Integrating ongoing biodiversity monitoring: potential benefits and methods

Pierre-Yves Henry · Szabolcs Lengyel · Piotr Nowicki · Romain Julliard · Jean Clobert · Tatjana Čelik · Bernd Gruber · Dirk S. Schmeller · Valerija Babij · Klaus Henle

Abstract Halting the loss of biodiversity comes along with the need to quantify biodiversity composition and dynamics at large spatial and temporal scales. Highly standardized, international monitoring networks would be ideal, but they do not exist yet. If we are to assess changes in biodiversity now, combining output available from ongoing monitoring initiatives is the only option. However, integration of biodiversity information across schemes is still very poorly developed. In this paper, we outline practical issues to be considered when planning to combine existing monitoring information. First, we provide an overview of avenues for integration along the four dimensions that characterize a monitoring design: sample size, biological coverage, spatial coverage and temporal coverage. We also emphasize that complementarity in monitoring targets across schemes

P.-Y. Henry (✉)
Département Écologie et Gestion de la Biodiversité, UMR 5173 & UMR 7179,
Muséum National d'Histoire Naturelle, 55 Rue Buffon, 75005 Paris, France
e-mail: henry@mnhn.fr

S. Lengyel
Department of Ecology, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary
e-mail: szabolcslengyel@yahoo.com

P. Nowicki
Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland
e-mail: piotr.nowicki@uj.edu.pl

R. Julliard
Département Écologie et Gestion de la Biodiversité, UMR 7179, Muséum National d'Histoire
Naturelle, 55 Rue Buffon, 75005 Paris, France
e-mail: julliard@mnhn.fr

J. Clobert · D. S. Schmeller
Station d'Ecologie Expérimentale du CNRS à Moulis, USR 2936, 09200 Saint Girons, France
e-mail: jean.clobert@EcoEx-moulis.cnrs.fr

D. S. Schmeller
e-mail: dirk.schmeller@EcoEx-moulis.cnrs.fr

enables to describe complex processes of biodiversity dynamics, e.g. through relating species traits to the impacts of environmental changes. Second, we review some methods to overcome differences in designs among monitoring schemes, such as site selection, post-stratification and measurement error. Finally, we point out some commonly used statistical methods that are at hand for combining data or parameter estimates. We especially emphasize the possible levels of data integration (raw data, parameter estimates, or effect size estimates), and the largely under-exploited potential of meta-analysis methods and weighted analyses. This contribution aims to bolster the practice and use of integration of ongoing monitoring initiatives for biodiversity assessment.

Introduction

Facing the scientific consensus about the general decline of biodiversity (Balmford et al. 2003, 2005a, b; European Environment Agency 2007), policy makers have launched several international initiatives with the goal to halt, or at least to decrease, the pace of biodiversity loss due to human activities (Balmford et al. 2005b; Green et al. 2005; Henle and Schmeller submitted—this volume). To assess success and progress of these initiatives requires monitoring, i.e. the quantification of the composition and dynamics of biodiversity at large spatial and temporal scales (Balmford et al. 2005b). Several syntheses provide helpful guidance for planning and implementing biodiversity monitoring (e.g. Elzinga et al. 2001; Yoccoz et al. 2001; Parr et al. 2002; Margurran 2004; Buckland et al. 2005; Green et al. 2005; Nichols and Williams 2006; Teder et al. 2007; European Environment Agency 2007). On this basis, information on the states and trends of single species or single communities are widely available. However these separate, spatially restricted, single trajectories are not directly indicative of general trends of biodiversity (Balmford et al. 2003; Mace 2005; Pereira and Cooper 2006). For this conversion, there is a need to integrate single trajectories into indicators of biodiversity components over large spatial

T. Čelik · V. Babij
Jovan Hadži Institute of Biology, Scientific Research Centre of the Slovenian Academy of Science and Arts, Novi trg 2, P. O. Box 306, Ljubljana 1001, Slovenia
e-mail: tcelik@zrc-sazu.si

V. Babij
e-mail: valerija@zrc-sazu.si

B. Gruber · D. S. Schmeller
Department of Conservation Biology & Department of Computational Landscape Ecology,
UFZ – Helmholtz Center for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
e-mail: bernd.gruber@ufz.de

K. Henle
Department of Conservation Biology, UFZ – Helmholtz Center for Environmental Research,
Permoserstr. 15, 04318 Leipzig, Germany
e-mail: klaus.henle@ufz.de

and temporal scales (Balmford et al. 2003; Buckland et al. 2005; Green et al. 2005; European Environment Agency 2007). The integration of biodiversity monitoring is thus an essential step in the progress towards a unified, appropriately scaled, adaptive management of biodiversity (Parr et al. 2002; Nichols and Williams 2006).

Increased integration should benefit all parties interested in biodiversity monitoring. Researchers analyzing biodiversity changes access and understand increased levels of complexity. Their conclusions, and therefore, the summary statistics they produce for environmental policy making, have an increased predictive power and an increased range of inference (e.g. Nichols and Williams 2006). Policy makers benefit from more general and more robust recommendations, that apply at more relevant, extended geographical and temporal scales (e.g. Côté et al. 2005; Balmford et al. 2005b; Donald et al. 2007; European Environment Agency 2007). Integration also provides environmental managers with assessments of the general impact of their management actions (Caughley and Gunn, 1996; Sutherland et al. 2004; e.g. Devictor et al. 2007b; Stoner et al. 2007). Finally, individuals and organizations engaged in monitoring benefit from an increased awareness about, and legitimacy of, their activity with a better recognition of their role as major data providers for biodiversity assessment (e.g. Pereira and Cooper 2006; Danielsen et al. 2005; Schmeller et al. submitted). Overall, although the different parties may not be interested in the same end-products, the monitoring integration process benefits each of them. The exemplary case of bird monitoring integration illustrates how one integration initiative benefit all interested parties. Researchers gained a better comprehension of determinants of population and community trends (e.g. DeSante et al. 1999; Thomas et al. 2004; Devictor et al. 2007a), which yielded robust, easily understandable indicators of the impact of several human activities on biodiversity (Gregory et al. 2005; European Environment Agency 2007; e.g. fragmentation: Devictor et al. 2008; European Common Agricultural Policy: Donald et al. 2006; international conservation policies: Donald et al. 2007). The same monitoring data were also used for assessing the impact of management activities (e.g. Devictor et al. 2007b). In addition, the monitoring activities constituted formidable opportunities to increase the awareness of the civil society about biodiversity changes (Couvet et al. 2008; Bell et al. in press). To make the present paper useful for all these parties, we focussed on two common issues: 'why' and 'how' biodiversity information should be integrated.

To achieve integration, top-down and bottom-up approaches can be considered. A top-down approach is based on highly standardized, international monitoring networks. This approach can be illustrated by some large-scale, integrated environmental monitoring programs (Olsen et al. 1999; Parr et al. 2002). Benefits of such highly coordinated networks have been outlined in recent reviews (Balmford et al. 2003; Pereira and Cooper 2006; Teder et al. 2007). We emphasize two benefits here: top-down projects are one solution to fill gaps in data availability for important but under-monitored taxa, regions and/or ecosystems (Green et al. 2005); and top-down projects may also be suitable for optimizing sampling effort (Nichols and Williams 2006).

However, top-down, global networks do not exist, and they are still to be designed and implemented, what poses formidable logistic, administrative, financial and governance challenges. Hence, bottom-up approaches, such as combining available ongoing monitoring schemes, are the only realistic option to assess the global state and trend of biodiversity now and in the coming years (Balmford et al. 2005a, b; Danielsen et al. 2005). Not surprisingly, the few existing national or international biodiversity monitoring programs rely on bottom-up integration of ongoing monitoring schemes. This is also the strategy chosen by several research groups that are engaged in the production of biodiversity indicators (Balmford et al. 2005a; Olsen et al. 1999; European Environment Agency 2007; van Swaay et al. in press).

Likewise, the reporting phase of the EU Habitats Directive (Article 17, Council of the European Communities 1992) is a top-down, legal approach but the reporting itself is a bottom-up integration of ongoing monitoring in the Member States. The scientific value and political usefulness of this integration process will greatly depend on the understanding of the potentials and limitations of integration of similar and dissimilar monitoring schemes.

Despite its importance, integration of information across existing biodiversity monitoring schemes is still very poorly developed (Parr et al. 2002; Lengyel et al. in press a). According to the EuMon databases on monitoring practices in Europe (EuMon consortium 2006; Henle and Schmeller submitted—this volume; see also Lengyel et al. in press a; Schmeller et al. submitted), only 23 out of 547 monitoring schemes assemble data at an international or EU level. We suspect that the lack of international funding for species monitoring (e.g. EU funds are the main support for only 2.3% of schemes in Europe; EuMon consortium 2006), the reluctance of institutions to share data, and the diversity of approaches participates in discouraging large-scale integration of monitoring output. In addition, biodiversity monitoring schemes were launched for very different objectives, and with restricted geographical scopes, likely due to the lack of a unified and international vision of environmental policy in the past. Regardless of their differences, most biodiversity monitoring schemes contain a common core framework: they collect measures of biodiversity components for defined units of space and time (Buckland et al. 2005; Teder et al. 2007). This core framework can and should be the basis for integration, using meta-analysis tools designed for this purpose (Côté et al. 2005). Combining monitoring output across initiatives may compensate for the three main weaknesses of ongoing biodiversity monitoring (Mace 2005; Pereira and Cooper 2006), which are: (1) fragmentary biological and spatial coverage, (2) no direct compatibility of data sets among initiatives, and (3) insufficient integration of biodiversity monitoring.

