



HAL
open science

Analyzing patterns in population dynamics using repeated population surveys with three types of detection data

Guillaume Péron, Mathieu Garel

► **To cite this version:**

Guillaume Péron, Mathieu Garel. Analyzing patterns in population dynamics using repeated population surveys with three types of detection data. *Ecological Indicators*, 2019, 106, pp.105546. 10.1016/j.ecolind.2019.105546 . hal-02309872

HAL Id: hal-02309872

<https://cnrs.hal.science/hal-02309872v1>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **ANALYZING PATTERNS IN POPULATION DYNAMICS USING REPEATED**
2 **POPULATION SURVEYS WITH THREE TYPES OF DETECTION DATA**

3

4 Guillaume Péron (1)*

5 Mathieu Garel (2)

6 1. Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive
7 UMR5558, F-69622 Villeurbanne, France

8 2. Office National de la Chasse et de la Faune Sauvage, Direction de la Recherche et de
9 l'Expertise, Unité Ongulés Sauvages, 5 allée de Bethléem, Z.I. Mayencin, 38610 Gières,
10 France.

11 * corresponding author: guillaume.peron@univ-lyon1.fr

12

13 **Running title:** Repeated population surveys

14

15

16 **Highlights**

17 • **We generalize distance sampling to analyze multi-year multi-site population surveys**
18 **when sampling protocols vary across years and sites, using multiple observers and**
19 **time to detection in addition to distance information.**

20 • **Standard distance sampling may misrepresent population trends in the presence of**
21 **temporal variation in the availability to detection.**

22 • **The estimation of availability to detection is improved by combining multiple data**
23 **types.**

24 • **The new framework is costly in terms of number of parameters to estimate and**
25 **computing time, but compatible with the logistics of typical ungulate population**
26 **surveys.**

27

28

29

30 **ABSTRACT**

31 To facilitate the use of population counts as an index of population change, we describe a
32 generalization of the distance sampling methodology to analyze, in addition to distance to the
33 observer, two other ways to estimate imperfect detection probability: multiple observers and
34 time-to-detection, in a flexible manner, meaning that not all sites or years need to have distance
35 information or be surveyed in the same way every year. We also account for the effect of
36 partially-observed individual covariates, to account for the effect of group size on detection
37 probability. Finally, we separate the probability of availability to detection from the probability of
38 detection itself. We perform a thorough, illustrated assessment of the pros and cons of this
39 framework with simulations and real case studies. First, we compare to simple linear models,
40 illustrating the magnitude of the bias caused by imperfect detection. Second, we compare to
41 standard distance sampling, illustrating the bias caused by variation in the probability of
42 availability to detection. However, the availability to detection was weakly identifiable, meaning
43 that the ability to separate it from detection probability, and therefore debias the trend estimate,
44 depended on the data configuration. Combining distance with multiple observers and with time-
45 to-detection solved the weak identifiability in an applied case study. We recommend using both
46 the type of analysis we showcase, and a simple regression of the population count against time.
47 Discrepancies between results from simple and complex analyses can help identify sources of
48 bias in the former and loss of precision in the latter within the logistical constraints of local
49 wildlife management schemes.

50 **Key-words:** capture-recapture; demography; distance sampling; imperfect detection; indicator of
51 ecological change;

52

53 INTRODUCTION

54 The way animal populations change through time is an essential part of environmental
55 assessments, from local stock management schemes to global biodiversity indices. Population
56 counts often constitute the base data for these assessments. Yet population counts are well-known
57 to yield a flawed picture of population dynamics because of confounding factors such as
58 imperfect detection and counting errors (Anderson, 2003; Engeman, 2005; Gerrodette, 1987;
59 Harris, 1986; Link and Sauer, 1998). A broad range of methods have been proposed to overcome
60 this issue (Williams et al., 2002). Our first objective herein is to quickly review these methods
61 and some of the aspects we view as shortcomings. Second, we address those shortcomings, by
62 assembling together several add-on features that improve the performance of distance sampling
63 (Buckland et al., 2007, 1993). More precisely, we devise a version of distance sampling where
64 multiple observers can document the detection process independently (Alldredge et al., 2008;
65 Conn et al., 2012; Nichols et al., 2000), where each counting session can be divided into
66 secondary sessions (Alldredge et al., 2007; Amundson et al., 2014; Chandler et al., 2011), and
67 where availability to detection is modelled separately from detection itself (Burnham, 1993;
68 Chandler et al., 2011), thereby introducing a “robust design” (Kendall et al., 1997) philosophy to
69 distance sampling. We focus on studies that monitor population trends across a few locations over
70 the long term, as opposed to one-off surveys of numerous locations, and aim to document the
71 optimal sampling design and the risk of flawed inference when not accounting for confounding
72 factors when estimating population trends.

73 However, complex models tend to have low statistical power (lower precision) and to
74 exhibit estimation issues when applied to sparse datasets, meaning that special care needs to be
75 taken at the sampling design stage. In particular, we demonstrate a case of weak identifiability,
76 that is a case where the parameters are in theory all separately estimable, but their relative

77 contributions to the variance in the data becomes impossible to separate as the data get sparser
78 (Auger-Méthé et al., 2016; Barker et al., 2018; Fan et al., 2018; Garrett and Zeger, 2000). A
79 straightforward example of weak identifiability is when attempting to discriminate two categories
80 of individuals based on their size. The discriminatory power (model identifiability) weakens as
81 the difference between the two categories decreases below to the within-category variance, i.e.,
82 the parameter identifiability depends on the biological properties of the system (Garrett and
83 Zeger, 2000). In our case, the issue affected the separation of availability to detection and
84 detection when available, with consequences for the estimation of population trends when
85 availability was either very variable over time or negatively correlated to detection.

86 **A QUICK REVIEW OF THE METHODS TO ANALYZE PATTERNS IN POPULATION** 87 **DYNAMICS USING COUNT DATA**

88 **The Index of population size methodology (IPS)**

89 Hereafter the acronym “IPS” refers to methodologies that infer patterns in population dynamics
90 using the expected count, i.e., the product between the population abundance and the probability
91 of detection. Some IPS methods consist in averaging the count over several replicates, i.e., they
92 “average out” the sampling variance around the expected count (Loison et al., 2006). These
93 methods assume that the expected detection probability is the same everywhere and every time,
94 and that most of the noise around the expected count is caused by counting errors and other
95 stochastic, constant-mean processes. Alternatively, one may rely on linear models of the count
96 across space and time. Linear predictors and random effects would then control for factors of
97 variation in detection probability, such as observer proficiency, vegetation type, or weather (Link
98 and Sauer, 1998), thereby relaxing the assumption that the expected detection probability is the
99 same everywhere and every time.

100 The main issue with the otherwise simple and effective IPS approach is that, if a factor
101 jointly influences population abundance and detection probability, it will not be possible to tease
102 apart these two influences (Anderson, 2003). Furthermore, the factors of variation in detection
103 probability may not be *a priori* known and quantified, preventing their inclusion as explanatory
104 variables. Lastly, count data are often very noisy, in which case IPS methods can become
105 unreliable or request too many replicates to be tractable (Gerrodette, 1987; Harris, 1986).

