Check for updates

Diversity and Distributions

WILEY

Diversity and Distributions

RESEARCH ARTICLE OPEN ACCESS

Causal Analysis Reveals That Public Access to Land Can Shape Our View of Species' Geographic Distributions

Mingrui Li[1] | Richard H. ffrench-Constant[1,†] | Jonathan Bennie[2] | Richard Fox[1,3] | David B. Roy[1,4] | Chloë Smith[3] | Robin J. Boyd[1,4]

[1]Centre for Ecology and Conservation, University of Exeter, Cornwall, UK | [2]Centre for Geography and Environmental Science, University of Exeter, Cornwall, UK | [3]Butterfly Conservation, Dorset, UK | [4]UK Centre for Ecology and Hydrology, Oxfordshire, UK

Correspondence: Mingrui Li (ml902@exeter.ac.uk)

Received: 22 January 2026 | **Revised:** 28 April 2026 | **Accepted:** 3 July 2026

Editor: Luca Santini

Keywords: butterfly monitoring | citizen science | directed acyclic graph | piecewise path analysis | public accessibility

ABSTRACT

Aim: Citizen science data are widely used to study species' distributions. Using butterfly observations from South-West England, we sought to better understand what shapes the spatial footprint of recording.

Location: South-West England.

Methods: We combined directed acyclic graphs (DAGs) with piecewise path analysis to quantify the causal effects of public access and habitat (using land cover as a proxy) on the number of records for the 39 butterfly species that occur in this region.

Results: The length of public footpath per unit area emerged as the strongest driver of local recording outcomes, especially for common species, while human population density and area of public access land show weaker or more variable effects. The total effect of public access features was larger than the total effect of habitat for 22 of the 39 species.

Main Conclusions: Overall, our findings indicate that current butterfly occurrence data largely reflect where citizen scientists can go, rather than solely where species occur. Incorporating accessibility metrics can therefore improve the design and interpretation of citizen science recording schemes and surveys.

1 | Introduction

Unstructured citizen science observations of species have proved vital to assess insect trends and drivers of change in the UK (Powney et al. 2019; Outhwaite et al. 2020; Suggitt et al. 2023; Bourhis et al. 2025) and elsewhere (Soroye et al. 2020; Engelhardt et al. 2022; Shirey et al. 2024). These observations, which reveal where species occur across landscapes, often underpin effective biodiversity conservation (Hailay Gebremariam 2024). However, the resulting species' distributions reflect a combination of genuine species occurrence and recorder behaviour as citizen scientists are free to collect data wherever they like (Mondain-Monval et al. 2024). The spatial distribution of citizen science records may then primarily reflect where people can go (public access) or where observers expect species to occur.

Existing studies have shown that the recording behaviour of citizen scientists is influenced by many factors, such as recorder expertise (Bowler et al. 2022; Kühn et al. 2024), sampling site preference (Dennis and Thomas 2000; Li et al. 2025) and taxonomic preference (Goldstein and Stoudt 2025; Li et al. 2025). In this study, we focus on a single issue: where recorders choose to visit in order to make records. The recorder's choice is constrained by whether they have the opportunity to access specific locations. It is therefore important to understand the role of public access in shaping where observations are actually made.

†Richard H. ffrench-Constant has passed away.

Countries differ widely in the extent to which citizens have legal, or de facto, access to land, which may have a profound effect on the distribution of citizen science recording effort. For example, the Nordic countries have a strong tradition of *allemansretten*, the right to access and spend time in uncultivated land (King et al. 2024), while in the United States, access is largely limited to federal or state-owned land, with private property rights including the right to restrict public access (Robertson 2010). In England and Wales, public access to the countryside is provided by a network of public rights of way (footpaths and bridleways) to which the public have a legal right of access, and to certain classes of land to which they have unrestricted access ('right to roam') under the Countryside and Rights of Way (CROW) act of 2000 (Natural England 2025), as well as permitted access provided by certain landowners including the National Trust and Forestry England. The extent to which land is accessible thus varies spatially, providing a valuable system for investigating the effect of public accessibility on recorder effort.

A range of spatial factors can be regarded as components of public access to land. These include the density of footpath networks, the extent of open access areas, and the density of population living nearby. Areas with a denser network of public footpaths, shorter distances from paths or roads, and higher surrounding human population densities are generally more accessible to people and are therefore expected to attract more recording visits (Mair and Ruete 2016; Guetté et al. 2022; Mandeville et al. 2022). In addition, the extent of land that is legally or physically open to the public, such as parks, nature reserves and other access land, may influence where recorders are able to enter and make observations (Mandeville et al. 2022; Kirilenko 2024). Similarly, recording can also be influenced by transport links (Sicacha-Parada et al. 2021; Mandeville et al. 2022). Together, these factors describe the accessibility structure of the landscape, which in turn is expected to influence occurrence records of species.

The distribution of records is constrained by species distributions, as records cannot occur in locations where the species is absent. Species distributions are themselves shaped by habitat conditions (Guilherme et al. 2024; Annessi et al. 2025; Han et al. 2025) and climate (Rödder et al. 2021). Consequently, habitat and public access jointly influence the spatial distribution of butterfly records. While previous studies have reported correlations between public accessibility and recording patterns (Dennis and Thomas 2000; Romo et al. 2006; Mair and Ruete 2016; Tiago et al. 2017; Mandeville et al. 2022; Geurts et al. 2023), these analyses have remained correlative, and causal approaches have rarely been applied to disentangle the effects of habitat associations and public accessibility on record numbers.

Assessing the effects of this complex array of potential causal relationships on species occurrence records requires an analytical approach that explicitly accounts for confounding variables that induce a non-causal association between explanatory and response variables and for indirect effects (Boyd et al. 2023; Guzman et al. 2024; Chen et al. 2025). Correlations between habitat or public access features and recording cannot be interpreted causally due to confounding, as both are influenced by unobserved factors (Pearl 1995). This issue is not resolved by model-based variable selection (Arif and Macneil 2022), meaning that model coefficients may not reflect true causal relationships (Pearl 2009). Directed acyclic graphs (DAGs), one type of causal diagram, provide a transparent framework to represent these relationships (Pearl 2009). By using the DAG approach, we can identify as many potential confounders as possible and then estimate the effect of each variable on the outcome, which is the causal effect (Thoemmes and Rose 2014). If we do not adjust for the correct set of variables and their relationships, the coefficient results would be misleading (Lee et al. 2023; Boyd et al. 2025).

Here, we use citizen-science butterfly occurrence observations in Devon and Cornwall in United Kingdom as a case study to examine how accessibility and habitat jointly shape recording patterns. Specifically, we aim to address the question 'how strong are the causal effects of public access features on whether a citizen science record was made relative to habitat'?

2 | Methods

2.1 | Overview

We combined causal diagrams with piecewise path analysis to infer the causal effects of habitat and public access features on whether any given site was sampled. We first constructed a DAG to define the hypothesised causal relationships between public access, land use and the number of butterfly records. Then, we applied piecewise path analysis, representing each causal pathway as an independent linear model to quantify the total causal effects of public access and land use on the number of butterfly records.

2.2 | Study Area

The counties of Devon and Cornwall in southwest England form a peninsula covering approximately 10,200 km^2. The study area was divided into 1 km × 1 km grid cells (10,995 grids in total) based on the British National Grid, which served as the spatial unit of analysis (Figure 1). 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. 2022). 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).

2.3 | Data

Butterfly data came from occurrence records submitted by citizen scientists via the iRecord Butterflies mobile application, managed by the UK Centre for Ecology & Hydrology in partnership with Butterfly Conservation. We used data for Devon and Cornwall (Watsonian Vice Counties 1, 2, 3 and 4) in south-west England, comprising 51,045 species occurrence records spanning the period 14 July 2011–1 April 2024. Records from the Isles of Scilly and Lundy Island were excluded from