When integrating monitoring output from different schemes, two sources of information are sought for: similarities and complementarities among schemes. If different taxa, countries, or habitats exhibit a similar response to the same environmental change, then similarity among schemes indicates that researchers can make strong inferences on biodiversity state and trend. The common message is also clear what satisfies the need of simplicity for policy making. To the contrary, if biodiversity responses differ in intensity or in direction across, schemes, taxa, or habitats, the different schemes carry complementary information. By the identification of these major differences, researchers gain access to a higher level of understanding of the processes responsible for the changes. This also means that policy makers are informed of important determinants of biodiversity change that they may not want to ignore for their policies to be effective and efficient. From a statistical perspective, similarities are additive effects that explain an important part of the total variation. These additive effects can be the effects of time or of habitats. Complementarities are to be considered when an important part of the total variation is explained by interaction terms between additive effects. For instance, if an important part of the variation is explained by the interaction between the effects of time and habitats, that means that biodiversity changed differently in the different habitat. Actually, that is the message brought to policy makers by researchers working with the European Bird Indicator: over all bird species, abundance tended to be stable through time, but temporal trends strongly differed across habitats, with alarming declines in farmlands (Gregory et al. 2005). Last but not least, complementarity is also sought for when combining schemes that document different processes for a same biological component. Distribution changes can be extracted from information-poor but cheap data, such as presence-absence. But combination with information-rich and expensive data, such as demographic studies, is needed to identify the processes responsible for these distribution changes.

In this paper, we review and illustrate the information gained by biodiversity monitoring integration, and corresponding statistical tools. We also identify practical issues to be considered when combining existing monitoring schemes. With this contribution, we hope to help the practice of biodiversity monitoring and promote data integration. The better our capacity to integrate monitoring output, the better the assessment of states and trends of biodiversity, and thus the better the basis for conservation policies. We focus primarily on species monitoring and general methodological issues, whereas integration of habitat monitoring is developed by Lengyel et al. (in press b). The present paper comprises three sections addressing (i) the benefits of integrating information among monitoring schemes, (ii) the integration of monitoring schemes with different sampling designs, and (iii) commonly used statistical methods for integration of monitoring data. The latter section is presented to make clear what tools are at hand to implement the suggested integration pathways.

The benefits of integration: improving biodiversity coverage

The most obvious benefit of integrating existing information from separate monitoring schemes is an increase in the number of biodiversity components that are under survey. Coverage increase can progress along each of four dimensions: (i) the ability of monitoring to detect trends (statistical power), (ii) the biological components and mechanisms determining the states and trends of biodiversity, (iii) space and (iv) time.

Precision of estimates and statistical power

Precision of an estimate depends on the sample size (the number of sampling units available for estimation) and the natural variation of the measured parameter in time and space (i.e. variation within or among years, variation among populations or among habitats). The ability of monitoring to detect a change as significant (e.g. statistical power for the test of the effect of time) is a function of the precision of the estimate. Hence, to increase the chances of detecting significant sources of variation in biodiversity, one wants to maximize sample size. Thus, combining information from different monitoring schemes is a straightforward way to increase sample size, precision of estimates, and, eventually, statistical power, without increasing sampling effort per scheme. For instance, Hochachka et al. (2000) compared count data collected opportunistically by several observers with precise estimates of population size. They concluded that variability in population size was correctly retrieved with opportunistic data, and that the increase in sample size due to the use of all available data outweighed the cost of high among-observer variation. We can expect that the same conclusions would hold when combining data among monitoring schemes.

Biological coverage

In this section, we consider integration at increasing levels of biological data heterogeneity, starting from combining similar data on similar species and ending with combining output of species and habitat monitoring schemes.

A single biological process for a single (set of) species

The first, intuitive avenue for integration is to combine monitoring schemes that document the same biological process (e.g. survival rate, population size) for the same (set of)
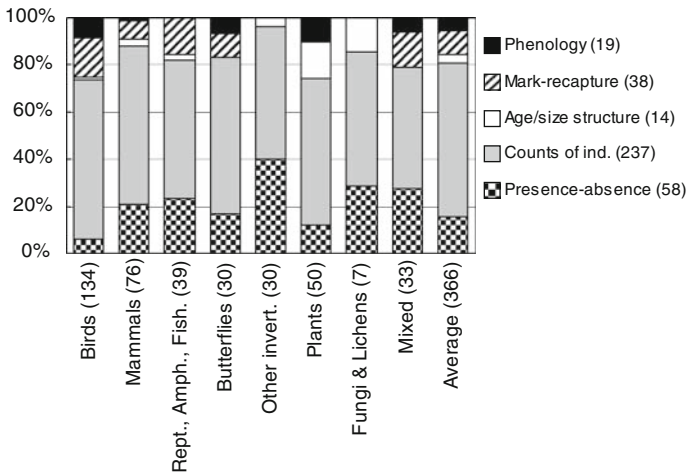
Fig. 1 Proportions (and numbers) of monitoring schemes collecting a given data type per taxonomic group in Europe (species database, EuMon consortium 2006). Scheme coordinators were asked to indicate only the main data type per scheme. The group 'Mixed' contains schemes that monitor more than one taxonomic group; 'Counts of individuals' includes plant densities. This figure gives a quantitative overview of the potential for integration of monitoring schemes collecting similar and complementary information within- and across-taxonomic groups

species. The benefits are increased precision (cf. previous section) but also increased generality of monitoring conclusions. Such an approach yielded, for instance, the first global evaluation of amphibian population trends, combining data from 936 populations of 157 species (Houlahan et al. 2001). At least within Europe, there is a large potential for such integration, with many schemes collecting the same data types on the same taxonomic groups (Fig. 1) or the same habitats (Lengyel et al. in press b).

The biological processes documented by monitoring schemes are largely determined by the type of the data collected. In species monitoring, four main data types are used (Fig. 1): presence/absence, counts (of individuals or species, including vegetation coverage), individual follow-up (capture-mark-recapture data), and measures of individual traits (e.g. age, size, state of individuals). Even if different monitoring schemes collect different data types, they still can be reduced to their smallest common denominator to document the same biological process, e.g. presence-absence data for geographic distribution or counts for population trend. Integration in this way extracts the information common to all monitoring schemes. Virtually all existing schemes could be combined in this way (Fig. 1). The browser available at the EuMon consortium (2006)'s website allows the identification of schemes that could be combined per species, taxonomic group, or habitat in Europe.

Different biological processes for a single (set of) species

When different monitoring schemes collect different data types for a same (set of) species, they contain information on different biological processes. A goal of integration is to structure the complementarity among these biological processes to gain a more subtle and operational characterization of biodiversity change. Consider the case of changes in population size. Most data can be used to analyze and estimate trends in population size (Strayer 1999; Pollock 2006). However, only individual follow-up or age/size–structure

dynamics contain the necessary information to explain the observed population trends in terms of demographic mechanisms. Such an integration approach was adopted to identify the demographic processes (survival versus recruitment) driving population trends in four common birds in France (Julliard 2004; see also DeSante et al. 1999). The identification of the driving process in turn is essential for the development of targeted management (e.g. Pradel and Henry 2007). Statistical methods were recently developed to combine capture-mark-recapture and count data to improve the analysis of population trends (e.g. Besbeas and Freeman 2006; Gauthier et al. 2007; Pradel and Henry 2007). In Europe, a large number of schemes collect time series of counts and capture-mark-recapture data for many bird species, as well as for several large mammals, reptiles, amphibians, and fishes (Fig. 1), underlining the great potential for more integration of monitoring data. Again, the online browser of the EuMon's database (EuMon consortium 2006) can be used to identify schemes that collect complementary information at the species level.

If the biological target of monitoring exhibits phenological fluctuations, the within-year temporal design should allow accounting for phenology in the analysis of among-year changes (e.g. Menzel et al. 2006; Roy et al. 2007). At least 43% of European species monitoring schemes have intra-annual replication in data collection (i.e. two or more monitoring sessions per year; EuMon consortium 2006) and, therefore, could document phenology (e.g. arrival of migrants, flowering). The combination of such schemes would greatly increase our ability to quantify the assumed effects of climate change on species and biological communities. Appropriately accounting for changes in phenology is an important step for explaining and quantifying changes in distribution and population size of mobile or multi-voltine organisms (cf. plants, invertebrates). To be suitable for this purpose, the temporal design must secure that within-year distribution of replicated samples encompasses the part of the year when phenological changes occur. For instance, in the case of monitoring settlement of migrating animals on their breeding grounds, monitoring visits should be distributed in time so that, over all monitored sites, data are collected before, during and after arrival of migrants. van Strien et al. (2008) give useful guidance on how to design an analysis of phenological change with monitoring data. For schemes without within-year replication, the problem is that true among-year variation of the measure (e.g. abundance) may be confounded by annual changes in phenology. In the context of integration, data from 'phenology' schemes should be used to assess the robustness of temporal trends derived from schemes that cannot account for phenological variation. Methods should be developed to allow the joint analysis of schemes with and without within-year replicated sampling.

A single biological process for different (sets of) species and taxonomic groups

Multi-species trends are usually obtained by combining single-species trends across species (e.g. Gregory et al. 2005). The resulting estimates has a broader biodiversity coverage than single-species approaches (as well as some other advantages; Balmford et al. 2005b; Mace 2005). The simplest method for combination is to consider that all species are equal, regardless of their characteristics (e.g. ecological function, life history traits), and to compute the mean.

