The geometric mean of relative abundance indices: a biodiversity measure with a difference

STEPHEN T. BUCKLAND,[1],† ANGELIKA C. STUDENY,[1] ANNE E. MAGURRAN,[2] JANINE B. ILLIAN,[1] AND STUART E. NEWSON[3]

[1]Centre for Research into Ecological and Environmental Modelling, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, Fife KY16 9LZ United Kingdom
[2]School of Biology, University of St Andrews, Fife KY16 8LB United Kingdom
[3]British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU United Kingdom

Abstract. The 2010 Biodiversity Target of the Convention on Biological Diversity (CBD), set in 2002, which stated that there should be 'a significant reduction of the current rate of biodiversity loss' by 2010, highlighted the need for informative and tractable metrics that can be used to evaluate change in biological diversity. While the subsequent Aichi 2020 targets are more wide-ranging, they also seek to reduce the rate of biodiversity loss. The geometric mean of relative abundance indices, G, is increasingly being used to examine trends in biological diversity and to assess whether biodiversity targets are being met. Here, we explore the mathematical and statistical properties of G that make it useful for judging temporal change in biological diversity, and we discuss its advantages and limitations relative to other measures. We demonstrate that the index reflects trends in both abundance and evenness, and that it is not prone to bias when detectability of individuals varies by species. We note that it allows data from different surveys to be combined to generate a composite index. However, the index exhibits high variance and unstable behaviour when rarely-recorded species are included in the analyses.

† **E-mail:** steve@st-andrews.ac.uk

INTRODUCTION

The 2010 Biodiversity Target of the Convention on Biological Diversity (CBD), agreed upon in 2002, marked a change in perspective on how biodiversity should be measured (Butchart et al. 2010). The target stated that there should be 'a significant reduction of the current rate of biodiversity loss' by 2010. While it has now been superseded by 20 targets for 2020 (CBD 2011), the focus is still on how to reduce biodiversity loss. Thus, effective measures of biodiversity trend from long-term datasets are required, to assess success or failure in meeting the targets (Pereira and Cooper 2006, Mace and Baillie 2007, Magurran et al. 2010). As these targets were agreed upon by nations, it is reasonable to assume that we need to measure the biodiversity of nations—in other words large geographic regions, as opposed to specific sites.

Key indicators of biodiversity have adopted G, the geometric mean of relative abundance indices, as their biodiversity measure. These include the international Living Planet Index (Loh et al.

2005, Hails et al. 2008; Fig. 1) and the UK's Wild Bird Indicators (BTO 2011, Gregory et al. 2008, Gregory and van Strien 2010; Fig. 2). Indices such as these are key to identifying whether targets such as the 2010 target have been met (Butchart et al. 2010). While Buckland et al. (2005) investigated the value of the geometric mean as a composite index, and compared it with other indices, its properties have not been investigated in the context of the CBD targets. In this paper, we explore the properties of G and make recommendations on how best to exploit these when drawing inference on biodiversity trends.

Many classical measures of biodiversity quantify one or a combination of the following components: number of species (termed species richness), total abundance, and evenness (Magurran 2004, Buckland et al. 2005). Evenness refers to the degree of uniformity of the species proportions p_i, where $p_i = n_i/n$ and is the proportion of the count n of individuals that is of species i. To help understand the advantages and limitations of G relative to classical measures based on these species proportions, we compare and contrast it with the Shannon index $H = -\sum p_i \log_e p_i$ (Shannon and Weaver 1949) and the transformation $-\log(D)$ of Simpson's index $D = \sum p_i^2$ (Simpson 1949).

THE GEOMETRIC MEAN AND RELATIVE ABUNDANCE

Consider first survey data from a single site. Suppose for species i in a community, we record counts n_{ij} for years $j = 1, ..., J$. Suppose further that we have a known number S of species in the
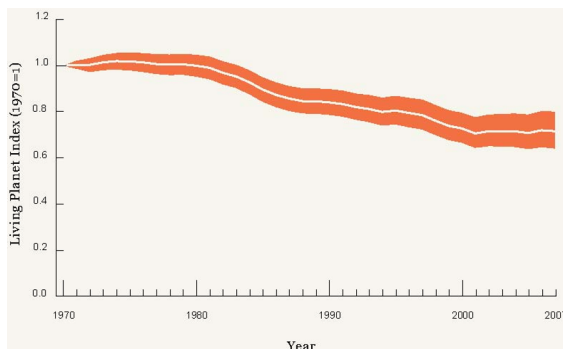

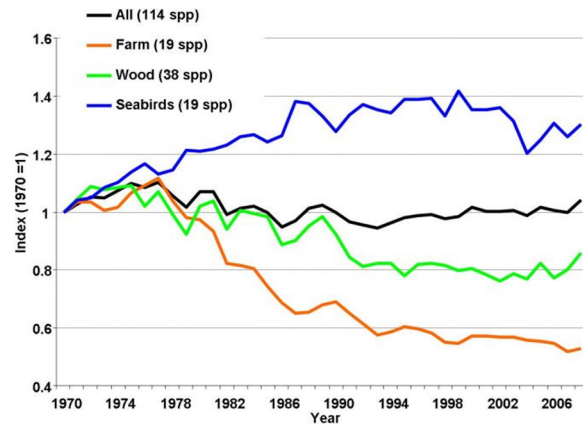Fig. 1. The Living Planet Index with 95% confidence limits. (Source: WWF, ZSL, GFN.)

Fig. 2. Estimated trends in biodiversity, as assessed by the geometric mean of annual relative abundance indices, of woodland birds, farmland birds and seabirds in the UK. Also shown is the trend corresponding to data on all species for which there are sufficient data (114 species). (Source: Defra, BTO, RSPB and JNCC.)

community, so that $i = 1, ..., S$. We convert our counts to measures of relative abundance by defining a baseline year, assumed here to be year 1, then calculating relative abundance for species i in year j as n_{ij}/n_{i1}. It is important to note that abundance of species i is thus relative to *that species'* abundance in the baseline year; it is *not* abundance of species i relative to the abundance of other species. This contrasts with the species proportions p_i, which are also sometimes called relative abundances, as they are assumed to reflect the abundance of each species relative to other species.

Note that relative abundance as defined above is best regarded as a multiplicative measure. If counts for one species increase from 100 to 200, while those for another species decrease from 200 to 100, the respective relative abundances of the two species at the second time point are 2 and 0.5. In additive terms, the first species has shown greater change than the second. However, regarded as a multiplicative measure, one rate is the inverse of the other ($0.5 = 1/2$), giving self-consistency (Gregory and van Strien 2010). When averaging a multiplicative measure, it is natural to use a geometric mean rather than an arithmetic mean. Equivalently, it is natural to convert to a log scale (for which $\log 0.5 = -\log 2$) and then to

take the arithmetic mean on this scale. Back-transforming this mean gives the geometric mean on the original scale.

