

1 **Optimizing monitoring schemes to detect trends in abundance over**  
2 **broad scales**

3

4 Gentile Francesco Ficetola<sup>1,2,6\*</sup>, Antonio Romano<sup>3,6</sup>, Sebastiano Salvidio<sup>4,6</sup>, Roberto  
5 Sindaco<sup>5,6</sup>

6

7 <sup>1</sup> Univ. Grenoble Alpes, CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble,  
8 France

9 <sup>2</sup> Department of Environmental Science and Policy, Università degli Studi di Milano. Via  
10 Celoria 26, 20133 Milano, Italy

11 <sup>3</sup> Consiglio Nazionale delle Ricerche, Istituto di Biologia Agroambientale e Forestale (CNR–  
12 IBAF), Area di Ricerca Roma 1, Via Salaria km 29,300, 00015 Monterotondo (RM), Italy

13 <sup>4</sup> DISTAV – Dipartimento di Scienze della Terra dell'Ambiente e della Vita, Università di  
14 Genova, Corso Europa 26, I-16132 Genova, Italy

15 <sup>5</sup> I.P.L.A., Istituto per le Piante da Legno e l'Ambiente, corso Casale 476, I-10132 Torino

16 <sup>6</sup> Societas Herpetologica Italica, c/o Museo Regionale di Scienze Naturali, via Giolitti 36,  
17 Torino, Italy.

18

19 \* corresponding author. francesco.ficetola@gmail.com

20

21 **Abstract**

22

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

46

47 **Keywords:** abundance, species decline, detection probability, *N*-mixture models,

48 optimization, sub-transects, visual transects.

49

## 50 **Introduction**

51  
52 The detection and quantification of trends in species abundance (e.g. declines or increases) is  
53 essential to ascertain conservation status, to identify threatened species, and to assess the  
54 effectiveness of conservation strategies (Yoccoz, Nichols & Boulinier, 2001; Reynolds,  
55 Thompson & Russell, 2011). For instance, in IUCN redlists, quantitative measures of decline  
56 are essential to qualify a species as threatened under criterion A, and under several other sub-  
57 criteria (IUCN, 2001). Measuring trends of protected species is also a major commitment  
58 requested by environmental regulations. The EU Habitats Directive (Directive 92/43/EEC) is  
59 perhaps the most important tool for biodiversity conservation in Europe, as it protects >1200  
60 species and their habitats. The Directive requires that, every six years, Member States shall  
61 draw up a report on the conservation status of listed species, including measures of trends of  
62 populations, to evaluate the efficiency of protection measures undertaken. Trends of species'  
63 ranges, as well as habitat extent changes, can be estimated by remote-sensing data (Tracewski  
64 *et al.*, 2016), but remote-sensing cannot capture the effect of local pressures determining the  
65 decline of populations (e.g. diseases, disturbance, exploitation, modification of  
66 microhabitats...), nor provide data on population changes (changes in the number of  
67 populations or in the number of individuals per population).

68         Obtaining quantitative measures of species trends over broad areas is challenging, as it  
69 generally requires repeated surveys of many sites covering significant portions of species'  
70 ranges. Volunteers can provide data at low cost, thus helping the measurement of the trends of  
71 widespread species living in easily accessible regions (Kéry *et al.*, 2009; Bonardi *et al.*, 2011;  
72 Sewell *et al.*, 2012; Griffiths *et al.*, 2015; Petrovan & Schmidt, 2016). However, there are  
73 cases in which volunteer data are not available, particularly for species requiring specific  
74 monitoring protocols or living in difficult to access areas. For instance, the Italian reptiles and

75 amphibians have some of the highest levels of richness and endemism in Europe (Sillero *et*  
76 *al.*, 2014). The most abundant and widespread species range through the whole territory, thus  
77 enabling monitoring through volunteers (Bonardi *et al.*, 2011), while many endemic and  
78 threatened species are restricted to small insular or mountainous areas with limited  
79 accessibility (Sindaco *et al.*, 2006), where consistent volunteer sampling is unlikely. Under  
80 these conditions, there is the need to develop a pragmatic and efficient strategy for  
81 monitoring, optimizing the limited available resources.

