographic information, reference collection, or the iNaturalist user who uploaded the data; and finally "catalogNumber" contains the URL of iNaturalist observations. The Darwin Core file is available for each National Park in the repository of PETS package (https://github.com/leondap/pets/tree/main/data).

Aim 1—evaluating unstructured citizen science contribution in butterfly diversity monitoring

We assessed the Potential Extinction upon Time Series (PETS) to evaluate the role of citizen science in dispelling doubts about local extinctions. The PETS formula introduced by Labadessa et al. (2021) (Eq. 1) assesses the perception of local extinction in the past years due to effective species losses or to the absence of recent occurrence data.

$$PETS = \frac{\sum_{i=1}^{n} last\ year - last\ occ_i}{\sum_{i=1}^{n} \left(last\ year - first\ occ_i \right) + 1} \tag{1}$$

In PETS formula (Eq. 1) $first\ occ_i$ and $last\ occ_i$ are the years of the first and the last observation of the species i, respectively; $last\ year$ is the year of the assessment (end of the study, 2021 in this case) and it is the same for all species; n is the number of species recorded in the local butterfly community. The potential extinction index for each species is calculated based on the difference between the last year and the last record date (represented by the red bar in Fig. 2a), divided by the time since the first observation date (represented by the cyan bar in Fig. 2a). If all the species observed in the past have been observed in the last year, the PETS index is equal to zero. The output of PETS analysis includes the PETS index, the species list ordered by last observation date, and a graphic representation (Fig. 2b, c). In the graphs, each row on the Y axis represents a species and observation years on the X axis are marked as colored squares with the color indicating the source type. The species with more recent observations are displayed at the top, while older records are shown at the bottom.

For each National Park we calculated two PETS values: (i) The PETS0 index was obtained by considering literature data only (from CkMap) and (ii) the PETS1 index which includes literature and citizen science data (from both CkMap and iNaturalist). We used the "pets" R function of the newly created PETS R package which is freely available at: https://github.com/leondap/pets. The difference between PETS0 and PETS1 (ΔPETS) represents the contribution of iNaturalist records in dispelling the perception of local extinction in each National Park butterfly community.

We checked if the Parks characterized by a high fraction of records missing the year of collection (NA data) also showed a higher PETS0, PETS1 and ΔPETS by using Spearman's correlation.

Aims 2 and 3—the interplay between user effort, species traits and documented biodiversity

We assessed if users who put in different levels of effort produce records with a different value in establishing local extinction and in assessing local diversity. We scored the effort as the number of records for each user in the studied National Parks and transformed it by square root (effort in Eq. 2). To assess the value of each record we calculated its

contribution in establishing the PETS values (PETSc) as follows: for each National Park, we iteratively removed one iNaturalist record from the dataset and recalculated the PETS1 index without that single datum (PETS1ir). Then we scored PETSc as the absolute value of the difference between PETS1ir and PETS1. A Generalized Linear Mixed Model (GLMM) was used to verify if single records produced by users who put in more effort have a higher PETSc. National Parks and users were included as random factors. The contribution to the observed PETS1 was analyzed by using the following model.

$$PETSc \sim Effort + (1|Park) + (1|User) \qquad (2)$$

We selected a tweedie distribution for the response variable since it provides a flexible family to deal with non-negative highly right-skewed data as well as symmetric and heavy tailed data. We used the "glmmTMB" function of the glmmTMB R package (Brooks et al. 2017).

We then analyzed the relationship between butterfly species traits, user effort, and the frequency of species records in the entire Italian butterfly dataset. To do this, we used all 58,993 iNaturalist records for Italian butterflies identified by the authors. We first obtained the number of records for each species obtained by each user (records_sp_us, response variable in Eq. 3). Then, for each species we obtained two functional traits from Middleton-Welling et al. (2020) : (i) the wing index (WI), a measure of wing size, and a proxy for species visual appearance; (ii) a set of phenology traits (maximum and minimum number of flight months in Europe, first and last month of flight, number of generations) that were subjected to Principal Component Analysis (PCA) to obtain a single component (Ph, Fig. S1) (Dapporto et al. 2019), which represents species appearance due to the duration of adult flight period. After CkMap we also obtained (iii) the number of 10×10 km^2 UTM cells where the species have been recorded in Italy (Dis, representing a measure of species appearance based on their distribution). User effort was calculated as the number of butterfly records uploaded by each user.

A Generalized Linear Mixed Model (GLMM) was used to assess the effect of these three species features in determining the number of records per each species uploaded by each user. Interactions between species' traits and user effort were also included in the model. Species and users were included as random factors.

$$Records\_sp\_us \sim (WI + Ph + Dis) * Effort + (1|User) + (1|Species) \qquad (3)$$

Count data was analyzed using a Poisson family. We used the "glmmTMB" function of the glmmTMB R package. Type-III analysis-of-variance table was calculated using the "Anova" function of the car R package. The interactions have been visualized using the "plot_model" function of the sjPlot R package (Lüdecke 2023) with default settings.

Finally, we evaluated if comparable amounts of records from users with varying levels of effort result in different levels of species diversity. To do this, we arranged the users based on the increasing number of records they had. Then, we separated the data into ten quantiles by aggregating the observations from users who exhibit increasing levels of effort, until each quantile boundary was reached. This method ensured that each quantile contained the same number of records, but the first quantiles were comprised of data submitted by users who showed lower levels of effort compared to the latter ones. The species diversity for each quantile was calculated using Hill numbers: species richness (q=0), Shannon index (q=1) and Simpson index (q=2). These calculations were performed using the "hill_taxa" function of the hillR R Package (Li 2018). We used Spearman's tests to

identify possible correlations between diversity values and the different level of user effort across the ten quantiles.

