Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

A Monte Carlo method to account for sampling error in multi-species indicators

Leo L. Soldaat[a,*], Jeroen Pannekoek[a], Richard J.T. Verweij[a], Chris A.M. van Turnhout[b], Arco J. van Strien[a]

[a] Statistics Netherlands, P.O. Box 24500, 2490 HA, The Hague, The Netherlands
[b] Sovon Dutch Centre for Field Ornithology, P.O. Box 6521, 6503 GA, Nijmegen, The Netherlands

ARTICLE INFO

ABSTRACT

The usefulness of biodiversity indicators strongly increases if accompanied by measures of uncertainty. In the case of indicators that combine population indices of species, however, the inclusion of the uncertainty of the species indices has shown to be hard to realize, usually due to imperfections in monitoring programmes. Missing values and time series of different lengths preclude the use of analytical approaches, whereas bootstrapping across sites requires the raw abundance data on the site level, which may not always be available. Sometimes bootstrapping across species rather than sites is opted for, but this approach ignores the uncertainty attached to species indices. We developed a method to account for sampling error of species indices in the calculation of multi-species indicators based on Monte Carlo simulation of annual species indices. The construction of confidence intervals enables various trend assessments, like testing for linear or smooth trends, testing for changes between two time points, testing the significance of a suspected change-point and testing for differences between two multi-species indicators. Here, we compare our method with conventional methods and illustrate the benefits of our approach using Dutch breeding bird indicators.

1. Introduction

In order to realize the international ambition to slow and eventually halt the ongoing global decline in biodiversity, as expressed in the context of the Convention on Biological Diversity (Butchart et al., 2010; Secretariat of the Convention of Biological Diversity, 2014), it is indispensable to have reliable instruments to measure progress towards set targets. Biodiversity indicators are increasingly used to monitor trends in biodiversity at various habitats and scales (Biala et al., 2012; Butchart et al., 2010; Szabo et al., 2012; Van Strien et al., 2016), the most popular being the combined population trends of individual species (Brereton et al., 2011; Freeman et al., 2001; Gregory et al., 2005; Loh et al., 2005). Such multi-species indicators (MSI) have the advantage of being relatively insensitive to the fluctuations of individual species, thus helping scientists, conservationists and decision makers to better understand the dominant factors influencing biodiversity in a region, country, continent or the entire biosphere. Until now the development of MSIs has mainly focused on methods to calculate the mean index of species, of which the geometric mean of species indices appears one of the most appropriate to use (Buckland et al., 2005, 2011;

Lamb et al., 2009; Van Strien et al., 2012). Popular examples of MSIs include the global Living Planet Index (Collen et al., 2009; Loh et al., 2005), the European Grassland Butterfly Indicator (Van Swaay et al., 2013), and the European Wild Bird Indicators (Gregory et al., 2005; Gregory and Van Strien, 2010).

The usefulness of MSIs and trends in MSIs is strongly increased if accompanied by proper measures of uncertainty. Without these, it becomes problematic to test whether changes in the indicator are statistically significant and/or to test the found trend against other indicators. The main sources of uncertainty in MSIs are sampling error and process noise. Sampling error refers to the uncertainty of the species indices, which in most monitoring programmes must be considered as "sampling error in a broad sense": the "pure sampling error" caused by sampling only part of the population, complemented by sources of variation like measurement bias, imperfect detection and missing values. This part of the variation in time series is also called "observation error" (e.g. Dennis et al., 2006). Process noise refers to the interannual variation between indices, the "process" being the trend in population numbers which usually is the main objective of a monitoring programme. Surprisingly, although the sources of uncertainty of MSIs are

theoretically well-known it often proves a challenge to construct confidence intervals (CIs) for both MSIs and trends therein that take into account both sampling error and process noise. We know of three common methods, none of which is completely satisfying:

(1) CI based on bootstrapping across species

In this approach (for instance Collen et al., 2009; Craigie et al., 2010; Eaton et al., 2016) the trend of each species is considered as a replicate of the MSI. This approach is useful to assess the robustness of the MSI against species selection, but it neglects sampling error in the species indices. In addition, it suffers from a conceptual drawback: it is questionable to include interspecific variation in the confidence intervals of MSIs. The rationale of testing against variation between species is that the species are randomly sampled from a large group, but this rationale is unjustified as species to represent an MSI are typically deliberately selected. In addition, bootstrapping species may yield wide confidence intervals if the trend of even a single species deviates from the trend of the other selected species for the MSI. Consequently, even evident shifts in the mean of the MSI may remain statistically insignificant.

(2) CI based on interannual variation

This approach is used for the European Wild Bird Indicators and the Living Planet Index (Butchart et al., 2010; Gregory and Van Strien, 2010; Loh et al., 2005), amongst others. Again, in these indicators sampling error is neglected and confidence intervals for trends in MSIs only include the interannual variation. For the European Wild Bird Indicators (Gregory et al., 2005) an analytical approach is presented to calculate CIs for the MSI that takes into account sampling error. However, this approach cannot be extended to trend assessments and it fails whenever a species index is missing for a particular year. Thus, as is the case for other indicators, sampling error is neglected in the trend assessment for European Wild Bird Indicators, even when available. The latter is inevitable, as the TrendSpotter software used for trend calculation cannot include standard errors of yearly MSIs (Soldaat et al., 2007; Visser, 2004). TrendSpotter can efficiently model flexible trends and their CIs by applying the Kalman filter. Unfortunately, only relative weighting factors can be attached to the MSIs. Absolute weighting factors like the standard errors of the MSI would not lead to proper CIs for the calculated trends.

