Population variability of species can be deduced from citizen science records: a case study using British butterflies

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Population variability in species can be deduced from opportunistic citizen science records: a case study using British butterflies

S U Z A N N A   C .   M A S O N ,[1,2]   JANE K. HILL,[2] CHRIS D. THOMAS,[2] GARY D. POWNEY,[1] RICHARD FOX,[3] TOM BRERETON[3] and TOM H. OLIVER[4]   [1]NERC Centre for Ecology and Hydrology, Wallingford, UK, [2]Department of Biology, University of York, York, UK, [3]Butterfly Conservation, Wareham, Dorset, UK and [4]School of Biological Sciences, University of Reading, Reading, UK

Abstract. 1. Abundance data are the foundation for many ecological and conservation projects, but are only available for a few taxonomic groups. In contrast, distribution records (georeferenced presence records) are more widely available. Here we examine whether year-to-year changes in numbers of distribution records, collated over a large spatial scale, can provide a measure of species' population variability, and hence act as a metric of abundance changes.

2. We used 33 British butterfly species to test this possibility, using distribution and abundance data (transect counts) from 1976 to 2012.

3. Comparing across species, we found a strong correlation between mean year-to-year changes in total number of distribution records and mean year-to-year changes in abundance ($N = 33$ species; $r^2 = 0.66$). This suggests that annual distribution data can be used to identify species with low versus high population variability.

4. For individual species, there was considerable variation in the strength of relationships between year-to-year changes in total number of distribution records and abundance. Between-year changes in abundance can be identified from distribution records most accurately for species whose populations are most variable (i.e. have high annual variation in numbers of records).

5. We conclude that year-to-year changes in distribution records can indicate overall population variability within a taxon, and are a reasonable proxy for year-to-year changes in abundance for some types of species. This finding opens up more opportunities to inform ecological and conservation studies about population variability, based on the wealth of citizen science distribution records that are available for other taxa.

Key words. Abundance, biogeography, biological records, inter-annual variation, population size, range extent.

Introduction

The long-term monitoring of population dynamics is an important aspect of ecology, and allows examination of

factors driving species' abundance trends, such as the effects of weather (Roy *et al.*, 2001), habitat (Lemoine *et al.*, 2007), disease (Daszak *et al.*, 2003) and human impacts (Lewis & Vandewoude, 2015). Monitoring abundance trends of species thus helps to identify species at risk, develop conservation strategies to halt population declines (Brown *et al.*, 1995), and identify increasing populations of pests to implement control strategies

Correspondence: Suzanna Mason, NERC Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK. E-mail: sazmas@ceh.ac.uk

(Petrovskii *et al.*, 2014). Measuring population variability is essential to explore the influence of environmental factors, such as climatic cycles or food availability, on population dynamics (van Schaik & van Noordwijk, 1985; Lynam *et al.*, 2004). In addition, population variability may be an important determinant of the likelihood that populations will survive in habitat fragments, and variability may indicate the sensitivity of populations to climatic fluctuations (Pimm *et al.*, 1998; Vucetich *et al.*, 2000; Oliver *et al.*, 2012). Nevertheless, collecting abundance data may be time-consuming and expensive, and thus many taxonomic groups lack information on abundance trends and population dynamics. In contrast, many more species have large datasets of distribution records (i.e. unique records of the presence of species at a given location and date). Such data are available for a wide range of taxonomic groups, tend to cover wide areas, span many years, and are often collected as part of 'citizen science' projects (Devictor *et al.*, 2010; Pocock *et al.*, 2015).

It is well-known that there is a positive relationship between species' abundances and distributions (Brown, 1984; Gaston *et al.*, 2000) and very abundant species tend to have larger ranges (Holt *et al.*, 1997). Abundance-distribution relationships are general patterns in ecology, but there are many forms of the relationship (Gaston, 1996), and these relationships are not necessarily linear (Hartley, 1998). In spite of this complexity, strong relationships have been found between distribution and abundance, which are evident over time, large spatial scales and different taxonomic groups (Zuckerberg *et al.*, 2009; Roney *et al.*, 2015). These relationships allow occupancy changes (changes in the likelihood of a species' presence) to be used to estimate population trends (Tempel & Gutiérrez, 2013), broad biodiversity changes to be assessed across multiple taxonomic groups (Oliver *et al.*, 2015a), and long-term trends in the frequency of species' occurrences to be modelled (Pearce & Boyce, 2006). These long-term occurrence trends have been shown to be reasonable proxies for abundance trends for both birds (Kamp *et al.*, 2016) and butterflies (Warren *et al.*, 2001; Oliver *et al.*, 2015a). But, there is little information on the capacity of distribution data to describe other aspects of population dynamics, such as population variability, which is an important factor affecting extinction risk (Inchausti & Halley, 2003; Mace *et al.*, 2008).

A challenge for ecologists is deriving an accurate measure of population variability when standardised abundance estimates are lacking. The positive associations between distribution size and abundance suggest that distribution records could potentially be used in analyses inferring species' population dynamics, by acting as proxies for abundance data. If there are strong and predictable relationships between year-to-year changes in abundance and year-to-year changes in distribution records, then distribution records could provide a useful metric for ecologists to study the factors affecting population variability in a much wider range of taxa than is currently possible.

In this study, we examine the relationships between abundance and distribution to assess whether year-to-year changes in the number of distribution records are strongly related to year-to-year changes in abundance. We study British butterflies because there are long-term and fine-scale data on both distribution and abundance, allowing robust testing of these relationships. We predict that year-to-year changes in abundance will be strongly positively related to year-to-year changes in distribution records, because increasing numbers of individuals would be expected to result in an increased likelihood of a species being recorded. In addition, as a population increases in size, density-dependent dispersal would be expected to result in individuals moving away from areas of high population density, thereby increasing the number of sites where species can be observed (Gaston *et al.*, 2000).

Within this broad topic, we examine three issues. The first is whether it is possible to identify species with higher or lower population variability using distribution data – a between-species comparison. We do this by calculating average between-year changes in the numbers of distribution records over time, and comparing these estimates with measures of variability that are based on fixed-transect population count data. Secondly, we assess whether distribution records can be used as proxies for inter-annual changes in abundance in each species separately – a within-species analysis. Finally, we identify the characteristics of species for which distribution data provide a proxy for abundance, concentrating on three attributes that can be deduced from the distribution records themselves (i.e. not requiring additional ecological or population dynamic data, which are lacking for many taxa). We selected these metrics because they are likely to be linked to our statistical capacity to detect year-to-year variation in abundance from distribution records: (i) the total number of distribution records for a species, (ii) how aggregated these records are in space (using a metric of 'fractal dimension' of distribution records) and (iii) the average size of the year-to-year changes in distribution records (i.e. how much annual variation there is in distribution records for a species). We refer to these metrics as 'biogeographical attributes', but recognise that they are also influenced by variation in recording intensity across species and over time. We also examine the effect of the spatial scale of the study area on the relationship between year-to-year changes in distribution records and year-to-year changes in abundance, by comparing data analysed at national (UK study area, 30,2800 km^2) and regional (county study area, 440 km^2) levels, given that population fluctuations may be synchronous in their dynamics at one spatial scale but not others (Sutcliffe *et al.*, 1996).

