

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK, 1970–2015

Charlotte L. Outhwaite *et al.*<sup>#</sup>

Here, we determine annual estimates of occupancy and species trends for 5,293 UK bryophytes, lichens, and invertebrates, providing national scale information on UK biodiversity change for 31 taxonomic groups for the time period 1970 to 2015. The dataset was produced through the application of a Bayesian occupancy modelling framework to species occurrence records supplied by 29 national recording schemes or societies ( $n = 24,118,549$  records). In the UK, annual measures of species status from fine scale data (e.g.  $1 \times 1$  km) had previously been limited to a few taxa for which structured monitoring data are available, mainly birds, butterflies, bats and a subset of moth species. By using an occupancy modelling framework designed for use with relatively low recording intensity data, we have been able to estimate species trends and generate annual estimates of occupancy for taxa where annual trend estimates and status were previously limited or unknown at this scale. These data broaden our knowledge of UK biodiversity and can be used to investigate variation in and drivers of biodiversity change.

### Background & Summary

Knowledge on the status and trends of biodiversity is essential for the conservation of threatened species and for the monitoring of progress towards biodiversity targets<sup>1</sup>. To date, UK scale analysis of annual biodiversity status has been restricted to well-studied taxa such as birds<sup>2</sup>, butterflies<sup>3</sup>, bats<sup>4</sup>, Odonata<sup>5</sup>, some moths<sup>6</sup> and a subset of “priority” species<sup>7</sup>. As a result, there are many taxa for which only coarse-scale measures of change are available with most invertebrate groups being a major gap due to a lack of abundance data. However, species occurrence records are fine-grained data available for many taxa and offer an alternative data source that can be used for estimating annual measures of biodiversity change.

Occurrence records are presence-only data documenting observations of species at known dates and locations. Within the UK, vast amounts of such occurrence data, known as biological records, are collected by volunteers and collated by recording schemes and societies and have been used extensively to produce species atlases and assess species range shifts<sup>8</sup>. These data offer greater taxonomic breadth than structured abundance data and have rarely been used to detect long-term change<sup>9–13</sup>. This limited use is due partly to the unstructured collection process which results in at least four forms of bias: the uneven detectability of species across space and time, uneven sampling effort per visit, uneven spatial coverage and uneven recording intensity over time<sup>14,15</sup>. These biases present challenges when estimating temporal change, however, methodological techniques have been developed that attempt to account for some forms of bias. One technique that has been increasingly used for the analysis of occurrence records is occupancy modelling<sup>7,16–19</sup>.

Occupancy models incorporate the data collection process to account for imperfect detection<sup>17,20,21</sup>. When compared with other methods developed for the estimation of trends from occurrence records, occupancy models have been shown to be the most capable of addressing associated biases, if the detection process is appropriately specified<sup>21</sup>. However, their use has been limited to taxa that have a high recording intensity including birds<sup>22</sup>, dragonflies<sup>5,19</sup> and butterflies<sup>17,23</sup>. Outhwaite *et al.* extended a previous Bayesian occupancy modelling framework, to increase the precision of occupancy estimates via the use of a random walk prior on the year effect of the state model<sup>24</sup>. This formulation allows information to be shared between years in a natural way and facilitates the application of such models to datasets of a low recording intensity that were not previously considered for practical

<sup>#</sup>A full list of authors and their affiliations appears at the end of the paper. \*email: [charlotte.outhwaite.14@ucl.ac.uk](mailto:charlotte.outhwaite.14@ucl.ac.uk)

use. Therefore, through the application of a modelling framework based on that of Outhwaite *et al.*<sup>24</sup> we have produced a 45-year dataset of annual occupancy estimates for 5,293 UK bryophyte, lichen and invertebrate species.

This dataset presents a long-term measure of change in species occupancy at a national (UK and GB) and nation-specific scale (England, Scotland, Wales and Northern Ireland) using fine grained ( $1 \times 1$  km) data. This represents new information for the taxa covered by this study. By providing the outputs of this analysis we hope to promote research into UK biodiversity change, particularly for those taxonomic groups that have received less attention in the context of national-scale trends at fine scales. It is hoped that these data will provide the basis of future aggregate measures of UK biodiversity change and enable the investigation of drivers of change.

## Methods

The raw data underpinning these models were occurrence records collated from 29 UK or Great Britain (GB) based recording schemes and societies, with additional data from the Biological Records Centre, Wallingford, and the iRecord database (<https://www.brc.ac.uk/irecord/>). These data were standardised to ensure all datasets met the required criteria (see methods) and had undergone review by experts of each species group. The standardised data were then organised into detection histories to enable their use within an occupancy modelling framework. Our Bayesian occupancy model was fitted for each species and provided annual estimates of occupancy for each country analysed. From these estimates, growth rates of species occupancy were calculated. The full workflow is described in Fig. 1 with the following sections describing each part of the workflow. The outputs from this study include 1000 samples from the posterior distribution of the annual occupancy estimates for each country analysed, large-scale trend estimates for each species in the form of annual growth rates as well as additional metadata.

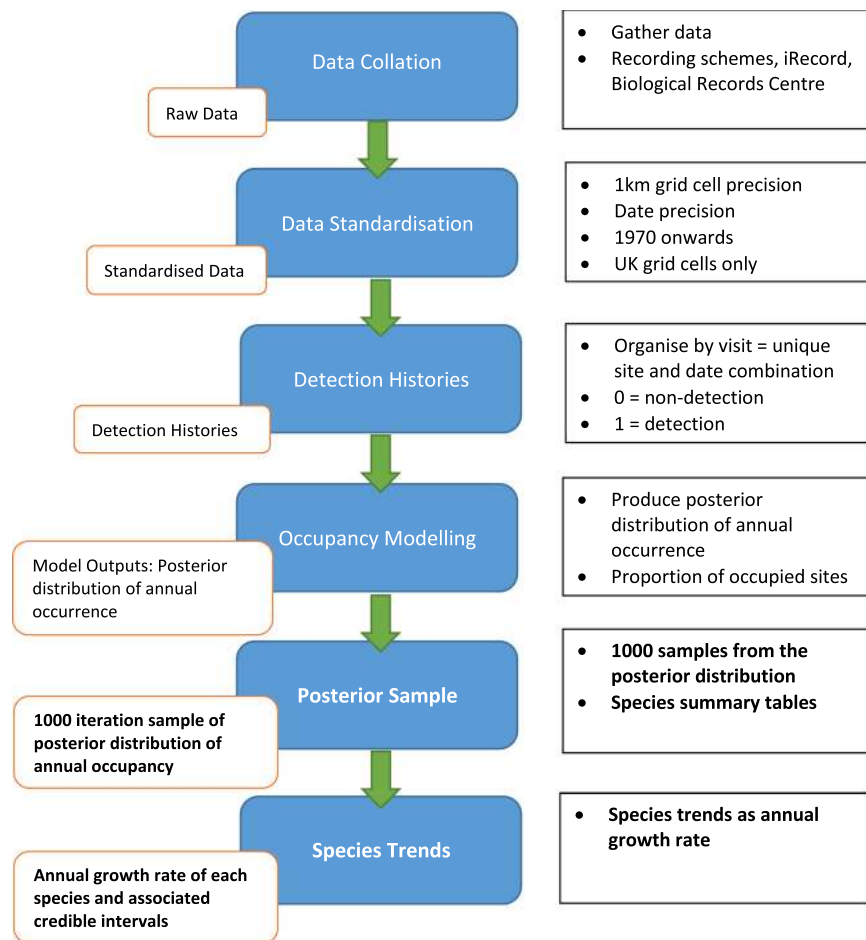
**Data collation.** Data were collated from recording schemes and societies that support recording networks and collect and verify occurrence records on UK species. Twenty-nine schemes granted the use of their data for this analysis. For some taxonomic groups, mainly where the number of records provided by the scheme was low, further data were acquired from the databases of the Biological Records Centre and from the iRecord system. iRecord is a website and associated mobile phone application that has been designed to support recording schemes and societies in the collation, management, quality assurance and sharing of wildlife observations (<https://www.brc.ac.uk/irecord/>). Only records that had been marked as “accepted” by an iRecord verifier were used<sup>25</sup>. These sources can be considered different routes for accessing the same form of data. Some of the raw data are available through the NBN Atlas (<https://nbnatlas.org/>, Online-only Table 1), although in some cases the publicly-available data is at a coarse spatial resolution or limited in time compared to that used here. For some schemes, the data had to be requested and supplied directly by the BRC or the scheme. Users desiring access to these data should first check the availability via the NBN Atlas and/or contact the scheme directly (contact information can be found on the relevant BRC scheme pages (<https://www.brc.ac.uk/recording-schemes>)).