Results

Aim 1—evaluating unstructured citizen science contribution in discarding local extinction

We obtained 47,356 records from the Italian National Parks, consisting of 39,929 occurrences from CkMap and 7427 from iNaturalist. While all observations collected by iNaturalist were provided with the year of collection, 50.6% of the records in CkMap, a total of 20,191, were not provided with this information and therefore not useful for analyzing time series.

The results of the PETS indexes for each National Park are displayed in Table S1 and arranged based on the contribution of iNaturalist in reducing the perception of local extinction, as indicated by the difference between PETS0 and PETS1 (ΔPETS).

The results show notable variations among the National Parks in terms of potential extinction of butterfly communities and the contribution of unstructured citizen science in reducing it (Fig. 1). Based on CkMap data only, nine Parks showed that over half of time series were represented by unconfirmed presences (PETS0 > 0.5, Fig. 1). When iNaturalist data were added to the occurrence datasets, only two Parks maintained a PETS1 higher than 0.5. The PETS index dropped from an average of 0.449 ± 0.227 (standard deviation) (for PETS0) to 0.279 ± 0.154 (standard deviation) (for PETS1), which indicates a marked reduction in the lack of knowledge about recent species occurrence (ΔPETS) of 0.170 ± 0.120 (standard deviation) (Fig. 1).

As a remarkable example, the Gargano National Park was found to have a rich butterfly fauna which was studied in two main campaigns during 1940 and 1950s. However, over half of the species were unrecorded since 1980 (Fig. 2b) resulting in a high PETS0 value of 0.524. Citizen science activities allowed for the confirmation of 56 species in the last 5 years of research (2017–2021) resulting in a large ΔPETS of 0.326 (Table S1 and Fig. 1 and 2b, c).

There were no significant correlations between the percentage of NA data in literature (records for species without a precise year of observation) and different PETS evaluations (PETS0, rho = 0.237, P = 0.255; PETS1, rho = 0.110, P = 0.599; ΔPETS, rho = 0.297, P = 0.149).

PETS graphs for all National Parks, and the butterfly species lists including the first and the last observations, can be found in the Supplementary Results document (Appendix 1).

Aims 2 and 3—the interplay between user effort, species traits and documented biodiversity

The GLMM analyzing the effect of user effort showed a significant positive relationship with the contribution of each observation to determine the PETSc values (Estimate = 0.099, Standard error = 0.035, z value = 2.869, P = 0.004). This demonstrates that single records from more committed users have a higher likelihood of reducing the perception of local extinction.
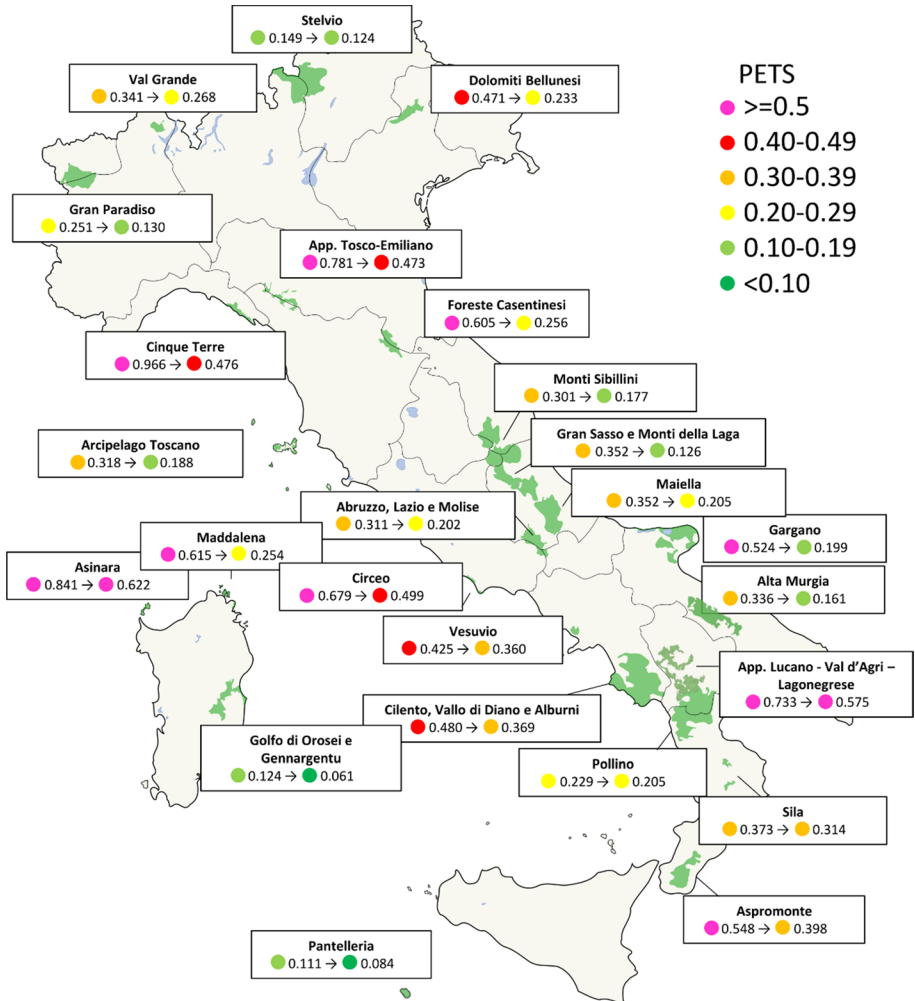
Fig. 1 The effect of iNaturalist data in reducing the potential exctinction for each Italian National Parks. In the map the National Parks are marked in green with the results obtained for PETS0 (left) and PETS1 (right) divided in classes and marked by dots colored from green, low potential extinction, to fuchsia, high potential extinction