Thus we can define the index in year j as

$$G_j = \exp\left(\frac{1}{S}\sum_{i=1}^{S} \log \frac{n_{ij}}{n_{i1}}\right). \quad (1)$$

To allow inference to be drawn on a region given data from sample sites, a randomized survey design should be used (Theobald et al. 2007). Such a design allows regional abundance of species within the community to be estimated, as in our case study below. We can then calculate the index based on abundance estimates instead of counts:

$$G_j = \exp\left(\frac{1}{S}\sum_{i=1}^{S} \log \frac{\hat{N}_{ij}}{\hat{N}_{i1}}\right) \quad (2)$$

where $\hat{N}_{ij}$ is the estimated abundance of species i in year j for the surveyed region.

PROPERTIES OF THE GEOMETRIC MEAN INDEX

Because the index G is a mean of trends in relative abundance, it is natural to assume that it reflects only changes in abundance. However, the use of a geometric mean rather than an arithmetic mean has implications that are best illustrated through an example. Suppose we have a 'community' of just three species, censused at two time points, giving the abundances of Table 1. We show the percentage change in several biodiversity measures between the two time points in Table 2. On an additive scale, abundance has not changed between years 1 and 2. However, the effect of using the geometric mean is that we

Table 1. Abundances of three species at two time points of a hypothetical community. Also shown are relative abundances at the second time point (relative to the first), and species proportions.

Species, i	1	2	3
Abundance of species i in year 1, N_{i1}	800	160	40
Abundance of species i in year 2, N_{i2}	900	80	20
Relative abundance of species i in year 2, N_{i2}/N_{i1}	1.125	0.500	0.500
Proportion of species i in year 1, $p_{i1} = N_{i1}/\sum_{i'} N_{i'1}$	0.800	0.160	0.040
Proportion of species i in year 2, $p_{i2} = N_{i2}/\sum_{i'} N_{i'2}$	0.900	0.080	0.020

Table 2. Biodiversity measures for the hypothetical community of Table 1.

Measure	Year 1	Year 2	% change
Arithmetic mean of abundance	333.3	333.3	0.0%
Geometric mean of relative abundance G	1.000	0.655	−34 .5%
Shannon index H	0.600	0.375	−37.5%
Log-transformed Simpson's index $-\log(D)$	0.405	0.202	−50.0%

work on a multiplicative scale. The relatively small changes in absolute terms of the rare species represent large percentage changes, so changes in these rare species dominate the index G. We conclude that there has been a substantial reduction in biodiversity, by this measure; the practical effect of working on a multiplicative scale is that trends in evenness generate trends in the index even when overall abundance is not changing.

To clarify this further, suppose we have abundances N_{ij} of species i ($i = 1, ..., S$) in year j. Assuming that the abundances are known, the index G_j for year j is given by

$$G_j = \exp\left(\frac{1}{S}\sum_{i=1}^{S} \log \frac{N_{ij}}{N_{i1}}\right)$$
$$= \exp\left(\frac{1}{S}\sum_{i=1}^{S} (\log N_{ij} - \log N_{i1})\right).$$

Note that $G_1 = 1$, and so the above represents a multiplicative change between years 1 and j, G_j/G_1. Assuming that the total number of species S is constant, we can infer that $G_j < G_1$ if and only if the mean of the log abundances in year j is less than the mean of the log abundances in year 1. Now consider the multiplicative change in species proportions between years 1 and j: $p_{ij}/p_{i1} = \exp\{(\log N_{ij} - \log N_j) - (\log N_{i1} - \log N_1)\}$, where $N_j = \sum_i N_{ij}$ is total abundance in year j. If overall abundance remains constant, then $N_j = N_1$ and hence $p_{ij}/p_{i1} = \exp\{\log N_{ij} - \log N_{i1}\}$ and

$$G_j = \exp\left(\frac{1}{S}\sum_{i=1}^{S} \log p_{ij} - \frac{1}{S}\sum_{i=1}^{S} \log p_{i1}\right).$$

The mean of the log species proportions has the key property of an evenness measure, in that it attains its maximum value when all the species proportions are equal: $p_{ij} = 1/S$ for all i (Smith and Wilson 1996). Thus when overall abundance and

number of species are constant, G_j may be regarded as a measure of the change in evenness from year 1 to year j. When overall abundance is changing, changes in G reflect changes in both abundance and evenness.

Advantages

Limpert et al. (2001) discuss advantages of using a geometric mean in a more general context, and Buckland et al. (2005), Lamb et al. (2009), Gregory and van Strien (2010), and O'Brien et al. (2010) consider its merits in the context of wildlife surveys.

The index G_j reflects trends in abundance: if all species are declining at the same rate (so that there is no trend in evenness), then G_j will decline at this rate. By contrast, the Shannon and Simpson's indices will show no trend. Similarly, if species trends are variable and predominantly negative, and are uncorrelated with how common each species is, G_j will decline, while the Shannon and Simpson's indices will again remain roughly constant. However, as noted above, G_j also reflects trends in evenness.

The index G_j is unaffected if detectability varies by species, as it is based on within-species trends; if detectability of individuals of a given species does not change over time, we do not need to estimate detectability to avoid bias, regardless of whether detectability varies among species. To see this, denote the probability that an individual of species i is detected, given that it is on a surveyed plot, by π_i, independent of year. Denote the estimated expected count of species i in year j on a random plot within the region by $\hat{E}(n_{ij})$. Given a randomized survey design, we can estimate this quantity. Estimated abundance $\hat{N}_{ij}$ is then $M \times \hat{E}(n_{ij})/\pi_i$ where M is the total number of plots in the region (whether sampled or not). When we substitute this into Eq. 2, the unknown probability π_i cancels, allowing us to evaluate the index. By contrast, the Shannon and Simpson's indices are biased when detectability varies by species, unless counts are corrected using species-specific estimates of detectability (Buckland et al. 2010). However, G_j is likely to suffer greater bias than the Shannon and Simpson's indices if there is a trend in detectability over time within species—unless we estimate and correct for detectability. (A trend in detectability that is common to all species does not affect measures

based on species proportions.)

Two important advantages arise because G_j is based on within-species trends, standardized to a baseline year. First it makes no difference whether we use counts of individuals or biomass to quantify abundance, provided there is no trend over time within species in mean weight of individuals. (For heavily-exploited fish stocks for example, there may be a downward trend in the mean size of fish (Fisher et al. 2010), so that G_j calculated from biomass would show greater reduction than G_j calculated from counts of individuals.) Second we can readily combine trends obtained from different surveys, which is not possible for classical measures of biodiversity. Thus the Living Planet Index combines trends in relative abundance of nearly 5,000 populations, representing nearly 1,700 species of mammal, bird, reptile, amphibian and fish, while the UK's Wild Bird Indicators combine trends from a number of different surveys. The geometric mean is thus a very natural method to adopt when we wish to construct composite indices across surveys, regions or communities.

A consequence of the above two advantages is that we can combine relative abundance trends from surveys that use different units of measurement. For example, trends in a plant species might be quantified using percent cover, those for a bird species using counts, and those for a fish species using biomass. We can legitimately combine these different trends into a composite index.