The occurrence records used in this analysis are presence-only data of a species and consist of a what, when and where: what species was observed, when it was observed and where it was observed. In most cases, one scheme is associated with one taxonomic group. In the case of data from the Bees, Wasps and Ants Recording Society, the dataset was split into three separate datasets for analysis, one for each taxonomic group, as these taxa are not considered to be recorded as a complete entity by all members of the society. The data collation step resulted in the generation of 31 “raw” datasets, one for each taxonomic group assessed.

**Data standardisation.** The resolution of record location and date vary in these data, particularly for data from earlier decades. Record location is represented by a British or Irish grid reference, but the resolution differed between records (e.g. 10 m, 100 m, 1 km, 2 km, 10 km). The date format of a record also varied or was unknown. The most precise records state on which day a species was recorded, but some older records state a year or range of years. These differences mean that the datasets needed to be standardised to ensure a collection of records with the same level of spatial and temporal precision. For our model, records are required with day level precision to maximise the number of replicates within a year from independent visits. Replicate visits within a closure period (here one year) are essential for estimating species detectability within the occupancy model<sup>26</sup>. Records where the date of the record was unknown were removed from the dataset. A  $1 \times 1$  km grid cell precision for location was chosen as this would provide the greatest number of spatial replicates across the time period of interest for most taxa. Any records with a more precise location were scaled up to 1 km resolution. Any records with a less precise location were removed from the dataset. Only records from 1970 onwards were retained within the datasets as, in general, the number of records at a  $1 \times 1$  km precision before 1970 was low for most taxa. A check was also carried out to ensure that only records from grid cells within the UK (England, Northern Ireland, Scotland and Wales) were retained, excluding data from the Channel Islands, the Republic of Ireland and the Isle of Man.

The supplied species names for each taxonomic group were also checked. Any records that were made to a taxonomic level higher than species were excluded from the dataset (but see below regarding species aggregates). Scheme organisers were contacted to aid in the checking of the species lists to ensure that synonyms representing the same species were not used in isolation. Spelling mistakes were also identified and corrected where necessary. In specific cases, certain species were modelled as aggregates of species. This was carried out upon advice from the scheme organisers and was due to changes in taxonomy during the period of interest resulting in records where species identity could not be certain, difficulties in identification of species by recorders, or differences in what people actually record under a specific name. After these checks were carried out, any duplicate records were then removed from the dataset.

The standardisation process resulted in 31 datasets covering 10,750 species and 24,118,549 individual species records (Online-only Table 1). Note that the coverage of countries within the UK varies between schemes. Some



**Fig. 1** The workflow to produce the datasets presented in this study. Major steps are shown in the blue boxes with the data type generated within the orange boxes. Associated details are given next to each step. Datasets provided alongside this paper are shown in bold.

groups, therefore, were only analysed at the scale of Great Britain (GB, includes England, Scotland and Wales) rather than at the UK scale (includes England, Scotland, Wales and Northern Ireland). The spatial coverage of the records within each of these standardised datasets can be seen in the maps provided in Supplementary Fig. S1.

**Organisation of detection histories.** The standardised data were organised into detection histories as first applied to presence-only data by Kéry *et al.*<sup>16</sup>. In this format the data were reorganised into visits (unique combinations of 1 × 1km grid cell and date), with 0s or 1s assigned to denote whether a species was detected (1) or not detected (0) during each visit. This step enables the use of presence-only data within a framework that requires information on non-detection. Detections were extracted directly from the data as a record of a species at a known date and location. There are two types of non-detection: true absences and false absences: where the species has been overlooked. Non-detections were inferred from the detection of other species within that taxonomic group when the focal species was not observed. For example, if ant species A was not detected during a visit, but ant species B, C and D were detected then this would be classified as a non-detection of ant species A. We assume that species A was available phenologically to be detected at the same time as other species. It is possible to include phenological information into the detection submodel<sup>27</sup>: our preliminary investigations showed that including phenology made little difference to the resulting estimates but dramatically slowed the convergence time, so we report results without terms for phenology. Our long-term trend estimates therefore assume that the distribution of recording effort throughout the year has remained approximately constant over time. List length is the number of species recorded during a visit, this parameter is used in the occupancy modelling framework as a proxy for sampling effort<sup>28</sup> (see later section). The detection history dataset and the list length of each visit were fed into the occupancy modelling framework for the analysis.

The organisation of the standardised data into detection histories was carried out using the function “*formatOccData*” in the R package *sparta*<sup>29</sup>. *sparta* is an R package that contains various methods for the analysis of unstructured occurrence records and is freely available on GitHub (<https://github.com/BiologicalRecordsCentre/sparta>). An example of how to use this function to generate model-ready datasets can be found in Supplementary File 1.

**The occupancy model.** The occupancy model used here is based on the “random walk” model of Outhwaite *et al.*<sup>24</sup>. The name refers to the use of a random walk prior on the year effect of the state model, which was found to improve precision in occurrence estimates, particularly for datasets with a low recording intensity (most of the input datasets in this study). This development has enabled the much broader application of occupancy modelling than was previously possible.