Materials and methods

Study species

We studied 33 species of British butterfly (See Table 1), including northern and southern species, and resident and migrant species, over the period 1976 to 2012. This

Table 1. The goodness of fit of the inter-annual distribution-abundance relationships for 33 butterflies.

Species	r^2	ΣD	Mean $|\Delta D|$	Fractal D
Aglais io	0.36	407 408	0.10	0.338
Aglais urticae	0.60	442 648	0.13	0.322
Anthocharis cardamines	0.22	220 768	0.11	0.302
Aphantopus hyperantus	0.18	177 673	0.08	0.300
Argynnis aglaja	0.37	28 184	0.10	0.468
Argynnis paphia	0.27	31 324	0.13	0.516
Aricia agestis	0.65	44 785	0.16	0.441
Boloria selene	0.03	20 723	0.11	0.480
Callophrys rubi	0.25	31 394	0.12	0.448
Celastrina argiolus	0.85	165 545	0.26	0.365
Coenonympha pamphilus	0.35	144 788	0.08	0.311
Erynnis tages	0.39	31 119	0.10	0.543
Favonius quercus	0.07	30 622	0.14	0.413
Gonepteryx rhamni	0.34	184 215	0.11	0.390
Hipparchia semele	0.18	22 647	0.09	0.716
Lasiommata megera	0.43	87 900	0.11	0.384
Limenitis camilla	0.48	17 988	0.16	0.520
Lycaena phlaeas	0.66	150 387	0.14	0.332
Maniola jurtina	0.11	459 084	0.07	0.309
Melanargia galathea	0.16	66 946	0.10	0.479
Ochlodes sylvanus	0.29	135 278	0.09	0.276
Pararge aegeria	0.10	400 596	0.10	0.375
Pieris brassicae	0.44	458 225	0.13	0.353
Pieris napi	0.25	399 295	0.10	0.303
Pieris rapae	0.32	474 880	0.11	0.370
Polygonia c-album	0.58	223 318	0.14	0.353
Polyommatus coridon	0.47	17 523	0.10	0.669
Polyommatus icarus	0.55	226 639	0.12	0.328
Pyrgus malvae	0.28	17 215	0.12	0.577
Pyronia tithonus	0.13	278 385	0.08	0.331
Thymelicus sylvestris	0.11	134 606	0.09	0.257
*Vanessa atalanta**	0.62	384 283	0.18	0.338
*Vanessa cardui**	0.92	183 430	0.53	0.342

Latin names with an asterisk (*) indicate migratory species. Presented are the Pearson's r^2 values of the relationship between year-to-year $\log_{10}$ change in abundance and year-to-year $\log_{10}$ change in total number of distribution records. We checked r_s values and found them to all be positive, indicating that the direction of the relationships below were always positive. Biogeographical attribute values are also included for each species: total number of distribution records (ΣD), mean absolute year-to-year change in $\log_{10}$ distribution records, fractal dimension (Fractal D).

study period was selected to maximise the geographical coverage of data, the length of the time-series of data analysed and the number of species analysed. We excluded species without 37 years of abundance and distribution data. Species that were subject to targeted, intensive surveying effort during certain years of the study period were also excluded (*Hesperia comma*; Thomas & Jones, 1993; *Boloria euphrosyne*; Brereton, 1998; and *Satyrium w-album*; Thomas, 2010), because large differences in the level of recording effort between years could bias results.

Distribution records

We computed year-to-year changes in distribution records based on data collected by volunteers for the Butterflies for the New Millennium (BNM) recording scheme,

surveying sites in the study area (see below) on an opportunistic basis using unstructured sampling (Fox *et al.*, 2015). A distribution record is an observation (recorded presence) of an individual species at a location on a particular date. Recording efforts are generally unstructured (there are no fixed or assigned times, places or methods for recording) and opportunistic, with little to no guidance given to recorders as to how, when and where to record, meaning that recording is influenced heavily by recorder behaviour (Boakes *et al.*, 2010; Isaac & Pocock, 2015). Recorder behaviour can vary due to encouragement to record in under-represented regions for the purposes of atlas creation or other targeted survey efforts. Despite these attempts to encourage, spatial and temporal variation in opportunistic recording effort remains high. Due to increased recruitment of recorders over time numbers of distribution records have increased (see Figure S2), which is why we de-trended the data prior to analysis.

The spatial and temporal resolution of BNM distribution records varies; we excluded records with spatial resolution coarser than a 10 km × 10 km grid square or with date ranges spanning more than 1 year. The study area was the UK, Isle of Man and Channel Islands (3028 hectads in total). We analysed a total of 5,873,182 distribution records from 1976 to 2012, after all filtering processes (see below). The majority of distribution records are independent of abundance data (UK Butterfly Monitoring Scheme (UKBMS) transect), but the distribution dataset did contain some records sourced from transects. Therefore, distribution records were excluded if they occurred within the 1 km grid cell that contained a UKBMS transect (based on the centroid of the digitised transect route). This led to 1604 1 km cells being excluded; approximately 5.3% of the UK land area and 26.2% (2 089 886) of records. Year-to-year changes in $\log_{10}$ distribution records were calculated for each study species over the 37-year study period by subtracting the number of distribution records ($\log_{10}$-transformed) in year t-1 from the number of records in year t.

Abundance data

We analysed abundance data from the UKBMS national collated index (www.ukbms.org). The UKBMS calculates their index from counts from weekly transect walks along fixed routes undertaken during the recording period (April-September) every year since 1976 (see http://www.ukbms.org/Methods.aspx for full details). Counts are taken from sites in Great Britain and Northern Ireland (1854 transect sites in total). Counts for missing weeks are estimated by the UKBMS by considering the area of a GAM curve fitted to observed weekly count data throughout the year (Rothery & Roy, 2001). The UKBMS national collated index from 1976 to 2012 is created using a log-linear model, with a transect site and year effect (Brereton *et al.*, 2011), as shown below:

$$\log_{10}(c_{ij}) = x_i + y_i$$

Where c is the expected count for site i in year j, and where x_i and y_j give the means for the ith site and the jth year. The index is then scaled to a mean of 2, for the purposes of comparing abundance trends across species. This produces a $\log_{10}$-transformed abundance index, which we used in our calculation of population variability. We computed year-to-year changes in $\log_{10}$ abundance by subtracting the abundance index value ($\log_{10}$-transformed) for year t-1 from the value for year t.

