



Desirable mathematical properties of indicators for biodiversity change

A.J. van Strien^{a,b,*}, L.L. Soldaat^a, R.D. Gregory^c

^a Statistics Netherlands, P.O. Box 24500, 2490 HA The Hague, The Netherlands

^b Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

^c The Royal Society for Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, UK

ARTICLE INFO

Article history:

Received 24 January 2011

Received in revised form 2 July 2011

Accepted 10 July 2011

Keywords:

Indicators
Biodiversity
Geometric mean
Simpson index
Shannon index

ABSTRACT

Numerous indicator approaches are found in the scientific literature to describe changes in biodiversity. It is however far from clear which indicators are most appropriate and which are less suitable to summarize trends in biodiversity. One reason for this lack of clarity is that so far the mathematical properties of indicator approaches have had little attention. In this paper, we derive a number of desirable mathematical properties of indicators from economic price-index theory and apply these in the form of tests to 10 metrics to summarize changes in biodiversity. The metrics species richness, Simpson index, Shannon index, Buckland's modified Shannon index and Sørensen's similarity coefficient violate the monotonicity and proportionality test. The percentage of increasing minus declining species also fails the proportionality test, and in the case where trends are assessed relative to the preceding year, this metric also violates the identity test. Most of these indicators are sensitive to spatial scale. The arithmetic and geometric mean of population indices and the mean abundance have better mathematical performance, but the first two are sensitive to appearing and disappearing species in the system surveyed. The metric mean abundance however can only be applied under particular conditions and has some undesirable properties. Unlike the arithmetic mean, the geometric mean is not sensitive to the base year chosen and has the most favourable mathematical properties of the indicators evaluated. The geometric mean can be straightforwardly extended to take into account unequal values of species if desired.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The state of biodiversity in the world is deteriorating and there is no reduction in the rate of decline (Butchart et al., 2010). Not only has the number of species with risk of extinction increased, but also many common species have declining populations. Moreover, substantial shifts in species composition in many ecosystems have occurred with habitat specialists as typical “losers” and habitat generalists as “winners”, a phenomenon called biotic homogenisation (McKinney and Lockwood, 1999).

Information on changes in biodiversity is of political relevance since world leaders agreed to reduce the rate of biodiversity loss (Secretariat of the Convention on Biological Diversity, 2003). Such information is usually provided through so-called headline indicators (e.g. see www.twentyten.net and European Environmental Agency, 2007). The purpose of headline indicators is to provide comprehensible information to policy-makers and the general public about progress towards meeting the political targets (Gregory et al., 2008). Because species are central to biodiversity, we here

focus on headline indicators that summarize trend information from multiple species using data collected on a large number of sites.

Numerous indicator approaches are found in the scientific literature to describe overall changes in particular species groups and in biodiversity more generally. Classical metrics are species richness, the Shannon index and Simpson's index to describe changes in the diversity of species (Spellerberg, 1991). To describe biotic homogenisation, similarity measures such as the Sørensen or the Bray–Curtis index have become popular in the last decades (Rooney et al., 2004, 2007). Of more recent date are measures to combine multiple species trends based on yearly indices, such as the geometric mean of species indices used for the European Farmland Bird Index (Gregory et al., 2005). Although different indicator approaches have been applied widely, it is still unclear which biodiversity indicators are most appropriate and which are less suitable to unambiguously summarize trends in biodiversity. This is because biodiversity indicators must satisfy a variety of requirements. First, headline indicators should be ecologically relevant i.e. they need to summarize degradation or improvement of ecosystem quality adequately. This implies that the indicator should perform as expected in ecological scenarios with respect to species extinction, species invasion or biotic homogenisation (Lamb et al., 2009). This is not solely a scientific criterion, but includes a subjective judgment,

* Corresponding author at: Statistics Netherlands, P.O. Box 24500, 2490 HA The Hague, The Netherlands. Tel.: +31 70 3374214; fax: +31 70 3373984.

E-mail addresses: asin@cbs.nl, lnmasint@wxs.nl (A.J. van Strien).

because interested parties may value the changes of particular species and species groups in a system differently (van Turnhout et al., 2007). Thus, it is essential to interact with stakeholders in the development of a biodiversity indicator to ensure that indicators are ecologically and policy relevant, user driven and easily understood (Green et al., 2005; Gregory et al., 2005). Preferably, indicators describing trends in species groups can be generalised to trends in biodiversity more generally and can be connected with causes of trends (van Strien et al., 2009).

Furthermore, the making of an indicator should be feasible in practice. Obviously, one cannot monitor all species groups in all habitat types in all localities. At best, it is possible to monitor a selection of species groups, preferable species sensitive to environmental change over relatively short time-scales. The data for the indicator should be realistic to collect to ensure regular updates (Gregory et al., 2005).