When among-species heterogeneity is high, across-species integration can provided valuable information due to the complementarity among species traits: which set of species tells us what (Kati et al. 2004; Thomas et al. 2004)? Biodiversity indicators focussing on habitats (e.g. Gregory et al. 2005; van Swaay et al. in press), habitat specialization (e.g. Henle et al. 2004, Devictor et al. 2007a), functional traits (e.g. Elzinga et al. 2001;

Statzner et al. 2001; Dziock et al. 2006), trophic levels (e.g. Pauly and Watson 2005), or any other species traits (e.g. Dolédec et al. 1999; Henle et al. 2004; Pavoine et al. 2005) are designed to gain information from these differences among species. The main limitation when combining data from different species or taxonomic groups is the lack of scientific/ theoretical knowledge to interpret the resulting composite biodiversity indices (Buckland et al. 2005; Green et al. 2005; Nichols and Williams 2006). For instance, the Living Planet Index (Loh et al. 2005) combines all available data into a single index, whatever realms, habitats and life-history traits of the groups are. The index is thus easy to define, but its biological meaning can be questioned. Another approach is to rely on a theoretical framework that formally links different taxonomic groups (e.g. Marine Trophic Index relying on explicit trophic networks; Pauly and Watson 2005). For terrestrial ecosystems, such a theory-based integration framework is still largely lacking (but see Pettorelli et al. 2005).

Integrating monitoring according to causes of change

An intuitive goal when combining monitoring datasets is to search for a common response across species or taxonomic groups to a given cause of environmental change (e.g. pollution, land-use, climate change, invasive species, Table 1; Elzinga et al. 2001; Henle et al. 2004, 2008; Balmford et al. 2005a; Gregory et al. 2005; but see limits of the approach in Nichols and Williams 2006). There is a great potential for integration per cause of change since 85% of species (Table 1) and habitat (Lengyel et al. in press a) monitoring schemes in Europe claim to document at least one possible cause.

The first benefit of integration per cause of change is to increase the robustness of conclusions on the causes of biodiversity change, and their respective intensity. Meta-analysis tools are specifically developed to derive such conclusions about the average effect of, e.g. climate change or habitat fragmentation, from independent, small-scale correlative tests with monitoring data (Côté et al. 2005).

The second benefit of integration per cause of change comes from testing for differences among species, and among taxonomic groups, in their response to one same cause of

Table 1 Proportions of species monitoring schemes documenting a given cause of change per taxonomic group in Europe (species database, EuMon consortium 2006)

Taxonomic group	Land use	Fragmentation	Climate change	Pollution	Invasive species	Nb. schemes
Birds	0.79	0.28	0.48	0.31	0.19	95
Mammals	0.83	0.58	0.15	0.13	0.10	48
Reptiles, amphibians, & fishes	0.88	0.67	0.33	0.55	0.55	33
Butterflies	0.82	0.57	0.57	0.14	0.14	28
Other invertebrates	0.78	0.41	0.41	0.52	0.33	27
Plants	0.82	0.48	0.27	0.34	0.39	44
Fungi & lichens	0.57	0.29	0.57	0.86	0.14	7
Several taxonomic groups	0.89	0.70	0.33	0.26	0.33	27
Nb. schemes	228	146	107	92	72	309

These figures give a quantitative overview of the potential for within- and among-taxonomic group integration per cause of change. Scheme coordinators could declare more than one cause of change per scheme

change. Understanding these differences should increase the robustness of biodiversity assessment conclusions, and the adequacy of corresponding management policies. Monitoring and among species differences have been successfully used to develop predictors of birds sensitivity to habitat loss and fragmentation (Henle et al. 2004). Another example is the combination of monitoring time series on butterflies and birds. Butterflies are shorter-lived and more specialized than birds. They are therefore expected to react more rapidly and at smaller spatial scales, whereas the longer-lived and less specialized birds would react more smoothly and at a broader spatial scale (Thomas 1995, 2005). Hence, butterflies would document finer-grained changes, whereas birds would integrate changes over larger spatial and temporal scales. An integrated monitoring of birds and butterflies would thus provide a complementary understanding of biodiversity changes (e.g. Thomas et al. 2004; see Kati et al. 2004 for a similar recommendation for conservation purposes). In an extreme case, species from different taxonomic groups could even behave more similarly than species from the same taxa. For instance, generalist birds and butterflies may respond similarly to some environmental changes, whereas specialists could exhibit different responses (e.g. review by Henle et al. 2004 for fragmentation response of generalists and specialists).

The third benefit of integration is to challenge observational, correlative results about causes of change from surveillance monitoring (sensu Nichols and Williams 2006) with results from monitoring schemes using an appropriate experimental design. Theoretically, only monitoring schemes with well-planned, experimental designs can demonstrate that a given cause of change actually explains the temporal or spatial trends observed (Nichols and Williams 2006). However, surveillance monitoring data is the only source of material available for evaluating large-scale changes in biodiversity, identifying putative causes of change, and measuring the intensity of these changes at the relevant spatial scale through post-hoc correlative evidence. In Europe, a large part of schemes lack any experimental design (72% of species schemes, 48% of habitat schemes; EuMon consortium 2006). When different areas are monitored, with some areas affected and others not affected by the change, correlative comparisons can come close to an experimental design (e.g. different forms of land use; Henle 2005). Combining monitoring schemes with and without experimental designs would benefit each type of monitoring: experiment-based monitoring would gain in spatial and temporal range of inference (external validity), whereas correlative-based monitoring would gain in inferential power about the role of underlying causes of change.

Integrating species and habitat monitoring

Monitoring of biodiversity is needed both at the level of species and habitats (Balmford et al. 2003, 2005a; Lengyel et al. in press a; habitat being also called environment, Teder et al. 2007). Habitat monitoring is the monitoring of habitat characteristics, with habitats defined as distinguishable and repeatable assemblages of species (see Lengyel et al. in press a). Thus, an integration of species monitoring and habitat monitoring has a high potential to provide a better insight in biodiversity changes. On the one hand, the states and trends of habitats provide information on the potential states and trends of their constitutive species. For example, if the coverage of a habitat is reduced by 10% per year, species depending on this habitat may be expected to also decrease by at least 10% per year. On the other hand, since habitats are most often defined as assemblages of plant species, species monitoring will be informative on the states and trends of habitats. Evaluations of the

number of species lost through deforestation are based on this rationale (Hughes et al. 1997). Obviously, these approaches are rather crude, and they can be refined with eco-logical data if available (e.g. adjustment for density–area relationship, transitory increase of density in remaining habitat fragments, habitat specialization per species). Since several environmental policies rely on the assumed tight relationship between species and habitats, and use them for assessing their conservation status (e.g. Habitats Directives, Council of the European Communities 1992), integration of species and habitat monitoring schemes are essential for the evaluation of these policies. Actually, this integration effort has been requested by the European Commission to the Member States for the production of national reports on states and trends of Habitats Directive species (Article 17, Council of the European Communities 1992). As an additional example of the benefits of integrating habitat and species monitoring, Devictor et al. (2008) combined a standardized, European-scale geo-referenced database of habitats (CORINE Landcover) and breeding bird survey data, providing the first large-scale empirical evidence of the positive relationship between landscape disturbance and homogenization of bird communities.

Spatial coverage

Integration of existing monitoring schemes through space has three main benefits: (i) it increases spatial coverage without increasing sampling effort, (ii) it secures that spatial variation in biodiversity components can be accounted for, and (iii) it facilitates directing new monitoring schemes to areas not yet covered.

Monitoring schemes often have a moderate spatial coverage. In Europe, 52% of species schemes (Fig. 2) and 55% of habitat schemes (Lengyel et al. in press a) are restricted to a small area or a region within a country. The federation of local, regional and national monitoring schemes is an efficient way of increasing spatial coverage without increasing monitoring effort. The EU Bird Indicators are based on such an integration of national
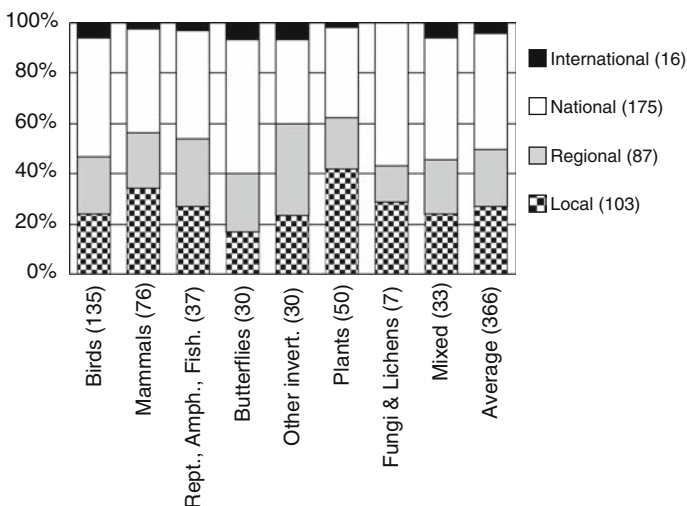

Fig. 2 Spatial coverage of monitoring schemes per taxonomic group in Europe (species database, EuMon consortium 2006). This figure gives a quantitative overview of the potential for integration of monitoring schemes across spatial scales (cf. assembling local and regional schemes into national and international monitoring)

monitoring schemes (Gregory et al. 2005; European Environment Agency 2007). The EU Butterfly Indicators are being developed with the same integrated structure, also including regional schemes when national monitoring is lacking (European Environment Agency 2007; van Swaay et al. in press).

State and trends of biodiversity vary through space. Thus, extrapolation of measures from one localized monitoring scheme to a wider area may often not be warranted. A better practice is to rely on integration of monitoring output through interpolation across different monitored regions. This rationale is included in the construction of most Red Lists of species. A great advantage of spatial interpolation from existing monitoring schemes is to allow biodiversity estimation even for areas not monitored (e.g. Jiguet et al. 2005). The predicted values for these areas come along with estimates of their precision, i.e. of their reliability. Local environmental authorities then benefit of robust estimates of states and trends of biodiversity for all the areas under their responsibility, even those not monitored. However, beyond a certain distance, data from one site are useless to predict biodiversity at another site. This distance, i.e. the limit between interpolation and extrapolation, is the maximum distance at which the measured biodiversity component is spatially autocorrelated. In this situation, new monitoring sites (or schemes) are needed to fill monitoring gaps.