(3) CI based on bootstrapping of sites

This approach properly takes into account sampling error and can be applied in a randomized monitoring scheme like the British Farmland Bird Indicator (Freeman et al., 2001). Bootstrapping on the site level, however, cannot be applied if sites are not a random sample of the population, as in many volunteer-based monitoring programmes. Obviously, bootstrapping of sites can also not be applied when data are not available on the site level, for example when MSIs are constructed using time series obtained from the literature (as in the Living Planet Index) or from national reports (as in the European Wild Bird Indicators).

An approach to take into account sampling error in MSIs that, to our knowledge, has not been explored so far is the use of standard errors of the species indices. In this paper we describe Monte Carlo procedures to generate confidence intervals for MSIs and trends in MSIs based on the standard errors of species indices. The method overcomes the above-mentioned conceptual and practical obstacles, and offers several opportunities for testing and comparing trends in MSIs. Here, we first use conventional approaches to calculate an MSI with confidence intervals from an ideal simulated data set, without missing values. Subsequently, we apply our method to the same simulated data, and compare the outcome to the results of the conventional approaches for validation.

Thereafter we illustrate our method using Dutch breeding bird data. Finally, we show how the method can be used to test for change-points in the MSI and trend differences between MSIs and some additional possibilities for trend assessment.

2. Methods

2.1. Calculating MSIs and confidence intervals by Monte Carlo simulation

The starting point of the Monte Carlo (MC) method is a data set with species indices and standard errors, for instance calculated with the TRIM software (Pannekoek and Van Strien, 2005). The index value in some pre-defined base year is set to 100 with standard error zero (step 1 in Fig. 1). The indices in the other years are expressed as percentage of the base year and their standard errors are a function of the variance in the specific year and the base year. Our method assumes that the standard errors are adjusted for the effect of serial correlation between years, which is the standard approach in most monitoring programmes. The indices are approximately log-normal distributed (Pannekoek and Van Strien, 2005) and the standard errors are used for MC simulation. Each available yearly index for each species is simulated 1000 times by drawing from a normal distribution $N(\mu, \sigma)$ with μ = the natural logarithm of the index and σ = the standard error of the index on the log scale (step 2). The standard error of the index on the log scale is assessed by the Delta-method (see e.g. Agresti, 1990) as SE(log scale) = SE(index scale)/index. After simulation the same base year (index = 100) is chosen in each simulation for each species and the other years are expressed as a percentage of the base year (step 3 and 4; for the imputation of missing indices, see below). The mean and standard error of the 1000 MSIs in each year are calculated and back-transformed to the index scale (step 5). The arbitrarily chosen number of 1000 simulations is a trade-off between computational efficiency and accuracy, to insure consistency in estimates across runs. This number could be increased if large variability in the results is observed between different runs.

2.1.1. Missing data

A complication in the procedure described in Section 2.1 arises if some species have missing indices. In practice these missing values will often occur at the beginning or end of the time series (due to differences in monitoring schemes between species). These missing indices must be imputed in order to set the same base year for each species, which is necessary to calculate geometric mean indices. We apply chain indexing (Crawford, 1991) to impute missing species indices, using the relative year-to-year population development in species without missing values (step 3 in Fig. 1). Thus, if all species without missing data show a mean increase of, say, 10% from year t to t + 1, this percentage is used for imputing the missing data points in species with missing data for year t + 1. Note that in each MC simulation different imputed values will be generated. After imputation we proceed with the common procedure to calculate MSIs and standard errors. These standard errors for years with missing species indices do not include the uncertainty caused by imputation.

2.1.2. Handling extreme cases

Using the geometric mean in biodiversity indicators has many advantages (Buckland et al., 2005, 2011; Van Strien et al., 2012), but the downside is that it makes such indicators oversensitive to strongly fluctuating, eruptive, strongly increasing or strongly decreasing species. For such species, index values may show extreme yearly changes or may become extremely large, zero, or close to zero. Especially small indices may have strong and unwanted (as they usually represent very low population numbers) effects on the MSI.