The model is a hierarchical model split into two distinct sub models: the state model and the observation model. The state model describes the true occupancy state,  $z_{it}$ , of site  $i$  in year  $t$  and is defined by Eqs (1) and (2).  $z_{it}$  will be 1 when a site is occupied and 0 if the site is not occupied.  $z_{it}$  takes a Bernoulli distribution:

$$z_{it} \sim \text{Bernoulli}(\psi_{it}). \quad (1)$$

where the logit of the probability of occurrence,  $\psi_{it}$ , varies with year and site:

$$\text{logit}(\psi_{it}) = \log\left(\frac{\psi_{it}}{1 - \psi_{it}}\right) = b_t + u_i, \quad (2)$$

$b_t$  and  $u_i$  denote year and site effects respectively.

For the model used here, the state model year effect was split into four regions to allow the estimation of occupancy for each country within the UK, as well as the aggregate occupancy at UK and/or GB level. This means that, instead of having a single year effect in the state model as shown in Eq. 2, there is a year effect associated with each country, hereafter termed region. Specifically, let  $r(i)$  be the region (England, Northern Ireland, Scotland or Wales) in which site  $i$  is located, then:

$$\text{logit}(\psi_{it}) = \log\left(\frac{\psi_{it}}{1 - \psi_{it}}\right) = b_{r(i)} + u_i, \quad (3)$$

where  $b_{r(i)}$  is the year effect for year  $t$  in region  $r$  in which site  $i$  is found.

The observation sub model describes the data collection process and is conditional on the true occupancy state  $z_{it}$ .  $p_{itv}$  represents the probability that a species will be observed on a single visit, given the species is present at that site. The observation,  $y_{itv}$ , is then described as being drawn from a Bernoulli distribution conditional on the true occupancy state:

$$y_{itv} | z_{it} \sim \text{Bernoulli}(p_{itv} \cdot z_{it}) \quad (4)$$

This means that a species can only be detected at a given site if it is truly present. We therefore assume that there are no false positive observations (for example incorrect species identifications) within the dataset. Given that our occurrence records are curated and verified by recording schemes and their organisers, this is likely to be a reasonable assumption. However, our long-term trend estimates will be biased if there is a directional trend in the rate of misidentification. A model extension has been developed that can deal with false positives<sup>30</sup>, but it has a small effect on overall occupancy.

Variation in detection probabilities  $p_{itv}$ , per visit are described as follows by Outhwaite *et al.*<sup>24</sup>:

$$\text{logit}(p_{itv}) = \log\left(\frac{p_{itv}}{1 - p_{itv}}\right) = a_t + c \log L_{itv}, \quad (5)$$

where  $a_t$  is a year effect and  $L_{itv}$  is the list length, that is the number of species recorded during a single visit. In this form,  $c$  represents the change in the detectability of the focal species as the list length increases. In using this formulation, the assumption is that there is likely to be a positive relationship between the number of species recorded on a visit and the probability of a species being detected. The suggestion being that more time was spent looking and so greater sampling effort expended. However, Eq. (5) imposes a specific mathematical form on the relationship between list length and species detectability, and this form may not be justified for all the species considered here. This continuous option is also likely to result in higher assumed detection in the south due to a general higher species richness than occurs in the north. Therefore, rather than using a continuous specification of list length we have chosen to use a categorical specification in which detectability is classified according to whether a species is recorded on a list of length 1, 2–3 or 4+ records. This alternative classification of list length was considered by Van Strien *et al.*<sup>17</sup> as a more flexible alternative to the continuous specification where detectability does not follow an increase with list length. It also does not assume that each list was a complete list of species recorded during that visit. As we were looking to apply this method across many thousands of species, a single option applied broadly to all groups was used, although we recognise that this may be less suitable for the few high richness groups considered in this study. In the model implemented here, Eq. (5) is replaced with the following:

$$\text{logit}(p_{itv}) = \log\left(\frac{p_{itv}}{1 - p_{itv}}\right) = a_t + \beta_1 * \text{datatype2}_{itv} + \beta_2 * \text{datatype3}_{itv}, \quad (6)$$

where  $\beta_1$  and  $\beta_2$  estimate differences in  $\text{logit}(p_{itv})$  for a list length of 2–3 (datatype2) and of 4+ (datatype3) respectively, relative to a list length of one.

This model is run in a Bayesian framework which requires unknown parameters to be assigned a prior distribution. The prior distribution describes our knowledge of the system before the data were collected. In the model

formulation of Outhwaite *et al.*<sup>24</sup> vague, uninformative priors are set on all parameters except for the year effect of the state model. The preferred prior on this parameter uses a random walk to describe the change in occurrence as similar to that of the previous year with some variation. Here, we apply this to the year effect for year  $t$  in region  $r$ ,  $b_{tr}$ :

$$b_{tr} \sim \begin{cases} \text{Normal}(\mu_{br}, 10^4) & \text{for } t = 1 \\ \text{Normal}(b_{t-1r}, \sigma_{br}^2) & \text{for } t > 1 \end{cases} \quad (7)$$

$$\text{where, } \mu_{br} \sim \text{Normal}(0, 100), \quad (8)$$

$$\text{and } \sigma_{br} \sim |\text{Student-}t \text{ on 1 degree of freedom}| \quad (9)$$

See Outhwaite *et al.*<sup>24</sup> for further details on the random walk prior.

Priors on all other parameters are set out as in the original paper including the use of the recommended half-Cauchy hyperpriors. These are set as shown in Eq. 9 as the modulus of a Student's  $t$ -distribution on 1 degree of freedom. Information on the setting of initial values can be found in the original paper where the procedure outlined was followed.

The models were fitted using the function “*occDetFunc*” from the R package *sparta*, selecting the random walk model with half-Cauchy hyperpriors and using the categorical specification of list length<sup>29</sup>. Parameters set for the model running process included  $nyr = 2$ , this means that any sites with fewer than two years of data are dropped from the dataset. Models were fitted to data for the period 1970 to 2015.

The *sparta* package uses a Markov Chain Monte Carlo (MCMC) algorithm to fit the models, using JAGS<sup>31</sup> via the function *occDetFunc*. This process can be computationally expensive, particularly when datasets consist of a large number of records and/or a large number of species. For small to medium datasets, models were fitted using a computer cluster hosted at CEH, Wallingford. Using this process, species were run in parallel across multiple cores. Large datasets, including the moths, dragonflies, bryophytes and lichens were run on the much larger NERC JASMIN supercomputer. For groups run on the CEH cluster, the MCMC algorithm was run for 40,000 iterations per species with a burn in of 20,000 and a thinning rate of three. This was sufficient to obtain convergence for most of the parameters of interest for most species. Convergence was assessed using the Rhat value<sup>32</sup>, where a value below 1.1 is considered sufficient<sup>33</sup>. For those groups run on JASMIN, the greater size of the datasets meant that these groups took longer to run than those run on the CEH cluster. These groups were run for 20,000 iterations in total with a burn-in of 10,000 and a thinning rate of three. As these groups generally had more data per species, convergence was reached in fewer iterations, so this was considered an acceptable compromise to reduce the overall run time. For a general idea of run time, small data sets with few species take just a few hours when run in parallel on these systems, but large datasets with many species took several weeks. An example of how to run an occupancy model using the *occDetFunc* function in the *sparta* package can be found in Supplementary File 1.