Accounting for phylogeny

Our butterfly species share evolutionary lineages, and this must be taken into account when analysing species together in models. All multispecies analyses conducted in this study accounted for the non-independence of species using phylogenetically informed linear models with estimated Pagel's λ, using the *pgls* function of the caper package in R (Pagel, 1999; Orme *et al.*, 2013), and the butterfly phylogeny of Brooks *et al.* (2016). These models are interpreted by p values indicating the difference between the phylogenetic correlation λ value (estimated using maximum likelihood) and the upper and lower bounds: 1 (indicating phylogenetic dependence) and 0 (indicating phylogenetic independence). In all our analyses, the phylogenetic correlation was not significantly different from the lower bound, indicating phylogenetic independence, and so we conclude that phylogeny did not influence our linear models.

Examining relationships between abundance and distribution records

First, we explored whether yearly changes in $\log_{10}$ distribution records (as above) were correlated with yearly changes in $\log_{10}$ abundance (as above), in a multispecies analysis with a control for phylogenetic independence (see section above). We computed overall mean change values for both variables for each species over the 37-year study period. In both cases (distribution-record and abundance changes), we calculated the average absolute magnitude of the year-to-year changes, rather than directional changes (positive or negative). We analysed year-to-year changes rather than absolute numbers each year to de-trend the data, and to remove any temporal trends in recording effort. This analysis tests whether species with high population variability (on transects) also have high variability in terms of numbers of distribution records.

Second, we examined each species separately. We calculated the strength of the relationships between year-to-year changes in $\log_{10}$ distribution records and changes in $\log_{10}$ abundance using r^2 values from least squares regressions. This relationship is hereafter termed the inter-annual distribution-abundance relationship and, for each study species, it reflects the extent to which yearly changes in $\log_{10}$ numbers of distribution records can be used to predict population size changes (from transect data).

Third, we examined the influence of three independent biogeographical attributes on these inter-annual distribution-abundance relationships to identify species for which distribution records were adequate proxies for population change. These attributes were: total number of distribution records; fractal dimension of a species' range; and overall variability in distribution records. We computed the total number of distribution records collected at any spatial resolution (10 m to 10 km grid) for a species during the study period (1976–2012). Fractal dimension is a metric of how 'well-filled' a species' range is, based on the proportion of 10 km grid cells with records within each occupied 100 km grid cell (Wilson *et al.*, 2004). For each species, we calculated the total area of all occupied 10 km and 100 km grid cells, and regressed these values against the length of the grid cells (10 km and 100 km

respectively; all values $\log_{10}$ transformed).The slope of the regression gives a measure (fractal dimension) of how 'well-filled' a species range is at 10 km scale, where a slope of 0 indicates a completely filled range, and a slope of two indicates a minimally filled range (see Figure S1 for two exemplar species; *Thymelicus sylvestris,* with the most well-filled range and *Hipparchia semele* with the most minimally filled range). For overall variability in distribution records we used the mean year-to-year change in $\log_{10}$ distribution records over the study period.

A phylogenetic multivariate regression was then fitted with the three biogeographical attributes as explanatory variables and the r^2 value of each species' inter-annual distribution-abundance relationship as the response variable. We fitted a fourth term to the model, the quadratic term of mean year-to-year change in $\log_{10}$ distribution records, to account for its apparent non-linear relationship with goodness-of-fit (r^2) values when relationships were visually inspected by plotting the data. We tested a full model, then removed non-significant terms using a stepwise deletion approach.

We selected these three attributes to test because autoecological information may be limited for other taxonomic groups, but these attributes can be easy from distribution datasets. Because butterflies, however, do have detailed autoecological information, we tested the influence of dispersal ability on the inter-annual distribution-abundance relationship in PGLS models, using two metrics: dispersal rankings based on expert opinion (Cowley *et al.*, 2001) and a mobility score calculated from indices of ecological information (Dennis *et al.*, 2004). We found no significant relationship between dispersal ability and the strength of the inter-annual distribution-abundance relationship (see Table S2).

Comparison of national and regional inter-annual distribution-abundance relationships

To investigate whether the goodness of fit of the inter-annual distribution-abundance relationships varied with spatial scale, we repeated our analysis of this relationship at a regional level, for the county of Dorset. We compared r^2 values from national and regional inter-annual distribution-abundance relationships for a sub-set of 23 butterfly species for the period 1983-2009 (maximum time period containing abundance data for species in Dorset). Dorset was selected because of its extensive history of surveying butterflies (Robertson *et al.*, 1988; Thomas *et al.*, 2001).

Results

Relationship between variability in abundance and distribution records across species

Across the 33 study species, there was a strong positive relationship between the mean year-to-year changes in
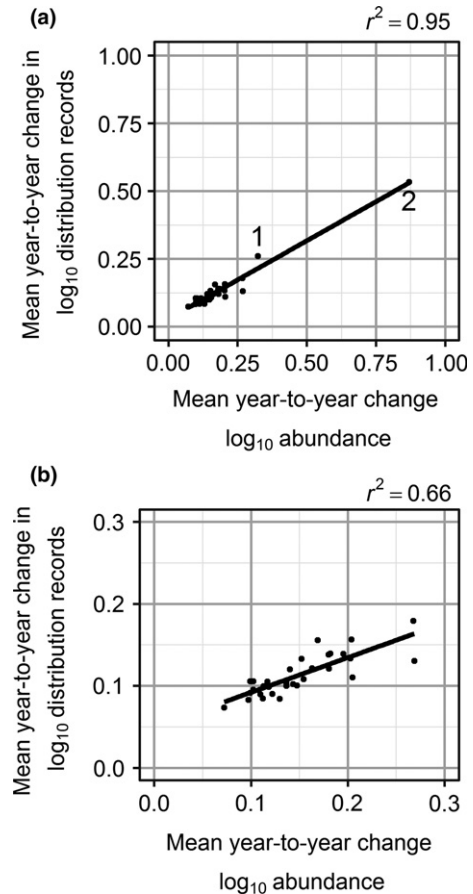

Fig. 1. Regressions of the mean year-to-year change in $\log_{10}$ distribution records against the mean year-to-year change in $\log_{10}$ abundance: (a) including all 33 species, with two outlier species labelled as follows: *Celastrina argiolus* (1) and *Vanessa cardui* (2); and (b) for 31 species, excluding these outlier species.