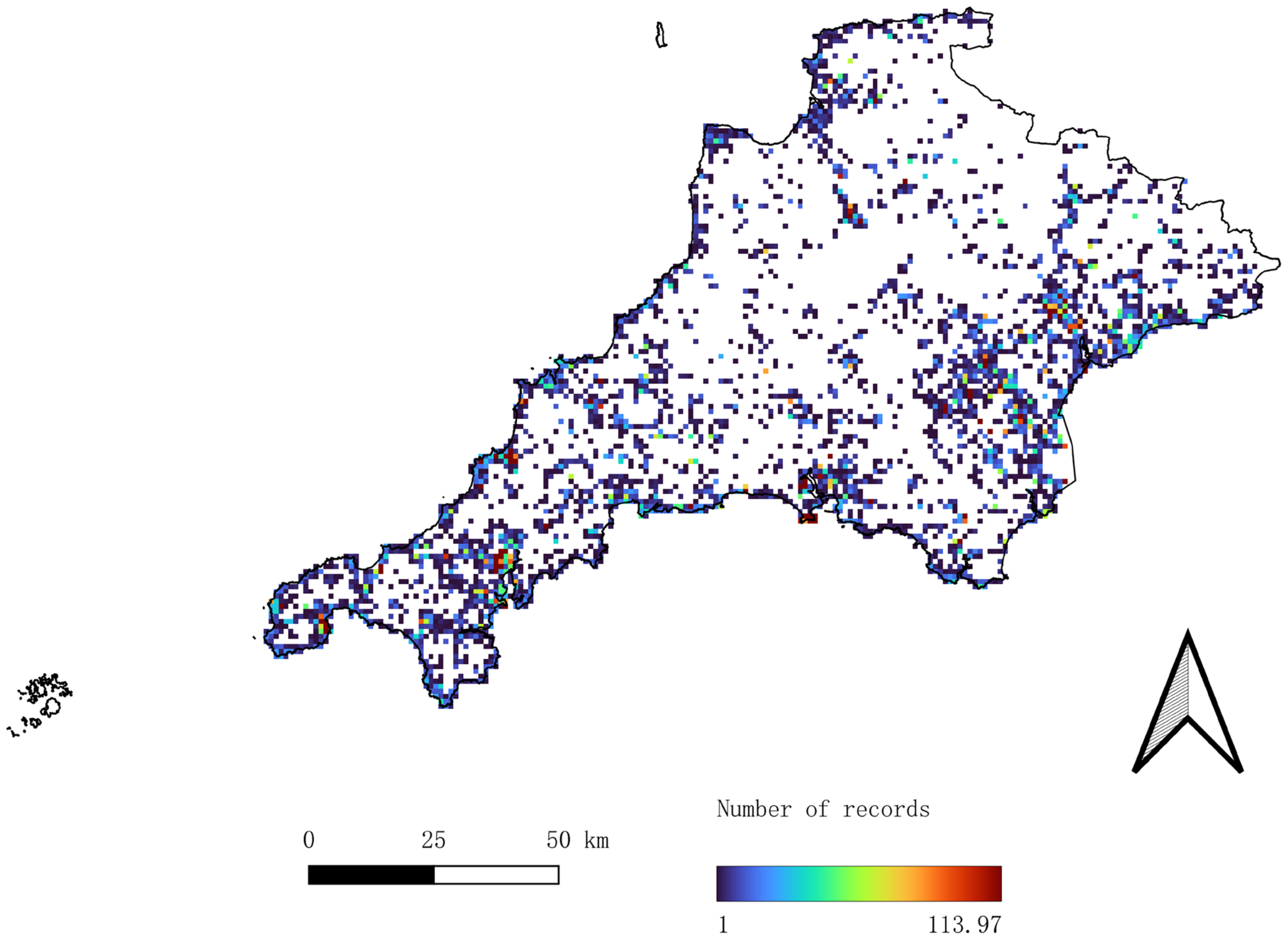

FIGURE 1 | Spatial distribution of total occurrence records across 1 km² grid squares in the study area: The colour scale was truncated to the 2nd–98th percentile to improve visualisation and reduce the influence of extreme values. Records from the Isles of Scilly and Lundy Island were excluded from the analysis.

the analysis (Figure 1). All submitted records were included in the analysis. The vast majority of records had been verified and accepted by local taxonomic experts, with only a very small number (< 0.05%) remaining unconfirmed but classified as plausible.

To quantify public access, we used human population density in 1 km² grid squares (Population_density), area of publicly accessible land from online databases and the mapped routes of public rights of way (footpaths, bridleways and byways). A detailed description of all datasets is provided in Appendix S1.

Butterfly species were associated with the habitats in which each is typically found by expert opinion (Table 2). Since data on microhabitats are not available, we used land cover as a proxy. We used historical data for butterfly distributions in Devon and Cornwall to distinguish between species that are relatively rare and common in the study area. The number of 1 km² grids in which each species was recorded in the study area in the period 1995–2005 was calculated from the national recording scheme for UK butterflies, Butterflies for the New Millennium (BNM) (Table 2). This dataset includes observational records for butterflies from a range of sources, and the date period was chosen to provide complete taxonomic coverage and good geographical coverage of butterflies. Species rarity status was defined as rare if they occupied < 500 grids and common if > 500 in this baseline data. In terms of habitat types, we used the 10 basic types (aggregate classes) of land cover (Table 1) in the UKCEH Land Cover Map for 2023 (Morton et al. 2024), plus hedgerow length (Hedge_length) and extent of sand dunes per kilometre square (Sand_dunes). Hedgerow data were derived from UKCEH Land Cover Plus: Hedgerows 2016–2021 (England) (Broughton et al. 2024). Sand dune data were accessed from the Priority Habitats Inventory (England) published by Natural England (1) (Priority Habitat Inventory (England) 2025).

Following data preparation, all relevant environmental datasets were imported into QGIS version 3.34 (QGIS.ORG 2025) for spatial processing and Python code was used for data processing. The study area excluded the Isles of Scilly and Lundy Island. A 1 km² resolution grid was generated across the entire study area using the Ordnance Survey British National Grid (BNG) coordinate reference system. All subsequent spatial analyses were conducted within this gridded framework. For each grid square, Footpath_length, Public_access, Population_density, Hedge_length, Footpath_alongside_hedge, and Distance_from_coast were calculated (see Table 3 for definitions of each variable).

Species occurrence records from the iRecord Butterflies dataset were summarised to 1 km grid square level. For each grid square,

TABLE 1 | Aggregation of UKCEH land cover classes into 10 habitat categories used in butterfly analyses: Defining the habitat variables used in all models.

UKCEH aggregate class (AC)	Definition
Broadleaf woodland	Broadleaved mixed and yew woodland
Coniferous woodland	Coniferous woodland
Arable	Arable and horticulture
Improved grassland	Improved grassland
Semi-natural grassland	Neutral grassland/ Calcareous grassland/ Acid grassland/Fen marsh and swamp
Mountain, heath and bog	Dwarf shrub and heath/Bog/Inland rock
Saltwater	Saltwater
Freshwater	Standing open water and canals/ Rivers and streams
Coastal	Supralittoral rock/ Supralittoral sediment/ Littoral rock/ Littoral sediment
Built-up areas and gardens	Built-up areas and gardens

the total number of records per species was calculated, irrespective of the date of observation. It is important to note that many grid squares contained no butterfly records. However, these grid squares were retained in the final dataset to ensure a complete spatial representation of the study area. The absence of records in these squares was treated as zero observations, rather than excluding them from the analysis.

The final processed dataset was structured such that each row represented a single $1\,\text{km}^2$ grid square. Columns included land cover variables, public access variables, other relevant environmental predictors, and the observed number of records of each butterfly species within that grid square.

2.4 | Directed Acyclic Graph (DAG)

The development of the DAG followed four main steps (Lee et al. 2023). First, an initial causal model was formulated based on theoretical assumptions about the relationships among variables. Second, we tested the conditional independence relationships implied by the model. Third, the model was revised according to empirical evidence and expert judgement. Finally, this process was repeated until a causal structure was obtained that was consistent with prior ecological theory and with the results of conditional independence tests applied to intermediary DAGs. While these intermediary DAGs were empirically testable, the final DAG represents a theoretically plausible causal structure rather than one that is itself directly testable. Full details are provided in Appendix S1.

We first selected a sample species, Speckled Wood, and constructed the DAG based on this species. We then generalised this structure to other species, assuming that the set of public accessibility variables remains identical and is included for all species, while differences among species arise only in the specific land cover variables (i.e., habitat characteristics) included in the model. Within this framework, we assumed that land cover influences true abundance directly and therefore affects the number of records (occurrence data) indirectly, whereas public accessibility variables have a direct effect on the number of records.

2.5 | Piecewise Path Analysis

All relationships among variables were modelled using linear regressions within the framework of piecewise path analysis (Lefcheck 2016). For a continuous exposure X (explanatory variables), our estimate is the causal marginal mean function $\mu(x) = E[Y(x)]$, where $Y(x)$ denotes the potential outcome under exposure level x. Linear regression incorporating the relevant adjustment set estimates a best linear approximation to $\mu(x)$, corresponding to an average marginal causal slope over the observed exposure distribution (Hernán and Robins 2010).

In this approach, each response variable was modelled separately through its own linear regression equation, and the system of equations collectively represented the hypothesised causal network. Total causal effects were calculated manually as follows: effects along the same path were multiplied, and effects from different paths were summed, with signs preserved (Grace (2006), p65). Under a linear modelling framework, this definition is equivalent to estimating the total effect via a single regression of the response variable on the focal explanatory variable, and therefore does not require explicit modelling of all mediating pathways.

Footpath_length, Population_density, and Public_access were grouped as the total access effect, while all land cover variables associated with a given species were grouped as the total land cover effect. For total access effect and total land cover effect, absolute values were used to calculate a log-transformed ratio ($\log_{10}$) for comparison among different butterfly species. Confidence intervals for these total effects were subsequently estimated through internal resampling procedures (details available in Appendix S1).