We fitted the model described above to all 10,750 species within the standardised datasets, regardless of the number of records that were available for that species. Species occupancy each year was calculated as a derived parameter within the model as the proportion of occupied sites. This was calculated for each region covered by the model, therefore, estimates are available for each species for multiple regions depending on input data coverage (see Online-only Table 1). A posterior distribution of estimates for each year for each region was therefore generated by the MCMC process.

**Assessing species outputs.** As a combination of species rarity, the data standardisation process, and the implementation of the *nyr* parameter (see section on the occupancy model), the number of records available for each species varied considerably. As a result, model outputs for some species were based on very few records. As a model output based on just a few records cannot be considered to contain any valuable information, there was a requirement to set a threshold number of records that a species must have to be considered a part of the dataset described here. After model fitting, we therefore set a threshold of 50 records across the 45 year time period, increasing this threshold made very little difference to multispecies assessments (not presented here) so this value is maintained (see also Outhwaite *et al.*<sup>24</sup> for examples of species models that achieve useable results based on 50–500 records). Users can increase this threshold for their own use should that be considered appropriate: the number of records contributing to each species output has been provided alongside the data in the repository. Species were also removed if they contain a gap in the dataset where more than 10 consecutive years were lacking records, this was to prevent possible cases where the prior takes over during periods of no data (see supplementary of the original paper describing the random walk model<sup>24</sup>). This reduced the number of species from the 10,750 that models were fitted for to 5,293 that we consider to contain valuable information on species status. Considering that the number of records that contributed to the estimation of the occupancy and trend values, as well as the uncertainty around these estimates, is important, this information has also been provided within the data repository.

**The posterior distribution.** The output produced from the occupancy model is a posterior distribution of the occupancy parameter estimated as the proportion of occupied sites for each year, within each region, for each species. These estimates cover the years 1970 to 2015. Some input datasets ended prior to 2015 so estimates are produced for years where data are not available, although the uncertainty (which is quantified via the provided samples from the posterior distribution) will be greater during these years. To make the analysis of these outputs

manageable to users, we supply 1000 samples from the occupancy posterior distribution for each region as a part of this dataset.

**Species trends.** Long-term species trends were estimated as the percentage annual growth rate of occupancy using the following formula:

$$\text{annual growth rate} = \left( \left( \frac{f}{s} \right)^{\frac{1}{y}} - 1 \right) \times 100, \quad (10)$$

where,  $f$  was the occupancy in the final year,  $s$  was the occupancy in the starting year and  $y$  was the number of years. The growth rate was calculated for each of the 1000 samples which were then summarised using the mean and 95% quantiles. To avoid extrapolating beyond the scope of the data, the start and end years that were used to calculate the annual growth rate differed between species depending on which years had species observations contributing to the occupancy outputs. For example, if the input dataset for a species only had records of that species between 1974 and 2013, then the posteriors for these years were used as the final and starting years in the formula above. This was considered appropriate as when there are no data at the start of the time period, the prior of the model can have an influence on the result (see supplementary information by Outhwaite *et al.*<sup>24</sup> for further information). The start and end years per species are detailed alongside the trend estimates in the repository. The precision of the trend estimate is also supplied and is estimated as  $1/\text{variance}$  of the 1000 sample trends.

### Data Records

All outputs as a part of this dataset are freely available through the Natural Environment Research Council (NERC) Environmental Information Data Centre (EIDC) within the dataset entitled “Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK (1970–2015)”<sup>34</sup> and is freely available to download (<https://doi.org/10.5285/0ec7e549-57d4-4e2d-b2d3-2199e1578d84>).

Data presented within this dataset are in three forms:

1. **1000 samples of the posterior distribution** of the proportion of occupied sites for each region per species per year (one file per species).
2. Tables summarising the **mean occupancy per region and associated uncertainty** for each species (one file per species).
3. **Large-scale long-term species trends** derived from the posterior samples as the percentage annual growth rate (one file, row per species).

These data are accompanied by information on the input datasets used to generate these estimates (one file) and information on the origins and changes to species names (one file). All data files are provided in a .csv format.

**Samples of the posterior distribution of species occupancy.** The output produced from the model is a posterior distribution of the occupancy parameter estimated as the proportion of occupied sites for each year, for each region, for each species. These estimates cover the years 1970 to 2015 and encompass four regions for GB scale groups (GB, England, Scotland and Wales) and six regions for UK scale groups (UK, GB, England, Scotland, Wales and Northern Ireland). To make the analysis of these outputs manageable to users, we supply 1000 samples from the occupancy posterior distribution for each region. 1000 samples from the posterior were randomly selected for each species:year combination and across each region, these are supplied as a csv file for each species with a row per iteration within each region and a column per year (Table 1). These can be found in the “POSTERIOR\_SAMPLES” folder in the repository<sup>34</sup>. Values presented in the year columns of these tables represent the proportion of sites occupied by that species in that region and can be any value between zero and one.

**Model output summary tables.** Alongside the samples from the posterior distribution, we also supply the summary table from the model output for each species. This table includes the mean estimate for each year and the associated 95% credible intervals estimated from the complete posterior distribution (Table 2). It also includes the standard deviation and Rhat values for each estimate. The Rhat parameter estimates the convergence of the MCMC chains, a value of 1.1 is usually considered acceptable<sup>33</sup>. A summary table for each species is supplied as a csv file in the “SUMMARY\_TABLES” folder of the repository<sup>34</sup>. The numeric values in this table have been rounded to three decimal places.

**Species trends.** Species trends, calculated as the percentage annual growth rate are supplied alongside the associated credible intervals, the first and last years used to calculate these trends for each species, the number of years across which the trend estimate is calculated and the number of records of the species (Table 3). The precision of the estimate is also presented. These values are present in a single table in the “Species\_Trends.csv” file within the repository<sup>34</sup>. The numeric values in this file have been rounded to three decimal places.

**Accompanying metadata.** Another csv file details information on the input datasets used to generate the results shared (Table 4). This includes the number of records in each input dataset, the name of the recording scheme that provided data, the number of species covered by the input datasets and the number that are covered by the outputs supplied. This table also details the number of visits that meet each list length category specified within

Group	Species	Region	Iteration	1970	1971	1972	...	...	2015
Ants	Formica aquilonia	GB	1	0.01804	0.02296	0.02132			0.01886
Ants	Formica aquilonia	GB	2	0.00164	0.00205	0.00082			0.01599
Ants	Formica aquilonia	GB	3	0.02747	0.02706	0.02952			0.02501
		...							
		...							

**Table 1.** Example table showing the layout of the samples from the posterior distribution for a species. There is a row per iteration per region and a column per year. Additional columns detail the region, iteration, the species name and the taxonomic group that species belongs to. ‘...’ represents intervening years and regions not shown here.