$\log_{10}$ distribution records and mean year-to-year changes in $\log_{10}$ abundance (Fig. 1a, PGLS, r^2: 0.95, $F_{1,31} = 623.8$, $P = <0.001$), even when two outlier species were removed (Fig. 1b, PGLS, r^2: 0.66, $F_{1,29} = 55.35$, $P = <0.001$). Thus, species that show high variability in abundance also have high variability in distribution records, and there was little evidence for any phylogenetic signal (i.e. results were not significantly different between models based on estimated λ, and where λ was set to 0).

Measuring inter-annual distribution-abundance relationships within species

For each of our 33 study species, the relationships between year-to-year changes in $\log_{10}$ distribution records and year-to-year changes in $\log_{10}$ abundance produced an overall mean r^2 value of 0.36, indicating that year-to-year changes in distribution records of UK butterflies provide a moderate proxy for year-to-year abundance changes.
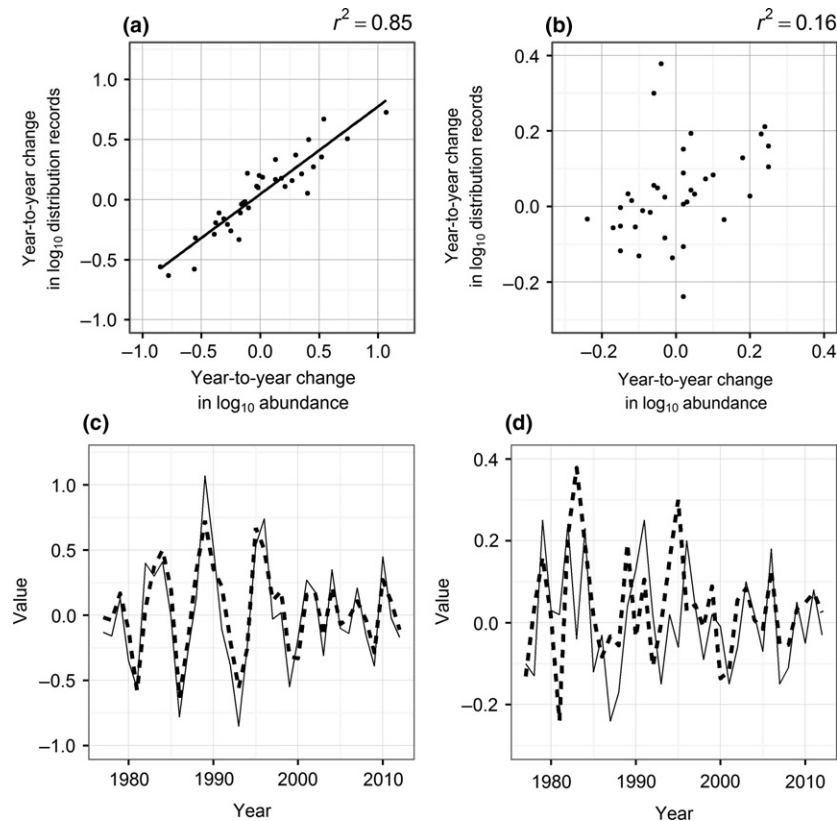
Fig. 2. Graphs show inter-annual distribution-abundance relationships for two exemplar species. Plots show year-to-year changes in $\log_{10}$ distribution records and year-to-year change in $\log_{10}$ abundance for *Celastrina argiolus* (a, c; left panels) and *Melanargia galathea* (b, d; right panels). Panels a and b show regressions (each point represents the change between a given pair of years), and panels c and d plot the same data together in a time-series (solid lines indicate year-to-year changes in $\log_{10}$ abundance, dashed line year-to-year changes in $\log_{10}$ distribution records).

Eight butterfly species had $r^2 > 0.5$, showing that distribution records were particularly informative in approximately 25% of study species. Nevertheless, there was considerable variation among species, with r^2 values varying between 0.03 and 0.92 (Table 1). Figure 2 highlights two exemplar species, where the relationship was strong (Holly blue, *Celastrina argiolus*, $r^2 = 0.85$) and one where the relationship was very weak (Marbled White, *Melanargia galathea*, $r^2 = 0.16$).

We selected these three attributes to test because autoecological information may be limited for other taxonomic groups, but these attributes can be easy calculated from distribution datasets. Because butterflies, however, do have detailed autoecological information, we tested the influence of dispersal ability on the inter-annual distribution-abundance relationship in PGLS models, using two metrics: dispersal rankings based on expert opinion (Cowley *et al.*, 2001) and a mobility score calculated from indices of ecological information (Dennis *et al.*, 2004). We found no significant relationship between dispersal ability and the strength of the inter-annual distribution-abundance relationship (see Table S2).

Influence of biogeographical attributes

The r^2 value for each species' inter-annual distribution-abundance relationship (i.e. relationships between year-to-year changes in $\log_{10}$ distribution records and year-to-year changes in $\log_{10}$ abundance; as in Fig. 2) was then analysed in relation to the biogeographical attributes of each species, which are provided in Table 1. We tested all these variables in a full model (PGLS, $r^2 = 0.64$, $F_{4,28} = 12.58$, AIC = -30.43, $P = <0.001$; Table 2a). Only mean absolute year-to-year changes in distribution records and its quadratic term significantly influenced inter-annual distribution-abundance relationships: total number of distribution records and fractal dimension did not, and were consequently dropped during model simplification. The best and most parsimonious model (PGLS, $r^2 = 0.63$, $F_{2,30} = 26.02$, AIC = -33.70, $P = <0.001$; Table 2b) revealed that the strength of the relationship (r^2 value) increased with overall variability in distribution records (Fig. 3). Thus, the results show that species with greater fluctuations in distribution records over time had stronger inter-annual distribution-abundance relationships

Table 2. (a) and (b) Results of full and best models, both with the goodness of fit (r^2 value) of the inter-annual distribution-abundance relationships as the response variable.

Coefficients	Estimate	Std. Error	t value	P
(a)				
Intercept	−0.333	0.193	−1.724	0.096
Mean year-to-year change in $\log_{10}$ distribution records	6.756	1.385	4.879	<0.001
Quadratic mean year-to-year change in $\log_{10}$ distribution records	−8.307	2.310	−3.597	0.001
Total number of species records	<0.001	<0.001	0.570	0.573
Fractal dimension	−0.025	0.290	−0.086	0.932
(b)				
Intercept	−0.317	0.128	−2.481	0.019
Mean year-to-year change in $\log_{10}$ distribution records	6.701	1.351	4.961	<0.001
Quadratic mean year-to-year change in $\log_{10}$ distribution records	−8.214	2.250	−3.660	<0.001

Table 2a shows the first, full model with the following explanatory variables: mean absolute year-to-year change in distribution records, total number of species records and fractal dimension. The model summary statistics were: $r^2 = 0.64$, $F_{4,28} = 12.58$, AIC = −30.43, $P = <0.001$. Table 2b shows the best model with only one explanatory variable: mean absolute year-to-year change in distribution records. Model summary statistics: $r^2 = 0.63$, $F_{2,30} = 26.02$, AIC = −33.70, $P = <0.001$. In both models, the quadratic term of the mean absolute year-to-year change in distribution records was included to account for the non-linear nature of the relationship, and model results with estimated λ were not significantly different from a model with λ set to 0 (Fig. 3).