82         When the detection probability is  $<1$  (as usually occurs in animal populations),  
83 estimating parameters such as presence/absence, abundance or trends may be problematic and  
84 requires taking into account the imperfect detection (Pollock *et al.*, 2002; Schmidt, 2004;  
85 Mazerolle *et al.*, 2007). However, formal approaches exist for estimating abundance from  
86 repeated counts in fixed sites, without marking individuals to identify the individuals (Royle  
87 & Nichols, 2003; Royle, 2004). These approaches, named hierarchical or  $N$ -mixture models,  
88 allow the estimation of population size and abundance trends for species that are imperfectly  
89 detected (Kéry *et al.*, 2009; Dail & Madsen, 2011; Zipkin *et al.*, 2014), and are thus  
90 appropriate to detect population declines. Such models are able to take into account the high  
91 heterogeneity of data collected over broad scales by a large number of observers, which is  
92 typical of many monitoring schemes, and are thus highly promising for population and even  
93 species assessments (Kéry *et al.*, 2009; Griffiths *et al.*, 2015). Nevertheless, as for any  
94 statistical approach, the power of  $N$ -mixture models is influenced by factors such as effect  
95 size and sample size. The analysis of synthetic data simulating ecological processes is a  
96 powerful framework, which allows evaluating the effect of sampling and analytical methods  
97 (Guillera-Arroita, Ridout & Morgan, 2010; Zurell *et al.*, 2010; Ficetola *et al.*, 2014). A few  
98 studies tried to identify under which conditions (number of sites, number of surveys) it is  
99 possible to successfully detect population declines (e.g. Guillera-Arroita, 2012; Sewell *et al.*,

100 2012; Kroll *et al.*, 2015), but these analysis mostly focused on occupancy models. Occupancy  
101 models do not provide explicit estimates of abundance (but see Royle & Nichols, 2003; Ellis,  
102 Ivan & Schwartz, 2014), and thus mostly deal with changes in number of populations, instead  
103 of estimating trends in abundance.

104         Given the monitoring required by the Habitat Directive, the Italian Herpetological  
105 Society was tasked by government agencies with identifying monitoring methodologies for  
106 the >70 Italian species of amphibians and reptiles of European concern (Stoch & Genovesi,  
107 2016), but discussion among experts lead to different proposals with regard to the best  
108 strategy for monitoring. The aim of this study was to identify the most effective monitoring to  
109 detect abundance trends (e.g. declines) through *N*-mixture models in each of multiple species  
110 with very heterogeneous features. When planning a monitoring scheme, the allocation of  
111 resources must be optimized to maximize the probability to address the specific management  
112 questions (Nichols & Williams, 2006; Wintle, Runge & Bekessy, 2010; Reynolds *et al.*, 2011)  
113 and, if the monitoring scheme includes multiple species, it is important to identify the optimal  
114 strategy for each of them (Guillera-Arroita *et al.*, 2010; Guillera-Arroita, 2012). The Italian  
115 amphibians and reptiles have very different features, some species being locally abundant and  
116 conspicuous, and others rare and difficult to detect (Supplementary Methods). We therefore  
117 assessed how species features (abundance and detection probability) and monitoring structure  
118 (e.g. number of surveys, number of sites) influence the detection of declines. Specifically, we  
119 asked: *i*) How do species abundance and detection probability determine our ability to detect  
120 declines? *ii*) How many sites should be surveyed for each target species? *iii*) For species with  
121 given abundance and detectability, is it better monitoring a few sites with repeated surveys, or  
122 many sites with a few surveys per site? *iv*) Is it better concentrating efforts on a few, large  
123 sites (e.g. long transects or large plots), or increasing sample size and analysing many small  
124 sub-transects (or sub-plots)? *v*) Under which conditions is it possible combining surveys on

125 multiple species to improve efficiency? Although analyses were initially motivated by the  
126 need of a national plan for the monitoring of Italian amphibians and reptiles (Stoch &  
127 Genovesi, 2016), our simulations represent realistic sampling covering a wide range of  
128 scenarios of abundance, detection probability, declines and monitoring schemes. Therefore  
129 our recommendations are applicable to the planning of monitoring of a large number of taxa  
130 and regions.