Finally, indicators should have desirable statistical and mathematical properties. Statistical properties concern the estimators of indicators which preferably have confidence intervals, allow statistical testing of changes and are unbiased and precise (Sutherland, 2006). Mathematical properties describe how logical the behaviour of an indicator is given the underlying species data. These properties concern the model chosen to summarize changes, for instance whether changes in the constituent species are averaged arithmetically or geometrically. The statistical properties of indicators have had considerable attention in the literature (e.g. Giavelli et al., 1986; Spellerberg, 1991; Fleishman et al., 2006), but the mathematical properties of indicators of biodiversity change are rarely acknowledged. Lamb et al. (2009) dealt with mathematical properties implicitly by examining the behaviour of indicators under different ecological scenarios. Buckland et al. (2005) explicitly addressed the importance of desirable mathematical properties of indicators, among others by stating that an indicator should have a monotone relation with changes in the constituent species. Here we extend the approach of Buckland et al. (2005) by discussing a number of desirable mathematical properties that so far were neglected by many researchers working with indicators. We believe that it is useful to take these properties into account in the choice of indicators and studying them strengthens the theoretical basis of indicators (Niemeijer and de Groot, 2008). We first outline a number of desirable mathematical properties derived from economic price index theory and then examine a set of well-known biodiversity indicators and test their performance regarding these properties, using a simple theoretical example with time series for a few species in one spatial unit.

2. Methods

2.1. Biodiversity indicators

We include in our evaluation indicators based on data about changes in the abundance of individual species that are then combined. Thus, measures based on occupancy were not considered (though see Section 4), nor were compound measures, such as threat status of species used for the Red List Index (Butchart et al., 2007, 2010). Note that we want to evaluate the appropriateness of these indicators as measures of change and do not consider their usefulness for other possible applications. The description of indicators follows Spellerberg (1991) or Buckland et al. (2005):

1. *Species richness*: this is the number of species in a system during a particular period, e.g. year. Species richness is a popular metric to express conservation value (Fleishman et al., 2006). This metric does not take into account the abundances of the species.
2. *Simpson's index*: this is a widely used diversity metric, which takes both the number of species into account and their abundances (Spellerberg, 1991). Let d_{ij} be the number of individuals in the system (abundance) in year j that belong to species i and $p_{ij} = d_{ij} / \sum_i d_{ij}$ the proportion of them from all species. Simpson's index then is $D_j = \sum_i p_{ij}^2$. As a diversity metric we use the transformed index $-\log_e D_j$. This produces low values when a few species dominate (low evenness) and high values when no single species dominates (high evenness).
3. *Shannon index*: this is again a widely used diversity metric, defined as $H_j = -\sum_i p_{ij} \log_e(p_{ij})$. Similar to the transformed Simpson's index, low values are achieved when a few species dominate and high values when no species dominate. If all species increase or decline with the same rate, both Simpson's index and the Shannon index remain unchanged however. To overcome this weakness as a biodiversity indicator, Buckland et al. (2005) designed the modified Shannon index.
4. *Modified Shannon index*: Buckland et al. (2005) defined $M_j = -\sum_i q_{ij} \log_e(q_{ij})$, with $q_{ij} = d_{ij} / \sum_i d_{i1}$. Thus abundances per species in all years are divided by the summed abundances of all species in year 1.
5. *Percentage increasing – percentage declining species of all species, relative to the first year*: this metric is frequently used in the “grey” literature to evaluate if more species decline than increase in a system. So this index does not consider the magnitude of change, but only its sign. A positive index reflects a higher number of species increasing compared to those declining, and vice versa.
6. *Percentage increasing – percentage declining species of all species, relative to the preceding year*: this metric is a variant on metric 5 and is an attempt to achieve an indicator tracking recent changes better than metric 5. Other variants are metrics relative to e.g. 5 or 10 years earlier and such metrics will have properties intermediate between metric 5 and 6.
7. *Arithmetic mean of abundance indices*: species indices are typically used in the framework of large-scale monitoring programmes. It is convenient to compute relative indices by dividing the abundance of a species in year j by its abundance in year 1, although the choice of another base year is allowed. The indices of m species then are averaged to obtain the arithmetic yearly mean: $R_j = (1/m) \sum_i (d_{ij}/d_{i1})$.
8. *Geometric mean of abundance indices*: instead of the arithmetic mean to aggregate species indices, some authors have used a geometric mean (Gregory et al., 2005; Loh et al., 2005). This is computed by taking the average of the log of indices of m species per year followed by a back-transformation: $G_j = \exp((1/m) \sum_i \log(d_{ij}/d_{i1}))$. The arithmetic and geometric mean of the base year are 100 by definition. Species are weighted equally so neither index reflects the potential different desirability of individual species. Both arithmetic mean and geometric mean are considered stable when positive and negative changes of indices, as well as the magnitude of these changes, are in balance. In a situation where species increase or decrease at the same fixed rate, then if more species decline than increase, the mean goes down and vice versa. In the case where one species increases from 100 to 200, and another species from 100 to 50, the geometric mean index of the two is 100, whereas the arithmetic mean index of the two is 125. The latter value seems odd when it is balancing doubling and halving species trends.
9. *Mean abundance of species*: this is the yearly mean abundance of all constituent species: $AB_j = (1/m) \sum_i d_{ij}$. This metric is not based on combining relative indices, but on summing absolute abundances of species. Alternatively, one might use the total abundance of a group of species rather than their mean.