Zero values of indices need special treatment anyhow, as the logarithm of zero is undefined. Often an arbitrary small amount (e.g. 0.1 or 1) is added to zero indices before log transformation. The effect on the
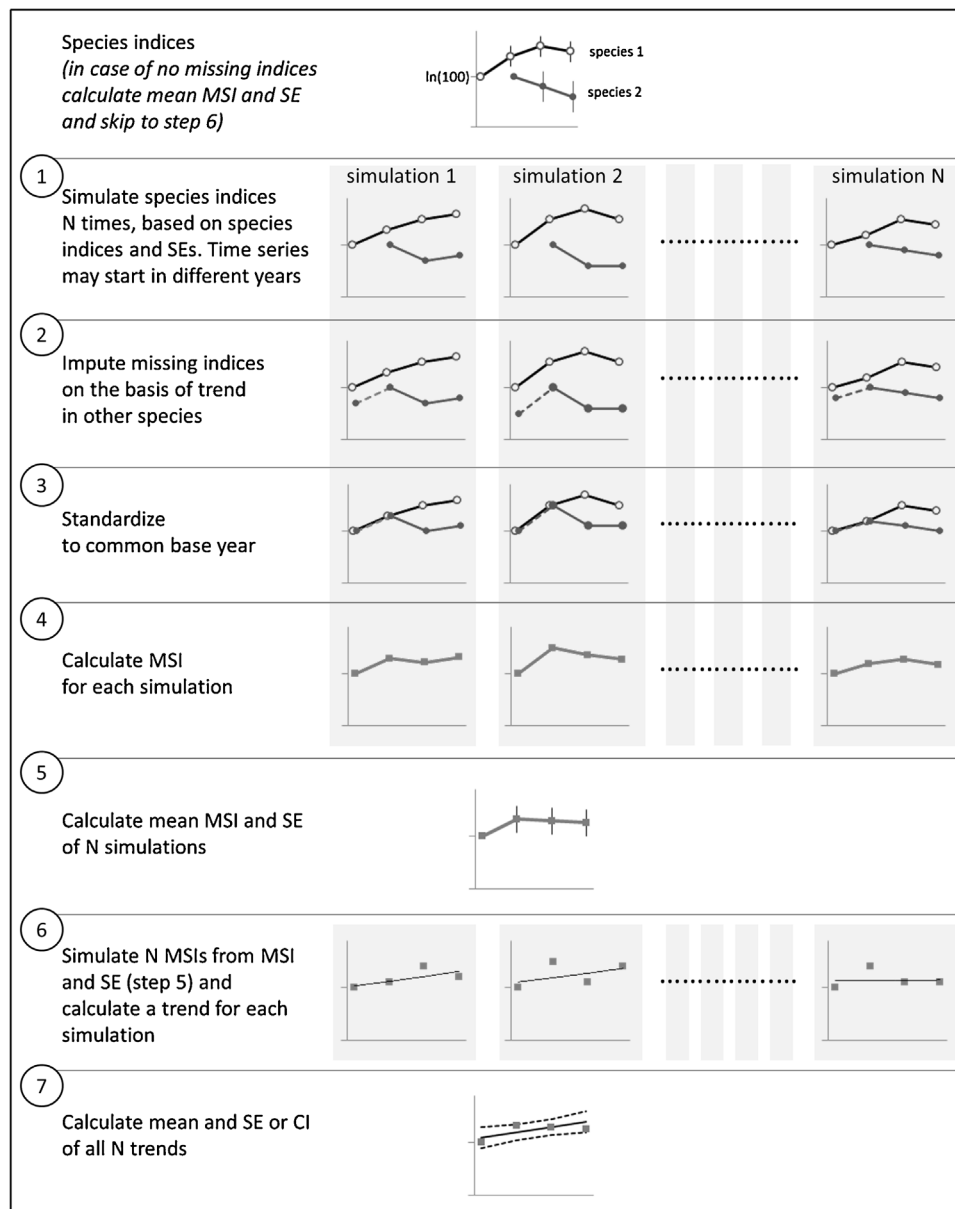
Fig. 1. Schematic representation of the Monte Carlo method to calculate standard errors (SE) and confidence intervals (CI) for multi-species indicators (MSI). The numbers refer to the consecutive steps in the Monte Carlo method (see text).

index and even more on the MSI, however, strongly depends on the chosen amount to add to zero. In the Living Planet Index, zero indices are replaced by 1% of the mean of all indices for a species (Collen et al., 2009; Loh et al., 2005). In addition, yearly changes in indices of a factor > 10 are set to missing values (McRae, pers. comm., 2015). The choice how to treat "noisy" indices is up to the researcher and may depend on the goals of the study. We apply the following rules:

- For strongly increasing (e.g. invasive) species, the last year is set to 100. All indices in the preceding years lower than 1 are truncated to 1 and their standard error is set to 0.
- For strongly decreasing species, the first year is set to 100. All indices in the subsequent years lower than 1 are truncated to 1 and their standard error is set to 0.
- Yearly changes of a factor > 10 are truncated to 10 (in case of an increase) or 0.1 (decrease). This rule dampens the extreme yearly variation in the MSI caused by strongly fluctuating (e.g. eruptive) species.

2.2. Trend classification and trend comparison by Monte Carlo simulation

The MSI pattern over time may show noisy behaviour. To find trends and uncertainty therein we also apply a MC simulation procedure. 1000 MSIs are simulated based on a normal distribution with μ = the natural logarithm of the MSI and σ = its standard error (as derived in Section 2.1) on the log scale (step 5 in Fig. 1). For each of the 1000 simulated MSIs two trend methods are applied: a linear OLS straight line and a smooth trend estimated by LOESS-regression (Cleveland, 1993) with default settings for span (0.75) and degree (2) (step 6). Standard calculus provides mean trends, standard errors and confidence limits of the trends, each of which are back-transformed to the index scale (step 7). The 1000 simulated MSIs offer the opportunity to compare trends and to test several additional trend parameters (see Section 3.3). We note that MSIs have value 100 in the base year chosen, with zero uncertainty (as a consequence of our indexing method). However, the trend models chosen here are not 'forced' to have a value 100 in the base year. Therefore, uncertainties in mean trends are not zero in base years.

Table 1
Trend classification. CI = confidence interval, CL = confidence limit.

Category	Criteria/description
Strong increase	lower CL > 1.05 *(significant increase of more than 5% per year)*
Moderate increase	1.00 < lower CL < 1.05 *(significant increase, but not significantly more than 5% per year)*
Stable	CI includes 1.00 AND 0.95 ≤ lower CL AND upper CL ≤ 1.05 *(no significant increase or decline, likely that changes are smaller than 5% per year)*
Uncertain	lower CL < 0.95 AND 1.05 < upper CL *(no significant increase or decline, unlikely that changes are smaller than 5% per year)*
Moderate decline	0.95 < upper CL < 1.00 *(significant decline, but not significantly more than 5% per year)*
Steep decline	upper CL < 0.95 *(significant decline of more than 5% per year)*

2.2.1. Trend classification

The 1000 linear or smooth trends were used to classify trends in categories in several ways. First, linear trends and standard errors were back-transformed from the additive log scale by applying the inverse of the Delta-method described in Section 2.1. On the resulting multiplicative scale a trend of 1.00 means no change, a trend of 1.05 means 5% increase per year and 0.95 means 5% decrease per year. A 95% confidence interval for the trends was calculated based on the normal distribution. The multiplicative trends were used to classify trends as in Table 1. Separate linear trends were calculated before and after a suspected change-point in each of the 1000 simulations as well as the difference between both trends. The mean and standard error of these 1000 differences were used to construct a 95% confidence interval. The trends are significantly different when this interval does not include 0. The smooth trends are primarily meant for presentation purposes, but can also be used to assess the difference between two years. For instance, the first and last year may be compared by calculating the 1000 differences in the smooth trend value between these years and the difference can then be tested statistically on the basis of the mean and confidence interval as described above. This yields an overall change e.g. "significant 40% increase from 1990 to 2014".