Group	Species	Region	Year	Mean	Lower_ CI	Upper_ CI	Standard_ deviation	Rhat
Ants	Formica aquilonia	GB	1970	0.034	0	0.081	0.03	2.937
Ants	Formica aquilonia	GB	1971	0.034	0	0.08	0.03	3.015
Ants	Formica aquilonia	GB	1972	0.034	0	0.08	0.03	3.101
			...					
Ants	Formica aquilonia	GB	2015	0.039	0.006	0.082	0.029	2.833

**Table 2.** An example summary table showing the layout and parameters included. There is a row per year per region. Information columns detail the taxonomic grouping, species name, region and year of the estimates. The remaining columns detail the statistics for that estimate including mean occupancy, 95% credible intervals, the standard deviation and the rhat statistic. ‘...’ represents intervening years not shown here.

the model. This information is taken after the datasets have been standardised and has been supplied in the “Dataset\_Information.csv” file within the repository<sup>34</sup>. This file contains all information in Online-only Table 1.

Information on the origin of the species names used is detailed in the “Species\_Names.csv” file, also found within the repository<sup>34</sup>. This includes information on why a model was not fitted for a species, advice on aggregations from schemes and any other changes made to species names (Table 5).

### Technical Validation

The model used here is based on the “random walk model” tested by Outhwaite *et al.*<sup>24</sup>. The authors tested this model, and other variants, on both simulated data and real world occurrence records (the kind used to produce this dataset). They showed that the random walk model improved the precision of the occupancy estimates and had low bias when estimating known species trends from simulated data. This model is, therefore, arguably the most appropriate for use in this study, particularly due to its improved application to datasets of a low-recording intensity which several the input datasets included here suffer from.

The input datasets were checked and standardised as described in the methods section. Species names within each taxon group dataset were checked by scheme organisers or were compared to online checklists to ensure no synonyms were present alongside preferred species names. Note that all the schemes providing data to these analyses maintain taxon registers integrated into their databases to ensure the taxonomic coherence of all data held, and to ensure conformity with the currently accepted taxonomic standard of that scheme. Scheme organisers also recommended the removal or aggregation of species where it was not certain which species the records were referring to (details in Species\_Names.csv file<sup>34</sup>). This could occur, for example, when a single species is split into two separate species. Those species may then be aggregated under one species name if records before the split cannot be identified as one of the two split species. Aggregate species can be identified by *agg.* within the species name. Any species where record identity was questioned, for example because species are very difficult to identify with confidence, were highlighted. These were retained within the dataset to fully inform list length, but models were not fitted for these species. These checks ensure that all data relevant to a taxon can be extracted from scheme databases, even if occurrence records were originally collected against synonyms or at a lower (infraspecific) rank than that of the species. It should be noted that, across the different taxon-focused schemes, decisions regarding the suitability of particular types of species occurrence data for modelling will vary, and these decisions are captured within the dataset metadata where provided (details in Species\_Names.csv file<sup>34</sup>). As an example, the Bees, Wasps and Ants Recording Scheme did not consider data relating to species within the *Lasius niger* aggregate as suitable for modelling: taxonomic changes have meant that data collected at different time points under the name “*Lasius niger*” cover two distinct species, and the scheme did not consider a trend at the aggregate level of these concepts to be ecologically meaningful. To give a contrasting example, the British Bryological Society were happy for a trend to be produced for the moss taxonomic concept *Ulota crispa sensu lato* (*s.l.*), an aggregate covering the species *Ulota crispa sensu*<sup>35</sup> and *Ulota bruchii* (see ref.<sup>36</sup> for a similar analysis using this aggregate), because this was felt to be both ecologically meaningful, and to make the best use of historic data. Ultimately, species trends produced using species occurrence data must deal with the trade-off between the taxonomic uncertainty attached to any given record and producing the most meaningful assessment of change given the available data. The expert opinion of those who collect and curate such data is an essential accompaniment

Group	Species	N_Years	First_Year	Last_year	N_Records	Mean_growth_rate	Lower_CI	Upper_CI	Precision
Ants	Formica aquilonia	34	1981	2014	94	-1.038	-4.160	1.861	0.448
Ants	Formica cunicularia	46	1970	2015	785	-2.059	-4.033	-0.221	1.101

**Table 3.** Table showing the layout of the species trends csv file. Associated information is provided including taxonomic grouping, species name, the number of years of data, the first and last years used to estimate growth rate, and the number of records of this species contributing to the occupancy estimates. The growth rate and 95% credible intervals are then supplied along with the precision of the estimate.

Group	Country coverage	First_Year	Last_Year	Scheme_Name	N_Records	N_Species	N_Species Outputs	N_Sites	N_Visits	LL_1	LL_2-3	LL4+
Ants	GB	1970	2015	Bees, Wasps and Ants Recording Society	34790	60	29	8988	19530	11749	5906	1875
Aquatic Bugs	GB	1970	2015	Aquatic Heteroptera Recording Scheme	61264	92	51	10254	21014	8701	6697	5616
Bees												

**Table 4.** Example rows of the Dataset\_information.csv table.

Group	Species	Name_origin	In_final_dataset	Reason_not_included	Details
Ants	Formica aquilonia	Direct from scheme dataset	Yes	NA	
Ants	Formica cunicularia	Direct from scheme dataset	Yes	NA	
Ants	Formica exsecta	Direct from scheme dataset	No	Didn't meet criteria	
Ants	Formica fusca	Direct from scheme dataset	Yes	NA	
Ants	Formica lemani	Direct from scheme dataset	Yes	NA	

**Table 5.** Example of the information provided in the “Species\_Names” csv file. This table includes information on all 10,750 species included in the study, the origin of a species name and any information detailing name changes or species aggregations where available.

to automated checks. For reference, species names were also checked using the *taxize* R package<sup>37,38</sup>. The scores generated using the *gnr\_resolve* function to check names against the GBIF Backbone Taxonomy register can be found in Supplementary File 2.

The number of records that make up the input datasets for each taxonomic group differed substantially, see “Dataset\_Information.csv”<sup>34</sup> or Online-only Table 1. This will impact the spatial coverage and the number of sites that estimates are based on (maps of the spatial coverage of the standardised input datasets are available in Supplementary Fig. S1).

Convergence of the model parameters of occupancy was assessed using the Rhat statistic. This measure is commonly used to assess the convergence of a model parameter with values less than 1.1 generally considered to be adequate. Due to the size of the datasets and the time taken to run all the models it was not possible to run all models to complete convergence. Set numbers of iterations were therefore undertaken according to the size of the dataset: 40,000 for smaller datasets and 20,000 for larger datasets. We have supplied a summary table for each species that details the mean occupancy values, the standard deviation of the estimates, 95% quantiles of occupancy and the Rhat value so users can check convergence of estimates as well as the uncertainty associated with the mean occupancy estimates, these can be found in the species specific csv files in the “SUMMARY\_TABLES” folder of the repository<sup>34</sup>.

