Ecological Indicators 85 (2018) 624–633

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

Evaluating and benchmarking biodiversity monitoring: Metadata-based indicators for sampling design, sampling effort and data analysis

Szabolcs Lengyel[a,*], Beatrix Kosztyi[a], Dirk S. Schmeller[b], Pierre-Yves Henry[c], Mladen Kotarac[d], Yu-Pin Lin[e], Klaus Henle[b]

[a] Department of Tisza Research, Danube Research Institute, Centre for Ecological Research, Hungarian Academy of Sciences, Bem tér 18 /c, 4032, Debrecen, Hungary
[b] Helmholtz Centre for Environmental Research – UFZ, Department of Conservation Biology, Permoserstr. 15., Leipzig, D-04318, Germany
[c] Centre d'Ecologie et des Sciences de la Conservation (CESCO UMR 7204), CNRS, MNHN, UPMC, Sorbonne Universites & Mecanismes Adaptatifs et Evolution (MECADEV UMR 7179), CNRS, MNHN, Sorbonne Universites, Muséum National d'Histoire Naturelle, 1 avenue du Petit Château, 91800, Brunoy, France
[d] Centre for the Cartography of Fauna and Flora, Klunova ulica 3, SI-1000, Ljubljana,Slovenia
[e] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, 10617, Taiwan

ARTICLE INFO

Keywords:
2020 target
Assessment
Biodiversity observation network
Biodiversity strategy
Citizen science
Conservation funding
Environmental policy
Evidence-based conservation
Statistical power
Surveillance

ABSTRACT

The biodiversity crisis has led to a surge of interest in the theory and practice of biodiversity monitoring. Although guidelines for monitoring have been published since the 1920s, we know little on current practices in existing monitoring schemes.

Based on metadata on 646 species and habitat monitoring schemes in 35 European countries, we developed indicators for sampling design, sampling effort, and data analysis to evaluate monitoring practices. We also evaluated how socio-economic factors such as starting year, funding source, motivation and geographic scope of monitoring affect these indicators.

Sampling design scores varied by funding source and motivation in species monitoring and decreased with time in habitat monitoring. Sampling effort decreased with time in both species and habitat monitoring and varied by funding source and motivation in species monitoring.

The frequency of using hypothesis-testing statistics was lower in species monitoring than in habitat monitoring and it varied with geographic scope in both types of monitoring. The perception of the minimum annual change detectable by schemes matched spatial sampling effort in species monitoring but was rarely estimated in habitat monitoring.

Policy implications: Our study identifies promising developments but also options for improvement in sampling design and effort, and data analysis in biodiversity monitoring. Our indicators provide benchmarks to aid the identification of the strengths and weaknesses of individual monitoring schemes relative to the average of other schemes and to improve current practices, formulate best practices, standardize performance and integrate monitoring results.

1. Introduction

The global decline of biodiversity and ecosystem services led to the adoption of several ambitious goals by the international community for 2010 and then again for 2020. Monitoring of biodiversity is instrumental in evaluating whether these goals are met. Although literature on how monitoring systems should be organized has been published since at least the mid-1920s (Cairns and Pratt, 1993), interest in the theory and practice of biodiversity monitoring has surged since 1990 (Noss, 1990; Yoccoz et al., 2001) and culminated in comprehensive, theory-based recommendations for monitoring (Balmford et al., 2003; Lindenmayer and Likens, 2009; Mace et al., 2005; Pocock et al., 2015).

Despite this growing knowledge, significant concerns regarding current practices remain (Lindenmayer and Likens, 2009; Walpole et al., 2009). A consistently voiced concern is that monitoring is not adequately founded in theory because many schemes are not designed to test hypotheses about biodiversity change even though their primary objective, almost exclusively, is to detect changes in biodiversity (Balmford et al., 2005; Nichols and Williams, 2006; Yoccoz et al., 2001). Although not all monitoring schemes require hypothesis-testing

* Corresponding author.
E-mail addresses: lengyel.szabolcs@okologia.mta.hu (S. Lengyel), cleo.deb@gmail.com (B. Kosztyi), dirk.schmeller@ufz.de (D.S. Schmeller), henry@mnhn.fr (P.-Y. Henry), mladen@ckff.si (M. Kotarac), yplin@ntu.edu.tw (Y.-P. Lin), klaus.henle@ufz.de (K. Henle).

https://doi.org/10.1016/j.ecolind.2017.11.012
Received 11 July 2017; Received in revised form 19 October 2017; Accepted 6 November 2017
1470-160X/

given the variety of their objectives (Pocock et al., 2015), there is also a general concern over the ability of monitoring schemes to adequately detect changes in biodiversity due to biased sampling designs, inadequate sampling effort, or low statistical power to detect changes (Di Stefano, 2001; Mihoub et al., 2017). Legg and Nagy (2006) and Lindenmayer and Likens (2009) warned that these shortcomings may lead to poor quality of monitoring, and, ultimately, to a waste of valuable conservation resources.