2.3. Validation on simulated data

In order to validate the Monte Carlo approach we applied the method to a computer-generated data set consisting of Poisson counts of three species during 20 years in 100 random sites and no missing data. In this ideal data set the randomness of the sites allowed for the application of a bootstrapping approach (cf. Freeman et al., 2001), whereas the non-missing data allowed the application of linear approximation (cf. Gregory et al., 2005). Our MC approach should yield approximately the same indices and confidence intervals as the two conventional methods. Linear approximation was applied to species indices and standard errors that were calculated with TRIM (Pannekoek and Van Strien, 2005).

2.4. Illustration on data of Dutch breeding birds

We illustrate our Monte Carlo approach using abundance data of breeding birds collected by skilled volunteers through the Breeding Bird Monitoring Programme (BMP) in the Netherlands (Van Turnhout et al., 2008). The Dutch BMP is based on territory mapping in fixed study sites. Bias due to imperfect detection is reduced in the BMP by prescribing that repeated mappings are performed at periods of high detection (both during the season as well as during the day). By clustering the repeated mappings (Bibby et al., 2000) the resulting number of territories for a species at a site comes as close as possible to the true population size. All common and scarce breeding birds in the Netherlands are covered. The number of study plots grew from about 1000 per year in 1990 to about 2000 at the end of the study period. The Dutch BMP has been used to assess trends in bird fauna (Le Viol et al., 2012; Van Turnhout et al., 2010), and the data also contribute to pan-European biodiversity indicators (Gregory et al., 2005). Population development of a species is assessed by calculating yearly indices and standard errors using the TRIM software (Pannekoek and Van Strien, 2005). We compiled indicators for woodlands (26 species), marshlands (including fresh water; 29 species) and open natural habitats (coastal dunes and saltmarshes, heathlands; 22 species). Typical breeding bird species for these habitats were selected using the Species Specialization Index (SSI), which uses the variance of average densities among twelve habitat classes (Julliard et al., 2006), derived from BMP data around 1990 (start year of trends). Species with an SSI of > 1.25 are considered as specialists for that particular habitat. Specialists depending largely on one particular habitat are considered to be most sensitive for changes in quality or quantity in that habitat, and therefore are the most suitable indicator species (Julliard et al., 2006). Rare species, for which no reliable densities could be calculated, were appointed on the basis of expert judgement. The species indices that are used as input for the indicator are based on study sites in that particular habitat only (most specialists occur in low densities in other habitats too).

3. Results

3.1. Method validation

Fig. 2 shows the MSIs of the computer-generated annual indices of the three fictitious species and their CIs derived by the Monte Carlo approach and by two alternative procedures: the analytical approach as advocated by Gregory et al. (2005) (Fig. 2a) and bootstrapping of sites (Fig. 2b). As expected, the three approaches yield practically the same MSIs and CIs. The CIs of the bootstrapping method are slightly smaller than for the Monte Carlo method.

3.2. Illustration on Dutch breeding bird data

Fig. 3 shows the standard output of the MC approach for typical breeding bird species in three habitat types. The width of the confidence intervals of the trends clearly varies between habitats and closely corresponds to the mean size of the standard errors of the MSIs. This is a direct consequence of the MC approach that takes into account the standard errors of the species indices when estimating standard errors for the MSIs (step 1–5 in Fig. 1). Linear trends and standard errors were assessed as described in step 6 and 7 in Fig. 1. Based on the standard errors the overall trends and trends for the last 10 years were classified as indicated in Fig. 3. The overall trends corresponded to the number of increasing, stable and decreasing species in the habitat types: the increasing trend in marshlands is caused by 17 increasing, 3 stable and 9 decreasing species; the stable trend in woodlands is caused by 11 increasing, 5 stable and 10 decreasing species; the decreasing trend in open habitats is caused by 5 increasing, 5 stable and 12 decreasing species.