131

## 132 **Methods**

133

### 134 **Rationale for models**

135

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

150 trend estimates: a) each of the sites is subdivided in 4 smaller sub-sites; b) combining the sites  
151 monitored for multiple species, to increase the sample size.

152

### 153 **Simulations and $N$ -mixture models**

154

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

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

172 First, we estimated the relative bias of  $N$ -mixture models in estimating the actual  
173 abundance of the species. For each simulation, we run single-season  $N$ -mixture models with  
174 Poisson error distribution on the data at  $t_1$  (Royle, 2004; fitted using pcount in unmarked:



175 Fiske & Chandler, 2011), and the population size estimated from the posterior distributions of  
176 the random variables using empirical Bayes methods (Royle & Dorazio, 2008). Relative bias  
177 was calculated as the median of  $\text{abs}[(\text{estimated population size} - \text{true population size})/\text{true}$   
178  $\text{population size}]$ , across all simulations. We also calculated the median Pearson's correlation  
179 between estimated and true population size.

180         Second, we tested whether  $N$ -mixture models for open populations (Dail & Madsen,  
181 2011) can successfully estimate the decline between  $t_1$  and  $t_2$ . In models, we assumed that  
182  $N_{[i,t_2]}=N_{[i,t_1]}\times\text{gamma}$ , where gamma is the rate of population change (Fiske & Chandler, 2011).  
183 We considered that a model successfully estimated the decline if gamma was significantly  $<1$   
184 (at  $\alpha=0.05$ ); models were built using the `pcountOpen` function in `unmarked`, assuming Poisson  
185 error. In  $N$ -mixture models, the default values of the upper bound used to approximate an  
186 infinite summation in the likelihood can provide inaccurate estimates- In each simulation we  
187 used as upper bound (70+the maximum observed species abundance), since preliminary  
188 analyses suggest that this value provides robust estimates (Table S1). For each combination of  
189 parameters, we measured the success of the approach (power) as the proportion of times the  
190 confidence intervals around gamma did not include 1.

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

195

## 196 **Simulations using dynamic models**

197

198 The first simulations were generated using a static approach, in which abundance at a site  $i$  at  
199 time  $t_1$  and  $t_2$  were unrelated. To assess whether our results are affected by the way we

200 generated data, we repeated analyses using a dynamic model. For each site, we first generated  
201 the initial abundance ( $N_{[i,1]}$ ) from a Poisson distribution with  $\lambda=\lambda_1$ . The decline  $D_i$  was then  
202 generated from a Poisson distribution with  $\lambda=D\times N_{[i,1]}$ . The final abundance was then  
203 calculated as:  $N_{[i,2]}=N_{[i,1]}-D_i$ . See appendix S2 for the R script.

204

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

206

207 We tested under which circumstances (e.g. for which species) it is better analyzing a few sites  
208 multiple times, or many sites with less surveys per site. Instead of 30 sites monitored 3-7  
209 times, we considered 70 sites monitored 3 times. The total number of surveys (210) is  
210 constant in these two schemes. We thus compared the efficiency of  $N$ -mixture models in  
211 detecting declines under a range of abundances and detectabilities. This analysis was limited  
212 to the moderate declines (10-20%), as 7 surveys on 30 sites almost always detected strong  
213 declines (see results).

214

### 215 **Alternative monitoring schemes**

216

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

221         Second, we tested whether combining sites from multiple syntopic species may  
222 improve performance. In previous approaches, we considered 30 surveys in relatively optimal  
223 sites, where the species is known be abundant. However, multiple target species may co-  
224 occur, and individuals of a given species can be detected at some of the sites surveyed for

225 other species. Instead of 30 sites with high mean abundance ( $\lambda_1$ ), we considered a total of 90  
226 sites: 30 sites are monitored specifically for the target species and, at time  $t_1$ , have high  
227 abundance ( $\lambda_1$ ), while 60 additional sites are monitored for other species within the range of  
228 target species. Among the 60 sites, in 20 the species has high abundance ( $\lambda_1$ ), in 20 the species  
229 is rare (abundance= $1/4\lambda_1$ ), in 20 the species is absent. At time  $t_2$ , the species suffers an overall  
230 decline, as described for the standard approach. In this latter analysis, models were built using  
231 zero-inflated Poisson distribution (ZIP) as in preliminary analyses ZIP models consistently  
232 showed lower Akaike's Information Criterion than Poisson or negative binomial models.