Even in the presence of spatial variation of biodiversity, policy makers may need a single indicator value for large regions that may contain several monitoring schemes. If sampling design and weighting issues are appropriately handled, the estimate of the global indicator should provide an unbiased assessment of biodiversity. In addition to the global picture, decision makers may need a finer-grained indicator to adjust local management recommendations to local conditions. In this case, the average indicator should be spatially disaggregated to identify areas of homogenous trend within the area of interest. From a statistical perspective, areas with contrasted temporal trends will be identified by significant interactions between the effects of time and of space (e.g. time * sites, or time * regions, or time * schemes; Fig. 3). This concept can be illustrated with climate warming in temperate regions. Spring arousal occurred earlier in recent warmer years, but this effect was stronger at northern than at southern latitudes (Menzel et al. 2006). Spatial disaggregation may also be considered at the habitat level. The European Bird Indicator can be computed over all species, but distinguishing trends per major habitat types revealed that the major concern was for farmland and grassland species (Gregory et al. 2005).

Spatial integration also stimulates the launching of new monitoring schemes in regions or countries that are not covered so far. Such new monitoring schemes have the possibility of benefiting from the experience of network partners. The federation of existing butterfly monitoring schemes had such a positive effect on the launching of new schemes (van Swaay et al. in press). From a logistical perspective, this is particularly helpful for identifying the monitoring design that makes the best compromise between local constraints and biodiversity monitoring goals (Yoccoz et al. 2003; Green et al. 2005; Schmeller et al. submitted).

Temporal coverage

Integrating different existing initiatives allows increasing temporal coverage. Similar field monitoring techniques have been used for decades. In Europe, at least 17 schemes have been running for more than 40 years, and two schemes even for more than one century
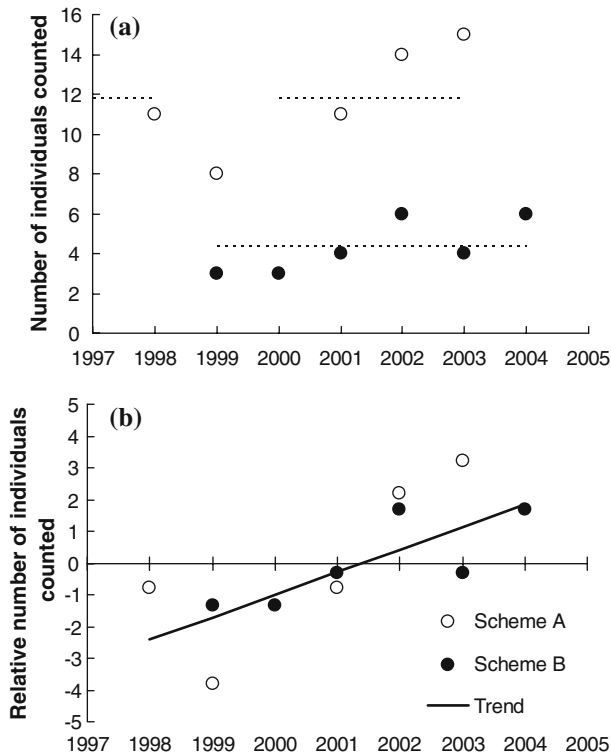
Fig. 3 Temporal trend analysis in the presence of missing data and differences of average value among schemes. (**a**) Two schemes, A (open dots) and B (filled dots), counted individuals from 1997 to 2005. To circumvent the problems of missing counts in some years (e.g. year 2000 for Scheme A), as well as the problem of systematic differences in relative abundance among schemes (cf. difference in means, indicated by dotted lines), the combined dataset can be analyzed with a log-linear regression model (i.e. generalized linear model with Poisson error structure for the dependent variable and a log link-function; Buckland et al. 2005), with the number of individuals counted as dependent variable, and the effects of year and scheme identity as additive explanatory variables. (**b**) Including a scheme effect accounts for systematic differences among schemes in relative abundance. The inclusion of the year effect is similar to averaging counts standardized for the effect of schemes per year (i.e. year-specific deviations from the scheme average). The average temporal trend in relative abundance can be estimated by including a linear effect of years in the model. Note that for the sake of clarity, the Y axes (counts) are on an arithmetic scale, whereas their natural scale should be logarithmic (Buckland et al. 2005)

(EuMon consortium 2006). Thus, by using similar monitoring data, monitoring assessment could go back far into the past by integrating data from old, abandoned monitoring schemes with ongoing and starting schemes (e.g. Loh et al. 2005).

Surveys are often not implemented with the same inter-annual frequency. Although this rises technical problems (see Section "Temporal design and missing data"), a benefit of having different time frequencies is to obtain complementary insights on the temporal patterns of the biological component of interest. Long-term monitoring with low survey frequency (i.e. wide temporal gaps) allows picking up long-term trends, while short-term monitoring with high temporal frequency (e.g. annual) allows picking up faster changes in population size. This is particularly critical when monitoring species with cyclic population dynamics (e.g. Krebs and Berteaux 2006).

Integration of monitoring schemes with different sampling designs

Setting clear goals for monitoring defines what biodiversity components should be monitored at which spatial and temporal scales (Elzinga et al. 2001; Yoccoz et al. 2001; Parr et al. 2002; Green et al. 2005; Teder et al. 2007). The choice of the sampling design then defines how samples are to be distributed in space and time to fulfil the monitoring goals. If the sampling design is not well planned, it can strongly impair the strength of the conclusions derived from monitoring data (Yoccoz et al. 2001; Balmford et al. 2003; Buckland et al. 2005). Combining information from schemes with different sampling designs is a way to partly compensate for potential defects in the design of some schemes. We consider here solutions to overcome or to make advantage of differences among schemes in three major components of sampling designs: (i) accurately accounting for spatial variation (cf. the methods to choose sites to be monitored), (ii) handling of missing data in time series, and (iii) measurement error. The interest of combining schemes with and without control samples (i.e. experimental designs) has already been addressed in the section "Integrating monitoring according to causes of change".

Spatial variation and choice of sampling sites

All monitoring schemes using site selection methods that secure an objective representation of spatial variation can be combined without any correction. This concerns the schemes where all sites are monitored (exhaustive monitoring), or where the subset of sites to be monitored is chosen randomly or systematically. However, the prevailing practice is to choose sites freely or according to expert knowledge (58% of schemes in Europe; EuMon consortium 2006). Since criteria underlying these choices are subjective and undefined, these monitored sites may provide a biased documentation of the monitored area. This problem is pointed out as one major weakness even for some widely recognized, long-term monitoring schemes (e.g. Buckland et al. 2005). In this situation, data have to be transformed a posteriori (or weighted, named post-stratification) so that the estimates and conclusions derived from the data provide a representation as unbiased as possible of the biodiversity change at the spatial and temporal scales of interest (see part on weights for "Different ranges of inference").

Stratification of sample collection is another method to optimize sampling effort according to specific monitoring goals while maintaining unbiased site selection. Stratification is similar to giving different weights at the design step. Stratification is used, for instance, when some habitats, regions, or species need to be sampled with a higher, but known and quantified, effort (e.g. Green et al. 2005; Henle et al. 2006). This is particularly the case for rare or localized species that are usually badly covered by fully systematic or random sampling designs. It is often preferred to stratify a priori the field effort among habitat types, and to monitor with a higher effort sites where the species is likely to occur. Integration of monitoring schemes with different stratification designs needs then to apply the inverse stratification when analyzing combined data. For instance, when computing the average estimate, if one habitat type was sampled twice more than others, data from this habitat type should be given a weight of 0.5, whereas others should be given a weight of 1.

Temporal design and missing data

When integrating different existing initiatives, the temporal design usually differs among monitoring schemes: their activities did not start or will not end in the same years, and

surveys are not implemented with the same inter-annual frequency (EuMon consortium 2006). A similar problem arises in 'adaptive monitoring', i.e. when new monitoring needs are identified while the monitoring is ongoing (e.g. Nichols and Williams 2006), or when defaults of the monitoring design need to be corrected for (e.g. Buckland et al. 2005). The dilemma is then whether to change the protocol, which will introduce heterogeneity in the monitoring design within the time series, or to keep using a suboptimal design but consistently through time. In Europe, 14% of species and habitat monitoring schemes declare to have implemented major modifications of their monitoring protocol after the monitoring had started (EuMon consortium 2006). The need to account for discontinuity in the time series is one of the important difficulties when integrating monitoring data (Balmford et al. 2003).

A solution to compensate for incomplete time series is to use statistical models that account for missing data (Olsen et al. 1999; Buckland et al. 2005; Gregory et al. 2005; but see Houlahan et al. 2001). Generalized linear models, with appropriate selection of data distribution, link-function and parameterization of the effects of schemes and year, intrinsically account for heterogeneity among schemes and through time (Fig. 3). For instance, for the EU Bird Indicators, counts of birds are analyzed with a log-linear model, which allows to estimate trends despite missing data (Gregory et al. 2005; van Swaay et al. in press; see Fig. 3). Note that interpolating values for missing data does not change estimates of the indicator or of the temporal trend.

When only a few different protocols are to be combined, another solution is to calibrate data among protocols from portions of the time series when two or more protocols were applied simultaneously within the same geographical area (e.g. the British Trust for Ornithology, Buckland et al. 2005; crocodile monitoring in Northern Australia, Webb et al. 1990).

Accounting for measurement error

The measurement error quantifies the range of statistical validity of the measure. The sampling design should allow taking into account this uncertainty of the measure when inferences are made from the data. When measurement error cannot be estimated in some of the datasets to be combined, a solution is to include independent estimates of this error in the statistical model for the joint analysis. Such methods are still under development (e.g. Hooten et al. 2007).