106 **Population reconstruction from individual-based data**

107 Because of the above shortcomings of the IPS approach, researchers have historically preferred to
108 “reconstruct” the population dynamics from estimates of vital rates, such as survival and
109 fecundity (Caswell, 2001; Williams et al., 2002; see also Besbeas et al., 2002). In this approach,
110 one uses individual-based data to compute, each year, the balance between the births and deaths,
111 and thereby the population growth rate, yielding an index of population abundance relative to the
112 abundance at the start of the study. The main advantage of this approach is the ability to
113 investigate individual and environmental variation in vital rates, and thereby obtain realistic
114 models of population dynamics likely to yield reliable short-term predictions (Gauthier et al.,
115 2016). The main issue is the cost and field-intensiveness, and the fact that the reconstructed
116 abundance is conditioned on the initial population estimate, i.e., it is an index relative to the
117 initial population abundance.

118 **Unmarked methods**

119 To avoid the shortcomings of the IPS and the cost and field-intensiveness of population
120 reconstruction, the “unmarked” philosophy (Fiske and Chandler, 2011) is currently gaining in
121 popularity. This refers to methods that do not require individual-based data from marked or
122 otherwise recognizable individuals, but that still separate the variance in the count data into a

123 sampling (detection) and a process (population dynamics) components. Distance sampling
124 (Buckland et al., 1993) is the first of these “unmarked” methodologies to have been widely used
125 for abundance and population trend estimation. In distance sampling, the decline in recorded
126 abundance with distance to the observer is attributed to a decline in detection probability, and
127 leveraged to correct the raw count data for imperfect detection. Another seminal model
128 underlying the unmarked philosophy is the N-mixture model (Royle, 2004). In the N-mixture
129 model, the sampling variance across replicated counts is modelled as the outcome of a binomial
130 process whose success rate is the individual detection rate.

131 Perhaps because they were so successful that they have been tested in a wide variety of
132 situations, these two approaches have revealed a few shortcomings. In particular, the N-mixture
133 approach may yield overestimated or infinite estimates of population size when detection
134 probability is small or when there are few replicates (Couturier et al., 2013; Dennis et al., 2015;
135 Veech et al., 2016). Recently, Barker et al. (2018) explained this pattern as a case of weak
136 identifiability. When the data are sparse, solutions with large abundance and low detection are as
137 likely as solutions with low abundance and large detection. In addition, the N-mixture model
138 requires that the detection probability is constant across replicate counts. This arguably prevents
139 the accurate description of the sampling process (Barker et al., 2018), even if the issue could in
140 theory be resolved by adding an additional hierarchical layer in the model (Zhao and Royle,
141 2019). Lastly, the N-mixture model fitting procedure in the Bayesian framework is sensitive to
142 the arbitrary choice of a maximum potential population size, requiring some biological insight
143 that may not always exist prior to the analysis (Couturier et al., 2013; Dennis et al., 2015).

144 Now regarding the distance sampling methodology, one of the lingering issues is that
145 crypsis and associated behaviors, vertical movements such as diving or climbing trees, and
146 temporary emigration out the survey area leads some individuals to be temporally unavailable to

147 detection. They are still part of the population, but their detection probability is temporarily zero.
148 Buckland et al. (1993) introduced the familiar g_0 term to describe this availability probability.
149 This parameter must however be documented separately, for example with telemetry data
150 (Couturier et al., 2013; Marques et al., 2013), which can however be quite costly and field-
151 intensive. In addition, distance sampling assumes that animal occurrences are equally likely at
152 any point in the study area, and in particular that the animals do not avoid the observer's location.
153 If that assumption is not met, the estimated detection function does not monotonically decrease
154 with distance from the observer nor start at $g_0 = 1$ (Borchers and Cox, 2017). This discrepancy
155 can be accommodated by combining the analysis of forward and perpendicular distances in
156 transect-based distance sampling (Borchers and Cox, 2017). However, this type of improvement
157 to the basic distance sampling framework is not always easy to implement in the field. The
158 alternative solution, that we will further develop, is to combine distance with additional
159 "detection data" from double observer protocols (Borchers et al., 2006; Sollmann et al., 2015) or
160 time-to-detection protocols (Amundson et al., 2014). Lastly, another lingering criticism of
161 distance sampling is that for a long time, software implementations were only geared towards
162 obtaining snapshots of the population abundance, not monitoring fluctuations in abundance over
163 multiple years or sites. In particular, the software did not facilitate the borrowing of information
164 across years and sites.

165 **OUR MODEL**

166 The model was motivated by surveys of mountain ungulate populations in France, i.e., gregarious
167 herbivorous large mammals that live in rough terrain with impaired observer visibility, that are
168 surveyed on a yearly basis, from the ground, at a few representative locations, initially to monitor
169 how the populations recovered from historical over-harvesting, now mostly to adaptively manage

170 their harvest and monitor the effect of epizootics. Because we ended up assembling in a flexible
171 way many of the model features that we reviewed above, we expect our framework to be relevant
172 in other situations as well. We first review the three types of “detection data” that we consider,
173 then we describe the likelihood function that allows their joint analysis, a few necessary post-hoc
174 manipulations to compute derived quantities, and finally we thoroughly discuss sampling design
175 optimization, weak identifiability, and statistical power, using application cases and simulations.

176 **Three types of detection data for unmarked animals**

177 The first type of detection data is distance to the observer – our model is a generalization of the
178 distance sampling model. In our implementation, distance may be recorded exactly, or binned
179 into classes of approximate distance. Importantly, when counting animals that are grazing the
180 distant opposite slope of a valley, distance is not always relevant as an information about
181 detection probability, i.e., the visibility is sometimes good enough that all the animals have
182 almost the same detection probability. Therefore, it is interesting to be able to combine distance
183 sampling with other sources of information about detection, in a flexible way that allows the joint
184 analysis of locations where distance is the main source of information about detection, and
185 locations where distance conveys little information.

186 The second type of detection data comes from the multiple-observer protocol (Borchers
187 et al., 2006; Nichols et al., 2000). For each detected individual or group of individuals, the series
188 of detection or non-detection by several observers generates an history of detection akin to a
189 capture-recapture history. Distance then becomes an individual covariate associated to each
190 individual capture-recapture history. In a nutshell, the proportion of observers that detected an
191 individual informs the detection probability of that individual, and this can be averaged across
192 individuals for more reliable inference. Importantly, we need to consider the risk that observers

193 influence each other (Borchers et al., 2006), e.g., by noticing when the others take out their
194 notebook or look intensively in a given direction. For this reason, we advocate (and we
195 implement in our model) a removal design for the multiple-observer protocol (Nichols et al.,
196 2000). We establish an order among the observers. Observer $n+1$ can only add new detections
197 that observer n did not make. In addition to avoiding positive observer bias, the removal design
198 requires less post-session communication between observers than the full multiple-observer
199 protocol and is thus more straightforward to implement.

200 The third and last type of detection data is generated by a time-to-detection protocol
201 (Alldredge et al. 2007; a.k.a. removal sampling protocol *sensu* Fiske and Chandler 2011). For this
202 protocol, we assume that the time to detection scales to the instant detection probability. In
203 practice, we may discretize the detection process by dividing the count period into secondary
204 occasions. Then, the series of detections and non-detections during the secondary occasions
205 constitutes a capture-recapture history for each detected individual, similar to the robust design
206 with within-session closure assumption (Kendall et al., 1997). However, once an individual has
207 been detected once, its probability of detection is drastically improved because the observers now
208 know that this individual is present and roughly where it is. For this reason, we also implement a
209 removal design for the time-to-detection protocol.