233

### 234 **Heterogeneity of detection probability**

235

236 Monitoring is often performed under highly heterogeneous conditions, thus detection  
237 probability is rarely constant (Guillera-Arroita, 2017). We assessed whether our results are  
238 robust to heterogeneity of  $p$  across sites and surveys. We assumed that  $p$  was related to an  
239 environmental variable (in our example, "humidity") and assumed that  $p$  increases with  
240 humidity, as often occurs with amphibians (Mazerolle *et al.*, 2007). For each combination of  
241  $D$  and  $\lambda$ , we considered five values of mean detection probability ( $p_M$ ), and three scenarios of  
242 heterogeneity of detection: homogeneous detection ( $p$  ranging between  $0.918 \times p_M$  and  
243  $1.04 \times p_M$ ); heterogeneous detection ( $p$  between  $0.639 \times p_M$  and  $1.22 \times p_M$ ); highly heterogeneous  
244 detection ( $p$  between  $0.290 \times p_M$  and  $1.516 \times p_M$ ). In all scenarios,  $p$  followed a logistic curve  
245 with increasing values at increasing humidity, and mean  $p = p_M$ . Simulated data were then  
246 analysed including humidity as detection covariate (see Appendix S2 for the R script, and Fig.  
247 S1 for the actual detection probabilities).

248

## 249 **Results**

250

## 251 **Estimates of population size**

252

253 The relative bias in the estimation of population size was high (>50%) for species with low  
254 abundance, low detection probability ( $p$ ), and sampled in few repeat visits (Fig. S2a-c). Bias  
255 decreased if  $p$  was high and if many surveys were performed (Fig. S2a-c). The correlation  
256 between true and estimated population size was weak (approx. 0.35) if  $p$  was 0.05 and just  
257 three replicated surveys were performed, but correlation quickly increased when more surveys  
258 were performed, or if  $p$  increased, while it was essentially unrelated to population size (Fig.  
259 S2d-f).

260

## 261 **Detection of decline**

262

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

270

271 Increasing the number of sites quickly increased the power of analyses (i.e., the  
272 proportion of simulations where  $\lambda$  was significantly smaller than 1; Fig. 2). For  
273 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

274 very large number of sites (240-500) would require surveys if the aim is detecting a 10%  
275 decline in species with limited abundance and  $p \leq 0.1$  (Fig. 2).

276

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

278

279 Performing 7 surveys on 30 sites was preferable to performing 3 surveys on 70 sites only for  
280 species with low detectability ( $< 0.1$ ). If detectability was  $\geq 0.1$ , it was preferable surveying  
281 many sites with less repeats (Fig. 3). If detectability was high and the species was abundant  
282 (e.g.  $\lambda = 100$  and  $p \geq 0.33$ ) the performance of the two approaches was similar, as both  
283 successfully detected declines in  $\sim 100\%$  of simulations (compare Fig. 1 with Table S4).

284

### 285 **Alternative monitoring schemes**

286

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

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

294

### 295 **Heterogeneity of detection probability**

296

297 Heterogeneity of  $p$  did not influence the success of models in detecting population declines  
298 (Fig. 5, Tables S7-S8). Even with the strongest heterogeneity of detectability (Fig. S1), the

299 mean difference in performance between the analysis of data with constant  $p$  and  
300 heterogeneous  $p$  was just 0.4% (SE=3%).

301

## 302 **Discussion**

303

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

311

### 312 **Performance of N-mixture models**

313

314 Models showed a good performance in the estimation of population size and declines.  
315 Estimated population sizes were well correlated with true population size, and the difference  
316 between estimated and true values remained reasonable, except if detection probability was  
317 very low and only a few surveys per site were performed (Fig. S2). Open-population models  
318 thus successfully detected population declines with the majority of combinations of  
319 population size, detectability and severity of decline. The severity of decline was a major  
320 determinant of the capacity of models to detect them: strong declines ( $\geq 30\%$ ) were almost  
321 always identified, even in the least detectable species.

