

# Optimizing monitoring schemes to detect trends in abundance over broad scales

G. F. Ficetola<sup>1,2,3</sup>, A. Romano<sup>3,4</sup> , S. Salvidio<sup>3,5</sup> & R. Sindaco<sup>3,6</sup>

1 Laboratoire d'Écologie Alpine (LECA), CNRS, Univ. Grenoble Alpes, Grenoble, France

2 Department of Environmental Science and Policy, Università degli Studi di Milano, Milano, Italy

3 Societas Herpetologica Italica, Torino, Italy

4 Consiglio Nazionale delle Ricerche, Istituto di Biologia Agroambientale e Forestale (CNR-IBAF), Monterotondo, Italy

5 DISTAV – Dipartimento di Scienze della Terra dell'Ambiente e della Vita, Università di Genova, Genova, Italy

6 I.P.L.A., Istituto per le Piante da Legno e l'Ambiente, Torino, Italy

## Keywords

abundance; species decline; detection probability; *N*-mixture models; optimization; sub-transects; visual transects; monitoring schemes.

## Correspondence

Gentile F. Ficetola, Department of Environmental Science and Policy, Università degli Studi di Milano. Via Celoria 26, 20133 Milano, Italy.

Email: francesco.ficetola@gmail.com

Editor: Res Altwegg

Associate Editor: Michael Schaub

Received 21 December 2016; accepted 24 April 2017

doi:10.1111/acv.12356

## Abstract

Measuring population changes and trends is essential to identify threatened species, and is requested by several environmental regulations (e.g. European Habitat Directive). However, obtaining this information for small and cryptic animals is challenging, and requires complex, broad-scale monitoring schemes. How should we allocate the limited resources available for monitoring, to maximize the probability of detecting declines? The analysis of simulated data can help to identify the performance of monitoring scenarios across species with different features. We simulated data of populations with a wide range of abundance, detection probability and rate of decline, and tested under which circumstances open-population *N*-mixture models can successfully detect the decline of populations. We tested multiple monitoring strategies, to identify the ones having the highest probability of detecting declines. If 30 sites are surveyed, strong declines ( $\geq 30\%$ ) can be successfully spotted for nearly all the simulated species, except the species with lowest abundance and detection probability. Weaker declines are successfully identified only in species that are easy to detect and have high abundance. Increasing the number of sites quickly increases model power, but hundreds of sites would require monitoring to measure trends of the least detectable species. For most of species, performance of monitoring was improved by: surveying many sites with a few replicates per site; surveying many small sites instead of a few large sites; combining data from sites monitored for multiple species. Our findings show that one single monitoring approach cannot be appropriate for all the species, and that surveying efforts should be modulated across them, according to their detection probabilities and abundances. We provide quantitative values on how the number of surveys and the number of sites to be surveyed can be assigned to different species, and emphasize the need of planning to maximize the performance of monitoring.

## Introduction

The detection and quantification of trends in species abundance (e.g. declines or increases) is essential to ascertain conservation status, to identify threatened species, and to assess the effectiveness of conservation strategies (Yoccoz, Nichols & Boulinier, 2001; Reynolds, Thompson & Russell, 2011). For instance, in IUCN redlists, quantitative measures of decline are essential to qualify a species as threatened under criterion A, and under several other sub-criteria (IUCN, 2001). Measuring trends of protected species is also a major commitment requested by environmental regulations. The EU Habitats Directive (Directive 92/43/EEC) is perhaps the most important tool for biodiversity conservation in Europe, as it protects >1200 species and

their habitats. The Directive requires that, every 6 years, Member States shall draw up a report on the conservation status of listed species, including measures of trends of populations, to evaluate the efficiency of protection measures undertaken. Trends of species' ranges, as well as habitat extent changes, can be estimated by remote-sensing data (Tracewski *et al.*, 2016), but remote-sensing cannot capture the effect of local pressures determining the decline of populations (e.g. diseases, disturbance, exploitation, modification of microhabitats...), nor provide data on population changes (changes in the number of populations or in the number of individuals per population).

Obtaining quantitative measures of species trends over broad areas is challenging, as it generally requires repeated surveys of many sites covering significant portions of

species' ranges. Volunteers can provide data at low cost, thus helping the measurement of the trends of widespread species living in easily accessible regions (Kéry *et al.*, 2009; Bonardi *et al.*, 2011; Sewell *et al.*, 2012; Griffiths *et al.*, 2015; Petrovan & Schmidt, 2016). However, there are cases in which volunteer data are not available, particularly for species requiring specific monitoring protocols or living in difficult to access areas. For instance, the Italian reptiles and amphibians have some of the highest levels of richness and endemism in Europe (Sillero *et al.*, 2014). The most abundant and widespread species range through the whole territory, thus enabling monitoring through volunteers (Bonardi *et al.*, 2011), while many endemic and threatened species are restricted to small insular or mountainous areas with limited accessibility (Sindaco *et al.*, 2006), where consistent volunteer sampling is unlikely. Under these conditions, there is the need to develop a pragmatic and efficient strategy for monitoring, optimizing the limited available resources.

When the detection probability is  $<1$  (as usually occurs in animal populations), estimating parameters such as presence/absence, abundance or trends may be problematic and requires taking into account the imperfect detection (Pollock *et al.*, 2002; Schmidt, 2004; Mazerolle *et al.*, 2007). However, formal approaches exist for estimating abundance from repeated counts in fixed sites, without marking individuals to identify the individuals (Royle & Nichols, 2003; Royle, 2004). These approaches, named hierarchical or  $N$ -mixture models, allow the estimation of population size and abundance trends for species that are imperfectly detected (Kéry *et al.*, 2009; Dail & Madsen, 2011; Zipkin *et al.*, 2014), and are thus appropriate to detect population declines. Such models are able to take into account the high heterogeneity of data collected over broad scales by a large number of observers, which is typical of many monitoring schemes, and are thus highly promising for population and even species assessments (Kéry *et al.*, 2009; Griffiths *et al.*, 2015). Nevertheless, as for any statistical approach, the power of  $N$ -mixture models is influenced by factors such as effect size and sample size. The analysis of synthetic data simulating ecological processes is a powerful framework, which allows evaluating the effect of sampling and analytical methods (Guillera-Arroita, Ridout & Morgan, 2010; Zurell *et al.*, 2010; Ficetola *et al.*, 2014). A few studies tried to identify under which conditions (number of sites, number of surveys) it is possible to successfully detect population declines (e.g. Guillera-Arroita, 2012; Sewell *et al.*, 2012; Kroll *et al.*, 2015), but these analyses mostly focused on occupancy models. Occupancy models do not provide explicit estimates of abundance (but see Royle & Nichols, 2003; Ellis, Ivan & Schwartz, 2014), and thus mostly deal with changes in number of populations, instead of estimating trends in abundance.