210 *In summary*, we record the first secondary occasion at which an individual is observed,
211 the first observer in an ordered series who recorded it, and at which distance. But we can make do
212 with just one or two of these information bits.

213 **Group size**

214 Because mountain ungulates (our motivation for the new development) often live in groups, the
215 statistical unit in our model is the group of animals, or the cluster *sensu* Buckland et al. (1993).

216 One of our concerns is the effect of group size on detection probability, and in particular the way
 217 in which covariation between abundance and group size may flaw the IPS methodology. In other
 218 words, if group size increases with abundance (Pépin and Gerard, 2008; Toïgo et al., 1996), and
 219 detection probability increases with group size, the observed population growth rate may be
 220 artificially inflated, potentially leading to over-optimistic management decisions. Each detected
 221 group is described by two group covariates: the group size and the distance to the observer. The
 222 group size data is considered error-free; there is no counting error on individual groups, or partial
 223 availability of groups. To deal with counting errors or partial availability of groups, see Clement
 224 et al. (2017), but this feature is not supported in our framework.

225 **Model likelihood**

226 We denote θ the set of model parameters (Table 1) and \mathbf{Y} the detection data. \mathbf{Y} is stratified across
 227 K sites, T years, $U_{k,t}$ within-year visits to site k in year t , $V_{k,t,u}$ robust design-style secondary
 228 occasions within visit u to site k in year t , and $O_{k,t,u}$ observers. As noted above, $U_{k,t}$, $V_{k,t,u}$, and
 229 $O_{k,t,u}$ can change across sites, years, and visits, allowing for a flexible study design. For example
 230 $O_{k,t,u} = 1$ means that only one observer participated in the survey of site k , year t , and visit u .
 231 The likelihood $L(\theta|\mathbf{Y})$ describes the probability to record \mathbf{Y} as a function of θ . For each detected
 232 group i , we know the site k , the year t , the visit u , the secondary session v_i , the observer o_i , the
 233 distance d_i , and the group size g_i . From these data we can compute the probability $P_{k,t,u,i}$ that the
 234 group was detected, as the product of four terms: the probability that the group was available for
 235 detection, the probability that it was not detected until observer o_i , the probability that observer
 236 o_i did not detect it until subsession v_i , and the probability that the observer o_i eventually detected
 237 it during subsession v_i .

Eq. 1	$ \begin{aligned} P_{k,t,u,i} = & \overbrace{\varphi_{k,t,u}}^{\text{Available to detection}} \cdot \overbrace{\left[\prod_{o=1}^{o_i-1} (1 - p_{k,t,u,o}(g_i, d_i))^{v_i} \right]}^{\text{Not detected until observer } o_i} \\ & \cdot \underbrace{\left[(1 - p_{k,t,u,o_i}(g_i, d_i))^{v_i-1} \right]}_{\text{Not detected until subsession } v_i} \cdot \underbrace{p_{k,t,u,o_i}(g_i, d_i)}_{\text{Detected by } o_i \text{ at } v_i} \end{aligned} $
-------	---

238 The product between the first pair of brackets is replaced by a one if $o_i = 1$. All the notation is
 239 summarized in Table 1.

240 The product of all the $P_{k,t,u,i}$ terms corresponds to the overall probability to detect the
 241 groups that were detected, in the way they were detected. Then we need to account for the groups
 242 that were not detected. This is the only place where the population abundance enters the
 243 likelihood. The challenge is however that the group size and distance from the observer are,
 244 obviously, not known for the groups that were not detected. As is routinely done in this type of
 245 situation, we tackled this as a simple extrapolation problem, by assuming that non-detected
 246 groups were drawn from the same stochastic model as detected groups, but that they were on
 247 average farther and smaller than detected groups. We introduced the distribution of distances to
 248 the observer, denoted $Pr(d|k)$, and the distribution of group sizes, denoted $Pr(g|k, t, u)$. In the
 249 present implementation, $Pr(d|k)$ only depended on the configuration of the site. We informed it
 250 by a separate field of view analysis in a GIS software. For $Pr(g|k, t, u)$, based on
 251 recommendations by Ver Hoef & Boveng (2007) and on the observation that there was an excess
 252 of solitary animals relative to the negative-binomial distribution, we used a one-inflated negative-
 253 binomial distribution of group sizes. We included the three parameters of that distribution in the
 254 list of parameters to be estimated (Table 1).

255 Lastly, we implemented two ways to model the relationship between distance and
 256 detection probability. First, as is often the case in practice (Miller, 2015), the link between

257 detection probability and distance could follow a half-normal function. The spread parameter,
 258 a.k.a. half-detection distance, denoted $D_{k,t,o,u}$, was made to vary log-linearly with group size.
 259 Alternatively, we also implemented a histogram-like shape, i.e., a piecewise staircase function. In
 260 this case, the effect of group size on detection probability was additive to the effect of distance on
 261 the logit-log scale. In both cases, the result was the function $p_{k,t,u,o}(g, d)$ giving the site-, year-,
 262 visit-, and observer-specific detection probability as a function of group size and distance to the
 263 observer. With all this notation, we can then write the probability that one group went undetected
 264 as:

Eq. 2

$$Q_{k,t,u} = \overbrace{(1 - \varphi_{k,t,u})}^{\text{Group was not available}} + \overbrace{\varphi_{k,t,u} \cdot \int_g \int_d \left[\prod_{o=1}^{O_{k,t,u}} (1 - p_{k,t,u,o}(g_i, d_i))^{V_{k,t,u}} \right] \Pr(g|k, t, u) \Pr(d|k)}^{\text{Group was available but not detected}} dd dg$$

265 The integration over all possible group sizes and distances to the observer addresses the fact that
 266 the group size and the distance to the observer are not known but are drawn from the same
 267 distribution as the detected groups, after correcting for detection biases. In practice we computed
 268 this integral using a numerical quadrature (a.k.a. Riemann sum approximation). The probability
 269 that the total number of groups in site k during year t is $N_{k,t}$ can then be expressed as a binomial
 270 law, with number of trials $N_{k,t}$, number of successes $N_{k,t} - C_{k,t,u}$ where $C_{k,t,u}$ is the number of
 271 detected groups during visit u , and success probability $Q_{k,y,u}$. The complete joint likelihood over
 272 all sites, years, and visits is then finally:

Eq. 3

$$L(\boldsymbol{\theta}|\mathbf{Y}) \propto \prod_{k,t,u} \left[\overbrace{\left(\prod_{i=1}^{C_{k,t,u}} P_{k,t,u,i} \cdot \Pr(g_i|k, t, u) \right)}^{\text{Detected groups}} \frac{\overbrace{N_{k,t}!}^{\text{Undetected groups}}}{C_{k,t,u}! (N_{k,t} - C_{k,t,u})!} Q_{k,y,u}^{N_{k,t} - C_{k,t,u}} \right]$$

273 Throughout, detection and availability probabilities can be made to vary with site-
 274 specific covariates (e.g., elevation, land ownership), visit-specific covariates (e.g., cloud cover,

275 temperature), linear temporal trends across years, and site- and time- random effects. Random
 276 effects are however not made available in the enclosed R-package (but see cat application case
 277 below).

