Article (refereed) - postprint

Pagel, Jörn; Anderson, Barbara J.; O'Hara, Robert B.; Cramer, Wolfgang; Fox, Richard; Jeltsch, Florian; Roy, David B.; Thomas, Chris D.; Schurr, Frank M. 2014. **Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records**. *Methods in Ecology and Evolution*, 5 (8). 751-760. 10.1111/2041-210X.12221

Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records

Jörn Pagel[1,2,*], Barbara J. Anderson[3,4], Robert B. O'Hara[5], Wolfgang Cramer[6], Richard Fox[7], Florian Jeltsch[1], David B. Roy[8], Chris D. Thomas[4], Frank M. Schurr[2,9]

[1]Plant Ecology and Conservation Biology, University of Potsdam, 14469, Potsdam, Germany

[2]Institut des Sciences de l'Evolution, UMR 5554, Université Montpellier 2, Montpellier, cedex 05, France

[3]Rutherford Discovery Fellow, Landcare Research, Private Bag 1930, Dunedin 9054, New Zealand

[4]Department of Biology (Area 18), University of York, York, YO10 5YW, UK

[5]Biodiversity and Climate Change Research Centre (BIK-F), 60325, Frankfurt am Main, Germany

[6]Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), Aix Marseille Université, CNRS, IRD, Avignon Université, Aix-en-Provence, cedex 04, France

[7]Butterfly Conservation, Manor Yard, East Lulworth, Dorset, BH20 5QP, UK

[8]NERC Centre for Ecology & Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK

[9]Landscape Ecology and Vegetation Science, University of Hohenheim, 70599, Stuttgart, Germany

*Correspondence author. E-mail: jpagel@uni-potsdam.de

Word count: 7286 (main text, acknowledgements, references, tables and figure captions)

Running title: Range-wide variation in abundance trends

SUMMARY

1. Species' abundances vary in space and time. Describing these patterns is a cornerstone of macroecology. Moreover, trends in population size are an important criterion for the assessment of a species' conservation status. Since abundance trends are not homogeneous in space, we need to quantify variation in abundance trends across the geographical range of a species. A basic difficulty exists in that data sets that cover large geographic areas rarely include population abundance data at high temporal resolution. Whilst both broad-scale geographic distribution data and site specific population trend data are becoming more widely available, approaches are required that integrate these different types of data.

2. We present a hierarchical model that integrates observations from multiple sources to estimate spatio-temporal abundance trends. The model links annual population densities on a spatial grid to both long-term count data and to opportunistic occurrence records from a citizen-science programme. Specific observation models for both data types explicitly account for differences in data structure and quality.

3. We test this novel method in a virtual study with simulated data and apply it to the estimation of abundance dynamics across the range of a butterfly species (*Pyronia tithonus*) in Great Britain between 1985 and 2004. The application to simulated and real data demonstrates how the hierarchical model structure accommodates various sources of uncertainty that occur at different stages of the link between observational data and the modelled abundance. Thereby, it accounts for these uncertainties in the inference of abundance variations.

4. We show that by using hierarchical observation models that integrate different types of commonly available data sources we can improve estimates of variation in species abundances across space and time. This will improve our ability to detect regional trends and can also enhance the empirical basis for understanding range dynamics.

Key words: atlas data, Bayesian statistics, biogeography, butterflies, citizen science program, conservation biology, count data, macroecology, state-space model

INTRODUCTION

Species distribution data are of central importance to ecology. Analysing spatial patterns of species' occurrence is the natural first step of studies that assess global change impacts on biodiversity and design conservation strategies (Dawson *et al.* 2011). Including the temporal dimension in macro-ecological data is critical to the development of macro-ecology as a predictive science (Fisher *et al.* 2010). Indeed, we need data on the spatio-temporal variation of not only occurrence but also abundance in order to understand the population demographics that underlie species niches and range dynamics (Schurr *et al.* 2012) and conservation biogeography (Whittaker *et al.* 2005). In particular, the detection of abundance trends is an important component of assessing the conservation status of species according to Red List criteria (IUCN 2011). Since threats are not equally distributed across the geographical range of species, and conservation actions are commonly deployed within administrative units rather than globally, we need to quantify

abundance trends in different parts of the range. However, standardised monitoring data, from which abundance trends can be inferred directly, rarely have sufficient spatial and temporal coverage (Whittaker *et al.* 2005).

A promising way to overcome data restrictions is the combination of different data types from various sources that contain information on the occurrence and abundance of a species across space and time (Scholes *et al.* 2008). In fact, recent initiatives like the Global Biodiversity Observation Network of the Group on Earth Observations (GEO BON) explicitly call for a 'hierarchical sampling approach' that combines large amounts of relatively simple data, like occurrence records, with more extensive data, like systematic abundance surveys (Scholes *et al.* 2008). Citizen science programs that provide geographically explicit data on large spatial and temporal scales can be particular valuable for the assessment of biodiversity trends (Devictor *et al.* 2010). Many of these programs deliver haphazardly collected species lists (Roberts *et al.* 2007) where volunteer recorders report the detection/non-detection of certain species from a target group at occasional and irregular site visits. These records are characterized by an uneven geographical and temporal distribution of surveys, non-standardized observer efforts per site visit and possible biases in species' reporting and detection which has to be carefully dealt with in order to avoid biased trend estimates (van Strien et al. 2013). The use of these data for the estimation of abundance trends and their combination with data from other sources requires flexible statistical models that explicitly account for differences in data structure and quality and that can handle and quantify the sources of uncertainty associated with each data type.

Recently, a range of hierarchical statistical modelling approaches have been developed that explicitly distinguish the data-generating observation processes from the processes that drive the variation of ecological state variables such as occurrence and abundance (Royle & Dorazio 2008): Occupancy models that estimate species occurrence from presence-absence data with imperfect detection (McKenzie *et al.* 2006) are now common and have been successfully applied to opportunistic detection/non-detection data from citizen science programs (e.g. Kery *et al.* 2010, van Strien *et al.* 2013). Analogous hierarchical models have been used to estimate of abundance form imperfect count surveys (e.g. Royle & Dorazio 2006, Royle *et al.* 2007, Kery *et al.* 2009). Some authors have argued for an advantage of using abundance as a state variable even for the analysis of presence-absence data, since variation in abundance is likely the most important cause of heterogeneous species detectability in presence-absence surveys (Royle & Nichols 2003, Dorazio 2007). Conroy *et al.* (2008) have shown how a functional relationship between abundance and detectability can be estimated by combining (repeated) presence-absence records with capture-mark-recapture data for the same locations and how this relationship can consequently be used to predict local abundance at sites with only presence-absence records.

In this study we advance this approach in order to estimate spatial variation in abundance trends from a combination of widespread opportunistic occurrence records and local abundance surveys. The presented method links both detection/non-detection data and count data to a spatially explicit state-space model of abundance variation. The hierarchical model thereby infers a relationship between the abundance state variable and detectability and in turn allows the detection/non-detection data to inform the estimation

of abundance trends. We (1) assess the reliability of this method by testing it in simulated data scenarios, and (2) demonstrate its application in a case study that estimates abundance trends for of a butterfly species (the Gatekeeper, *Pyronia tithonus*) in Great Britain.

MATERIALS AND METHODS

A hierarchical model of abundance variation

The model aims to estimate population densities of a focal species in all cells *i* of a regular spatial grid and in all years *t* within an observation time period. We generally consider two types of data: (i) standardized count surveys within a certain subarea of a grid cell and (ii) opportunistic occurrence records that can be geographically referenced to a grid cell and report the detection/non-detection of the focal species. Count data will typically only be available for a small subset of grid cells and occurrence records will come from a highly variable number of recorder visits per cell and year (including zero visits).

For the estimation of abundance variation from these heterogeneous data, the hierarchical model integrates specific observation models for both data types with a state-space model that describes the spatio-temporal variation of population densities $\mathbf{\Lambda}$ (Fig. 1). For the basic model concept presented here, we do not consider the effect of environmental covariates on the observation processes or the species' population density, but we will address their potential inclusion in the discussion. In the following, we specify basic observation models that describe the likelihood of either count data or

detection/non-detection records conditional on the population density **Λ**, as well as the state-space model of **Λ**. As outlined below, we will interpret **Λ** as a relative measure of population density rather than an absolute measure.

Modelling count data

We assume a single number of individuals $y_{j,t}$ that was counted at a survey site j located within grid cell i of size A. Survey sites may differ in their area a_j. A reasonable starting point for modeling count data is a Poisson distribution with a rate parameter λ that is proportional to the sampled area a_j and the total number of individuals $N_{i,t}$ in the grid cell. Yet, a common feature of count data is that the sample variance is larger than assumed for a Poisson distribution (i.e. larger than the mean). This overdispersion may arise from various factors, including a spatially aggregated distribution of individuals and sampling variability (Linden & Mantyniemi 2011, Kotze et al. 2012). Without aiming to resolve individual factors that contribute to overdispersion and for technical convenience we model count data $y_{j,t}$ by a mixed lognormal-Poisson distribution

$$y_{j,t} \sim Poisson(\lambda_{j,t}) \qquad \text{(eqn. 1)}$$

$$\lambda_{j,t} \sim LogNormal(\mu_{j,t}, \sigma_\lambda^2)$$

which is parameterized via the log-scale mean $\mu_{j,t} = a_j \cdot \Lambda i_{,t}$, so that the median count is the product of a relative measure of population density $\Lambda_{i,t}$ and the sampled area a_j. σ_λ^2 is an estimated variance parameter that represents overdispersion with respect to the Poisson distribution.

We chose to link the count data to a relative measure of population density, because the estimation of absolute population sizes would rely on estimating the rate ω at which individuals are detected. Robust methods to estimate ω have been proposed for cases where a closed (i.e. constant during the sampling period) population is counted repeatedly (Royle *et al.* 2007, Royle & Dorazio 2008). Since here we do not assume that data from repeated counts are available, we instead normalize the estimated population density by the (unknown) average per individual detection rate E[ω]. From the expected count at a survey site $E[y_{j,t}] = (a_j/A)\cdot N_{i,t}\cdot E[\omega]$, it follows that the estimated relative population density is proportional to population size:

$$\Lambda_{i,t} = \frac{N_{i,t}}{A} \frac{E[\omega]}{\exp(\sigma_\lambda^2/2)} \qquad \text{(eqn. 2)}$$

Due to this normalization the observation model does not explicitly account for variation in ω. Yet, the inference of trends from estimated relative population densities Λ does not assume that detection rates are constant, but just that there is no distinct spatial or temporal pattern to this variation (Link & Sauer 1998).

Modelling detection/non-detection data

Detection/non-detection histories for each cell *i* and year *t* consist of the total number of recorder visits $J_{i,t}$ and the respective number of visits $x_{i,t}$ that report the focal species' presence. We model this as a binomial process $x_{i,t} \sim Binomial(J_{i,t}, \psi_{i,t})$ with *detectability* $\psi_{i,t}$ denoting the per-visit probability to record a presence. To formulate a likelihood of detection/non-detection data conditional on the state variable Λ we describe a functional

relationship between detectability $\psi_{i,t}$ and population density (hereafter called *density-detectability-curve*). A basic model of the dependence of ψ on the population size N can be derived from a binomial model for the number of detected individuals per visit $n \sim Binomial(N, r)$ with per-individual detection probability r (Royle & Nichols 2003). The detectability, i.e. the probability to encounter at least one individual during a visit, is then $\Pr(n = 1) = 1 - (1 - r)^N$. An equivalent formulation of this relationship with respect to relative population density Λ is

$$\psi_{i,t} = \Pr(n = 1) = 1 - \exp(-\alpha \cdot \Lambda_{i,t}) \qquad \text{(eqn. 3)}$$

with the new parameter $\alpha = -\ln(1 - r) \cdot N/\Lambda$. The saturation rate α describes how fast detectability approaches one when population density Λ increases and can be interpreted as a relative measure of sampling effort that is scaled by the proportionality between Λ and absolute abundance N (cf. eqn. 2).

However, the binomial model makes the assumption that all individuals of a cell were independently detected (McCarthy *et al.* 2013) and does not account for additional variation in detectability. In practice, $\psi_{i,t}$ may vary widely due to factors like weather and habitat conditions as well as observer effort and skill. In order to reflect additional sources of uncertainty, we modify the detectability by a multiplicative random term $\varphi_{i,t}$ so that $\psi_{i,t} = \varphi_{i,t}(1 - \exp(-\alpha \cdot \Lambda_{i,t}))$. Furthermore, we allow also $\varphi_{i,t}$ to depend on population densities and formulate a linear regression of $\text{logit}(\varphi_{i,t})$ on $\ln(\Lambda_{i,t})$. This specific functional form of the density-detectability-curve was motivated by a preliminary data analysis for our case study, which we discus in Appendix S1 in the Supporting Information.

Altogether, the probability of obtaining $x_{i,t}$ presence records from $J_{i,t}$ recorder visits, conditional on a relative population density $\Lambda_{i,t}$, saturation rate α and regression parameters β_0, β_1, and σ_φ^2, is

$$\Pr\left(x_{i,t} = k \middle| J_{i,t}, \Lambda_{i,t}, \alpha, \beta_0, \beta_1, \sigma_\varphi^2\right) = \binom{J_{i,t}}{k} \psi_{i,t}^{k} \left(1 - \psi_{i,t}\right)^{J_{i,t}-k} \qquad \text{(eqn. 4)}$$

$$\psi_{i,t} = \varphi_{i,t} \left(1 - \exp(-\alpha \cdot \Lambda_{i,t})\right)$$

$$\text{logit}(\varphi_{i,t}) = \beta_0 + \beta_1 \cdot \ln(\Lambda_{i,t}) + \varepsilon_{i,t}$$

$$\varepsilon_{i,t} \sim \textit{Normal}\ (0, \sigma_\varphi^2)$$

Modelling population density

Spatio-temporal variation of relative population density $\mathbf{\Lambda}$ is modelled by a lognormal distribution. We account for zero-inflation due to complete absence in parts of the study area by introducing an indicator variable $\boldsymbol{I}$ that denotes species presence. Thus, variation of $\Lambda_{i,t}$ is described by a zero-inflated lognormal distribution, where $\Pr(I_{i,t} = 1)$ is the occurrence probability and $\mu_{i,t}$ and σ^2 are the log-scale mean and variance of a lognormal distribution of $\Lambda_{i,t}$ conditional on the species being present ($I_{i,t} = 1$):

$$\Pr\left(\Lambda_{i,t} = x \middle| I_{i,t}, \mu_{i,t}, \sigma\right) = \begin{cases} \Pr\left(I_{i,t} = 0\right) & x = 0 \\ \Pr\left(I_{i,t} = 1\right) \dfrac{1}{\sqrt{2\pi\sigma x}} \exp\left(\dfrac{\left(\ln x - \mu_{i,t}\right)^2}{2\sigma^2}\right) & x > 0 \end{cases} \qquad \text{(eqn. 5)}$$

Models for both occurrence probability $\Pr(I_{i,t} = 1)$ and log-scale mean density $\mu_{i,t}$ include spatially autocorrelated random effects as well as temporal random effects on annual mean incidence and density:

$$\text{logit}[\Pr(I_{i,t} = 1)] = \mu_{Inc} + \Delta_{i,t} + \varepsilon.Inc_t \qquad \text{(eqn. 6)}$$

$$\mu_{i,t} = \mu_D + \rho\Delta_{i,t} + \varepsilon.D_t \qquad \text{(eqn. 7)}$$

The temporal random effects $\varepsilon.Inc$ and $\varepsilon.D$ are normally distributed with zero mean and variances σ_{Inc}^2 and σ_D^2, respectively. For spatial effects $\Delta_{i,t}$ we do not consider environmental covariates but include a year-specific parabolic effect of geographical latitude L:

$$\Delta_{i,t} = \gamma_{1,t} \cdot L_i + \gamma_{2,t} \cdot L_i^2 + \delta_{i,t} \qquad \text{(eqn. 8)}$$

The motivation for this latitudinal effect is mainly to constrain estimates of $\mathbf{\Lambda}$ for poorly sampled areas beyond the species' range (cf. Fig. 3). For the spatially auto-correlated random effects $\delta_{i,t}$ we use an intrinsic conditionally autoregressive (CAR) model (Besag *et al.* 1991):

$$\delta_{i,t} | \boldsymbol{\delta}_{-i,t} \sim Normal(\delta.\text{bar}_{i,t}, \nu/n_i) \qquad \text{(eqn. 9)}$$

where $\delta.\text{bar}_{i,t}$ is the mean $\Sigma_j \delta_{j,t}/n_i$ over all n_i cells that are adjacent to i. Note that, conditional on the variance parameter ν which is constant across years, random effects $\boldsymbol{\delta}_t$ in different years t are independent of each other.

Simulation study

We conducted a simulation study to test the performance of the presented model for a range of data scenarios. On an artificial landscape grid of 50×50 cells we simulated dynamic abundance patterns in a changing environment. In a 'Virtual Ecologist' approach (Zurell et al. 2010) we then sampled observation data from the simulated abundance patterns, used these data to estimate the spatio-temporal variation in population densities and assessed model estimates based on the known, simulated population dynamics. Imperfect count and detection/non-detection data were drawn from probability distributions (conditional on the simulated population density) as specified by the respective observation models above. For count data we deliberately set the proportionality factor between simulated abundances and estimated relative abundances (cf. eqn. 2) to one, in order to facilitate the comparison of true and estimated population densities. The sampling scheme for the *standard* data scenario covers an observation period of 20 years and was designed to mimic the data availability of the butterfly case study (see below) with respect to the total amount as well as the temporal and spatial heterogeneity of both data types (see Appendix S2 for details on the simulation of abundance and virtual data). We also created a set of reduced data scenarios by shortening the observation period to the last ten years and/or by reducing the number of sites with count data to either 10% or 25% of the number in the standard scenario.

Case study

The Gatekeeper (*Pyronia tithonus*, sometimes called Hedge Brown) is a common butterfly species in sheltered grasslands of England and Wales that has expanded its range northwards in recent years (Mair *et al.* 2012). For the estimation of abundance trends across its range in Great Britain, we used two data sets: (i) presence records from the Butterflies for the New Millennium (BNM) project and (ii) transect count data from the UK Butterfly Monitoring Scheme (UKBMS). Based on these data, we estimated relative population densities for the 2689 cells of a 10 km (i.e. cells of area 100 km²) UTM grid across Great Britain and for all years from 1985 to 2004.

UKBMS abundance indices

The UKBMS is a long-term monitoring program that conducts systematic counts of butterflies in a standardized survey on permanent transects in the United Kingdom. In each of 26 weeks from the beginning of April until the end of September surveyors count all butterflies within a 5 m wide corridor around each transect. Pollard & Yates (1993) provide further details and validation of the sampling scheme. The length of individual transects varies between 1.5 and 3 km and hence the specific surveyed area a_j differs between transects. Since surveys require suitable weather conditions, the precise timing of the weekly counts is irregular and occasional weeks are missing for a given transect. For our analysis we use an annually aggregated index of butterfly abundance (*IBA*) that interpolates between temporally irregular samples. For each transect and year this index is calculated from the series of all counts $n_1, n_2, ..., n_T$ at days $d_1, d_2, ..., d_T$ as

$IBA = \sum n_k(d_{k+1} - d_{k-1})/2$ (see Rothery & Roy 2001 and Dennis *et al.* 2013 for a discussion of alternative indices). The *IBA* represents for each transect the cumulative counts throughout one season and thereby integrates over the phenology. Consequently, the *IBA* does not enable us to directly estimate the absolute abundance of butterfly individuals. Instead we use it to estimate butterfly days, i.e., the expected total number to have been counted if a transect had been sampled every day. Since our case study directly uses the *IBA* as count data y (eqn. 1), the modelled relative densities $\mathbf{\Lambda}$ are likewise defined relative to butterfly-days per year (see the discussion for implications this has for trend estimation). The number of transects from which data were available varied between 80 and 151 per year, with a median of 124.

BNM *occurrence records*

Extensive data on the occurrence of butterfly species across Great Britain were collected in the Butterflies for the New Millennium project (BNM, Asher *et al.* 2001). The underlying raw data, which we use in our analysis, consist of record cards that were submitted mainly by volunteer observers since 1970. These records list all species observed at one field visit and typically originate from opportunistic rather than systematic recording (Asher *et al.* 2001). However, records occasionally pool observations for a whole month or year. We excluded any record that could not be assigned to a single visit of a site. As a simple treatment to account for selective recording of rare and interesting species, we also removed all records which report only one species (van Strien *et al.* 2013). The remaining data for the years 1985–2004

comprise a total of 510,209 recorder visits. Aggregation of the recorder data to the 100 km² grid then gives for each cell the total number of times J at which the cell has been visited within a year and the respective number of visits x that recorded a presence (cf. eqn. 4). On average about half of all grid cells (1408) were visited at least once each year and the number of visits per year in these cells varied between 1 and 555, with a median of 6 (Fig 3a).

Bayesian parameter estimation

For both the simulation study and the butterfly case study, estimates of all parameters in the hierarchical model, including spatio-temporal density estimates $\mathbf{\Lambda}$, were generated by a Markov chain Monte Carlo (MCMC) algorithm. We used OpenBUGS (version 3.2.1, Lunn *et al.* 2009) and ran three independent MCMC chains with 100,000 iterations each, the first 75,000 of which were discarded as burn-in. Convergence of the MCMC sampler after the burn-in period was checked by calculating the multivariate scale reduction factor of Gelman & Rubin (1992). Samples of the high-dimensional state vector $\mathbf{\Lambda}$ were only stored for every 50^{th} iteration in view of memory limitations (Link & Eaton 2012). Computation times for the different data scenarios were 18–35 hours per MCMC chain (Intel i5 2.4 GHz CPU). Details of model estimation and the OpenBUGS code are presented in Appendix S3.

RESULTS

Simulation study

For the simulation study we tested the model's capability of estimating the relationship between population density and occurrence-detectability form the combination of both data sets and using this relationship to estimate population densities also in grid cells for which no count data was available. Results for the *standard* data scenario show that the estimated density-detectability-curve, as predicted from posterior estimates of parameters α, β_0 and β_1 (cf. eqn. 4), matches well the 'true' occurrence-detectability-curve applied in sampling of the simulation data (Fig. 2a). Evaluating estimated population densities against the simulated 'true' population densities shows no systematic under- or overestimation across the range of simulated population densities, but a decreased precision for grid cells where occurrence data stems from only a few reorder visits per year (Fig. 2b).

In order to compare the accuracy of these estimates across the different data scenarios, we calculated the predictive deviance for population densities in all grid cells in the last 5 years of the observation period. We summarized the posterior sample for each population density $\Lambda_{i,t}$ by the three parameters of a zero-inflated lognormal distribution (cf. eqn. 5): the mean incidence $I_{i,t}$, i.e. the fraction of non-zero samples $\Lambda_{i,t} > 0$, and the log-scale mean $\mu_{i,t}$ and variance $\sigma^2_{i,t}$ of a lognormal distribution fitted to all non-zero posterior samples. The predictive deviance is then calculated from the likelihood of the true population density $\Lambda^*_{i,t}$ under the posterior distribution as $-2 \cdot \ln[Pr(\Lambda^*_{i,t} \mid \mu_{i,t}, \sigma^2_{i,t}, I_{i,t})]$. To further investigate the relationship between accuracy of

model estimates and the number of recorder visits per site and year (#visits) we then compared, for each data scenario, the mean predictive deviance across grid cells that had no count data but different #visits (Fig. 3c). For all data scenarios the accuracy of estimated population density increases with #visits, particularly for grid cells that received more than 25 visits. Shortening the overall length of the observation period had very little effect, whereas decreasing the total number of sites with count data per year decreased accuracy. However, a severe loss of accuracy only occurred after reducing the count data to 10%, whereas a reduction to 25% had almost no effect.

Abundance trends of the Gatekeeper

Parameter estimates for the Gatekeeper are given in Table 1. As in the simulation study, a comparison of posterior estimates of population densities $\mathbf{\Lambda}$ for different years (Fig. 3) indicates that precision increases with data availability: In general, both the number of monitoring transects and the number of recorder visits increased during the study period (Fig. 3a) and consequently the variance in the posterior distributions of local population densities becomes smaller in later years (Fig. 3e). Additional to temporal variation in data availability, there are also geographical differences in the data coverage that likewise result in more uncertain estimates in regions from which only few records were reported. In the following, we derive estimates of local and global trends in the Gatekeeper's occurrence and abundance from the full posterior distributions of $\mathbf{\Lambda}$ at each site and thereby account for this heterogeneous precision.

Local changes

We evaluated the detection of changes in local abundance between two 5 year periods: 1990–1994 and 2000–2004. Thus, we calculate the posterior distributions of both 5 year means $\overline{\Lambda}_i^{1990..1994}$ and $\overline{\Lambda}_i^{2000..2004}$ for each grid cell i (Fig. 4a,b). The probability of abundance increase or decrease, respectively, is then calculated from these posterior distributions as $\Pr(\textit{Increase}) = \Pr\left(\overline{\Lambda}_i^{2000..2004} > \overline{\Lambda}_i^{1990..1994}\right)$ and $\Pr(\textit{Decrease}) = \Pr\left(\overline{\Lambda}_i^{2000..2004} < \overline{\Lambda}_i^{1990..1994}\right)$. The results indicate a likely increase of Gatekeeper abundance in the central northern part of its range, in the Southwest (Cornwall and South Wales) and in an area of northward range expansion at the west coast, whereas abundance widely decreases in other parts of the range (Fig. 4c). In total, a likely increase ($\Pr(\textit{Increase}) > 0.95$) is detected for 80 grid cells (8.000 km²) and a likely decrease ($\Pr(\textit{Decrease}) > 0.95$) is detected for 261 grid cells (26.100 km²).

General trends in range size and abundance

General trends in the Gatekeeper's abundance and range size in Great Britain can be inferred by summarizing the posterior estimates across all sites for each year. The estimated densities $\Lambda_{i,t}$ for year t serve to estimate the global relative abundance as $\sum_i \Lambda_{i,t} \cdot 100\text{km}^2$. The range size (measured at a resolution of 100 km²) can be calculated directly from the zero-inflation component as $\sum_i I_{i,t}$. Range size is estimated to slightly increase throughout the study period, although range sizes begin to stabilize after 1995

(Fig. 5a). In contrast, global relative abundance declines after an interim peak in 1995 (Fig. 5b).

DISCUSSION

Lessons from the British butterfly data

In our analysis of the Gatekeeper data we integrated two major data sources of British butterfly populations. Previous studies based the detection of large scale range shifts on presence-absence maps for distinct time periods that were summarized from occurrence records (e.g. Parmesan et al. 1999), whereas abundance indices from transect sites were used to study population trends at the site level or lumped over the entire range (e.g. Roy *et al.* 2001, Rothery & Roy 2001, Dennis *et al.* 2013). A few studies have used both types of data (Cowley *et al.* 2001; Warren *et al.* 2001; Mair *et al.* 2012), but they analyzed them separately and combined the respective findings only qualitatively. For instance, Mair *et al.* (2012) estimated trends in range size (from BNM atlas data) and total abundance (from collated UKBMS transect counts) for the Gatekeeper between the periods 1970–1982, 1995–1999 and 2005–2009 and found an increase in range size between the first intervals and a decrease of abundance between the later intervals. Our findings are in line with these general trends, but draw a more detailed picture of the spatial pattern of abundance trends.

Similar to previous analyses of UKBMS transect data, our case study estimated relative population densities corresponding to an index that summarizes count data from weekly surveys. Estimates of population trends from such indices commonly assume that

these indices are proportional to actual population size, i.e. that the average flight activity per individual is constant (e.g. Rothery & Roy 2001) and that variation in detection rates is small relative to variation in abundance across sites (Isaac *et al.* 2011). Studies on the covariance of transect counts and independently measured population sizes have confirmed strong correlations across time and space (e.g. Pollard 1977; Collier et al. 2008). Although there is strong support for spatio-temporal variability in the phenology of butterflies in the UK (Hodgson *et al.* 2011) this does not necessarily imply systematic variation in the length of the flight period. Nonetheless, the relationship between transect counts and population size deserves further investigation in order to provide reliable estimates of absolute population sizes. A more explicit link between weekly counts and annual population abundance, however, would have to account for possible phenological shifts, eventually by a model of population dynamics (Zonneveld 1991, see Gross *et al.* 2007 for a discussion). In a range-wide application, the modelling of intra-annual population dynamics would likely increase model complexity beyond practical limits. In addition to ignoring intra-annual dynamics, our analysis also does not explicitly resolve spatial heterogeneity of the surveyed transects, which are divided into subsections by a classification of habitat type. The location of both the transects and of the areas visited for presence records are non-randomly selected by the recorders and are likely to be biased towards habitat types where the occurrence of butterflies is expected. The presented model framework does not explicitly account for this bias and instead assumes – for the estimation of relative abundance variation – that the selection of sampling locations for both data types favours certain habitat types in a similar way. Extending the model towards a separate analysis of transect subsections and attribution of occurrence

records to different habitat types (when recorded at sufficiently high spatial resolution) would be possible, e.g. by modelling the expected counts (λ) and the detectability (ψ) also as a function of within-cell habitat distribution. This would be particularly relevant for studies aiming to quantify small-scale habitat-abundance relationships. For the detection of relative abundance trends, the neglect of within-cell spatial heterogeneity, just as the integration over the phenology, represents a trade-off with model complexity and computational constraints. Clearly, the extent to which range-wide estimates of abundance trends can reasonably aggregate the available data – either spatially or temporally – will have to be assessed each time the model is applied to other data sets and study species.

Potential for widespread application

Our assessment of different data scenarios indicates that, on the one hand, a network of abundance surveys as extensive as in the UKBMS scheme might not be a prerequisite for a application of the method, but that a more moderate number of about 25 sites may suffice to infer a relationship between abundance and detectability. On the other hand, the accuracy of abundance estimates for sites without count data strongly depends on the number of occurrence records per year. In the most recent years of our study period about 40% of all grid cells received five or more recorder visits per year; about 15% had more than 25 visits (cf. Fig 2c, 3a). From our findings, having many repeated visits per grid cell for such a substantial fractions of the study region appears desirable, if detection/non-detection data is expected to inform about spatial variation in abundance trends.

461 However, to what extent an abundance-detectability-curve can be inferred from a
462 given amount of abundance and occurrence data will depend on a range of additional
463 factors, including the accuracy of the abundance data, the spatial match of both data types
464 and the degree of unexplained variation in detectability among occurrence records: For
465 instance, using a systematic two-phase sampling of occurrence data and capture-mark-
466 recapture data at distinct sites, Conroy *et al.* (2008) estimated an abundance-detectability
467 relationship from as few as six surveyed sites and eight occurrence records per site. If
468 comprehensive information on the environmental conditions (e.g. weather or habitat type)
469 and the sampling effort (time spent in the field, recorder skill) is available for the
470 occurrence records, such covariates can be used to explain additional variation in
471 detectability (e.g. Kery *et al*. 2009; van Strien *et al.* 2013) and thereby be expected to
472 facilitate the estimation of an abundance-detectability relationship. A more detailed
473 analysis of the variation of detection rates, in both count and occurrence data, would also
474 be required to control for possible temporal trends in detection rates (van Strien *et al.*
475 2013). The model framework could also be extended to accommodate other data types of
476 different structure. For instance, if repeated counts of closed populations or capture-
477 mark-recapture data are available instead of or additional to the simple count surveys,
478 then the integration of these data would not only allow a more direct estimation of their
479 own observation errors (Royle & Dorazio 2008) but also inform better about absolute
480 population sizes and thereby facilitate the estimation of observation errors of other
481 surveys and of the relationship between abundance and detectability for the occurrence
482 records (Conroy *et al.* 2008).

In the form presented here, we expect the model to be readily adaptable for other taxa for which numerous detection/non-detection data overlap in their spatial coverage with a moderate amount of standardized local abundance surveys. Possible examples include the Rothamsted Insect Survey (Harrington & Woiwod 2007) that runs a long-term light trap network for moths, in parallel with geographic distribution records of moths collected by volunteer recorders for Butterfly Conservation; and the British Trust for Ornithology Breeding Birds Survey count data, in combination with their Atlas data. Using opportunistic occurrence records for the estimation of spatio-temporal abundance variation could not only improve the assessment of conservation status but also enhance the empirical basis to fundamental research in biogeography. For instance, recent approaches to understand species range dynamics from demographic process like hybrid species distribution models (e.g. Anderson *et al.* 2009; Cabral & Schurr 2010) and dynamic range models (Pagel & Schurr 2012) are often restricted by their need for data on large-scale abundance variation for parameterisation and validation.

ACKNOWLEDGEMENTS

For helpful discussion we thank the members of the UKPopNet (NERC R8-H12-01 and English Nature) working group “Bayesian distribution models: dynamics, processes and projections". We are indebted to all the recorders who contribute to the BNM and UKBMS data schemes. The UKBMS is operated by the Centre for Ecology & Hydrology and Butterfly Conservation and funded by a multi-agency consortium including the Countryside Council for Wales, Defra, the Joint Nature Conservation Committee,

505 Forestry Commission, Natural England, the Natural Environment Research Council, and
506 Scottish Natural Heritage. JP acknowledges support from the University of Potsdam
507 Graduate Initiative on Ecological Modelling (UPGradE), the German Federal Agency for
508 Nature Conservation (FKZ 806 82 270 - K1) and from the German Research Foundation
509 (SCHU 2259/3-1 and SCHU 2259/5-1). Funding for R.B.O'H was provided by the
510 Biodiversity and Climate Research Centre (BiK-F), which is part of the LOEWE
511 programme ‘Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer
512 Exzellenz' of Hesse's Ministry of Higher Education, Research and the Arts, Germany.

REFERENCES

Anderson, B.J., Akçakaya, H.R., Araújo, M.B., Fordham, D.A., Martinez-Meyer, E., Thuiller, W. & Brook, B.W. (2009) Dynamics of range margins for metapopulations under climate change. *Proceedings of the Royal Society B*, **276**, 1415–1420.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. & Jeffcoate, S. (2001) *The Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press, Oxford, UK.

Besag, J., York, J. & Mollie, A. (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–20.

Cabral, J.S. & Schurr, F.M. (2010) Estimating demographic models for the range dynamics of plant species. *Global Ecology and Biogeography*, **19**, 85–97.

Collier, N., Mackay, D.A., Benkendorff, K. (2008) Is relative abundance a good indicator of population size? Evidence from fragmented populations of a specialist butterfly (Lepidoptera: *Lycaenidae*). *Population Ecology*, **50**, 17–23.

Conroy, M.J., Runge, J.R., Barker, R.J., Schofield, M.R. & Fonnesbeck, C.J. (2008) Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. *Ecology*, **89**, 3362–3370.

Cowley, M.J.R., Thomas, C.D., Roy, D.B., Wilson, R.J., León-Cortés, J.L., Gutiérrez, D., Bulman, C.R., Quinn, R.M., Moss, D. & Gaston, K. J. (2001) Density–distribution relationships in British butterflies. I. The effect of mobility and spatial scale. *Journal of Animal Ecology*, **70**, 410–425.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011) Beyond Predictions: Biodiversity Conservation in a Changing Climate. *Science*, **332**, 53–58.

Dennis, E.B., Freeman, S.N., Brereton, T. & Roy, D.B. (2013) Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. *Methods in Ecology and Evolution*, **4**, 637–645.

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.

Dorazio, R.M. (2007) On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology*, **88**, 2773–2782.

Fisher, J.A.D., Frank, K.T. & Leggett, W.C. (2010) Dynamic macroecology on ecological time-scales. *Global Ecology and Biogeography*, **19**, 1–15.

Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.

548 Gross, K., Kalendra, E.J., Hudgens, B.R. & Haddad, N.M. (2007) Robustness and
549 uncertainty in estimates of butterfly abundance from transect counts. *Population Ecology*,
550 **49**, 191–200.

551 Harrington, R., & Woiwod, I., (2007) Foresight from hindsight: The Rothamsted Insect
552 Survey. *Outlooks on Pest Management*, **18**, 9–14.

553 Hodgson, J.A., Thomas, C.D., Oliver, T.H., Anderson, B.J., Brereton, T.M. & Crone,
554 E.E. (2011) Predicting insect phenology across space and time. *Global Change Biology*,
555 **17**, 1289–1300.

556 Isaac, N.J.B., Cruickshanks, K.L., Weddle, A.M., Rowcliffe, J.M., Brereton, T.M.,
557 Dennis, R.L.H., Shuker, D.M., Thomas, C.D. (2011) Distance sampling and the challenge
558 of monitoring butterfly populations. *Methods in Ecology and Evolution*, **2**, 585–594.

559 IUCN Standards and Petitions Subcommittee (2011) *Guidelines for Using the IUCN Red*
560 *List Categories and Criteria*, version 9.0.
561 URL http://www.iucnredlist.org/documents/RedListGuidelines.pdf.
562 [accessed 8 July 2013]

563 Kery, M., Dorazio, R.M., Soldaat, L., van Strien, A., Zuiderwijk, A. & Royle, J.A. (2009)
564 Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*,
565 **46**, 1163–1172.

566 Kery, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Häflinger, G., & Zbinden, N.
567 (2010) Site-occupancy distribution modeling to correct population-trend estimates
568 derived from opportunistic observations. *Conservation Biology*, **24**, 1388–1397.

569 Kotze, D.J., O'Hara, R.B. & Lehvävirta, S. (2012) Dealing with varying detection
570 probability, unequal sample sizes and clumped distributions in count data. *PLoS ONE*, **7**,
571 e40923.

572 Linden, A. & Mantyniemi, S. (2011) Using the negative binomial distribution to model
573 overdispersion in ecological count data. *Ecology*, **92**, 1414–1421.

574 Link, W.A. & Eaton, M.J. (2012) On thinning of chains in MCMC. *Methods in Ecology*
575 *and Evolution*, **3**, 112–115.

576 Link, W.A. & Sauer, J.R. (1998) Estimating population change from count data:
577 application to the North American breeding bird survey. *Ecological Applications*, **8**, 258–
578 269.

579 Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009) The BUGS project: Evolution,
580 critique, and future directions. *Statistics in Medicine*, **28**, 3049–3067.

581 Mair, L., Thomas, C.D., Anderson, B.J., Fox, R., Botham, M. & Hill, J.K. (2012)
582 Temporal variation in responses of species to four decades of climate warming. *Global*
583 *Change Biology*, **18**, 2439–2447.

584 McCarthy, M.A., Moore, J.L., Morris, W.K., Parris, M.P., Garrard, G.E., Vesk, P.A.,
585 Rumpff, L., Giljohann, K.M., Camac, J.S., Sana Bau, S., Friend, T., Harrison B. & Yue,
586 B. (2013) The influence of abundance on detectability. *Oikos*, **122**, 717-726.

587 MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E.
588 (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species*
589 *Occurrence.* Academic Press, San Diego, CA.

590 Pagel, J. & Schurr, F.M. (2012) Forecasting species ranges by statistical estimation of
591 ecological niches and spatial population dynamics. *Global Ecology and Biogeography*,
592 **21**, 293–304.

593 Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J.K., Thomas, C.D., Descimon, H.,
594 Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennent, W.J., Thomas, J.A. &
595 Warren, M. (1999) Poleward shifts in geographical ranges of butterfly species associated
596 with regional warming. *Nature*, **399**, 579–583.

597 Pollard, E. (1977) A method for assessing changes in the abundance of butterflies.
598 *Biological Conservation*, **12**, 115–134.

599 Pollard, E. & Yates, T.J. (1993) *Monitoring butterflies for ecology and conservation: the*
600 *British Butterfly Monitoring Scheme*. Chapman & Hall, London, UK.

601 Roberts, R. L., Donald, P. F. & Green, R. E. (2007) Using simple species lists to monitor
602 trends in animal populations: new methods and a comparison with independent data.
603 *Animal Conservation*, **10**, 332–339.

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

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

609 Royle, J.A, & Dorazio, R.M. (2006) Hierarchical models of animal abundance and
610 occurrence. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 249–
611 263.

612 Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology:*
613 *The Analysis of Data from Populations, Metapopulations, and Communities*. Academic
614 Press, San Diego, CA.

615 Royle, J.A, Kery, M., Gautier, R., & Schmid, H. (2007) Hierarchical spatial models of
616 abundance and occurrence from imperfect survey data. *Ecological Monographs*, **77**, 465–
617 481.

618 Royle, J.A. & Nichols, J. D. (2003) Estimating abundance from repeated presence-
619 absence data or point counts. *Ecology*, **84**, 777–790.

620 Scholes, R.J., Mace, G.M., Turner, W., Geller, G.N., Jürgens, N., Larigauderie, A.,
621 Muchoney, D., Walther, B.A. & Mooney, H.A. (2008) Toward a global biodiversity
622 observing system. *Science*, **321**, 1044–1045.

623 Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B., Hartig, F.,
624 Kissling, W.D., Linder, H.P., Midgley, G.F., Schröder, B., Singer, A. & Zimmermann, N.
625 E. (2012) How to understand species' niches and range dynamics: a demographic
626 research agenda for biogeography. *Journal of Biogeography*, **39**, 2146–2162.

627 van Strien, A.J., van Sway, C.A.M., & Termaat, T. (2013) Opportunistic citizen science
628 data of animal species produce reliable estimates of distribution trends if analysed with
629 occupancy models. *Journal of Applied Ecology*, **50**, 1450–1458.

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

634 Whittaker, R.J., Araújo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005)
635 Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–
636 23.

637 Zonneveld, C. (1991) Estimating death rates from transect counts. *Ecological*
638 *Entomology*, **16**, 115–121.

639 Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T.,
640 Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010) The virtual
641 ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.

642

643 Table 1 Overview of model parameters and their posterior estimates for the Gatekeeper
644 case study (only for scalar, i.e. non-vector, parameters). MCMC SE quantifies the Monte
645 Carlo sampling error in terms of the time-series standard error of the posterior mean.

Parameter	Description	Posterior mean	Posterior standard deviation	MCMC SE
State variable				
$\Lambda_{i,t}$	Relative population density in cell *i* and year *t*		(see Fig 3c-e)	
Observation process of count data				
$\lambda_{j,t}$	Expected count at site *j* in year *t*			
σ_λ^2	Overdispersion in count data	1.77	0.07	0.01
Observation process of detection/non-detection data				
$\psi_{i,t}$	Detectability in cell *i* and year *i*			
α	Saturation rate of density-detectability-curve	0.0156	0.0007	0.0001
β_0	Regression coefficients of detectability	−3.77	0.05	0.01
β_1		0.461	0.007	0.002
σ_φ^2	Variance in detectability	0.285	0.032	0.008
Hyperparameters of spatio-temporal variation in population density				
μ_{Inc}	Mean incidence	−0.56	0.27	0.05
σ_{Inc}^2	Inter-annual variance of overall incidence	0.88	0.26	0.03
μ_{Dens}	Mean (log) population density	−5.93	0.24	0.04
σ^2_D	Inter-annual variance of overall log-density	1.09	0.40	0.02
$\mu_{\gamma 1}$	Mean latitude effects across all years	−12.01	0.27	0.05
$\mu_{\gamma 2}$		−11.38	0.32	0.07
ν	Variance of the CAR model	4.48	0.16	0.04
ρ	Proportionality factor of spatial effects	0.392	0.020	0.003
σ^2	Spatially uncorrelated variance of log density	0.0088	0.0021	0.0005

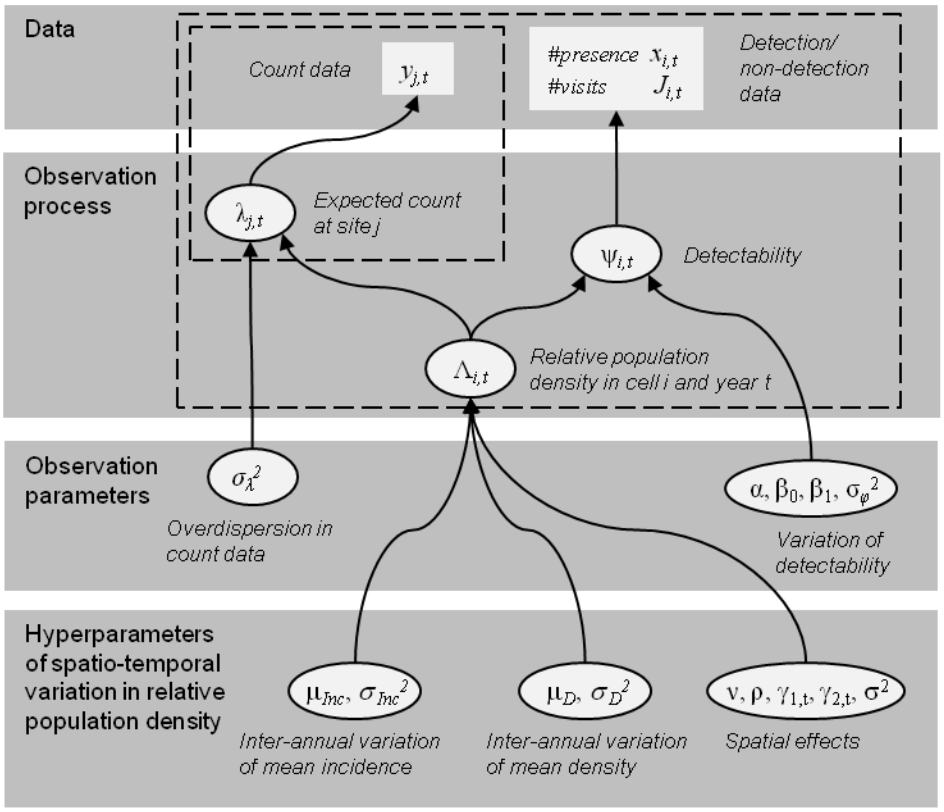
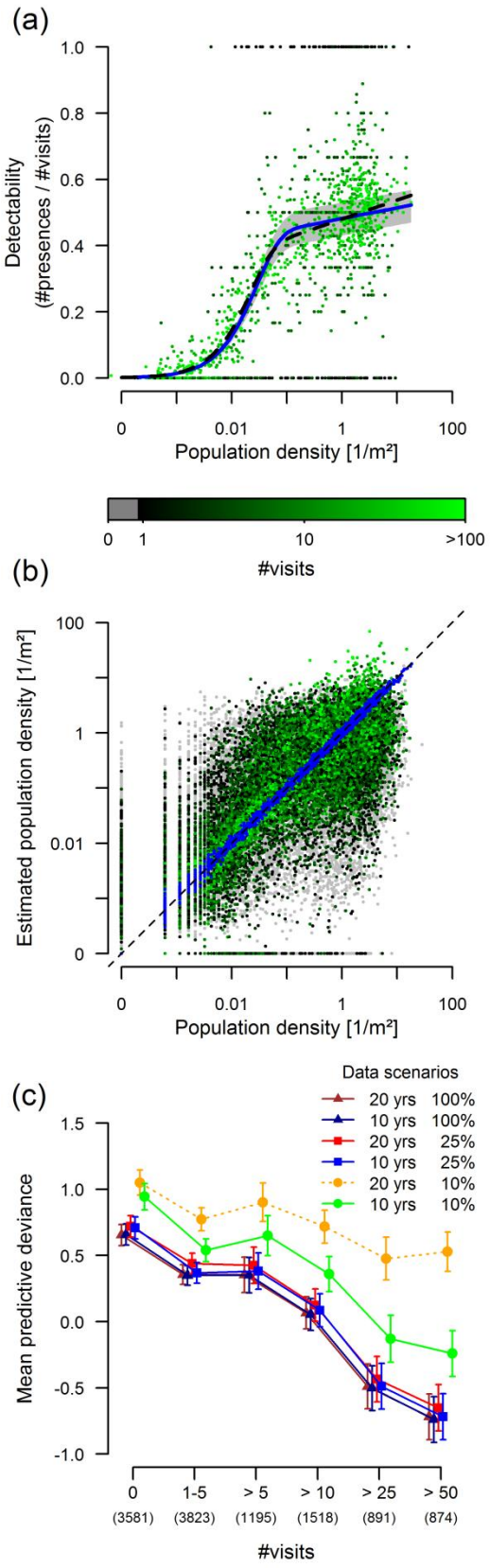
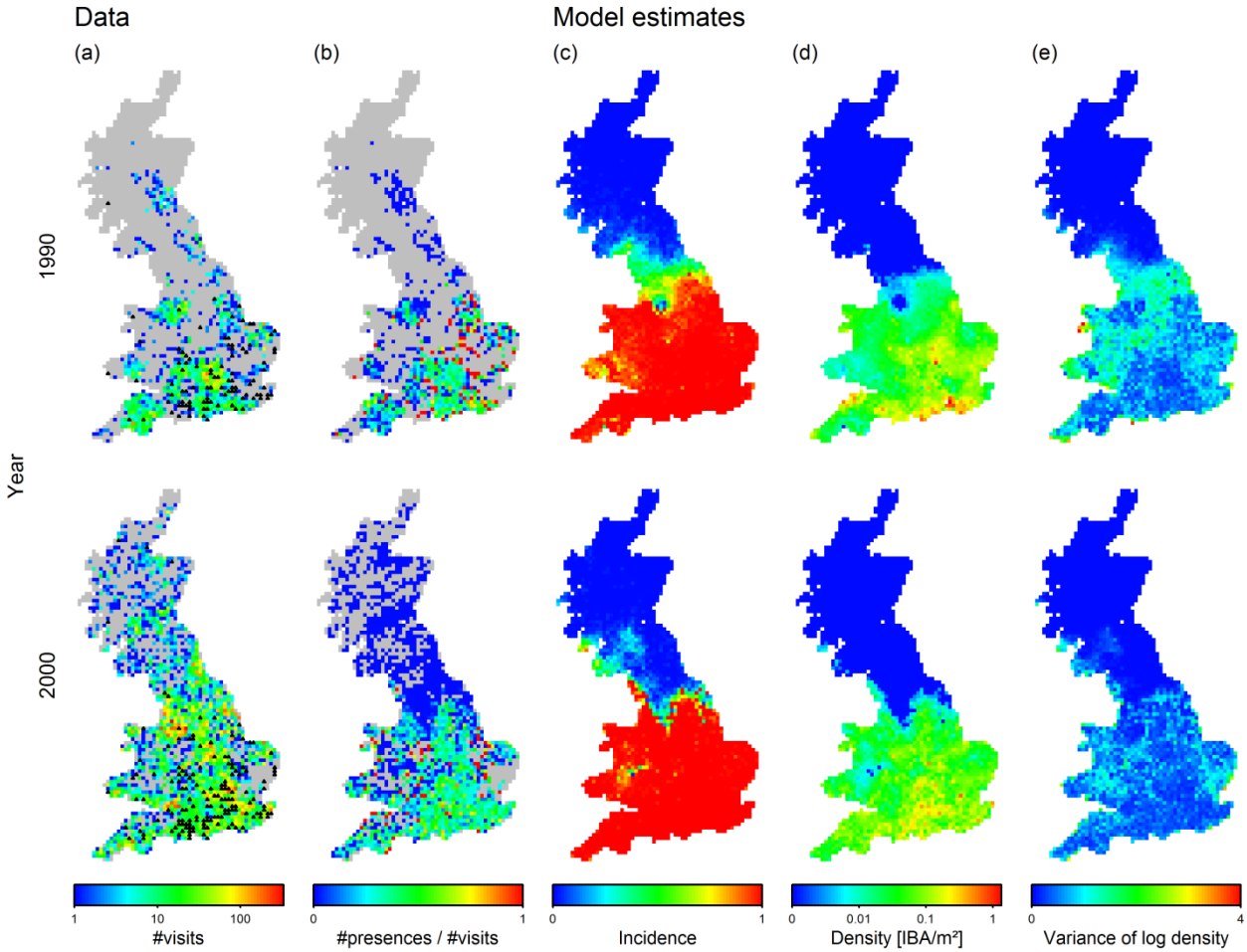


Figure 1 Structure of the hierarchical statistical model. The directed acyclic graph (DAG) describes conditional relationships between data and parameters at different levels. For each grid cell the observation models describe the likelihood of presence records and of count data from transects (if any) within this grid cell conditional on the local relative population density $\Lambda_{i,t}$ and a set of observation parameters. The variation of $\Lambda_{i,t}$ across grid cells *i* and years *t* is constrained by a set of hyperparameters that describe spatial and temporal random effects (see text for model details and Table 1 for an overview of all model parameters).

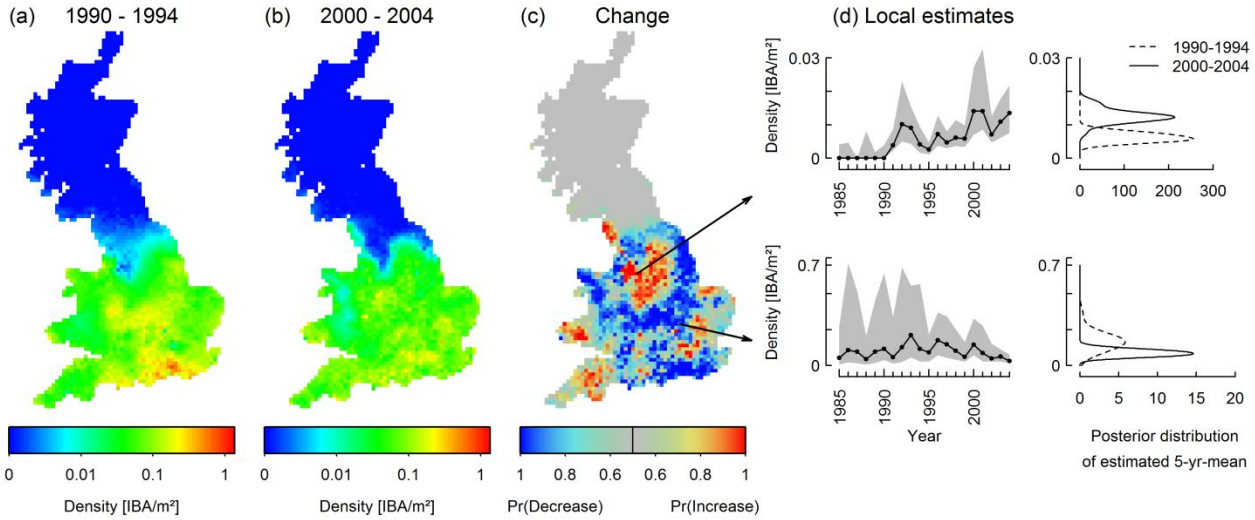

(a)
Detectability
(#presences / #visits)
1.0
0.8
0.6
0.4
0.2
0.0
0
0.01
1
100
Population density [1/m²]
0
1
10
>100
#visits
(b)
Estimated population density [1/m²]
100
1
0.01
0
0
0.01
1
100
Population density [1/m²]
(c)
Data scenarios
20 yrs 100%
10 yrs 100%
20 yrs 25%
10 yrs 25%
20 yrs 10%
10 yrs 10%
Mean predictive deviance
1.5
1.0
0.5
0.0
-0.5
-1.0
0
(3581)
1-5
(3823)
> 5
(1195)
> 10
(1518)
> 25
(891)
> 50
(874)
#visits

656

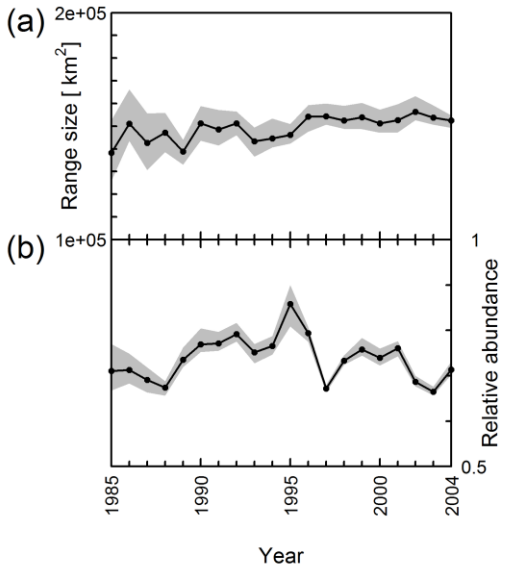
Figure 2 Results for the *standard* data scenario of the simulation study (a,b) and comparison with reduced data scenarios (c). (a) Median (blue line) and 90% credibility interval (shaded) of the estimated relationship between population density and occurrence-detectability compared to the (true) occurrence-detectability-curve applied in sampling of the simulation data (black dashed line). Points show the simulated occurrence data as the fraction of presence records obtained for cells with different population densities and numbers of recorder visits (#visits, see colour scale). (b) Estimated (posterior median) vs. true population density for sites with count data (blue) and different #visits (see colour scale). (c) Effects of shortening the observation period and reducing the proportion of cells with count data on the mean predictive deviance of population densities in the last five years of the observation period for cells with different #visits (in categories; number of cell-year combinations per category given in brackets). Error bars show the standard error of the mean. For one scenario (dotted line) the convergence criteria were not completely met after 100.000 iterations of the MCMC sampler.



672 Figure 3 Data and model estimates on distribution and abundance variation of the
673 Gatekeeper butterfly in 1990 (upper row) and 2000 (lower row). The individual columns
674 show (a) the number of recorder visits in each grid cell for the respective year and the
675 location of UKBMS transects for which count data was available (black triangles); (b) the
676 fraction of reported presences for the Gatekeeper among all reports from a cell; (c) the
677 estimated occurrence probability $Pr(\Lambda > 0)$; (d) the posterior median of estimated relative
678 population density $\boldsymbol{\Lambda}$ (log scale); (e) the posterior variance of log($\boldsymbol{\Lambda}$). Note that relative
679 population density is measured by an annual index of butterfly abundance (IBA) that
680 integrates over the phenology of butterfly activity (see *UKMBS abundance indices* for
681 details).



682 Figure 4 Spatial variation in abundance trends of the Gatekeeper butterfly in Great
683 Britain. (a-b) Posterior median of 5-yr-means of abundance for the periods 1990–1994
684 and 2000–2004. (c) Comparison of the estimated 5-yr-means result in a map which gives
685 for each grid the estimated probability of a decrease or increase of local abundances
686 between both periods. (d) Examples of time series of estimated relative population
687 density (left, grey areas depict 95% credibility intervals) and respective posterior
688 distributions of the estimated 5-yr-means (right) for two grid cells, where local
689 abundance either increases (top) or decreases (bottom) between 1990–1994 and 2000–
690 2004.



691

692 Figure 5 Estimated annual variation in range size (a) and total abundance (b) of the
693 Gatekeeper butterfly in Great Britain. For both graphs the lines depict the posterior
694 median of model estimates and shaded areas comprise the central 95% credibility
695 interval. Note that estimated range sizes correspond to a spatial resolution of 10×10 km².
696 The shown abundances are not absolute values but refer to a relative measure of the
697 butterfly-day-index (see text for details) and are presented on a logarithmic scale.

Supporting Information: Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records.

S.1 Pre-analysis of the relationship between population density and detectability

The estimation of population density from detection/non-detection data is based on inferring how population density (or abundance) influences the detectability of a species' presence. Here, we review approaches that have been applied in previous studies to model this relationship, and test different models in a preliminary analysis of the count and occurrence data from our case study on the Gatekeeper butterfly in Great Britain.

Models for the density-detectability-curve

A model for a principal relationship between abundance N and the probability ψ to detect a species' presence was presented by Royle & Nichols (2003). Their model (hereafter called RN model) derives from a binomial model for the number of encountered individuals x if individuals have per-individual detection probability r. With the number of encountered individuals $x \sim Binomial(N, r)$ the detectability can be calculated as the probability to encounter at least one individual:

$$\psi = \Pr(x > 1|r, N) = 1 - (1 - r)^N \qquad \text{(eqn. S.1.1)}$$

An alternative approach starts from describing sampling as a Poisson process, where the rate at which individuals are encountered is a product of the abundance N and a measure of sampling intensity α. With the number of encountered individuals $x \sim Poisson(\alpha \cdot N)$ the probability to encounter at least one individual is

$$\psi = \Pr(x > 1|\alpha, N) = 1 - exp(-\alpha \cdot N) \qquad \text{(eqn. S.1.2)}$$

This is equivalent to eqn. S.1.1 with $\alpha = -\ln(1 - r)$. In the following we use the formulation of eqn. S.1.2 for the RN model. Conveniently, if abundance is not described as the total number of individuals but by some relative measure, as in our case study, the proportionality factor between relative and absolute abundance will simply scale the estimate of α.

McCarthy *et al.* (2013) demonstrate how the linear increase of the rate parameter in eqn. S.1.2 with abundance implies the assumption of independent encounters of individuals, which is likely violated in many applications. They propose a generalization (hereafter MC model)

$$\psi = 1 - \exp(-\lambda) \qquad \text{(eqn. S.1.3)}$$

$$\ln(\lambda) = \beta_0 + \beta_1 \cdot \ln(N)$$

which is equivalent to the RN model for scaling exponent $\beta_1 = 1$ (and with $\beta_0 = \ln(\alpha)$). A scaling exponent $\beta_1 < 1$ describes non-independent detection due to increased clustering of individuals. Notably, the MC model is equivalent to a linear regression with a complementary log log link[1] function, i.e. $\text{cloglog}(\psi) = \beta_0 + \beta_1 \cdot \ln(N)$. Other studies have used the (more common) logit link and applied a logistic regression model

$$\text{logit}(\psi) = \beta_0 + \beta_1 \cdot N \qquad \text{(eqn. S.1.4)}$$

to describe the relationship between abundance and detectability (e.g. Tanadini & Schmidt 2011). Additional to these previously proposed functional relationships, we consider an additional model that combines the basic RN model with a logistic regression. Therefore we add a multiplicative term φ and formulate a logistic regression of φ on the abundance N:

$$\psi = \varphi \cdot \{1 - \exp(-\alpha \cdot N)\} \qquad \text{(eqn. S.1.5)}$$
$$\text{logit}(\varphi) = \beta_0 + \beta_1 \cdot N$$

Data analysis

In order to investigate which functional form of the density-detectability-curve is most appropriate for our case study, we performed a preliminary analysis based on count data (UKBMS abundance indices) and occurrence data (detection/non-detection data). In the preliminary analysis we only use data for grid cells *i* and years *t* for which both data types are available. To study the relationship between abundance indices and detectability ψ we formulate a binomial model $x_{i,t} \sim Binomial(J_{i,t}, \psi_{i,t})$ for the number of presence records $x_{i,t}$ among all visits $J_{i,t}$ of a cell and use the different models outlined above to describe $\psi_{i,t}$ as a function of population density. For this preliminary analysis, an index of relative population density on the grid cell level $\Lambda_{i,t}$ is calculated from the UKBMS abundance indices by simply dividing the sum of all counts from one cell and year by the total transect area. Considered models for the relationship between ψ and Λ include the RN model, the MC model, the logistic regression (LR) and our extension of the RN model by a multiplicative random effect with (MR.LR) or without (MR) an additional dependence on Λ. For the logistic regression model (LR) and the regression part of the MR.LR model we additionally include alternative versions that use $\ln(\Lambda_{i,t})$ as covariate (denoted LR.log resp. MR LR.log). We used a maximum-likelihood approach to estimate the parameters of each model and to calculate Akaike's Information criterion (AIC) for each candidate model.

[1]The complementary log log link $p = 1 - \exp\{-\exp(\boldsymbol{\beta X})\}$ dates back to Fisher (1922), where it was introduced in the very related context of estimating the number of micro-organisms in a sample of soil or water from the distribution of organism's presence and absence in diluted sub-samples.

Table S.1.1 Overview of the different models for the density-detectability-curve and their maximum-likelihood estimation for the pre-analysis of the Gatekeeper data.