The Italian data on butterflies collected on iNaturalist also showed that larger species with a wider distribution received a higher number of records (Table 2). The three species features had significant interactions with user effort (Table 2). Users with high effort (Fig. 3a) tended to record species with a wider distribution more frequently. This relationship was less evident for users with a low engagement (Fig. 3a). Similarly, the relationship between flight period and number of records was steeper for users with high engagement (Fig. 3b). Wingspan showed a different trend since users with higher effort tended to record smaller species more frequently while less engaged users showed a steeper and opposite trend, thus reporting mainly large species (Fig. 3c).

We found that a similar number of records uploaded by users with high effort encompassed a higher diversity in terms of the number of detected species (richness, q=0) and
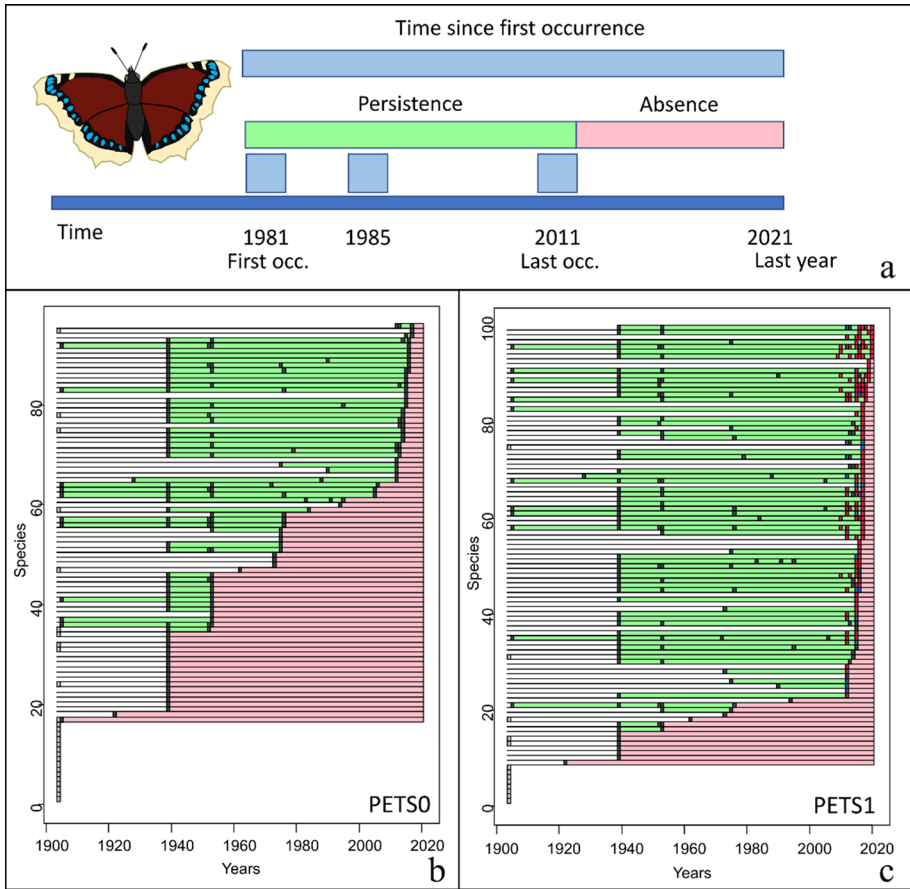
Fig. 2 **a** The rationale of the PETS algorithm to compute potential extinction based on three records (1981, 1985, and 2011) from 1981 to 2021. Same abbreviations as in Eq. 1. **b**, **c** The time series graphs produced by PETS analysis on the butterfly data of the Gargano National Park where each species is represented with a row with its records as in a. **b** The results for PETS0 where only literature records (CkMap, dark grey dots) are used to assess the potential for extinction and **c** the result for PETS1 where data from iNaturalist (in red) are also added. The years with both kinds of records for any given species are reported in blue

Table 2 The effect of the three variables of species appearance on the number of observations per user in a GLMM. Interactions with user effort (number of records per observer) are also reported		

Variable	Chisq	P
Distribution	23.326	<0.001
Flight period	0.152	0.697
Wingspan	6.306	0.012
Distribution*User effort	508.641	<0.001
Flight period*User effort	47.809	<0.001
Wingspan*User effort	19.882	<0.001
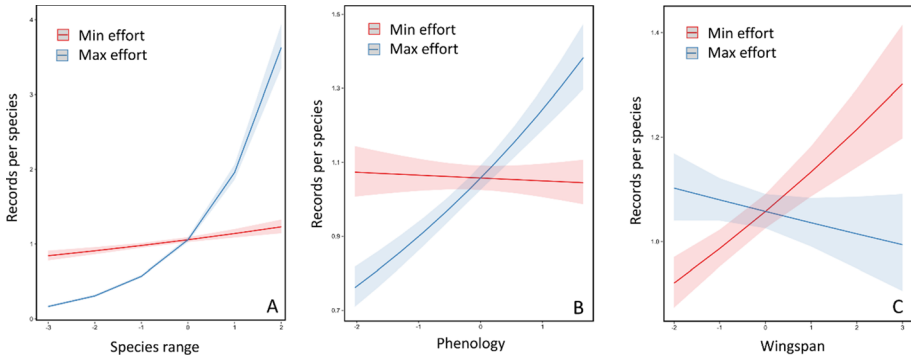
Fig. 3 Marginal effects of interaction terms in GLMMs between species features and user effort visualized as predicted trends for users showing minimum and maximum effort, species range (**A**), phenology (**B**) and Wingspan (**C**)