278 Our model is a generalization of distance sampling because if we remove the multiple
 279 observer and time-to-detection information ($O_{k,t,u} = V_{k,t,u} = 1$), if we fix all the $\varphi_{k,t,u}$ to one,
 280 and if we remove all the dependencies on g , we arrive at a likelihood of the form explained by
 281 Buckland, Rexstad, Marques, & Oedekoven (2015). By contrast, our model does not belong to
 282 the N-mixture class of models because the binomial error structure applies *within*, not *across* sites
 283 and visits.

284 To obtain the maximum-likelihood estimates of the model parameters, we find the
 285 minimum of $-\log L(\theta|Y)$. For that optimization we recommend the genetic algorithm with
 286 derivatives (Mebane and Sekhon, 2011), because in our experience there are many local minima
 287 in the negative log-likelihood. The preferred combination of model features should be selected
 288 using the Akaike Information Criterion (Burnham and Anderson, 2002), although to our
 289 knowledge there are no goodness-of-fit tests readily available for this type of model.

290 **Post-hoc manipulations**

291 The above model fitting procedure yields an estimate for the number of groups $N_{k,t}$. To compute
 292 the population abundance, denoted $M_{k,t}$, we multiplied the number of groups by the expected
 293 group size, corrected for detection biases, using the following formula:

$$\text{Eq. 4} \quad \hat{M}_{k,t} = \max_{u=1 \dots U_{k,t}} \left(\sum_{i=1}^{C_{k,t,u}} g_i \right) + \left(\hat{N}_{k,t} - \max_{u=1 \dots U_{k,t}} C_{k,t,u} \right) \frac{\sum_{g=1}^{+\infty} (g \cdot \widehat{\text{Pr}}(g|k, t, u) \cdot \hat{R}_{k,t}(g))}{\sum_{g=1}^{+\infty} (\widehat{\text{Pr}}(g|k, t, u) \cdot \hat{R}_{k,t}(g))}$$

294 $\max_{u=1\dots U_{k,t}} \left(\sum_{i=1}^{C_{k,t,u}} g_i \right)$ is the maximum number of individuals counted in site k during year t .
 295 $R_{k,t}(g)$ is the probability of not detecting a group of size g but of unknown distance to the
 296 observer. $R_{k,t}(g)$ is computed with an equation similar to Eq. 2. In practice, the sum over g was
 297 stopped after a large g chosen so that $\widehat{\Pr}(g|k, t, u) \cdot \widehat{R}_{k,t}(g)$ was negligible.

298 To estimate temporal trends in population abundance, we *a posteriori* regressed $\widehat{M}_{k,t}$
 299 against year t . We considered the random effect of site k on the intercept, and we weighed the
 300 Poisson-distributed regression by the inverse of the sampling variance of $\widehat{M}_{k,t}$. The slope of the
 301 regression represents the log-linear temporal increase or decrease in abundance. Tools for model
 302 building, model fitting, and post-processing are provided in the R-package *chamois* for R
 303 (Supplementary Data file).

304 SIMULATION STUDIES

305 Demonstrating bias in simpler methods

306 For this section, we designed a scenario specifically to challenge the IPS methodology and fully
 307 illustrate its shortcomings. At the start of a 6-year period, 240 animals were equally distributed
 308 across 8 separate sites. The abundance decreased in a similar fashion in all sites, reaching a total
 309 of 80 animals at the end of the 6 years. Over the 6 years, the average detection probability
 310 increased. The half-detection distance increased linearly from 150 to 665m and mean group size
 311 increased linearly from 1.7 to 3.2, while the log-scale effect of group size on the half-detection
 312 distance was +0.5. By contrast, the availability probability decreased from 0.80 to 0.70, which
 313 partly compensated the increase in detection probability. Each year, each site was visited 3 times
 314 by 2 observers. The 8 sites were treated as spatial replicates in the analysis.

315 These parameters values were purposely chosen so that the expected population count
316 slightly increased over the years, whereas the actual population size decreased. Accordingly, the
317 IPS methodology failed to detect the underlying population decrease (Table 2).

318 This scenario was also expected to challenge the N-mixture approach, because the non-
319 independence of animals in groups and the two-step detection process (availability and detection)
320 violated the binomial variance assumption. In addition, the simulated counts were quite small
321 especially at the end of the simulation, which the relatively large simulated effort (24 replicates
322 per year) may not adequately compensate for. We tentatively analyzed the simulated datasets with
323 the N-mixture methodology. We used the `unmarked` package for R (Fiske & Chandler, 2011), and
324 specifically the option `siteCovs` of the function `unmarkedFramePCount` to code for year
325 effects in the routine `pcount`. This way we directly estimated the temporal trend in abundance
326 as part of the list of parameters of the N-mixture model. The performances of the N-mixture were
327 slightly improved compared to the IPS method, but still featured a large proportion of type I and
328 type II errors (Table 2). Type II errors (false negatives) likely stemmed from the poor fit of the
329 model to the data, and in particular the fact that we specified a model that aggregated the effects
330 of the temporal variation in detection, in group size, and in availability to detection, instead of
331 separating them. Type I errors (false positives) likely stemmed from the occurrence of unrealistic
332 estimates due to the identifiability issues that we reviewed above.

333 The simulation scenario was also expected to challenge the standard distance sampling
334 methodology, because the probability of detection at distance 0 was below zero and varied over
335 time. Nevertheless, we tentatively implemented distance sampling using the `Distance` package for
336 R (Miller, 2015), and more precisely the `ds` function, with default options for the shape of the
337 decrease in detection with distance, and using the `region.table` option to code for the

338 different years, the `sample.table` option to code for the different sites, and the `obs.table`
339 option to code for the different visits (Miller, 2015). Thereby we obtained one overall estimate of
340 abundance per year, which we then post-processed in a generalized linear model to compute the
341 estimated temporal trend. The distance methodology performed very well, with only very few
342 type II errors to report. However, because temporal variation in availability was not modelled, the
343 magnitude of the population decline was, as expected, consistently under-estimated (RMSE =
344 35%).

345 The new methodology, which as a reminder is a generalization of distance sampling,
346 improved on the trend estimate (RMSE = 15%) by separating availability and detection. It
347 however exhibited a slightly larger rate of type II error than distance sampling (Table 2), despite
348 fully using the double observer data, indicating a loss of precision caused by the added number of
349 parameters to estimate.