There is little information, however, on the prevalence of these potential methodological weaknesses in current practices of biodiversity monitoring. Descriptions of current practices are available for monitoring schemes in North America (Marsh and Trenham, 2008), and for European schemes of habitat monitoring (Lengyel et al., 2008a) and bird monitoring (Schmeller et al., 2012), however, these descriptions do not evaluate strengths or weaknesses in monitoring. Monitoring schemes are rarely known well enough for a comprehensive evaluation of current practices (Henle et al., 2010a; Schmeller et al., 2009), partly because monitoring schemes are designed for many different objectives at different spatial and temporal scales (Geijzendorffer et al., 2015; Jarzyna and Jetz, 2016; Pocock et al., 2015). Therefore, the performance of biodiversity monitoring in terms of the criteria regarded by the critiques as insufficiently considered in monitoring has not yet been assessed. Consequently, little is known about whether and how performance varies among programs by spatial and temporal scales or socio-economic drivers. Moreover, it is rarely known whether and how programs evaluate their performance, either by expert judgement on their ability to detect trends or by estimating their statistical power to detect changes (Geijzendorffer et al., 2015; Nielsen et al., 2009). Hence, there is a need to provide monitoring coordinators with standard indicators of performance so that they can evaluate their programs and revise their practices to address potential weaknesses. A clear understanding of performance in existing monitoring schemes also provides crucial information to the institutions running and funding monitoring schemes as well as to policy-makers using information from biodiversity monitoring.

Here we present an overview of current practices in biodiversity monitoring in Europe by focusing on properties that have been frequently mentioned in critiques of biodiversity monitoring. We used metadata on monitoring schemes to develop indicators for sampling design, sampling effort and type of statistical analysis. While monitoring schemes have been established for many different purposes, these three properties are regarded as generally relevant in determining the scientific quality of the information derived from biodiversity monitoring (Lindenmayer and Likens, 2009; Nichols and Williams, 2006; Yoccoz et al., 2001). Sampling design, an indicator of how well the spatial and temporal distribution of data collection is founded in sampling theory (Balmford et al., 2003), is essential for accuracy, i.e., closeness of measured trends and real trends in biodiversity. Sampling effort, the number of measurements made, is central to precision, i.e., the ability to measure the same value under identical conditions. Finally, to translate collected data into information relevant for further use, such as conservation or policy, appropriate statistical analysis of data is required to detect changes or trends with a given level of uncertainty, and confidence in the estimates should be based on the ability of the scheme to detect changes (Legg and Nagy, 2006).

Although these three indicators are generally relevant in any type of monitoring, monitoring schemes differ in their objectives and many different types of monitoring schemes exist (Pocock et al., 2015). For example, schemes in Europe have been started as early as the 1970s, are motivated by different reasons, funded by different sources, and their geographic scope ranges from local to continental (Lengyel et al., 2008a; Schmeller et al., 2012). To account for these socio-economic differences and to increase the useability of our indicators in different monitoring schemes, we evaluated the variation in indicators as a function of starting year, funding source, motivation, and geographic scope. Finally, we show how our indicators can be used by coordinators as benchmarks to assess their schemes relative to the average practice and to identify options for improvement of their monitoring schemes. We present different benchmark values for the three indicators to be meaningful for schemes monitoring different species groups and habitat types.

Table 1
Scores allocated to different levels of variables describing the sampling design used in species and habitat monitoring schemes in Europe. Please see Supporting information for justification of score values.

Object monitored	Variable	Response option	Score
Species	Monitored property	Population trend	0
		Distribution trend	1
		Community/ecosystem trend	2
		Population + distribution trend	1
		Population + community trend	2
		Distribution + community trend	3
		All three of the above	3
	Data type	Presence/absence	0
		Age/size structure	1
		Phenology	1
		Counts	2
		Mark-recapture	3
	Information on population structure	No	0
		Yes	1
	Stratification of sampling design	No	0
		Yes	1
	Experimental design	Not used	0
		Before/after comparison	1
		Controlled experiment	2
		Before/after plus control	3
	Selection of sampling sites	Expert/personal knowledge or other criteria	0
		Exhaustive, random, or systematic	1
	Detection probability	Not quantified	0
		Quantified	1
Habitats	Monitored property	Species composition (quality)	0
		Distribution (quantity)	1
		Both of the above (quality and quantity)	2
	Data type	Species presence/absence	0
		Species abundance	1
	Documentation of spatial variation	Not reported/no spatial aspect	0
		Field mapping	1
		Remote sensing	2
	Extent of monitoring	Certain habitat types in an area	0
		All habitat types in area	1
	Stratification of sampling design	Not stratified	0
		Stratified	1
	Experimental design	Not used	0
		Used	1
	Selection of sampling sites	Expert/personal knowledge or other criteria	0
		Exhaustive, random, or systematic	1

2. Methods

2.1. Definition and dataset

We used Hellawell's (1991) definition of "biodiversity monitoring" as the repeated recording of the qualitative and/or quantitative