322 Detection probability and mean abundance of species also were extremely important.

323 As expected, detecting declines was particularly challenging in rare species with low

324 detectability (Guillera-Arroita, 2012; Kroll *et al.*, 2015). In the most extreme scenario (7.5  
325 individuals/site,  $p=0.05$ ), models failed to detect weak declines, and their performance  
326 remained limited even with strong declines (Fig. 1). Such combination of low abundance and  
327 difficult detection is frequent in predators, such as snakes and some large lizards  
328 (Supplementary Methods). For these species, the expected count of individuals per survey is  
329  $<1$  (0.375 if  $p=0.05$  and  $N=7.5$ ; Kéry, 2010), thus in the majority of surveys no individuals are  
330 detected, making trend estimation difficult.

331         Our analysis might suffer some limitations. Results were robust to heterogeneity in  
332 detectability (Fig. 5), if the variables determining species detection in the different surveys are  
333 correctly specified into the models. However, in the real world model misspecification is  
334 possible, for instance if the variables influencing species detection are not recorded and  
335 integrated into models, and this might undermine the performance of models. Furthermore, all  
336 populations from the same simulation were drawn from Poisson distributions with constant  
337 lambda, but within a given species population size may vary over orders of magnitudes across  
338 the populations, and strong heterogeneity of lambda among sites can reduce model  
339 performance (Fig. 4i-j; Guillera-Arroita *et al.*, 2014). Finally, the selection of sites (not just  
340 their number) is an additional important issue. Ideally, site selection should be spatially  
341 random to allow inference over broad scales, but this might be impossible for certain species.  
342 Additional studies should assess the effect of spatial selection of sites on trend estimation.

343

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

345

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

348           *i) Surveying more sites with fewer surveys.* The trade-off between number of sites to be  
349 surveyed, and number of surveys per sites, is well known. If species have very low detection  
350 probability, it is better surveying more times a few sites, compared to surveying many sites a  
351 few times (MacKenzie *et al.*, 2006; Guillera-Arroita, 2012; Guillera-Arroita, 2017).  
352 Performing many surveys is particularly useful for species with low or intermediate  
353 abundances and difficult detection. However, the advantage of performing many surveys per  
354 site quickly vanishes in species that have both high abundance and high detection  
355 probabilities (Fig. 3), because in these cases the power of the estimator is high (>70%) even  
356 with a limited number of surveys (Fig. 1). In most of cases, if  $p > 0.1$  it is better allocating  
357 resources on the monitoring of many sites, even though this may require increasing travel  
358 costs.

359           *ii) Surveying many small sites* instead of a few large sites may improve detection of  
360 trends for several species (Fig. 4a-d). The performance increase was greater for species with  
361 low abundance and  $p \geq 0.15$ . This probably occurs because such an approach increases sample  
362 size, without notably increasing the number of surveys without detections. Nevertheless, our  
363 analysis assumed that the short transects are independent. If a large number of small transects  
364 is obtained just by dividing long transects (e.g. transforming 1-km transects into four 250-m  
365 transects), sub-transects are not independent. This may violate statistical assumptions  
366 (pseudoreplication), and it is even possible that the same individual is counted in different  
367 transects (Chandler & Royle, 2013). It is thus important to ensure the independence of  
368 transects, or to adopt analytical frameworks allowing to take spatial dependence into account,  
369 such as multi-scale models (Chandler & Royle, 2013; Kroll *et al.*, 2015).

370           *iii) Combining data from target and non-target sites.* We first assumed that, for each  
371 species, several sites where the species is relatively frequent were monitored. In the real  
372 world, multiple species can be detected in most of sites. For instance, in bird and reptile



373 surveys, it is standard to map all species, and not just a single one, so it may be possible  
374 combining observations of co-occurring species. This has the advantage of only requiring that  
375 surveyors record more species, without additional costs. In several cases, the combined  
376 transect approach improved the detection of declines, particularly for the least abundant  
377 species (Fig. 4f-h). However, this approach has drawbacks. First, if sites are selected to be  
378 representative for a given species, it is not a priori clear that they will also be representative  
379 for other species. Therefore the selection of sites and of pairs of co-occurring species should  
380 be performed with care, on the basis of pilot surveys (Tulloch *et al.*, 2016). The combined  
381 method showed poor performance for species with highest abundance (Fig. 4i-j), perhaps  
382 because of the very high heterogeneity of abundance among sites.