the evenness of recorded individuals among species (Shannon index, q=1; and Simpson index, q=2) (q=0: Rho=0.893, P<0.001; q=1: Rho=0.939, P<0.001; q=2: Rho=0.939, P<0.001; Fig. 4).
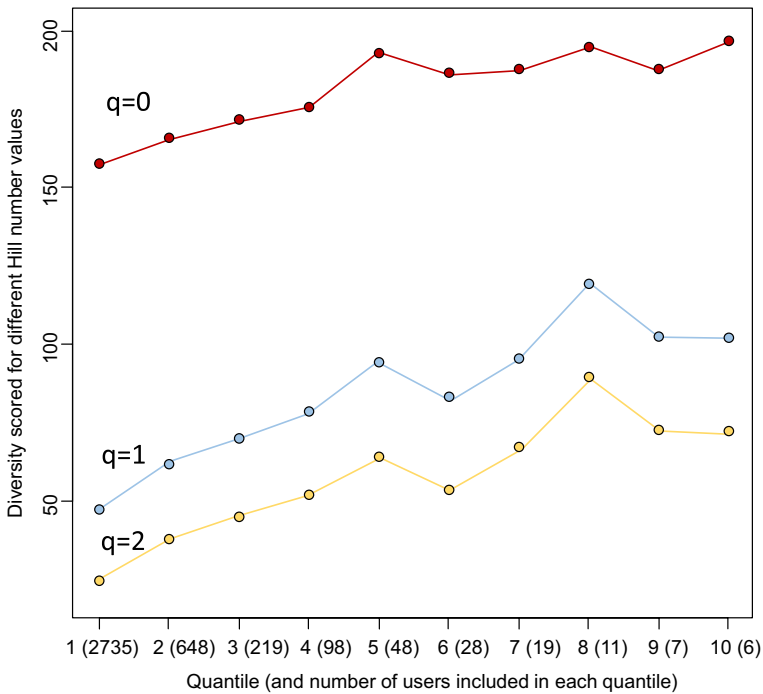
Fig. 4 The relationship between three indexes of diversity obtained after Hill's numbers (q=0: richness, q=1: Shannon index, q=2: Simpson index) for ten quantiles containing a similar number of observations by user showing an increasing effort. The number of users included in each quantile is indicated in parentheses

Discussion

We evaluated the impact of citizen science records on reducing the perception of local extinctions in butterfly communities in Italian National Parks, which have an extraordinary diversity within the European and Mediterranean regions. The records by citizen scientists confirmed that several potential local extinctions were actually due to lack of recent records. Additionally, we found that observers with varying levels of effort on iNaturalist had varying contribution to this process, primarily because they recorded different levels of butterfly diversity and responded differently to various aspects of species appearance. These findings provide crucial information for National Parks to develop effective strategies for promoting citizen science initiatives to monitor butterfly populations over time.

Aim 1—the potential extinction upon time series approach

The establishment of the targeted butterfly monitoring scheme (BMS) citizen science project (https://butterfly-monitoring.net/) has allowed for a precise tracking of the overall decline in butterfly populations over the past decades (e.g. Warren et al. 2021). The BMS has also helped detecting the effects of climate change on butterfly distribution and community composition, as well as the correlation between population trends and functional traits and phylogeny (Parmesan et al. 1999; Devictor et al. 2012; Bonelli et al. 2022; Halsch et al. 2021; Melero et al. 2022). However, long-term data for the BMS is only available for a few European countries and parts of North America (Warren et al. 2021).

Local butterfly extinctions have been documented globally, even in areas without monitoring schemes (Finland: van Bergen et al. 2020; Panama: Basset et al. 2015; California: Preston et al. 2012; Italy: Bonelli et al. 2011, 2022). In most cases, this evidence was based on exceptional datasets, mostly from the past decades, which allowed for the evaluation of a few butterfly communities. The PETS index can integrate knowledge from multiple sources such as literature, museum data, expert collections, standardized monitoring, and unstructured citizen science to evaluate the possibility of local extinctions, even in the absence of exceptional datasets. Our dataset shows that published records are scarce or outdated for most Italian National Parks, and the occurrence of butterflies is not confirmed for more than one-third of the time since their first sighting. Although there is no correlation between the fraction of records without a collection date in literature and the contribution of citizen science in reducing the perception of extinction, the large fraction of records without a precise date greatly hinders the possibility to obtain complete time series. In the light of the current biodiversity crisis, we recommend that researchers include precise data in their observations, especially considering the lack of well-established rules for writing faunistic papers. In this regard, citizen science records are less affected by the lack of collection data.