When combining trends from different surveys, the issue that the surveys may span different time periods must be addressed. If for example a new survey started in 2008, the 2008 index from the new survey can be scaled to equal the composite index in 2008, ensuring that it does not affect trends up to 2008, but does subsequently. A similar approach can be used if for example a previously-common species becomes too rare to include in the index; in the final year that it is included, the index is calculated both with and without the data for this species, and the latter rescaled to match the former. Any subsequent estimates are scaled by the same amount.

Disadvantages

A major limitation of G_j is that it cannot be

calculated if any of the relative abundance estimates are zero. Thus if a species is not recorded in a given year, the index cannot be evaluated. We could add a small quantity to zeros (O'Brien et al. 2010), but the index is sensitive to the quantity chosen, and has poor precision if rarely recorded species are included. Hence typically, species with small sample sizes in some or all years are excluded from analysis. Thus it is not a useful measure if primary interest is in rare, or rarely recorded, species, or in species that are not consistently present in the community. However, most biodiversity measures perform poorly in such cases. The index G_j gives equal weight to all species (Zar 1999), and so is sensitive to changes in rarely-recorded species, but also has poor stability and precision when such species are included; by contrast, indices such as the Shannon and Simpson's are very insensitive to changes in rarely-recorded species, which allows them to have high stability and precision. If we include rarely-recorded species when calculating G_j, problems are reduced but not eliminated by developing a model for counts, and replacing the observed counts by the corresponding predicted counts before evaluating the index.

Related to the above problem, we assume that the number of species S in the community is known. We avoid this issue by monitoring only species that are observed in our samples in sufficient numbers, so that trends reflect the subset of species we analyse, not the whole community. Periodically, the species included may need to be revised, and the time series of estimates recalculated. This is likely to be necessary if conditions change or if some previously dominant species become rare or vice versa. As noted above, by appropriate rescaling of the index, species can drop into or out of the index to accommodate changes in the community.

If it is feasible to conduct species-specific surveys of the rare species of interest, then the surveys can be designed to ensure adequate sample sizes of rare species. Because the index is derived by taking an average of within-species relative abundance estimates, we can estimate trends for each species independently, before combining them. This is the philosophy underlying the Living Planet Index. However, for highly diverse communities, where most species are unrecorded or seldom recorded, as with tropical arthropods for example (Coddington et al. 2009), the index G_j is likely to be of limited value.

Note that G_j is entirely a relative measure—relative to the baseline year. It is only useful for looking at time trends within a site or region, not for comparing sites or communities (although of course we may want to compare within-site time trends across sites or communities).

TESTING FOR TREND

When the survey design comprises a set of sample plots based on a randomized scheme, a natural and straightforward way to quantify precision of biodiversity measures is to use the nonparametric bootstrap, with plots as the resampling unit. If for example we have a stratified scheme based on a random sample of 1 km squares from each stratum, we can generate a bootstrap resample by sampling the 1 km squares with replacement within each stratum, so that the number of sampled squares in each stratum is the same as for the original sample. (When a plot is selected for a resample, the entire time series of data from that plot is included.) We then analyse the resample exactly as for the original sample. This is repeated for a large number of resamples (typically around 1000 or more), and the variation in bootstrap estimates used to quantify precision. For example if we use the percentile method to estimate 95% confidence limits for the annual biodiversity measure, and we have say 999 bootstrap resamples, we order the bootstrap estimates of biodiversity for a given year from smallest to largest, and extract the 25th smallest and 25th largest estimates as confidence limits (Buckland 1984).

A disadvantage of this approach when using the geometric mean of relative abundances is that the first point (corresponding to the baseline year) has zero variance (the index is necessarily 1), then confidence intervals steadily get wider and less useful over time. This effect is evident in the Living Planet Index (Fig. 1). A related issue is that, if the baseline year is year 1 of a scheme, and the effort in that year was low relative to later years, the estimated diversity for later years is relative to year 1 for which there is poor

precision, compromising the precision of the entire time series. Even if precision was good in the baseline year, precision for a subsequent year is driven by variance both in the baseline year and in the subsequent year.

The 2010 target itself suggests a solution to this problem. It states that there should be 'a significant reduction of the current rate of biodiversity loss' by 2010. The slope or first derivative of the trend curve quantifies the rate of change in biodiversity. If we wish to draw inference about change in the rate of change, then this corresponds to the second derivative of the trend curve. Fewster et al. (2000) estimated the second derivative numerically to identify time points at which the slope of the trend changed for species-specific trends. They used the bootstrap as described above to obtain confidence intervals for the second derivative, and identified years in which the confidence interval did not span zero as likely change-points. Buckland et al. (2005) applied the same approach to the geometric mean of relative abundances, to identify years in which there was a change in the rate of change in the overall biodiversity measure. These results are independent of choice of baseline year (i.e., they are unaffected if the baseline year is changed), and confidence interval length does not increase with increasing length of the time series.

Unless an index incorporates relative abundance estimates from a large number of datasets, it is likely to show short-term fluctuations. Figs. 1 and 2 illustrate this; the Living Planet Index is based on nearly 5000 time series of relative abundance estimates, so that the trend curve is very smooth. The Wild Bird Indicators of Fig. 2 however show short-term fluctuations. If we only wish to draw inference on longer-term trends in biodiversity, we may wish to smooth the short-term fluctuations out by applying some scatterplot smoother. This can be done either by smoothing the species-specific relative abundance trends before the geometric mean is calculated or by directly smoothing the index once it has been derived from the (unsmoothed) abundance estimates. Fewster et al. (2000) and Buckland et al. (2005) used generalized additive models (Hastie and Tibshirani 1990) to obtain smoothed trends in this way. Other possible methods include kernel regression, locally weighted regression and running-median smoothers (Hastie and Tibshirani 1990). By using smoothed time series, we improve precision for detecting long-term trends, while reducing the number of change-points detected: those corresponding to short-term variation in trend, caused perhaps by weather effects, will no longer be identified.

Very often, the raw data are not available when a composite index is formed. Provided the estimated trends are available for each species separately, one option is to implement the bootstrap by resampling species instead of sites (Buckland et al. 2005). This treats species as a random effect, and would be appropriate if the species were a random subset of the species in the region and community of interest. When there is no sub-sampling of species (so that all relevant species that are encountered are recorded), the method tends to generate pessimistic estimates of precision, so that real changes in trend are more likely to go undetected. Butchart et al. (2010) formed a composite index from nine different indicators of trend, and used the bootstrap to quantify precision, presumably by resampling these nine indicators.

Example: Biodiversity Trends in UK Birds

We consider data from the British Breeding Bird Survey (BBS) for the years 1994–2008 (Newson et al. 2005, 2008, Freeman et al. 2007). We exclude data for 2001, as access to many sites was not possible in that year due to an outbreak of foot-and-mouth disease. Volunteer observers survey 1 km^2 plots, selected according to a stratified random sampling scheme, where the sampling rate is proportional to the number of available volunteers in each stratum. Stratification is by regions which correspond roughly to UK counties.