Given the monitoring required by the Habitat Directive, the Italian Herpetological Society was tasked by government agencies with identifying monitoring methodologies for the  $>70$  Italian species of amphibians and reptiles of European concern (Stoch & Genovesi, 2016), but discussion among experts lead to different proposals with regard to the best strategy for monitoring. The aim of this study was to identify the most effective monitoring to detect abundance trends

(e.g. declines) through  $N$ -mixture models in each of multiple species with very heterogeneous features. When planning a monitoring scheme, the allocation of resources must be optimized to maximize the probability to address the specific management questions (Nichols & Williams, 2006; Wintle, Runge & Bekessy, 2010; Reynolds *et al.*, 2011) and, if the monitoring scheme includes multiple species, it is important to identify the optimal strategy for each of them (Guillera-Arroita *et al.*, 2010; Guillera-Arroita, 2012). The Italian amphibians and reptiles have very different features, some species being locally abundant and conspicuous, and others rare and difficult to detect (Data S1). We therefore assessed how species features (abundance and detection probability) and monitoring structure (e.g. number of surveys, number of sites) influence the detection of declines. Specifically, we asked: (1) How do species abundance and detection probability determine our ability to detect declines? (2) How many sites should be surveyed for each target species? (3) For species with given abundance and detectability, is it better monitoring a few sites with repeated surveys, or many sites with a few surveys per site? (4) Is it better concentrating efforts on a few, large sites (e.g. long transects or large plots), or increasing sample size and analyzing many small sub-transects (or sub-plots)? (5) Under which conditions is it possible combining surveys on multiple species to improve efficiency? Although analyses were initially motivated by the need of a national plan for the monitoring of Italian amphibians and reptiles (Stoch & Genovesi, 2016), our simulations represent realistic sampling covering a wide range of scenarios of abundance, detection probability, declines and monitoring schemes. Therefore, our recommendations are applicable to the planning of monitoring of a large number of taxa and regions.

## Methods

### Rationale for models

In simulations, we generated artificial species, with abundance and detection probability reflecting values observed in natural populations. The features of simulated species were initially intended to cover the actual abundance and detection probability of European amphibians and reptiles, but analogous values are also observed during monitoring of several other taxa (see Data S1). Each artificial species had a specific combination of mean abundance per site and detection probability (total: 30 combinations). Average abundance values were 7.5, 15, 30, 50, 100 and 300 individuals per site, detection probability values were set at 0.05, 0.1, 0.15, 0.33 and 0.5. For each species, we first assumed a 'basic' monitoring, which was performed at 30–500 relatively large sites where the target species was known to be present; each site was surveyed at two time points (e.g. in two different years). This corresponds to proposals of monitoring each site once every 6 years to fulfill the Habitat Directive requirements (Stoch & Genovesi, 2016). Example of sites may include 1-km visual transects, but also ponds surveyed for aquatic and semiaquatic species or forest plots. We also simulated

the performance of two alternative strategies, proposed to improve trend estimates: (1) each of the sites is subdivided in four smaller sub-sites; (2) combining the sites monitored for multiple species, to increase the sample size.

## Simulations and $N$ -mixture models

We first simulated population size  $N$  at 30 sites at the beginning of the study (time =  $t_1$ ) from a Poisson distribution, with  $\lambda_1$  = mean abundance (range = 7.5–300; average number of individuals). The expected count at site  $i$  was the product of abundance  $N_i$  and the detection probability of the species  $p$  (range = 0.05–0.5). To simulate the observation process, at each survey the number of individuals observed at site  $i$  was drawn from a binomial distribution with  $n = N_i$  and probability =  $p$  (Kéry, 2010); each site was surveyed 3, 5 or 7 times within each season. Due to convergence failure in some runs, we did not test combinations with mean abundance = 300,  $p \geq 0.33$ .

The sites were re-surveyed at time  $t_2$ , when the species suffered a decline  $D$ . At time  $t_2$ , population size was estimated from a Poisson distribution with  $\lambda_2 = \lambda_1 - D \times \lambda_1$ . We considered four possible rates of decline, from limited to dramatic ( $D = 10, 20, 30$  and 50%), while  $p$  remained constant. We repeated 150 simulations per combination of parameter sets (mean abundance,  $p$ , number of surveys and  $D$ ; 276 combinations). Abundance of site  $i$  at time  $t_2$  was not specifically related to the abundance at the same site at  $t_1$ , as they were independently drawn from two distinct distributions differing for mean abundance. Therefore, this approach assumed an overall decline of the species, but some populations may be more abundant at  $t_2$  than at  $t_1$ .

First, we estimated the relative bias of  $N$ -mixture models in estimating the actual abundance of the species. For each simulation, we run single-season  $N$ -mixture models with Poisson error distribution on the data at  $t_1$  (Royle, 2004; fitted using `pcount` in `unmarked`; Fiske & Chandler, 2011), and the population size estimated from the posterior distributions of the random variables using empirical Bayes methods (Royle & Dorazio, 2008). Relative bias was calculated as the median of  $\text{abs}[(\text{estimated population size} - \text{true population size})/\text{true population size}]$ , across all simulations. We also calculated the median Pearson's correlation between estimated and true population size.

Second, we tested whether  $N$ -mixture models for open populations (Dail & Madsen, 2011) can successfully estimate the decline between  $t_1$  and  $t_2$ . In models, we assumed that  $N_{[i,t_2]} = N_{[i,t_1]} \times \gamma$ , where  $\gamma$  is the rate of population change (Fiske & Chandler, 2011). We considered that a model successfully detected the decline if  $\gamma$  was significantly  $< 1$  (at  $\alpha = 0.05$ ); models were built using the `pcountOpen` function in `unmarked`, assuming Poisson error. In  $N$ -mixture models, the default values of the upper bound used to approximate an infinite summation in the likelihood can provide inaccurate estimates. In each simulation, we used as upper bound (70 + the maximum observed species abundance), since preliminary analyses suggest that this value provides robust estimates (Table S1). For each combination of parameters, we measured the success of the approach (power) as the

proportion of times the confidence intervals around  $\gamma$  did not include 1.