Not surprisingly, the long-term linear trends for the MSIs in the three habitat types differ significantly. Comparing the differences between the 1000 long-term trends in each habitat showed that the yearly change in marshlands was 3.15% larger than in woodlands (p < 0.001) and 7.64% larger than in open habitats (p < 0.001) and the mean yearly change in woodlands was 4.50% larger than in open habitats (p < 0.001). Over the last 10 years these differences were less
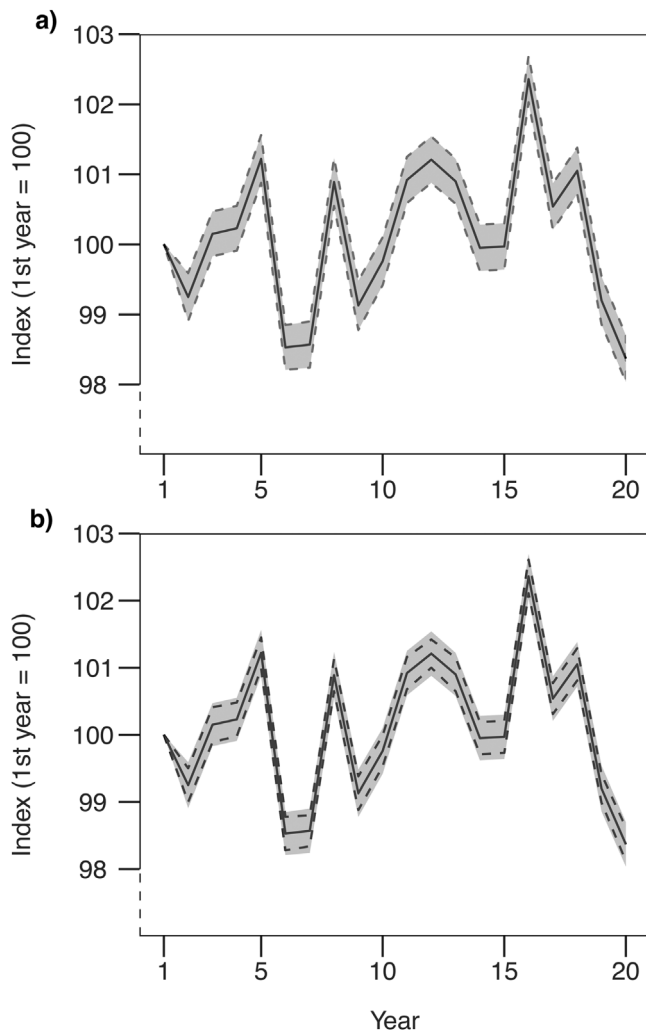
Fig. 2. Comparison of the Monte Carlo approach with (a) linear approximation and (b) bootstrapping of sites in a computer-generated dataset with indices of three species over 20 years. The mean of the simulated annual indices (black line) is the same for each method. The confidence intervals derived by the Monte Carlo approach is indicated by the grey area. Confidence intervals derived by linear approximation and bootstrapping of sites are indicated by dotted lines.
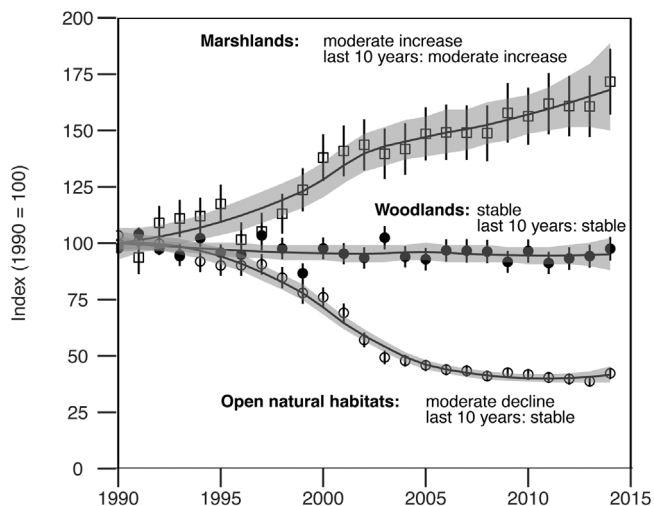

Fig. 3. Multi-species indices (point symbols) and standard errors (error bars) for typical breeding birds of three habitats as calculated in step 1–5 in Fig. 1. The black line and shaded area represent the smoothed trend and confidence interval as calculated in step 6 and 7 in Fig. 1. The trend classifications are based on linear trends (see text).
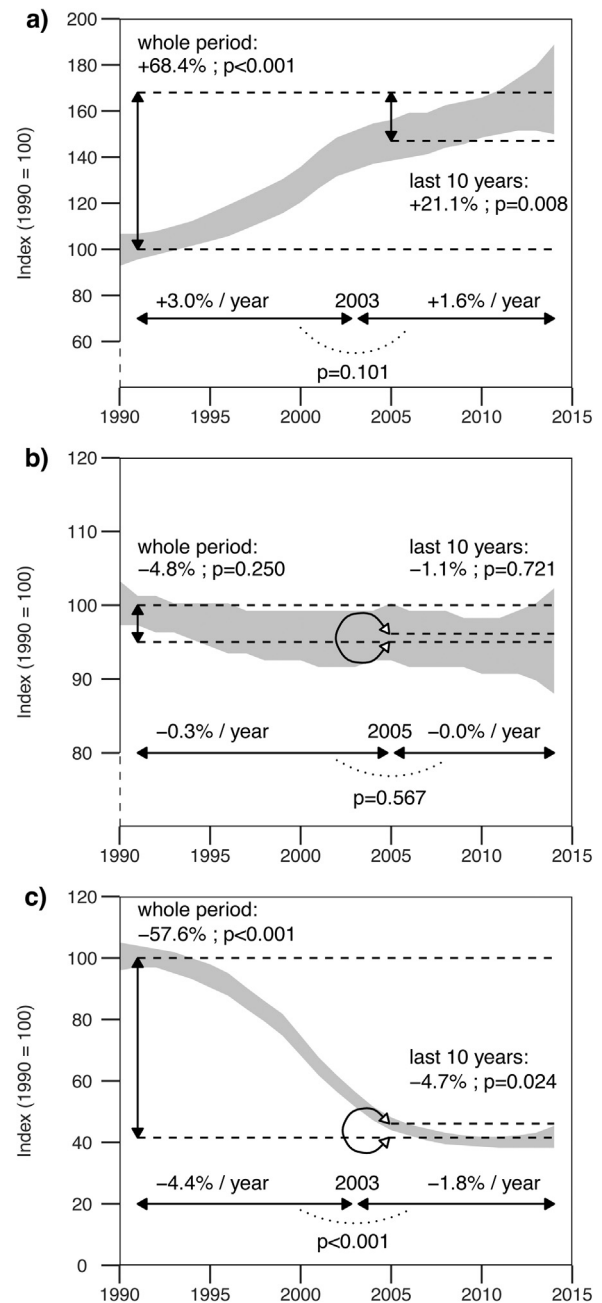

Fig. 4. Decomposition of the smooth trends of typical birds of a) marshlands, b) woodlands and c) open natural habitats.

pronounced, as may be expected from Fig. 3: comparing the 1000 short-term trends in each habitat showed that the differences between marshlands and woodlands (1.9%, p = 0.091) and between woodlands and open habitats (0.9%, p = 0.259) were not significant. Only the difference in the short-term trends between marshlands and open habitats was statistically significant (2.7%, p = 0.022).