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

395

396 **Improving detection of declines: where should we increase resources?**

397

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

412 In simulations we assumed that populations are monitored during two years (Stoch &  
413 Genovesi, 2016), but monitoring during multiple years may improve the detection of trends. A  
414 rotating panel design, where each site is surveyed in multiple years but not all the years, can  
415 be a good compromise between number of surveys per sites and reliability of trend estimation  
416 (McDonald, 2003). Furthermore, detecting declines is easier if larger populations are  
417 monitored, and if detection probabilities are higher. Resources may be allocated to surveying  
418 larger sites (e.g. 2-km transects instead of 1-km transects), as this would increase the number  
419 of individuals that may be contacted (Kery & Royle, 2016), or to spending longer time per  
420 site, thus enhancing detection.

421

422 **Conclusions**

423

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

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

440

#### 441 **Acknowledgements**

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

446

#### 447 **References**

448

449 Bonardi A., Manenti R., Corbetta A., Ferri V., Fiacchini D., Giovine G., Macchi S.,  
450 Romanazzi E., Soccini C., Bottoni L., Padoa Schioppa E. & Ficetola G.F. (2011).  
451 Usefulness of volunteer data to measure the large scale decline of "common" toad  
452 populations. *Biol. Conserv.* **144**, 2328-2334.

453 Chandler R.B. & Royle J.A. (2013). Spatially explicit models for inference about density in  
454 unmarked or partially marked populations. *Annals of Applied Statistics* **7**, 936-954.

455 Dail D. & Madsen L. (2011). Models for estimating abundance from repeated counts of an  
456 open metapopulation. *Biometrics* **67**, 577-587.

457 Ellis M.M., Ivan J.S. & Schwartz M.K. (2014). Spatially Explicit Power Analyses for  
458 Occupancy-Based Monitoring of Wolverine in the U.S. Rocky Mountains. *Conserv. Biol.*  
459 **28**, 52-62.

460 Ficetola G.F., Cagnetta M., Padoa-Schioppa E., Quas A., Razzetti E., Sindaco R. & Bonardi A.  
461 (2014). Sampling bias inverts ecogeographical relationships in island reptiles. *Global*  
462 *Ecol. Biogeogr.* **23**, 1303–1313.

463 Fiske I. & Chandler R. (2011). unmarked: An R package for fitting hierarchical models of  
464 wildlife occurrence and abundance. *J. Stat. Softw.* **43**, 1-23.

465 Griffiths R.A., Foster J., Wilkinson J.W. & Sewell D. (2015). Science, statistics and surveys: a  
466 herpetological perspective. *J. Appl. Ecol.* **52**, 1413-1417.

467 Guillera-Arroita G. (2012). Designing studies to detect differences in species occupancy:  
468 power analysis under imperfect detection. *Methods in Ecology and Evolution* **3**, 860-869.

469 Guillera-Arroita G. (2017). Modelling of species distributions, range dynamics and  
470 communities under imperfect detection: advances, challenges and opportunities.  
471 *Ecography* **40**, 281-295.

472 Guillera-Arroita G, Ridout M.-S. & Morgan B.J.T. (2010). Design of occupancy studies with  
473 imperfect detection. *Methods in Ecology and Evolution* **1**, 131-139.

474 Guillera-Arroita G, Lahoz-Monfort J.J., MacKenzie D.I., Wintle B.A. & McCarthy M.A.  
475 (2014). Ignoring imperfect detection in biological surveys is dangerous: A response to  
476 'Fitting and interpreting occupancy models'. *Plos One* **9**, e99571.

477 IUCN (2001). IUCN Red List Categories and Criteria: Version 3.1. IUCN: Gland, Switzerland  
478 and Cambridge, UK.

479 Kery M. & Royle J.A. (2016). *Applied Hierarchical Modeling in Ecology: Analysis of*  
480 *distribution, abundance and species richness in R and BUGS: Volume 1:Prelude and*  
481 *Static Models*. Academic Press: Burlington, MA.