The first round of analyses considered 30 sites per species, but 30 sites are rarely sufficient to detect weak declines (Sewell *et al.*, 2012; Kéry & Royle, 2016). To test how increasing the number of surveyed sites improves the detection of declines, we repeated analyses considering 60, 120, 240 and 500 sites.

## Simulations using dynamic models

The first simulations were generated using a static approach, in which abundance at a site  $i$  at time  $t_1$  and  $t_2$  were unrelated. To assess whether our results are affected by the way we generated data, we repeated analyses using a dynamic model. For each site, we first generated the initial abundance ( $N_{[i,t_1]}$ ) from a Poisson distribution with  $\lambda = \lambda_1$ . The decline  $D_i$  was then generated from a Poisson distribution with  $\lambda = D \times N_{[i,t_1]}$ . The final abundance was then calculated as:  $N_{[i,t_2]} = N_{[i,t_1]} - D_i$  (see Appendix S1 for the R script).

## Trade-offs between number of sites and number of surveys

We tested under which circumstances (e.g. for which species) it is better analyzing a few sites multiple times, or many sites with less surveys per site. We considered 30 sites monitored 7 times, versus 70 sites monitored three times. The total number of surveys (210) is constant in these two schemes. We thus compared the efficiency of  $N$ -mixture models in detecting declines under a range of abundances and detectabilities. This analysis was limited to the moderate declines (10–20%), as seven surveys on 30 sites almost always detected strong declines (see Results).

## Alternative monitoring schemes

First, we tested the impact of monitoring many small sub-sites. Instead of 30 sites with high mean abundance ( $\lambda_1$ ), we considered 120 sub-sites, each with mean abundance =  $\frac{1}{4}\lambda_1$ . This scenario mimics, for instance, surveying the same territory, but through 120 transects with length 250-m, instead of 30 transects with length 1-km.

Second, we tested whether combining sites from multiple syntopic species may improve performance. In previous approaches, we considered 30 surveys in relatively optimal sites, where the species is known to be abundant. However, multiple target species may co-occur, and individuals of a given species can be detected at some of the sites surveyed for other species. Instead of 30 sites with high mean abundance ( $\lambda_1$ ), we considered a total of 90 sites: 30 sites are monitored specifically for the target species and, at time  $t_1$ , have high abundance ( $\lambda_1$ ), while 60 additional sites are monitored for other species within the range of target species. Among the 60 sites, in 20 the species has high abundance ( $\lambda_1$ ), in 20 the species is rare (abundance =  $\frac{1}{4}\lambda_1$ ), and in 20 the species is absent. At time  $t_2$ , the species suffers an overall decline, as described for the standard approach. In this

latter analysis, models were built using zero-inflated Poisson (ZIP) distribution as in preliminary analyses ZIP models consistently showed lower Akaike’s information criterion than Poisson or negative binomial models.

**Heterogeneity of detection probability**

Monitoring is often performed under highly heterogeneous conditions, thus detection probability is rarely constant (Guilera-Arroita, 2017). We assessed whether our results are robust to heterogeneity of  $p$  across sites and surveys. We assumed that  $p$  was related to an environmental variable (in our example, ‘humidity’) and assumed that  $p$  increases with humidity, as often occurs with amphibians (Mazerolle *et al.*, 2007). For each combination of  $D$  and  $\lambda$ , we considered five values of mean detection probability ( $p_M$ ), and three scenarios of heterogeneity of detection: homogeneous detection ( $p$  ranging between  $0.918 \times p_M$  and  $1.04 \times p_M$ ); heterogeneous detection ( $p$  between  $0.639 \times p_M$  and  $1.22 \times p_M$ ) and highly heterogeneous detection ( $p$  between  $0.290 \times p_M$  and  $1.516 \times p_M$ ). In all scenarios,  $p$  followed a logistic curve with increasing values at increasing humidity, and mean  $p = p_M$ . Simulated data were then analyzed including humidity as detection covariate (see Appendix S1 for the R script, and Fig. S1 for the actual detection probabilities).

**Results**

**Estimates of population size**

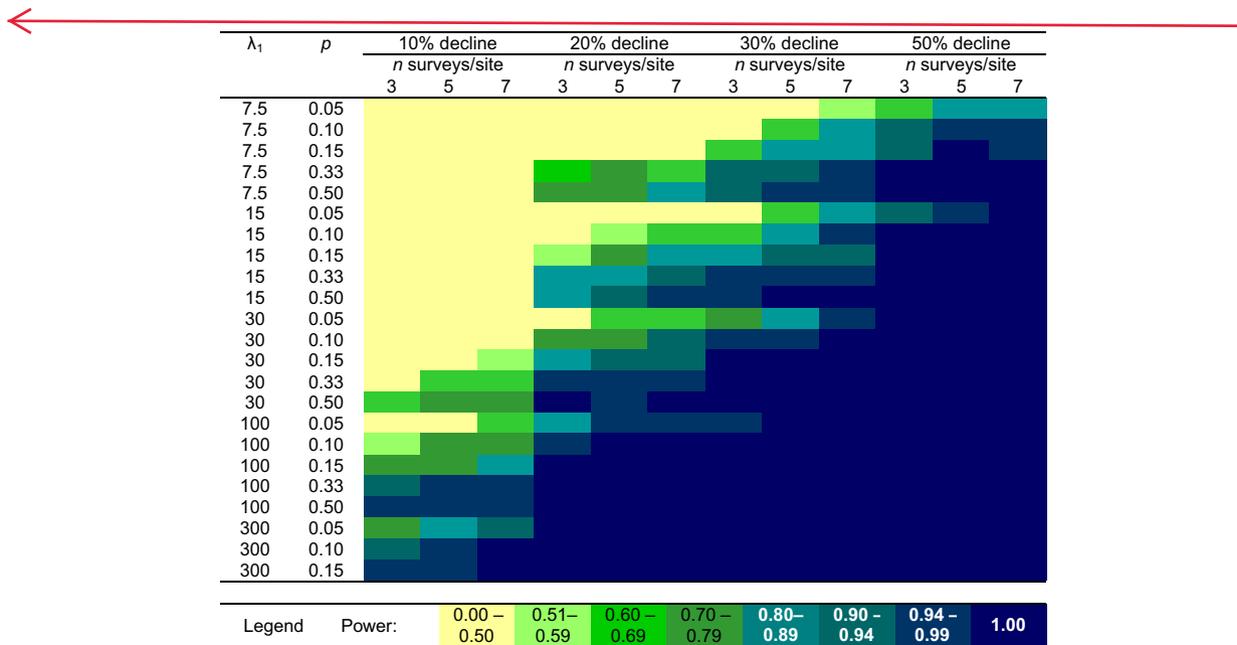
The relative bias in the estimation of population size was high (>50%) for species with low abundance, low detection

probability ( $p$ ), and sampled in few repeat visits (Fig. S2a–c). Bias decreased if  $p$  was high and if many surveys were performed (Fig. S2a–c). The correlation between true and estimated population size was weak (*c.* 0.35) if  $p$  was 0.05 and just three replicated surveys were performed, but correlation quickly increased when more surveys were performed, or if  $p$  increased, while it was essentially unrelated to mean population size (Fig. S2d–f).