A common source of measurement error in monitoring schemes based on counting individuals (or species) is the fact that the observer cannot detect all individuals (or species) present during monitoring visits. In other words, the detection probability is usually lower than one. A specific sampling design based on repeated sampling needs to be implemented so that monitoring data can be adjusted for fluctuations in detection probability. When integrating monitoring schemes with and without detection probability design, two approaches can be followed. First, it is common practice to ignore detection probability (e.g. Rosenstock et al. 2002; EuMon consortium 2006; Henry and Jarne 2007). This practice may be reasonable if a pre-analysis showed that detection probability can be considered constant trough space and time, or that variation is random and cannot generate spurious trends. However, this may rarely be the case and the many possible sources of variation in detection probability can critically confound the conclusions of monitoring (e.g. Yoccoz et al. 2001; Buckland 2006; Henle et al. 2006; Kull et al. submitted—this volume). Second, uncertainty in the measure can be systematically quantified by additional information (e.g. extra field-work). If such post-hoc measures are not feasible (too

technical or time consuming, Danielsen et al. 2005; but see Schmeller et al. submitted), a solution is to incorporate independent estimates of the error in the joint analytical model. Estimates of detection probability can be extracted from monitoring schemes with appropriate sampling designs and incorporated in analyses of data from monitoring schemes with inappropriate sampling designs (e.g. MacKenzie et al. 2005; Hooten et al. 2007; Schaub et al. 2007). More generally, Bayesian models provide a promising analytical framework for such combinations of heterogeneous data, or the integration of extra-biological knowledge in the statistical analysis. Applications of these methods to monitoring data are under development.

Statistical methods for integration

There are two main ways to integrate information from different monitoring schemes: combining data or combining estimates. Combining raw data into a single dataset is possible when data are compatible, i.e. when they are measured in the same unit (or can be reduced to the same unit) and they quantify the same biological process (Table 2). When data types differ but still document the same biodiversity indicator, a solution is to combine estimates of the indicator across datasets. Two supplementary methodological issues are also considered hereafter. Whatever the data to be integrated, if contributions to the global indicator should not be equal among monitoring schemes, species, or regions (etc.), data or estimates need to be weighted. Finally, when different monitoring datasets document a similar biodiversity component, cross-validation could be used to assess the robustness of the conclusions.

Combining data

When the measurement unit is the same among different monitoring schemes, raw datasets can be combined easily (Table 2). For simultaneous analysis with the same parametric statistical model, data need to follow the same theoretical distribution. Then, combined data can be jointly analyzed to produce an estimate averaged across all monitoring data in regard to the parameter of interest. Summary statistics are straightforward to compute from the integrated dataset. For instance, Julliard et al. (2004a) estimated national population growth rates for bird species with a single model combining data from two separate monitoring schemes, one counting individuals detected acoustically and the other counting captured individuals. Although the numbers of individuals per sampling effort could not be compared because counting techniques were different, data still followed a similar theoretical distribution (Poisson distribution) and documented the same biological parameter, population growth rate (estimated by the slope for the linear effect of year; Fig. 3). Such approaches provide access to parameter estimates across all datasets with a single analysis, despite differences in sampling units and scales among monitoring schemes. When combining heterogeneous data, the general model may not fit satisfactorily all data (cf. overdispersion). In this case, it would not be warranted to combine all data into a single analysis. Estimates of the population growth rate should be extracted separately from each dataset and then combined with meta-analysis methods.

When the nature of the data collected differs among monitoring schemes, the simplest method for data combination is to reduce the complexity of information to the lowest common level (common denominator). For instance, if a set of individual follow-up, counts of individuals, and presence-absence data is available (Fig. 1), the lowest complexity level would be presence-absence (e.g. Roberts et al. 2007). Combining data in this

Table 2 Three levels of combining information from different monitoring schemes so that combined data can be analyzed with a single statistical model

Data to be combined across datasets	Conditions of application				Example
	Measurement unit in original datasets	Biological parameter to be analysed (i.e. Y) in original datasets	Biological effect to be tested (i.e. X) in original datasets	Statistical method	
Y i.e. raw data for dependent variable in the statistical model	Same	Same	Can differ	Classical	Population growth rate documented with counts of individuals
$\hat{Y}$ i.e. estimates of Y	Can differ	Same	Can differ	Classical + weights for precisions of estimates or Meta-analysis methods	Population growth rate documented with counts of individuals and presence-absence data
$\hat{z}$ i.e. estimate of effect size, the standardized estimate of the intensity of an effect of X (independent variable) on Y	Can differ	Can differ	At least one X needs to be the same	Meta-analysis methods or Classical + weights for precisions of estimates	Linear effect of time (temporal trend) on population growth rate documented with counts of individuals and presence-absence data

Corresponding conditions of application and examples are given

way is rather straightforward. However, much of the original information and precision contained in the data is lost (Strayer 1999). In this case, combining estimates instead of the raw data would make a more optimal use of the information to be integrated (see following section).

If data heterogeneity is so high that no common quantitative currency can be defined, better than nothing is to synthesize the sparse, available information on states and trends into standardized ratings. This is how the IUCN evaluates extinction risk status with the help of standardized criteria assessed by independent experts (Miller et al. 2007). These criteria are then used as raw data for biodiversity assessment (e.g. Butchart et al. 2005).

Combining estimates and meta-analysis

When different data types are collected, parameter estimates rather than original data can be integrated. Estimates become the dependent variable in the joint analysis. The difference between analyzing raw data or estimates is that error of the measurement is usually ignored for raw data. Raw data are analyzed as if they were known without error (i.e. perfect measurement). To the contrary, measurement error for estimates is known; it is measured by the standard error. Then, a proper analysis using estimates as dependent variables should simultaneously account for estimates of the mean and of the standard error. Estimates to be integrated can characterize state or trend of a biodiversity component ($\hat{Y}$ in Table 2), or the response of this state or trend to an external factor ($\hat{z}$ in Table 2).

Combining estimates of dependent variables

When measurement units differ, information from each monitoring scheme can be summarized as the estimate of a single biological parameter for each separate data set. Then, integration is achieved by analyzing these estimates with a single statistical model. For instance, trends in population size can be estimated both from counts of individuals per unit of time or with presence-absence data (e.g. Strayer 1999; Pollock 2006). Both estimates can be combined to obtain an integrated, average estimate of the population growth rate. For example Julliard et al. (2004a) estimated population growth rates from different data types (point counts versus numbers of individuals captured) for a large set of species. Then, they tested with a single ANOVA model whether among-species variation in population growth rate could be explained by species traits, while accounting for differences of estimate precision among species.

When producing summary statistics from combined estimates, the recommended method is the geometric mean (instead of the arithmetic mean), i.e. averaging on a log-scale and exponentiation of the average (Buckland et al. 2005). The formula for computing standard errors for geometric means is provided in Appendix A of Gregory et al. (2005). An interesting property of the geometric mean is that its temporal trend is invariant with respect to the weights attributed to each monitoring scheme (or species; Buckland et al. 2005).

An illustrative study is the estimation of the average trend of breeding bird populations per major habitat in Europe (Gregory et al. 2005). In 2000, up to 18 EU countries maintained a national breeding bird survey and counted individuals per species but with different methods. Thus, data could not be combined into a single dataset from which trends could be estimated. The integration procedure comprised three steps. First, each country produced national estimates of population growth rate per year for each species. Second, these estimates were combined with independent estimates of national population

sizes to produce yearly estimates of the European population size, allowing the computation of population growth rates at the European level for each species, with missing data accounted for by interpolation. Finally, species were attributed to broad habitat categories based on expert knowledge, and estimates of population size changes were averaged across species by the geometric mean to produce estimates of trend per habitat in Europe. For several other groups, e.g. butterflies, but also raptors, large mammals, bats, beetles, a number of schemes suitable for similar integrative analyses exist (EuMon consortium 2006), the integration of which would considerably improve our understanding of biodiversity conservation needs.

Combining estimates of the effect of explanatory variables

The idea behind meta-analysis is that results of independent studies are treated as input units for the analysis of a general pattern (Gurevitch et al. 2001). Such an approach allows combining information coming from various monitoring schemes regardless of the differences in their sampling designs, objects monitored, data characteristics, and to some extent even statistical methods applied. If statistical analyses applied on each separate dataset included the same effect (the same independent variable), then the average effect can be computed to infer the average pattern across all datasets (Table 2). Meta-analysis methods use the effect-size concept to integrate estimates of effects across analyses (e.g. Hedges and Olkin 1985; Cooper and Hedges 1994; Osenberg et al. 1999; Gurevitch et al. 2001). The effect size is a standardized estimate of the magnitude of the effect of an explanatory variable. A common metric of effect size z is the estimate of the slope for the explanatory variable, divided by the standard error of the slope estimate (see Osenberg et al. 1999 for other metrics). Effect size is computed independently for each monitoring. The mean effect size is then computed by summing effect size estimates from all monitoring schemes and dividing this sum by the square-root of the number of degrees of freedom (i.e. number of monitoring schemes—1). If the supposed cause of change has an effect, the mean effect size will depart from 0. Whatever the magnitude of the true effect in each monitoring scheme, the expectation of the test statistic will be negative if there is a general negative effect, or positive if there is a general positive effect. The statistical power of the resulting meta-analysis will depend on the magnitude and precision of the effects in the various monitoring, but power should be reasonable in the case of small to moderate effects in all monitoring schemes. A meta-analysis has a good probability of detecting the effect of the cause of change over all observations, which is not the case for separate tests on each single dataset. Another important advantage of meta-analysis is the possibility to identify different patterns of response across monitoring schemes with tests of homogeneity of effect size. This statistical framework allows estimating average trends across monitoring schemes, as well as discriminating sets of regions with contrasted trends.