(although the effect of variability in records was non-linear and asymptoted at roughly 0.8; Fig. 3). Two species (*Celastrina argiolus* and *Vanessa cardui*) potentially had strong effects on our analyses (Fig. 3c), but excluding these two species did not alter our conclusions (Table S1).

Comparison of national and regional inter-annual distribution-abundance relationships

The strength of inter-annual distribution-abundance relationships computed for species at a regional level (Dorset) were strongly positively correlated with those computed at the national level (PGLS, $r^2 = 0.53$, $F_{1,21} = 23.25$, $P = <0.001$; Fig. 4). Therefore, we conclude that any differences in population synchrony between national and regional scales had little influence on the strength of inter-annual distribution-abundance relationships for butterfly species.
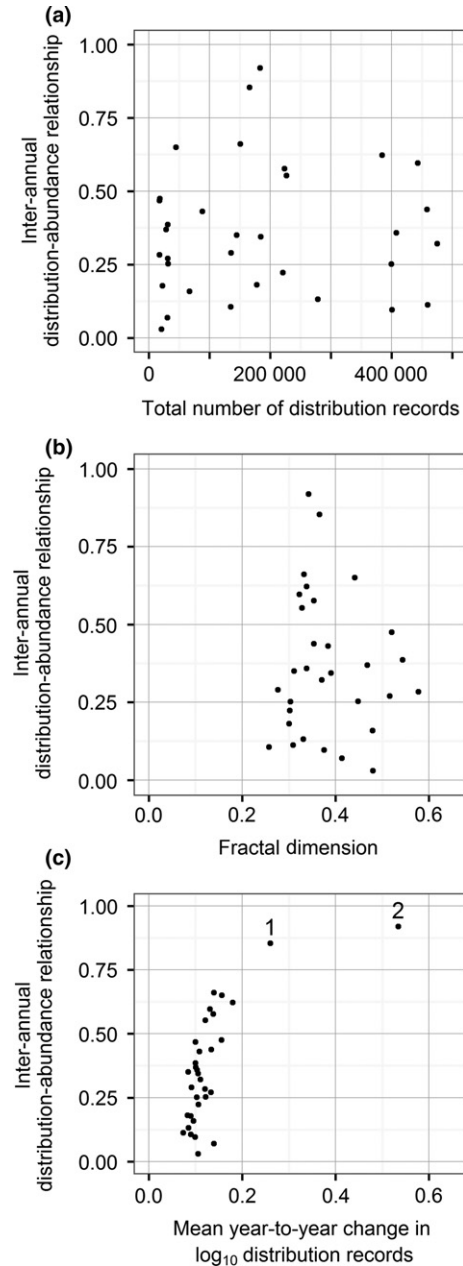

Fig. 3. Regressions of the inter-annual distribution-abundance relationship and three biogeographical attributes of the species: (a) total number of distribution records (PGLS, $r^2 = <0.01$, $F_{1,31} = 0.09$, $P = 0.76$); (b) fractal dimension (PGLS, $r^2 = 0.02$, $F_{1,31} = 0.61$, $P = 0.44$) and (c) mean absolute year-to-year change in $\log_{10}$ distribution records (PGLS, $r^2 = 0.63$, $F_{2,30} = 26.02$, $P = <0.001$). Each dot represents a species; the numbered data points on panel (c) are *Celastrina argiolus* (1) and *Vanessa cardui* (2).

Discussion

We found that citizen-collected distribution data can be used to extract information about population variability, in the absence of bespoke abundance monitoring
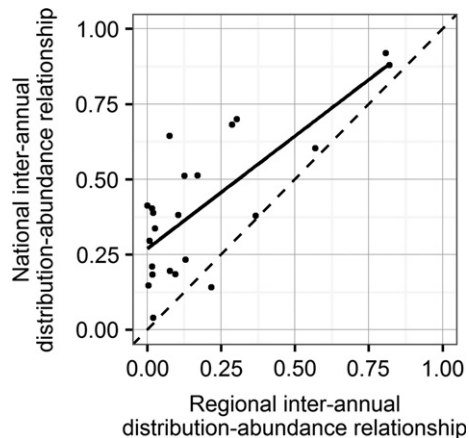
Fig. 4. Regression of a national inter-annual distribution-abundance relationship against a regional inter-annual distribution-abundance relationship (region = county of Dorset) calculated for 23 butterfly species, indicated by the solid black line. The dashed line indicates the 1:1 line.

programmes. In particular, mean year-to-year changes in distribution records were positively related to mean year-to-year changes in abundance (with outlier species removed, r^2 value: 0.66; Fig. 1). Thus, we were able to identify species with low and high between-year population variability quite accurately, using distribution data. This result supports the ability of unstructured citizen science data to reflect population-dynamic patterns found in long-term abundance data, and hence citizen science data may be useful in multispecies studies for which it is necessary have an overall measure of population variability (Robertson *et al.*, 2015; Gandiwa *et al.*, 2016) where abundance data are lacking. The ability to recognise species with the highest levels of population variability may help identify species that are at greatest risk of stochastic extinction following habitat fragmentation (Pimm *et al.*, 1998; Vucetich *et al.*, 2000; Oliver *et al.*, 2012), and the most variable species may potentially be the most responsive to yearly variation in climatic conditions (Maclean *et al.*, 2008; Howard *et al.*, 2015) and to parasitoids or other natural enemies (Robertson *et al.*, 2015). The findings from these analyses imply that information from citizen science data can provide useful input to landscape-scale conservation planning and to climate-change risk assessments.

When we considered each species in turn, there was considerable variation in the strength of relationships between year-to-year changes in distribution records and abundance among our study species; although these associations were always positive, averaging an r^2 of 0.36 across all species (Table 1). These relationships suggest that there is also some potential to use the distribution records of individual species to infer their population dynamics in greater detail (rather than as one metric for overall variability in the time-series). But, this is only feasible for some species: only eight out of 33 species having 'strong' relationships ($r^2 > 0.5$) between year-to-year

abundance and distribution changes. Thus it should not be presumed that distribution records can be used as a substitute for population data in the assessment of inter-annual change for all species.