Due to a general lack of data, the likelihood of local extinctions in PETS0 appears to be quite high, with many species remaining unrecorded in National Parks for decades. This highlights the need for field investigations to confirm the presence of previously recorded species. Such efforts can be costly, but the contribution of citizen science can help reduce the costs. The use of the PETS algorithm revealed that iNaturalist data can play an important role in recording butterfly populations, reducing the lack of knowledge about persistence to an average of 11%. Another significant finding is the

considerable variability in PETS0 and PETS1, as well as the difference between them (delta), among different National Parks. This variability is largely dependent on the time since the first and last faunistic study of each park, and to some extent on the level of citizen science activity.

Aim 2 and 3—the effect of species traits and user effort on iNaturalist occurrences

Single observations uploaded on iNaturalist by users who put in more effort contribute more to evaluating the potential for local extensions in Italian National Parks. This is expected if more committed users tend to record a higher proportion of butterfly diversity, being more focused on taking pictures of different species and being differently affected by species appearance.

In general, our findings aligned with previous research on birds by Callaghan et al. (2021), which demonstrated that the availability of unstructured citizen science data depends on species appearance. For butterflies, we found that larger species with a longer flight period and broader geographical distribution are more likely to have a greater number of records available on the iNaturalist platform. A higher number of records for species showing a wider distribution and a longer flight period cannot be considered as a bias but as a desirable property, since these trends are the basis of high quality data obtained from structured monitoring schemes (e.g., in transect counts). However, this property is only shown by iNaturalist users uploading a high number of observations, as documented by the strong interactions between phenology and species range with user effort. In the case of less engaged users, these correlations are less strict than for highly engaged users. This result could be due to the fact that they do not use iNaturalist frequently and their observations constitute too small samples to be affected by phenology and distribution.

While a positive relationship linking upload frequency with phenology and distribution is a desired property of data, the tendency to document more often the occurrence of large species is a typical bias of citizen science (Kral-O'Brien et al. 2020; Isaac et al. 2011; Moranz 2010, Dennis et al. 2006). This expected behavior is generally confirmed in our analysis because larger species scored a higher number of records. However, the interaction between user engagement and butterfly size showed a significant effect. In fact, the decision to upload an observation does not only depend on the probability of encountering a given species, but also on other factors, such as the personal appreciation for that species (e.g. Callaghan et al. 2021; Isaac and Pocock 2015). Also in this case, highly committed users provide more accurate data, as they do not seem to be selectively attracted to bigger and more visible butterfly species.

The preference for capturing pictures of both large and small butterflies by highly engaged users is likely to contribute to the higher diversity observed in their records, both in terms of species richness and evenness. It is possible that these users may learn more about the taxonomy of the butterfly group they are interested in and photograph rarer or less conspicuous species. Additionally, these users may also search more widely to find species with limited distributions, and document butterflies during different seasons. Furthermore, it is possible that a highly committed user may actively search for species not encountered yet, which may further contribute to a more diverse sample of species captured in photographs.

Final remarks

Protected areas play a critical role in conserving biodiversity, promoting sustainability, and raising public awareness of the importance of natural capital and ecosystem services (Bastian 2013; Geldmann et al. 2013; Millennium Ecosystem Assessment 2013; Stolton et al. 2015; Chowdhury et al. 2023). Involving citizens in biodiversity monitoring through citizen science has been shown to be an effective and efficient way to gather data and information (Fontaine et al. 2021; Mannino and Balistreri 2018; Dennis et al. 2017; Zapponi et al. 2017). We documented that citizen science data can also be used to complement existing literature data to more accurately determine the possibility for local extinction and community erosion. This information can then be used by National Parks to prioritize their conservation efforts and save financial resources. National Parks should encourage citizens to participate in both structured and unstructured projects to gather standard and opportunistic data. This can be done through events such as bioblitzes, where people are educated about the importance of monitoring biodiversity and encouraged to upload their observations to platforms like iNaturalist.

It is important to be aware of the limitations of citizen science, including the unequal contributions and quality of data provided by differently engaged users. To address this, National Parks should also promote activities that educate and engage the general public, such as workshops focused on taxa identification with the help of expert taxonomists. In Italy, this has already begun with the hosting of the first Italian BMS workshop in the Sila National Park in 2019, which has since been replicated in five other National Parks. This increased knowledge is likely to result in higher quality data and less influence from aesthetic preferences (Callaghan et al. 2021; Barbato et al. 2021; Randler 2021). Additionally, individuals who are highly engaged in unstructured citizen science are more likely to participate in targeted projects, such as the globally successful Butterfly Monitoring Scheme (Warren et al. 2021). The Italian National Parks are also committed to carrying out pollinator monitoring, including butterfly counts, through Environment Ministry funding with the involvement of volunteers and experts. Improving taxonomy knowledge through citizen science can also help to address the shortage of taxonomists as outlined by the Red List of Taxonomists, a European Commission-funded initiative to increase awareness of the available expertise for preserving insect biodiversity (Hochkirch et al. 2022).

Data availability The datasets generated and analyzed during the current study are available in the GitHub repository, https://github.com/leondap/pets.

Declarations

Competing interests The authors declare no competing interests.