When only qualitative information is available for the tested effect or cause of change (cf. significant positive, non-significant, significant negative; e.g. Parmesan and Yohe 2003), non-parametric tests can be used to identify whether the proposed cause of change has, on average, a significant effect over all monitoring schemes (Cooper and Hedges 1994). As for any analytical method, meta-analysis cannot compensate for all defaults of the data. For instance, they will not compensate for biases in data availability (cf. sampling or publication biases, non-independence of data-points; e.g. Møller and Jennions 2001; Côté et al. 2005). As for the design of monitoring schemes, the design of meta-analyses has to be planned carefully to secure accurate contributions to biodiversity assessment.

Given all these methodological possibilities, and their suitability for monitoring integration, it is surprising that meta-analysis methods remained so rarely used for biodiversity assessment from monitoring data (Sutherland et al. 2004; Balmford et al. 2005a). Two explanatory variables would be particularly good candidates for meta-analysis: the effects of time, and of given causes of change. Nearly all monitoring schemes aim at testing for temporal trends in the measured biodiversity component. The effect size for time would be the very first candidate for the application of meta-analyses in the context of biodiversity assessment. Two temporal effects can be analyzed: the unconstrained effect of years, or the linear effect of years, i.e. the linear trend throughout the time series (Fig. 3b). Estimates of the slopes for the linear effect of time can be combined among analyses to obtain the averaged, global population growth rate, and to test for a global temporal trend. A good example of meta-analysis of the effect of time is the assessment of the world-wide trend in coral reef coverage (see Côté et al. 2005). Note that in the case of the analysis of linear effect of time, an analysis of time effect estimates, weighted for estimate precisions, would be similar to a meta-analysis of effect size estimates (Table 2).

The second important application of meta-analysis is the analysis of causes of environmental change across monitoring datasets (Côté et al. 2005). Classical components of the global change (Table 1) are documented by several tens of monitoring schemes in the same and/or complementary taxonomic groups. For instance, coordinators of 27 bird monitoring schemes, 28 mammal schemes and 16 butterfly schemes considered that they could assess the effect of fragmentation with their monitoring data. Hence, there is a large potential for coordinated analysis of independent datasets and meta-analysis of size effects for widespread causes of change. Another good candidate for meta-analysis of monitoring data is the study of climate change. The effect of climate warming is often tested for with the effect of yearly temperatures. Since many different monitoring schemes use this same explanatory variable in their analysis of time series, estimates of the effect size for the yearly temperature can be combined across monitoring schemes. Combining all these estimates into a single meta-analysis provides a robust, general test for the response to yearly temperature, as an indicator of climate warming, across all datasets (Menzel et al. 2006). Other examples of explanatory variables used in meta-analysis of monitoring data are common human-induced or natural disturbances (e.g. Pons et al. 2003; Côté et al. 2005).

Compensating for differences in biodiversity coverage and monitoring designs: the use of weights

When combining information from different monitoring schemes, the issue of differences in biodiversity coverage and monitoring designs among schemes arises (Balmford et al. 2005a; Buckland et al. 2005). Are all species, countries or estimates equally indicative of biodiversity state or change? Should some have higher contributions to the global biodiversity index than others? The monitoring goals (should) answer these questions. Once priorities are set, a common practice to implement these choices is to apply weights to the data or estimates prior to statistical testing or averaging. Hereafter, we present some standard weights used when combining biodiversity monitoring data. Weights have two natures: weights that formally adjust for differences in precision, and weights that are used to compensate for biased measures of the parameter of interest (e.g. over-/under-sampling) or to intentionally bias the contribution of different data to an indicator (e.g. differences in contribution among species, taxa, habitats). This second type of weights is a pragmatic

response to an important need, but they have no methodological background. Standard methods to simultaneously use these two types of weights (precision and bias) in a single analysis remain to be proposed.

Different precisions of estimates

If estimates have different precisions (i.e. standard errors, SE), the weight to be used should be the inverse of the squared standard error for each estimate (i.e. $1/(SE)^2$; e.g. Julliard et al. 2004b). In this way, when testing for a temporal trend with estimates from different monitoring schemes, differences in precision of trend estimates per scheme are accounted for. A moderate but precisely estimated decline will contribute more to the global estimate and to the test of a temporal trend than a very steep but largely imprecise decline. If standard error estimates are not available, surrogates of precision, such as the squared number of monitored sites, or the monitored area per scheme (Côté et al. 2005), may be used as weights. Note that the size effect statistic z used in meta-analysis already accounts for estimate precision.

Different geographical contributions

If population sizes differ across monitored geographical regions to be combined, a suitable weight would be the proportion of the total population size held per region (e.g. Côté et al. 2005; van Swaay et al. in press). For instance, for the European Bird Indicators (Gregory et al. 2005), weights are the percentage of the total European population size held per country per species.

Different ranges of inference

For assessing states and trends for a species or a taxonomic group, it is important that all inhabited habitats and/or biogeographic regions are accounted for. This is typically achieved during the planning phase by selecting monitoring sites that provide an unbiased coverage of habitat composition. However, if no sampling design is used, it is likely that habitats will not be equally represented. To obtain an unbiased, average trend across all regions, weights need to be applied to the data so that each habitat is represented according to its actual surface area. For example, when producing national trends for butterfly population sizes in the Netherlands (van Swaay et al. 2002), indices of population size per monitored site were post-stratified according to habitat availability at the national scale. Such a procedure was necessary because butterflies and transects were not equally distributed over the country and habitats.

Post-stratification is also to be used when a biodiversity component is known to vary through space (which is likely to be true in most cases). Data from regions with contrasted trends need to be appropriately weighted so that the overall estimate is an unbiased combination of spatial variations in the trend (e.g. Olsen et al. 1999; Houlahan et al. 2001).

Different species/taxonomic groups

If different species or taxonomic groups are to be combined, several weighting rationales can be considered. First, *no* weighting is used when biological knowledge of the relationship among species and taxa is insufficient (Buckland et al. 2005). In practice, the same

weight is given to all species and taxonomic groups (e.g. Living Planet Index, Loh et al. 2005). Second, weights can be used to give priority to a given biological property, e.g. degree of specialization, rarity, originality, ecosystem function, or trophic level (e.g. Butchart et al. 2005; Pavoine et al. 2005), or to policy goals or conservation priorities (Yoccoz et al. 2001; Nichols and Williams 2006; Miller et al. 2007; Schmeller et al. in press a, b). These ad hoc weights are to be defined according to monitoring goals. The analysis may also need to account for phylogenetic non-independence across monitored species. The same response to a given environmental change from phylogenetically distant species is more convincing about the general impact of the change than the same response exhibited by closely related species (Helmus et al. 2007). Thus, comparisons among distant species should be given a higher weight than comparisons between closely related species. Several data transformations exist so that among-species comparisons are independent of phylogenetic relationships (Harvey and Pagel 1991; Faith et al. 2004; Pavoine et al. 2005). Another phylogenetic correction approach useful for relating species trends to environmental changes is to compare pairs of closely related species but that have contrasting ecological requirements or life-history traits (Møller and Birkhead 1992).

Cross-validation and robustness of conclusions

For a given dataset, at the end of the model selection (or effect selection in a stepwise regression), there will always be one final model, i.e. the model that supposedly makes the best compromise between good description of the data and parsimony (low number of parameters). The final statistical model, or the dataset, however, may be of poor generality. To evaluate the robustness of the conclusions, i.e. the external validity of the statistical analysis, a method is to use cross-validation. Part of the data is used for identifying the best statistical model, and the remaining part of the data is used to challenge this best model. This process is repeated several times. Cross-correlation coefficients quantify the departure between model predictions and observed data (Hastie et al. 2001). This approach has been used for example to gain insights into the effects of hunting on mallard (*Anas platyrhynchos*) populations in the USA (Nichols and Hines 1983).

When integrating data from different monitoring schemes, several datasets are at hand. The external validity of the model could be evaluated by computing cross-correlations across the different datasets. For instance, Breeding Bird Survey data from one set of countries could be used to parameterize the statistical model. Then, data from the remaining countries could be used for challenging the final statistical model by cross-correlation. This approach is particularly useful for assessing the robustness of spatial interpolations of biodiversity measures. If cross-correlation coefficients are high, the selected statistical model has a high predictive power, and it can also be concluded that the same major effects apply in the different sub-datasets. In other words, biodiversity states or trends are similar across schemes. At the opposite end, if cross-correlation coefficients are low, it means that important causes of biodiversity variation (i.e. effects) are still missing in the final statistical model.

Recommendations for future monitoring integration

From our overview and understanding of the monitoring practices, we suggest four priorities for future integration of ongoing biodiversity monitoring. (i) The experience of bird and butterfly monitoring (Gregory et al. 2005; European Environment Agency, 2007; van

Swaay et al. in press) should be used to develop similar bottom-up, international, federative monitoring programmes that produce indicators for other taxonomic groups. The number of existing schemes (EuMon consortium 2006) suggests that most vertebrates groups would be suitable, as well as several macro-invertebrates (e.g. beetles, Odonata, and Orthoptera), and plants as a whole (with a possible group focussing on orchids). (ii) The next integration step would be the production of indicators combining information from different taxonomic groups, e.g. for trophic chains per ecosystems. Several monitoring schemes already monitor different taxonomic groups simultaneously (Fig. 1). Land-use and fragmentation are the first causes of biodiversity change that could be assessed with such multi-taxa indicators (Table 1). Much research is still needed in this area for the definition of scientifically sound and user-friendly indicators for terrestrial ecosystems. (iii) Intellectual property and differences in sampling designs should no longer be a barrier to data exchange. Better than nothing is to exchange meta-data, i.e. estimates derived per scheme with standard statistical procedures. Finally, (iv) statistical tools (cf. meta-analysis methods, interpolation models, models mixing different data sources, cross-validation) should be further developed and fully enjoyed by biostatisticians implementing the integration of data from monitoring. Policy makers would benefit from more robust conclusions, at more appropriate spatial and temporal scales.