Plots are surveyed using line transect sampling (Buckland et al. 2001). In each assigned plot, an observer walks along two parallel 1 km transects and records every bird detected in one of four categories—within 25 m of the line, between 25 and 100 m of the line, beyond 100 m, or flying over. In accordance with Newson et al. (2008), we only consider the first two categories here. Each plot is visited twice during the breeding season. We use data from the early visit only, except for

summer migrants, for which we use data from the late visit only.

Following standard distance sampling methods, we assume that all birds on the line (i.e., at distance 0) are detected and that the probability of detection then drops with increasing distance from the line (Buckland et al. 2001, 2004). This fall-off is modeled by specifying and fitting a detection function, which was chosen to be half-normal here. A single model across all years and sites was fitted for each species by maximum likelihood estimation. We used AIC to select between three models. In the first, year was included as a factor covariate, allowing detectability to vary by year. In the second, year was included as a continuous covariate, allowing a trend in detectability. For the third, detectability had no dependence on year. Given estimates of the detection probabilities for individuals of species i in year j, abundance of the UK population for species i in year j can be estimated as

$$\hat{N}_{ij} = \sum_r \frac{A_r}{m_{jr}a}\Big[\sum_s \sum_k \frac{1}{\hat{\pi}_{ijksr}}\Big]$$

where $\hat{\pi}_{ijksr}$ is the estimated detection probability of the kth detected bird at site s in region (stratum) r. Within a plot, we have two strips each of length 1 km and half-width 100 m, giving $a = 0.4$ km^2 as the survey area covered per plot, m_{jr} is the number of plots visited in year j in region r, and A_r is the size of that region.

We present analyses for a group of 23 species classified as woodland/park/garden birds in Gregory et al. (2005) (Table 3). The list was restricted to those species considered to be representative; they include both specialist (with respect to habitat use) and non-specialist species.

A generalized additive model (GAM) was used to smooth the time series of density estimates for each species. To fit the GAM, we calculated mean counts (weighted to allow for stratification) from the original data and included an offset term for the detectability conversion to density estimates. A gamma error distribution was assumed for the mean counts, together with a log link function. The smoother uses thin-spline regression and was given an upper limit of 3 df where the actual df is determined during the model fitting procedure by cross-validation (Wood 2006). The density estimates were then

Table 3. UK bird species classified by experts as representative of the woodland/park/garden bird community (Gregory et al. 2005).

Common name	Scientific name
Sparrowhawk	*Accipiter nisus*
Buzzard	*Buteo buteo*
Great spotted woodpecker	*Dendrocopus major*
Jay	*Garrulus glandarius*
Goldcrest	*Regulus regulus*
Blue tit	*Cyanistes caeruleus*
Great tit	*Parus major*
Coal tit	*Periparus ater*
Long-tailed tit	*Aegithalos caudatus*
Chiffchaff	*Phylloscopus collybita*
Willow warbler	*Phylloscopus trochilus*
Blackcap	*Sylvia atricapilla*
Garden warbler	*Sylvia borin*
Wren	*Troglodytes troglodytes*
Blackbird	*Turdus merula*
Song thrush	*Turdus philomelos*
Mistle thrush	*Turdus viscivorus*
Spotted flycatcher	*Muscicapa striata*
Robin	*Erithacus rubecula*
Redstart	*Phoenicurus phoenicurus*
Dunnock	*Prunella modularis*
Tree pipit	*Anthus trivialis*
Chaffinch	*Fringilla coelebs*

Notes: We omit wryneck (*Jynx torquilla*), which has not been recorded at any BBS site. Data from the early visit only was used for all species except tree pipit, chiffchaff, willow warbler, blackcap, garden warbler, spotted flycatcher and redstart, for which data from the late visit only were used.

scaled up to give UK abundance estimates for each species. Relative abundance estimates and species proportions for all species, calculated from both the smoothed and the unsmoothed abundance estimates, can be found in the Appendix.

In Fig. 3, we show the geometric mean of relative abundance estimates plotted against year, and also the Shannon and Simpson's indices calculated from the species proportions, where both the relative abundance estimates and the species proportions were calculated from smoothed absolute abundance estimates. Change-points in trends were determined through a numerical approximation of the second derivative of the curve (Fewster et al. 2000). A nonparametric bootstrap as described above was used to quantify precision, and 95% percentile confidence intervals for both indices are shown in Fig. 3. Confidence intervals were also calculated for the second derivative, and points for which the interval did not span zero are indicated. These reveal likely change-points in long-term trends in biodiversity.
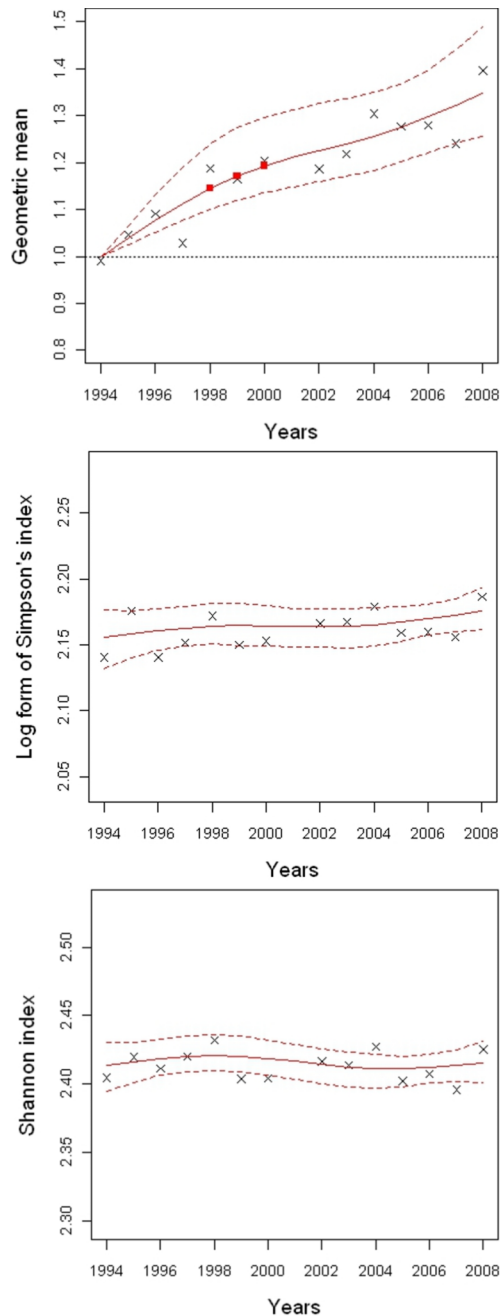
The geometric mean for the UK woodland and park bird community has increased appreciably since 1994, although the rate of increase slowed in the late 1990s. Neither the Shannon index nor Simpson's index identify this increase, because it reflects increased abundance rather than increased evenness of the community. Nor does either index show significant changepoints.