Inferring abundance change from distribution data

Many species are declining or facing range retractions (Hayhow *et al.*, 2016), and it is important to monitor their population trends. Species with highly variable population dynamics tend to be at high risk of extinction (Pimm *et al.*, 1998; Vucetich *et al.*, 2000; Oliver *et al.*, 2012) and thus our measure of variability in distribution records has ecological value, with the potential to assist conservation assessments by helping to identify species at risk of extinction or habitats in need of management (Meyer *et al.*, 2015; Sánchez-Hernández *et al.*, 2015). Our multispecies analysis (Fig. 1) indicates that it is possible to derive robust estimates of population variability using distribution data alone.

Despite the promising results, there are two caveats that we should highlight. In this study, we examined only one taxonomic group with a high level of recording effort by citizen scientists. We also included only species with data in every year of the study period, excluding rare/less well-studied species. The value of other distribution datasets with lower recording effort may not be so informative. Kamp *et al.* (2016) found that reducing the number of distribution records resulted in poorer abundance trend estimates for Danish birds. Even without reducing the sample size, population trends were misclassified for 50% of the species they considered. Thus, using distribution data to infer population changes may require quite mature citizen science schemes, with substantial numbers of distribution records. Given that butterflies are a data-rich taxonomic group in the UK it remains unknown whether other groups will have sufficient data to replicate these results. Datasets which may have sufficient data for this method are butterflies in other countries, or other taxa in the UK, for which standardised abundance monitoring schemes are lacking, e.g. dragonflies.

Our second caveat is that more detailed population-dynamic interpretations of distribution data only seem possible for some species. Our finding that citizen science distribution data explain an average of only 34% of the year-to-year variation in abundance is unlikely to be sufficient to build meaningful models for examining the sensitivity of populations to environmental drivers, such as specific climate variables. For example Malinowska *et al.* (2014) were unable to detect impacts of extreme weather events on populations of ectothermic species from distribution records, despite evidence of these impacts from population data (e.g. Oliver *et al.*, 2015b). In addition, while we have removed species which have unusually high levels of recording effort due to species-specific surveys, not all species are necessary reliably monitored by UKBMS, which could result in poor year-to-year

distribution-abundance relationships. For example the purple hairstreak butterfly (*Favonius quercus*) occurs in tree canopies, and is therefore difficult to monitor from ground-based surveys. Other species may suffer from limited recording for other reasons, such as occurring in restricted locations or not being identified correctly due to confusion with other morphologically similar species.

Biogeographical attributes

Despite the above caveats, we conclude that year-to-year changes in distribution records represented an adequate proxy for abundance change in species with large fluctuations in their occurrence from year to year (Fig. 3, Table 1). Species with large year-to-year fluctuations in their occurrences, such as migrants, may offer the greatest statistical power to deduce population changes from distribution data. Even though two migrant species and the holly blue butterfly *Celastrina argiolus* demonstrate the strongest inter-annual distribution-abundance relationships, the mean year-to-year change in distribution records was also an important variable in predicting the strength of the year-to-year distribution-abundance relationship for other species. Therefore, mean year-to-year change in distribution records may help to identify non-butterfly species where citizen science distribution data could be used as a 'replacement' for direct population data. We found that total numbers of records and fractal dimension did not significantly influence the inter-annual distribution-abundance relationship. The most parsimonious explanation for this is that these variables are not important, and that our hypotheses, that our statistical capacity to detect year-to-year variation in abundance from distribution records was linked to the total number of distribution records, and fractal dimension, were wrong. We had predicted that a large total number of records would mean greater statistical power to find the inter-annual distribution-abundance relationship. The lack of a significant relationship between the inter-annual distribution-abundance relationship and total number of distribution records could be because patterns of year-to-year change in distribution records can be similar those in abundance even when numbers of observations are low. Recorder behaviour may have biased our results, as recorders may not record widespread common species on an ad hoc basis instead favouring notable records (e.g. rare species), this contrasts the abundance data that were collected following a structured survey design where all species seen are recorded. This could lead to mismatch in abundance and distribution patterns even for inter-annual changes, as recording effort varies temporally. Finally, the lowest total number of distribution records in our study was quite high (see Table 1), therefore the concerns with low sample size were not an issue here. Nevertheless, the issue may be important to other more poorly recorded taxonomic groups.

Fractal dimension of species' distribution also did not impact the inter-annual distribution-abundance relationship. This might be because even if a range is fragmented, distribution recorders and transect volunteers still find and document species in those locations. In addition, if a species is known to be fragmented (which usually indicates rareness or being at risk of extinction), there may be a recording bias towards it (Isaac & Pocock, 2015), which results in good information for that species. Therefore, species with a high fractal dimension may still have a positive inter-annual distribution-abundance relationship. Although it should be noted that species which are very poorly studied, and therefore likely rare and in fragmented habitats, were not been included due to the selection criteria. The study species also had ranges which were relatively well-filled, with fractal dimension scores ranging from 0.257 to 0.716 (maximum possible value is 2). It is possible that fractal dimension is an important factor for highly fragmented species, and there may have been insufficient variation in this attribute to be important to the inter-annual distribution-abundance relationship. Similarly we found no relationship between the inter-annual distribution-abundance relationship and dispersal for butterflies (Table S2). If these variables lack significant explanatory power even for a well-studied taxon, then this suggests that they will have limited use for identifying species in other taxa for which our method may be appropriate.

Population synchrony and inter-annual distribution-abundance relationships

The success of year-to-year changes in distribution records mirroring abundance changes in migratory species suggests that population synchrony over large areas may play a role, and so we examined the impact of scale on the inter-annual distribution-abundance relationship by comparing national and county-level analyses. Weak relationships at the national level may occur if species' population dynamics are asynchronous, such that abundances and distributions may be closely linked locally, but a 'good year' in one region might occur when it is a 'bad year' in another region, obscuring any overall pattern at a national scale. Yet, when we repeated our national-scale analysis for a much smaller region (the county of Dorset), the results were similar: goodness of fit scores across species for the inter-annual distribution-abundance relationship for Dorset were correlated with those for the same species at the national level (Fig. 4). The majority of species had lower r^2 values for the regional analyses, probably due to reduced data quantity. The spatial scales at which abundance and distribution changes are linked deserve more attention, but our preliminary conclusion is that reducing the extent of the study region considered does not

improve the inter-annual distribution-abundance relationship.

Conclusions

Our key finding that (mean year-to-year changes in) citizen-collected distribution data can provide useful information on population variability suggests that it may be possible to expand our methods to other taxonomic groups, or to populations of butterflies in countries that do not have standardised population monitoring schemes. Such measures of variability can inform habitat, landscape and regional conservation decision-making. The use of distribution data for more detailed analyses of inter-annual population change is only likely to be possible, however, for species that have highly variable numbers of records between years. For these species, it may be possible to analyse year-to-year population changes across much longer time periods than covered by transect data and hence identify how populations are influenced by the effects of specific weather variables, density dependence, and any other process that operates at a large geographical and temporal scale. Further investigation is required, however, in the feasibility of extending these methods to other taxonomic groups.