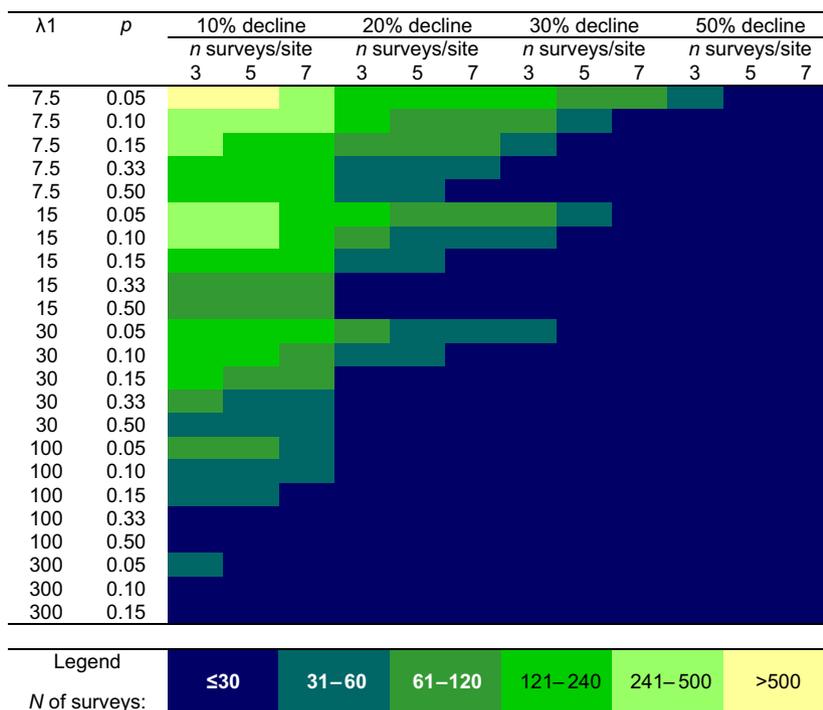
**Detection of decline**

The success of open-population  $N$ -mixture models in detecting declines was variable across species abundances,  $p$ , number of surveys per site and rate of decline (Fig. 2). Strong declines ( $\geq 30\%$ ) were successfully detected for nearly all the species, except in least abundant ones with very low  $p$  (Fig. 1). Weak declines (10–20%) were successfully detected only in species with high abundance and easy to detect (Fig. 1). Results did not depend on the way we simulated data, as conclusions remained nearly identical when we generated population abundance data using dynamic models (Table S2).

Increasing the number of sites quickly increased the power of analyses (i.e. the proportion of simulations where lambda was significantly smaller than 1; Fig. 2). For instance, if 120 sites were surveyed 5–7 times,  $N$ -mixture models were able to detect declines in most of cases, except with weak decline, low  $p$  and abundance (Table S3). Nevertheless, a very large number of sites (240–500) would require surveys if the aim is detecting a 10% decline in species with limited abundance and  $p \leq 0.1$  (Fig. 2).



**Figure 1** Success of open-population  $N$ -mixture models in detecting the decline of species with a range of mean abundance at the first monitoring ( $\lambda_1$ ) and detection probability ( $p$ ). The success is measured as the proportion of simulations in which models detected a significant species decline. For all combinations, we considered 30 sites.



**Figure 2** Number of sites that need to be surveyed, to detect a given decline with success  $\geq 80\%$ : results of simulations.  $\lambda_1$ : mean abundance at the first monitoring.

**Trade-offs between number of sites and number of surveys**

Performing seven surveys on 30 sites was preferable to performing three surveys on 70 sites only for species with low detectability ( $<0.1$ ). If detectability was  $\geq 0.1$ , it was preferable surveying many sites with less repeats (Fig. 3). If detectability was high and the species was abundant (e.g.  $\lambda = 100$  and  $p \geq 0.33$ ), the performance of the two approaches was similar, as both successfully detected declines in  $\sim 100\%$  of simulations (compare Fig. 1 with Table S4).

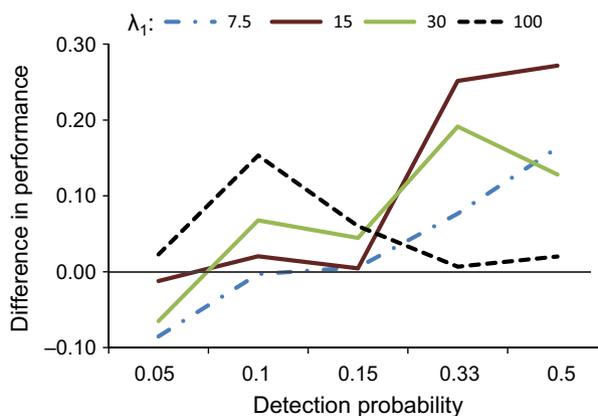
**Alternative monitoring schemes**

Monitoring 120 smaller sub-transects instead of 30 large sites slightly improved the success of analyses. The increase was particularly relevant for species with low abundance and high detectability (Fig. 4a–e).

Combining sites selected for the target species with sites monitored from other syntopic species generally improved the success of analyses (Fig. 4f–h). However, there were conditions under which the heterogeneous data approach did not show better performance, particularly if the target species was very abundant and detectability low (Fig. 4i,j).