Note that the indices could be calculated from unsmoothed abundance estimates, followed by smoothing of the indices. We found that inference was little affected by which of these two options was adopted. An advantage of smoothing the abundance estimates first is that, if any species has a zero count, we cannot calculate the geometric mean of unsmoothed counts. However, inclusion of species that can have zero counts compromises precision and stability of results, so this advantage is of little significance.

Discussion

Nichols and Williams (2006) argue that surveillance monitoring—i.e. monitoring trends unguided by a priori hypotheses—is frequently an inefficient option. They favor targeted monitoring. While it is true that species-specific trends from omnibus surveys are often estimated with poor precision, so that management action is triggered only belatedly, in a biodiversity context, precision is improved by combining data from a number of species into a composite index. Further, it is generally not possible to implement targeted monitoring of sufficient species to allow effective monitoring of the biodiversity of an entire region.

We have shown that trends in G reflect trends in both abundance and evenness. This raises the question of how a decline in G should be interpreted. In the absence of other measures, we cannot know whether the decline reflects a decline in overall abundance or a decline in evenness, or some combination of both. Indeed, it may be that one of these two components is increasing, but more than offset by a decline in

Fig. 3. Geometric mean of relative abundance estimates (top), Simpson's index on a log scale ($-\log(D)$, middle), and the Shannon index (bottom) for the woodland/park/garden community. All were calculated from abundance estimates that had first been smoothed using generalized additive models. Unsmoothed index values are shown as crosses. A square indicates a point at which the long-term trend in biodiversity has changed for the worse (corresponding

←

to a 95% confidence interval for the second derivative that is entirely in the positive range). Dashed lines show pointwise 95% confidence limits for the indices.

the other. In our example, we show that further interpretation is possible, if we also calculate additional measures. Fig. 3 shows a strong positive trend in G, while both the Shannon and Simpson's indices suggest little change, indicating that the increase in G is primarily due to an upward trend in abundance rather than in evenness.

By using a measure such as G that is sensitive to changes in abundance of rarely-recorded species, precision may be poor relative to measures that are insensitive to such changes, such as the Shannon and Simpson's indices. Studeny et al. (2011) address this insensitivity by embedding the Shannon and Simpson's indices in a parametric family of indices, where a parameter controls the relative weighting given to rare and common species in the community. This allows a trade-off between precision and sensitivity to changes amongst rarely-recorded species.

The effect of invasive species on biodiversity trend estimates should be borne in mind. The absence of trend in a biodiversity index might mask a reduction in native species, offset by an increase in invasive alien species. A simple solution to this issue is to omit alien species. However, in the context of climate change and of consequent changes in natural ranges, it may be difficult to distinguish undesirable aliens from welcome additions to the natural fauna or flora of a region.

We conclude that G provides an effective way of combining time trends in relative abundance across species and surveys to assess whether biodiversity targets have been met. Although conceptually a measure of trends in abundance, it also reflects trends in evenness. Further properties are that it is not prone to bias when detectability varies by species, and it allows data from different surveys to be combined to generate a composite index, even when units of measurement differ between surveys. On the other hand, the index exhibits high variance and unstable behaviour when rarely-recorded species are included in the analyses. While inclusion of such species in classical measures such as the Shannon and Simpson's indices causes no statistical difficulties, there is also negligible gain from including them, as noted above. For these reasons we do not believe that the exclusion of rarely-recorded species is necessarily problematic for regional biodiversity monitoring, and that this constraint should not be considered a major failing of the index G.

The deeper question is which aspect of the assemblage we want our chosen metric to emphasize. As Gaston (2011) argues, common species are important contributors to ecosystem function, structure, biomass and energy turnover. By contrast, rare species generally are not. For example, if a species that occurs in 100% of sites declines by 50%, then we might expect a greater impact on ecosystem function than if a species that occurs in only 1% of sites declines by 50%. Thus, one view might be that it is the trends in the common species that we most need to monitor. However, rare or declining species, such as some pollinators (Fitzpatrick et al. 2007), may play important functional roles. Moreover, the relationship between species richness and function is complex (Tilman et al. 2006, Hector and Bagchi 2007, Creed et al. 2009, Jain et al. 2010). Conservation biologists often target rare species or wish to protect areas of high species richness, and in this context the ability to evaluate trends in low abundance taxa or track richness may be crucial (Gotelli and Colwell 2010). G is a useful tool for biodiversity monitoring, but to use it effectively, it is essential to appreciate both what it can, and what it cannot, do.

Acknowledgments

Literature Cited

BTO. 2011. Developing bird indicators. ⟨http://www.bto.org/science/monitoring/developing-bird-indicators⟩

Buckland, S. T. 1984. Monte Carlo confidence intervals. Biometrics 40:811–817.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling. Oxford University Press, Oxford, UK.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2004. Advanced distance sampling. Oxford University Press, Oxford, UK.

Buckland, S. T., A. E. Magurran, R. E. Green, and R. M. Fewster. 2005. Monitoring change in biodiversity

through composite indices. Philosophical Transactions of the Royal Society of London B 360:243–254.

Buckland, S. T., A. C. Studeny, A. E. Magurran, and S. E. Newson. 2010. Biodiversity monitoring: the relevance of detectability. Pages 25–36 in A. E. Magurran and B. J. McGill, editors. Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford, UK.

Butchart, S. H. M., et al. 2010. Global biodiversity: indicators of recent declines. Science 328:1164–1168.

CBD. 2011. Strategic plan for biodiversity 2011–2020 and the Aichi Targets. ⟨http://www.cbd.int/doc/strategic-plan/2011-2020/Aichi-Targets-EN.pdf⟩

Coddington, J. A., I. Agnarsson, J. A. Miller, M. Kuntner, and G. Hormiga. 2009. Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. Journal of Animal Ecology 78:573–584.

Creed, R. P., R. P. Cherry, J. R. Pflaum, and C. J. Wood. 2009. Dominant species can produce a negative relationship between species diversity and ecosystem function. Oikos 118:723–732.

Fewster, R. M., S. T. Buckland, G. M. Siriwardena, S. R. Baillie, and J. D. Wilson. 2000. Analysis of population trends for farmland birds using generalized additive models. Ecology 81:1970–1984.

Fisher, J. A., K. T. Frank, and W. C. Leggett. 2010. Breaking Bergmann's rule: truncation of northwest Atlantic marine fish body sizes. Ecology 91:2499–2505.

Fitzpatrick, U., T. E. Murray, R. J. Paxton, J. Breen, D. Cotton, V. Santorum, and M. J. F. Brown. 2007. Rarity and decline in bumblebees—a test of causes and correlates in the Irish fauna. Biological Conservation 136:185–194.

Freeman, S. N., D. G. Noble, S. E. Newson, and S. R. Baillie. 2007. Modelling population changes using data from different surveys: the Common Birds Census and the Breeding Bird Survey. Bird Study 54:61–72.

Gaston, K. J. 2011. Valuing common species. Science 327:154–155.

Gotelli, N., and R. K. Colwell. 2010. Estimating species richness. Pages 39–54 in A. E. Magurran and B. J. McGill, editors. Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford, UK.

