



Modelling spatial deficits in citizen science recording: a case study from UK butterflies

Mingrui Li¹ · Richard H. Ffrench-Constant¹ · Richard Fox^{1,2} · David B. Roy^{1,3} · Robin J. Boyd^{1,3} · Jonathan Bennie⁴

Received: 27 February 2026 / Accepted: 11 September 2026
© The Author(s) 2026

Abstract Citizen science data are increasingly used to infer species distributions and monitor biodiversity at large spatial scales. However, recording effort is often uneven across space, potentially biasing inferences about species' distributions. We analysed 51,045 butterfly records submitted via the iRecord Butterflies app in Devon and Cornwall, in the southwest of the United Kingdom. We used negative binomial models to examine how land cover and three public access variables influenced the number of butterfly records in two groups of species: wider countryside species (WS) and habitat specialists (HS). Results show that the area of public access land was a consistently important explanatory variable of recording effort across both WS and HS. Footpath density showed a significantly stronger effect on number of records for WS than for HS. Moreover, predicted recording deficits, defined as the difference between model predictions of record density under current and maximum access scenarios, were widespread. Deficits tended to occur in areas predicted to provide suitable butterfly habitat but with relatively low public access, whereas highly accessible areas were more likely to be recorded. Recording deficits for HS are predicted in areas with a high density of semi-natural habitats but low public access; by contrast for WS recording deficits are predicted in built-up areas and gardens, where these common species may be widespread but recorder effort is highly localised. Implications for insect conservation. Improving the spatial representativeness of citizen science data may therefore require complementary strategies: encouraging recording close to residences for WS and securing access to restricted semi-natural habitats for HS. By explicitly quantifying effects of public access, our framework moves beyond intuitive explanations and offers empirical, operational guidance for identifying and addressing spatial data gaps, thereby helping to improve the spatial representativeness and reliability of citizen science data.

Implications for insect conservation Improving the spatial representativeness of citizen science data may therefore require complementary strategies: encouraging recording close to residences for WS and securing access to restricted semi-natural habitats for HS. By explicitly quantifying effects of public access, our framework moves beyond intuitive explanations and offers empirical, operational guidance for identifying and addressing spatial data gaps, thereby helping to improve the spatial representativeness and reliability of citizen science data.

Keywords Public access · Citizen science data · Butterfly records · Sampling bias · Public footpath length

Richard H. Ffrench-Constant passed away.

✉ Mingrui Li
ml902@exeter.ac.uk

¹ Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, Exeter, Cornwall TR10 9FE, UK

² Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset BH20 5QP, UK

³ UK Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK

⁴ Centre for Geography and Environmental Science, University of Exeter, Penryn Campus, Penryn, Exeter, Cornwall TR10 9FE, UK

Introduction

Unstructured citizen (community) science provides large-scale data for investigating insect spatial distributions and responses to global drivers (Powney and Isaac, 2015; Soroye et al., 2020; Díaz-Calafat et al., 2024) and represents an important evidence base for informing insect conservation policies (Dyer et al., 2017; León-Cortés et al., 2024; Roy et al., 2024; Sánchez-Fernández et al., 2026). However, compared with structured sampling data, unstructured citizen science data do not solely represent true species abundance, but instead reflect the joint influence of species occurrence and recorder behaviour (Mondain-Monval et

al., 2024). Thus, the spatial distribution of citizen science records may partly reflect where recorders are able to go based on public access to land. At the same time, limited public access at recording sites may directly restrict or hinder recording activity (Jingu and Ogawa, 2025). However, the extent to which public access influences the number of species records, and the mechanisms through which this occurs, remain poorly quantified.

The extent to which land is legally or practically accessible to the public varies substantially between countries. In the United Kingdom (UK), legal rights to access differ between the constituent nations; in Scotland the public has a statutory right of access to most land, while in England and Wales there are explicit legal frameworks for public rights of access, such as the Countryside and Rights of Way Act 2000, which grants the public the right to access certain types of rural land and formally designated public rights of way (Rischel, 2024; Natural England, 2025). In contrast, land tenure in the United States is rooted in a strong tradition of private property rights, and generally lacks a legal framework equivalent to a “freedom to roam” (Robertson, 2010). As a result, public land access exhibits spatial variation, which may in turn shape the spatial distribution of citizen science records.

Public access can be quantified by metrics such as the density of public footpath networks and the extent of open access areas, as well as by proxy measures such as the local human population density (Choi et al., 2020). Accessibility is expected to be higher in areas with dense public footpath networks, short distances to paths or roads, and high surrounding human population densities, which increase the availability of potential recorders (Mair and Ruete, 2016; Guetté et al., 2022; Mandeville et al., 2022). Collectively, these factors capture landscape accessibility and are expected to shape observed patterns of species occurrence in citizen science data.

Beyond public access, citizen science observations inherently reflect the habitats occupied by species, as records can only occur where the species is present and are therefore constrained by the species’ distribution. For both widespread and range-restricted butterfly species, the influence of land cover on their occurrence and abundance has been widely documented (Krämer et al., 2012; Börschig et al., 2013; Dapporto and Dennis, 2013; Habel et al., 2023). Therefore, when assessing the effects of public access in citizen science data, land cover must be incorporated into models to account for underlying ecological constraints (Han et al., 2021; Cooper et al., 2024; Kasiske et al., 2025; Wang et al., 2025).

Whilst it is already recognised that limited public access can hinder recording by citizen scientists, the magnitude of this effect has rarely been quantified. Here, we model the

effects of public access on butterfly records in the southwest, explicitly quantifying the influence of public footpaths and open access land. We hypothesise that restrictions on public access disproportionately limit the number of records for both widespread and non-widespread species. By disentangling accessibility-driven recording bias from ecological patterns, this approach helps identify species and areas that are under recorded and provides a predictive framework for guiding future recording effort.

Methods

Study area

The counties of Devon and Cornwall in southwest England form a peninsula covering approximately 10,200 km². The region is predominantly rural, with land use dominated by agriculture. Major urban centres include Plymouth, Torbay and Exeter. Elevation ranges from sea level to approximately 600 m, with granite upland areas characterised by rough grazing, heathland, moorland and blanket bog, while lowland areas typically comprise a mosaic of arable farming, livestock pasture, semi natural woodland and commercial forestry (Gatis et al., 2025). Important butterfly habitats include coastal dunes and grasslands, semi natural grasslands, woodland and heathland. Hedgerows and suburban gardens also provide key habitats for several species (Spalding, 2022).

Data

All data processing, modelling, and prediction were conducted at the level of 1 km² grid square defined by the British National Grid, which served as the fundamental spatial unit of the study. Within each grid square, butterfly occurrence records, public access variables, and land cover characteristics were summarised. Butterfly record counts were then modelled using species-specific negative binomial generalized linear models with public access and land cover variables as predictors. Model outputs were used to predict record numbers for each grid square under original geographic conditions and under a high accessibility scenario defined as the 95th percentile of public access. All statistical analyses were conducted in Python (version 3.12.4; Python Software Foundation). Finally, predicted outputs were visualised in QGIS (version 3.34) (Qgis.org, 2025) to identify areas with low sampling effort and potential recording gaps.