350 **Quantifying the loss of precision**

351 The relatively large rate of type II error in our method indicates that correcting for known sources
352 of bias with our new framework comes at a cost in terms of loss of precision. Therefore, the effort
353 needed to fully accommodate confounding factors, should any occur, ought to be anticipated at
354 the study design stage. To investigate this further, we simulated a range of scenarios where the
355 IPS methodology was expected to perform well. That way, we could compare the statistical
356 power of our method to that of the simplest method with the lowest number of parameters,
357 providing a direct quantification of the loss of precision, and a guideline for sampling design. We
358 simulated K sites with initially 100 animals per site, so $K*100$ animals in total at the start of the
359 simulations. The population decreased by 5% per year over a 6-year period. We parameterized
360 the scenarios so that half of the decline was accounted for by a decline in the number of groups

361 per site and the remaining half was caused by a decline in the number of animals per group. Each
362 year, O observers visited each site U times for 6 years. At each visit, they divided the count in $V =$
363 3 secondary occasions, following the time-to-detection protocol that we described under “Our
364 model” above. Observers did not record distance; instead, the inference was entirely based on the
365 time to detection and multiple-observer data. Detection probability increased with group size with
366 a slope of 0.1 on the logit-log scale. The intercept of the detection-size relationship was kept
367 constant over the years. In other words, the only source of temporal variation in nuisance
368 parameters was through the change in group size with year. We ran 100 simulations per
369 combination of K , U and O . We computed the proportion of replicates in which the population
370 decrease was effectively detected.

371 As expected, the IPS method performed very well in this scenario with limited variation
372 in nuisance parameters (Fig. 1). The loss of precision by our new method relative to the IPS did
373 not appear large enough to prevent real-world applications (Fig. 1; red curves vs. blue curves).
374 For example, monitoring 8 sites for 6 years was enough to be able to detect a 5% yearly rate of
375 decrease (Fig. 1). This is a sample size typical of many ungulate monitoring schemes. Clearly, the
376 IPS methodology would have reached the same objective with a much smaller effort (3 sites
377 monitored over 3 years). But it would not have been able to detect the effect of confounding
378 factors should any be present.

379 Another issue that these simulations put to the fore was weak identifiability. When the
380 availability probability was < 0.3 (very low), the procedure converged towards a solution with
381 $\hat{\varphi} = 1$ and $\hat{p} = \varphi p$. The probability of availability was consistently over-estimated at boundary
382 one and that bias was propagated to the detection probability, which was under-estimated (Fig.
383 A4a). This is a typical weak identifiability issue, whereby the parameters φ and p are separately
384 estimable only when the data are dense. When $p < 0.3$, not enough groups are detected. In

385 Application case #2 (below), we demonstrate that incorporating additional sources of detection
386 data, as we advocate in this study, resolved the issue in a real-life application.

387 Lastly, these simulations demonstrate that the double observer protocol was never cost-
388 effective in terms of precision compared to doubling the number of surveyed sites or the number
389 of replicates per site.

390

391 **REAL STUDY CASES**

392 **Application case #1: Pyrenean chamois**

393 This case study aimed at empirically comparing the new method to the population reconstruction
394 method. The latter is expected to perform best so is used as a reference point. The objective is to
395 demonstrate the good performance of the new method at a fraction of the cost of the population
396 reconstruction method. In the Bazès study area (foothills of the Pyrenees mountains; 43.00°N,
397 0.23°W), the Pyrenean chamois (*Rupicapra pyrenaica*) population has experienced a mass
398 mortality event in the summer of 2001 that was attributed to an intoxication with an insecticide
399 (Gibert et al., 2004). Since then, breeding success has remained low. The monitoring program
400 involved up to 27 visits per year since 1998. At each visit, the distance sampling protocol was
401 applied from the same hiking trail each time. In the meantime, chamois were captured and
402 marked every year, and then marked individuals were resighted during the population surveys.

403 When applying our new framework, we used the Akaike Information Criterion to select
404 the presence or absence of temporal trends in detection probability, availability probability, and
405 group size. We also asked whether availability probability changed during the 2001 events, as
406 would be expected if the mass mortality event was associated to a change in movement rates.

407 When analyzing the capture-recapture data, we used two methods. We used the Arnason-
408 Schwarz-Gerard model (Arnason et al., 1991; Schwarz and Seber, 1999) to estimate population
409 size each year based on the year-specific estimated detection probability for marked individuals
410 and the number of detected individuals (marked and unmarked). We also reconstructed the
411 population trajectory using a matrix population model (Caswell, 2001) with 10 age-classes. The
412 demographic parameters (a.k.a. vital rates) in the matrix population model were estimated from
413 the capture-recapture data with E-Surge (Choquet et al., 2009). Further detail can be found in
414 Richard et al. (2017).

415 The model without temporal variation had 14.8 AIC points more than the model with
416 fully year-specific detection probability and an effect of the 2001 events on availability
417 probability. The half-detection distance varied across years between 247 and 611 meters. The
418 lowest detection probabilities corresponded to years with staffing issues. Availability probability
419 was 0.57 (\pm standard error: 0.14) during normal years and 0.87 (\pm standard error: 0.74) during the
420 2001 intoxication event, suggesting lower movement rates. Both our new method and the two
421 capture-recapture analyses yielded the same estimated population trajectory (Fig. 2), indicating
422 the good performance of the unmarked approach in this case relative to the much more costly
423 mark-recapture approach. The two-way coefficients of determination (r^2) between the year-
424 specific population size estimates from the three methods were both 0.66.

425

426 **Application case #2: Mediterranean mouflon**

427 This case study was specifically designed to test the new framework in the field. We wanted to
428 quantify how the precision of the population abundance estimate increased when we combined
429 distance sampling, multiple-observer, and time-to-detection in a single framework, compared to

430 when we used only one type of detection data. Incidentally, the case study also yielded an
431 unambiguous demonstration of how combining multiple types of detection data resolved the
432 above-mentioned weak identifiability issue.

433 In 2014, Mediterranean mouflons (*Ovis gmelini musimon x Ovis sp.*) were counted at
434 three locations from fixed points in the Caroux-Espinouse national hunting and wildlife reserve
435 (southwestern France; 43°38'N, 2°56'E). The environment was low scrub with forest patches. On
436 seven or eight occasions (depending on the site), two observers conducted 15-min scans
437 separated into 3 secondary sessions of 5 minutes. They noted which observer first recorded the
438 animals, during which scan, and at what distance from their vantage point, yielding 138 different
439 detection events of mouflon groups. We compared the standard errors of the parameter estimates
440 when discarding the observer information, the scan information, or both.

441 Discarding either the time-to-detection information or the double-observer information
442 led to a two to three-times increase in standard errors (Fig. 3). The time-to-detection information
443 improved precision slightly more than the double-observer information. Based on these results
444 we rank the observation protocols by order of increasing precision as follows: distance sampling,
445 time-to-detection, and multiple observers. Importantly, when we discarded the time-to-detection
446 information, the availability probability was estimated at boundary one. In other words, we
447 resolved the weak identifiability issue by collecting time-to-detection information in this case.

448

449 **Application case #3: Feral cat**

450 This case study was chosen to illustrate the challenges associated with temporal variation in
451 nuisance parameters and the adequate performance of the new analytical protocol even when only
452 distance information is available. Feral cats (*Felis silvestris catus*) have been introduced to the

453 Kerguelen archipelago (southern Indian Ocean); their abundance is a key information for a range
454 of projects in community ecology and conservation biology. We focused on one study area (the
455 2.8km-long Pointe Morne transect; 49°22S,70°26E) where the cat population was surveyed on 19
456 occasions between 2013 and 2016 (and still ongoing) using distance sampling. We considered
457 only the adult cats and did not use the information about the size of occasional family groups. At
458 each occasion, observers walked the transect back and forth until they obtained at least 30 cat
459 sightings, later reduced to 20 sightings. They waited at least 45 minutes between the back and the
460 forth, and at least two hours before starting again, sometimes the next day. We treated each back-
461 and-forth as a primary occasion *sensu* our model, but introduced a slight modification in that the
462 population abundance was constrained to remain constant in the model across the up to 19 back-
463 and-forth walks that together constituted a separate occasion.