10. *Sørensen's similarity coefficient*: this metric enables one to describe the similarity in species composition in a system over time. Sørensen's index to compare two different years in one system is: $C_j = 1 - (\sum_i |d_{ij} - d_{i1}| / (\sum_i d_{ij} + \sum_i d_{i1}))$, thus the differences in abundance between year 1 and year j are being compared and divided by the total abundance of all species taken together in both years. Values of C_j range from zero, for no similarity in species composition, to 1 for identical species composition and abundances. The complement of this index is equal to the widely used Bray–Curtis metric (Spellerberg, 1991; Rooney et al., 2004).

2.2. Desirable properties of indicators

Taking into account the mathematical properties of indicators is not exclusively important for biodiversity but also applies to other statistical fields. It has had much attention in the first half of the 20th century in economic price index theory when numerous indicators were proposed to evaluate changes in the cost of living (Balk, 2008). Examining the mathematical behaviour of different price index approaches has been of great help in achieving consensus among economists and mathematicians on the best approach. We build on this work in price index theory by deriving desirable mathematical properties from Balk (2008), which were then tailored to our purposes in assessments of biodiversity change. Our theoretical starting point is that all species are equally valued and any increase is interpreted as positive and any decline as negative. This interpretation ignores any changes in species evenness. The evenness would increase if a few abundant species decline while the other species remain constant, and considering this positive contradicts our starting point. Any species entering the system is considered beneficial and species disappearing from the system are considered losses. Then the following requirements (or tests) seem natural for a measure to assess changes in biodiversity over time:

(T1) *Monotonicity*. If all species are declining in a system, the indicator is also declining. This includes the desired behaviour of the indicator when species disappear from the ecosystem: the indicator then should decline as well. Similar but opposite behaviour is expected for an indicator describing increases instead of declines.

(T2) *Proportionality*. If all individual species abundances are changed by a common factor, the indicator is also changed by this factor.

(T3) *Identity*. If all individual species abundances in a particular year are equal to those in another year, the indicator value is also equal.

(T4) *Base year invariance*. Changes in the indicator are not sensitive to the choice of the base year chosen for indices of the constituent species. This is relevant only for metrics based on indices.

(T5) *Oversensitivity to appearing and disappearing species*. Changes in the indicator are not dominated by new species entering the ecosystem or species that disappear from the ecosystem.

(T6) *Spatial scale invariance*. Changes in the indicator are not sensitive to the spatial scale chosen.

2.3. Test of indicators

We examined requirement T1–T5 for all indicators using a simple numerical example with only 1 spatial unit, 3 species and 5 years. T6 was only evaluated using the scientific literature. The abundance of species in our example is both given in numbers of individuals per year and in indices reflecting changes relative to the first year (Table 1). Simpson's index, the Shannon index, the modified Shannon index, the percentage of declining species, mean abundance and Sørensen similarity were all based on abundance data in Table 1. Geometric and arithmetic means were derived

Table 1

Example dataset with population numbers and indices for 3 species. Indices are values relative to the first year, which is set at 100.

	Year 1	Year 2	Year 3	Year 4	Year 5
Population abundance					
Species 1	2000	1000	4000	1000	4000
Species 2	1000	500	500	500	2000
Species 3	250	125	250	125	500
Indices					
Species 1	100	50	200	50	200
Species 2	100	50	50	50	200
Species 3	100	50	100	50	200

Table 2

Results for each indicator based on data in Table 1.

	Year 1	Year 2	Year 3	Year 4	Year 5
Expected indicator value	100	50	– ^b	50	200
Species richness	100	100	100	100	100
Simpson's index	100	100	44	100	100
Shannon index	100	100	62	100	100
Modified Shannon index	100	90	27	90	39
% Increasing–% declining species using first year	– ^a	–100	0	–100	100
% Increasing–% declining species using preceding year	– ^a	–100	67	–67	100
Arithmetic mean	100	50	117	50	200
Geometric mean	100	50	100	50	200
Mean abundance	100	50	146	50	200
Sørensen similarity	100	67	69	67	69

^a Not defined.