Butterfly occurrence records were sourced from the iRecord Butterflies mobile application, which collects observations submitted by citizen scientists. The scheme is run by the UK Centre for Ecology & Hydrology in partnership

Table 1 Square-level public access variables

Variable name	Description
Footpath length	Total length (km) of public rights of way (footpaths, bridleways and byways) within each grid square
Public access area	Total public access area (km ²) within each grid square
Population density	Human population density (people per km ²) within each grid square

with Butterfly Conservation. We used iRecord Butterflies data to focus on spatial variation within a single opportunistic, unstructured recording system, rather than combining datasets collected under different recording protocols. Our analysis focused on Devon and Cornwall in southwest UK, corresponding to Watsonian Vice Counties 1, 2, 3 and 4 (Dandy, 1971). The dataset comprised 51,045 butterfly occurrence records collected between 14 July 2011 and 1 April 2024. All available records were retained for analysis because the vast majority had been verified and accepted by local taxonomic experts, while only a very small proportion (<0.05%) remained unconfirmed but were classified as plausible.

Public access was quantified using three indicators (see definition in Table 1 and spatial distribution in Fig. 1): Footpath length, public access area and human population density. In terms of land cover, ten aggregated classes of land cover in the UKCEH Land Cover Map for 2023 (Morton et al., 2024), hedgerow length per kilometre square (Hedge length) and area of sand dunes per kilometre square (Sand

dunes) were used. Each butterfly species recorded in the Devon and Cornwall dataset was assigned to the land cover types in which it is typically found, based on expert ecological knowledge provided by one of the authors, who has extensive experience in regional butterfly recording (Table 2). The analysis was restricted to wider countryside species (WS) and habitat specialists (HS) defined by ASHER et al. (2001). A summary of data availability for each species, including the total number of records and the number of occupied 1 km² British National Grid monads, is provided in Table S2. All other relevant dataset information is available in the supplementary material.

Statistical modelling and prediction

For each species with a sufficient sample size (more than 20 records), we modelled number of records per 1 km² grid square as a function of land cover and public access using negative binomial generalized linear models. All explanatory variables were Z score standardised prior to modelling. Pairwise Pearson correlation coefficients were calculated among all predictor variables before model fitting to assess potential multicollinearity, with correlations $|r| > 0.5$ considered indicative of potentially strong correlation. For each species, two models were fitted: (1) a full model, including public access variables and species-specific land cover variables; (2) a no public access model, including land cover variables only. Model comparison was primarily based on differences in AIC and McFadden's pseudo R^2 .

Fig. 1 Spatial patterns of access-related variables. The panels show **A** public access area, **B** population density, **C** footpath length, and **D** built-up areas and gardens, all expressed per km². Although built-up areas and gardens is classified as a land cover category, it is shown here because of its conceptual relevance to public access. Warmer colours represent higher values of each variable, and the scale bar applies to all panels

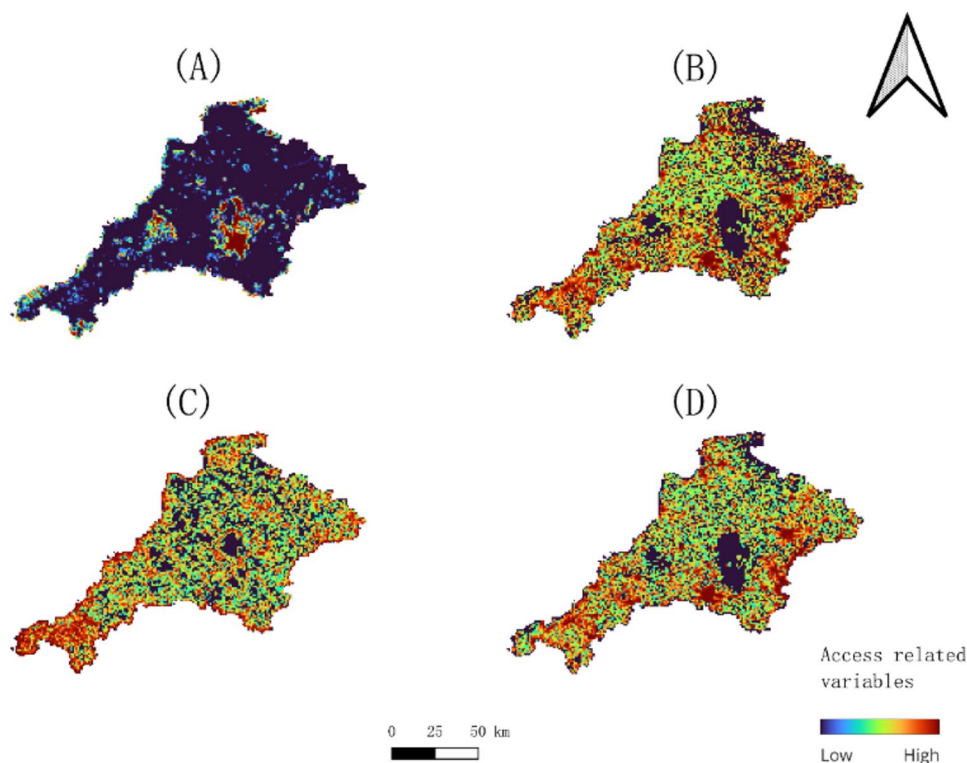


Table 2 Information on butterfly species in Devon and Cornwall. Land cover types listed represent the main habitats associated with each species in Devon and Cornwall