464 Model parameters were allowed to vary across primary occasions using a random-effect
465 structure implemented with the Gauss-Hermite quadrature within a Nelson-Mead optimization
466 algorithm (Appendix B). We implemented Gaussian random effects acting on the log-transformed
467 half detection distance and on the logit-transformed availability probability. We implemented a
468 basic AIC selection procedure to select between models $\varphi(r)p(r)$, $\varphi(\cdot)p(r)$, and $\varphi(\cdot)p(\cdot)$,
469 where φ and p denote availability and detection probabilities, respectively, a dot denotes a time-
470 constant model, and r denotes a time random effect. Each random effect added a single parameter
471 to the parameter count for the AIC. Note that random effects are currently not available in the
472 ‘chamois’ user interface.

473 For comparative purposes we also applied the IPS methodology (Poisson regression) and
474 the standard distance sampling methodology, which in this case meant pooling data together from
475 up to 19 primary occasions (back-and-forth walks). We acknowledge that the fact that each
476 sampling occasion would then last several days violates the assumptions of the distance sampling

477 methodology. Our objective was indeed to determine whether this represented an issue or not, by
478 comparing the results from the standard distance sampling to the results from our new approach.

479 We obtained better precision with the new framework (Table 3: Distance vs. $\varphi(r)p(r)$)
480 because we borrowed information across sampling occasions and we exploited the repeated
481 survey structure, instead of pooling data across primary occasions. Thus, applying the standard
482 distance methodology to primary occasions that spanned over several days did not introduce a
483 major bias, only a loss of precision caused by a loss of information. In this case, contrary to the
484 other cases we presented above, our new framework thus made it possible to increase precision
485 by way of more efficient use of information, rather than lose precision by way of adding more
486 parameters.

487 The IPS methodology underestimated the population trend compared to the other
488 methods (Table 3: IPS vs. Distance and $\varphi(r)p(r)$). This is because of temporal variation in
489 nuisance parameters, which the IPS methodology did not correct for. Thus, this case study
490 unambiguously illustrates the importance of accurately representing temporal variation in
491 nuisance parameters when using population counts to infer population trends. Here, the nuisance
492 was mostly caused by variation in half-detection distance, but in the previous sections we
493 illustrated the role of availability to detection as well.

494 In the present case, we could not separate the probability of availability from the
495 detection probability (weak identifiability; p was estimated at boundary 1). From the results of
496 the mouflon case study and the simulations, we recommend either implementing a double
497 observer protocol, changing the survey area so that it is possible to implement a time-to-detection
498 protocol, or drastically increasing the number of replicates, in order to be able to identify p and
499 assess whether temporal variation in p may bias the inference in Table 3.

500 DISCUSSION

501 The methods in this study build on previous efforts to jointly analyze several sources of
502 “detection data” in studies of population abundance and population trend: distance sampling,
503 time-to-detection, and multiple observers (Amundson et al., 2014; Chandler et al., 2011; Conn et
504 al., 2012; Fiske and Chandler, 2011). Motivated by studies into mountain ungulates population
505 dynamics, we identified a need for an approach that worked for a small number of locations
506 monitored over long periods of time, when group size influenced detection and the rate of
507 temporary emigration out of terrain-limited survey areas varied over time. In addition, long-term
508 ecological monitoring schemes increasingly need to adapt their sampling effort in the face of
509 variation in financial, institutional, and volunteer support, and as a result there is a need for a
510 flexible analytical framework. We do not recommend choosing flexibility for the sake of it when
511 designing a study. But, when variation in sampling effort is inevitable, it is critical that analyses
512 effectively accommodate it. Furthermore, we implemented a fully expanded version of the
513 likelihood function, allowing the incorporation of partially observed individual covariates and
514 individual and temporal random effects, whereas previous approaches used closed-form
515 likelihood functions based on summary statistics (Fiske and Chandler, 2011). We acknowledge
516 that this decision is computationally costly: our implementation is at least 10,000 times slower
517 than a closed-form likelihood. It also requires careful care to avoid local minima in the
518 likelihood. But with simulations and real case studies we demonstrated that these features could
519 be critical to control the effect of confounding factors in population trends. Finally, a last source
520 of concern is that the bias/precision trade-off was not always in favor of our method (Fig. 1).
521 However, our simulation studies clearly showed that there are situations in which our method was
522 the only one to yield unbiased results about population trend, because the assumptions and data
523 requirements of simpler approaches were not met (Table 2). Application case #2 (mouflon),

524 which we specifically designed to test the new approach in the field, also clearly demonstrated
525 that our new method solved a weak identifiability issue, namely made it possible to separate the
526 availability and detection probabilities which otherwise would have been confounded. When
527 availability and detection covary through time, we need to separate them to avoid biases in
528 population trend estimates.

529 In our view, the loss of precision caused by the increased number of parameters in our method
530 relative to the IPS does not prevent the use of the method in real-life management cases,
531 especially when the loss of precision is taken into consideration at the sampling design stage. We
532 however recommend applying both the IPS methodology and our new method, maybe in a
533 dashboard-like suite of indicators of population change. Discrepancies between the IPS and the
534 new method would make it compelling that population trend estimation remains a difficult task
535 when the data are sparse at the beginning of a long-term program. These discrepancies would
536 quantify either the biasing effect of confounding factors, or the loss of precision associated with
537 the increased number of parameters in our new method. It is also possible to perform a
538 simulation-based statistical power analysis, as implemented in the “chamois” R-package
539 (Appendix B), to plan ahead the sampling design and determine when the results from the new
540 method are expected to reach statistical significance depending on the biological parameters.

541

542 **ACKNOWLEDGEMENTS**

543 We thank Q. Richard for the capture-recapture analysis of the chamois data. We warmly thank the
544 computing lab at IN2P3. This work was supported by the French National Park system. We thank
545 all French national park staff for their collective insight into mountain ungulate monitoring, and
546 in particular R. Bonet, M. Canut, J. Cavailhes, M. Delorme, T. Faivre, G. Farny, L. Imberdis, A.

547 Jailloux, R. Papet, E. Sourp. For the mouflon and chamois studies we thank wildlife technicians
548 J. Appolinaire, J. Duhayer, and C. Itty. For the cat study we thank the French Polar Institute
549 (IPEV) for financial support (Program n°279) and all field workers (Y Chervaux, F. Egal, A.
550 Lec'hvien and C. Brunet). A very warm thank you generally goes to everyone who contributed to
551 the field efforts over the years.

552

553 **AUTHOR'S CONTRIBUTIONS**

554 GP designed the study, wrote the R scripts, performed the statistical analyses and simulation
555 studies, and wrote the manuscript with inputs from MG. MG designed the ungulate application
556 cases and procured the data.