Author contributions

SCM, GDP, THO, CDT, JKH designed the study, SCM carried out all analyses and calculations, with the exception of national UKBMS indices. THO provided code and data for regional UKBMS indices, RF provided species information for exclusion, SCM initiated the writing of the manuscript and all authors commented on and helped write the paper.

Acknowledgements

Data accessibility

The data for the butterfly species that were used in this study are archived by the National Biodiversity Network Gateway https://data.nbn.org.uk/ (BNM presence records) and the UKBMS http://www.ukbms.org (UKBMS transect count data and TRMOBs index for 1976-2012).

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: doi: 10.1111/icad.12242:

Figure S1. The ranges of a) the small skipper butterfly, *Thymelicus sylvestris,* a species with a well-filled range (fractal dimension: 0.257), and b) the Grayling butterfly, *Hipparchia semele* a species with the most minimally filled range of the 33 species studied (fractal dimension: 0.716).

Figure S2. The annual total number of distribution records for all 33 study butterfly species across the study period 1972–2012.

Table S1. Results of the best model explaining the inter-annual distribution-abundance relationship with two outlying species, *Celastrina argiolus* and *Vanessa cardui* removed from the analysis. One biogeographical attribute was included as an explanatory variable: mean absolute year-to-year change in distribution records. PGLS model results: $r^2 = 0.43$, $F_{1,28} = 20.86$, AIC = −31.11, $P = <0.001$.

Table S2. Results of two models examining the relationship between dispersal ability and the inter-annual distribution-abundance relationship. The explanatory variable in Model 1 is a dispersal ranking from Cowley *et al.* (2001), PGLS model results: $r^2 = 0.13$, $F_{1,26} = 3.776$, AIC = −17.52, $P = 0.063$; and the explanatory variable in Model 2 is a dispersal score from Dennis *et al.* (2004) PGLS model results: $r^2 = 0.08$, $F_{1,26} = 2.119$, AIC = −15.78, $P = <0.157$.

References

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. & Mace, G.M. (2010) Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, **8**, e1000385.

Brereton, T. (1998) Pearl-bordered fritillary national survey: 1998 update. *Butterfly Conservation News*, **69**, 5.

Brereton, T., Roy, D.B., Middlebrook, I., Botham, M. & Warren, M. (2011) The development of butterfly indicators in the United Kingdom and assessments in 2010. *Journal of Insect Conservation*, **15**, 139–151.

Brooks, S.J., Self, A., Powney, G.D., Pearse, W.D., Penn, M. & Paterson, G.L.J. (2016) The influence of life history traits on the phenological response of British butterflies to climate variability since the late-19th century. *Ecography*, https://doi.org/10.1111/ecog.02658.

Brown, J.H. (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, **124**, 255.

Brown, J.H., Mehlman, D.W. & Stevens, G.C. (1995) Spatial variation in abundance. *Ecology*, **76**, 2028–2043.

Cowley, M.J., Thomas, C.D. & Roy, D.B. (2001) Density-distribution relationships in British butterflies. I. The effect of mobility and spatial scale. *Journal of Animal Ecology*, **70**, 410–425.

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2003) Infectious disease and amphibian population declines. *Diversity and Distributions*, **9**, 141–150.

Dennis, R.L.H., Hodgson, J.G., Grenyer, R., Shreeve, T.G. & Roy, D.B. (2004) Host plants and butterfly biology. Do host-plant strategies drive butterfly status? *Ecological Entomology*, **29**, 12–26.

Devictor, V., Whittaker, R.J. & Beltrame, C. (2010) Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, **16**, 354–362.

Fox, R., Brereton, T.M., Asher, J., August, T.A., Botham, M.S., Bourn, N.A.D., Cruickshanks, K.L., Bulman, C.R., Ellis, S., Harrower, C.A., Middlebrook, I., Noble, D.G., Powney, G.D., Randle, Z., Warren, M.S. & Roy, D.B. (2015) *The State of the UK's Butterflies 2015*. Butterfly Conservation, Wareham, Dorset.

Gandiwa, E., Heitkonig, I.M.A., Eilers, P.H.C. & Prins, H.H.T. (2016) Rainfall variability and its impact on large mammal populations in a complex of semi-arid African savanna protected areas. *Tropical Ecology*, **57**, 163–180.

Gaston, K.J. (1996) The multiple forms of the interspecific abundance-distribution relationship. *Oikos*, **76**, 211–220.

Gaston, K.J., Blackburn, T.M., Greenwood, J.D., Gregory, R.D., Quinn, R.M. & Lawton, J.H. (2000) Abundance-occupancy relationships. *Journal of Applied Ecology*, **37**, 39–59.

Hartley, S. (1998) A positive relationship between local abundance and regional occupancy is almost inevitable (but not all positive relationships are the same). *Journal of Animal Ecology*, **67**, 992–994.

Hayhow, D.B., Burns, F., Eaton, M.A., Al Fulaij, N., August, T.A., Babey, L., Bacon, L., Bingham, C., Boswell, J., Boughey, K.L., Brereton, T., Brookman, E., Brooks, D.R., Bullock, D.J., Burke, O., Collis, M., Corbet, L., Cornish, N., De Massimi, S., Densham, J., Dunn, E., Elliott, S., Gent, T., Godber, J., Hamilton, S., Havery, S., Hawkins, S., Henney, J., Holmes, K., Hutchinson, N., Isaac, N.J.B., Johns, D., Macadam, C.R., Mathews, F., Nicolet, P., Noble, D.G., Outhwaite, C.L., Powney, G.D., Richardson, P., Roy, D.B., Sims, D., Smart, S., Stevenson, K., Stroud, R.A., Walker, K.J., Webb, J.R., Webb, T.J., Wynde, R. & Gregory, R.D. (2016) *State of Nature 2016*. The State of Nature partnership.

Holt, R.D., Lawton, J.H., Gaston, K.J. & Blackburn, T.M. (1997) On the relationship between range size and local abundances: back to the basics. *Oikos*, **78**, 183–190.

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D. & Willis, S.G. (2015) The drivers of avian abundance: patterns in the relative importance of climate and land use. *Global Ecology and Biogeography*, **24**, 1249–1260.

Inchausti, P. & Halley, J. (2003) On the relation between temporal variability and time in animal populations persistence. *Journal of Animal Ecology*, **72**, 899–908.

Isaac, N.J.B. & Pocock, M.J.O. (2015) Bias and information in biological records. *Biological Journal of the Linnean Society*, **115**, 522–531.

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T. & Donald, P.F. (2016) Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, **22**, 1–12.