482 Kéry M. (2010). *Introduction to WinBUGS for ecologists*. Academic Press: Burlington, MA.

483 Kéry M., Dorazio R.M., Soldaat L., Van Strien A., Zuiderwijk A. & Royle J.A. (2009). Trend  
484 estimation in populations with imperfect detection. *J. Appl. Ecol.* **46**, 1163-1172.

485 Kroll A.J., Garcia T.S., Jones J.E., Dugger K., Murden B., Johnson J., Peerman S., Brintz B.  
486 & Rochelle M. (2015). Evaluating Multi-Level Models to Test Occupancy State  
487 Responses of Plethodontid Salamanders. *PLoS ONE* **10**, e0142903.

488 MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.A. & Hines J.E. (2006).  
489 *Occupancy estimation and modeling: inferring patterns and dynamics of species*  
490 *occurrence*. Academic Press: Burlington, MA.

491 Mazerolle M.J., Bailey L.L., Kendall W.L., Royle J.A., Converse S.J. & Nichols J.D. (2007).  
492 *Making great leaps forward: Accounting for detectability in herpetological field studies.*  
493 *J. Herpetol.* **41**, 672-689.

494 McDiarmid R.W., Foster M.S., Guyer C., Gibbons J.W. & Chernoff N. (2012). *Reptile*  
495 *biodiversity. Standard methods for inventory and monitoring*. University of California  
496 Press: Berkley and Los Angeles, CA.

497 McDonald T.L. (2003). Review of Environmental Monitoring Methods: Survey Designs.  
498 *Environ. Monit. Assess.* **85**, 277-292.

499 Nichols J.D. & Williams B.K. (2006). Monitoring for conservation. *Trends Ecol. Evol.* **21**,  
500 668-673.

501 Petrovan S.O. & Schmidt B.R. (2016). Volunteer Conservation Action Data Reveals Large-  
502 Scale and Long-Term Negative Population Trends of a Widespread Amphibian, the  
503 Common Toad (*Bufo bufo*). *Plos One* **11**, 12.

504 Pollock K.H., Nichols J.D., Simons T.R., Farnsworth G.L., Bailey L.L. & Sauer J.R. (2002).  
505 Large scale wildlife monitoring studies: statistical methods for design and analysis.  
506 *Environmetrics* **13**, 105-119.

507 Purse B.V. & Golding N. (2015). Tracking the distribution and impacts of diseases with  
508 biological records and distribution modelling. *Biol. J. Linn. Soc.* **115**, 664-677.

509 Reynolds J.H., Thompson W.L. & Russell B. (2011). Planning for success: Identifying  
510 effective and efficient survey designs for monitoring. *Biol. Conserv.* **144**, 1278-1284.

511 Royle J.A. (2004). *N*-mixture models for estimating population size from spatially replicated  
512 counts. *Biometrics* **60**, 108-115.

513 Royle J.A. & Nichols J.D. (2003). Estimating abundance from repeated presence-absence data  
514 or point counts. *Ecology* **84**, 777-790.

515 Royle J.A. & Dorazio R.M. (2008). Hierarchical Modeling and Inference in Ecology: The  
516 analysis of data from populations, metapopulations and communities. Academic Press:  
517 Lonon.

518 Schmidt B.R. (2004). Declining amphibian populations: The pitfalls of count data in the study  
519 of diversity, distributions, dynamics, and demography. *Herpetol. J.* **14**, 167-174.

520 Schmidt B.R. (2005). Monitoring the distribution of pond-breeding amphibians when species  
521 are detected imperfectly. *Aquatic Conserv: Mar. Freshw. Ecosyst.* **15**, 681-692.

522 Sewell D., Guillera-Arroita G., Griffiths R.A. & Beebee T.J.C. (2012). When Is a Species  
523 Declining? Optimizing Survey Effort to Detect Population Changes in Reptiles. *Plos One*  
524 7.