Model	Formula	Estimated parameters			ΔAIC
		ln(α)	β_0	β_1	
RN	$\psi = 1 - \exp(-\alpha \cdot \Lambda)$	2.38			21238
MC	$\psi = 1 - \exp(-\lambda)$ $\ln(\lambda) = \beta_0 + \beta_1 \cdot \ln(\Lambda)$		-0.79	0.075	154
LR	$\text{logit}(\psi) = \beta_0 + \beta_1 \cdot \Lambda$		-0.83	0.38	292
LR.log	$\text{logit}(\psi) = \beta_0 + \beta_1 \cdot \ln(\Lambda)$		-0.55	0.092	150
MR	$\psi = \varphi \cdot \{1 - \exp(-\alpha \cdot \Lambda)\}$ $\text{logit}(\varphi) = \beta_0$	5.97	-0.76		52
MR.LR	$\psi = \varphi \cdot \{1 - \exp(-\alpha \cdot \Lambda)\}$ $\text{logit}(\varphi) = \beta_0 + \beta_1 \cdot \Lambda$	6.06	-0.80	0.27	30
MR.LR.log	$\psi = \varphi \cdot \{1 - \exp(-\alpha \cdot \Lambda)\}$ $\text{logit}(\varphi) = \beta_0 + \beta_1 \cdot \ln(\Lambda)$	6.36	-0.62	0.058	0

Results and Interpretation

Among the considered models, the modification of the RN model by a multiplicative term with additional dependence on ln(Λ) (MR.LR.log) clearly performs best in describing the relationship between abundance data and detection/non-detection data for the Gatekeeper (Tab. S.1.1). When fitted to the data, this model describes a rapid sigmoid increase of detectability for smaller population densities, whereas detectability increases much slower and approximately linear to ln(Λ) for larger population densities (Fig. S.1.1). While the mechanisms underlying this relationship cannot be fully resolved by this analysis, a possible interpretation can be deduced from the specific structure of the detection/non-detection data. The citizen science program that provides the opportunistic records is targeted not specifically at the focal species but at all butterflies in Great Britain, which vary in both their habitat requirements and phenology. Consequently, recorder visits occur in habitat types and at times within the season, where the detection of the focal species is highly unlikely irrespective of its relative population density in the area. A potential rate of not reporting the species when detected might add to this. This substantial probability of non-detection (resp. reporting) even for high population density is reflected by the saturation of the sigmoid part of the density-detectability-curve at values far below one. Interestingly, the best model still predicts a positive effect of population density on detectability at higher densities (beyond saturation of the RN model component). Under the given interpretation, this could indicate a ‘spill-over-

effect', where a very high population density increases the chance to encounter the species outside the preferred habitat type.

While the emergence of the found density-detectability-curve clearly demands further investigation, for our study on estimating trends in relative population densities, we conclude that the MR.LR.log model proves most suitable to describe the relationship between population density and detectability of the Gatekeeper in the opportunistic occurrence records.

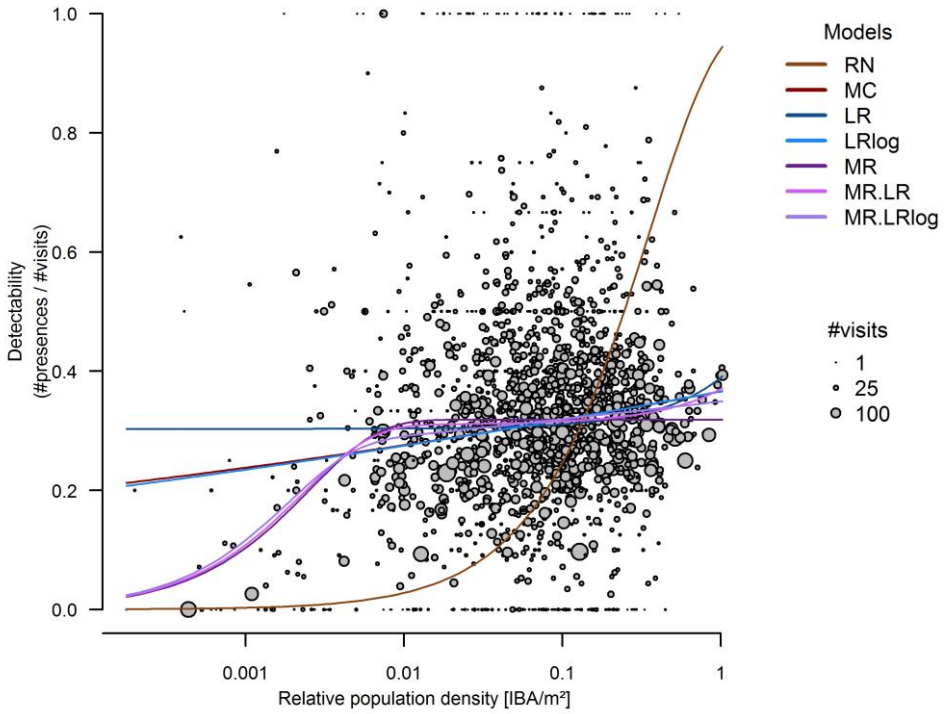


Figure S.1.1 Estimated density-detectability-curves based on different models for the functional relationship between relative population density and detectability of the Gatekeeper butterfly.

S.2 Simulation of virtual data

The simulation study was designed as a virtual ecologist study (Zurell *et. al* 2010) to test the presented model framework for the estimation of spatial and temporal abundance variation from observation data. We therefore generated virtual data from a dynamic abundance pattern in three steps:

(i) Creating a spatially heterogeneous and dynamic virtual landscape

(ii) Simulation of population dynamics

(iii) Probabilistic sampling of (imperfect) observation data from the simulated 'true' abundance pattern

The simulation of spatial population dynamics in a dynamic landscape (i–ii) was based on a modified model from a previous virtual ecologist study (Pagel & Schurr 2012).

(i) Artificial landscape

We generated a dynamic artificial landscape with an extent of 50×50 grid cells and a cell size of 10×10 km². Environmental variation across the landscape was represented as variation of the intrinsic population growth rate r (see model description below, eqn. S.2.1) in space and time. We used fractal Brownian motion (Hurst index = 0.5) to generate a spatially auto-correlated static landscape and added a humped-shaped latitudinal effect. To represent temporal dynamics, i.e. environmental change, the optimum of this latitudinal effect was shifted towards the northern border of the model landscape. We generated yearly maps of intrinsic population growth rates for a spin-off period of 50 years (without environmental change) and subsequent 50 years of gradual environmental change. Finally, growth rates were scaled so that on average 25% of the model landscape had positive growth rates ($r > 0$).

(ii) Population dynamics

We simulated spatio-temporal population dynamics by a stochastic grid-based simulation model that combines local (within-cell) population dynamics with dispersal between grid cells. As a description of population dynamics within cells we used the stochastic Ricker model

$$\log(N_{i,t+1}) = \log(\tilde{N}_{i,t}) + r_{i,t} - h\,\tilde{N}_{i,t} + \varepsilon_{i,t} \qquad \text{(eqn. S.2.1)}$$

Stochasticity is introduced by the error term ε being an *iid* normal random variable $\varepsilon \sim Normal(0, \sigma_P^2)$. $\tilde{N}$ denotes the post-dispersal population size. Dispersal was described by a mixture-dispersal-kernel, where a fraction f_{LDD} of dispersal units is subject to long-distance dispersal following an exponential kernel $f(r) = 1/\alpha \cdot \exp(-r/R)$, with mean dispersal distance

R. The dispersal kernel was then integrated over both the cell of origin *j* and the target cell *i* to obtain dispersal probabilities $P_{j\to i}(f_{LDD},\alpha)$ between spatially discrete cells and to calculate post-dispersal population sizes as

$$\tilde{N}_{i,t} = \sum_j P_{j\to i}\left(f_{LDD},\alpha\right)\cdot N_{j,t} \qquad \text{(eqn. S.2.2)}$$

Note that with the parameterisation of the Ricker model (eqn. S.2.1) in terms of intrinsic growth rate *r* and competition intensity *h*, the carrying capacity ($K = r/h$) likewise varied across the artificial landscape, which mainly drives spatial variation in simulated abundances. These abundances were finally divided by the grid cell area of 100 km² to calculate population densities $\Lambda_{i,t}$ (Fig. S.2.1a). Parameter values used in the simulation were $h = 0.0002$; $\sigma_P^2 = 0.1$; $f_{LDD} = 0.05$; R = 5 km. The model was initialized by assigning to each cell a population size equal to its carrying capacity *K* (if positive) at the beginning of the spin-off period.

(iii) Virtual data collection

The sampling scheme for the virtual data was designed to mimic the data availability (the distribution of transects and of recorder visits across cells) in the butterfly case study. Hence, we randomly assigned to each cell *A* in the artificial landscape a 'sister cell' *B* of the British 10×10 km² grid and used the number of annual recorder visits and the characteristics of transects (if there are any) in *B* to generate observation data from the population density in *A* (see Fig. S.2.1b for examples of annual sampling schemes). Data was sampled for the last 20 years of the simulation period and for each year the number of presence records and the count data were randomly sampled from probability distributions as given by the observation models: The number of presence records was drawn from a binomial distribution with sample size equal to the assigned number of recorder visits and a per-visit-probability of a presence record calculated from the population density $\Lambda_{i,t}$ (cf. eqn. 4 in the main text, see Fig. S.2.1c for examples of generated data). Count data $y_{j,t}$ was drawn from a lognormal-Poisson distribution with the mean calculated by multiplying the population densities $\Lambda_{i,t}$ with the assigned transect area (cf. eqn. 1 in the main text). We deliberately set the proportionality factor between simulated abundances and estimated relative abundances (cf. eqn. 2 in the main text) to one. The following parameter values were used to simulate the data: $\alpha = 0.005$; $\beta_0 = -1$; $\beta_1 = 0.1$; $\sigma_\varphi = 0.05$; $\sigma_\lambda = 0.1$.