Common name	Scientific name	Species type	Land cover types
Brimstone	<i>Gonepteryx rhamni</i>	WS	Broadleaf woodland Freshwater Hedge
Brown Argus	<i>Aricia agestis</i>	WS	Semi-natural grassland Sand dunes
Comma	<i>Polygonia c-album</i>	WS	Broadleaf woodland Improved grassland Semi-natural grassland Built-up areas and gardens Hedge
Common Blue	<i>Polyommatus icarus</i>	WS	Semi-natural grassland
Dark Green Fritillary	<i>Speyeria aglaja</i>	HS	Semi-natural grassland Sand dunes
Dingy Skipper	<i>Erynnis tages</i>	HS	Semi-natural grassland Mountain heath and bog Sand dunes
Gatekeeper	<i>Pyronia tithonus</i>	WS	Arable Improved grassland Semi-natural grassland Built-up areas and gardens Hedge
Grayling	<i>Hipparchia semele</i>	HS	Mountain heath and bog Coastal
Green Hairstreak	<i>Callophrys rubi</i>	HS	Semi-natural grassland Mountain heath and bog
Green-veined White	<i>Pieris napi</i>	WS	Broadleaf woodland Semi-natural grassland Mountain heath and bog Freshwater Built-up areas and gardens, Hedge
Grizzled Skipper	<i>Pyrgus malvae</i>	HS	Semi-natural grassland Sand dunes
Heath Fritillary	<i>Melitaea athalia</i>	HS	Broadleaf woodland Semi-natural grassland Mountain heath and bog
High Brown Fritillary	<i>Fabriciana adippe</i>	HS	Semi-natural grassland
Holly Blue	<i>Celastrina argiolus</i>	WS	Broadleaf woodland Improved grassland Semi-natural grassland Built-up areas and gardens Hedge
Large Skipper	<i>Ochlodes sylvanus</i>	WS	Broadleaf woodland Semi-natural grassland Sand dunes Hedge
Large White	<i>Pieris brassicae</i>	WS	Arable Semi-natural grassland Built-up areas and gardens Hedge
Marbled White	<i>Melanargia galathea</i>	WS	Semi-natural grassland
Marsh Fritillary	<i>Euphydryas aurinia</i>	HS	Semi-natural grassland Mountain heath and bog
Meadow Brown	<i>Maniola jurtina</i>	WS	Arable Improved grassland Semi-natural grassland Built-up areas and gardens Hedge
Orange-tip	<i>Anthocharis cardamines</i>	WS	Semi-natural grassland Freshwater Built-up areas and gardens, Hedge
Peacock	<i>Aglais io</i>	WS	Improved grassland Semi-natural grassland Built-up areas and gardens Hedge
Pearl-bordered Fritillary	<i>Boloria euphrosyne</i>	HS	Broadleaf woodland Semi-natural grassland
Purple Hairstreak	<i>Favonius quercus</i>	WS	Broadleaf woodland Improved grassland Semi-natural grassland Hedge
Ringlet	<i>Aphantopus hyperantus</i>	WS	Broadleaf woodland Semi-natural grassland Hedge
Silver-studded Blue	<i>Plebejus argus</i>	HS	Semi-natural grassland Mountain heath and bog Sand dunes
Silver-washed Fritillary	<i>Argynnis paphia</i>	HS	Broadleaf woodland Hedge
Small Copper	<i>Lycaena phlaeas</i>	WS	Improved grassland Semi-natural grassland
Small Heath	<i>Coenonympha pamphilus</i>	WS	Semi-natural grassland Mountain heath and bog Sand dunes
Small Pearl-bordered Fritillary	<i>Boloria selene</i>	HS	Semi-natural grassland Coastal Sand dunes
Small Skipper	<i>Thymelicus sylvestris</i>	WS	Semi-natural grassland Sand dunes
Small Tortoiseshell	<i>Aglais urticae</i>	WS	Improved grassland Semi-natural grassland Built-up areas and gardens Hedge
Small White	<i>Pieris rapae</i>	WS	Arable Semi-natural grassland Built-up areas and gardens
Speckled Wood	<i>Pararge aegeria</i>	WS	Broadleaf woodland Semi-natural grassland Built-up areas and gardens Hedge
Wall	<i>Lasiommata megera</i>	WS	Semi-natural grassland Mountain heath and bog Coastal Hedge
White Admiral	<i>Limenitis camilla</i>	HS	Broadleaf woodland
Wood White	<i>Leptidea sinapis</i>	HS	Broadleaf woodland

After model fitting, we extracted the standardised regression coefficients of the public access variables together with their 95% confidence intervals. Coefficients were further transformed to the exponential scale to facilitate interpretation of effect sizes. Statistical significance was assessed using p values. Residual spatial autocorrelation was also assessed using Moran's I calculated from the Pearson residuals of each species-specific negative binomial model.

Spatial neighbours were defined as grid-cell centroids within 1500 m of each focal British National Grid monad centroid. This threshold captures directly adjacent and diagonal neighbouring monads in the regular 1 km² grid while allowing coastal and boundary cells to have fewer neighbours. Moran's I was calculated for both the full model and the model excluding public access variables.

We then predicted the number of species occurrence records expected under contrasting public access scenarios for each square (Fig. 2). Under the observed scenario, predicted numbers of records were generated using the observed values of public access and land cover variables within each 1 km² grid square, standardised on the same scale as used during model fitting (Fig. 2, panel A and B). We also evaluated scenarios under the assumption that the relationships among variables remain unchanged when public access is increased to high levels, represented by the

95th percentile values across the study area as proxies for maximum access (Fig. 2, panel C). Land cover variables were retained at their observed values within each 1 km² grid square to simulate potential recording intensity under conditions of high accessibility. For each species, the difference between predicted values under the observed and high public access scenarios was then calculated to quantify the potential effect of changes in public access on number of records (Fig. 2, panel D).

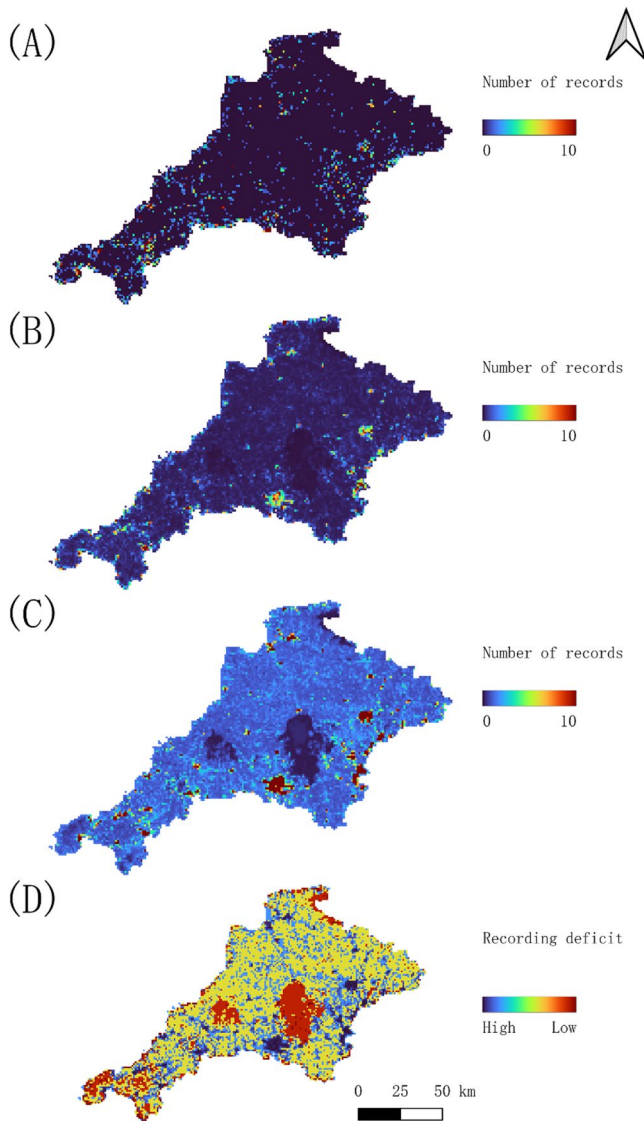


Fig. 2 Illustration of the four map types produced in this study, shown using Speckled Wood as an example at 1 km² grid square resolution across Devon and Cornwall. **A** Observed number of records; **B** Predicted number of records under observed public access; **C** Predicted number of records under the 95th percentile public access scenario; **D** Recording deficit, calculated as predicted records under observed access minus predicted records under the 95th percentile access scenario. Specifically, warmer colours indicate a lower recording deficit in panel D