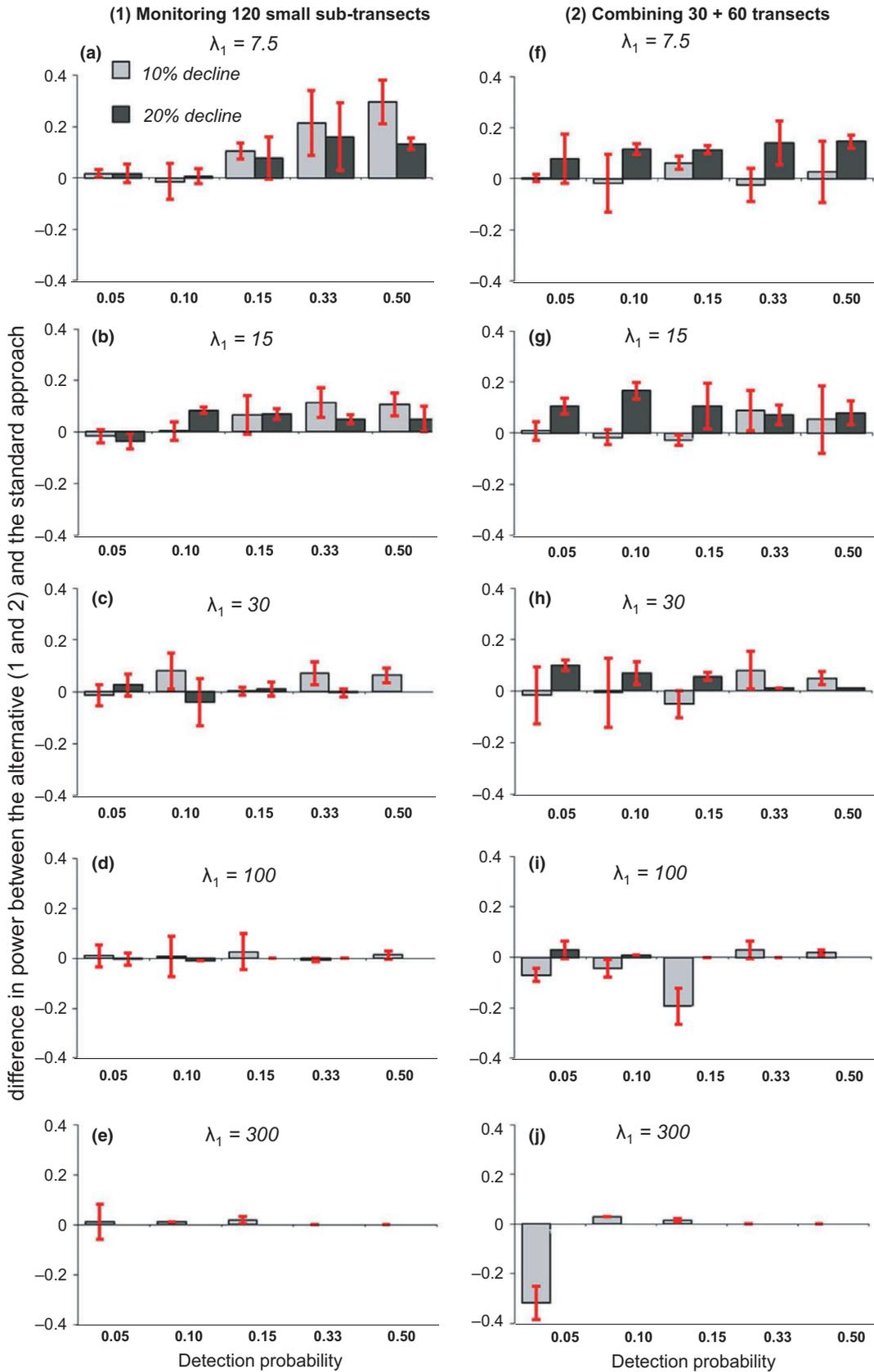
**Heterogeneity of detection probability**

Heterogeneity of  $p$  did not influence the success of models in detecting population declines (Fig. 5; Tables S7,S8). Even

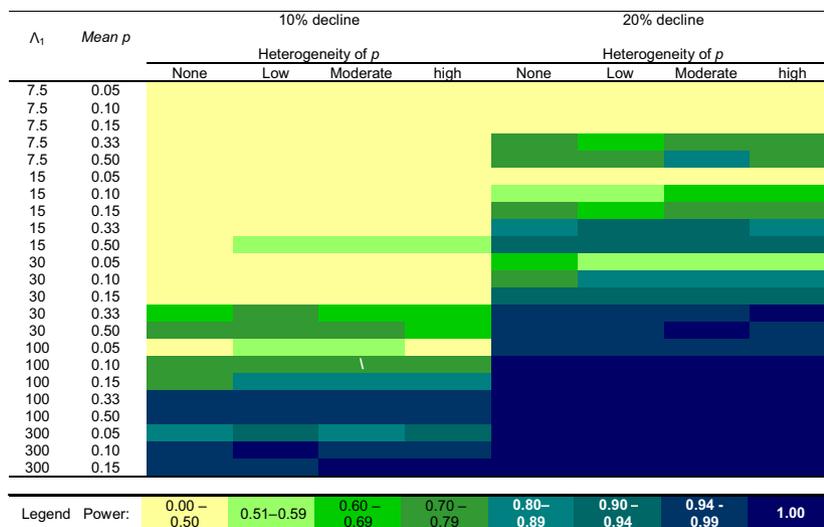


**Figure 3** Outcome of alternative approaches to surveys: differences between surveying 70 sites three times, and surveying 30 sites seven times. The plot represents the difference in power between models analyzing 70 sites three times, and models analyzing 30 sites seven times. Positive values are the conditions under which it is preferable surveying more sites with less surveys and vice-versa.  $\lambda_1$ : mean abundance at the first monitoring (see Table S4 for actual power values).

with the strongest heterogeneity of detectability (Fig. S1), the mean difference in performance between the analysis of data with constant  $p$  and heterogeneous  $p$  was just 0.4% (SE = 3%).



**Figure 4** Difference in power between the alternative and the standard approaches to surveys. (a–e): The alternative approach is surveying 120 small sites (initial abundance:  $\frac{1}{4}\lambda_1$ ). (f–j): The alternative approach is combining 30 targeted sites + 60 sites surveyed for syntopic species. The standard approach is surveying 30 large sites. If values are positive, the alternative approach is preferable to the standard approach and vice-versa. Each bar is the average of models analyzing monitoring schemes with 3, 5 and 7 surveys per site; error bars are standard deviations of these three models.  $\lambda_1$ : mean abundance at the first monitoring (see Tables S5,S6 for the actual power values).