557

558 **SUPPLEMENTARY FILES**

559 Fig. A1: Same as Fig. 1 from the main text, but decrease of 10% per year over 3 years

560 Fig. A2: Same as Fig. 1 from the main text, but decrease of 10% per year over 6 years

561 Fig. A3: Same as Fig. 1 from the main text, but decrease of 5% per year over 6 years

562 Fig. A4: Proportional bias in estimated availability probability (A), detection probability (B), and
563 abundance (C) across all scenarios from Fig. 1 of the main text.

564 Appendix B: Using the Gauss-Hermite quadrature to fit random effect distance models

565 Supplementary Data file: beta-version R-package, tested on Windows operating systems only.

566 This includes a user's manual with installation instructions.

567

568 **LITERATURE CITED**

- 569 Alldredge, M.W., Pacifici, K., Simons, T.R., Pollock, K.H., 2008. A novel field evaluation of the
570 effectiveness of distance and independent observer sampling to estimate aural avian
571 detection probabilities. *J. Appl. Ecol.* 45, 1349–1356.
- 572 Alldredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A., Shriner, S.A., 2007. Time-of-
573 detection method for estimating abundance from point-count surveys. *Auk* 124, 653–664.
- 574 Amundson, C.L., Royle, J.A., Handel, C.M., 2014. A hierarchical model combining distance
575 sampling and time removal to estimate detection probability during avian point counts. *Auk*
576 131, 476–494.
- 577 Anderson, D.R., 2003. Response to Engeman: Index values rarely constitute reliable information.
578 *Wildl. Soc. Bull.* 31, 288–291.
- 579 Arnason, N.A., Schwarz, C.J., Gerrard, J.M., 1991. Estimating Closed Population-Size and
580 Number of Marked Animals From Sighting Data. *J. Wildl. Manage.* 55, 716–730.
- 581 Auger-Méthé, M., Field, C., Albersen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D., Mills
582 Flemming, J., 2016. State-space models' dirty little secrets: even simple linear Gaussian
583 models can have estimation problems. *Sci. Rep.* 6, 26677.
- 584 Barker, R.J., Schofield, M.R., Link, W.A., Sauer, J.R., 2018. On the reliability of N-mixture
585 models for count data. *Biometrics* 74, 369–377.
- 586 Besbeas, P., Freeman, S.N., Morgan, B.J.T., Catchpole, E.A., 2002. Integrating Mark-Recapture-
587 Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters.
588 *Biometrics* 58, 540–547.
- 589 Borchers, D.L., Cox, M.J., 2017. Distance sampling detection functions: 2D or not 2D?
590 *Biometrics* 73, 593–602.
- 591 Borchers, D.L., Laake, J.L., Southwell, C., Paxton, C.G.M., 2006. Accommodating Unmodeled

- 592 Heterogeneity in Double-Observer Distance Sampling Surveys. *Biometrics* 62, 372–378.
- 593 Buckland, S.T., Anderson, D., Burnham, K., Laake, J., 1993. Distance sampling: estimating
594 abundance of biological populations. Chapman and Hall, London.
- 595 Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2007.
596 Advanced distance sampling. Oxford University Press, Oxford UK.
- 597 Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S., 2015. Model-Based Distance
598 Sampling: Full Likelihood Methods, in: Distance Sampling: Methods and Applications.
599 Springer, Cham, pp. 141–163.
- 600 Burnham, K.P., 1993. A theory for combined analysis of ring recovery and recapture data, in:
601 Lebreton, J., North, P. (Eds.), *Marked Individuals in the Study of Bird Populations*.
602 Birkhauser Verlag, Basel, Switzerland.
- 603 Burnham, K.P., Anderson, D.R., 2002. Model selection and multi-model inference: a practical
604 information-theoretic approach. Springer, New York.
- 605 Caswell, H., 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer
606 Associates, Sunderland, MA.
- 607 Chandler, R., Royle, J., King, D., 2011. Inference about density and temporary emigration in
608 unmarked populations. *Ecology* 92, 1429–1435.
- 609 Choquet, R., Rouan, L., Pradel, R., 2009. Program E-SURGE: a software application for fitting
610 multievent models, in: Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.), *Modeling
611 Demographic Processes in Marked Populations*. Springer US, Environmental and Ecological
612 Statistics, Springer, New York, pp. 845–865.
- 613 Clement, M.J., Converse, S.J., Royle, J.A., 2017. Accounting for imperfect detection of groups
614 and individuals when estimating abundance. *Ecol. Evol.* 7, 7304–7310.
- 615 Conn, P.B., Laake, J.L., Johnson, D.S., 2012. A hierarchical modeling framework for multiple

- 616 observer transect surveys. *PLoS One* 7, e42294.
- 617 Couturier, T., Cheylan, M., Bertolero, A., Astruc, G., Besnard, A., 2013. Estimating abundance
618 and population trends when detection is low and highly variable: A comparison of three
619 methods for the Hermann's tortoise. *J. Wildl. Manage.* 77, 454–462.
- 620 Dennis, E.B., Morgan, B.J.T., Ridout, M.S., 2015. Computational aspects of N-mixture models.
621 *Biometrics* 71, 237–246.
- 622 Engeman, R.M., 2005. Indexing principles and a widely applicable paradigm for indexing animal
623 populations. *Wildl. Res.* 32, 203–210.
- 624 Fan, J., Liu, H., Wang, Z., Yang, Z., 2018. Curse of Heterogeneity: Computational Barriers in
625 Sparse Mixture Models and Phase Retrieval. <http://arxiv.org/abs/1808.06996>.
- 626 Fiske, I.J., Chandler, R.B., 2011. unmarked: An R package for fitting hierarchical models of
627 wildlife occurrence and abundance. *J. Stat. Softw.* 43, 1–23.
- 628 Garrett, E.S., Zeger, S.L., 2000. Latent Class Model Diagnosis. *Biometrics* 56, 1055–1067.
- 629 Gauthier, G., Péron, G., Lebreton, J., Grenier, P., Oudenhove, L. Van, 2016. Partitioning
630 prediction uncertainty in climate-dependent population models. *Proc. R. Soc. B* 283,
631 20162353.
- 632 Gerrodette, T., 1987. A power analysis for detecting trends. *Ecology* 68, 1364–1372.
- 633 Gibert, P., Appolinaire, J., ONCFS SD65, 2004. Intoxication d'isards au Lindane dans les Hautes-
634 Pyrénées. *Faune Sauvag.* 261, 42-47. In French.
- 635 Harris, R.B., 1986. Reliability of trend lines obtained from variable counts. *J. Wildl. Manage.* 50,
636 165–171.
- 637 Kendall, W.L., Nichols, J.D., Hines, J.E., 1997. Estimating temporary emigration using capture-
638 recapture data with Pollock's robust design. *Ecology* 78, 563–578.
- 639 Link, W., Sauer, J., 1998. Estimating population change from count data: application to the North