References

Aristeidou M, Herodotou C, Ballard HL et al (2021) Exploring the participation of young citizen scientists in scientific research: the case of iNaturalist. PLoS One 16:e0245682. https://doi.org/10.1371/journal. pone.0245682

Balletto E, Bonelli S, Cassulo L (2007) Insecta lepidoptera papilionoidea In: Ruffo S, Stoch F (eds) Check- list and distribution of the Italian Fauna. 10.000 terrestrial and inland water species 2nd and revised edition. Memorie del Museo Civico di Storia Naturale di Verona, pp. 257–261

Barbato D, Benocci A, Guasconi M, Manganelli G (2021) Light and shade of citizen science for less charis- matic invertebrate groups: quality assessment of iNaturalist nonmarine mollusc observations in central Italy. J Molluscan Stud 87:eyab033. https://doi.org/10.1093/mollus/eyab033

Basset Y, Barrios H, Segar S et al (2015) The butterflies of Barro Colorado Island, Panama: local extinction since the 1930s. PLoS One 10:e0136623. https://doi.org/10.1371/journal.pone.0136623

Bastian O (2013) The role of biodiversity in supporting ecosystem services in Natura 2000 sites. Ecol Ind 24:12–22

Bonelli S, Cerrato C, Loglisci N, Balletto E (2011) Population extinctions in the Italian diurnal lepidoptera: an analysis of possible causes. J Insect Conserv 15:879–890. https://doi.org/10.1007/s10841-011-9387-6

Bonelli S, Cerrato C, Barbero F et al (2022) Changes in Alpine Butterfly Communities during the last 40 years. Insects 13:43. https://doi.org/10.3390/insects13010043

Brooks ME, Kristensen K, van Benthem KJ et al (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J 9:378–400

Callaghan CT, Ozeroff I, Hitchcock C, Chandler M (2020) Capitalizing on opportunistic citizen science data to monitor urban biodiversity: a multi-taxa framework. Biol Conserv 251:108753

Callaghan CT, Poore AGB, Hofmann M et al (2021) Large-bodied birds are over-represented in unstructured citizen science data. Sci Rep 11:19073. https://doi.org/10.1038/s41598-021-98584-7

Cambria VE, Campagnaro T, Trentanovi G et al (2021) Citizen science data to measure human use of green areas and forests in European cities. Forests 12:779. https://doi.org/10.3390/f12060779

Campanaro A, Hardersen S, De Zan LR et al (2017) Analyses of occurrence data of protected insect species collected by citizens in Italy. Nat Conserv 20:265–297

Capotorti G, Zavattero L, Anzellotti I et al (2012) Do national parks play an active role in conserving the natural capital of Italy? Plant biosystems-an international journal dealing with all. Asp Plant Biol 146:258–265

Carpaneto GM, Campanaro A, Hardersen S et al (2017) The LIFE project monitoring of insects with public participation(MIPP): aims, methods and conclusions. Nat Conserv 20:1–35

Ceballos G, Ehrlich PR, Barnosky AD et al (2015) Accelerated modern human–induced species losses: entering the sixth mass extinction. Sci Adv 1:e1400253. https://doi.org/10.1126/sciadv.1400253

Chowdhury S, Fuller RA, Ahmed S et al (2023a) Using social media records to inform conservation planning. Conserv Biol. https://doi.org/10.1111/cobi.14161

Chowdhury S, Jennions MD, Zalucki MP et al (2023b) Protected areas and the future of insect conservation. Trends Ecol Evol 38:85–95. https://doi.org/10.1016/j.tree.2022.09.004

Courtney SP, Hill CJ, Westerman A (1982) Pollen carried for long periods by butterflies. Oikos 38:260–263. https://doi.org/10.2307/3544030

Crall AW, Newman GJ, Jarnevich CS et al (2010) Improving and integrating data on invasive species collected by citizen scientists. Biol Invasions 12:3419–3428. https://doi.org/10.1007/s10530-010-9740-9

Cruden RW, Hermann-Parker SM (1979) Butterfly pollination of *Caesalpinia pulcherrima*, with observations on a psychophilous Syndrome. J Ecol 67:155–168. https://doi.org/10.2307/2259342

Dapporto L, Cini A, Vodă R et al (2019) Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. Mol Ecol Resour 19:1623–1636. https://doi.org/10.1111/1755-0998.13059

Dennis RLH, Shreeve TG, Isaac NJB et al (2006) The effects of visual apparency on bias in butterfly recording and monitoring. Biol Conserv 128:486–492. https://doi.org/10.1016/j.biocon.2005.10.015

Dennis EB, Morgan BJT, Brereton TM et al (2017) Using citizen science butterfly counts to predict species population trends. Conserv Biol 31:1350–1361. https://doi.org/10.1111/cobi.12956

Devictor V, Van Swaay C, Brereton T et al (2012) Differences in the climatic debts of birds and butterflies at a continental scale. Nat Clim Change 2:121–124

Dirzo R, Young HS, Galetti M et al (2014) Defaunation in the anthropocene. Science 345:401–406. https://doi.org/10.1126/science.1251817

Echeverria A, Ariz I, Moreno J et al (2021) Learning plant biodiversity in nature: the use of the citizen–science platform iNaturalist as a collaborative tool in secondary education. Sustainability 13:735. https://doi.org/10.3390/su13020735

Fischer EE, Cobb NS, Kawahara AY et al (2021) Decline of amateur *Lepidoptera* collectors threatens the future of specimen-based research. Bioscience 71:396–404. https://doi.org/10.1093/biosci/biaa152

Fontaine C, Fontaine B, Prévot A-C (2021) Do amateurs and citizen science fill the gaps left by scientists? Curr Opin Insect Sci 46:83–87. https://doi.org/10.1016/j.cois.2021.03.001

Franzén M, Johannesson M (2007) Predicting extinction risk of butterflies and moths (Macrolepidoptera) from distribution patterns and species characteristics. J Insect Conserv 11:367–390. https://doi.org/10.1007/s10841-006-9053-6

Gallo T, Waitt D (2011) Creating a successful citizen science model to detect and report invasive species. Bioscience 61:459–465. https://doi.org/10.1525/bio.2011.61.6.8