**Figure 5** Effect of heterogeneity of detection probability on the success of  $N$ -mixture models in detecting species decline. The figure reports the results of simulations performing five surveys/site (see Tables S7,S8 for results with three and seven surveys/site).

## Discussion

An appropriate planning of surveys, optimizing the efforts, is often the key to detect population changes, if they occur (Reynolds *et al.*, 2011). This is critical for cryptic animals for which we need quantitative estimates of declines. Our analysis shows that, keeping constant sampling efforts, the capacity of monitoring schemes to accurately detect declines is strongly variable, being related to parameters such as local abundance and detectability (Fig. 1). Our results can be used to identify the optimal approach for the monitoring of a given species, if information on average abundance and detectability is available.

### Performance of $N$ -mixture models

Models showed a good performance in the estimation of population size and declines. Estimated population sizes were well correlated with true population size, and the difference between estimated and true values remained reasonable, except if detection probability was very low and only a few surveys per site were performed (Fig. S2). Open-population models thus successfully detected population declines with the majority of combinations of population size, detectability and severity of decline. The severity of decline was a major determinant of the capacity of models to detect them: strong

declines ( $\geq 30\%$ ) were almost always identified, even in the least detectable species.

Detection probability and mean abundance of species also were extremely important. As expected, detecting declines was particularly challenging in rare species with low detectability (Guillera-Arroita, 2012; Kroll *et al.*, 2015). In the most extreme scenario (7.5 individuals/site,  $p = 0.05$ ), models failed to detect weak declines, and their performance remained limited even with strong declines (Fig. 1). Such combination of low abundance and difficult detection is frequent in predators, such as snakes and some large lizards (Data S1). For these species, the expected count of individuals per survey is  $< 1$  (0.375 if  $p = 0.05$  and  $N = 7.5$ ; Kéry, 2010), thus in the majority of surveys no individuals are detected, making trend estimation difficult.

Our analysis might suffer some limitations. Results were robust to heterogeneity in detectability (Fig. 5), if the variables determining species detection in the different surveys are correctly specified into the models. However, in the real world model misspecification is possible, for instance if the variables influencing species detection are not recorded and integrated into models, and this might undermine the performance of models. Furthermore, all populations from the same simulation were drawn from Poisson distributions with constant lambda, but within a given species population size may vary over orders of magnitudes across the populations,

and strong heterogeneity of lambda among sites can reduce model performance (Fig. 4i,j; Guillera-Arroita *et al.*, 2014). Finally, the selection of sites (not just their number) is an additional important issue. Ideally, site selection should be spatially random to allow inference over broad scales, but this might be impossible for certain species. Additional studies should assess the effect of spatial selection of sites on trend estimation.

### How can we improve the detection of declines?

Several strategies can improve the detection of declines, some of which do not require a strong increase in resources.

- 1 *Surveying more sites with fewer surveys*: The trade-off between number of sites to be surveyed, and number of surveys per sites, is well known. If species have very low detection probability, it is better surveying more times a few sites, compared to surveying many sites a few times (MacKenzie *et al.*, 2006; Guillera-Arroita, 2012, 2017). Performing many surveys is particularly useful for species with low or intermediate abundances and difficult detection. However, the advantage of performing many surveys per site quickly vanishes in species that have both high abundance and high detection probabilities (Fig. 3), because in these cases the power of the estimator is high (>70%) even with a limited number of surveys (Fig. 1). In most of cases, if  $p > 0.1$  it is better allocating resources on the monitoring of many sites, even though this may require increasing travel costs.
- 2 *Surveying many small sites* instead of a few large sites may improve detection of trends for several species (Fig. 4a–d). The performance increase was greater for species with low abundance and  $p \geq 0.15$ . This probably occurs because such an approach increases sample size, without notably increasing the number of surveys without detections. Nevertheless, our analysis assumed that the short transects are independent. If a large number of small transects is obtained just by dividing long transects (e.g. transforming 1-km transects into four 250-m transects), sub-transects are not independent. This may violate statistical assumptions (pseudoreplication), and it is even possible that the same individual is counted in different transects (Chandler & Royle, 2013). It is thus important to ensure the independence of transects, or to adopt analytical frameworks allowing to take spatial dependence into account, such as multi-scale models (Chandler & Royle, 2013; Kroll *et al.*, 2015).
- 3 *Combining data from target and non-target sites*: We first assumed that, for each species, several sites where the species is relatively frequent were monitored. In the real world, multiple species can be detected in most of sites. For instance, in bird and reptile surveys, it is standard to map all species, and not just a single one, so it may be possible combining observations of co-occurring species. This has the advantage of only requiring that surveyors record more species, without additional costs. In several cases, the combined transect approach improved the

detection of declines, particularly for the least abundant species (Fig. 4f–h). However, this approach has drawbacks. First, if sites are selected to be representative for a given species, it is not *a priori* clear that they will also be representative for other species. Therefore, the selection of sites and of pairs of co-occurring species should be performed with care, on the basis of pilot surveys (Tulloch *et al.*, 2016). The combined method showed poor performance for species with highest abundance (Fig. 4i,j), perhaps because of the very high heterogeneity of abundance among sites.

- 4 *Improving detection probability*: The activity of most of animals is strongly tied to seasons and weather conditions, which determine detectability (Mazerolle *et al.*, 2007; McDiarmid *et al.*, 2012; Griffiths *et al.*, 2015). Focusing surveys in the periods when animals are more active improves their detection, thereby increasing model performance (Fig. 1). Nevertheless, in the real world detectability is not easily predictable *a priori* and it is influenced by site survey conditions. Pilot studies can allow measuring species detectability and then perform sampling under the best conditions. Spending more time at each site is an additional strategy to increase detection. The situation is particularly challenging if habitat features influence both species abundance and detection probability. For instance, lizard density may be higher in densely vegetated sites, but this reduces detection (Kéry, 2010). It is thus pivotal recording and integrating site and survey-covariate (e.g. day, weather conditions) into models (Schmidt, 2005; Mazerolle *et al.*, 2007; Kéry *et al.*, 2009; Kéry, 2010).