3.3. Additional trend assessments

Additional trend assessments showed that the percentage change between the first and the last year as well as the percentage change over the last ten years of the time series were significant in marshlands and open natural habitats, but not in woodlands (Fig. 4). Finally, trend changes within the time series were assessed by inspection of several suspected change-points. In marshlands and open natural habitats a

significant change-point was found in 2003. In woodlands no significant change-point could be detected (only 2005 is shown in Fig. 4, but several other suspected change-points were also tested).

4. Discussion

4.1. Validation and applications

The Monte Carlo method we describe is a straight-forward and conceptually sound method to estimate confidence intervals around multi-species indicators. The method produces almost exactly the same results as both the analytical approach and bootstrapping, as we demonstrated for the simulated data set. Small differences are caused by the stochastic nature of both the MC approach and bootstrapping. The validation test could only be performed in an ideal simulated data set with random sites and without missing species indices, because of the prerequisites of the conventional methods. The MC approach, however, can be applied in many, less restricted situations where species indices with log-normal error distribution are to be combined. An additional advantage is the multitude of ways in which the development of the MSIs in time can be assessed and tested on the basis of the data sets generated by the MC method. These properties make our approach a valuable tool for conservationists and decision makers in evaluating the progress made in nature policy.

Although the main purpose of our method is to calculate confidence limits around trends in multi-species indicators, one can also apply the smoothing procedure to year indices and standard errors of single species. Of course, other flexible trend methods can be applied on the species level as well, like smoothing splines, generalized additive models or the Kalman filter (as applied in the TrendSpotter software mentioned above). However, these methods cannot include the uncertainty of the species indices.

4.2. Demands on sampling error and type of inference

We illustrated the MC method on indices and standard errors that were calculated by TRIM. TRIM takes into account 'pure sampling error' caused by selection of only a part of all possible sites, but also complementary variation caused by differing population development between sites and by the imputation of missing values. TRIM uses an efficient implementation of glm-Poisson regression (Pannekoek and van Strien, 2005) to account for missing data. This model-based approach assumes that the yearly observed counts for each of the species in the BMP are random variables due to natural variation (and possibly measurement error) around the model expected value. Specifically, in Poisson regression the counts are assumed to follow a Poisson distribution which entails that the variance of the counts equals their expected value. In the TRIM-implementation of Poisson regression this assumption is relaxed in the sense that the variance is taken to be *proportional* to the expected value, with a proportionality factor that is estimated from the data (overdispersion relative to the Poisson assumption). The indices are calculated by TRIM using the observed counts as input and the standard errors of the indices are thus determined by the variance of these observed counts (expected value times overdispersion factor). This model-based type of inference is different from a design-based inference in which the counts are considered fixed quantities rather than random variables, but random variation is introduced by considering the sites as a random sample from some population. Estimates of indices are then random variables because they are based on a random selection of sites and these estimates will differ between samples. Standard error estimation will in this case be based on the observed variation between sites rather than on the assumption (model-based) of random variation between counts according to some distribution. Standard errors of species indices may differ between both types of inference. However, as the MC approach presented in this paper starts with the standard errors of indices it can be applied both in a setting of model-based as well as design-based inference.

4.3. Serial correlation

As our Monte Carlo approach assumes that the standard errors of the species indices already incorporate the effect of serial correlation between years, the simulations do not incorporate this serial correlation. In other words, the simulation of a species index or an MSI in a specific year is not *directly* dependent on the simulated values in the previous year. In the examples in this study serial correlation is *indirectly* accounted for. As the standard errors of the species indices calculated by TRIM (Pannekoek and Van Strien, 2005) incorporate the effect of serial correlation, this effect propagates into the standard errors of the MSIs. Likewise, by simulating MSIs for trend assessment the effect of serial correlation propagates into the standard errors of the trend estimates. However, the MSIs themselves are possibly also serially correlated and this unknown serial correlation cannot be accounted for in the simulation. The standard errors of the trends should thus be considered as best approximations with the available data, that neglect possible serial correlation between MSIs. Incorporating serial correlation between MSIs in the MC simulations would be a further refinement of the method, but is beyond the scope of this study. In addition, it will be technically complex and based on our experience with the effect of serial correlation on the standard errors of species trends, we expect the effects to be relatively small.

4.4. Features and possible complications

As for the species indices, for the MSI a base year with index 100 and the standard error fixed to zero must be chosen. The base year for the MSI is not necessarily the first year. In case of less reliable count data in the beginning of the time series (e.g. due to more missing values), one may choose to select the last year as base year, which will cause the CIs of the MSIs and of smooth trends to be wider at the beginning of the time series. In other cases some intermediate base year may be of interest. In general, the confidence intervals around the base year are narrower, and widen as temporal autocorrelation decreases (e.g. Buckland et al., 2011; Studeny et al., 2013). Any choice of a base year is computationally equally valid and can be made with the output of the MC simulations. Choosing a different base year only changes the visual display of the confidence limits of the MSIs and smooth trends, but it does not influence the linear trend calculation and classification.