In order to test whether the effect of public access differed significantly between common and rare species, we calculated the weighted mean total causal effect of public access (weighted by number of $1\,\text{km}^2$ grids recorded in Devon and Cornwall 2011–2024, Table 2), and compared the two groups using a Mann Whitney U test. In addition, we also estimated the weighted mean (using relative abundance as above) causal effects for each individual access variable and tested for significant differences between common and rare species using Mann Whitney U tests. Results were summarised using boxplots showing the weighted mean ± SD, illustrating the variation in causal effects across species within each group.

TABLE 2 | Information on butterfly species in Devon and Cornwall.

Species, common name	Scientific name	Habitat (land cover) types	Number of 1 km² grids recorded in Devon and Cornwall 1995–2005	Species rarity status
Speckled wood	*Pararge aegeria*	Broadleaf woodland, semi-natural grassland, hedge	2422	Common
Meadow brown	*Maniola jurtina*	Arable, improved grassland, semi-natural grassland, built up areas and gardens, hedge	2421	Common
Red admiral	*Vanessa atalanta*	Broadleaf woodland, coniferous woodland, arable, improved grassland, semi-natural grassland, mountain heath and bog, saltwater, freshwater, coastal, built up areas and gardens, hedge	2412	Common
Gatekeeper	*Pyronia tithonus*	Arable, improved grassland, semi-natural grassland, built up areas and gardens, hedge	2107	Common
Green-veined white	*Pieris napi*	Broadleaf woodland, semi-natural grassland, mountain heath and bog, freshwater, built up areas and gardens, hedge	2086	Common
Small white	*Pieris rapae*	Arable, semi-natural grassland, built up areas and gardens	2079	Common
Large white	*Pieris brassicae*	Arable, semi-natural grassland, built up areas and gardens, hedge	2072	Common
Small tortoiseshell	*Aglais urticae*	Improved grassland, semi-natural grassland, hedge	1928	Common
Peacock	*Aglais io*	Improved grassland, semi-natural grassland, built up areas and gardens, hedge	1770	Common
Painted lady	*Vanessa cardui*	Semi-natural grassland, mountain heath and bog, coastal, built up areas and gardens, hedge	1691	Common
Orange tip	*Anthocharis cardamines*	Semi-natural grassland, freshwater, built up areas and gardens, hedge	1488	Common
Common blue	*Polyommatus icarus*	Semi-natural grassland	1350	Common
Holly blue	*Celastrina argiolus*	Improved grassland, semi-natural grassland, built up areas and gardens, hedge	1299	Common
Small copper	*Lycaena phlaeas*	Improved grassland, semi-natural grassland	1262	Common
Ringlet	*Aphantopus hyperantus*	Broadleaf woodland, semi-natural grassland, hedge	1245	Common
Wall	*Lasiommata megera*	Semi-natural grassland, mountain heath and bog, coastal, hedge	1117	Common

(Continues)

Species, common name	Scientific name	Habitat (land cover) types	Number of 1 km^2 grids recorded in Devon and Cornwall 1995–2005	Species rarity status
Comma	*Polygonia c-album*	Broadleaf woodland, improved grassland, semi-natural grassland, built up areas and gardens, hedge	1079	Common
Large skipper	*Ochlodes sylvanus*	Broadleaf woodland, semi-natural grassland, sand dunes, hedge	989	Common
Brimstone	*Gonepteryx rhamni*	Broadleaf woodland, freshwater, hedge	984	Common
Clouded yellow	*Colias croceus*	Arable, improved grassland, semi-natural grassland, built up areas and gardens	955	Common
Small skipper	*Thymelicus sylvestris*	Semi-natural grassland, sand dunes	838	Common
Marbled white	*Melanargia galathea*	Semi-natural grassland	659	Common
Small heath	*Coenonympha pamphilus*	Semi-natural grassland, mountain heath and bog, sand dunes	623	Common
Silver-washed fritillary	*Argynnis paphia*	Broadleaf woodland, hedge	607	Common
Green hairstreak	*Callophrys rubi*	Semi-natural grassland, mountain heath and bog	409	Rare
Purple hairstreak	*Favonius quercus*	Broadleaf woodland, improved grassland, semi-natural grassland, hedge	375	Rare
Small pearl-bordered fritillary	*Boloria selene*	Semi-natural grassland, coastal, sand dunes	344	Rare
Dark green fritillary	*Speyeria aglaja*	Semi-natural grassland, sand dunes	242	Rare
Grayling	*Hipparchia semele*	Mountain heath and bog, coastal	208	Rare
Pearl-bordered fritillary	*Boloria euphrosyne*	Broadleaf woodland, semi-natural grassland	198	Rare
Marsh fritillary	*Euphydryas aurinia*	Semi-natural grassland, mountain heath and bog	185	Rare
Dingy skipper	*Erynnis tages*	Semi-natural grassland, mountain heath and bog, sand dunes	182	Rare
Brown argus	*Aricia agestis*	Semi-natural grassland, sand dunes	112	Rare
Silver-studded blue	*Plebejus argus*	Semi-natural grassland, mountain heath and bog, sand dunes	104	Rare
White admiral	*Limenitis camilla*	Broadleaf woodland	94	Rare
Grizzled skipper	*Pyrgus malvae*	Semi-natural grassland, sand dunes	82	Rare
Wood white	*Leptidea sinapis*	Broadleaf woodland	70	Rare

(Continues)

Species, common name	Scientific name	Habitat (land cover) types	Number of 1 km^2 grids recorded in Devon and Cornwall 1995–2005	Species rarity status
High brown fritillary	*Fabriciana adippe*	Semi-natural grassland	65	Rare
Heath fritillary	*Melitaea athalia*	Broadleaf woodland, semi-natural grassland, mountain heath and bog	15	Rare

Because the response variable contained a high proportion of zeros, the data were strongly zero-inflated, which compromises the reliability of standard error and confidence interval estimates based on the linear regression assumptions. Although zero inflation does not bias point estimates (assuming a correctly specified mean model), it affects uncertainty estimation. We therefore attempted bootstrap resampling, but excessive zeros caused some bootstrap fits to fail. As a result, we did not report confidence intervals.

3 | Results

3.1 | DAG Results

The final DAG for Speckled Wood is shown in Figure 1 and Table 4 shows every causal route with theoretical justification. Following multiple rounds of conceptual refinement and revision, the resulting DAG did not imply any testable conditional independencies, meaning that no conditional independence relationships could be evaluated using the available data. However, we consider the DAG to be reasonable on theoretical grounds. The DAG was therefore used as an explicit assumption framework to guide the subsequent analyses and summarised as follows: an unobserved variable, Land use history in 1 km^2 grid square (Land_use_history), was introduced as a common cause with direct effects on all variables except Distance_from_coast, number of records for each species, and True_abundance. Distance_from_coast was included as a background variable in the DAG to support the causal structure, acting as a common cause of multiple variables and resolving unsupported conditional independence assumptions (see Supporting Information; Table 4). In addition, Hedge_length was assumed to directly influence percentage of Footpath_alongside_hedge, while the latter variable also had a direct effect on number of records among all the species. Land cover variables influence the number of records both through True_abundance and through Detectability. The detailed procedure for constructing and refining the DAG for Speckled Wood is provided in Appendix S1.

3.2 | Piecewise Path Analysis

As true abundance and detectability act as mediators of the effects of land cover on the number of records, they were not included as explicit components of the piecewise pathway framework. Our interest lies in the total effect of land cover variables on recording intensity, which can be recovered without explicitly modelling true abundance and detectability in this part of the analysis (Figure 2). In other words, the total effect of land cover variables on the number of records inherently captured both the effect of true abundance and detectability on the number of records and the effects of other land cover variables on true abundance and detectability (Figure 2).

When estimating bootstrap confidence intervals for the causal effects, the response variable (the number of records per species) contained a high proportion of zeros. Consequently, some bootstrap replicates led to linear model fitting failures for individual path models, resulting in missing estimates. To overcome this issue, we instead extracted the corresponding path-specific causal coefficients and manually combined them to derive the overall total causal effect.

3.2.1 | Total Access Effect and Total Land Cover Effect

Figure 3 shows the log-transformed ($\log_{10}$, sign-preserved) ratio of total effects of access variables to land cover variables for each species' number of records. Since the calculation used absolute values, Figure 3 reflects only the relative strength of the total access and total land cover effects (and not the direction of these separate effects). Positive values indicate species for which the number of records is more strongly influenced by access factors and dashed line represented equal total effects from both factor types ($\log_{10} = 0$). In total, 22 species' records were more strongly impacted by access.

When averaged across species groups (Figure 4), there was a significant difference between common ($N = 24$, weighted mean ± SD = 0.1237 ± 0.0980) and rare species ($N = 15$, weighted mean ± SD = 0.0218 ± 0.0268) in their total causal effect of public access (Mann–Whitney $U = 337$, $p < 0.001$), with common species showing substantially higher effects (Cliff's $\delta = 0.87$). This suggests that common species' occurrence data is more affected by public accessibility. Nevertheless, some widespread species including Comma and Large White showed negative ratios, indicating that even common butterflies respond heterogeneously to public access and land cover causal effect (Figure 3).

3.2.2 | Total Effect of Each Access Variable

From Figures 5 and 6, footpath length had the strongest positive influence on recording activity. Common species showed much higher weighted mean effects (0.0877 ± 0.0734) than rare species

TABLE 3 | Square-level variables.

Variable name	Description
Footpath_length	Total length (km) of public rights of way (footpaths, bridleways and byways) within each grid square
Public_access	Total public access area (km^2) within each grid square
Population_density	Human population density (people per km^2) within each grid square
Hedge_length	Total hedgerow length (km) within each grid square
Distance_from_coast	Shortest distance (km) from grid square centroid to British coastline, using the BOUNDARY-LINE Digital Data as a proxy to exclude inland estuaries
Footpath_alongside_hedge	Percentage of footpaths within 5 m of hedgerow, calculated as (footpath length within 5 m of hedgerow/total footpath length) × 100

(0.0070 ± 0.0078), and this difference was highly significant (Mann Whitney $U = 348$, $p < 0.001$, Cliff's $\delta = 0.93$). Public access area also had a broadly positive effect, with common species again exhibiting higher weighted mean values (0.0540 ± 0.0729) than rare species (0.0250 ± 0.0259). This difference was also significant (Mann Whitney $U = 251$, $p = 0.042$, Cliff's $\delta = 0.39$). By contrast, human population density had much weaker effects, with weighted means close to zero for both common (0.0086 ± 0.0326) and rare species (0.0003 ± 0.0008), and no significant difference between groups (Mann Whitney $U = 168$, $p = 0.740$, Cliff's $\delta = -0.07$). Detailed total effect estimates for each species are provided in the Supporting Information.

4 | Discussion

Our main results are that access is relatively more important than land cover variables for the recording of common butterfly species (Figure 4), and that the causal influence of public access and land cover differs among species (Figures 5 and 6). For example, Speckled Wood (0.3285) and Meadow Brown (0.3101) were frequently observed in semi-open habitats such as forest margins and hedgerows (Bergerot et al. 2012; Wix et al. 2019), which tend to coincide with accessible areas. However, after accounting for land cover variation, this could be explained by volunteers and the public more frequently travelling along roads and trails, sampling at accessible locations, which results in a systematically inflated number of records at these locations (Mair and Ruete 2016; Bowler et al. 2022; Mandeville et al. 2022; Dimson and Gillespie 2023; Mondain-Monval et al. 2024). In other words, occurrence data are strongly influenced by citizen-science sampling effort associated with public accessibility, rather than representing the species' true distribution. However, some rare species were dominated by public access causal effects while others were dominated by land cover causal effects (Figure 3). The relative effects indicate that recording opportunities for rare species are not driven by a single mechanism but jointly shaped by habitat constraints and spatial accessibility. This implies that inferences about rare species may require careful consideration of both habitat structure and accessibility-driven sampling biases.

When examining accessibility variables separately, common species on average showed consistently positive causal effects in areas with high human accessibility (Figure 6). This pattern was particularly evident for Footpath_length and Public_access_area, whereas Population_density exhibited much weaker effects. Previous studies have shown that volunteers tend to follow existing paths and trails rather than sampling randomly within populated areas (Mair and Ruete 2016; Geurts et al. 2023). Therefore, the stronger effect of footpath length may be explained by its ability to better represent opportunities for recorders to move and observe within semi-natural environments. Although higher human population density may increase the potential number of observers and is often associated with a greater provision of footpaths to facilitate human access, densely populated areas are often associated with higher levels of urbanisation, lower butterfly abundance and species richness and fewer natural or semi-natural habitats (Merckx and Van Dyck 2019; Vaz et al. 2023; Collins et al. 2024; Albalawi 2025; Sanllorente et al. 2025), which might weaken its positive causal effect on butterfly recording intensity.

The number of records of rare species exhibited lower mean causal effects of accessibility variables (Figure 6), indicating that recording of these butterflies is not strongly dependent on size of accessibility area, footpath length or human population density. This is likely because their scarcity and/or conservation status attracts dedicated attention and targeted surveys, meaning they are still relatively well recorded (Goldstein et al. 2024; Habel et al. 2025). Another possible explanation is that rare species are more restricted to specific habitats and are therefore only observed in accessible locations when these coincide with suitable habitat.

There are several strengths to using a DAG. First, we were able to ensure that the model could provide less biassed estimates of the effect of individual variables on record numbers while controlling for confounding influences. Secondly, variable selection for the DAG was guided not only by statistical diagnostics but also by discussions with co-authors, whose ecological expertise helped ensure that the resulting structure was ecologically plausible (Tennant et al. 2021; Lee et al. 2023). Moreover, two new key variables emerged as essential components of the causal structure. The first was an unobserved variable, Land_use_history, identified as a common cause influencing most other variables. Its inclusion likely reflects the profound and long-term influence of human activities and ecological processes on both

TABLE 4 | Theoretical justifications of assumed causal effects.

Cause	Effect	Theoretical justification
Land_use_history	Public_access	Historical patterns of land ownership, settlement and infrastructure development determine today's public access rights and trail networks.
Land_use_history	Broadleaf_woodland	Historical land management (e.g., forest retention, clearance and rotation) influences the present distribution and extent of broadleaf woodland.
Land_use_history	Semi_natural_grassland	Semi-natural grasslands persist where long-term low-intensity use has prevented conversion to intensive agriculture or urban land, reflecting past land use decisions.
Land_use_history	Footpath_length	Networks of footpaths evolve from repeated human travel and land use over time; long-established routes persist and expand based on historical access and settlement patterns.
Land_use_history	Hedge_length	Hedgerows originate from historical enclosure and field boundary establishment; their length and density reflect past land division and management practices.
Land_use_history	Footpath_alongside_hedge	Hedgerows and adjacent footpaths co-occur because historical field boundaries often guided route placement, linking historical field structure and movement pathways.
Land_use_history	Population_density	Long-term settlement locations based on historical resource availability and infrastructure continue to influence present-day population density.
True_abundance	Number of records	Sites with higher true abundance yield more opportunities for detection and recording; observers are more likely to record species where individuals are more frequent.
Broadleaf_woodland	True_abundance	Broadleaf woodland provides structural habitat and resources for many species, supporting higher true abundance compared to non-wooded areas.
Semi_natural_grassland	True_abundance	Semi-natural grasslands offer key ecological niches and foraging resources, supporting higher true abundance of grassland-associated species.
Hedge_length	True_abundance	Hedges increase habitat heterogeneity and resource availability, thus supporting higher true abundance of species that use linear woody habitat.
Footpath_length	Number of records	Longer footpath networks increase the probability of observer movement through habitat, thus raising the likelihood of species being recorded.
Population_density	Number of records	Higher human population density increases the number of potential observers and recording effort, leading to more records.
Footpath_alongside_hedge	Number of records	Footpaths adjacent to hedgerows concentrate observer movement along ecologically rich edges, increasing recording opportunities.
Distance_from_coast	Number of records	Proximity to coast shapes accessibility and human visitation patterns, with coastal areas often attracting more observers and thus higher recording rates.
Public_access	Number of records	Bigger public access area increases human presence and effort in an area, elevating the probability of species detection and recording.
Distance_from_coast	Footpath_length	Historical settlement and recreation planning in coastal areas influence the development of footpath networks, often increasing footpaths closer to the coast.

(Continues)

TABLE 4 | (Continued)

Cause	Effect	Theoretical justification
Distance_from_coast	Hedge_length	Coastal climatic and land use gradients influence hedgerow persistence and establishment, as settlement and farming intensity vary with distance from the coast.
Distance_from_coast	Semi_natural_grassland	Environmental gradients from coast to inland affect habitat types; semi-natural grassland occurrence reflects both ecological gradients and historical coastal land use.
Distance_from_coast	Public_access	Coastal locations tend to have higher recreational access due to tourism and settlement patterns, linking distance from coast with public access area.
Distance_from_coast	Population_density	Distance from coast affects settlement density, as coastal zones historically attracted human habitation and economic activity.
Distance_from_coast	Broadleaf_woodland	Climatic and soil gradients associated with coastal proximity affect the distribution of woodland types, with broadleaf woodland varying along the coast–inland gradient.
Hedge_length	Footpath_alongside_hedge	Where hedgerows are extensive, adjacent footpaths are more likely to be established, as field boundaries historically guided route placement.
Broadleaf_woodland	Detectability	Habitat structure, such as vegetation density, openness and edge effects, is likely to influence the probability of detecting and recording butterflies.
Semi_natural_grassland	Detectability	Semi-natural grasslands are species habitat thus will affect detectability
Hedge_length	Detectability	Species are also often recorded along hedgerows, which can increase their detectability.
Detectability	Number of records	High detectability means more records for a given abundance

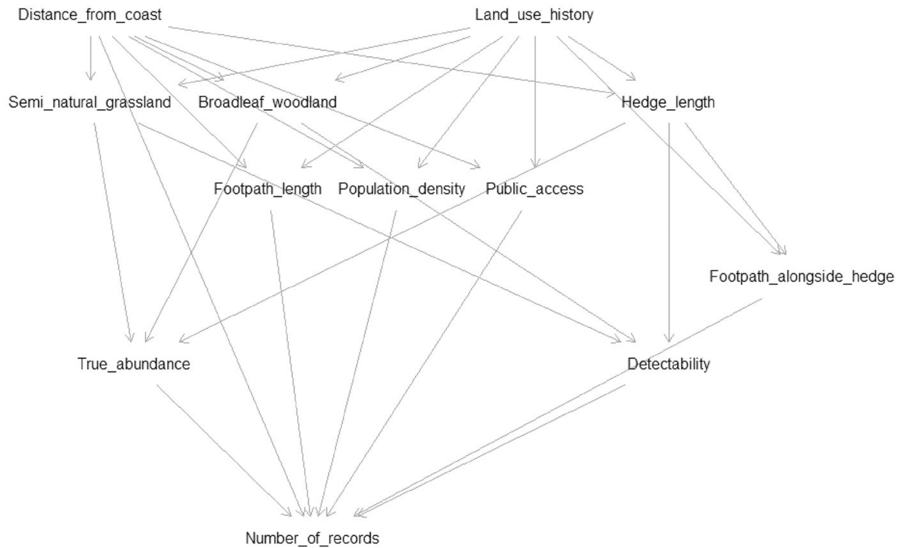

FIGURE 2 | Hypothesised DAG for speckled wood showing direct and indirect effects of land cover, access and spatial variables on the number of records.

landscape patterns and accessibility (Wurz et al. 2022). The second was Distance_from_coast, which was recognised as a driver affecting both accessibility and land cover variables, and was also assigned a direct effect on the number of records. This variable captures the combined influence of the coastal–inland gradient in the study area on multiple dimensions, including accessibility, human population distribution, and land cover composition (Zhai et al. 2025).

One limitation of this methodological approach is that, after multiple rounds of model refinement, the final DAG no longer contained any testable conditional independences. This situation is not uncommon in DAG modelling (Tennant et al. 2021), and means that the structure could not be further validated statistically. In such cases, the reliability of the causal model depends largely on theoretical reasoning and expert knowledge rather than empirical tests. Nevertheless, a well-constructed expert-informed DAG can still provide a robust basis for identifying minimal adjustment sets to reduce bias (Tennant et al. 2021; Boyd et al. 2025; Byrnes and Dee 2025). In this study, our aim was to look at the relative causal effect of public access on butterfly occurrence data. Therefore, the DAG was primarily used to identify as many relevant confounders as possible.

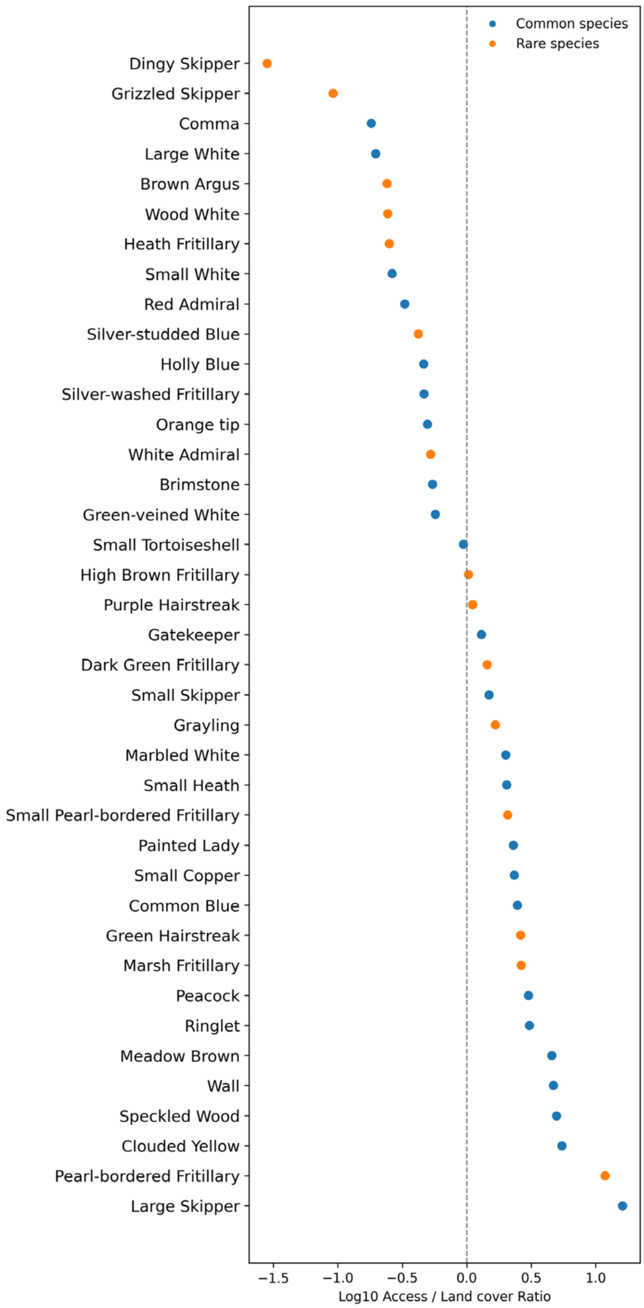

FIGURE 3 | Log_{10} ratio of access to land cover total effects per species. Dots at right side of the dotted line represent species for which the distribution of records is more strongly influenced by public access than landcover, those at left side the line for which land cover is more important.

One important implication of our study is that footpath length should be explicitly considered when analysing citizen science butterfly recording scheme data, particularly for common species. We found that footpath length had a stronger causal effect on record numbers than other accessibility variables (e.g., public access area or human population density). This suggests that observers disproportionately record in areas with more extensive path networks, potentially inflating numbers of records of common species in such areas (Mair and Ruete 2016; Bowler et al. 2022). Moreover, this also highlights a key interpretative issue: the spatial pattern of records, particularly at fine spatial scales, may represent recorder accessibility rather than species occurrence.

Moreover, while public access area also influenced record numbers for both common and rare butterfly species (e.g., by facilitating visits to large accessible reserves), it does not resolve the issue of under sampling on private land. Many potential butterfly habitats on private estates remain under recorded because observers lack permission or logistical access (Green et al. 2023). Therefore, combined efforts to improve land access and implement targeted surveys on private land and in low accessibility habitats are necessary to document species' distributions more comprehensively.

Distance_from_coast is a key variable in the study region as much recreational activity in the region is concentrated at the coastline. This variable not only captures the combined effects of land cover and climatic gradients (Mingarro et al. 2021) but also reflects human accessibility, for example coastal areas often attract substantial human activity due to tourism and recreation. The negative causal effect of distance from the coast and butterfly observation numbers for many species (see Appendix S1) mirrors both recreational use and the human population distribution pattern in this part of Britain, where human density and activity are concentrated around the coast while inland areas

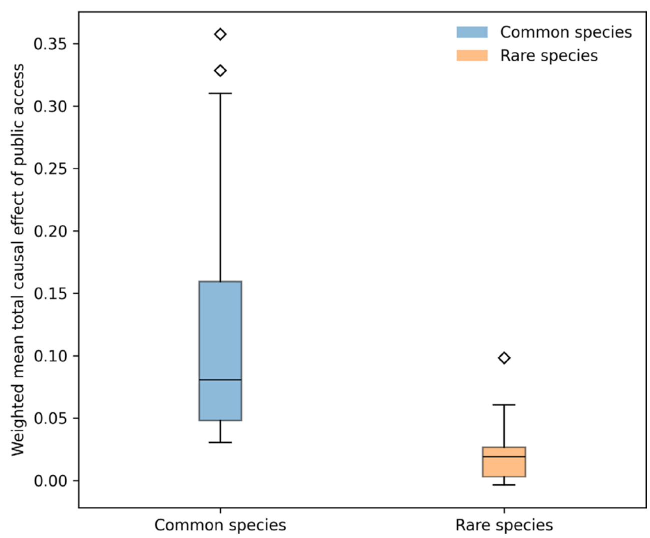

FIGURE 4 | Weighted mean total causal effect of public access on number of records for common and rare butterfly species.

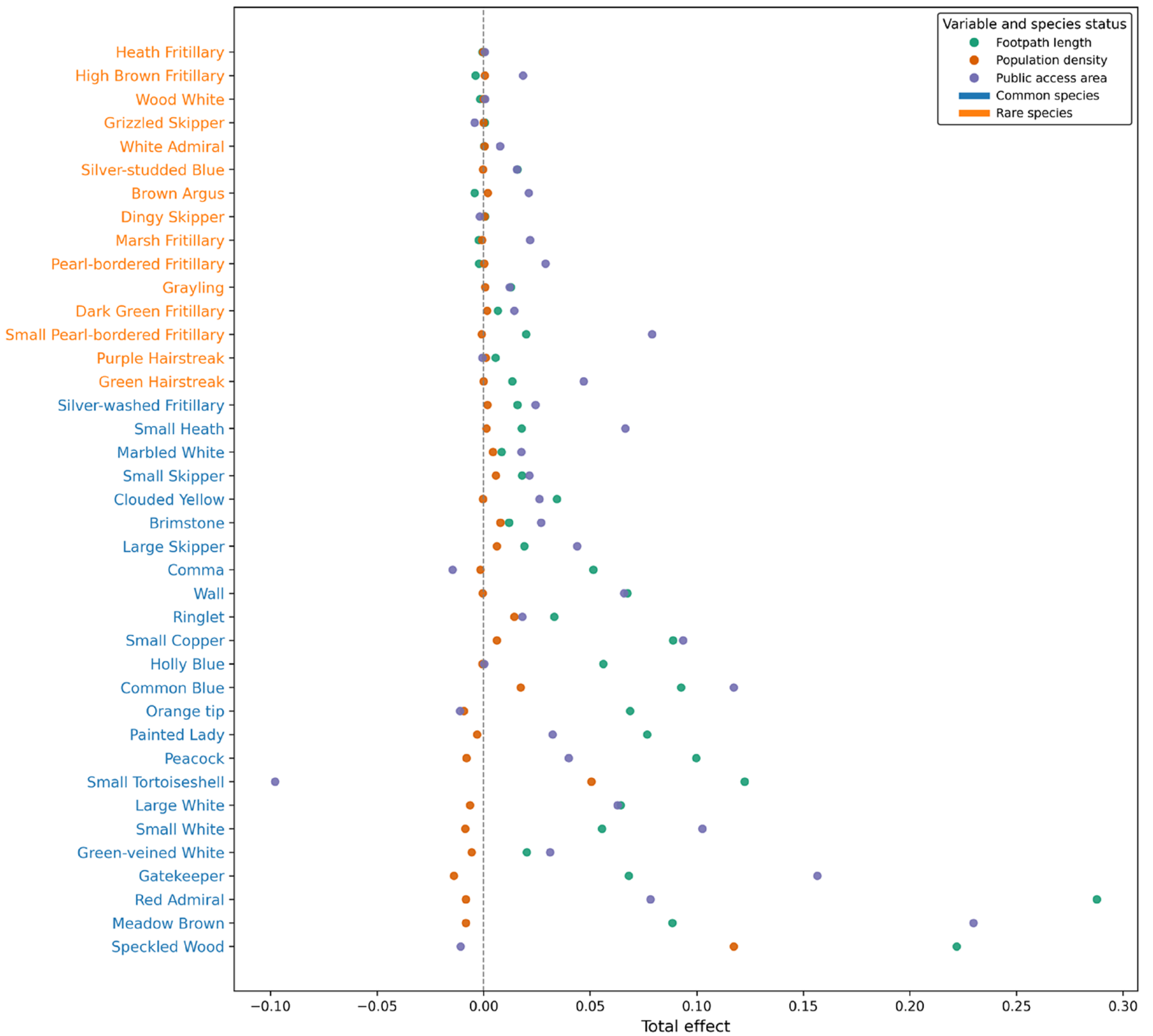

FIGURE 5 | Total effect of footpath length, human population density and public access area on the number of records of each butterfly species.

are more sparsely populated. However, this relationship is likely to be region specific and will not be applicable everywhere, so its inclusion should be evaluated according to regional geography.

The DAG enabled us to identify key common causes, such as Distance_from_coast, that influence multiple variables. Thus, incorporating causal reasoning into sampling design involving multiple geographical or environmental indicators can effectively help identify hidden variables. Although it is impossible to identify all relevant factors with complete certainty, the introduction of a causal inference framework can substantially reduce bias and enhance the robustness of ecological analyses based on citizen science data (Tennant et al. 2021).

This study has several limitations, which also provide directions for future research. First, because excessive zero inflation prevented the use of bootstrap resampling within the piecewise path analysis framework, we were unable to obtain confidence intervals for the estimated causal effects. The total effects were therefore derived analytically, following standard principles consistent with conventional approaches. Second, this study did not incorporate zero inflated negative binomial models (ZINB). Although ZINB is a preferred method for handling zero-inflated count data (Cooper et al. 2024; Darst et al. 2024), the current piecewise pathway analysis framework cannot directly incorporate ZINB and therefore relies on generalised linear models (GLM) (Kerner et al. 2023). This limitation may reduce the precision of parameter estimates but will not bias the results. Third, this study focused primarily on the effects of public access variables and did not explicitly consider other variables in the DAG, such as species preference (Soroye et al. 2018; Goldstein et al. 2024) or incorporate phenological factors (Dennis et al. 2024; Mowbray et al. 2024). Fourth, when calculating the total effects, the use of absolute values followed by log10 transformation allows for comparison of the relative importance of public access and land cover, but it prevents direct

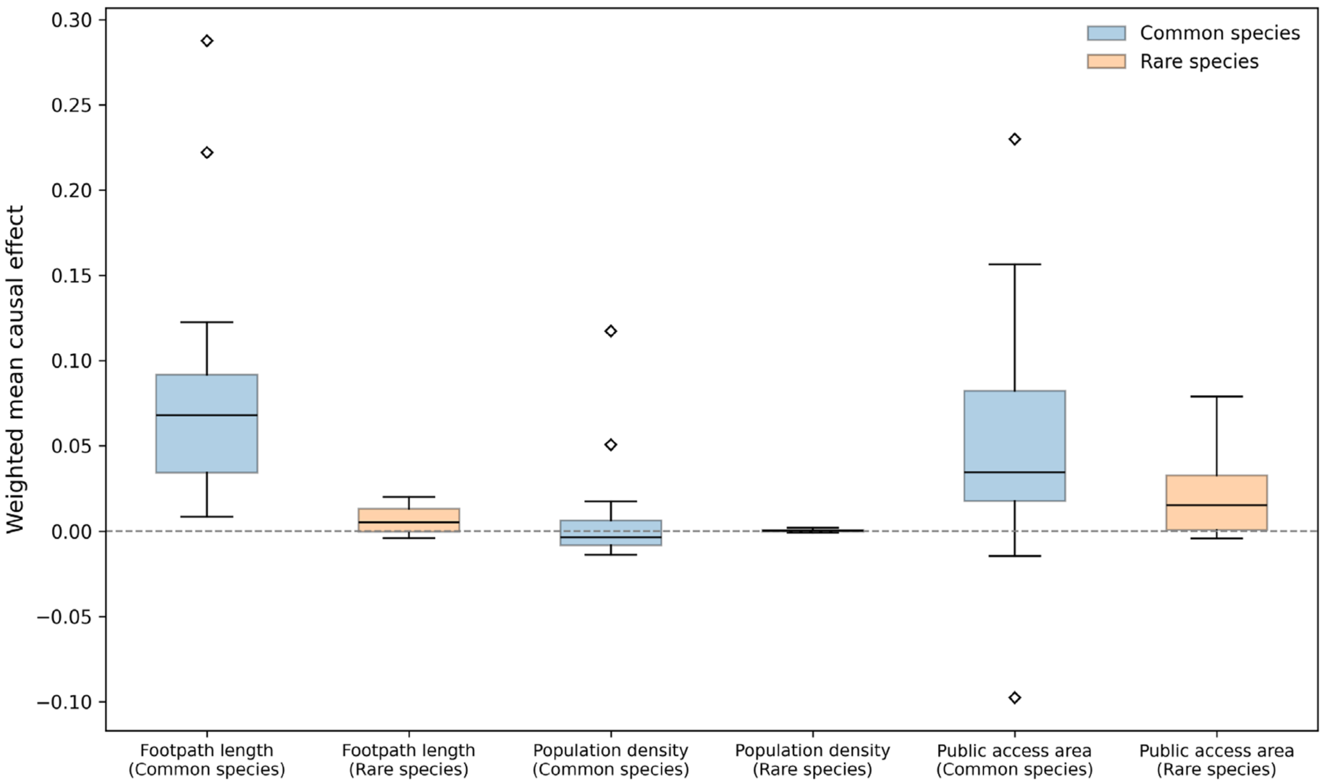

FIGURE 6 | Weighted mean and variation causal effects of each public access variable for common and rare butterfly species.

interpretation of the direction (positive or negative) of causal effects. This limitation is particularly relevant for rare species, for which the total effects may vary substantially in both magnitude and direction. Fifth, the causal relationships identified based on the DAG could not be evaluated at finer spatial resolutions due to data limitations and the use of coarser spatial resolutions may reduce the explanatory power of the identified variables. Lastly, our modelling approach was based on unstructured, opportunistic data, where observers had full discretion over when and what to record (Soroye et al. 2018). Consequently, the framework may not be directly applicable to structured citizen science schemes that employ systematic protocols with predetermined transects or recording locations, such as the UK Butterfly Monitoring Scheme (UKBMS). However, our approach could be used to address issues of nonresponse or site dropout in professional monitoring programmes when randomly selected survey locations fall on inaccessible land.

Future research could integrate the observation process and ecological processes within the existing framework more closely. In addition to public accessibility and land cover, recorder related factors should be explicitly incorporated into models (August et al. 2020), including variables such as preference for recording previously unseen species (Soroye et al. 2018; Goldstein et al. 2024), sampling completeness (Sánchez-Fernández et al. 2021), an emphasis on rare species (Habel et al. 2025), species apparency (Dennis et al. 2006) and observer skill level (Bowler et al. 2022). Moreover, future research should explore how to integrate DAG with piecewise pathway analysis, using ZINB or alternative models, to handle biodiversity databases characterised by a large proportion of zero values and to validate the robustness of causal inferences.

5 | Conclusions

Our study quantifies the relative causal effect of public accessibility on the number of butterfly records made by citizen scientists. We find that the strength of accessibility effects varied among taxa, with length of public rights of way emerging as the dominant driver of recording patterns. Higher numbers of records of common species largely reflect recorder behaviour in areas with dense footpath networks. Incorporating accessibility measures such as footpath length into both survey design and analysis can therefore improve data representativeness, while the persistent under sampling of private land and inaccessible habitats highlights the need for improved land access and targeted monitoring.

Author Contributions

M.L.: conceptualisation, methodology, formal analysis, visualisation, writing – original draft, and writing – review and editing. R.H.f.-C.: conceptualisation, methodology, supervision, and writing – review and editing. J.B.: conceptualisation, methodology, supervision, and writing – review and editing. R.J.B.: conceptualisation, methodology, and writing – review and editing. R.F.: writing – review and editing. D.B.R.: writing – review and editing. C.S.: data curation, resources, and writing – review and editing. All living authors read and approved the final version of the manuscript.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Paper2_datashare.zip at https://figshare.com/s/6f101b717a0c4135abff.

Peer Review

For transparency, the peer review documents associated with this article are available at https://doi.org/10.1111/ddi.70242.

References

Albalawi, B. 2025. "Urbanization and Habitat Loss: An Overview of Rapidly Growing Cities in Saudi Arabia." *Frontiers in Sustainable Cities* 7: 1636228.

Annessi, M., F. Marzialetti, M. Marconi, A. T. R. Acosta, and A. DI Giulio. 2025. "Butterfly Diversity in Relation to the CORINE Land Cover of the Culuccia Peninsula (Sardinia, Italy)." *Journal of Insect Conservation* 29: 17.

Arif, S., and M. A. Macneil. 2022. "Predictive Models Aren't for Causal Inference." *Ecology Letters* 25: 1741–1745.

August, T., R. Fox, D. B. Roy, and M. J. Pocock. 2020. "Data-Derived Metrics Describing the Behaviour of Field-Based Citizen Scientists Provide Insights for Project Design and Modelling Bias." *Scientific Reports* 10: 11009.

Bergerot, B., T. Merckx, H. Van Dyck, and M. Baguette. 2012. "Habitat Fragmentation Impacts Mobility in a Common and Widespread Woodland Butterfly: Do Sexes Respond Differently?" *BMC Ecology* 12: 5.

Bourhis, Y., A. E. Milne, C. R. Shortall, et al. 2025. "Trait Mediation Explains Decadal Distributional Shifts for a Wide Range of Insect Taxa." *Nature Communications* 16: 8131.

Bowler, D. E., N. Bhandari, L. Repke, et al. 2022. "Decision-Making of Citizen Scientists When Recording Species Observations." *Scientific Reports* 12: 11069.

Boyd, R. J., M. Botham, E. Dennis, et al. 2025. "Using Causal Diagrams and Superpopulation Models to Correct Geographic Biases in Biodiversity Monitoring Data." *Methods in Ecology and Evolution* 16: 332–344.

Boyd, R. J., M. Harvey, D. B. Roy, et al. 2023. "Causal Inference and Large-Scale Expert Validation Shed Light on the Drivers of SDM Accuracy and Variance." *Diversity and Distributions* 29: 774–784.

Broughton, R. K., R. Burkmar, M. Mccracken, et al. 2024. "UKCEH Land Cover Plus: Hedgerows 2016–2021 (England)." NERC EDS Environmental Information Data Centre.

Byrnes, J. E., and L. E. Dee. 2025. "Causal Inference With Observational Data and Unobserved Confounding Variables." *Ecology Letters* 28: e70023.

Chen, P.-Y., X. Dong, J. Durand, and H.-W. Yuan. 2025. "Impact of Emergent Macrophyte Mowing on an Aquatic Insect Community in Urban Ponds: A Case Study in an Artificial Wetland in Southeast Asia." *Wetlands* 45: 1–13.

Collins, C. M., H. Audusseau, C. Hassall, N. Keyghobadi, P. A. Sinu, and M. E. Saunders. 2024. "Insect Ecology and Conservation in Urban Areas: An Overview of Knowledge and Needs." *Insect Conservation and Diversity* 17: 169–181.

Cooper, J. E. J., K. E. Plummer, I. Middlebrook, and G. M. Siriwardena. 2024. "Using Butterfly Survey Data to Model Habitat Associations in Urban Developments." *Journal of Applied Ecology* 61: 773–783.

Darst, A. L., T. S. Mitchell, M. R. Verhoeven, et al. 2024. "Diversity of Bumble Bees and Butterflies in Minnesota Roadsides Depends on Floral Diversity and Abundance but Not Floral Native Status." *Insect Conservation and Diversity* 17: 690–701.

Dennis, E. B., B. J. Morgan, C. A. Harrower, N. A. Bourn, and R. Fox. 2024. "Incorporating Phenology to Estimate Species-Population Trends From Snapshot Citizen-Science Data." *Journal of Agricultural, Biological, and Environmental Statistics* 31: 1–17.

Dennis, R., T. Shreeve, N. Isaac, et al. 2006. "The Effects of Visual Apparency on Bias in Butterfly Recording and Monitoring." *Biological Conservation* 128: 486–492.

Dennis, R., and C. Thomas. 2000. "Bias in Butterfly Distribution Maps: The Influence of Hot Spots and Recorder's Home Range." *Journal of Insect Conservation* 4: 73–77.

Dimson, M., and T. W. Gillespie. 2023. "Who, Where, When: Observer Behavior Influences Spatial and Temporal Patterns of iNaturalist Participation." *Applied Geography* 153: 102916.

Engelhardt, E. K., M. F. Biber, M. Dolek, et al. 2022. "Consistent Signals of a Warming Climate in Occupancy Changes of Three Insect Taxa Over 40 Years in Central Europe." *Global Change Biology* 28: 3998–4012.

Gatis, N., D. Carless, D. Luscombe, R. Brazier, and K. Anderson. 2022. "An Operational Land Cover and Land Cover Change Toolbox: Processing Open-Source Data With Open-Source Software." *Ecological Solutions and Evidence* 3: e12162.

Geurts, E. M., J. D. Reynolds, and B. M. Starzomski. 2023. "Not All Who Wander Are Lost: Trail Bias in Community Science." *PLoS One* 18: e0287150.

Goldstein, B. R., and S. Stoudt. 2025. "Evidence of Novelty and Specialization Behavior in Participatory Science Reporting." *Oikos* 2025: e10938.

Goldstein, B. R., S. Stoudt, J. M. Lewthwaite, V. Shirey, E. Mendoza, and L. M. Guzman. 2024. "Logistical and Preference Bias in Participatory Science Butterfly Data." *Frontiers in Ecology and the Environment* 22: e2783.

Grace, J. B. 2006. *Structural Equation Modeling and Natural Systems*. Cambridge University Press.

Green, R. E., A. A. Dayer, and A. E. Johnson. 2023. "Citizen Scientist Participation in Research on Private Lands Positively Impacts Multiple Conservation Behaviors." *Citizen Science: Theory and Practice* 8: 53.

Guetté, A., S. Caillault, J. Pithon, et al. 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.

Guilherme, F., J. R. Vicente, M. A. Carretero, and P. FARINHA-Marques. 2024. "Mapping Multigroup Responses to Land Cover Legacy for Urban Biodiversity Conservation." *Biological Conservation* 291: 110508.

Guzman, L. M., E. Elle, L. A. Morandin, et al. 2024. "Impact of Pesticide Use on Wild Bee Distributions Across the United States." *Nature Sustainability* 7: 1324–1334.

Habel, J. C., T. Schmitt, P. Huemer, J. Rüdisser, P. Gros, and W. Ulrich. 2025. "Selective Observation Causes Differences in Citizen Science Butterfly Data." *Basic and Applied Ecology* 87: 46–54.

Hailay Gebremariam, G. 2024. "A Systematic Review of Insect Decline and Discovery: Trends, Drivers, and Conservation Strategies Over the Past Two Decades." *Psyche: A Journal of Entomology* 2024: 5998962.

Han, D., C. Wang, J. She, Z. Sun, and L. Yin. 2025. "Effects of Vegetation Heterogeneity on Butterfly Diversity in Urban Parks: Applying the Patch-Matrix Framework at Fine Scales." *Sustainability* 17: 6289.

Hernán, M. A., and J. M. Robins. 2010. *Causal Inference*. CRC.

Kerner, J. M., J. Krauss, F. Maihoff, L. Bofinger, and A. Classen. 2023. "Alpine Butterflies Want to Fly High: Species and Communities Shift Upwards Faster Than Their Host Plants." *Ecology* 104: e3848.

King, K., H. Tråsavik, and S. Sareen. 2024. "Take a Hike: Spatializing Allemannsretten and Transportation Accessibility for Outdoor Recreation in the Greater Stavanger Region, Norway." *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography* 78: 324–337.

Kirilenko, A. 2024. "Biodiversity of Accessible Greenspace for Vulnerable Population Groups: Citizen Science Data Analysis." *Sustainability* 16: 10492.

Kühn, E., A. Harpke, T. Schmitt, J. Settele, and I. Kühn. 2024. "Counting Butterflies: Are Old-Fashioned Ways of Recording Data Obsolete?" *Journal of Insect Conservation* 28: 577–588.

Lee, K. J., J. B. Carlin, J. A. Simpson, and M. Moreno-Betancur. 2023. "Assumptions and Analysis Planning in Studies With Missing Data in Multiple Variables: Moving Beyond the MCAR/MAR/MNAR Classification." *International Journal of Epidemiology* 52: 1268–1275.

Lefcheck, J. S. 2016. "piecewiseSEM: Piecewise Structural Equation Modelling in r for Ecology, Evolution, and Systematics." *Methods in Ecology and Evolution* 7: 573–579.

Li, M., R. J. Boyd, C. Smith, et al. 2025. "Citizen Scientists as Butterfly Predators: Using Foraging Theory to Understand Individual Recorder Behaviour." *Ecological Modelling* 510: 111344.

Mair, L., and A. Ruete. 2016. "Explaining Spatial Variation in the Recording Effort of Citizen Science Data Across Multiple Taxa." *PLoS One* 11: e0147796.

Mandeville, C. P., E. B. Nilsen, and A. G. Finstad. 2022. "Spatial Distribution of Biodiversity Citizen Science in a Natural Area Depends on Area Accessibility and Differs From Other Recreational Area Use." *Ecological Solutions and Evidence* 3: e12185.

Merckx, T., and H. Van Dyck. 2019. "Urbanization-Driven Homogenization Is More Pronounced and Happens at Wider Spatial Scales in Nocturnal and Mobile Flying Insects." *Global Ecology and Biogeography* 28: 1440–1455.

Mingarro, M., J. P. Cancela, A. Burón-Ugarte, et al. 2021. "Butterfly Communities Track Climatic Variation Over Space but Not Time in the Iberian Peninsula." *Insect Conservation and Diversity* 14: 647–660.

Mondain-Monval, T., M. Pocock, S. Rolph, T. August, E. Wright, and S. Jarvis. 2024. "Adaptive Sampling by Citizen Scientists Improves Species Distribution Model Performance: A Simulation Study." *Methods in Ecology and Evolution* 15: 1206–1220.

Morton, R. D., C. G. Marston, A. W. Neil, and C. S. Rowland. 2024. "Land Cover Map 2023 (10m Classified Pixels, GB)." NERC EDS Environmental Information Data Centre.

Mowbray, S., J. Bennie, M. W. Rhodes, D. A. Smith, and R. H. Ffrench-Constant. 2024. "Eyespot Variation and Field Temperature in the Meadow Brown Butterfly." *Ecology and Evolution* 14: e10842.

Natural England. 2025. *CRoW Act 2000 Access Layer*. Natural England.

Outhwaite, C. L., R. D. Gregory, R. E. Chandler, B. Collen, and N. J. B. Isaac. 2020. "Complex Long-Term Biodiversity Change Among Invertebrates, Bryophytes and Lichens." *Nature Ecology & Evolution* 4: 384–392.

Pearl, J. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82: 669–688.

Pearl, J. 2009. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3: 96–146.

Powney, G. D., C. Carvell, M. Edwards, et al. 2019. "Widespread Losses of Pollinating Insects in Britain." *Nature Communications* 10: 1018.

Priority Habitat Inventory (England). 2025. *Natural England*. Priority Habitat Inventory.

QGIS.ORG. 2025. *QGIS Geographic Information System (Version 3.34)*. QGIS Association.

Robertson, H. G. 2010. "Public Access to Private Land for Walking: Environmental and Individual Responsibility as Rationale for Limiting the Right to Exclude." *Georgetown International Environmental Law Review* 23: 211.

Rödder, D., T. Schmitt, P. Gros, W. Ulrich, and J. C. Habel. 2021. "Climate Change Drives Mountain Butterflies Towards the Summits." *Scientific Reports* 11: 14382.

Romo, H., E. GARCÍA-Barros, and J. M. Lobo. 2006. "Identifying Recorder-Induced Geographic Bias in an Iberian Butterfly Database." *Ecography* 29: 873–885.

Sánchez-Fernández, D., R. Fox, R. L. Dennis, and J. M. Lobo. 2021. "How Complete Are Insect Inventories? An Assessment of the British Butterfly Database Highlighting the Influence of Dynamic Distribution Shifts on Sampling Completeness." *Biodiversity and Conservation* 30: 889–902.

Sanllorente, O., E. Blanco-Urdillo, A. Sánchez-Tójar, and J. Diego Ibáñez-Álamo. 2025. "A Systematic Review and Meta-Analysis on Urban Arthropod Diversity." *Insect Conservation and Diversity* 18: 447–464.

Shirey, V., N. Neupane, R. Guralnick, and L. Ries. 2024. "Rising Minimum Temperatures Contribute to 50-Years of Occupancy Decline Among Cold-Adapted Arctic and Boreal Butterflies in North America." *Global Change Biology* 30: e17205.

Sicacha-Parada, J., I. Steinsland, B. Cretois, and J. Borgelt. 2021. "Accounting for Spatial Varying Sampling Effort due to Accessibility in Citizen Science Data: A Case Study of Moose in Norway." *Spatial Statistics* 42: 100446.

Soroye, P., N. Ahmed, and J. T. Kerr. 2018. "Opportunistic Citizen Science Data Transform Understanding of Species Distributions, Phenology, and Diversity Gradients for Global Change Research." *Global Change Biology* 24: 5281–5291.

Soroye, P., T. Newbold, and J. Kerr. 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 Gazette* 73: 64.

Suggitt, A. J., C. J. Wheatley, P. Aucott, et al. 2023. "Linking Climate Warming and Land Conversion to Species-Range Changes Across Great Britain." *Nature Communications* 14: 6759.

Tennant, P. W., E. J. Murray, K. F. Arnold, et al. 2021. "Use of Directed Acyclic Graphs (DAGs) to Identify Confounders in Applied Health Research: Review and Recommendations." *International Journal of Epidemiology* 50: 620–632.

Thoemmes, F., and N. Rose. 2014. "A Cautious Note on Auxiliary Variables That Can Increase Bias in Missing Data Problems." *Multivariate Behavioral Research* 49: 443–459.

Tiago, P., A. CEIA-Hasse, T. A. Marques, C. Capinha, and H. M. Pereira. 2017. "Spatial Distribution of Citizen Science Casuistic Observations for Different Taxonomic Groups." *Scientific Reports* 7: 12832.

Vaz, S., S. Manes, G. Khattar, et al. 2023. "Global Meta-Analysis of Urbanization Stressors on Insect Abundance, Richness, and Traits." *Science of the Total Environment* 903: 165967.

Wix, N., M. Reich, and F. Schaarschmidt. 2019. "Butterfly Richness and Abundance in Flower Strips and Field Margins: The Role of Local Habitat Quality and Landscape Context." *Heliyon* 5: e01636.

Wurz, A., I. Grass, D. C. Lees, et al. 2022. "Land-Use Change Differentially Affects Endemic, Forest and Open-Land Butterflies in Madagascar." *Insect Conservation and Diversity* 15: 606–620.

Zhai, J., L. Pu, L. Qie, et al. 2025. "Changes of Land Use and Landscape Pattern Along Sea-Land Gradient in Developed Coastal Region: A Case Study of Jiangsu Province, China." *Ecological Indicators* 176: 113635.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** Supporting Information.