### Improving detection of declines: where should we increase resources?

The decision of where increasing resources is a multi-factorial decision, as it depends on environmental policy, funding availability, and on the need to optimize the available resources (Wintle *et al.*, 2010; Reynolds *et al.*, 2011). In the case of species listed under the Habitat Directive, Member States are committed to measure population trends of all the species. To achieve this target, each species should receive monitoring efforts sufficient to the detection of eventual declines. A potential target might be the detection of eventual declines  $\geq 20\%$  over 6 years with power  $> 80\%$ . In this case, species with limited detectability and abundance should be monitored intensively (see Fig. 2 for values). A first possibility is increasing the number of sites surveyed. Surveying 120 sites would allow detecting declines for most of species (Table S3), but detecting declines of species with very low abundance and detectability remains challenging even with hundreds sites. Furthermore, increasing the number of sites is not always feasible. Actually, the most endangered species often have very restricted distributions, thus limiting the number of potential sites. In this situation, the only strategy is increasing the number of surveys per site.

In simulations we assumed that populations are monitored during 2 years (Stoch & Genovesi, 2016), but monitoring during multiple years may improve the detection of trends.

A rotating panel design, where each site is surveyed in multiple years but not all the years, can be a good compromise between number of surveys per sites and reliability of trend estimation (McDonald, 2003). Furthermore, detecting declines is easier if larger populations are monitored, and if detection probabilities are higher. Resources may be allocated to surveying larger sites (e.g. 2-km transects instead of 1-km transects), as this would increase the number of individuals that may be contacted (Kery & Royle, 2016), or to spending longer time per site, thus enhancing detection.

## Conclusions

Even within a given class of vertebrates, species can have very heterogeneous lifestyles, life histories, abundances and activity periods, so it is impossible conceiving one single monitoring scheme valid for all the species. Instead, the available resources should be modulated to optimize the probability of detecting declines. Monitoring a few sites may be enough for abundant species with high detectability, and the saved resources may be allocated to increase efforts toward the rarest and least detectable species. Approaches that allow the detection of multiple species, such as visual transects, trapping or point counts, may be particularly useful, because can increase sample size for rare species.

Nevertheless, detecting declines provides limited information on the driving factors and on conservation strategies to prevent them. Instead of just monitoring the species, a management-targeted monitoring may be particularly effective to deliver conservation solutions. For instance, a large body of *a priori* information is available on the stressors that determine the trends of amphibians and reptiles. Integrating the surveillance of populations and their threatening factors can be challenging (Sutherland, Roy & Amano, 2015), but would be pivotal to develop more immediate and efficient conservation strategies (Nichols & Williams, 2006; Purse & Golding, 2015).

## Acknowledgments

The comments of two reviewers improved an early version of the paper. This study was stimulated by constructive discussions with many colleagues of the Societas Herpetologica Italica. GFF is a member of the LECA, which is part of Labex OSUG@2020 (Investissements d'avenir-ANR10 LABX56).

## References

- Bonardi, A., Manenti, R., Corbetta, A., Ferri, V., Fiacchini, D., Giovine, G., Macchi, S., Romanazzi, E., Soccini, C., Bottoni, L., Padoa Schioppa, E. & Ficetola, G.F. (2011). Usefulness of volunteer data to measure the large scale decline of 'common' toad populations. *Biol. Conserv.* **144**, 2328–2334.
- Chandler, R.B. & Royle, J.A. (2013). Spatially explicit models for inference about density in unmarked or partially marked populations. *Ann. Appl. Stat.* **7**, 936–954.
- Dail, D. & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* **67**, 577–587.
- Ellis, M.M., Ivan, J.S. & Schwartz, M.K. (2014). Spatially explicit power analyses for occupancy-based monitoring of wolverine in the U.S. Rocky Mountains. *Conserv. Biol.* **28**, 52–62.
- Ficetola, G.F., Cagnetta, M., Padoa-Schioppa, E., Quas, A., Razzetti, E., Sindaco, R. & Bonardi, A. (2014). Sampling bias inverts ecogeographical relationships in island reptiles. *Global Ecol. Biogeogr.* **23**, 1303–1313.
- Fiske, I. & Chandler, R. (2011). unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *J. Stat. Softw.* **43**, 1–23.
- Griffiths, R.A., Foster, J., Wilkinson, J.W. & Sewell, D. (2015). Science, statistics and surveys: a herpetological perspective. *J. Appl. Ecol.* **52**, 1413–1417.
- Guillera-Arroita, G. (2012). Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods Ecol. Evol.* **3**, 860–869.
- Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* **40**, 281–295.
- Guillera-Arroita, G., Ridout, M.-S. & Morgan, B.J.T. (2010). Design of occupancy studies with imperfect detection. *Methods Ecol. Evol.* **1**, 131–139.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., MacKenzie, D.I., Wintle, B.A. & McCarthy, M.A. (2014). Ignoring imperfect detection in biological surveys is dangerous: a response to 'Fitting and interpreting occupancy models'. *PLoS ONE* **9**, e99571.
- IUCN. (2001). *IUCN Red List categories and criteria: version 3.1*. Gland and Cambridge: IUCN.
- Kéry, M. (2010). *Introduction to WinBUGS for ecologists*. Burlington: Academic Press.
- Kery, M. & Royle, J.A. (2016). *Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS: vol. 1: prelude and static models*. Burlington: Academic Press.
- Kéry, M., Dorazio, R.M., Soldaat, L., Van Strien, A., Zuiderwijk, A. & Royle, J.A. (2009). Trend estimation in populations with imperfect detection. *J. Appl. Ecol.* **46**, 1163–1172.
- Kroll, A.J., Garcia, T.S., Jones, J.E., Dugger, K., Murden, B., Johnson, J., Peerman, S., Brintz, B. & Rochelle, M. (2015). Evaluating multi-level models to test occupancy state responses of plethodontid salamanders. *PLoS ONE* **10**, e0142903.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.A. & Hines, J.E. (2006). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Burlington: Academic Press.
- Mazerolle, M.J., Bailey, L.L., Kendall, W.L., Royle, J.A., Converse, S.J. & Nichols, J.D. (2007). Making great leaps