The choices we made to calculate the MSI after simulating species indices may have profound effects on the resulting MSI. First, truncating indices < 1 and setting their standard errors to 0 may strongly reduce CIs of MSIs and trends. The methodological rationale behind this choice is that small indices often have relatively large standard errors that in some situations may cause large stochastic variation in the simulated indices and MSIs. From an ecological point of view it is not desirable that these small indices have large influence on the MSI, as they represent relatively low population sizes. An additional advantage of the truncation rule is that we do not have to choose a small (but usually influential) value to replace zero-indices in order to be able to calculate a logarithm. A second choice that may affect the outcome of MSI-calculation is the chaining method used to impute missing species indices. Ter Braak et al. (1994) point out some risks of applying chain indexing to handle missing values. However, they refer to situations where chain indexing is used to handle missing site counts and not to species indices as in MSIs. The missing site counts in the BMP, as in most monitoring programmes, are imputed by the loglinear regression model and the increased uncertainty of the indices is accounted for in the SEs. Nevertheless, chain indexing may cause bias in MSIs, especially when the MSI consists of few species with large year-to-year fluctuations, analogous to "small samples" and "scarce species" mentioned by Ter Braak et al. (1994). So we recommend to always check the species

indices visually before inclusion in the MSI. A special case where chain indexing will always be inappropriate is when a species is not monitored every year, but e.g. every two or three years. Then, the chaining method cannot calculate any year-to-year ratio for that species and the species will not contribute to the MSI. In most cases, however, missing species indices will be concentrated in the beginning of the time series, due to differences between monitoring schemes. One should be aware that the MSI for these years only represents the development of the species for which data are available. A more realistic but often laborious alternative to chain indexing is to model the missing indices on the basis of data from the species itself. Whatever method is opted for, we advise to compare MSIs with and without the species with incomplete time series. If necessary, the MSI should be recalculated for a period of years with less missing species indices. Whatever the outcome, we advise to report the pattern of missing species indices and discuss its effect on the MSI. In the examples in this paper no missing species indices were encountered. A third influential choice concerns the species selected for the indicator. Some species may have relatively large standard errors, reflecting the inability of the monitoring programme to follow the population development accurately. In our indicators we judge species indices by the mean coefficient of variation (CV) of the yearly indices (the mean of all standard error/index ratios). The exclusion of species based on statistical arguments comes with the danger of ignoring ecologically or otherwise interesting species – usually the rare ones. We recommend to always test the effect on the MSI of stricter versus more liberal levels at which maximum allowed CV is set.

If base years differ between species the standard error of the MSI in a specific year becomes a function of the sampling errors of the species indices in that year and in the different base years. So, unlike the standard error of a species index the standard error of the MSI can no longer be interpreted as the variation between that year and the base year of the MSI. A simple way to avoid this would be to choose the same base year for each species. However, as mentioned in the methods section this may lead to extremely high indices in case of species with low abundances in the base year with undesired effects on the MSI. We therefore recommend to standardize the base year between species whenever possible, and otherwise to clearly specify the differences in base years.

5. Conclusion

Our Monte Carlo simulation approach is a straightforward, easy to apply and conceptually sound method to take into account sampling error in multi-species indicators. Unlike bootstrapping of sites, it does not require the raw abundances per species for each of the surveyed plots: it can be applied whenever standard errors of the year indices of individual species are available. As a consequence, different approaches for index calculation between species are allowed. Contrary to analytical approaches to assess the standard error of the MSI, the Monte Carlo method still works in the case that not all yearly index numbers of all species are available. The method not only works in case the annual indices represent relative species abundance; it also works if the indices relate to trends in site occupancy, biomass, etc. Given the advantages of the Monte Carlo method over other methods, we recommend its use for the calculation of confidence intervals and trends of multi-species indicators whenever possible. In line with this recommendation monitoring programmes should always attempt to collect and report necessary information on the standard errors of species indices.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind.2017.05.033.

An R-script with the MC method is freely available at https://www.cbs.nl/en-gb/society/nature-and-environment/indices-and-trends–trim-/msi-tool. The script enables the calculation of MSIs and various trend assessments based on a simple input database with species indices and standard errors. Several parameters for the analysis can be set, such as the number of MC simulations, the desired base year, the level for omitting species with high CV, the year of a suspected change-point, the period for the calculation of short-term trends and the level for truncation of low indices.

References

Agresti, A., 1990. Categorical Data Analysis. John Wiley, New York.

Biala, K., Condé, S., Delbaere, B., Jones-Walters, L., Torre-Marín, A., 2012. Streamlining European biodiversity indicators 2020: Building a future on lessons learnt from the SEBI 2010 process. EEA Technical report No 11/2012.

Bibby, C.J., Burgess, N.D., Hill, D.A., Mustoe, S., 2000. Bird Census Techniques. Academic Press, London.

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

Buckland, S.T., Magurran, A.E., Green, R.E., Fewster, R.M., 2005. Monitoring change in biodiversity through composite indices. Phil. Trans. R. Soc. B: Biol. Sci. 360, 243–254.

Buckland, S.T., Studeny, A.C., Magurran, A.E., Illian, J.B., Newson, S.E., 2011. The geometric mean of relative abundance indices: a biodiversity measure with a difference. Ecosphere 2, art100.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.F., Leverington, F., Loh, J., McGeoch, M.A., Mcrae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.C., Watson, R., 2010. Global biodiversity: indicators of recent declines. Science 328, 1164–1168.

Cleveland, W.S., 1993. Visualising Data. Hobart Press, Summit, New Jersey.