525 Sillero N., Campos J., Bonardi A., Corti C., Creemers R., Crochet P.A., Isailović J.C., Denoël  
526 M., Ficetola G.F., Gonçalves J., Kuzmin S., Lymberakis P., de Pous P., Rodríguez A.,  
527 Sindaco R., Speybroeck J., Toxopeus B., Vieites D.R. & Vences M. (2014). Updated  
528 distribution and biogeography of amphibians and reptiles of Europe based on a  
529 compilation of countrywide mapping studies. *Amphibia-Reptilia* **35**, 1-31.

530 Sindaco R., Doria G., Razzetti E. & Bernini F., eds (2006). Atlas of Italian Amphibians and  
531 Reptiles. Polistampa: Firenze.

532 Stoch F. & Genovesi P., eds (2016). Manuali per il monitoraggio di specie e habitat di  
533 interesse comunitario (Direttiva 92/43/CEE) in Italia: specie animali. ISPRA: Roma.

534 Sutherland W.J., Roy D.B. & Amano T. (2015). An agenda for the future of biological  
535 recording for ecological monitoring and citizen science. *Biol. J. Linn. Soc.* **115**, 779-784.

536 Tracewski L., Butchart S., Di Marco M., Ficetola G.F., Rondinini C., Symes A., Wheatley H.,  
537 Beresford A. & Buchanan G. (2016). Toward quantification of the impact of 21st century  
538 deforestation on the extinction risk of terrestrial vertebrates. *Conserv. Biol.* **30**, 1070-  
539 1079.

540 Tulloch A.I.T., Chadès I., Dujardin Y., Westgate M.J., Lane P.W. & Lindenmayer D. (2016).  
541 Dynamic species co-occurrence networks require dynamic biodiversity surrogates.  
542 *Ecography*, n/a-n/a.

543 Wintle B.A., Runge M.C. & Bekessy S.A. (2010). Allocating monitoring effort in the face of  
544 unknown unknowns. *Ecol. Lett.* **13**, 1325-1337.

545 Yoccoz N.G., Nichols J.D. & Boulinier T. (2001). Monitoring of biological diversity in space  
546 and time. *Trends Ecol. Evol.* **16**, 446-453.

547 Zipkin E.F., Thorson J.T., See K., Lynch H.J., Grant E.H.C., Kanno Y., Chandler R.B.,  
548 Letcher B.H. & Royle J.A. (2014). Modeling structured population dynamics using data  
549 from unmarked individuals. *Ecology* **95**, 22-29.

550 Zurell D., Berger U., Cabral J.S., Jeltsch F., Meynard C.N., Munkemuller T., Nehrbass N.,  
551 Pagel J., Reineking B., Schroder B. & Grimm V. (2010). The virtual ecologist approach:  
552 simulating data and observers. *Oikos* **119**, 622-635.

553

554



555 **Figure legends.**

556

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

561

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

564

565 **Figure 3.** Outcome of alternative approaches to surveys: differences between surveying 70  
566 sites 3 times, and surveying 30 sites 7 times. The plot represents the difference in power  
567 between models analyzing 70 sites 3 times, and models analyzing 30 sites 7 times. Positive  
568 values are the conditions under which it is preferable surveying more sites with less surveys  
569 and vice-versa.  $\lambda_1$ : mean abundance at the first monitoring. See Table S4 for actual power  
570 values.

571

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

580

581 **Figure 5.** Effect of heterogeneity of detection probability on the success of  $N$ -mixture models  
582 in detecting species decline. The figure reports the results of simulations performing 5 surveys  
583 / site. See Tables S7-S8 for results with 3 and 7 surveys / site.

584

585

586 **Supporting Information**

587 Additional Supporting Information may be found in the online version of this article.

588

589 **Supplementary methods.** : Examples of species with abundance and detection probability  
590 values within the range of values tested in simulations

591 **Supplementary references**

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

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

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

595 **Table S4.** Results of simulations considering 70 sites surveyed 3 times per season.

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

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

599 **Table S7.** Results of simulations with heterogeneous detection probability (3 surveys / site).

600 **Table S8.** Results of simulations with heterogeneous detection probability (7 surveys / site).

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

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

605

606 **Appendix S2.** Scripts and table for the generation of all the results