Geldmann J, Barnes M, Coad L et al (2013) Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. Biol Conserv 161:230–238

Groom Q, Desmet P, Reyserhove L et al (2019) Improving darwin core for research and management of alien species. Biodivers Inform Sci Stand 3:e38084

Habel JC, Samways MJ, Schmitt T (2019) Mitigating the precipitous decline of terrestrial european insects: requirements for a new strategy. Biodivers Conserv 28:1343–1360. https://doi.org/10.1007/s10531-019-01741-8

Hallmann CA, Sorg M, Jongejans E et al (2017) More than 75% decline over 27 years in total flying insect biomass in protected areas. PLoS One 12:e0185809

Halsch CA, Shapiro AM, Fordyce JA et al (2021) Insects and recent climate change. Proc Natl Acad Sci 118:e2002543117. https://doi.org/10.1073/pnas.2002543117

Heigl F, Kieslinger B, Paul KT et al (2019) Opinion: toward an international definition of citizen science. Proc Natl Acad Sci 116:8089–8092. https://doi.org/10.1073/pnas.1903393116

Hochkirch A et al (2022) European red list of insect taxonomists. Publications Office of the European Union, Luxembourg. https://doi.org/10.2779/364246

Isaac NJ, Pocock MJ (2015) Bias and information in biological records. Biol J Linn Soc 115:522–531

Isaac NJ, Cruickshanks KL, Weddle AM et al (2011) Distance sampling and the challenge of monitoring butterfly populations. Methods Ecol Evol 2:585–594

Jennersten O (1984) Flower visitation and pollination efficiency of some north european butterflies. Oecologia 63:80–89. https://doi.org/10.1007/BF00379789

Krabbenhoft CA, Kashian DR (2020) Citizen science data are a reliable complement to quantitative ecological assessments in urban rivers. Ecol Ind 116:106476

Kral-O'Brien KC, Karasch BM, Hovick TJ et al (2020) Morphological traits determine detectability bias in north american grassland butterflies. Ecosphere 11:e03304. https://doi.org/10.1002/ecs2.3304

Labadessa R, Cagnetta G, Desaphy J-F et al (2021) Using occurrence data to evaluate extinction reveals a strong resilience of butterflies in a National Park of Southern Europe (Alta Murgia National Park). J Insect Biodivers 28:1–12

Li D (2018) hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill numbers. J Open Source Softw 3:1041. https://doi.org/10.21105/joss.01041

Lobo JM (2016) The use of occurrence data to predict the effects of climate change on insects. Curr Opin Insect Sci 17:62–68. https://doi.org/10.1016/j.cois.2016.07.003

Lüdecke D (2023) sjPlot: Data visualization for statistics in social science. R package version 2.8.15 https://CRAN.R-project.org/package=sjPlot

Lundmark C (2003) BioBlitz: getting into backyard biodiversity. BioScience 53:329. https://doi.org/10.1641/0006-3568(2003)053[0329:BGIBB]2.0.CO;2

Maistrello L, Dioli P, Bariselli M et al (2016) Citizen science and early detection of invasive species: phenology of first occurrences of *Halyomorpha halys* in Southern Europe. Biol Invasions 18:3109–3116. https://doi.org/10.1007/s10530-016-1217-z

Mannino AM, Balistreri P (2018) Citizen science: a successful tool for monitoring invasive alien species (IAS) in marine protected areas. the case study of the Egadi Islands MPA (Tyrrhenian Sea, Italy). Biodiversity 19:42–48

Marshall SA (2008) Field photography and the democratization of arthropod taxonomy. Am Entomol 54:4

McDermott Long O, Warren R, Price J et al (2017) Sensitivity of UK butterflies to local climatic extremes: which life stages are most at risk? J Anim Ecol 86:108–116. https://doi.org/10.1111/1365-2656.12594

Melero Y, Evans LC, Kuussaari M et al (2022) Local adaptation to climate anomalies relates to species phylogeny. Commun Biology a5:1–9. https://doi.org/10.1038/s42003-022-03088-3

Middleton-Welling J, Dapporto L, García-Barros E et al (2020) A new comprehensive trait database of European and Maghreb butterflies, *Papilionoidea*. Sci Data 7:351

Millennium Ecosystem Assessment (2013) Bridging scales and knowledge systems: concepts and applications in ecosystem assessment. Island Press, Washington

Montgomery GA, Dunn RR, Fox R et al (2020) Is the insect apocalypse upon us? How to find out. Biol Conserv 241:108327. https://doi.org/10.1016/j.biocon.2019.108327

Moranz RA (2010) The effects of ecological management on tallgrass prairie butterflies and their nectar sources. Oklahoma State University, Stillwater

Noriega JA, Hortal J, Azcárate FM et al (2018) Research trends in ecosystem services provided by insects. Basic Appl Ecol 26:8–23. https://doi.org/10.1016/j.baae.2017.09.006

Nugent J (2018) Inaturalist. Sci Scope 41:12–13

Oberhauser KS, Prysby MD (2008) Citizen Science: creating a research army for conservation. Am Entomol 54:103–105. https://doi.org/10.1093/ae/54.2.103

Parmesan C, Ryrholm N, Stefanescu C et al (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. Nature 399:579–583. https://doi.org/10.1038/21181

Preston KL, Redak RA, Allen MF, Rotenberry JT (2012) Changing distribution patterns of an endangered butterfly: linking local extinction patterns and variable habitat relationships. Biol Conserv 152:280–290. https://doi.org/10.1016/j.biocon.2012.03.011