Conclusion

Monitoring data in Europe are scattered and heterogeneous (EuMon consortium 2006; e.g. Lengyel et al. in press a), but contain a massive amount of information on biodiversity changes and drivers of these changes. This information would be much more valuable for biodiversity assessment if it were more easily accessible, e.g. if assembled in meta-databases. We hope that this contribution will encourage researchers to develop biodiversity monitoring integration across schemes, and policy makers to support and rely more on output from integrated monitoring.

References

Balmford A, Green RE, Jenkins M (2003) Measuring the changing state of nature. Trends Ecol Evol 18:326–330. doi:10.1016/S0169-5347(03)00067-3

Balmford A, Crane P, Dobson AP, Green RE, Mace G (2005a) The 2010 challenge: data availability, information needs and extraterrestrial insights. Philos Trans R Soc Lond B Biol Sci 360:221–228. doi:10.1098/rstb.2004.1599

Balmford A, Bennun L, Ten Brink B, Cooper D, Côté IM, Crane P et al (2005b) The Convention on Biological Diversity's 2010 target. Science 307:212–213. doi:10.1126/science.1106281

Bell S, Marzano M, Cent J, Kobierska H, Podjed D, Vandzinskaite D et al (in press) What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. Biodivers Conserv

Besbeas P, Freeman SN (2006) Methods for joint inference from panel survey and demographic data. Ecology 87:1138–1145. doi:10.1890/0012-9658(2006)87[1138:MFJIFP]2.0.CO;2

Buckland ST (2006) Point-transect surveys for songbirds: robust methodologies. Auk 123:345–357. doi:10.1642/0004-8038(2006)123[345:PSFSRM]2.0.CO;2

Buckland ST, Magurran AE, Green RE, Fewster RM (2005) Monitoring change in biodiversity through composite indices. Philos Trans R Soc Lond B Biol Sci 360:243–254. doi:10.1098/rstb.2004.1589

Butchart SHM, Stattersfield JA, Baillie J, Bennun LA, Stuart SN, Akçakaya HR et al (2005) Using Red List Indices to measure progress towards the 2010 target and beyond. Philos Trans R Soc Lond B Biol Sci 360:255–268. doi:10.1098/rstb.2004.1583

Caughley G, Gunn A (1996) Conservation biology in theory and practice. Blackwell Science, Cambridge

Cooper HD, Hedges LV (1994) Handbook of research synthesis. Russel Sage Foundation, New York

Côté IM, Gill JA, Gardner TA, Watkinson AR (2005) Measuring coral reef decline through meta-analyses. Philos Trans R Soc Lond B Biol Sci 360:385–395. doi:10.1098/rstb.2004.1591

Council of the European Communities (1992) EC Habitats Directive. Council Directive 92/43/EEC of 21 May 1992 on the Conservation of natural habitats and of wild fauna and flora. Official Journal of the European Communities, Series L, 206:750. http://circa.europa.eu/Public/irc/env/monnat/library. Accessed 14 May 2008

Couvet D, Jiguet F, Julliard R, Levrel H, Teyssèdre A (2008) Enhancing citizen contributions to biodiversity science and public policy. Interdiscip Sci Rev 31:95–103

Danielsen F, Burgess ND, Balmford A (2005) Monitoring matters: examining the potential of locally-based approaches. Biodivers Conserv 14:2507–2542. doi:10.1007/s10531-005-8375-0

DeSante DF, O'Grady DR, Pyle P (1999) Measures of productivity and survival derived from standardized mist-netting are consistent with observed population changes. Bird Study 46:178–188

Devictor V, Julliard R, Couvet D, Lee A, Jiguet F (2007a) Functional homogenization effect of urbanization on bird communities. Conserv Biol 21:741–751. doi:10.1111/j.1523-1739.2007.00671.x

Devictor V, Godet L, Julliard R, Couvet D, Jiguet F (2007b) Can common species benefit from protected areas? Biol Conserv 139:29–36. doi:10.1016/j.biocon.2007.05.021

Devictor V, Julliard R, Clavel J, Jiguet F, Lee A, Couvet D (2008) Functional biotic homogenization of bird communities in disturbed landscapes. Glob Ecol Biogeogr 17:252–261. doi:10.1111/j.1466-8238.2007.00364.x

Dolédec S, Statzner B, Bournard M (1999) Special traits for future biomonitoring across ecoregions: patterns along a human-impacted river. Freshw Biol 42:737–758. doi:10.1046/j.1365-2427.1999.00509.x

Donald PF, Sanderson FJ, Burfield IJ, van Bommel FPJ (2006) Further evidence of continent-wide impacts of agricultural intensification on European farmland birds, 1990–2000. Agric Ecosyst Environ 116:189–196. doi:10.1016/j.agee.2006.02.007

Donald PF, Sanderson FJ, Burfield IJ, Bierman SM, Gregory RD, Waliczky Z (2007) International conservation policy delivers benefits for birds in Europe. Science 317:810–813. doi:10.1126/science.1146002

Dziock F, Henle K, Foeckler F, Follner K, Scholz M (2006) Biological indicator systems in floodplains—a review. Int Rev Hydrobiol 91:271–291. doi:10.1002/iroh.200510885

Elzinga CL, Salzer DW, Willoughby JW, Gibbs JP (2001) Monitoring plant and animal populations: a handbook for field biologists. Blackwell Science, Malden, MA

EuMon consortium (2006) DaEuMon: a database of animals, plants and habitats monitoring schemes in Europe. http://eumon.ckff.si/. Accessed 14 May 2008

European Environment Agency (2007) Halting the loss of biodiversity by 2010: proposal for a first set of indicators to monitor progress in Europe. Office for Official Publications of the European Communities. http://reports.eea.europa.eu. Accessed 14 May 2008

Faith DP, Reid CAM, Hunter J (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. Conserv Biol 18:255–261. doi:10.1111/j.1523-1739.2004.00330.x

Gauthier G, Besbeas P, Lebreton JD, Morgan BJT (2007) Population growth in snow geese: a modeling approach integrating demographic and survey information. Ecology 88:1420–1429. doi:10.1890/06-0953

Green RE, Balmford A, Crane PR, Mace GM, Reynolds JD, Turner RK (2005) A framework for improved monitoring of biodiversity: responses to the World Summit on Sustainable Development. Conserv Biol 19:56–65. doi:10.1111/j.1523-1739.2005.00289.x

Gregory RD, van Strien A, Vorisek P, Gmelig Meyling AW, Noble DG, Foppen RPB et al (2005) Developing indicators for European birds. Philos Trans R Soc Lond B Biol Sci 360:269–288. doi:10.1098/rstb.2004.1602

Gurevitch J, Curtis PS, Jones MH (2001) Meta-analysis in ecology. Adv Ecol Res 32:199–247. doi:10.1016/S0065-2504(01)32013-5

Harvey PH, Pagel M (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Data mining, inference and prediction. Springer-Verlag, New York

Hedges LV, Olkin I (1985) Statistical methods for meta-analysis. Academic Press, Orlando

Helmus MR, Bland TJ, Williams CK, Ives AR (2007) Phylogenetic measures of biodiversity. Am Nat 169:E68–E83. doi:10.1086/511334

Henle K (2005) Lessons from Europe. In: Lannoo M (ed) Amphibian declines: the conservation status of United States species. University of California Press, Berkeley

Henle K, Schmeller DS (this volume) Biodiversity conservation strategies and monitoring in Europe. Biodivers Conserv (submitted)

Henle K, Alard D, Clitherow J, Cobb P, Firbank L, Kull T et al (2008) Identifying and managing the conflicts between agriculture and biodiversity conservation in Europe—a review. Agric Ecosyst Environ 124:60–71. doi:10.1016/j.agee.2007.09.005

Henle K, Davies KF, Kleyer M, Margules C, Settele J (2004) Predictors of species sensitivity to fragmentation. Biodivers Conserv 13:207–251. doi:10.1023/B:BIOC.0000004319.91643.9e

Henle K, Dziock F, Foeckler F, Follner K, Husing V, Hettrich A et al (2006) Study design for assessing species environment relationships and developing indicator systems for ecological changes in floodplains—the approach of the RIVA project. Int Rev Hydrobiol 91:292–313. doi:10.1002/iroh.200610886

Henry P-Y, Jarne P (2007) Marking hard-shelled gastropods: tag loss, impact on life-history traits, and perspectives in biology. Invertebr Biol 126:138–153. doi:10.1111/j.1744-7410.2007.00084.x

Hochachka WM, Martin K, Doyle F, Krebs CJ (2000) Monitoring vertebrate populations using observational data. Can J Zool 78:521–529. doi:10.1139/cjz-78-4-521

Hooten MB, Wikle CK, Dorazio RM, Royle JA (2007) Hierarchical spatiotemporal matrix models for characterizing invasions. Biometrics 63:558–567. doi:10.1111/j.1541-0420.2006.00725.x

Houlahan JE, Findlay CS, Meyer AH, Kuzmin SL, Schmidt BR (2001) Global amphibian population declines—reply. Nature 412:500. doi:10.1038/35087661

Hughes JB, Daily GC, Ehrlich PR (1997) Population diversity: its extent and extinction. Science 278:689–692. doi:10.1126/science.278.5338.689

Jiguet F, Julliard R, Couvet D, Petiau A (2005) Modeling spatial trends in estimated species richness using breeding bird survey data: a valuable tool in biodiversity assessment. Biodivers Conserv 14:3305–3324. doi:10.1007/s10531-004-0448-y

Julliard R (2004) Estimating the contribution of survival and recruitment to large scale population dynamics. Anim Biodivers Conserv 27:417–426