Gregory, R. D., and A. van Strien. 2010. Wild bird indicators: using composite population trends of birds as measures of environmental health. Ornithological Science 9:3–22.

Gregory, R. D., P. Vorisek, D. G. Noble, A. van Strien, A. Klvanova, and M. Eaton. A. W. Gmelig Meyling, A. Joys, R. P. B. Foppen, and I. J. Burfield. 2008. The generation and use of bird population indicators in Europe. Bird Conservation International. 18:S223–S244.

Gregory, R. D., A. van Strien, P. Vorisek. A. W. Gmelig Meyling, D. G. Noble, R. P. B. Foppen, and D. W. Gibbons. 2005. Developing indicators for European birds. Philosophical Transactions of the Royal Society of London B. 360:269–288.

Hails, C., S. Humphrey, J. Loh, and S. Goldfinger, editors. 2008. Living planet report. World Wide Fund for Nature, Gland, Switzerland.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman and Hall, London, UK.

Hector, A., and R. Bagchi. 2007. Biodiversity and ecosystem multifunctionality. Nature 448:188–190.

Jain, M., C. Prager, D. Flynn, C. Devan, G. Hart, F. Ahrestani, D. Bunker, M. Palmer, S. Smukler, J. Sircely, and S. Naeem. 2010. Commonness and rarity of species: Does species' rank influence contribution to functional diversity? Nature Proceedings ⟨http://dx.doi.org/10.1038/npre.2010.5243.1⟩

Lamb, E. G., E. Bayne, G. Holloway, J. Schieck, S. Boutin, J. Herbers, and D. L. Haughland. 2009. Indices for monitoring biodiversity change: are some more effective than others? Ecological Indicators 9:432–444.

Limpert, E., W. A. Stahel, and M. Abbt. 2001. Lognormal distributions across the sciences: keys and clues. BioScience 51:341–352.

Loh, J., R. E. Green, T. Ricketts, J. Lamoreux, M. Jenkins, V. Kapos, and J. Randers. 2005. The Living Planet Index: using species population time series to track trends in biodiversity. Philosophical Transactions of the Royal Society of London B 360:289–295.

Mace, G. M., and J. E. M. Baillie. 2007. The 2010 biodiversity indicators: challenges for science and policy. Conservation Biology 21:1406–1413.

Magurran, A. E. 2004. Measuring biological diversity. Blackwell, Oxford, UK.

Magurran, A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. J. Somerfield, and A. Watt. 2010. Long-term data sets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology and Evolution 25:574–582.

Newson, S. E., K. L. Evans, D. G. Noble, J. J. D. Greenwood, and K. J. Gaston. 2008. Use of distance sampling to improve estimates of national population sizes for common and widespread breeding birds in the UK. Journal of Applied Ecology 45:1330–1338.

Newson, S. E., R. J. W. Woodburn, D. G. Noble, S. R. Baillie, and R. D. Gregory. 2005. Evaluating the Breeding Bird Survey for producing national population size and density estimates. Bird Study 52:42–54.

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. Trends in Ecology and Evolution

21:668–673.

O'Brien, T. G., J. E. M. Baillie, L. Krueger, and M. Cuke. 2010. The Wildlife Picture Index: monitoring top trophic levels. Animal Conservation 13:335–343.

Pereira, H. M., and H. D. Cooper. 2006. Towards the global monitoring of biodiversity change. Trends in Ecology and Evolution 21:123–129.

Shannon, C. E., and W. Weaver. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois, USA.

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

Smith, B., and J. B. Wilson. 1996. A consumer's guide to evenness indices. Oikos 76:70–82.

Studeny, A. C., S. T. Buckland, J. B. Illian, A. Johnston, and A. E. Magurran. 2011. Goodness-of-fit measures of evenness: a new tool for exploring changes in community structure. Ecosphere 2:art15.

Theobald, D. M., D. L. Stevens, D. White, N. S. Urquhart, A. R. Olsen, and J. B. Norman. 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. Environmental Management 40:134–146.

Tilman, D., P. B. Reich, and J. M. H. Knops. 2006. Biodiversity and ecosystem stability in a decade-long grassland experiment. Nature 441:629–632.

Wood, S. N. 2006. Generalized additive models—an introduction with R. CRC, Boca Raton, Florida, USA.

Zar, J. H. 1999. Biostatistical analysis. Fourth edition. Prentice Hall, Upper Saddle River, New Jersey, USA.
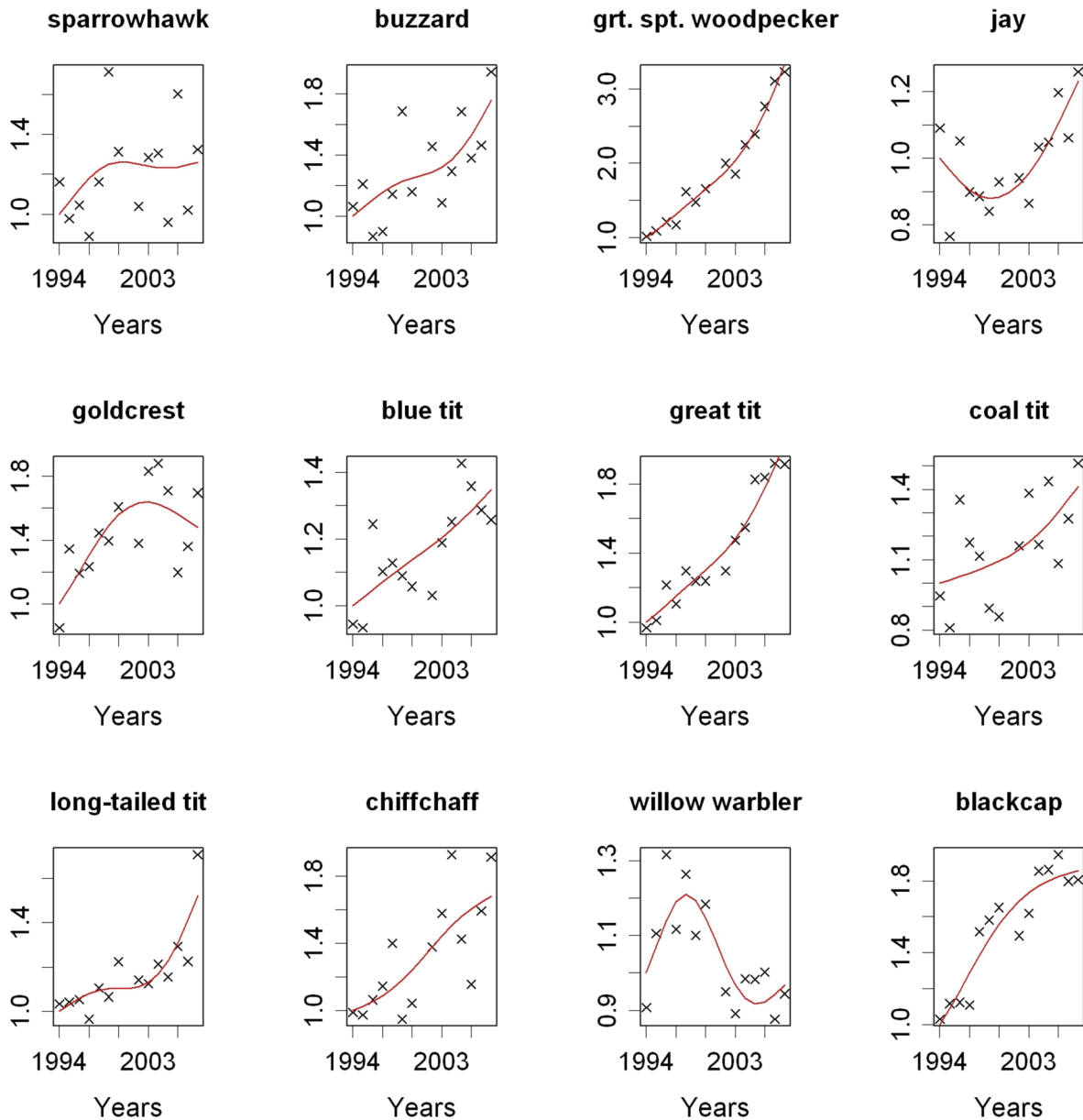
Appendix