^b Varies by indicator.

from indices in Table 1. In a separate test to examine the effect of a newly appearing species in the system, we again used the data from Table 1, but slightly modified (see heading of Table 4). To facilitate comparison, all indicators are indexed with base year value set at 100, except % increasing–% declining species. Expected indicator values were derived according to T2. Details of the computations are given in Appendix A.

For simplicity's sake, we worked only with a reference year here and not with other references, such as species abundance in the pre-industrial state. In addition, we examine changes in one species group at a time (e.g. birds or butterflies) and do not consider protocols to aggregate data from several species groups (as Scholes and Biggs, 2005).

3. Results

Obviously, species richness does not reflect any changes in abundance of species unless species enter or leave the system, thus this metric violates test T1 (monotonicity) and T2 (proportionality) (Tables 2 and 3). Identical species composition produces equal values for species richness and test T3 (identity) is satisfied. Species richness is certainly sensitive to appearing and disappearing species (T5), but not oversensitive: if one out of 3 species appears in the system in year 2, indices are not extremely different (compare species richness in Tables 2 and 4). But changes in species richness are intrinsically scale-dependent (T6) (Sax and Gaines, 2003; Keil et al., 2010). For instance, the invasion of a non-native species in a country leads to an increase of the species richness of the country, while at the same time it may lead to local declines in species richness if the invader increases at the expense of native species.

Although all three species in the example decline in years 1–2 (Table 1), both Simpson's index and the Shannon index remain unchanged in these years, thus these two metrics also do not satisfy tests T1 and T2 (Table 2 and 3). The modified Shannon index is designed to overcome this shortcoming of the Shannon index and

Table 3

Scores of indicators for each mathematical requirement. X = substantial difficulty to satisfy test. T4 is only applicable for geometric and arithmetic means of indices.

	(T1) Monotonicity	(T2) Proportionality	(T3) Identity	(T4) Base year invariance	(T5) Oversensitivity (dis)appearing species	(T6) Spatial scale invariance
Species richness	X	X				X
Simpson's index	X	X				X
Shannon index	X	X				X
Modified Shannon index	X	X				X
% Increasing–% declining species using first year		X				
% Increasing–% declining species using preceding year		X	X			
Arithmetic mean				X	X	
Geometric mean					X	
Mean abundance						
Sørensen similarity	X	X				X

indeed this metric declines in years 1–2 (Table 2). But the magnitude of change of the modified Shannon in these years differs from those of the three species, thereby violating T2. Even worse, the doubling of species abundances in year 5 as compared to year 1 is not accompanied by any increase in the modified Shannon index but instead by a decline (Table 2), so T1 is violated. A three-fold increase of the three species would even produce negative values for this metric. The negative values of the modified Shannon arise because the proportions q_{ij} do no longer sum to 1 as in the unmodified Shannon and values of $\log_e(q_{ij})$ easily become positive when species increase. Although the modified Shannon index might be considered an improvement compared to the original Shannon index because it captures declines in all species, it fails to capture increases in species.

Simpson's index, Shannon index and modified Shannon index all fulfil test T3 (Table 3). At first sight the Shannon index and the modified Shannon index are susceptible to (dis)appearing species (T5): they cannot be calculated when the abundance of a species equals zero, because the logarithms of p_{ij} and q_{ij} are not defined. But this is easily treated by setting the products $p_{ij} \log_e(p_{ij})$ and $q_{ij} \log_e(q_{ij})$ at zero when a species is absent. These three indicators then are not much affected by the appearance of a new species in the dataset (compare years 1 and 2 in Tables 2 and 4), and the same is true for disappearance. Similar to species richness, also changes in Simpson's index, the Shannon index and the modified Shannon index are susceptible to spatial scale (T6). At low spatial level, lower values of these metrics arise when a few species are becoming dominant and evenness declines. But when the identity of these dominant species varies between locations, evenness does not necessarily decline at higher spatial scale.

Table 4

Results for each indicator in case of a newly appearing species in year 2. Results are based on data in Table 1 with one modification: species 3 in year 1 had value 0 for all indicators, except for arithmetic mean and geometric mean for which value 1 was used.