Collen, B., Loh, J., Whitmee, S., McRae, L., Amin, R., Baillie, J.E., 2009. Monitoring change in vertebrate abundance: the living planet index. Conserv. Biol. 23, 317–327.

Craigie, I.D., Baillie, J.E., Balmford, A., Carbone, C., Collen, B., Green, R.E., Hutton, J.M., 2010. Large mammal population declines in Africa's protected areas. Biol. Conserv. 143, 2221–2228.

Crawford, T.J., 1991. The calculation of index numbers from wildlife monitoring data. In: Goldsmith, F.B. (Ed.), Monitoring for Conservation and Ecology. Chapman & Hall, London, United Kingdom.

Dennis, B., Ponciano, J.M., Lele, S.R., Taper, M.L., Staples, D.F., 2006. Estimating density dependence, process noise and observation error. Ecol. Monogr. 76, 323–341.

Eaton, M.A., Burns, F., Isaac, N.J.B., Gregory, R.D., August, T.A., Barlow, K.E., Brereton, T., Brooks, D.R., Al Fulaij, N., Haysom, K.A., Noble, D.G., Outhwaite, C., Powney, G.D., Procter, D., Williams, J., 2016. The priority species indicator: measuring the trends in threatened species in the UK. Biodiversity 16, 108–119.

Freeman, S.N., Baillie, S.R., Gregory, R.D., 2001. Statistical analysis of an indicator of population trends in farmland birds. BTO Research Report No. 251.

Gregory, R.D., Van Strien, A., 2010. Wild bird indicators: using composite population trends of birds as measures of environmental health. Ornithol. Sci. 9, 3–22.

Gregory, R.D., Van Strien, A., Vorisek, P., Meyling, A.W.G., Noble, D.G., Foppen, R.P., Gibbons, D.W., 2005. Developing indicators for European birds. Phil. Trans. R. Soc. B: Biol. Sci. 360, 269–288.

Julliard, R., Clavel, J., Devictor, V., Jiguet, F., Couvet, D., 2006. Spatial segregation of specialists and generalists in bird communities. Ecol. Lett. 9, 1237–1244.

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

Le Viol, I., Jiguet, F., Brotons, L., Herrando, S., Lindström, A., Pearce-Higgins, J.W., Reif, J., Van Turnhout, C., Devictor, V., 2012. More and more generalists: two decades of changes in the European avifauna. Biol. Lett. 8, 780–782.

Loh, J., Green, R.E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., Randers, J., 2005. The living planet index: using species population time series to track trends in

biodiversity. Phil. Trans. R. Soc. B: Biol. Sci. 360, 289–295.

Pannekoek, J., van Strien, A.J., 2005. TRIM 3 manual. TRends and Indices for Monitoring data. Research paper no. 0102. Voorburg, The Netherlands. Available freely at: https://www.cbs.nl/en-gb/society/nature-and-environment/indices-and-trends-trim

Secretariat of the Convention on Biological Diversity 2014. Global Biodiversity Outlook 4. Montréal, 155 pp.

Soldaat, L., Visser, H., Van Roomen, M., Van Strien, A., 2007. Smoothing and trend detection in waterbird monitoring data using structural time-series analysis and the Kalman filter. J. Ornithol. 148 (2), 351–357.

Studeny, A.C., Buckland, S.T., Harrison, P.J., Illian, J.B., Magurran, A.E., Newson, S.E., 2013. Fine-tuning the assessment of large-scale temporal trends in biodiversity using the example of British breeding birds. J. Appl. Ecol. 50, 190–198.

Szabo, J.K., Butchart, S.H.M., Possingham, H.P., Garnett, S.T., 2012. Adapting global biodiversity indicators to the national scale: a red list index for Australian birds. Biol. Conserv. 148, 61–68.

Ter Braak, C.J.F., van Strien, A.J., Meijer, R., Verstrael, T.J., 1994. Analysis of monitoring data with many missing values: which method? Hagemeijer, E.J.M., Verstrael, T.J. (Eds.), Bird Numbers 1992. Monitoring and Ecological Aspects. Proceedings of the 12th International Conference of the International Bird Census Committee and European Ornithological Atlas Committee. Statistics Netherlands/SOVON (Voorburg/Beek-Ubbergen, The Netherlands) 663–673.

Van Strien, A.J., Soldaat, L.L., Gregory, R.D., 2012. Desirable mathematical properties of indicators for biodiversity change. Ecol. Indic. 14, 202–208.

Van Strien, A.J., Gmelig Meyling, A.W., Herder, J.E., Hollander, H., Kalkman, V.J., Turnhout, S., van der Hoorn, B., van Strien-van Liempt, W.T.F.H., Poot, M.J.M., van Swaay, C.A.M., van Turnhout, C.A.M., Verweij, R.J.T., Oerlemans, N.J., 2016. Modest recovery of biodiversity in a western European country: the Living Planet Index for the Netherlands. Biol. Conserv. 200, 44–50.

Van Swaay, C., et al., 2013. The European Grassland Butterfly Indicator: 1990–2011. EEA Technical report No 11/2013.

Van Turnhout, C.A.M., Willems, F., Plate, C., van Strien, A., Teunissen, W., van Dijk, A., Foppen, R., 2008. Monitoring common and scarce breeding birds in the Netherlands: applying a posthoc stratification and weighting procedure to obtain less biased population trends. Catalan J. Ornithol. 24, 15–29.

Van Turnhout, C.A.M., Foppen, R.P.B., Leuven, R.S.E.W., Van Strien, A.J., Siepel, H., 2010. Life-history and ecological correlates of population change in Dutch breeding birds. Biol. Conserv. 143, 173–181.

Visser, H., 2004. Estimation and detection of flexible trends. Atmos. Environ. 38, 4135–4145.