Julliard R, Jiguet F, Couvet D (2004a) Common birds facing global changes: what makes a species at risk? Glob Change Biol 10:148–154. doi:10.1111/j.1365-2486.2003.00723.x

Julliard R, Jiguet F, Couvet D (2004b) Evidence for the impact of global warming on the long-term population dynamics of common birds. Proc R Soc Lond B Biol Sci 271:S490–S492. doi:10.1098/rsbl.2004.0229

Kati V, Devillers P, Dufrêne M, Legakis A, Vokou D, Lebrun P (2004) Testing the value of six taxonomic groups as biodiversity indicators at a local scale. Conserv Biol 18:667–675. doi:10.1111/j.1523-1739.2004.00465.x

Krebs CJ, Berteaux D (2006) Problems and pitfalls in relating climate variability to population dynamics. Clim Res 32:143–149. doi:10.3354/cr032143

Kull T, Sammul M, Tali K, Lanno K, Henle K (this volume) Necessity and reality of monitoring European vascular plant diversity. Biodivers Conserv (submitted)

Lengyel S, Déri E, Varga Z, Horváth R, Tóthmérész B, Henry P-Y et al (in press, a) Habitat monitoring in Europe: a description of current practices. Biodivers Conserv

Lengyel S, Kobler A, Kutnar L, Framstad E, Henry P-Y, Babij V et al (in press, b) A review and a framework for the integration of biodiversity monitoring at the habitat level. Biodivers Conserv

Loh J, Green RE, Ricketts T, Lamoreux J, Jenkins M, Kapos V et al (2005) The Living Planet Index: using species population time series to track trends in biodiversity. Philos Trans R Soc Lond B Biol Sci 360:289–295. doi:10.1098/rstb.2004.1584

Mace GM (2005) A user's guide to biodiversity indicators. European Academies Science Advisory Council, The Royal Society, London

MacKenzie DI, Nichols JD, Sutton N, Kawanishi K, Bailey LL (2005) Improving inferences in population studies of rare species that are detected imperfectly. Ecology 86:1101–1113. doi:10.1890/04-1060

Margurran AE (2004) Measuring biological diversity. Blackwell Science, Oxford

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Aha R et al (2006) European phenological response to climate change matches the warming pattern. Glob Change Biol 12:1969–1976. doi:10.1111/j.1365-2486.2006.01193.x

Miller RM, Rodriguez JP, Aniskowicz-Fowler T, Bambaradeniya C, Boles R, Eaton MA et al (2007) National threatened species listing based on IUCN criteria and regional guidelines: current status and future perspectives. Conserv Biol 21:684–696. doi:10.1111/j.1523-1739.2007.00656.x

Møller AP, Birkhead TR (1992) A pairwise comparative method as illustrated by copulation frequency in birds. Am Nat 139:644–656. doi:10.1086/285348

Møller AP, Jennions MD (2001) Testing and adjusting for publication bias. Trends Ecol Evol 16:580–586. doi:10.1016/S0169-5347(01)02235-2

Nichols JD, Hines JE (1983) The relationship between harvest and survival rates of mallards: a straight-forward approach with partitioned data sets. J Wildl Manage 47:334–348. doi:10.2307/3808506

Nichols JD, Williams BK (2006) Monitoring for conservation. Trends Ecol Evol 21:668–673. doi:10.1016/j.tree.2006.08.007

Olsen AR, Sedransk J, Edwards D, Gotway CA, Liggett W, Rathbun S et al (1999) Statistical issues for monitoring ecological and natural resources in the United States. Environ Monit Assess 54:1–45. doi:10.1023/A:1005823911258

Osenberg CW, Sarnelle O, Cooper SD, Holt RD (1999) Resolving ecological questions through meta-analysis: goals, metrics, and models. Ecology 80:1105–1117

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42. doi:10.1038/nature01286

Parr TW, Ferretti M, Simpson IC, Forsius M, Kovacs-Lang E (2002) Towards a long-term integrated monitoring programme in Europe: network design in theory and practice. Environ Monit Assess 78:253–290. doi:10.1023/A:1019934919140

Pauly D, Watson R (2005) Background and interpretation of the 'Marine Trophic Index' as a measure of biodiversity. Philos Trans R Soc Lond B Biol Sci 360:415–423. doi:10.1098/rstb.2004.1597

Pavoine S, Ollier S, Dufour AB (2005) Is the originality of a species measurable? Ecol Lett 8:579–586. doi:10.1111/j.1461-0248.2005.00752.x

Pereira HM, Cooper HD (2006) Towards the global monitoring of biodiversity change. Trends Ecol Evol 21:123–129. doi:10.1016/j.tree.2005.10.015

Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ, Stenseth NC (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol Evol 20:503–510. doi:10.1016/j.tree.2005.05.011

Pollock JF (2006) Detecting population declines over large areas with presence-absence, time-to-encounter, and count survey methods. Conserv Biol 20:882–892. doi:10.1111/j.1523-1739.2006.00342.x

Pons P, Henry P-Y, Gargallo J, Prodon R, Lebreton J-D (2003) Bird survival after fire in Mediterranean shrublands: combining capture-recapture data over several bird species. Popul Ecol 45:187–196. doi:10.1007/s10144-003-0155-7

Pradel R, Henry P-Y (2007) Potential contributions of capture-recapture to the estimation of population growth rate in restoration projects. Ecoscience 14:432–439. doi:10.2980/1195-6860(2007)14[432:PCOCTT]2.0.CO;2

Roberts RL, Donald PF, Green RE (2007) Using simple species lists to monitor trends in animal populations: new methods and a comparison with independent data. Anim Conserv 10:332–339. doi:10.1111/j.1469-1795.2007.00117.x

Rosenstock SS, Anderson DR, Giesen KM, Leukering T, Carter MF (2002) Landbird counting techniques: current practices and an alternative. Auk 119:46–53. doi:10.1642/0004-8038(2002)119[0046:LCTCPA]2.0.CO;2

Roy DB, Rothery P, Brereton T (2007) Reduced-effort schemes for monitoring butterfly populations. J Appl Ecol 44:993–1000. doi:10.1111/j.1365-2664.2007.01340.x

Schaub M, Gimenez O, Sierro A, Arlettaz R (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Conserv Biol 21:945–955. doi:10.1111/j.1523-1739.2007.00743.x

Schmeller DS, Gruber B, Bauch B, Lanno K, Budrys E, Babij V et al (in press, a) Determination of national conservation responsibilities in regions with multiple political jurisdictions. Biodivers Conserv

Schmeller DS, Bauch B, Gruber B, Juškaitis R, Budrys E, Babij V et al (in press, b) Determination of conservation priorities in regions with multiple political jurisdictions. Biodivers Conserv

Schmeller DS, Henry P-Y, Julliard R, Clobert J, Lengyels S, Gruber B, Dziock F, Kull T, Tali K, Bauch B, Kobler A, Déri E, Babij V, Nowicki P, Papastergiadou E, van Swaay CAM, Henle K Advantages of volunteer-based biodiversity monitoring in Europe (submitted)

Statzner B, Bis B, Dolédec S, Usseglio-Polatera P (2001) Perspectives for biomonitoring at large spatial scales: a unified measure for the functional composition of invertebrate communities in European running waters. Basic Appl Ecol 2:73–85. doi:10.1078/1439-1791-00039

Stoner C, Caro T, Mduma S, Mlingwa C, Sabuni G, Borner M (2007) Assessment of effectiveness of protection strategies in Tanzania based on a decade of survey data for large herbivores. Conserv Biol 21:635–646. doi:10.1111/j.1523-1739.2007.00705.x

Strayer DL (1999) Statistical power of presence-absence data to detect population declines. Conserv Biol 13:1034–1038. doi:10.1046/j.1523-1739.1999.98143.x

Sutherland WJ, Pullin AS, Dolman PM, Knight TM (2004) The need for evidence-based conservation. Trends Ecol Evol 19:305–308. doi:10.1016/j.tree.2004.03.018

Teder T, Moora M, Roosaluste E, Zobel K, Partel M, Koljalg U et al (2007) Monitoring of biological diversity: a common-ground approach. Conserv Biol 21:313–317. doi:10.1111/j.1523-1739.2006.00575.x

Thomas JA (1995) Why small cold-blooded insects pose different conservation problems to birds in modern landscapes. Ibis 137:S112–S119. doi:10.1111/j.1474-919X.1995.tb08431.x

Thomas JA (2005) Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. Philos Trans R Soc Lond B Biol Sci 360:339–357. doi:10.1098/rstb.2004.1585

Thomas JA, Telfer MG, Roy DB, Preston CD, Greenwood JJD, Asher J et al (2004) Comparative losses of British butterflies, birds and plants and the global extinction crisis. Science 303:1879–1881. doi:10.1126/science.1095046

Van Strien AJ, Plantenga WF, Soldaat LL, Van Swaay CAM, WallisDeVries MF (2008) Bias in phenology assessments based on first appearance data of butterflies. Oecologia 156:227–235

van Swaay CAM, Plate CL, van Strien AJ (2002) Monitoring butterflies in the Netherlands: how to get unbiased indices. Proc Exp Appl Entomol 13:21–27

van Swaay CAM, Nowicki P, Settele J, van Strien AJ (in press) Butterfly Monitoring in Europe—methods, applications and perspectives. Biodivers Conserv

Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. Trends Ecol Evol 16:446–453. doi:10.1016/S0169-5347(01)02205-4

Yoccoz NG, Nichols JD, Boulinier T (2003) Monitoring of biological diversity—a response to Danielsen et al. Oryx 37:410. doi:10.1017/S0030605303000747

Webb GJW, Dillon ML, McLean GE, Manolis SC, Ottley B (1990) Monitoring the saltwater crocodile (*Crocodylus porosus*) population in the Northern Territory of Australia. Proc Work Meet IUCN-SSC Crocodile Spec Group 9:329–380