- forward: accounting for detectability in herpetological field studies. *J. Herpetol.* **41**, 672–689.
- McDiarmid, R.W., Foster, M.S., Guyer, C., Gibbons, J.W. & Chernoff, N. (2012). *Reptile biodiversity. Standard methods for inventory and monitoring*. Berkeley and Los Angeles: University of California Press.
- McDonald, T.L. (2003). Review of environmental monitoring methods: survey designs. *Environ. Monit. Assess.* **85**, 277–292.
- Nichols, J.D. & Williams, B.K. (2006). Monitoring for conservation. *Trends Ecol. Evol.* **21**, 668–673.
- Petrovan, S.O. & Schmidt, B.R. (2016). Volunteer conservation action data reveals large-scale and long-term negative population trends of a widespread amphibian, the common toad (*Bufo bufo*). *PLoS ONE* **11**, 12.
- Pollock, K.H., Nichols, J.D., Simons, T.R., Farnsworth, G.L., Bailey, L.L. & Sauer, J.R. (2002). Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* **13**, 105–119.
- Purse, B.V. & Golding, N. (2015). Tracking the distribution and impacts of diseases with biological records and distribution modelling. *Biol. J. Linn. Soc.* **115**, 664–677.
- Reynolds, J.H., Thompson, W.L. & Russell, B. (2011). Planning for success: identifying effective and efficient survey designs for monitoring. *Biol. Conserv.* **144**, 1278–1284.
- Royle, J.A. (2004). *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**, 108–115.
- Royle, J.A. & Dorazio, R.M. (2008). *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. London: Academic Press.
- Royle, J.A. & Nichols, J.D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology* **84**, 777–790.
- Schmidt, B.R. (2004). Declining amphibian populations: the pitfalls of count data in the study of diversity, distributions, dynamics, and demography. *Herpetol. J.* **14**, 167–174.
- Schmidt, B.R. (2005). Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **15**, 681–692.
- Sewell, D., Guillera-Arroita, G., Griffiths, R.A. & Beebee, T.J.C. (2012). When is a species declining? Optimizing survey effort to detect population changes in reptiles *PLoS ONE* **7**, e43387.
- Sillero, N., Campos, J., Bonardi, A., Corti, C., Creemers, R., Crochet, P.A., Isailović, J.C., Denoël, M., Ficetola, G.F., Gonçalves, J., Kuzmin, S., Lymberakis, P., de Pous, P., Rodríguez, A., Sindaco, R., Speybroeck, J., Toxopeus, B., Vieites, D.R. & Vences, M. (2014). Updated distribution and biogeography of amphibians and reptiles of Europe based on a compilation of countrywide mapping studies. *Amphibia-Reptilia* **35**, 1–31.
- Sindaco, R., Doria, G., Razzetti, E. & Bernini, F. (Eds). (2006). *Atlas of Italian amphibians and reptiles*. Firenze: Polistampa.
- Stoch, F. & Genovesi, P. (Eds). (2016). *Manuali per il monitoraggio di specie e habitat di interesse comunitario (Direttiva 92/43/CEE) in Italia: specie animali*. Roma: ISPRA.
- Sutherland, W.J., Roy, D.B. & Amano, T. (2015). An agenda for the future of biological recording for ecological monitoring and citizen science. *Biol. J. Linn. Soc.* **115**, 779–784.
- Tracewski, L., Butchart, S., Di Marco, M., Ficetola, G.F., Rondinini, C., Symes, A., Wheatley, H., Beresford, A. & Buchanan, G. (2016). Toward quantification of the impact of 21st century deforestation on the extinction risk of terrestrial vertebrates. *Conserv. Biol.* **30**, 1070–1079.
- Tulloch, A.I.T., Chadès, I., Dujardin, Y., Westgate, M.J., Lane, P.W. & Lindenmayer, D. (2016). Dynamic species co-occurrence networks require dynamic biodiversity surrogates. *Ecography* **39**, 1185–1196.
- Wintle, B.A., Runge, M.C. & Bekessy, S.A. (2010). Allocating monitoring effort in the face of unknown unknowns. *Ecol. Lett.* **13**, 1325–1337.
- Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends Ecol. Evol.* **16**, 446–453.
- Zipkin, E.F., Thorson, J.T., See, K., Lynch, H.J., Grant, E.H.C., Kanno, Y., Chandler, R.B., Letcher, B.H. & Royle, J.A. (2014). Modeling structured population dynamics using data from unmarked individuals. *Ecology* **95**, 22–29.
- Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Munkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schroder, B. & Grimm, V. (2010). The virtual ecologist approach: simulating data and observers. *Oikos* **119**, 622–635.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Figure S1.** Relationship between humidity and detection probability in simulations with heterogeneous detection probability.

**Figure S2.** Relative bias of population size estimated through *N*-mixture models, and correlations between true and estimated population sizes.

**Table S1.** Effect of varying values of the upper bound *K* on the parameters estimated.

**Table S2.** Results obtained using dynamic models to simulate populations.

**Table S3.** Results of simulations considering 120 sites.

**Table S4.** Results of simulations considering 70 sites surveyed three times per season.

**Table S5.** Results of simulations considering 120 small transects.

**Table S6.** Results of simulations 30 targeted transects + 60 occasional transects surveyed for synanthropic species.

**Table S7.** Results of simulations with heterogeneous detection probability (three surveys/site).

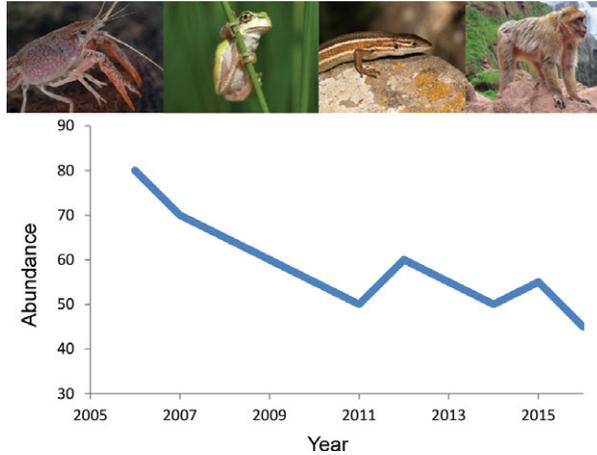
**Table S8.** Results of simulations with heterogeneous detection probability (seven surveys/site).

**Data S1.** Examples of species with abundance and detection probability values within the range of values tested in simulations.

**Appendix S1.** Scripts and table for the generation of all the results.

# Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only. It will not be published as part of main article.



How should we allocate the limited resources available for monitoring, to maximize the probability of detecting species declines? Through the analysis of simulated data, we assessed the performance of multiple monitoring scenarios across species with different features, and showed that the probability of detecting declines increases with species detection probability, in species showing high local abundance, and in monitoring schemes surveying many populations. Surveying efforts should be modulated across species to maximize the effectiveness of monitoring, and we provide quantitative values on how the number of surveys and the number of sites to be surveyed can be assigned to different species.