Lemoine, N., Bauer, H.G., Peintinger, M. & Bohning-Gaese, K. (2007) Effects of climate and land-use change on species abundance in a central European bird community. *Conservation Biology*, **21**, 495–503.

Lewis, J. & Vandewoude, S. (2015) The effects of urbanization on population density, occupancy, and detection probability of wild felids. *Ecological Applications*, **25**, 1880–1895.

Lynam, C.P., Hay, S.J. & Brierley, A.S. (2004) Interannual variability in abundance of North Sea jellyfish and links to the North Atlantic Oscillation. *Limnology and Oceanography*, **49**, 637–643.

Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akçakaya, H.R., Leader-Williams, N., Milner-Gulland, E.J. & Stuart, S.N. (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, **22**, 1424–1442.

Maclean, I.M.D., Austin, G.E., Rehfisch, M.M., Blew, J., Crowe, O., Delany, S., Devos, K., Deceuninck, B., Günther, K., Laursen, K., Van Roomen, M. & Wahl, J. (2008) Climate change causes rapid changes in the distribution and site abundance of birds in winter. *Global Change Biology*, **14**, 2489–2500.

Malinowska, A.H., Strien, A.J.Van, Verboom, J. & Wallisdevries, M.F. (2014) No evidence of the effect of extreme weather events on annual occurrence of four groups of ectothermic species. *PLoS ONE*, **9**, e110219.

Meyer, S., Robertson, B.C., Chilvers, B.L. & Krkošek, M. (2015) Population dynamics reveal conservation priorities of the threatened New Zealand sea lion Phocarctos hookeri. *Marine Biology*, **162**, 1587–1596.

Oliver, T.H., Gillings, S., Girardello, M., Rapacciuolo, G., Brereton, T.M., Siriwardena, G.M., Roy, D.B., Pywell, R. & Fuller, R.J. (2012) Population density but not stability can be predicted from species distribution models. *Journal of Applied Ecology*, **49**, 581–590.

Oliver, T.H., Isaac, N.J.B., August, T.A., Woodcock, B.A., Roy, D.B. & Bullock, J.M. (2015a) Declining resilience of ecosystem functions under biodiversity loss. *Nature Communications*, **6**, 10122.

Oliver, T.H., Marshall, H.H., Morecroft, M.D., Brereton, T., Prudhomme, C. & Huntingford, C. (2015b) Interacting effects of climate change and habitat fragmentation on drought-sensitive butterflies. *Nature Climate Change*, **5**, 1–6.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N. & Pearse, W. (2013) *caper: Comparative Analyses of Phylogenetics and Evolution in R*. R package version 0.5.2. < https://CRAN. R-project.org/package = caper > 21st April 2017.

Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Pearce, J.L. & Boyce, M.S. (2006) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, **43**, 405–412.

Petrovskii, S., Petrovskaya, N. & Bearup, D. (2014) Multiscale approach to pest insect monitoring: random walks, pattern formation, synchronization, and networks. *Physics of Life Reviews*, **11**, 467–525.

Pimm, S.L., Jones, H.L. & Diamond, J. (1998) On the risk of extinction. *The American Naturalist*, **132**, 757–785.

Pocock, M.J.O., Roy, H.E., Preston, C.D. & Roy, D.B. (2015) The biological records centre: a pioneer of citizen science. *Biological Journal of the Linnean Society*, **115**, 475–493.

Robertson, G.S., Bolton, M., Morrison, P. & Monaghan, P. (2015) Variation in population synchrony in a multi-species seabird community: response to changes in predator abundance. *PLoS ONE*, **10**, 1–15.

Robertson, P.A., Woodburn, M.I.A. & Hill, D.A. (1988) The effects of woodland management for pheasants on the abundance of butterflies in Dorset, England. *Biological Conservation*, **45**, 159–167.

Roney, N.E., Kuparinen, A. & Hutchings, J.A. (2015) Comparative analysis of abundance-occupancy relationships for species-at-risk at both broad taxonomic and spatial scales. *Canadian Journal of Zoology*, **93**, 515–519.

Rothery, P. & Roy, D.B. (2001) Application of generalized additive models to butterfly transect count data. *Journal of Applied Statistics*, **28**, 897–909.

Roy, D.B., Rothery, P., Moss, D., Pollard, E. & Thomas, J.A. (2001) Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. *Journal of Animal Ecology*, **70**, 201–217.

Sánchez-Hernández, J., Cobo, F. & Amundsen, P.A. (2015) A simple approach for assigning the conservation status of freshwater fish based on their natural variability - a case study from the Iberian Ecoregion. *Fisheries Management and Ecology*, **22**, 125–133.

van Schaik, C.P. & van Noordwijk, M.A. (1985) Interannual variability in fruit abundance and the reproductive seasonality in Sumatran Long-tailed macaques (Macaca fascularis). *Journal of Zoology*, **206**, 533–549.

Sutcliffe, O.L., Thomas, C.D. & Moss, D. (1996) Spatial synchrony and asynchrony in butterfly population dynamics. *Journal of Animal Ecology*, **65**, 85–95.

Tempel, D.J. & Gutiérrez, R.J. (2013) Relation between occupancy and abundance for a territorial species, the california spotted owl. *Conservation Biology*, **27**, 1087–1095.

Thomas, J. A. (2010) *Butterflies of Britain and Ireland*. British Wildlife Publishing, Gillingham.

Thomas, C.D., Bodsworth, E.J., Wilson, R.J., Simmons, A.D., Davies, Z.G. & Conradt, L. (2001) *Ecological and Evolutionary Processes at Expanding Range Margins*, **411**, 577–581.

Thomas, A.C.D. & Jones, T.M. (1993) Partial recovery of a skipper butterfly (Hesperia comma) from population refuges: lessons for conservation in a fragmented landscape. *Journal of Animal Ecology*, **62**, 472–481.

Vucetich, J.A., Waite, T.A., Qvarnemark, L. & Ibargüen, S. (2000) Population variability and extinction risk. *Conservation Biology*, **14**, 1704–1714.

Warren, M.S., Hill, J.K., Thomas, J.A., Asher, J., Fox, R., Huntley, B., Roy, D.B., Telfer, M.G., Jeffcoate, S., Harding, P., Jeffcoate, G., Willis, S.G., Greatorex-Davies, J.N., Moss, D. & Thomas, C.D. (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, **414**, 65–69.

Wilson, J.W., Thomas, C.D., Fox, R., Roy, D.B. & Kunin, W.E. (2004) Spatial patterns in species distribution reveal biodiversity change. *Nature*, **432**, 393–396.

Zuckerberg, B., Porter, W.F. & Corwin, K. (2009) The consistency and stability of abundance–occupancy relationships in large-scale population dynamics. *Journal of Animal Ecology*, **78**, 172–181.