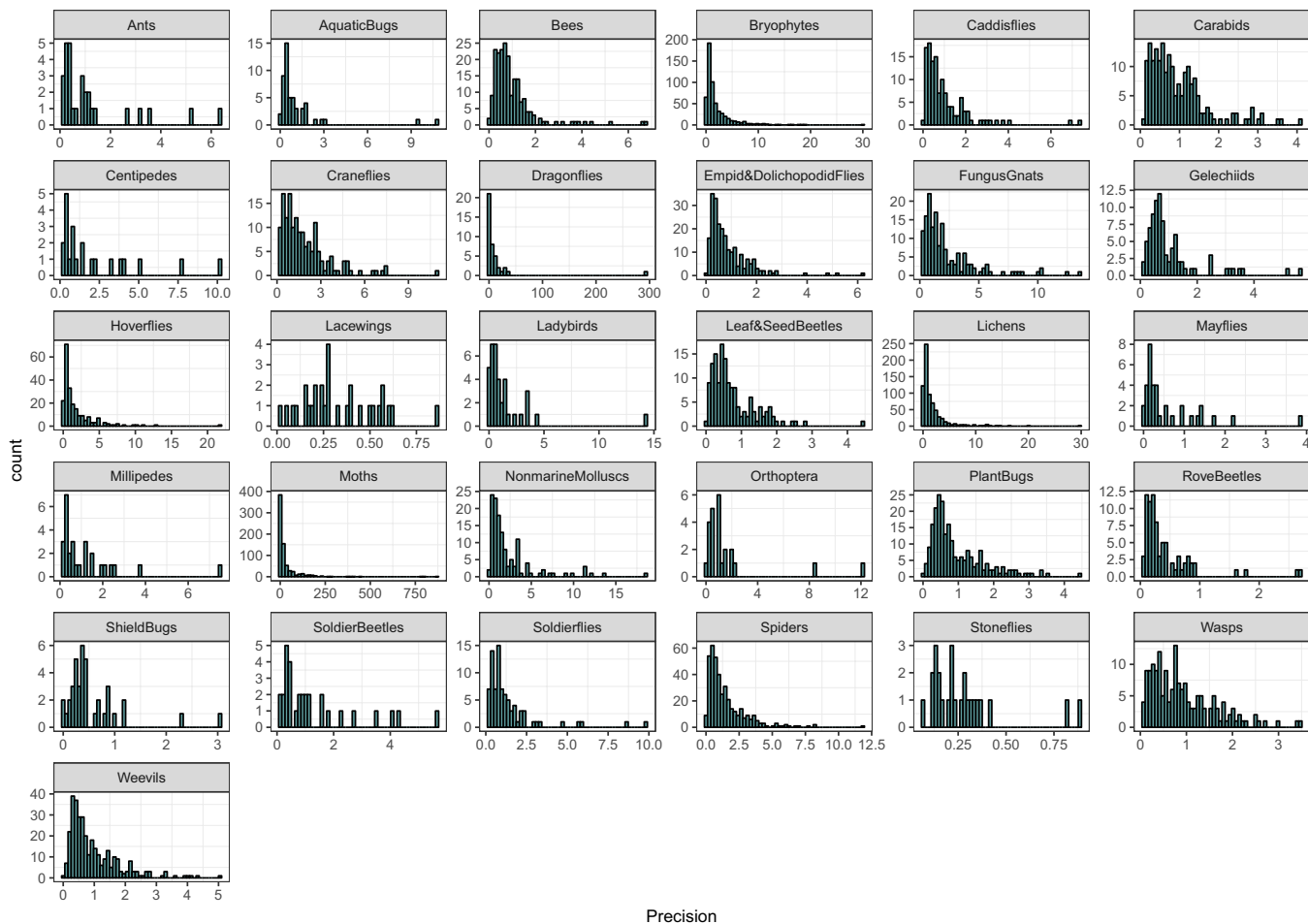
The number of records per species within each dataset also varied considerably. In some cases, data standardisation and the removal of sites visited in only a single year (using the *nyr* model parameter) meant that some species were left with very few records. A column detailing the number of records per species after filtering has been supplied to ensure users are aware of the number of records contributing to species estimates. These values can be found in the *N\_records* column of the “Species\_Trends.csv” file<sup>34</sup>.

It was not possible to validate these estimates against an independent source of distribution or occupancy trends, since this is the first time that such information has been produced. As a form of statistical validation, we explore the precision of the trend estimates. Precision of the trend estimates are presented within the *Species\_Trends.csv* file<sup>34</sup> but are highly variable (Fig. 2) reflecting variance in the number of records available for each species<sup>39</sup>.

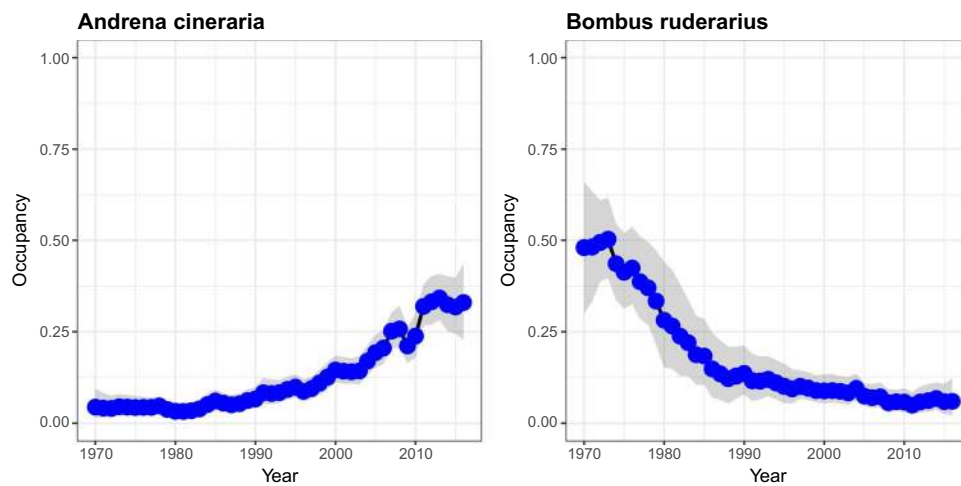
### Usage Notes

This dataset can be used to assess change in occupancy of single species or an aggregation of species. Plotting the mean estimate from the summary table alongside the associated credible intervals for a species will give you a plot of the occupancy estimates for that species over time (Fig. 3).





**Fig. 2** Histogram of the precision estimates of the estimated species trends (annual percentage growth rates). Note that the axes vary across groups.



**Fig. 3** Example plots of species occupancy estimates over time and associated 95% credible intervals for two bee species.