Results

Among the 36 butterfly species analysed (Table 2), 34 were successfully fitted with models (Table 3). Negative binomial models could not be fitted for Heath Fritillary and Silver-studded Blue because the optimisation returned non-finite log likelihood values, indicating numerical instability and failure to converge, likely arising from extremely sparse records and highly restricted spatial occurrence. Delta AIC values were positive for only one species and negative for all others, indicating that models including public access variables generally achieved lower AIC which indicated reduced information loss and better relative predictive performance. McFadden's R^2 values of models including both access and land cover variables were greater than 0.05 for all fitted species, suggesting acceptable model performance. Footpath length was significantly associated with the number of records for 32 species, of which three showed negative relationships. Human population density was significantly associated with the number of records for 24 species, of which nine showed negative relationships. Public access area was significantly associated with the number of records for 28 species, of which two showed negative relationships. Detailed results are provided in Table 3 and Figure S1. Pairwise correlations among the three public access variables were all below $|r| = 0.5$, indicating no strong correlation among these predictors (Table S3). Residual spatial autocorrelation was detected in the species-specific negative binomial models. Moran's I calculated from Pearson residuals showed significant positive spatial autocorrelation for all fitted models, including both the full models and the models excluding public access variables (Table S1). This indicates that neighbouring 1 km² grid cells tended to have more similar residuals than expected under spatial independence.

WS exhibited more consistently positive responses to footpath length than HS (Fig. 3), where WS had a higher mean coefficient (0.29) than HS (0.02), and this difference was statistically significant based on Welch's t-test ($t=2.19$, $df=11.80$, $p=0.025$). For population density, mean coefficients were slightly positive for WS (0.03) but more negative for HS (mean = -1.50), although difference

Table 3 Model fit and public access effects from negative binomial models for each butterfly species

Common name	McFadden's R^2	Δ AIC	Footpath length coefficient (95% CI)	Footpath length significance	Population density coefficient (95% CI)	Population density significance	Public access area coefficient (95% CI)	Public access area significance
Brimstone	0.11	-59.88	0.06±0.07		0.21±0.05	***	0.19±0.07	***
Brown Argus	0.47	-97.69	0.18±0.12	**	0.32±0.08	***	0.78±0.18	***
Comma	0.13	-151.91	0.32±0.04	***	-0.03±0.06		-0.08±0.09	
Common Blue	0.14	-1183.33	0.42±0.04	***	0.32±0.03	***	0.64±0.06	***
Dark Green Fritillary	0.16	-95.46	0.27±0.07	***	0.28±0.08	***	0.61±0.16	***
Dingy Skipper	0.12	1.32	0.17±0.13	*	0.12±0.17		-0.07±0.33	
Gatekeeper	0.14	-347.29	0.23±0.03	***	-0.15±0.05	***	0.32±0.06	***
Grayling	0.23	-141.19	0.45±0.08	***	0.27±0.09	***	0.54±0.12	***
Green Hairstreak	0.18	-275.52	0.27±0.06	***	0.04±0.18		0.88±0.11	***
Green-veined White	0.08	-163.13	0.21±0.04	***	-0.27±0.09	***	0.37±0.10	***
Grizzled Skipper	0.43	-17.60	0.27±0.10	***	-1.56±2.79		-1.75±1.75	*
High Brown Fritillary	0.30	-104.49	-1.07±0.61	***	0.23±0.16	**	1.24±0.22	***
Holly Blue	0.22	-160.43	0.31±0.04	***	-0.03±0.05		0.02±0.08	
Large Skipper	0.09	-190.20	0.27±0.05	***	0.24±0.05	***	0.47±0.10	***
Large White	0.22	-278.07	0.25±0.03	***	-0.14±0.05	***	0.24±0.07	***
Marbled White	0.08	-153.64	0.14±0.07	***	0.25±0.05	***	0.57±0.11	***
Marsh Fritillary	0.11	-143.31	0.15±0.08	***	-14.08±5.01	***	0.36±0.13	***
Meadow Brown	0.12	-419.42	0.23±0.03	***	-0.04±0.05		0.36±0.05	***
Orange tip	0.11	-305.84	0.38±0.04	***	-0.25±0.07	***	0.02±0.09	
Peacock	0.12	-341.29	0.34±0.03	***	-0.13±0.06	***	0.16±0.07	***
Pearl-bordered Fritillary	0.28	-97.05	-0.24±0.21	*	-1.16±1.84		0.90±0.17	***
Purple Hairstreak	0.10	-17.07	0.25±0.08	***	0.21±0.10	***	-0.01±0.27	
Ringlet	0.09	-206.46	0.29±0.04	***	0.27±0.04	***	0.17±0.09	***
Silver-washed Fritillary	0.11	-22.29	0.11±0.06	***	0.10±0.07	**	0.15±0.08	***
Small Copper	0.15	-669.87	0.45±0.04	***	0.18±0.04	***	0.56±0.07	***
Small Heath	0.18	-313.96	0.29±0.05	***	0.17±0.07	***	0.66±0.09	***
Small Pearl-bordered Fritillary	0.27	-428.61	0.39±0.06	***	-0.28±0.39		1.18±0.12	***
Small Skipper	0.08	-198.21	0.25±0.05	***	0.25±0.04	***	0.44±0.10	***
Small Tortoiseshell	0.16	-487.25	0.46±0.03	***	-0.18±0.06	***	-0.11±0.09	*
Small White	0.19	-491.97	0.23±0.03	***	-0.14±0.05	***	0.45±0.06	***
Speckled Wood	0.17	-820.52	0.44±0.03	***	-0.06±0.05	*	0.24±0.06	***
Wall	0.16	-465.80	0.52±0.04	***	0.12±0.07	***	0.40±0.09	***
White Admiral	0.29	-18.42	-0.06±0.25		0.18±0.23		0.52±0.18	***
Wood White	0.07	-6.92	-0.52±0.47	*	-2.16±3.27		-0.03±0.30	

McFadden's R^2 refers to the fit of the full model including all public access variables. Δ AIC represents the difference between the full model and a reduced model excluding all public access variables and negative values indicate stronger support for the full model. Coefficients are untransformed log-scale estimates from the full negative binomial model; positive values indicate higher predicted records with increasing predictor values, and negative values indicate the opposite

Significance levels are shown as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

between WS and HS was not significant ($t = 1.32$, $df = 11.03$, $p = 0.105$) (Fig. 3). For public access area, mean coefficients were positive for both WS (0.31) and HS (0.38), with HS showing greater variability and more negative outliers, but no significant group difference ($t = -0.27$, $df = 12.21$, $p = 0.604$) (Fig. 3).

For WS, there is a predicted recorder deficit in areas dominated by built-up areas and gardens (Fig. 4, panel A; Fig. 1, panel D) but with few public footpaths or access land (Fig. 1, panel A and C), forming a clear suburban deficit

that represents the predominant spatial pattern for these species. For HS, no comparable suburban recorder deficit is predicted (Fig. 4, panel B; Fig. 1, panel D). Instead, their recorder deficit is concentrated in inland countryside areas characterised by high proportions of semi-natural habitats (e.g., broadleaf woodland, lowland heath, and sand dunes) but limited public access area (Fig. 4, panel B; Fig. 1, panel A). In contrast, a generally low recorder deficit is predicted across most moorland areas, including Dartmoor, Bodmin Moor, West Penwith, and the Lizard (Fig. 4, panel B).