	Year 1	Year 2	Year 3	Year 4	Year 5
Species richness	100	150	150	150	150
Simpson's index	100	125	55	125	125
Shannon index	100	135	84	135	135
Modified Shannon index	100	125	19	125	29
% Increasing–% declining species using first year	– ^a	–33	33	–33	100
% Increasing–% declining species using preceding year	– ^a	–33	67	–67	100
Arithmetic mean	100	4200	8417	4200	16800
Geometric mean	100	315	630	315	1260
Mean abundance	100	36	106	36	144
Sørensen similarity	100	65	65	65	63

^a Not defined.

The percentage of increasing minus declining species relative to the first year satisfies test T1, because when all species decline, this metric also declines and the same accounts for increases (Table 2). But the values in Table 2 differ much from the abundance changes in the three species, so this indicator does not satisfy T2 (Table 3). Test T3 is satisfied (compare years 2 and 4 in Table 2). Test T3 is however not satisfied if the percentage of increasing minus declining species is chosen relative to the preceding year, because values for years 2 and 4 differ even though all three species have similar values (Table 2). That is because by definition the values of this metric do not solely depend on the species abundances in a particular year, but also on the abundances in the preceding year. Both these two metrics are based on increasing minus decreasing species and are neither oversensitive to disappearing species, nor to newly appearing species (T5; Table 4). New species are taken into account indirectly using the percentage of all species. In addition, both indicators are insensitive to spatial scale (T6), because trends in abundance measured at low spatial scale add up to trends in abundances at higher spatial scale, so changes at different spatial scales are unlikely to be opposite.

In contrast to all earlier discussed indicators, the changes in arithmetic and geometric mean are all equivalent to the changes in the three underlying species (Table 2), so these two indicators satisfy tests T1, T2 as well as T3 (Table 3). The geometric mean does not depend on the base year chosen (T4), but the arithmetic mean violates this test. If instead of the first year, the third year is chosen as base year for the constituent species, the arithmetic mean does no longer increase from 100 to 117 between years 1 and 3 (Table 2), but instead declines from 100 to 86 (see Appendix B for computational details). In addition, both geometric mean and arithmetic mean are oversensitive to disappearing and appearing species (T5; Table 4), but in a different manner. The geometric mean is sensitive to species that are disappearing from the system. The yearly indices of the species then go to very low values and this may affect the geometric mean substantially. If the index reaches zero, one has to substitute a small constant for zero to avoid taking the logarithm of zero (this is also required for the arithmetic mean if the base year has an index of zero; see Table 4). The same sensitivity accounts for new species in the ecosystem, which have zero abundance in the years prior to their appearance. The arithmetic mean is however very sensitive to strong increases of species newly entering the system. Indices for such species achieve high values if the first year with low abundances is set at value 100 and the arithmetic mean is strongly influenced by such high values in single species (Table 4). One may try to reduce the impact of appearing species by not using the first year but the last year as base year. But this has the drawback of constantly changing indicator values when data for a new year are added to the dataset because values of the arithmetic mean depend on the base year (T4). Any other base year than the first or the last year may fulfil better, but this works only until

a new and increasing species appears and another suitable base year of the arithmetic mean must be found. The geometric mean is less influenced by newly appearing species (apart from the need to avoid the log of zero) (Table 4). That is because the mean of the logarithm of the indices rather than the mean of the indices themselves dampens extreme changes in species. Similar to the percentage of increasing minus declining species, the geometric and arithmetic means are not dependent on spatial scale (T6).

The indicator mean abundance passes T1, T2, T3, T4, and is not sensitive to (dis)appearing species (T5), because such species will have low abundances anyway (Tables 2–4), nor to spatial scale (T6). The Sørensen index is lower in year 5 than in year 1, despite the increase of all species (Table 2). Halving the abundance of all species (year 2) or doubling the species (year 5) produces more or less equal values (Table 2). Thus, both T1 and T2 are violated (Table 3). Test T3 and T5 are satisfied, but according to Rooney et al. (2007) trends in similarity are scale-dependent (T6) (Tables 3 and 4).

4. Discussion

Using a number of tests, we have examined the mathematical behaviour of a number of widely used biodiversity indicators. Note that if an indicator violates one or more of these tests, it does not imply that the indicator is without any value. Many of the indicators examined have proved useful for specific ecological questions or are useful diagnostic tools (see e.g. Spellerberg, 1991; Rooney et al., 2007). Nevertheless, violation of the tests may hamper the applicability of a metric for our purposes, i.e., to serve as a headline indicator whose magnitude of change can be unambiguously interpreted as the degree of progress made towards meeting political targets for biodiversity change.