When using the outputs provided within this dataset, users need to consider the uncertainty assessments supplied alongside the data. Those species where the uncertainty assessments we consider unreliable (those with fewer than 50 records and gaps of 10 years between records) have been removed from this dataset. However, users are urged to make their own judgement on whether the uncertainties are small enough to provide useful information in the context in which they are being used. Uncertainty can be established by summarising 95% quantiles of

the posterior samples or from the supplied 95% credible intervals in the species occupancy summary tables and species trends table. Data users should also make sure to take a note of convergence of parameters contributing to the estimates when using these outputs.

The fitting of the models and analysis of the outputs produced can be carried out using two R packages that have been developed for this purpose. *sparta* is an R package that has been developed to carry out methods for the estimation of species trends from occurrence records<sup>29</sup>. This package is freely available on GitHub: <https://github.com/BiologicalRecordsCentre/sparta>.

The posterior samples for species can be used to generate aggregate indicators of change in occupancy over time with associated measures of uncertainty for groups of species or for specific regions. Using the posterior samples means that uncertainties can be propagated throughout the analysis. Another R package, *BRCindicators*, has been developed to estimate species trends and generate indicators of change over time from the outputs produced from *sparta* or similar methods. This package is also available on GitHub: <https://github.com/BiologicalRecordsCentre/BRCindicators>.

### Code availability

Code used for taxa specific input data standardisation is not presented. Species name checks and changes were taxa specific and required a lot of manual processing after consultation with scheme organisers. Information on species aggregations, removals and name changes are, however, detailed in the “Species\_Names.csv” spreadsheet.

Functions for organising data into detection histories and for fitting the specified occupancy model are available in the R package *sparta*<sup>29</sup>. The function *formatOccData* was used to arrange the data into detection histories and to calculate the list length of visits. The function *occDetFunc* was used to run the models. Note that in order to run these models using *sparta*, JAGS must be downloaded separately in order to carry out the MCMC sampling<sup>31</sup>. An example workflow detailing function and model specifications has been supplied within Supplementary File 1. This PDF document runs through each subsection of the methods, except the raw data processing, providing the code used and examples of the outputs produced as a result. Raw data processing was not included since processes were group specific and raw data could not be supplied alongside the outputs due to data provider restrictions.

Received: 17 June 2019; Accepted: 10 October 2019;

Published online: 05 November 2019

### References

- Tittensor, D. P. *et al.* A mid-term analysis of progress toward international biodiversity targets. *Science* (80-) **346**, 241–244 (2014).
- Gregory, R. & van Strien, A. Wild bird indicators: using composite population trends of birds as measures of environmental health. *Ornithol. Sci* **9**, 3–22 (2010).
- Brereton, T., Roy, D. B., Middlebrook, I., Botham, M. & Warren, M. The development of butterfly indicators in the United Kingdom and assessments in 2010. *J. Insect Conserv.* **15**, 139–151 (2010).
- Barlow, K. E. *et al.* Citizen science reveals trends in bat populations: The National Bat Monitoring Programme in Great Britain. *Biol. Conserv.* **182**, 14–26 (2015).
- Powney, G. D., Cham, S. S. A., Smallshire, D. & Isaac, N. J. B. Trait correlates of distribution trends in the Odonata of Britain and Ireland. *PeerJ* **3**, e1410 (2015).
- Fox, R., Conrad, K. F., Parsons, M. S., Warren, M. S. & Woiwod, I. P. The state of Britain's larger moths (2006).
- Eaton, M. A. *et al.* The priority species indicator: measuring the trends in threatened species in the UK. *Biodiversity* 1–12, <https://doi.org/10.1080/14888386.2015.1068222> (2015).
- Powney, G. D. & Isaac, N. J. B. Beyond maps: a review of the applications of biological records. *Biol. J. Linn. Soc.* n/a–n/a, <https://doi.org/10.1111/bij.12517> (2015).
- Ball, S., Morris, R., Rotheray, G. & Watt, K. Atlas of the Hoverflies of Great Britain (Diptera, Syrphidae) (2011).
- Powney, G. D., Rapacciuolo, G., Preston, C. D., Purvis, A. & Roy, D. B. A phylogenetically-informed trait-based analysis of range change in the vascular plant flora of Britain. *Biodivers. Conserv.* **23**, 171–185 (2013).
- Stroh, P. A. *et al.* A Vascular Plant Red List for England. (Botanical Society of Britain and Ireland, 2014).
- Fox, R. *et al.* Long-term changes to the frequency of occurrence of British moths are consistent with opposing and synergistic effects of climate and land-use changes. *J. Appl. Ecol.* **51**, 949–957 (2014).
- Pescott, O. L. *et al.* Ecological monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and Ireland. *Biol. J. Linn. Soc.* **115**, 505–521 (2015).
- Isaac, N. J. B. & Pocock, M. J. O. Bias and information in biological records. *Biol. J. Linn. Soc.* **115** (2015).
- Boakes, E. H. *et al.* Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol.* **8**(6), e1000385 (2010).
- Kéry, M., Gardner, B. & Monnerat, C. Predicting species distributions from checklist data using site-occupancy models. *J. Biogeogr.* no-no, <https://doi.org/10.1111/j.1365-2699.2010.02345.x> (2010).
- van Strien, A. J., van Swaay, C. A. M. & Termaat, T. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *J. Appl. Ecol.* **50**, 1450–1458 (2013).
- Woodcock, B. A. *et al.* Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nat. Commun.* **7**, 12459 (2016).
- van Strien, A. J. *et al.* Modest recovery of biodiversity in a western European country: The Living Planet Index for the Netherlands. *Biol. Conserv.* **200**, 44–50 (2016).
- MacKenzie, D. I. *et al.* Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. (Academic Press, 2006).
- Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P. & Roy, D. B. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods. Ecol. Evol.* **5**, 1052–1060 (2014).
- Kéry, M. *et al.* Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conserv. Biol.* **24**, 1388–97 (2010).
- Fox, R. *et al.* *The State of the UK's Butterflies 2015* (2015).
- Outhwaite, C. L. *et al.* Prior specification in Bayesian occupancy modelling improves analysis of species occurrence data. *Ecol. Indic.* **93**, 333–343 (2018).
- August, T. *et al.* Emerging technologies for biological recording. *Biol. J. Linn. Soc.* **115**, 731–749 (2015).