Fig. A1. Relative abundance estimates for sparrowhawk, buzzard, great spotted woodpecker, jay, goldcrest, blue tit, great tit, coal tit, long-tailed tit, chiffchaff, willow warbler and blackcap. Smoothed estimates are shown by lines, while the points represent unsmoothed estimates.
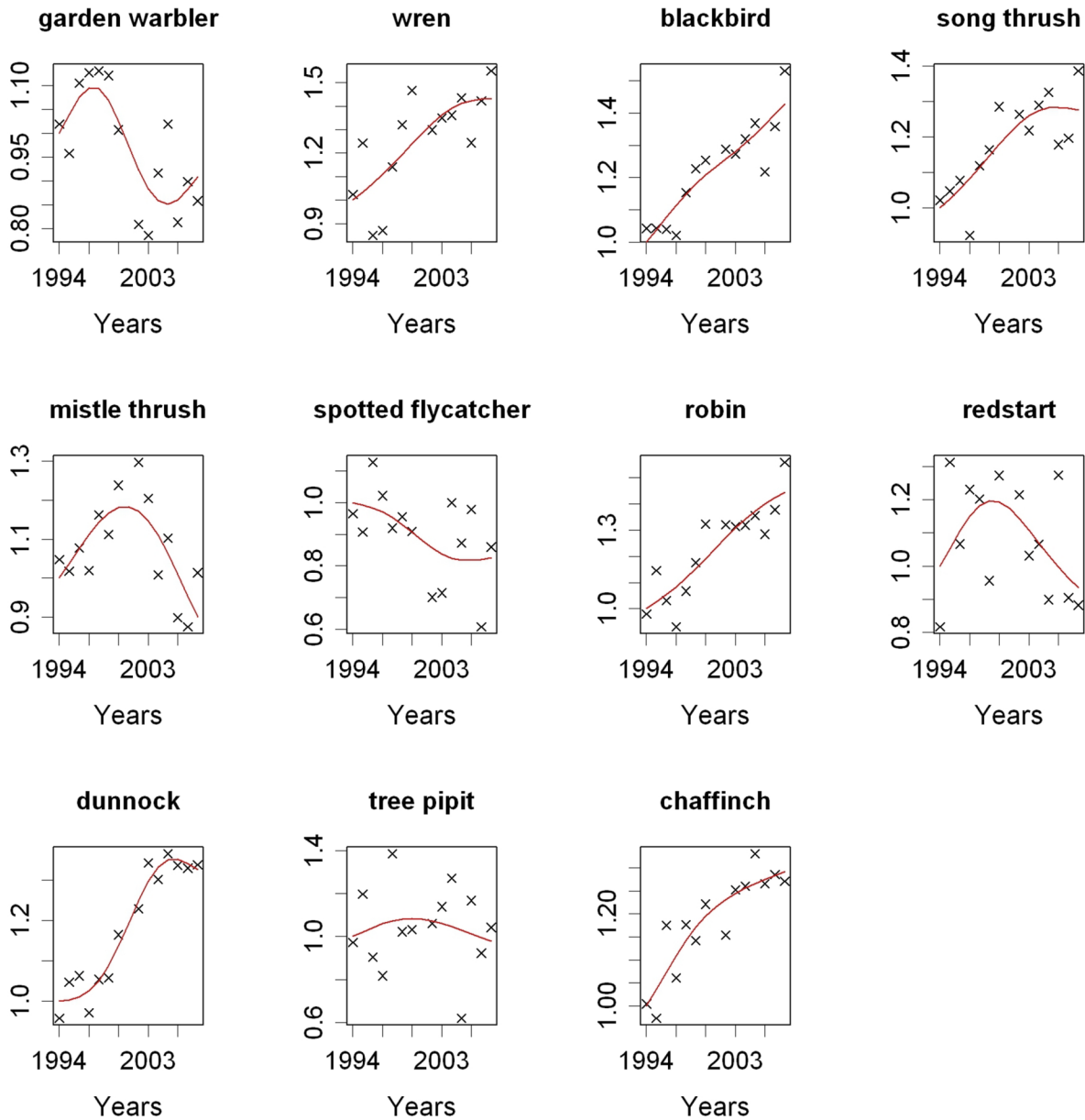
Fig. A2. Relative abundance estimates for garden warbler, wren, blackbird, song thrush, mistle thrush, spotted flycatcher, robin, redstart, dunnock, tree pipit and chaffinch. Smoothed estimates are shown by lines, while the points represent unsmoothed estimates.