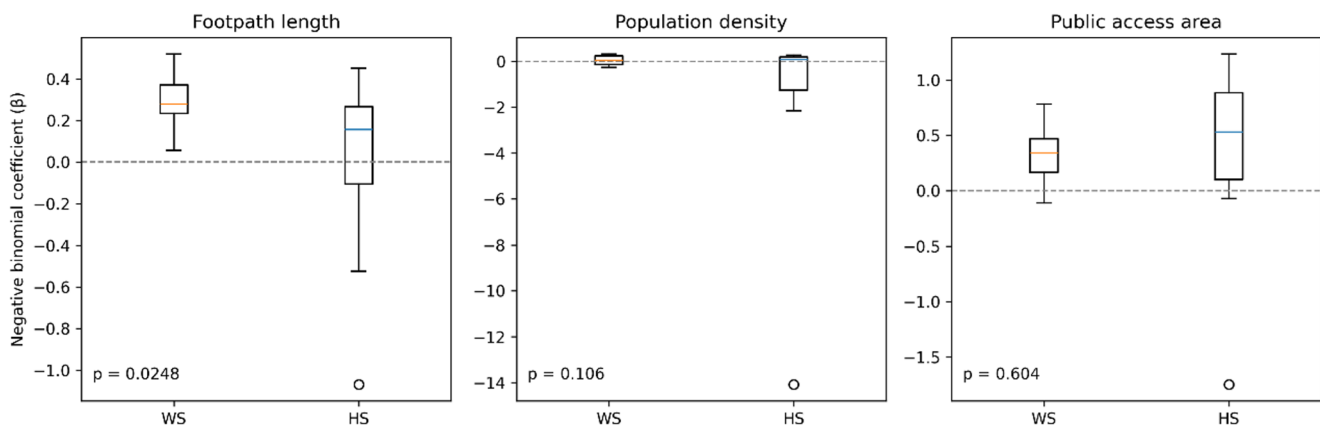


Fig. 3 Species-level coefficients of public access variables for WS and HS. Each boxplot is based on the distribution of species-specific coefficients within each classification (WS or HS) for **(a)** Footpath length, **(b)** Population density, and **(c)** Public access area; the orange central line

denotes the median, boxes represent the interquartile range, and whiskers show the data range. Reported p-values test differences in mean coefficients between WS and HS. The dashed horizontal line indicates zero effect

These patterns should be interpreted as relative spatial patterns within each species or species group, rather than as direct comparisons of absolute deficit values among species. Because a small number of grid cells had extreme predicted values, species-specific relative scales were used to highlight where under-recording was most pronounced within each species; the underlying grid-level predicted deficit values are provided as supplementary TIFF files.

Discussion

Based on negative binomial models predicting the spatial distribution of number of records of butterfly species in the southwest of the UK, we show how public access to land systematically shapes recording patterns across space. Species-level coefficients revealed that wider countryside species (WS) had a significantly stronger positive association with footpath length than habitat specialists (HS), while no significant group differences were detected for population density or public access area. Footpath length and public access area had significantly positive effects on the recording of most species in both groups. We also found that a predicted recording deficit, calculated as predicted records under observed access minus predicted records under the 95th percentile access scenario, was widespread and primarily driven by two factors: (1) whether a site was predicted to be suitable butterfly habitat based on land cover, and (2) whether recorders have physical access to that site. Consequently, a recording deficit tended to emerge in sites that were likely to provide good butterfly habitat for a species but had relatively low public access. Overall, the map indicates that sites with higher public access are more likely to be observed and recorded by citizen scientists, corresponding to the red areas in the recording deficit maps (Fig. 4).

For WS, predicted recording deficits were evident in areas dominated by built-up areas and gardens but with limited footpath length or public access area (Fig. 4, panel A). This is likely because, in urban and suburban areas where recorders are active, disproportionate recording of widespread butterfly species are derived from private gardens (Dennis et al., 2017; Pendl et al., 2021; Prudic et al., 2022; Plummer et al., 2023). However, although recorder density is typically higher in major urban areas, access to private gardens remains restricted (i.e. no public access is available). Recording is therefore largely confined to publicly accessible spaces, and/or the private gardens of recorders, resulting in highly localised and comparatively scarce records within residential landscapes (Jain et al., 2021; Pendl et al., 2021; Plummer et al., 2023). Moreover, in inland rural regions, the predicted recording deficit showed a fragmented pattern across the landscape (Fig. 4, panel A). This pattern reflects the uneven distribution of public access in rural landscapes (Fig. 1). For example, arable farmland is largely inaccessible to the public (Frigerio et al., 2020) and such limited access constrains recorder movement and likely contributes to reduced recording of WS in these areas (Videvall et al., 2016).

In contrast to WS, overall, for HS, predicted recording deficits are lower in coastal areas and in many upland moorland areas. Although these areas may host habitat specialist species, large portions of the landscape remain publicly accessible, especially along the coast with high footpath length (Fig. 1, panel C), as these regions tend to attract many visitors and recorders (Arkema et al., 2021). In built-up areas and gardens, however, low recording deficit occurs because many HS do not occur in people's gardens in the first place (Pla-Narbona et al., 2021; Donkersley et al., 2022). Instead, their recording deficits are mainly concentrated in inland rural areas, which often contain highly

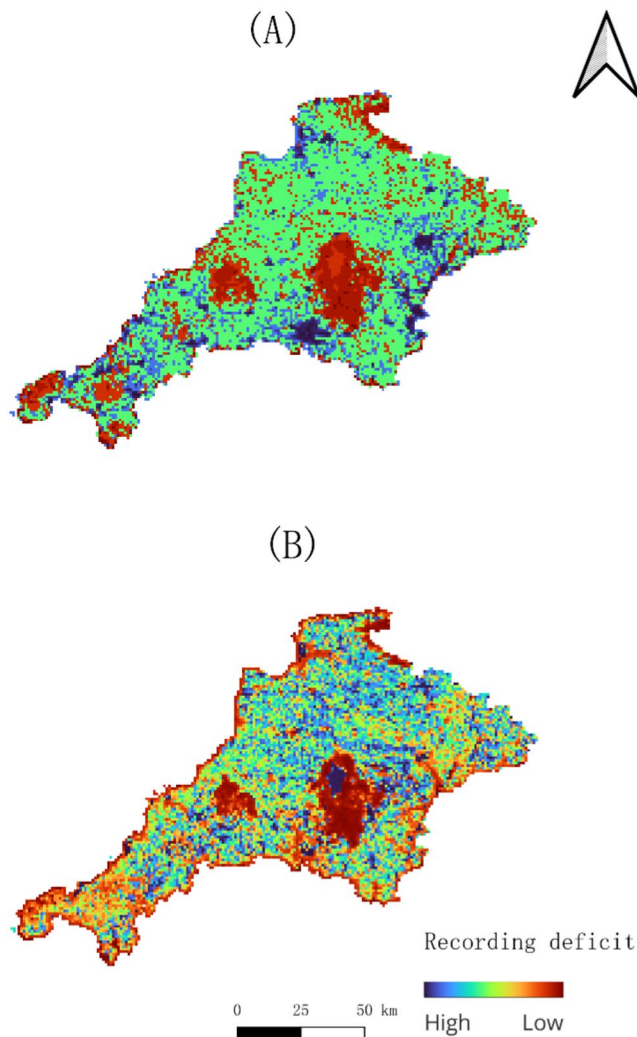


Fig. 4 Mean recording deficit across Devon and Cornwall at 1 km² grid square resolution for (A) WS and (B) HS. Colours indicate the relative magnitude of predicted recording deficit within each species group, with bluer colours indicating higher relative deficit. Colour scales are group-specific and should not be interpreted as directly comparable between panels. The scale bar and north arrow apply to both panels

aggregated semi-natural habitats (such as broadleaf woodland, lowland heath, and semi-natural grassland) but have relatively poor public access (Henle et al., 2016; Plummer et al., 2023). Comparisons between WS and HS species should also be interpreted cautiously because the two groups differ in species numbers and data density, with some habitat specialists having fewer records and occupied grid cells, leading to greater uncertainty in their species-level coefficient estimates.

Factors relating to human demographics may also be influencing the predicted recording deficit in built-up areas and gardens for some species of both HS and WS. Participation in biological recording, as with other forms of citizen science, lacks diversity, with severe biases based on socioeconomic status, ethnicity, gender, age and other

characteristics (Merenlender et al., 2016; Domhnaill et al., 2020; Paleco et al., 2021; Walker et al., 2021; Grade et al., 2022; Nesbit et al., 2024; Pateman et al., 2026).

Implications

The access-related variables yield several important implications for improving the spatial representativeness of citizen science data. Public access area represents a broadly applicable explanatory variable across WS and HS. The significantly stronger effect of footpath length for WS compared with HS suggests the value of targeting low footpath density areas to reduce spatial recording gaps for widespread butterfly species. Moreover, built-up areas and gardens may be more appropriately treated as a proxy for recording effort, rather than solely as a habitat variable (Fig. 1). This is because built-up areas typically have higher human population densities (Fig. 1) and therefore more potential recorders. At the same time, a substantial proportion of urban space consists of private gardens or privately owned land that are inaccessible to the public (Guzmán Fernández et al., 2020; Verheij et al., 2023; Michelle and Satwiko, 2025). At the same time, it is possible that private gardens provide higher quality habitat for some species than the surrounding countryside, and that such habitats are under recorded for WS. Thus, built-up areas and gardens are more indicative of where people are located and their limited access to sampling sites, not only the underlying quality or suitability of butterfly habitats.

To improve the spatial representativeness of recording data, two complementary strategies could be adopted. For WS, the public could be encouraged to record butterflies in and around their residential area, such as in private gardens, for example through initiatives such as the Big Butterfly Count (Dennis et al., 2017). For HS, targeted recording could be promoted by securing access permissions to currently restricted inland semi-natural habitats where public access is limited.

Our models suggest that the largest recording deficits are found in urban areas. Whilst this may be exacerbated by our restricted definition of public access parameters in urban areas it may also represent a socio-economic effect. The models therefore suggest that recording in urban areas warrants specific attention, particularly in residential settings where access to private gardens may limit spatial coverage (Dennis et al., 2017).

By quantifying public access, we can both better understand current recording biases and identify practical routes to address data gaps. Through explicitly incorporating public access into a negative binomial generalized linear model, we can separate ecological patterns from recording bias and estimate the potential distribution of records under

unrestricted access. Beyond individual species, aggregating predicted recording deficits across species could also help identify grid cells where species richness may be underestimated because multiple species are likely to be under-recorded, supporting more targeted recording strategies and improving the use of citizen science data in biodiversity assessment and conservation planning.

Future directions

We acknowledge some limitations of the data sources that we used for access related variables. Our model may predict a relatively high recording deficit due to low footpath density (Figs. 1 and 4). This might be because the footpath data may only include public rights of way and, therefore, do not include all the pavements along public roads in cities, towns and villages throughout the two counties. However, there may in fact be active recorders observing butterflies in private gardens within that grid. Conversely, the same grid could consist largely of high-density apartments or heavily sealed surfaces, where butterfly habitat is genuinely limited. Moreover, as the present model does not distinguish between different types of built-up areas and gardens, it may overlook important within-urban heterogeneity. These models could be adapted to divide built-up areas and gardens into more detailed categories and incorporate these into a refitted model in finer spatial scale and improved metrics of public access, thereby enhancing the separation of recording bias from genuine ecological signals in urban landscapes.

Another limitation of the study is the way that species have been associated with relative broad land use classes. Each of these classes encompasses a wide range of habitats and ecological communities. For example, here Marsh Fritillary is associated with semi-natural grassland but, in reality, the species can only occur in a small proportion of such grasslands which have a high density of its larval hostplants (Smee et al., 2011; Johansson et al., 2020). Thus, our approach may overestimate the extent of suitable habitat for some species and consequently influence estimates of recording deficit, which are derived from species–land-cover relationships. These relationships were based on expert judgement from a single specialist and may therefore introduce some subjectivity.

Residual spatial autocorrelation was detected in the fitted models, with Moran's *I* tests showing significant positive spatial autocorrelation in the Pearson residuals of both the full models and the models excluding public access variables (Table S1). This suggests that some spatially structured variation in recording intensity remained unexplained, and future work could incorporate spatial random effects or spatial smooth terms to better account for spatial structure.

Our cross-species comparisons were based on Welch's *t*-tests applied to species-level coefficient estimates. This approach does not propagate the uncertainty associated with individual species-level estimates, and results should therefore be interpreted as broad comparative patterns rather than formal hierarchical inference. Future work could use hierarchical models to estimate group-level differences while accounting for species-level uncertainty.

In addition, the current model does not explicitly incorporate other spatial or temporal sources of bias that may influence recording patterns, such as sampling preferences (Goldstein et al., 2024; Backstrom et al., 2025; Habel et al., 2025), the spatial distribution of recorders (Li et al., 2025), temporal variation in effort (Bishop et al., 2013; Massimino et al. 2021, Li et al., 2026; Sánchez-Fernández et al., 2026), or sampling completeness (Sánchez-Fernández et al., 2021). We argue that the predicted maps themselves likely capture, to some extent, the spatial patterns of these underlying biases, rather than reflecting only true species distributions or accessibility constraints. Future research should, therefore, more systematically disentangle the relationships among different driving factors, for example by assessing interactions between variables or by constructing explicit Directed Acyclic Graphs and quantifying both direct and indirect causal pathways (Boyd et al., 2025). Such approaches would allow a more precise evaluation of the relative contribution of public access to land on variation in the number of species records, while distinguishing its effects from other spatial or temporal sources of bias (Pearl, 2009).

Conclusions

Our findings demonstrate that the spatial patterns of species distributions recorded by citizen science are shaped not only by species' habitat but also by systematic differences in public access. Public access area emerges as a consistently important explanatory variable across both WS and HS. In contrast, the stronger effect of footpath length for WS highlights the differential sensitivity of species groups to access infrastructure and its role in shaping spatial recording patterns. Recording deficits are most likely to emerge where suitable butterfly habitat coincides with limited public access. Addressing this therefore involves differentiated strategies, whereby recording close to residences is prioritised for WS, while access to restricted semi-natural habitats is facilitated for HS. By incorporating public access into a negative binomial modelling framework, we provide an operational tool for identifying potential under recorded sites prior to resource allocation, and strengthening management

decision making, and supporting more equitable and effective monitoring strategies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10841-026-00826-4>.

Acknowledgements We thank Chloë Smith, Head of Biological Data at Butterfly Conservation, for providing access to the butterfly recording data used in this study.

Author contributions M.L. conducted the analyses, prepared the figures, and wrote the manuscript. R.H.f.-C. and J.B. conceived the study and contributed to methodological development, with J.B. providing additional methodological guidance. R.F. provided the data. D.B.R. and R.J.B. contributed comments and revisions. All authors reviewed and approved the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial, or not for profit sectors.

Data availability The datasets generated and analysed during the current study are available in the Figshare repository (<https://doi.org/10.6084/m9.figshare.31421861>).

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arkema KK, Fisher D, Wyatt KH, Wood SA, Payne HJ (2021) Advancing sustainable development and protected area management with social media-based tourism data. *Sustainability*. <https://doi.org/10.3390/su13052427>
- Asher J, Warren M, Fox R, Harding P, Jeffcoate G, Jeffcoate S (2001) *The millennium atlas of butterflies in Britain and Ireland*
- Backstrom LJ, Drake RL, Worthington H, Johnston A (2025) Detections of rare species lead citizen scientists to initiate data recording. *Divers Distrib* 31:e70103
- Bishop TR, Botham MS, Fox R, Leather SR, Chapman DS, Oliver TH (2013) The utility of distribution data in predicting phenology. *Methods Ecol Evol* 4:1024–1032
- Börschig C, Klein A-M, VON Wehrden H, Krauss J (2013) Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic Appl Ecol* 14:547–554
- Boyd RJ, Botham M, Dennis E, Fox R, Harrower C, Middlebrook I, Roy DB, Pescott OL (2025) Using causal diagrams and super-population models to correct geographic biases in biodiversity monitoring data. *Methods Ecol Evol* 16:332–344
- Choi D-A, Park K, Rigolon A (2020) From XS to XL urban nature: examining access to different types of green space using a 'just sustainabilities' framework. *Sustainability (Basel)* 12:6998
- Cooper JEJ, Plummer KE, Middlebrook I, Siriwardena GM (2024) Using butterfly survey data to model habitat associations in urban developments. *J Appl Ecol* 61:773–783
- Dandy JE (1971) Watsonian vice-counties of Great Britain. *Taxon* 20:627
- Dapporto L, Dennis RLH (2013) The generalist-specialist continuum: Testing predictions for distribution and trends in British butterflies. *Biol Conserv* 157:229–236
- Dennis EB, Morgan BJ, Brereton TM, Roy DB, Fox R (2017) Using citizen science butterfly counts to predict species population trends. *Conserv Biol* 31:1350–1361
- Díaz-Calafat J, Jaume-Ramis S, Soacha K, Álvarez A, Piera J (2024) Revealing biases in insect observations: a comparative analysis between academic and citizen science data. *PLoS One* 19:e0305757
- Domhnaill CM, Lyons S, Nolan A (2020) The citizens in citizen science: demographic, socioeconomic, and health characteristics of biodiversity recorders in Ireland
- Donkersley P, Ashton LA, Lamarre GPA, Segar ST (2022) Global insect decline is the result of wilful political failure: a battle plan for entomology. *Ecol Evol*. <https://doi.org/10.1002/ece3.9417>
- Dyer RJ, Gillings S, Pywell RF, Fox R, Roy DB, Oliver TH (2017) Developing a biodiversity-based indicator for large-scale environmental assessment: a case study of proposed shale gas extraction sites in Britain. *J Appl Ecol* 54:872–882
- Frigerio D, Per Richtera, Prüse E, B., Vohland K (2020) Citizen science in the natural sciences. *The Science of Citizen Science*, 79–96
- Gatis N, Luscombe Carlessd, Brazier D, R., Anderson K (2025) An operational land cover and land cover change toolbox processing open-source data with open source software (dataset). University of Exeter
- Goldstein BR, Stoudt S, Lewthwaite JM, Shirey V, Mendoza E, Guzman LM (2024) Logistical and preference bias in participatory science butterfly data. *Front Ecol Environ* 22:e2783
- Grade AM, Chan NW, Gajbhiye P, Perkins DJ, Warren PS (2022) Evaluating the use of semi-structured crowdsourced data to quantify inequitable access to urban biodiversity: a case study with eBird. *PLoS One* 17:e0277223
- Guetté A, Caillault S, Pithon J, Pain G, Daniel H, Marchadour B, Beaujouan V (2022) Who and where are the observers behind biodiversity citizen science data? Effect of landscape naturalness on the spatial distribution of French birdwatching records. *Land* 11:2095
- Guzmán Fernández K, Moreno-Calles AI, Casas A, Blancas J (2020) Contributions of urban collective gardens to local sustainability in Mexico City. *Sustainability*. <https://doi.org/10.3390/su12187562>
- Habel JC, Huemer P, Schmitt T, Rüdiger J, Ulrich W (2025) Strengths and shortcomings of citizen science data: lessons from Austrian butterflies. *J Insect Conserv* 29:1–13
- Habel JC, Schmitt T, Ulrich W, Gros P, Salcher B, Teucher M (2023) Landscape homogenisation and simplified butterfly community structure go on par across Northern Austria. *Landsc Ecol* 38:3237–3248
- Han D, Zhang C, Wang C, She J, Sun Z, Zhao D, Bian Q, Han W, Yin L, Sun R (2021) Differences in response of butterfly diversity and species composition in urban parks to land cover and local habitat variables. *Forests* 12:140
- Henle K, Stoev P, Penev L (2016) Nature Conservation-achievements and challenges within its first four years. *Nat Conserv* 14:1–5

- Jain A, Zeng Y, Webb EL (2021) Critical dependence of butterflies on a non-native host plant in the urban tropics. *Front Ecol Evol*. <https://doi.org/10.3389/fevo.2021.655012>
- Jingu S, Ogawa Y (2025) Effects of Accessibility on the Records of Biodiversity Observation by the Citizens of Urban Forests. *ISPRS Annals Photogrammetry Remote Sens Spat Inform Sci* 10:67–74
- Johansson V, Kindvall O, Askling J, Franzén M (2020) Extreme weather affects colonization–extinction dynamics and the persistence of a threatened butterfly. *J Appl Ecol*. <https://doi.org/10.1111/1365-2664.13611>
- Kasiske T, Klimek S, Dauber J, Harpke A, Kühn E, Levers C, Schwiender M, Settele J, Sietz D, Tetteh GO, Musche M (2025) Identifying typical patterns of land-use and landscape structure in citizen science butterfly monitoring. *Ecol Indic* 180:114317
- Krämer B, Poniatowski D, Fartmann T (2012) Effects of landscape and habitat quality on butterfly communities in pre-alpine calcareous grasslands. *Biol Conserv* 152:253–261
- León-Cortés JL, Bried JT, Roy DB (2024) Citizen science for insect conservation. *Routledge Handbook of Insect Conservation*. Routledge
- Li M, Boyd RJ, Smith C, Fox R, Roy D, Bennie J (2025) Citizen scientists as butterfly predators: using foraging theory to understand individual recorder behaviour. *Ecol Model* 510:111344
- Li M, Roy Bennie J, Fox DB, Boyd R, Smith RJ, C., Ffrench-Constant RH (2026) Systematic differences in phenology estimates from unstructured and structured biodiversity datasets. *Insect Conservation and Diversity*
- Mair L, Ruete A (2016) Explaining spatial variation in the recording effort of citizen science data across multiple taxa. *PLoS One* 11:e0147796
- Mandeville CP, Nilsen EB, Finstad AG (2022) Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use. *Ecol Solutions Evid* 3:e12185
- Massimino D, Harris SJ, Gillings S (2021) Phenological mismatch between breeding birds and their surveyors and implications for estimating population trends. *J Ornithol* 162:143–154
- Merenlender AM, Crall A, Drill S, Prysby MD, Ballard HL (2016) Evaluating environmental education, citizen science, and stewardship through naturalist programs. *Conserv Biol*. <https://doi.org/10.1111/cobi.12737>
- Michelle B, Satwiko P (2025) Striking a balance between nature and development: investigating private green spaces using ENVI-met. *J Reg City Plan*. <https://doi.org/10.5614/jpwk.2024.35.3.4>
- Mondain-Monval T, Pocock M, Rolph S, August T, Wright E, Jarvis S (2024) Adaptive sampling by citizen scientists improves species distribution model performance: a simulation study. *Method Ecol Evolut* 15(7):1206–1220
- Morton, R. D., Marston, C. G., Neil, A. W., Rowland CS (2024) ; Land cover map 2023 (10m classified pixels, GB).
- Naturalengland (2025); CRow Act 2000 access layer
- Nesbit R, Paarlberg LE, Jo S (2024) The decline of volunteering in the United States: is it the economy? *Nonprofit Volunt Sect Q* 54:583–613
- Paleco C, García Peter S, Salas Seoane N, Kaufmann J, Argyri P, Vohland K, Land-Zandstra A, Ceccaroni L, Lemmens R, Perelló J, Ponti M (2021) Inclusiveness and diversity in citizen science. *Sci Citiz Sci* 261:261–282
- Pateman, R. M., Fox, R., Dickinson, H., Haysom, K. A., Hughes, J., Kelly, L., Leech, E. I., Lowe, M., Merry, K. & Murphy, R. D. 2026. Increasing inclusivity in biological recording in the United Kingdom: progress to date and future priorities [Meeting report]. *Citizen Science: Theory and Practice*, 11.
- Pearl J (2009) Causal inference in statistics: an overview
- Pendl M, Hussain RI, Moser D, Frank T, Drapela TE (2021) Influences of landscape structure on butterfly diversity in urban private gardens using a citizen science approach. *Urban Ecosyst* 25:477–486
- Pla-Narbona C, Stefanescu C, Pino J, Cabrero-Sañudo FJ, García-Barros E, Munguira ML, Melero Y (2021) Butterfly biodiversity in the city is driven by the interaction of the urban landscape and species traits: a call for contextualised management. *Landsc Ecol* 37:81–92
- Plummer KE, Dadam D, Brereton TM, Dennis EB, Massimino D, Risely K, Siriwardena GM, Toms MP (2023) Trends in butterfly populations in UK gardens-new evidence from citizen science monitoring. *Insect Conserv Divers* 17:345–357
- Powney GD, Isaac NJB (2015) Beyond maps: a review of the applications of biological records. *Biol J Linn Soc* 115:532–542
- Prudic KL, Cruz TMP, Winzer JIB, Oliver JC, Melkonoff NA, Verbais H, Hogan A (2022) Botanical gardens are local hotspots for urban butterflies in arid environments. *Insects*. <https://doi.org/10.3390/insects13100865>
- QGIS.ORG (2025) QGIS geographic information system (Version 3.34). QGIS Association
- Rischel D (2024) Property, nature, and the freedom to roam. *Crit Rev Int Soc Polit Philos*. <https://doi.org/10.1080/13698230.2024.2387995>
- Robertson HG (2010) Public access to private land for walking: Environmental and individual responsibility as rationale for limiting the right to exclude. *Geo Int'l Environ L Rev* 23:211
- Roy HE, Martinou AF, Pocock MJ, Werenkraut V, Roy DB (2024) The global reach of citizen science for monitoring insects. *One Earth* 7:552–557
- Sánchez-Fernández D, Fox R, Dennis RL, Lobo JM (2021) How complete are insect inventories? An assessment of the British butterfly database highlighting the influence of dynamic distribution shifts on sampling completeness. *Biodivers Conserv* 30:889–902
- Sánchez-Fernández D, Mañas-Jordá S, García-Meseguer AJ, Acosta R, Ariño A, Baquero E, Bartomeus I, de Biurrun G, Bonada N, Galicia D, García-Barros E, García-París M, García-Roselló E, Gil-Tapetado D, Jurado-Angulo González M, López-Rodríguez P, Martínez MJ, Millán J, Monserrat A, Morano VJ, Pina Emunguirampallars, Romo Fprietoe, Sánchez-Campaña H, Figueroa C, Carpe, J. M. T., Yela, J. L., Lobo JM (2026) Priority areas for arthropod conservation in the Iberian Peninsula and Balearic Islands: Insights from a multi-taxon distributional database. *Insect Conservation and Diversity*, n/a
- Smee M, Smyth W, Tunmore M, Ffrench-Constant R, Hodgson D (2011) Butterflies on the brink: habitat requirements for declining populations of the marsh fritillary (*Euphydryas aurinia*) in SW England. *J Insect Conserv* 15:153–163
- Soroye P, Newbold T, Kerr J (2020) Climate change contributes to widespread declines among bumble bees across continents. *Science* 367:685–688
- Spalding A (2022) Butterflies of Cornwall. Atlas for the Twenty-First Century by Sarah Board, Tristram Besterman, Bob Dawson, Dick Goodere, Maggie Goodere and Cerin Poland. *Entomologist's Gaz* 73:64–64
- Verheij J, Ay D, Gerber J-D, Nahrath S (2023) Ensuring public access to green spaces in urban densification: the role of planning and property rights. *Planning Theory Pract* 24:342–365
- Videvall E, Öckinger E, Pettersson LB (2016) Butterfly monitoring using systematically placed transects in contrasting climatic regions - Exploring an established spatial design for sampling. *Nat Conserv* 14:41–62
- Walker DW, Smigaj M, Tani M (2021) The benefits and negative impacts of citizen science applications to water as experienced by participants and communities. *Wiley Interdiscip Rev Water* 8(1):e1488

Wang K, Qi Y, Hu Y, Han D (2025) Distribution and influences on butterfly diversity in urban park green spaces: a case study of Harbin, China. *Ecol Evol* 15:e71554

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.