26. Mackenzie, D. I. & Royle, J. A. Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology* **42**, 1105–1114 (2005).
27. van Strien, A. J., Termaat, T., Groenendijk, D., Mensing, V. & Kéry, M. Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists. *Basic Appl. Ecol.* **11**, 495–503 (2010).
28. Szabo, J. K., Vesk, P. A., Baxter, P. W. J. & Possingham, H. P. Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. *Ecol. Appl.* **20**, 2157–2169 (2010).
29. August, T. *et al.* sparta: Trend Analysis for Unstructured Data. R package version 0.1.40 (2018).
30. Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R. & Tingley, R. Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution* **8**, 1081–1091 (2017).
31. Plummer, M. JAGS Version 3.4.0 (2009).
32. Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **7**, 457–472 (1992).
33. Kéry, M. & Schaub, M. *Bayesian population analysis using WinBUGS: A hierarchical perspective*. (Elsevier, 2012).
34. Outhwaite, C. L. *et al.* Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK (1970–2015). NERC Environmental Information Data Centre, <https://doi.org/10.5285/0ec7e549-57d4-4e2d-b2d3-2199e1578d84> (2019).
35. Smith, A. J. E. *Moss Flora of Britain and Ireland*. (Cambridge University Press, 2004).
36. Blockeel, T. L., Bosanquet, S. D. S., Hill, M. O. & Preston, C. D. *Atlas of British & Irish Bryophytes*. (Pisces Publications, 2014).
37. Chamberlain, S. & Szocs, E. taxize - taxonomic search and retrieval in R. *F1000Research* **2**, 191 (2013).
38. Chamberlain, S. *et al.* taxize: Taxonomic information from around the web. *R package version 0.9.7* (2019).
39. Pocock, M. J. O. *et al.* Rapid assessment of the suitability of multi-species citizen science datasets for occupancy trend analysis. *bioRxiv*, <https://doi.org/10.1101/813626> (2019).

## Acknowledgements

We would like to acknowledge the contribution of the dedicated and skilled volunteers who collected the species records used within this research and the associated schemes and societies. A special thanks goes to those scheme organisers/participants who supplied data and checked species lists (the remainder are co-authors of this study): Mike Edwards, Chris Preston, Dave Smallshire, Steve Hewitt, Martin Drake and Peter Chandler. Thanks also to Kevin Walker for comments on a previous version of this manuscript and to Colin Harrower for assistance with and advice on data extraction. This work was funded by the Natural Environment Research Council (NERC), award number NE/L008823/1. This work was also supported by the UK Joint Nature Conservation Committee, the Natural Environment Research Council (through National Capability funding), by Defra and the Scottish Government under project WC1101 and by Defra, JNCC, the Welsh Government, Scottish Government and partners of the UK Pollinator Monitoring and Research Partnership under project BE0125. Additionally, the research was partly funded by Natural Environment Research Council and the Biotechnology and Biological Sciences Research Council (BBSRC) under research programmes NE/N018125/1/LTS-M ASSIST – Achieving Sustainable Agricultural Systems, and by the Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability. This work used JASMIN, the UK collaborative data analysis facility.

## Author contributions

N.J.B.I. and R.E.C. conceived the study. C.L.O. collated the data with assistance from S.R., H.E.R. and M.H. C.L.O., G.D.P. and T.A.A. ran the models and organised the model outputs. C.L.O. led the writing of the manuscript with input from G.D.P., T.A.A., R.E.C., O.L.P., R.F., K.W., D.B.R. and N.J.B.I. O.L.P., M.H., H.E.R., R.F., K.W., K.A., S.B., T.B., T.B., B.C.C., T.C., J.F., A.F., P.H., P.H., D.H., D.H., J.K., P.L., C.M., R.M., A.N., S.P., C.P., J.S., A.S., P.S., M.T. and I.W. supplied data via the recording schemes they are a part of, performed species name checks and provided advice on data use.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41597-019-0269-1>.

**Correspondence** and requests for materials should be addressed to C.L.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019

Charlotte L. Outhwaite<sup>1,2,3\*</sup>, Gary D. Powney<sup>1</sup>, Tom A. August<sup>1</sup>, Richard E. Chandler<sup>4</sup>,  
 Stephanie Rorke<sup>1</sup>, Oliver L. Pescott<sup>1,5</sup>, Martin Harvey<sup>1,6</sup>, Helen E. Roy<sup>1,7</sup>, Richard Fox<sup>8</sup>,  
 David B. Roy<sup>1</sup>, Keith Alexander<sup>9</sup>, Stuart Ball<sup>10</sup>, Tristan Bantock<sup>11</sup>, Tony Barber<sup>12</sup>,  
 Björn C. Beckmann<sup>1,13</sup>, Tony Cook<sup>14</sup>, Jim Flanagan<sup>15</sup>, Adrian Fowles<sup>16</sup>, Peter Hammond<sup>17</sup>,  
 Peter Harvey<sup>18</sup>, David Hepper<sup>19</sup>, Dave Hubble<sup>20</sup>, John Kramer<sup>21</sup>, Paul Lee<sup>22</sup>,  
 Craig MacAdam<sup>23,24</sup>, Roger Morris<sup>10</sup>, Adrian Norris<sup>25</sup>, Stephen Palmer<sup>26</sup>, Colin W. Plant<sup>27</sup>,  
 Janet Simkin<sup>28</sup>, Alan Stubbs<sup>21</sup>, Peter Sutton<sup>13</sup>, Mark Telfer<sup>29</sup>, Ian Wallace<sup>30</sup> & Nick J. B. Isaac<sup>1,2</sup>

<sup>1</sup>Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>2</sup>Centre for Biodiversity and Environment Research, University College London, Gower Street, London, WC1E 6BT, UK. <sup>3</sup>RSPB Centre for Conservation Science, RSPB, the Lodge, Sandy, Bedfordshire, SG19 2DL, UK. <sup>4</sup>Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK. <sup>5</sup>British Bryological Society, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>6</sup>Soldierflies and Allies Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>7</sup>UK Ladybird Survey, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>8</sup>National Moth Recording Scheme, Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset, BH20 5QP, UK. <sup>9</sup>Soldier Beetles, Jewel Beetles and Glow-worms Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>10</sup>Dipterists Forum, Hoverfly Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>11</sup>Terrestrial Heteroptera Recording Scheme - Shield bugs and allied species, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>12</sup>British Myriapod and Isopod Group, Centipede Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>13</sup>Grasshoppers and Related Insects Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, UK. <sup>14</sup>Aquatic Heteroptera Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>15</sup>Terrestrial Heteroptera Recording Scheme - Plant bugs and allied species, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>16</sup>Weevil and Bark Beetle Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>17</sup>Staphylinidae Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>18</sup>Spider Recording Scheme, British Arachnological Society, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>19</sup>Dragonfly Conservation Group, British Dragonfly Society, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>20</sup>Chrysomelidae Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>21</sup>Dipterists Forum, Cranefly Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>22</sup>British Myriapod and Isopod Group, Millipede Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>23</sup>Riverfly Recording Schemes: Ephemeroptera, c/o Buglife Scotland, Balallan House, 24 Allan Park, Stirling, FK8 2QG, UK. <sup>24</sup>Riverfly Recording Schemes: Plecoptera, c/o Buglife Scotland, Balallan House, 24 Allan Park, Stirling, FK8 2QG, UK. <sup>25</sup>Conchological Society of Great Britain and Ireland, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>26</sup>Gelechiid Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>27</sup>Lacewings and Allies Recording Scheme, 14 West Road, Bishops Stortford, Hertfordshire, CM23 3QP, UK. <sup>28</sup>British Lichen Society, c/o School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. <sup>29</sup>Ground Beetle Recording Scheme, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK. <sup>30</sup>Riverfly Recording Schemes: Trichoptera, c/o Biological Records Centre, Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK.