

# A regionally informed abundance index for supporting integrative analyses across butterfly monitoring schemes

Reto Schmucki<sup>1,2\*</sup>, Guy Pe'er<sup>3,4</sup>, David B. Roy<sup>5</sup>, Constantí Stefanescu<sup>6,7</sup>,  
Chris A.M. Van Swaay<sup>8</sup>, Tom H. Oliver<sup>5,9</sup>, Mikko Kuussaari<sup>10</sup>, Arco J. Van Strien<sup>11</sup>,  
Leslie Ries<sup>12,13</sup>, Josef Settele<sup>4,14</sup>, Martin Musche<sup>14</sup>, Jofre Carnicer<sup>6,15</sup>, Oliver Schweiger<sup>14</sup>,  
Tom M. Brereton<sup>16</sup>, Alexander Harpke<sup>14</sup>, Janne Heliölä<sup>10</sup>, Elisabeth Kühn<sup>14</sup> and  
Romain Julliard<sup>1</sup>

<sup>1</sup>MNHN-CNRS-UPMC, UMR7204-CESCO, Sorbonne Universités, 43 rue Buffon, CP 135, 75005 Paris, France;

<sup>2</sup>Centre de Synthèse et d'Analyse sur la Biodiversité, Immeuble Henri Poincaré, Domaine du Petit Arbois, Avenue Louis Philibert, 13857 Aix-en-Provence, France; <sup>3</sup>Department of Conservation Biology, UFZ - Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany; <sup>4</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany; <sup>5</sup>NERC Centre for Ecology & Hydrology, Wallingford, Oxfordshire OX10 8EF, UK; <sup>6</sup>CREAF, Cerdanyola del Valles, Catalonia 08193, Spain;

<sup>7</sup>Butterfly Monitoring Scheme - Museu de Ciències Naturals de Granollers, Granollers, Catalonia 08402, Spain;

<sup>8</sup>Dutch Butterfly Conservation and Butterfly Conservation Europe, PO Box 506, NL-6700 AM Wageningen, Netherlands; <sup>9</sup>School of Biological Sciences, University of Reading, Whiteknights, Reading, Berkshire RG6 6AS, UK;

<sup>10</sup>Natural Environment Centre, Finnish Environment Institute (SYKE), PO Box 140, FI-00251 Helsinki, Finland;

<sup>11</sup>Statistics Netherlands, PO Box 24500, 2490 HA The Hague, Netherlands; <sup>12</sup>Department of Biology, University of Maryland, College Park, MD 20740, USA; <sup>13</sup>National Socio-Environmental Synthesis Centre, 1 Park Place, Suite 300, Annapolis, MD 21401, USA; <sup>14</sup>Department of Community Ecology, UFZ - Helmholtz Centre for Environmental Research, Theodor-Lieser Str. 4, 06120 Halle, Germany; <sup>15</sup>Community and Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, Nijenborgh 7, 9747 AG Groningen, Netherlands; and <sup>16</sup>Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset BH20 5QP, UK

<sup>12</sup>Department of Biology, University of Maryland, College Park, MD 20740, USA; <sup>13</sup>National Socio-Environmental Synthesis Centre, 1 Park Place, Suite 300, Annapolis, MD 21401, USA; <sup>14</sup>Department of Community Ecology, UFZ - Helmholtz Centre for Environmental Research, Theodor-Lieser Str. 4, 06120 Halle, Germany; <sup>15</sup>Community and Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, Nijenborgh 7, 9747 AG Groningen, Netherlands; and <sup>16</sup>Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset BH20 5QP, UK

<sup>11</sup>Statistics Netherlands, PO Box 24500, 2490 HA The Hague, Netherlands; <sup>12</sup>Department of Biology, University of Maryland, College Park, MD 20740, USA; <sup>13</sup>National Socio-Environmental Synthesis Centre, 1 Park Place, Suite 300, Annapolis, MD 21401, USA; <sup>14</sup>Department of Community Ecology, UFZ - Helmholtz Centre for Environmental Research, Theodor-Lieser Str. 4, 06120 Halle, Germany; <sup>15</sup>Community and Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, Nijenborgh 7, 9747 AG Groningen, Netherlands; and <sup>16</sup>Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset BH20 5QP, UK

## Summary

1. The rapid expansion of systematic monitoring schemes necessitates robust methods to reliably assess species' status and trends. Insect monitoring poses a challenge where there are strong seasonal patterns, requiring repeated counts to reliably assess abundance. Butterfly monitoring schemes (BMSs) operate in an increasing number of countries with broadly the same methodology, yet they differ in their observation frequency and in the methods used to compute annual abundance indices.

2. Using simulated and observed data, we performed an extensive comparison of two approaches used to derive abundance indices from count data collected via BMS, under a range of sampling frequencies. Linear interpolation is most commonly used to estimate abundance indices from seasonal count series. A second method, hereafter the regional generalized additive model (GAM), fits a GAM to repeated counts within sites across a climatic region. For the two methods, we estimated bias in abundance indices and the statistical power for detecting trends, given different proportions of missing counts. We also compared the accuracy of trend estimates using systematically degraded observed counts of the Gatekeeper *Pyronia tithonus* (Linnaeus 1767).

3. The regional GAM method generally outperforms the linear interpolation method. When the proportion of missing counts increased beyond 50%, indices derived via the linear interpolation method showed substantially higher estimation error as well as clear biases, in comparison to the regional GAM method. The regional GAM method also showed higher power to detect trends when the proportion of missing counts was substantial.

\*Correspondence author. E-mail: reto.schmucki@mail.mcgill.ca

**4. Synthesis and applications.** Monitoring offers invaluable data to support conservation policy and management, but requires robust analysis approaches and guidance for new and expanding schemes. Based on our findings, we recommend the regional generalized additive model approach when conducting integrative analyses across schemes, or when analysing scheme data with reduced sampling efforts. This method enables existing schemes to be expanded or new schemes to be developed with reduced within-year sampling frequency, as well as affording options to adapt protocols to more efficiently assess species status and trends across large geographical scales.

**Key-words:** abundance indices, butterfly monitoring scheme, butterfly count, citizen science, flight period, insect conservation, missing data, pollard walk, sampling effort, seasonal pattern

## Introduction

Long-term, standardized monitoring programmes are key to assessing the state of biodiversity. They enable the impacts of environmental change on population abundance to be quantified, and provide evidence of the status of species and ecosystems against policy targets (Warren 1993; Van Swaay *et al.* 2011). Over the last two decades, the number of volunteer-based monitoring schemes has substantially increased and expanded to multiple taxa, including amphibians, birds, butterflies, mammals, plants, reptiles and other insects (Schmeller *et al.* 2009). With the growing prominence of citizen science initiatives, data contributed by large networks of non-professionals represent an important resource to assess species trends and build robust biodiversity indicators (Gregory *et al.* 2005; Van Swaay *et al.* 2008; Brereton *et al.* 2011; Sauer & Link 2011). These large data sets are particularly useful for measuring the impact of changes in climate and other environmental drivers, providing substantial insights into ecological processes that inform conservation initiatives (Schmeller *et al.* 2009).

A dominant goal in most long-term monitoring schemes is to produce reliable measures of density or abundance to adequately assess population change in time and space (Stephens *et al.* 2015). Collection of long time series is particularly relevant for invertebrates, which are prone to show important inter-annual fluctuation and where short-term population change must be interpreted with caution. For insect monitoring schemes, a key challenge in producing reliable abundance metrics is the link between species phenology and observable abundance of a specific life stage (e.g. adult butterflies). Like most insects, butterfly counts are characterized by strong seasonal patterns (Roy & Sparks 2000), determined by asynchrony of emergence, the longevity of individuals and the number of generations produced per year – all of which are species-specific attributes that vary over climatic regions and change annually in response to factors such as weather and biotic interactions. Such variability in phenology has been evidenced along a latitudinal gradient covering three different climatic regions in the United Kingdom (Hodgson *et al.* 2011).

When abundance counts are characterized by seasonal patterns, repeated counts are crucial for producing reli-

able abundance indices and to detect change in populations over time. By recommending weekly counts, the first butterfly monitoring scheme (BMS) established in 1976 in the UK (Pollard & Yates 1993) acknowledged the necessity of such repeated sampling over the monitoring season. With such data, abundance indices can then be calculated as the total number of individuals observed over all weekly counts from a site in a given year, providing a measure of accumulated ‘butterfly days’ over a defined time interval (Pollard & Yates 1993). This measure assumes that detection probability does not vary systematically and that counts are frequent, evenly spread in time, and cover the entire period of adult butterfly activity. In reality, transect counts produced by BMSs are often unevenly distributed in time or with periods of missing counts due to unsuitable weather conditions, unavailability of recorders, or protocols based on less frequent visits. Although weekly visits over the entire season are advisable, the requirement for such effort can also deter a wide range of potential recorders and makes it difficult to recruit new volunteers that could contribute to BMS development. Yet, expanding existing monitoring schemes to wider areas and establishing new schemes in unmonitored regions are critical to investigate broad geographical patterns and better understand the impacts of environmental change. For these reasons, new BMSs are increasingly being established with fewer samples per year. Examples include fortnightly (Israel), monthly (France) or a set number of visits within a peak period in the UK and many USA-based BMSs. Given the wide range in count frequency observed across BMSs (Table 1), a systematic comparison of the methods to estimate abundance indices is timely.

The challenge of estimating abundance indices from BMS data, however, increases when the number of visits per site decreases. This becomes particularly important when analyses are conducted across BMSs with different survey intensities. When only a few weekly visits are missed, values for missing counts can be estimated locally from the counts recorded on either side of the missing observation, hereafter referred to as the ‘linear interpolation’ method. Annual abundance indices can then be estimated from the area under the curve derived from the

**Table 1.** Details of butterfly monitoring schemes within the LOLA-BMS project, linking the sampling frequency suggested and the mean proportion of missing weeks observed between 2006 and 2012

| Butterfly monitoring scheme (BMS) | Year initiated | Number of sites (range/year) | Number of weeks | Sampling frequency | Missing counts % (SD) |
|-----------------------------------|----------------|------------------------------|-----------------|--------------------|-----------------------|
| United Kingdom                    | 1976           | 1409 (806–1013)              | 26              | Weekly             | 36 (3)                |
| Netherlands                       | 1990           | 707 (419–455)                | 26              | Weekly             | 46 (3)                |
| Catalonia (Spain)                 | 1994           | 96 (65–69)                   | 30              | Weekly             | 26 (3)                |
| Finland                           | 1999           | 86 (63–70)                   | 17              | Weekly             | 45 (3)                |
| Germany                           | 2005           | 670 (302–405)                | 26              | Weekly             | 48 (3)                |
| France                            | 2005           | 212 (86–108)                 | 26              | Monthly            | 82 (3)                |
| Israel                            | 2009           | 34 (22–29)*                  | 39              | Biweekly           | 68 (3)                |
| Ohio (US)                         | 1995           | 190 (82–113)                 | 25              | Weekly             | 36 (4)                |
| Illinois (US)                     | 1987           | 92 (57–66)                   | 11              | 6 visits           | 51 (1)                |

\*As of 2010 and 2012.

observed and the estimated weekly counts (Pollard & Yates 1993). While this method is commonly used by a number of BMSs to produce local abundance indices, an alternative approach using a Generalized Additive Model (GAM) fitted at the site level to estimate values for missing counts has been shown to improve reliability of the resulting indices (Rothery & Roy 2001). Nevertheless, a GAM can be highly sensitive to changes in survey intensity and it was therefore recommended only for sites with relatively few missing counts and where the week of peak abundance was sampled (Rothery & Roy 2001).

More recently, Dennis *et al.* (2013) proposed a two-stage modelling approach where a GAM is used to extract the seasonal pattern of the flight-period curve from multiple sites. The resulting curve can then be used to predict values for missing counts at local sites. When applied to BMS data collected in the UK and tested against simulated data with 30% of the weekly counts missing, the two-stage modelling approach showed improved precision of trend estimates. Yet, it is unknown how it performs when monitoring schemes require fewer visits and thus have a larger proportion of missing weeks. While the strength of the Dennis *et al.* (2013) approach over the others resides in its use of data collected at multiple sites to inform the annual pattern in abundance, it does not explicitly account for variability across climatic regions. However, deriving seasonal patterns from regions where emergence and longevity patterns are expected to be similar should considerably increase the predictive power of the two-stage model (Dennis *et al.* 2013). If this is true, strength can also be gained from integrating data across BMSs to better estimate seasonal patterns in the flight period across a species range and produce more accurate abundance indices locally.

This paper aims to assess the performance of the two-stage approach proposed by Dennis *et al.* (2013), but applied within specific climatic regions rather than across sites within a single programme. This adaptation, hereafter called the 'regional GAM', allows the method to be applied across multiple programmes, yet still accounts for variability in flight curves across climatic regions. We focus on exploring how two methods (regional GAM and

linear interpolation) perform in situations where sampling effort each season is relatively low and the proportion of missing counts is consequently high. Specifically, we investigate: (i) the accuracy of abundance indices derived from each method, and (ii) the relative statistical power for detecting a trend, given a range of scenarios of number of monitoring sites. We do so both for univoltine and multivoltine species with simulated count data and examine the impact of excluding data for which week of peak abundance was not sampled. In addition, we also examine the effect of increasing the proportion of missing weekly counts on trend estimates derived from both abundance indices, using observed count data collected for the Gatekeeper *Pyronia tithonus* (Linnaeus 1767), a common and widespread univoltine species with notable declining trends across Europe (Van Swaay *et al.* 2010). In carrying out these comparisons of the impacts of reduced-effort monitoring on model performance, we determine the consequences of having different levels of survey frequencies across BMSs and assess the ability of each approach to perform reliable large-scale analyses by integrating data across schemes with different sampling protocols.

## Materials and methods

### BUTTERFLY MONITORING SCHEMES

Most butterfly monitoring schemes are based on the protocol developed for the original BMS in the UK (Pollard & Yates 1993). When weather conditions are suitable for butterfly activity, observers count all individual butterflies detected along a fixed linear transect route divided into sections which aim to be a homogeneous habitat type or management unit (Van Swaay *et al.* 2008). Butterfly counts are conducted within a 2.5-m distance either side of the line transect and 5 m above and ahead of the observer. Regular counts are made by trained volunteers and reported annually to build long-term time series stored in national data bases.

In an effort to examine butterfly responses to global change, we compiled butterfly counts from eight BMSs distributed across Europe, North America and Israel (within the LOLA-BMS project). The resulting data set enables analyses to be conducted beyond political borders, similar to approaches already conducted

in the framework of Butterfly Conservation Europe (e.g. Van Swaay *et al.* 2012), and open opportunities to better account for ecologically meaningful gradients in our models. Nevertheless, as the number of visits varies between BMSs, differences in the proportion of missing weekly counts are likely to affect the error in abundance indices and restrict comparison of trend estimates between countries.

#### LINEAR INTERPOLATION METHOD

This method applies the trapezoidal rule to estimate the area under the flight curve derived from the observed butterfly counts (eqn 1). Thus, the linear interpolation method accounts for uneven distribution in time, which corresponds to using a linear interpolation from local counts to impute values for missing data. For a series of  $N$  counts  $y_1, y_2, \dots, y_N$ , recorded at time  $t_1, t_2, \dots, t_N$ , the area under the curve (abundance index) can be approximated by:

$$\text{Index} = \sum_{k=2}^N \frac{(y_k + y_{k-1})(t_k - t_{k-1})}{2} \quad \text{eqn 1}$$

While we applied the linear interpolation method in its generalized form, specific procedures have been developed to improve the reliability of this index. Thus, the Dutch BMS computes local abundance indices for distinct generations and restricts analysis to sites where a species was counted at least once within the flight period and where the time between subsequent counts was less than half the duration of the focal generation (Van Swaay, Plate & Van Strien 2002). Such procedures, however, might be too restrictive for BMSs with reduced sampling frequency.

#### REGIONAL GAM METHOD

Dennis *et al.* (2013) suggested a two-stage modelling approach where in a first step, a GAM is fitted across multiple sites to estimate an average flight curve per year, representing the overall variation in butterfly count over time within a specific region and year. Thus, count  $y$  recorded at site  $i$  at day  $t$  ( $y_{it}$ ) is modelled using a GAM with a Poisson distribution and log link function:

$$E[y_{it}] = \mu_{it} = \exp[n_i + s(t, f)] \quad \text{eqn 2}$$

where count  $y_{it}$  is a function of a site effect ( $n$ ) and a smoothing effect over time ( $t$ ) with  $f$  degree of freedom (eqn 2). Here, we estimated the GAM models with the *mgcv* package in R version 3.1.1 (Wood 2006; R Core Team 2014), using a penalized cubic regression spline as basis with the degree of smoothing estimated by general cross-validation (Wood 2006). In a second step, the resulting flight curve is standardized to one ( $\sum \mu_i = 1$ ) and used as an offset in a loglinear model predicting values for missing counts. A local annual abundance index can then be derived from the area under the curve obtained from the observed and the imputed weekly counts (eqn 1).

The regional GAM approach assumes a common flight period across sites. While the two-stage modelling approach was originally applied across all sites in the UK, the authors suggested using geographical stratification to enhance the realism of the assumption and improve the reliability of the method (Dennis *et al.* 2013). Stratification is particularly relevant for data collected over large spatial extents where flight periods are expected to vary across climatic regions (Table S1 in Supporting Information). Here, we

stratified the European BMS data available with the LOLA-BMS project with the bioclimatic regions defined in Metzger *et al.* (2013) and computed annual flight curves for each region (Fig. 1). The regional flight curves were then used to impute values for missing counts and compute local abundance indices (eqn 1).

We assessed and compared the performance of both the regional GAM and the linear interpolation methods with respect to the error in local abundance indices and the statistical power of trend analysis on collated indices derived from a loglinear model that accounts for site and year effects (Roy, Rothery & Breerton 2007). The methods were first examined against simulated butterfly count data, where variation in flight period patterns, temporal trends and proportion of missing data were controlled, and then with real data that we sequentially degraded by increasing the proportion of missing counts.

#### SIMULATED COUNT DATA

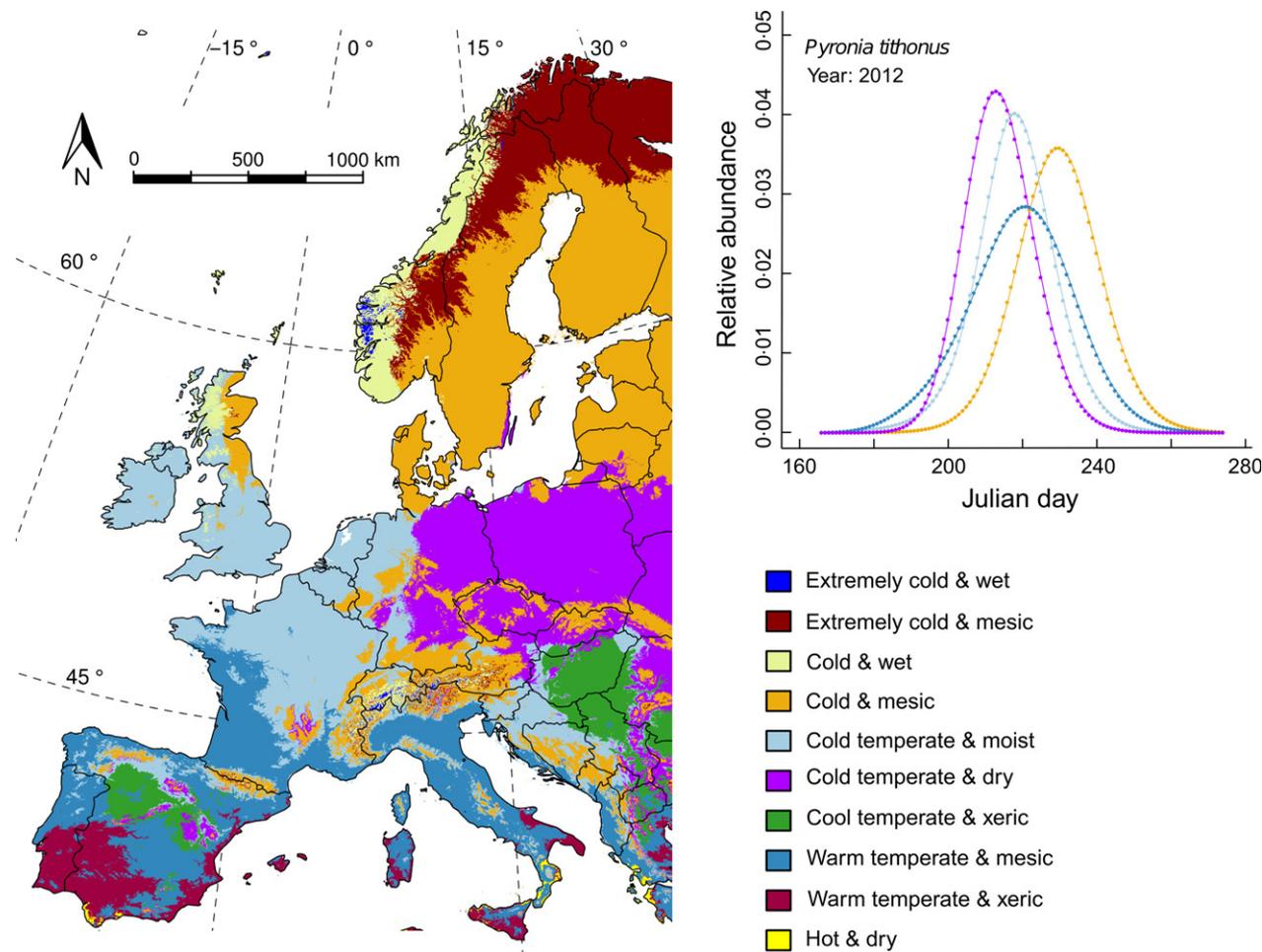
To simulate realistic butterfly count data, we generated an emergence curve for a univoltine species from a generalization of the Zonneveld model (eqn 3), a model that is widely used to describe and analyse the phenology of adult stage (Zonneveld 1991; Calabrese 2012). In this model, the probability distribution function is based on a generalized logistic distribution:

$$f_E(t; \mu, \beta, \delta) = \frac{\delta e^{(t-\mu)/\beta}}{\beta(1 + e^{(t-\mu)/\beta})^{\delta+1}} \quad \text{eqn 3}$$

where  $\mu$  and  $\beta$  are the location and scale parameters, respectively, and  $\delta$  affecting the skewness of the probability distribution and thereby the asymmetry in the emergence curve. To account for variation in seasonal patterns observed across sites and years within a climatic region, the location parameter ( $\mu$ ) of peak abundance was extracted from a normal random variable with a standard deviation of 2 days around a set date ( $190 \pm 4$  days [July 9th] for univoltine). In our simulation, scale ( $\beta = 3$ ) and skewness ( $\delta = 0.15$ ) were kept constant among sites and only the date of peak abundance was allowed to vary. This model was used to generate a probability distribution over the monitoring season, for each site and year. We simulated two data sets, one with 1000 sites monitored over 1 year and a second with 100 sites monitored over a 10-year period (i.e. 1000 independent distributions). For both data sets, we generated an initial total abundance per site, ranging between 35 and 1150 individuals, and in the second one we imposed a declining trend of 10% over the 10-year period. Butterfly counts were then extracted from a random Poisson distribution where the expected count was the total abundance for a specific site and year weighted by its corresponding probability distribution. From the resulting series, we kept 26 weekly counts (one per week from April to September) per site and year, the retained weekdays being selected randomly. To simulate data for a bivoltine species, we used the same procedure, but juxtaposed two non-overlapping probability distributions with two distinct peaks.

#### OBSERVED COUNT DATA

From the data available for the LOLA-BMS project, we extracted butterfly counts for *Pyronia tithonus* to compute abundance indices and trend estimates using both the regional GAM and the linear interpolation methods. This univoltine species was selected for its known decline across Europe (Van Swaay *et al.*



**Fig. 1.** Climate regions across Europe as defined in Metzger *et al.* (2013) with the flight curve of *Pyronia tithonus* observed in 2012 within four regions (Cold & mesic [yellow], Cold temperate & moist [pale blue], Cold temperate & dry [purple], Warm temperate & mesic [blue]).

2010) and for the quality of the data available to establish reliable reference values for both local abundance and temporal trend. We extracted the count data from two schemes where butterfly counts were recorded weekly (Netherlands and United Kingdom), restricting our selection to a 10-year period (2003–2012) and sites found in the ‘Cool temperate and moist’ bioclimatic region (Fig. 1), having at least 20 weekly visits per year and where the species was present for at least 8 years. Following those criteria, we gathered data for an average of 32 sites per year, ranging from 30 to 34, as some sites are not monitored every year. For the regional GAM, we derived the annual flight curves from the data available within the ‘Cold temperate and moist’ climate regions (Germany, Netherlands, United Kingdom, France and Catalonia north-east Spain, Fig. 1). To optimize computation, the GAM was fitted on a random subset of 300 sites where the focal species was detected and with at least four visits during a season.

#### REDUCING SAMPLING EFFORT – DATA DEGRADATION

For protocols recording butterfly count on a weekly basis from April to September, a ‘complete’ data set corresponds to 26 counts. Thus, we systematically degraded the observed and simulated data sets by increasing the number of missing counts from 20 to 80% of the 26 monitoring weeks. In reality, missing counts

are generated by two alternative processes, one determined by the monitoring protocol itself (weekly, biweekly or monthly counts), hereafter defined as ‘structural’, and one relating to sampling behaviour originating from the environment (unsuitable weather conditions) or the observer being unavailable. To account for both processes when degrading the data to a specific proportion of missing values, we applied a stratified randomization procedure. Stratification was applied to account for the structural effect of different sampling protocols (weekly, fortnightly and monthly) and distribute the missing counts across the monitoring season. Missing counts were then generated by randomly removing counts in each stratum. To assess the effect of restricting the analysis to sites where the week of peak abundance was recorded, we compared the results of two degradation processes, one where random sampling was constrained to keep the week where the peak abundance was observed (highest for bivoltine) and the other where this constraint was not applied. For each degradation process, we estimated the proportion of the flight curve that was sampled with the remaining counts (Table S2).

#### ERROR AND STATISTICAL POWER

To assess the sensitivity of local abundance indices to increasing number of missing counts, we ran the degradation process for each site of the first simulated data set (1000 sites). From the

degraded data (20, 30, 40, 50, 60, 70, 80% missing week), we computed abundance indices using both the linear interpolation and the regional GAM approach and estimated the percentage error of each index at the site level, using the undegraded data as a benchmark. Error was estimated for both univoltine and bivoltine species and where the degradation process was fully random or constrained to include the week of peak abundance.

From the second set of simulated data (100 sites, 10 years), we examined the statistical power of detecting declining trends (10%), given five levels of missing counts (0, 20, 40, 60, 80%) and for an increasing number of sites contributing to the collated indices (2–200). For each site and year, we ran 1000 degradation processes for each level of missing counts to estimate the overall trend in abundance with a loglinear model with fixed effect for year, a random intercept for site and a serial correlation over time. Models were fitted with the penalized quasi-likelihood approximation procedure available in the `glmmPQL` function of the MASS package in R version 3.1.1 (Venables & Ripley 2002; R Core Team 2014). Power analyses were conducted for both univoltine and bivoltine species and for constrained and fully random degradation processes.

Because real butterfly count data show important inter-annual variation and uneven temporal trends, we assessed the accuracy of trend estimation on observed count of *P. tithonus* that we degraded to increasing levels of missing counts (0–80%). From each level, we calculated the mean and the standard deviation of the trend obtained from 100 iterative degradation processes and quantified the error relative to the original data with the Root Mean Squared Error (RMSE). These metrics were calculated, given constrained or fully random degradation processes. Additional details and R-scripts are available in online Supporting Information.

## Results

Overall, the regional GAM showed better performance than the linear interpolation method. This was observed for both the error of local indices and the power of detecting a trend in collated indices. The benefit of using the regional GAM over the linear interpolation method was most apparent when butterfly counts covered <50% of the monitoring weeks.

For both methods, errors of local abundance indices increased with the proportion of missing weeks (Figs 2 and S1). When the regional GAM approach is used on data with the peak week, the error of the local abundance index was contained within 50% and evenly distributed around zero (Fig. 2). This pattern was observed for both univoltine and bivoltine species, although the error was slightly larger for bivoltine species (Fig. 2a,c). In contrast, the linear interpolation method was less robust, as the error of local abundance indices increased substantially with the proportion of missing counts and tended to be biased when the proportion exceeded 50% (Fig. 2). For univoltine species, the direction of the bias was positive when the indices were computed on data with the peak week. When the constraint on peak week was not applied, the indices derived with the linear interpolation method tended to display negative biases for both univoltine and

bivoltine species (Fig. S1). These results indicate that indices produced with the regional GAM approach are less prone to show systematic bias and that the expectation of the collated index (across sites) remains unbiased even with low sampling frequencies.

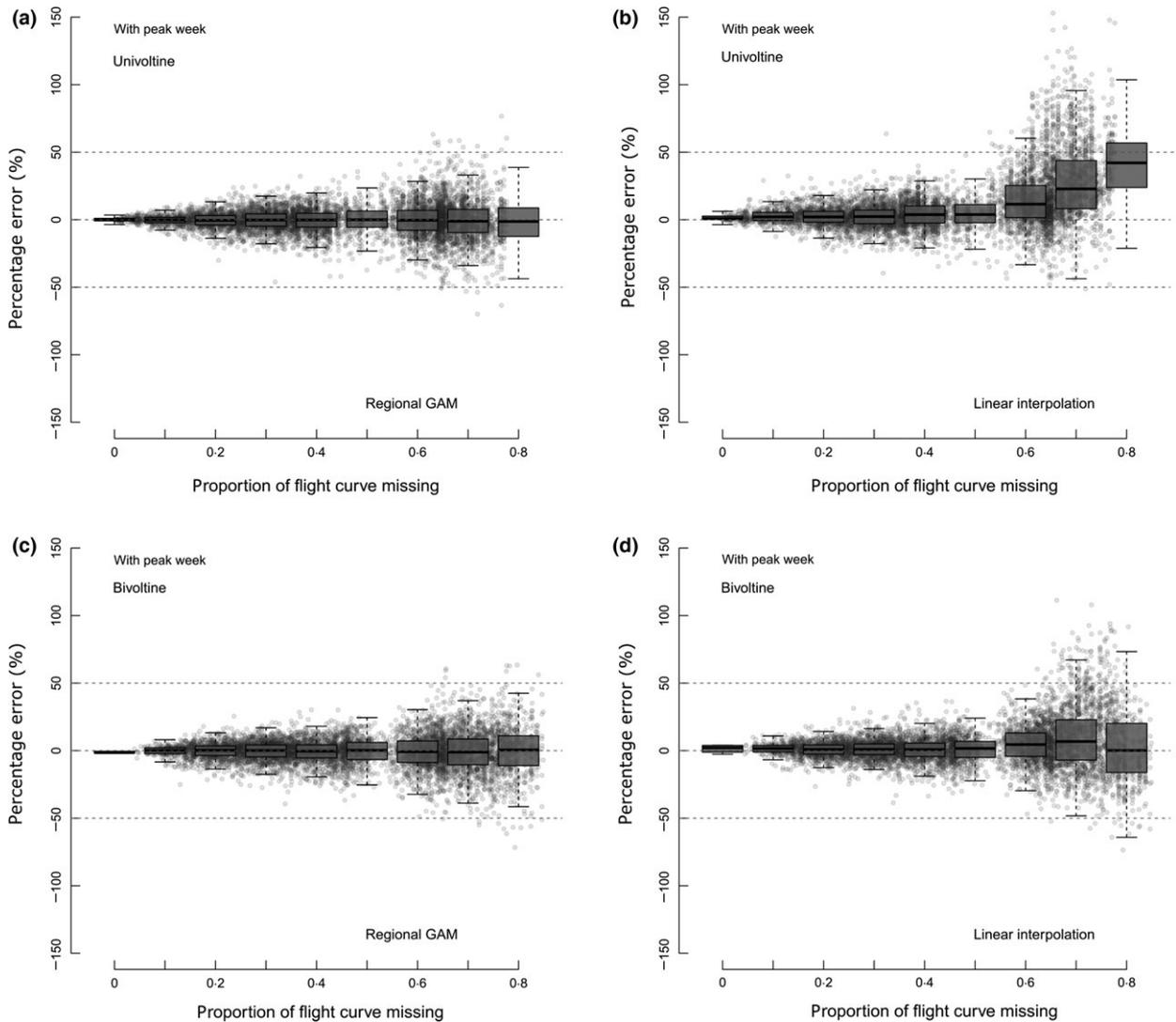
In terms of the statistical power of detecting a 10% decline observed over a 10-year period in simulated data, the collated abundance indices based on the regional GAM method require less sites to reach a power of 80% of trend detection than when based on the linear method (Fig. 3). This holds true for univoltine species, even when 80% of the monitoring weeks were missing (Fig. 3a). For bivoltine species, the regional GAM method required at least 60 sites to reach 80% of detection when 80% of the monitoring weeks were missing (Fig. 3c). With such data, the simulated trend was not detected with the linear interpolation method, even with 200 sites available (Table 2). Note that our results are strictly dependent on the variance included in the simulated data and must be interpreted as measures of the relative ability of the two methods to detect trends, not as a strict recommendation for the number of sites required to detect a decline (see Van Strien *et al.* 1997).