4.1. Use of indicators to reflect temporal changes

The tests T1, T2 and T3 determine the applicability of indicators to reflect changes in abundance. Species richness, Simpson's index, Shannon index, the modified Shannon index and Sørensen index violate T1 and T2 and may produce inappropriate values, which make them unsuitable as headline indicators for biodiversity change. Although both variations on the percentage of increasing minus decreasing species have a better performance than the indicators just mentioned (Table 3), their values appear strange and difficult to grasp, and they do not reflect the precise magnitude of trends among species (Table 2). The arithmetic and geometric mean as well as the mean abundance have no difficulties considering the interpretation of their temporal changes in abundance.

Test T5 is associated with the applicability of indicators to reflect changes in the number of species. Most of the indicators studied satisfactorily track any changes in the number of species over time. Some complications were encountered for the arithmetic and geometric mean though, because they can be susceptible to the introduction or disappearance of species.

4.2. Use of indicators for benchmarking

We expect that biodiversity indicators shall be increasingly used to compare regions and countries for benchmarking purposes. Test T4 and T6 are linked to the applicability of indicators for benchmarking. Because changes in species richness, Simpson's index, the Shannon index, the modified Shannon index and the Sørensen index are scale-dependent, one cannot compare them between regions unless their assessments are conducted on the same spatial scale. All other indicators can be compared straightforwardly between regions or countries because their trends are unlikely to differ by spatial scale. This is however different if indicators are based on measurements of occupancy instead of abundance,

because occupancy is intrinsically scale-dependent. Occupancy-based indicators require standardised spatial scales in order to make them comparable.

The sensitivity for the choice of the base year used for indices of the species may be another difficulty to compare indicators of different regions. Both the arithmetic and geometric means of indices can only be compared between regions if the same base year is used. If different base years are used for different regions, it is necessary for the arithmetic mean to reset the base year of all individual species indices. Thus, for benchmarking purposes one needs information about the underlying species indices. In contrast, for the geometric mean it is sufficient to reset the base year at the indicator level (see Appendix B for computational details).

4.3. Comparing indicators

Table 3 suggests that the mean abundance has the best mathematical performance. Nevertheless, its use is limited because mean abundance is often not an ecologically relevant indicator. Because it is based on absolute abundances rather than relative indices, it is easily dominated by changes in the most abundant species and this is often not in line with the desires of stakeholders. Its use is limited to those situations where an entire set of species has parallel declines or increases. An example is the overall decline in the abundance of common butterflies in the Netherlands (van Dyck et al., 2009).

Putting mean abundance aside, the arithmetic and geometric mean have the most favourable properties: they have a correct interpretation of changes in abundance and they do not suffer from sensitivity from spatial scale. They may be oversensitive to (dis)appearing species, but we believe this is an acceptable disadvantage for their use. A bigger problem is the sensitivity of arithmetic mean regarding the base year. In addition, the trends in arithmetic means cannot be unambiguously tested statistically when taking into account the underlying species indices. That is because the indices are usually heteroscedastic, i.e. the spread in indices tends to increase if indices have higher values. For many statistical tests homoscedasticity is a precondition (Sokal and Rohlf, 1981) which typically is achieved by a log-transformation of the data prior to testing. Yet using the log of indices is essentially testing the trend in the geometric mean rather than in the arithmetic mean. Out of all the indicators evaluated, we therefore favour the geometric mean, also because it is the natural scale in ecology. Populations grow geometrically, not arithmetically (Volterra, 1931), so it is more natural to think in terms of logarithmic declines and increases than in terms of absolute changes. The adequate mathematical performance of geometric means supports their current application to summarize trends in the abundance of species in the Living Planet Index (Loh et al., 2005), the European Farmland Bird Indicator (Gregory et al., 2005) and other biodiversity indicators to measure progress towards biodiversity targets (Butchart et al., 2010).

Our requirements (T1–T6) for indicators mainly apply to volume-indicators, i.e. indicators that reflect aggregated growth and decline of species populations. Mean abundance, arithmetic and geometric means have better scores than the other indicators simply because these three are all volume-indicators. To a lesser degree this holds for the two indicators focussed on increasing and declining species which are a very simple form of volume-indicators. The remaining indicators are only partly (modified Shannon), or not at all volume-indicators and describe changes in species composition rather than growth and decline of populations. Our indicator requirements however are not about changes in species composition and such requirements would be difficult to formulate because these depend on the situation. For this reason, we did not include evenness in our requirements, in

contrast to Buckland et al. (2005), and this explains why the traditional diversity indicators (Simpson and Shannon index) do not perform well here. We neglected evenness in species abundance deliberately, because we believe that a higher evenness does not necessarily mean higher biodiversity. The evenness in abundance of species is not so much a generally desirable mathematical property, but rather an aspiration that stakeholders might or might not have for a particular species community. For instance, if all farmland birds go down, but the most abundant species declines stronger than the other species, an index of evenness will increase and probably Simpson's index and Shannon index too. However, most stakeholders would feel that the indicator should reflect a decline. Furthermore, Sørensen's similarity is not useful as volume-indicator because a change in similarity has no direction: it only describes the deviation of the abundance of species in the system from those in the reference situation, but the same deviation may either result from deterioration of the ecosystem (all species in decline) or improvement (all species going up). Thus, biotic homogenisation as measured by a similarity index is not identical to biotic impoverishment (Rooney et al., 2007).