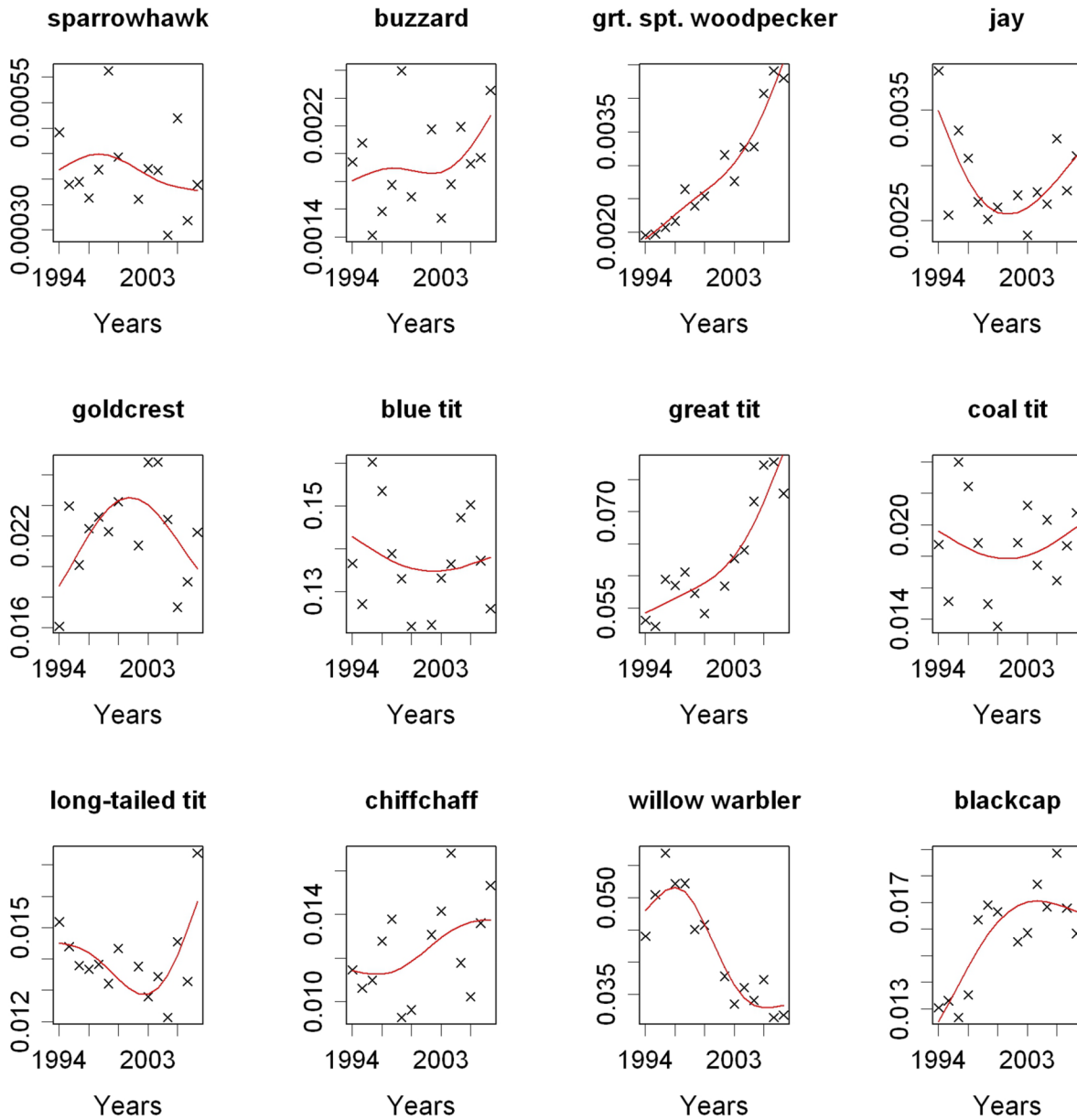
Fig. A3. Species proportions for sparrowhawk, buzzard, great spotted woodpecker, jay, goldcrest, blue tit, great tit, coal tit, long-tailed tit, chiffchaff, willow warbler and blackcap. Smoothed estimates are shown by lines, while the points represent unsmoothed estimates.
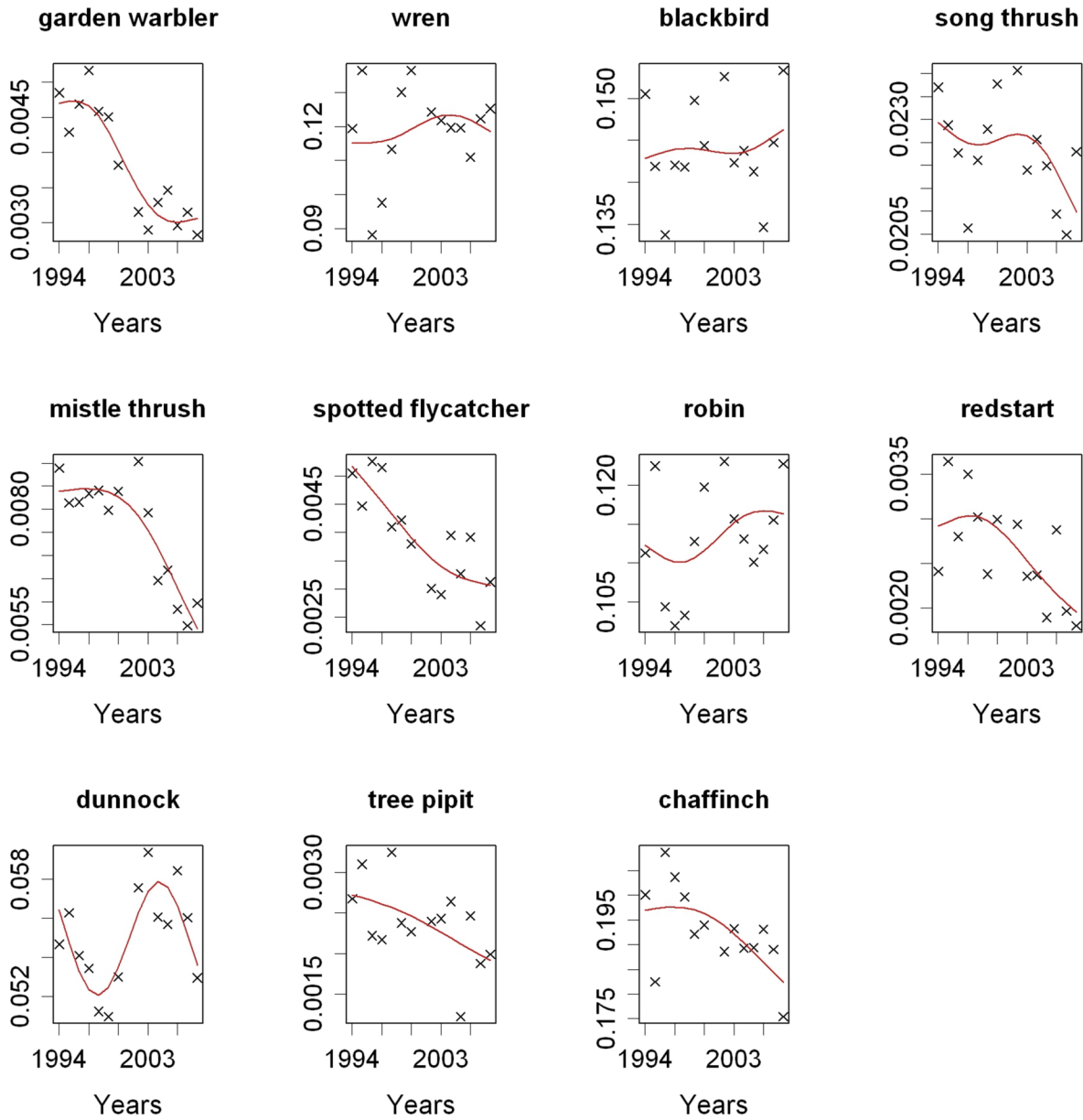
Fig. A4. Species proportions for garden warbler, wren, blackbird, song thrush, mistle thrush, spotted flycatcher, robin, redstart, dunnock, tree pipit and chaffinch. Smoothed estimates are shown by lines, while the points represent unsmoothed estimates.