When no constraint on peak week was imposed, the number of sites needed to reach a power of 80% with 80% of the weeks missing increases substantially for both univoltine and bivoltine species (Fig. S2a,c). In all the cases tested with the simulated data, statistical power obtained with the regional GAM approach was superior to the linear interpolation method when the proportion of missing counts exceeded 40% and was at least comparable below this threshold (Figs 3 and S2). In agreement with the patterns observed for the error in abundance indices, statistical power tended to increase when the peak week was observed. Nevertheless, the loss in power associated with the number of sites that would be excluded from the analysis is expected to be more important than the gain in precision (Table 2). Indeed, filtering for peak week would result in excluding three sites out of four in a protocol prescribing one visit per month.

When applied to observed counts (*Pyronia tithonus*), trend estimates derived from collated indices were more robust to data degradation when local indices were computed with the regional GAM compared with the linear interpolation method (Table 3). The improvement in the performance was most important when the proportion of missing counts exceeds 70% and peak week was observed (Table 3).

## Discussion

Because of their ectothermic nature and the short life cycles, butterflies are expected to respond rapidly to climate and land-use change, making them a good indicator group for a large number of terrestrial insect taxa (Thomas 2005; but see Musters, Kalkman & Van Strien 2013; Fleishman & Murphy 2009). butterfly monitoring schemes



**Fig. 2.** Percentage error of abundance index from regional GAM and linear interpolation methods applied on simulated data with increasing proportion of the flight curve missing for two types of life cycle: univoltine and bivoltine. Data were degraded with a stratified randomization procedure constrained to include the peak week. Boxes and whiskers indicate the 5th, 25th, 50th, 75th, 95th percentiles.

are being initiated in a growing number of countries and are thus developing into a large network of programmes, each contributing a substantial amount of data on species status over time and space (Van Swaay *et al.* 2012). With over 4000 transects monitored in more than 20 countries, BMSs represent the second largest network of terrestrial biodiversity monitoring after bird schemes, and undoubtedly the largest network focusing on insects. Their rapid expansion requires support by know-how, guiding both monitoring design and analysis. Here, we show that the regional GAM method is the best approach to compute abundance indices from BMS data when sampling frequency decreases, whereas the commonly used linear interpolation produces both higher errors and directional biases when data become too sparse. This is critical for BMSs with lower intensity survey such as biweekly protocol that will never have more than 50% completed surveys and even fewer if there is a spell of bad weather or

volunteers miss a survey for other reasons. Because the regional GAM can leverage data within climatic regions and offers the flexibility to account for other covariates (e.g. latitude, growing degree-days), it should also be considered the best approach when combining data across BMSs and conducting large-scale analyses.

Assessment and detection of change in species status over time requires abundance metrics that are robust to variation in sampling effort and species biology. While collecting and analysing insect monitoring data present many challenges associated with the complexity of their life cycle, the difficulty in identifying some taxa and the importance of inter-annual variation in their abundance, our results clearly show the benefit of the regional GAM approach when large proportions of weekly counts are missing. Previous methods for estimating abundance indices, such as the one derived from the linear interpolation method and the site GAM (Rothery & Roy 2001),

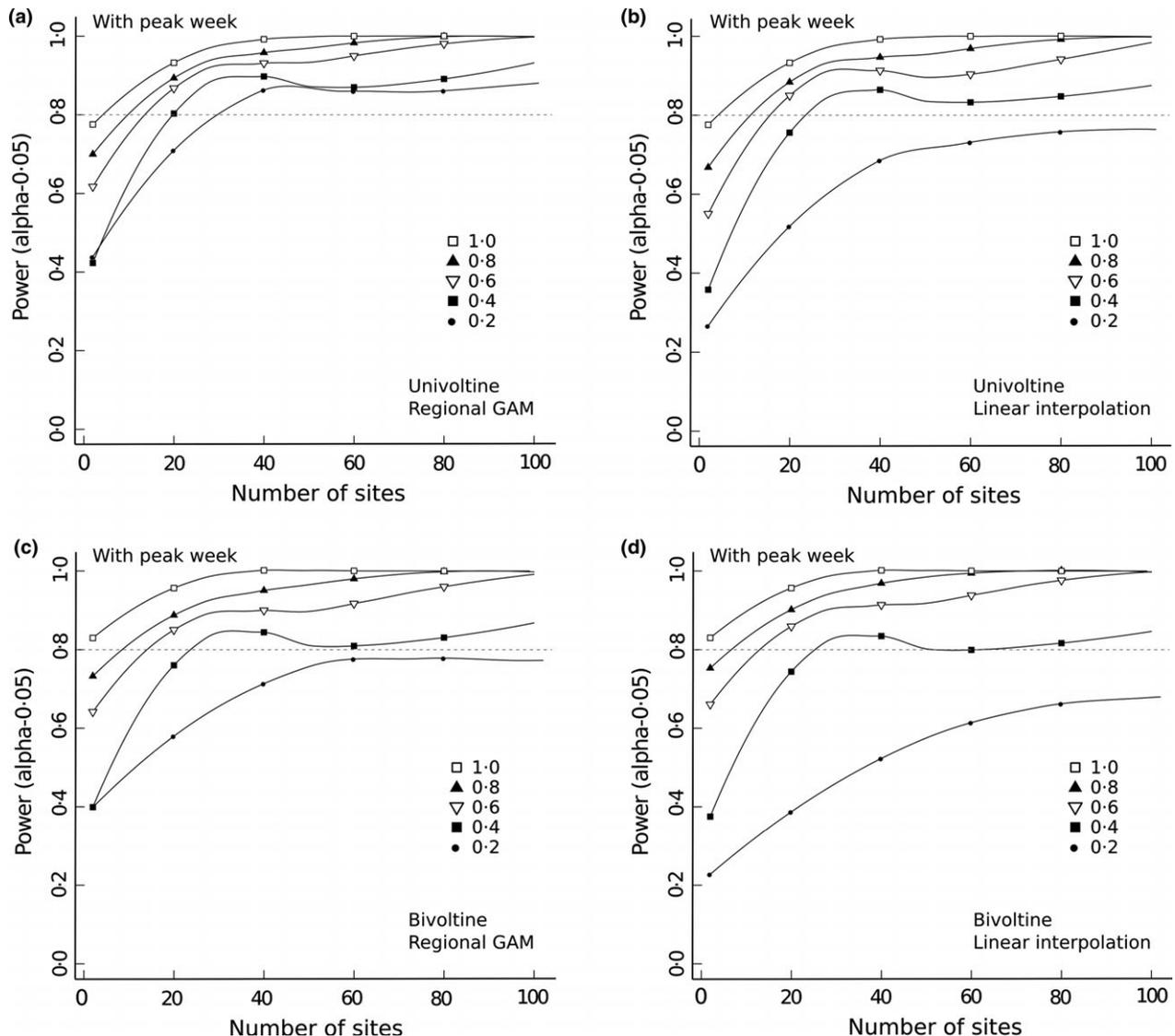


Fig. 3. Probability of detecting simulated 10% decline over 10 years with increasing number of sites, given a decreasing proportion of sampled weeks, when regional GAM and linear interpolation methods are applied with two types of life cycle: univoltine and bivoltine. Data were degraded with a stratified randomization procedure constrained to include the peak week.

fail to produce reliable indices for schemes characterized by reduced sampling efforts and where local count data provide poor information on species' flight period and seasonal patterns in butterfly counts.

By integrating information from multiple sites to model the seasonal pattern in butterfly counts, the two-stage modelling approach used in the regional GAM method greatly improves the accuracy of abundance indices and allows data from all the sites to be used, whereas in the past, sites without sufficient visits were systematically excluded from the analyses (e.g. 38% in the UKBMS, Dennis *et al.* 2013). Because the regional GAM method is based on the assumption that the seasonal pattern in butterfly abundance is shaped by drivers such as climate and inter-annual weather fluctuations, it is crucial that the specific flight curve is computed for climatically comparable regions. This is well illustrated when

contrasting the observed counts with predictions derived from the regional GAM applied to sites distributed across different climatic regions or within a single region (Table S1). The improved performance of the regional GAM highlights the benefit of computing local abundance by drawing from multiple BMSs and thereby providing better coverage of each climatic region. This is highly relevant for BMSs where some climatic regions only occur in a small section of the country, and therefore would most benefit from the additional information contained in an adjacent scheme that covers the same climatic region.

While the overall proportion of missing weekly counts generally predicts increasing errors, a stronger impact on the abundance index is the question of whether observations successfully encapsulate the species' flight period. Thus, when aiming to derive a reliable proxy for species abundance, what matters most is to optimize butterfly

**Table 2.** Mean number of sites (N) required for detecting 10% decline over 10 years with 80% statistical power ( $\alpha = 0.05$ ) with the regional GAM and linear interpolation methods applied on simulated data with increasing proportion of missing weeks for univoltine and bivoltine life cycles. Lower (lci) and upper (uci) confidence intervals (95%) on the number of sites were estimated from a bootstrap procedure. Data degradation was performed with stratified randomization procedures with and without constraint on the inclusion of the peak week

| Life cycle | Proportion of missing weeks | Regional GAM          |                    | Linear interpolation  |                    |
|------------|-----------------------------|-----------------------|--------------------|-----------------------|--------------------|
|            |                             | With peak N (lci-uci) | Random N (lci-uci) | With peak N (lci-uci) | Random N (lci-uci) |
| Univoltine | 0.0                         | 4 (2–6)               | 6 (4–7)            | 4 (2–6)               | 6 (4–7)            |
|            | 0.2                         | 10 (8–12)             | 11 (9–12)          | 11 (10–13)            | 13 (12–15)         |
|            | 0.4                         | 14 (12–15)            | 15 (13–17)         | 16 (14–18)            | 20 (18–22)         |
|            | 0.6                         | 20 (18–22)            | 33 (29–38)         | 24 (22–26)            | 99 (60–137)        |
|            | 0.8                         | 30 (27–33)            | 80 (63–*)          | * (132–*)             | *                  |
| Bivoltine  | 0.0                         | 2 (2–2)               | 2 (2–2)            | 2 (2–2)               | 2 (2–2)            |
|            | 0.2                         | 9 (6–11)              | 9 (7–12)           | 7 (4–9)               | 8 (6–11)           |
|            | 0.4                         | 14 (12–17)            | 15 (13–17)         | 13 (11–15)            | 16 (14–18)         |
|            | 0.6                         | 24 (21–26)            | 23 (21–26)         | 25 (23–28)            | 30 (26–*)          |
|            | 0.8                         | 148 (58–164)          | *                  | *                     | *                  |

\*More than 200 sites.

**Table 3.** Mean trend estimates for *Pyronia tithonus* computed on observed butterfly count data where the proportion of missing weeks was systematically increased with stratified randomization procedures with and without constraint on the inclusion of the peak week. Root Mean Squared Error (RMSE) was calculated relative to the original trend ( $-0.066$  per year)

| Degradation constraint | Proportion of missing weeks | Regional GAM   |       | Linear interpolation |       |
|------------------------|-----------------------------|----------------|-------|----------------------|-------|
|                        |                             | Mean (SE)      | RMSE  | Mean (SE)            | RMSE  |
| With Peak              | 0.2                         | -0.067 (0.002) | 0.003 | -0.067 (0.003)       | 0.003 |
|                        | 0.4                         | -0.067 (0.003) | 0.003 | -0.067 (0.003)       | 0.004 |
|                        | 0.6                         | -0.066 (0.004) | 0.004 | -0.067 (0.005)       | 0.007 |
|                        | 0.8                         | -0.065 (0.004) | 0.004 | -0.081 (0.006)       | 0.017 |
| Random                 | 0.2                         | -0.068 (0.003) | 0.003 | -0.067 (0.003)       | 0.003 |
|                        | 0.4                         | -0.068 (0.004) | 0.005 | -0.067 (0.005)       | 0.005 |
|                        | 0.6                         | -0.068 (0.006) | 0.007 | -0.067 (0.007)       | 0.007 |
|                        | 0.8                         | -0.066 (0.014) | 0.013 | -0.069 (0.018)       | 0.019 |

counts within the flight period. This enables programmes with reduced, but targeted, sampling effort to generate reliable estimates of abundance for specific subsets of species (Roy, Rothery & Brereton 2007; Roy *et al.* 2015). Nevertheless, such sampling protocols alone will fail to provide accurate information for species emerging before or after the specific sampling period. Therefore, monitoring schemes with reduced sampling effort can benefit from adjacent schemes with more intensive effort, but only when survey sites are scattered across the same climatic region. Another cost-effective approach to improve our ability to estimate yearly flight curves is to establish a subset of sites with higher frequency of visits. Determining the best balance between the number of low-frequency sites and the number of high-frequency sites should be a subject of future research.

When aiming to assess butterfly status across a large spatial extent, multiple data sets with different sampling effort need to be collated, making the choice of a method

to calculate a robust abundance index crucial. Furthermore, BMSs with low-intensity survey protocols, until now, have had no basis by which to choose between methods to compute reliable abundance indices, since all previous models were developed specifically for schemes with weekly visits and relatively few missing counts. Here, we show that in both cases, the regional GAM approach produces unbiased abundance indices and has more power to detect trends compared with the linear interpolation method. Our findings can provide guidance to new or growing monitoring schemes by showing the value of tweaking visitation protocols to reach an overall sampling intensity of 50%, including the impact of incidental missing visits. By optimizing the use of information contained in BMS data, the regional GAM method enables integrative analyses across sites with high- and low-intensity visits. This means that BMSs can recruit from a wider range of potential volunteers motivated to sample at different intensities and increase the area of land monitored. Maxi-

mizing the use of data in this way, through the inclusion of data with lower sampling effort and by sharing across schemes, should help support better conservation policy and decision-making from local to continental scales.

## Acknowledgements

This paper is the outcome of LOLA-BMS, a FRB-CESAB project co-financed by the FRB and EDF SA. GP also acknowledges support from the FP7 project EU BON (ref. 308454). LR was supported by NSF (grant no. DBI-1147049). The UKBMS is operated by the CEH, Butterfly Conservation and the British Trust for Ornithology funded by a consortium of government agencies. All BMSs are indebted to all volunteers who contribute data to the different schemes that enabled this study.

## Data accessibility

R-scripts and data associated with this paper are available in online supporting information.

## References

- Brereton, T.M., 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.
- Calabrese, J.M. (2012) How emergence and death assumptions affect count-based estimates of butterfly abundance and lifespan. *Population Ecology*, **54**, 431–442.
- 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.
- Fleishman, E. & Murphy, D.D. (2009) A realistic assessment of the indicator potential of butterflies and other charismatic taxonomic groups. *Conservation Biology*, **23**, 1109–1116.
- Gregory, R.D., Van Strien, A.J., Vorisek, P., Meyling, A.W.G., Noble, D.G., Foppen, R.P.B. & Gibbons, D.W. (2005) Developing indicators for European birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 269–288.
- Hodgson, J.A., Thomas, C.D., Oliver, T.H., Anderson, B.J., Brereton, T.M. & Crone, E.E. (2011) Predicting insect phenology across space and time. *Global Change Biology*, **17**, 1289–1300.
- Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Sayre, R., Trabucco, A. & Zomer, R. (2013) A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, **22**, 630–638.
- Musters, C.J.M., Kalkman, V. & Van Strien, A.J. (2013) Predicting rarity and decline in animals, plants, and mushrooms based on species attributes and indicator groups. *Ecology and Evolution*, **3**, 3401–3414.
- Pollard, E. & Yates, T.J. (1993) *Monitoring Butterflies for Ecology and Conservation - The British Butterfly Monitoring Scheme*. Chapman & Hall, London.
- R Core Team. (2014) *R: A Language and Environment for Statistical Computing*. R Core Team, Vienna, Austria.
- 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. & Brereton, T.M. (2007) Reduced-effort schemes for monitoring butterfly populations. *Journal of Applied Ecology*, **44**, 993–1000.
- Roy, D.B. & Sparks, T.H. (2000) Phenology of British butterflies and climate change. *Global Change Biology*, **6**, 407–416.
- Roy, D.B., Ploquin, E.F., Randle, Z., Risely, K., Botham, M.S., Middlebrook, I. et al. (2015) Comparison of trends in butterfly populations between monitoring schemes. *Journal of Insect Conservation*, **19**, 313–324.
- Sauer, J.R. & Link, W.A. (2011) Analysis of the North American breeding bird survey using hierarchical models. *The Auk*, **128**, 87–98.
- Schmeller, D.S., Henry, P.Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F. et al. (2009) Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, **23**, 307–316.
- Stephens, P.A., Pettorelli, N., Barlow, J., Whittingham, M.J. & Cadotte, M.W. (2015) Management by proxy? The use of indices in applied ecology. *Journal of Applied Ecology*, **52**, 1–6.
- Thomas, J.A. (2005) Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 339–357.
- Van Strien, A.J., Van De Pavert, R., Moss, D., Yates, T.J., Van Swaay, C.A.M. & Vos, P. (1997) The statistical power of two butterfly monitoring schemes to detect trends. *Journal of Applied Ecology*, **34**, 817–828.
- Van Swaay, C.A.M., Plate, C. & Van Strien, A.J. (2002) Monitoring butterflies in the Netherlands: how to get unbiased indices. *Proceedings of the Section Experimental and Applied Entomology of the Netherlands Entomological Society*, **13**, 21–27.
- Van Swaay, C.A.M., Nowicki, P., Settele, J. & Van Strien, A.J. (2008) Butterfly monitoring in Europe: methods, applications and perspectives. *Biodiversity and Conservation*, **17**, 3455–3469.
- Van Swaay, C.A.M., Wynhoff, I., Verovnik, R., Wiemers, M., López Munguira, M., Maes, D. et al. (2010) *Pyronia tithonus*. The IUCN Red List of Threatened Species. www.iucnredlist.org
- Van Swaay, C.A.M., Maes, D., Collins, S., Munguira, M.L., Šasić, M., Settele, J. et al. (2011) Applying IUCN criteria to invertebrates: how red is the Red List of European butterflies? *Biological Conservation*, **144**, 470–478.
- Van Swaay, C.A.M., Collins, S., Dušej, G., Maes, D., Munguira, M.L., Rakosy, L. et al. (2012) Dos and don'ts for butterflies of the habitats directive of the European Union. *Nature Conservation*, **1**, 73–153.
- Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S, Fourth*. Springer, New York.
- Warren, M.S. (1993) A review of butterfly conservation in central southern Britain: II. Site management and habitat selection of key species. *Biological Conservation*, **64**, 37–49.
- Wood, S.N. (2006) *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, Boca Raton.
- Zonneveld, C. (1991) Estimating death rates from transect counts. *Ecological Entomology*, **16**, 115–121.

Received 27 April 2015; accepted 23 October 2015

Handling Editor: Jörg Müller

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Details of data degradation and missing count.

**Appendix S2.** R-script for simulation.

**Appendix S3.** R-script with worked examples of index computation.

**Appendix S4.** BMS data for *Pyronia tithonus*.

**Figure S1.** Percentage error computed without constraint on data degradation.

**Figure S2.** Power analysis without constraint on data degradation.

**Figure S3.** Distribution of missing observation weeks per monitoring season in seven butterfly monitoring schemes (BMS).

**Table S1.** Regional GAM indices computed in climate region and at the national scale.

**Table S2.** Proportion of the flight curve sampled per degradation level.