Prudic KL, McFarland KP, Oliver JC et al (2017) eButterfly: leveraging massive online citizen science for butterfly conservation. Insects 8:53. https://doi.org/10.3390/insects8020053

Randler C (2021) Users of a citizen science platform for bird data collection differ from other birdwatchers in knowledge and degree of specialization. Global Ecol Conserv 27:e01580. https://doi.org/10.1016/j.gecco.2021.e01580

Riley JC (1986) Insects and the european mortality decline. Am Hist Rev 91:833–858. https://doi.org/10.2307/1873324

Rocha-Ortega M, Rodriguez P, Córdoba-Aguilar A (2021) Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. Ecol Entomol 46:718–728. https://doi.org/10.1111/een.13027

Rosenberg DM, Danks HV, Lehmkuhl DM (1986) Importance of insects in environmental impact assessment. Environ Manag 10:773–783

Roskov Y, Ower G, Orrell T et al (2019) Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist. Digital resource at http://www.catalogueoflife.org/annual-checklist/2019

Sánchez-Bayo F, Wyckhuys KAG (2019) Worldwide decline of the entomofauna: a review of its drivers. Biol Conserv 232:8–27. https://doi.org/10.1016/j.biocon.2019.01.020

Sanderson C, Braby MF, Bond S (2021) Butterflies Australia: a national citizen science database for monitoring changes in the distribution and abundance of australian butterflies. Austral Entomol 60:111–127. https://doi.org/10.1111/aen.12513

Schultz CB, Haddad NM, Henry EH, Crone EE (2019) Movement and demography of at-risk butterflies: building blocks for conservation. Ann Rev Entomol 64:167–184. https://doi.org/10.1146/annurev-ento-011118-112204

Stolton S, Dudley N, Avcıoğlu Çokçalışkan B et al (2015) Values and benefits of protected areas. Worboys GL (Ed), In: Protected area governance and management. p. 145–168 ANU Press, Canberra.

Stoudt S, Goldstein BR, de Valpine P (2022) Identifying engaging bird species and traits with community science observations. Proc Nat Acad Sci 119:e2110156119. https://doi.org/10.1073/pnas.2110156119

Sullivan BL, Wood CL, Iliff MJ et al (2009) eBird: a citizen-based bird observation network in the biological sciences. Biol Conserv 142:2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

Turvey ST, Crees JJ (2019) Extinction in the anthropocene. Curr Biol 29:R982–R986

van Bergen E, Dallas T, DiLeo MF et al (2020) The effect of summer drought on the predictability of local extinctions in a butterfly metapopulation. Conserv Biol 34:1503–1511. https://doi.org/10.1111/cobi.13515

Van Eupen C, Maes D, Herremans M et al (2021) The impact of data quality filtering of opportunistic citizen science data on species distribution model performance. Ecol Model 444:109453. https://doi.org/10.1016/j.ecolmodel.2021.109453

Van Eupen C, Maes D, Herremans M et al (2022) Species profiles support recommendations for quality filtering of opportunistic citizen science data. Ecol Model 467:109910. https://doi.org/10.1016/j.ecolmodel.2022.109910

van Swaay CA, Nowicki P, Settele J, Van Strien AJ (2008) Butterfly monitoring in Europe: methods, applications and perspectives. Biodivers Conserv 17:3455–3469

Warren MS, Maes D, van Swaay CAM et al (2021) The decline of butterflies in Europe: problems, significance, and possible solutions. Proc Natl Acad Sci 118:e2002551117. https://doi.org/10.1073/pnas.2002551117

Wei JW, Lee BPY-H, Wen LB (2016) Citizen Science and the Urban ecology of birds and butterflies—a systematic review. PLoS One 11:e0156425. https://doi.org/10.1371/journal.pone.0156425

Wieczorek J, Döring M, De Giovanni R et al (2009) Darwin Core. Accessed date February 1, 2023. Available: http://www.tdwg.org/standards/450/

Wieczorek J, Bloom D, Guralnick R et al (2012) Darwin core: an evolving community-developed biodiversity data standard. PLoS One 7:e29715. https://doi.org/10.1371/journal.pone.0029715

Yang LH, Gratton C (2014) Insects as drivers of ecosystem processes. Curr Opin Insect Sci 2:26–32. https://doi.org/10.1016/j.cois.2014.06.004

Zapponi L, Cini A, Bardiani M et al (2017) Citizen science data as an efficient tool for mapping protected saproxylic beetles. Biol Conserv 208:139–145. https://doi.org/10.1016/j.biocon.2016.04.035

Authors and Affiliations

Elia van Tongeren[1,2] · Ginevra Sistri[2] · Marco Bonifacino[2] · Mattia Menchetti[3] · Lorenzo Pasquali[2] · Vania Salvati[2] · Emilio Balletto[4] · Simona Bonelli[4] · Alessandro Cini[5] · Mariagrazia Portera[6] · Leonardo Dapporto[1,2]

✉ Elia van Tongeren
   elia.vantongeren@unifi.it

[1] NBFC, National Biodiversity Future Center, Palermo, Italy

[2] ZEN Lab, Department of Biology, University of Florence, Florence, Italy

[3] Institute of Evolutionary Biology, CSIC-UPF, Barcelona, Spain

[4] Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

[5] Department of Biology, University of Pisa, Pisa, Italy

[6] Department of Humanities and Philosophy, University of Florence, Florence, Italy