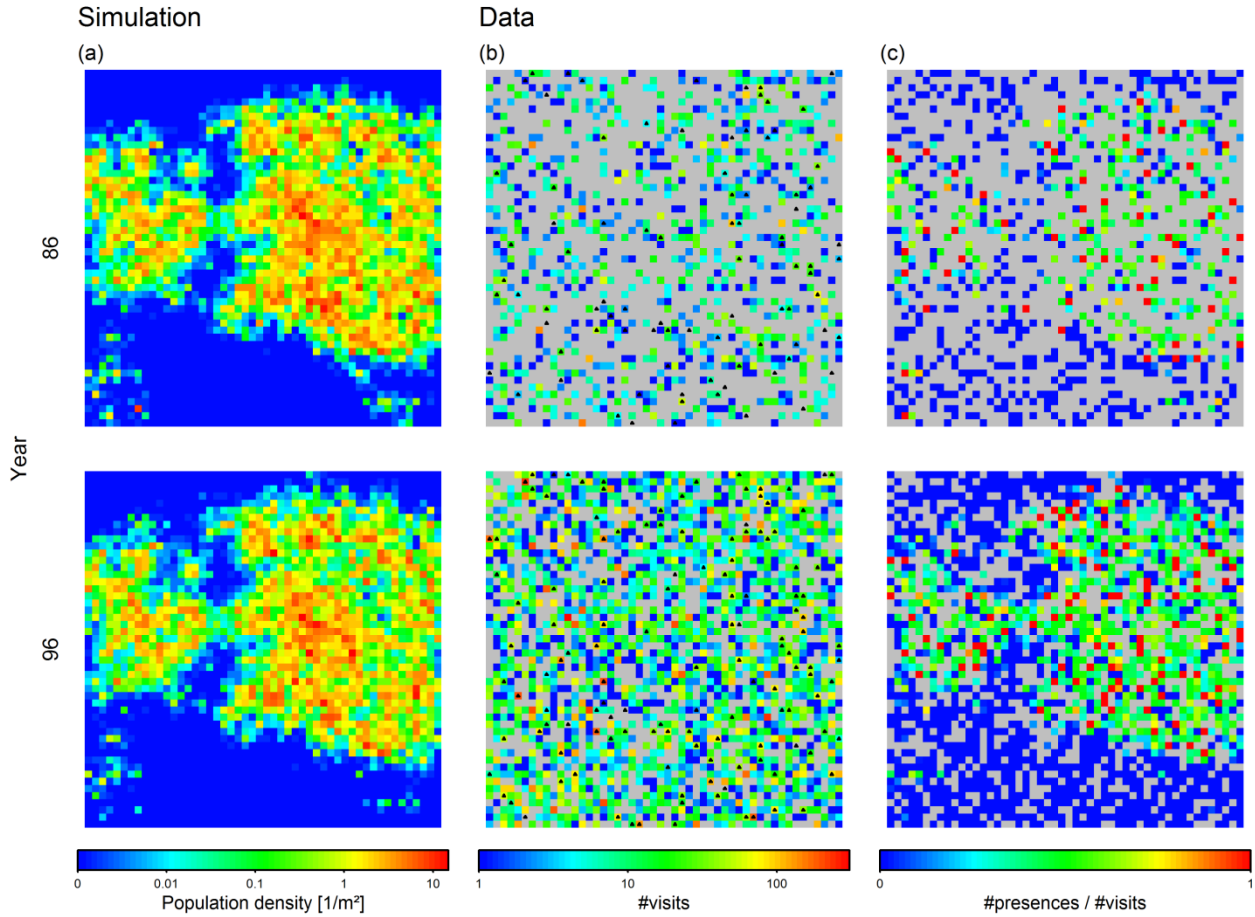


Figure S.2.1 Simulated population density and virtual data of the standard scenario for the years 86 (upper row) and 96 (lower row) of the simulation. The individual columns show (a) the simulated population density; (b) the number of recorder visits in each grid cell for the respective year and the location of sites for which count data was sampled (black triangles); (c) the fraction of presence records among all visits of a cell. For the presented years the quantity of data is equal to the available data for the Gatekeeper case study in the years 1990 and 2000 (see Fig. 3).

S.3 Bayesian model estimation with OpenBUGS

Here we give the OpenBUGs model code that was implemented for the parameter estimation of the presented hierarchical Bayesian model. Both the simulation study and the Gatekeeper case study used the same code. In order to facilitate the application of the model to other data sets we briefly describe the necessary pre-processing of the data and give an overview of the variables of the model and their relation to variable names used in the main text (Table S.3.1). The following overview lists the various data objects that need to be passed to the model:

State-space dimensions

The spatial and temporal dimensions of the state-space model of population densities are given by:

`n.sites` – the number of cells in the model grid

`n.yrs` – the length of the time period for which population densities are estimated

Spatial configuration

For the estimation of spatially correlated random effects one has to specify the adjacency of grid cells (eight-neighbour-rule). The implemented CAR model requires the following data format:

`n.NB[n.sites]` – a vector of length `n.sites` that gives for each site the number of neighbours

`NBvec[NBtot]` – a vector that lists consecutively for all cells the indices of their neighbouring cells

`NBtot` – the total length of `NBvec`

For illustration, imagine that (as in the rectangular grid of the simulation study) cell 1 has three adjacent cells (2, 51, 52) and cell 2 has five adjacent cells (1, 3, 51, 52, 53). Then `n.NB` = (3, 5, ...) and `NBvec` = (2, 51, 52, 1, 3, 51, 52, 53, ...) and the total length of `NBvec` equals the sum of `n.NB`. For further details see the GeoBUGS manual.

`Lat[n.sites]` – (normalized) geographical latitude of grid cell midpoints

Occurrence data

After aggregating the occurrence records to #visits and #presence per grid cell and year, data is passed to the model only for those cells and years for which the number of recorder visits is positive:

`n.rec` – the total number of grid cells (per year) with recorder visits

`visits[n.rec]` – the number of recorder visits

`presence[n.rec]` – the number of recorded presences

`rec.site[n.rec]` – the index of the grid cell where the records were sampled

`rec.year[n.rec]` – the index of the year when the records were sampled

Presence record data

The format of the count data is similar and comprises four vectors with one entry each for every count:

`n.S` – the total number of count data

`S.index[n.S]` – the result of the count survey

`S.area[n.S]` – the area of the sampled site

`S.site [n.S]` – the index of the cell where the site is located

`S.year [n.S]` – the index of the year when the survey was conducted

Table S.3.1 Overview of model parameters and the respective variable names in the OpenBUGS code.

Parameter	Description	Variable name in OpenBUGS code
$\ln(\alpha)$	Saturation rate of detection probability (log)	`log.alpha`
β_0	Regression coefficients of detection probability	`pi.b0`
β_1		`pi.b1`
σ^2_λ	Variance of detection probability	`sig.pi`
σ^2_φ	Overdispersion of count data	`sig.S`
$\lambda_{j,t}$	Expected count at site *j* in year *t*	`lambda.eff[n.S]`
$\Lambda_{i,t}$	Population density in cell *i* and year *t*	`Lambda[n.sites,n.yrs]`
Inc_t	Mean overall incidence in year *t*	`b0[n.yrs]`
μ_{Inc}	Mean incidence across all years	`mu.b0`
σ^2_{Inc}	Inter-annual variance of overall incidence	`sig.b0`
D_t	Mean log-density in year *t*	`c0[n.yrs]`
μ_D	Mean log-density across all years	`mu.c0`
σ^2_D	Inter-annual variance of overall log-density	`sig.c0`
$\mathbf{\Delta}_t$	Spatially autocorrelated random effects	`rho[n.sites,n.yrs]`
$\mu_{\gamma 1}$	Mean latitude effects across all years	`mu.b1`
$\mu_{\gamma 2}$		`mu.b2`
ν	Variance of the CAR model	`v`
ρ	Proportionality factor of spatial effects	`beta`
σ^2	Spatially uncorrelated variance of log density	`sig.dens`

OpenBUGS model code

```
model{

# spatio-temporal abundance variation
for(yr in 1:n.yrs){
  for (i in 1:n.sites){
    logit(pInc[i,yr]) <- b0[yr] + beta * sp[i,yr]
    Inc[i,yr] ~ dbern(pInc[i,yr])
    sp[i,yr] <- b1[yr]*Lat[i] + b2[yr]*Lat[i] *Lat[i] + rho[yr,i]
    muDen[i,yr] <- c0[yr] + sp[i,yr]
    LogD[i,yr] ~ dnorm(muDen[i,yr], prec.dens)

    Lambda[i,yr] <- Inc[i,yr]*exp(LogD[i,yr])
    }

  rho[yr,1:n.sites] ~ car.normal(NBvec[], weights[], n.NB[], tau)
  }
# constant weights for CAR
for(k in 1:NBtot) {weights[k] <- 1}

# temporal random effects
for(yr in 1:n.yrs){
  b0[yr] ~ dnorm(mu.b0,pr.b0)
  b1[yr] ~ dnorm(mu.b1,pr.b1)
  b2[yr] ~ dnorm(mu.b2,pr.b2)

  c0[yr] ~ dnorm(mu.c0,pr.c0)
  }

# presence records
# in loop over all gridcell-year combinations with recorder visits
for(rec in 1:n.rec) {
  presence[rec] ~ dbin(psi[rec],visits[rec])
  psi[rec] <- pi[rec] * (1 - exp(-exp(log.alpha) * Lambda[rec.site[rec],rec.year[rec]]))
  logit(pi[rec]) <- pi.b0 + pi.b1 * LogD[rec.site[rec],rec.year[rec]] + e.pi[rec]
  e.pi[rec] ~ dnorm(0,prec.pi)
  }

# count data
# in loop over all site-year combinations where count data were recorded
for (s in 1:n.S){
    S.mu[s] <- log(Lambda[S.site[s],S.year[s]] * S.area[s] + 0.001)
    lambda.eff[s] ~ dlnorm(S.mu[s],prec.S)
    S.index[s] ~ dpois(lambda.eff[s])
    }

# prior distributions
mu.b0 ~ dnorm(0,0.01)
mu.b1 ~ dnorm(0,0.01)
mu.b2 ~ dnorm(0,0.01)

mu.c0 ~ dnorm(0,0.01)

pr.b0 <- dgamma(0.001,0.001)
pr.b1 <- dgamma(0.001,0.001)
pr.b2 <- dgamma(0.001,0.001)
pr.c0 <- dgamma(0.001,0.001)

log.alpha ~ dnorm(0,0.01)
pi.b0 ~ dnorm(0,0.01)
pi.b1 ~ dnorm(0,0.01)
prec.pi <- 1 / (sig.pi*sig.pi)
sig.pi ~ dnorm(0,0.1) I(0,)

prec.dens <- 1 / (sig.dens*sig.dens)
sig.dens ~ dnorm(0,1) I(0,10)
prec.S <- 1 / (sig.S*sig.S)
sig.S ~ dnorm(0,1) I(0,10)


v ~ dnorm(0, 0.2) I(0,)
tau  <- 1/v
beta ~ dnorm(0, 0.01) I(0,)
}
```

Literature cited

Fisher, R.A. (1922) On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A*, **222**, 309–368.

McCarthy, M.A., Moore, J.L., Morris, W.K., Parris, M.P., Garrard, G.E., Vesk, P.A., Rumpff, L., Giljohann, K.M., Camac, J.S., Sana Bau, S., Friend, T., Harrison B. & Yue, B. (2013) The influence of abundance on detectability. *Oikos*, **122**, 717-726.

Pagel, J., Schurr, F.M. (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.

Royle, J.A. & Nichols, J. D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.

Tanadini, L.G. & Schmidt, B.R. (2011) Population Size Influences Amphibian Detection Probability: Implications for Biodiversity Monitoring Programs. *PLoS ONE*, **6**, e28244.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.