In our theoretical example, we have treated all species equally to create a simple situation to examine the behaviour of indicators. Stakeholders however may wish to distinguish between indices of higher and lower conservation interest, or sometimes even want to give unwanted species negative values (Nielsen et al., 2007). Gregory and van Strien (2010) considered the advantages and disadvantages of weighting multispecies indices in different ways. For instance, it does not make sense to take the geometric mean of indices of both habitat generalists and specialists if the increase of common generalist species and the decline of specialists are two sides of the same process of biological degradation. One possibility is to compose indicators for separate groups of species. An example is the European Farmland Bird Index which consists of species characteristic of agricultural habitats (Gregory et al., 2005). Using groups can however be difficult if species gradually differ in their degree of habitat specialisation or other characteristics that one wants to take into account. In such cases, weighting species according to these values might be more appropriate. Weighting can be easily incorporated in the calculation of the geometric mean by weighting the log of the indices rather than the indices themselves:

$$G_j = \exp \left(\frac{1}{s} \sum_i (\text{species value}_i) \log \left(\frac{d_{ij}}{d_{i1}} \right) \right)$$

in which s stands for the sum of all species values. Obviously, if species values are equal to 1, this formula reduces to the geometric mean. A species value of 2 implies that such a species is used twice in the indicator. More generally, a species value >1 acts as if pseudo-species are added to the indicator group, meaning that criterion T1 and T2 are still met for weighted geometric means. Negative values for species might also be used, if desired. Because of the use of indices, the weighted geometric mean is not affected by any differences in abundance between species, but only by changes in abundance within species. This is an advantage over the more conventional "site-based" approach to differentiate between species values. In the site-based approach, mean species values per site are calculated taking into account their abundances (see e.g. Diekmann, 2003; Devictor et al., 2008). The site-based indicator reduces to the mean abundance indicator if species values are equal. Consequently, in the site-based approach abundant species have a bigger influence on the results than rare species, which may be unintended (see Appendix C for an example). The effect of abundance may be circumvented in the site-based approach through using information about occurrences only, but then any changes in abundance within the species are ignored. Both

geometric mean and arithmetic mean have the advantage of taking into account different species values without including effects of abundance differences between species, though the arithmetic mean to a somewhat lesser extent (see example in Appendix C).

Buckland et al. (2005) and Lamb et al. (2009) in their reviews on indicators considered the traditional diversity indices (Simpson's index, Shannon index) also unsuitable as metrics of biodiversity change. Buckland et al. (2005) disapproved the arithmetic mean of indices on similar grounds as we do and favoured geometric means and the modified Shannon index. However, they did not comment on the poor performance of the modified Shannon index when species increase, which is one of our arguments against the modified Shannon index. Lamb et al. (2009) gave the arithmetic mean a higher score than the geometric mean, mainly based on the disadvantage of geometric means that adding a small constant may be necessary (see above). We however consider this not a substantial weakness, while we regard the sensitivity for the choice of the base year of arithmetic means as a weaker point in practice, a characteristic not identified by Lamb et al. (2009).

We conclude that it is important to consider the mathematical properties of indices during the selection process of biodiversity indicators. However, mathematical properties are not the only criteria to take into account. Statistical arguments, practicality and ease of communication are also relevant for the selection (see Section 1).

Acknowledgements

We thank Lodewijk van Duuren and Bert Balk for valuable comments on an earlier draft.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ecolind.2011.07.007](https://doi.org/10.1016/j.ecolind.2011.07.007).

References

- Balk, B., 2008. Price and Quantity Index Numbers. Models for Measuring Aggregate Change and Difference. Cambridge University Press, New York.
- 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* 360, 243–254.
- Butchart, S.H.M., Akçakaya, H.R., Chanson, J., Baillie, J.E.M., Collen, B., Quader, S., Turner, W.R., Amin, R., Stuart, S.N., Hilton-Taylor, C., 2007. Improvements to the Red List Index. *PLoS One* 2 (1), e140. doi:10.1371/journal.pone.0000140.
- Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharleman, J.P.W., et al., 2010. Global biodiversity: indicators of recent declines. *Science* 328, 1164–1168. doi:10.1126/science.1187512.
- Devictor, V., Julliard, R., Couvet, D., Jiguet, F., 2008. Birds are tracking climate warming, but not fast enough. *Proc. R. Soc. B* 275, 2743–2748. doi:10.1098/rspb.2008.0878.
- Diekmann, M., 2003. Species indicator values as an important tool in applied plant ecology—a review. *Basic Appl. Ecol.* 4, 493–506.
- European Environmental Agency, 2007. Halting the Loss of Biodiversity by 2010: Proposal for a First Set of Indicators to Monitor Progress in Europe. EEA Technical Report No. 11/2007, Copenhagen.
- Fleishman, E., Noss, R.F., Noon, B.R., 2006. Utility and limitations of species richness metrics for conservation planning. *Ecol. Indic.* 6, 543–553.
- Giavelli, G., Rossi, O., Sartore, F., 1986. Comparative evaluation of four species diversity indices related to two specific ecological situations. *Field Studies* 6, 429–438.
- Green, R.E., Balmford, A., Crane, P.R., Mace, G.M., Reynolds, J.D., Turner, R.K., 2005. A framework for improved monitoring of biodiversity: responses to the world summit on sustainable development. *Conserv. Biol.* 19, 56–65.
- Gregory, R.D., van Strien, A., Vorisek, P., Gmelig Meyling, G.W., Noble, D.G., Foppen, R.P.B., Gibbons, D.W., 2005. Developing indicators for European birds. *Phil. Trans. R. Soc. B* 360, 269–288.
- Gregory, R.D., Vorisek, P., Noble, D.G., van Strien, A., Klvanová, A., Gmelig Meyling, A.W., Joys, W., Foppen, R.P.B., Burfield, I.J., 2008. The generation and use of bird population indicators in Europe. *Bird Conserv. Int.* 18, S223–S244.
- 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.
- Keil, P., Biesmeijer, J.C., Barendregt, A., Reemer, M., Kunin, W.E., 2010. Biodiversity change is scale-dependent: an example from Dutch and UK hoverflies (Diptera, Syrphidae). *Ecography* 33, 1–10. doi:10.1111/j.1600-0587.2010.06554.x.

- 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.
- 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* 360, 289–295.
- McKinney, M.L., Lockwood, J.L., 1999. Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends Ecol. Evol.* 14, 450–453.
- Niemeijer, D., de Groot, R., 2008. A conceptual framework for selecting environmental indicator sets. *Ecol. Indic.* 8, 14–25.
- Nielsen, S.E., Bayne, E.M., Schieck, J., Herbers, J., Boutin, S., 2007. A new method to estimate species and biodiversity intactness using empirically derived reference conditions. *Biol. Conserv.* 137, 403–414.
- Rooney, T.P., Wiegmann, S.M., Rogers, D.M., Waller, D.M., 2004. Biotic impoverishment and homogenization in unfragmented forest understory communities. *Conserv. Biol.* 18, 787–798.
- Rooney, T.P., Olden, J.D., Leach, M.K., Rogers, D.A., 2007. Biotic homogenization and conservation prioritization. *Biol. Conserv.* 134, 447–450.
- Sax, D.F., Gaines, S.D., 2003. Species diversity: from global decreases to local increases. *Trends Ecol. Evol.* 18, 561–566.
- Scholes, R.J., Biggs, R., 2005. A biodiversity intactness index. *Nature* 434, 45–49.
- Secretariat of the Convention on Biological Diversity, 2003. Handbook of the Convention on Biological Diversity. Earthscan, London.
- Sokal, R.R., Rohlf, F.J., 1981. *Biometry*. W.H. Freeman and Company, New York.
- Spellerberg, I.A., 1991. *Monitoring Ecological Change*. Cambridge University Press, Cambridge.
- Sutherland, W.J., 2006. *Ecological Census Techniques. A Handbook*, second ed. Cambridge University Press, Cambridge.
- van Dyck, H., van Strien, A.J., Maes, D., van Swaay, C.A.M., 2009. Declines in common, widespread butterflies in a landscape under intensive human use. *Conserv. Biol.* 4, 957–965.
- van Strien, A.J., van Duuren, L., Foppen, R.P.B., Soldaat, L.L., 2009. A typology of indicators of biodiversity change as a tool to make better indicators. *Ecol. Indic.* 9, 1041–1048.
- van Turnhout, E., Hisschemöller, M., Eijsackers, H., 2007. Ecological indicators: between the two fires of science and policy. *Ecol. Indic.* 7, 215–228.
- Volterra, V., 1931. Variations and fluctuations of the number of individuals in animal species living together. In: Chapman, R.N. (Ed.), *Animal Ecology*. McGraw-Hill Book Comp., New York.