- 640 American Breeding Bird Survey. *Ecol. Appl.* 8, 258–268.
- 641 Loison, A., Appolinaire, J., Jullien, J.M., Dubray, D., 2006. How reliable are total counts to detect
642 trends in population size of chamois *Rupicapra rupicapra* and *R. pyrenaica*? *Wildlife Biol.* 1,
643 77–88.
- 644 Marques, T.A., Buckland, S.T., Bispo, R., Howland, B., 2013. Accounting for animal density
645 gradients using independent information in distance sampling surveys. *Stat. Methods Appl.*
646 22, 67–80.
- 647 Mebane, W.R.J., Sekhon, J.S., 2011. Genetic Optimization Using Derivatives: The *rgenoud*
648 package for R. *J. Stat. Softw.* 42, 1–26.
- 649 Miller, D.L., 2015. Distance Sampling detection function and abundance Estimation [WWW
650 Document]. <http://github.com/DistanceDevelopment/Distance/>
- 651 Nichols, J., Hines, J., Sauer, J., Fallon, F., 2000. A double-observer approach for estimating
652 detection probability and abundance from point counts. *Auk* 117, 393–408.
- 653 Pépin, D., Gerard, J.-F., 2008. Group dynamics and local population density dependence of group
654 size in the Pyrenean chamois, *Rupicapra pyrenaica*. *Anim. Behav.* 75, 361–369.
- 655 Richard, Q., Toïgo, C., Appolinaire, J., Loison, A., Garel, M., 2017. From gestation to weaning:
656 Combining robust design and multi-event models unveils cost of lactation in a large
657 herbivore. *J. Anim. Ecol.* 86, 1497–1509.
- 658 Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated
659 counts. *Biometrics* 60, 108–15.
- 660 Schwarz, C.J., Seber, G.A.F., 1999. Estimating Animal Abundance: Review III. *Stat. Sci.* 14,
661 427–456.
- 662 Sollmann, R., Gardner, B., Chandler, R.B., Royle, J.A., 2015. An open-population hierarchical
663 distance sampling model. *Ecology* 96, 325–331.

- 664 Toïgo, C., Gaillard, J.M., Michallet, J., 1996. La taille des groupes : un bioindicateur de l'effectif
665 des populations de bouquetin des Alpes (*Capra ibex ibex*) ? *Mammalia* 60, 463-472. In
666 French.
- 667 Veech, J.A., Ott, J.R., Troy, J.R., 2016. Intrinsic heterogeneity in detection probability and its
668 effect on N -mixture models. *Methods Ecol. Evol.* 7, 1019–1028.
- 669 Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson vs. negative binomial regression: how should
670 we model overdispersed count data? *Ecology* 88, 2766–72.
- 671 Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and management of animal
672 populations: modeling, estimation, and decision making. Academic Press, San Diego, CA.
- 673 Zhao, Q., Royle, J.A., 2019. Dynamic N -mixture models with temporal variability in detection
674 probability. *Ecol. Modell.* 393, 20–24.
- 675
- 676 Alldredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A., Shriver, S.A., 2007. Time-of-
677 detection method for estimating abundance from point-count surveys. *Auk* 124, 653–664.
- 678 Anderson, D.R., 2003. Response to Engeman: Index values rarely constitute reliable information.
679 *Wildl. Soc. Bull.* 31, 288–291.
- 680 Arnason, N.A., Schwarz, C.J., Gerrard, J.M., 1991. Estimating Closed Population-Size and
681 Number of Marked Animals From Sighting Data. *J. Wildl. Manage.* 55, 716–730.
- 682 Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D., Mills
683 Flemming, J., 2016. State-space models' dirty little secrets: even simple linear Gaussian
684 models can have estimation problems. *Sci. Rep.* 6, 26677.
- 685 Barker, R.J., Schofield, M.R., Link, W.A., Sauer, J.R., 2018. On the reliability of N -mixture
686 models for count data. *Biometrics* 74, 369–377.
- 687 Besbeas, P., Freeman, S.N., Morgan, B.J.T., Catchpole, E.A., 2002. Integrating Mark-Recapture-

- 688 Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters.
689 *Biometrics* 58, 540–547.
- 690 Borchers, D.L., Cox, M.J., 2017. Distance sampling detection functions: 2D or not 2D?
691 *Biometrics* 73, 593–602.
- 692 Borchers, D.L., Laake, J.L., Southwell, C., Paxton, C.G.M., 2006. Accommodating Unmodeled
693 Heterogeneity in Double-Observer Distance Sampling Surveys. *Biometrics* 62, 372–378.
- 694 Buckland, S.T., Anderson, D., Burnham, K., Laake, J., 1993. Distance sampling: estimating
695 abundance of biological populations. Chapman and Hall, London.
- 696 Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2007.
697 Advanced distance sampling. Oxford University Press, Oxford UK.
- 698 Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S., 2015. Model-Based Distance
699 Sampling: Full Likelihood Methods, in: *Distance Sampling: Methods and Applications*.
700 Springer, Cham, pp. 141–163.
- 701 Burnham, K.P., 1993. A theory for combined analysis of ring recovery and recapture data, in:
702 Lebreton, J., North, P. (Eds.), *Marked Individuals in the Study of Bird Populations*.
703 Birkhauser Verlag, Basel, Switzerland.
- 704 Burnham, K.P., Anderson, D.R., 2002. Model selection and multi-model inference: a practical
705 information-theoretic approach. Springer, New York.
- 706 Caswell, H., 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer
707 Associates, Sunderland, MA.
- 708 Chandler, R., Royle, J., King, D., 2011. Inference about density and temporary emigration in
709 unmarked populations. *Ecology* 92, 1429–1435.
- 710 Choquet, R., Rouan, L., Pradel, R., 2009. Program E-SURGE: a software application for fitting
711 multievent models, in: Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.), *Modeling*

- 712 Demographic Processes in Marked Populations. Springer US, Environmental and Ecological
713 Statistics, Springer, New York, pp. 845–865.
- 714 Clement, M.J., Converse, S.J., Royle, J.A., 2017. Accounting for imperfect detection of groups
715 and individuals when estimating abundance. *Ecol. Evol.* 7, 7304–7310.
- 716 Conn, P.B., Laake, J.L., Johnson, D.S., 2012. A hierarchical modeling framework for multiple
717 observer transect surveys. *PLoS One* 7, e42294.
- 718 Couturier, T., Cheylan, M., Bertolero, A., Astruc, G., Besnard, A., 2013. Estimating abundance
719 and population trends when detection is low and highly variable: A comparison of three
720 methods for the Hermann's tortoise. *J. Wildl. Manage.* 77, 454–462.
- 721 Dénes, F. V., Silveira, L.F., Beissinger, S.R., 2015. Estimating abundance of unmarked animal
722 populations: accounting for imperfect detection and other sources of zero inflation. *Methods*
723 *Ecol. Evol.* 6, 543–556.
- 724 Dennis, E.B., Morgan, B.J.T., Ridout, M.S., 2015. Computational aspects of N-mixture models.
725 *Biometrics* 71, 237–246.
- 726 Engeman, R.M., 2005. Indexing principles and a widely applicable paradigm for indexing animal
727 populations. *Wildl. Res.* 32, 203–210.
- 728 Fan, J., Liu, H., Wang, Z., Yang, Z., 2018. Curse of Heterogeneity: Computational Barriers in
729 Sparse Mixture Models and Phase Retrieval. <http://arxiv.org/abs/1808.06996>.
- 730 Fiske, I.J., Chandler, R.B., 2011. unmarked: An R package for fitting hierarchical models of
731 wildlife occurrence and abundance. *J. Stat. Softw.* 43, 1–23.
- 732 Garrett, E.S., Zeger, S.L., 2000. Latent Class Model Diagnosis. *Biometrics* 56, 1055–1067.
- 733 Gauthier, G., Péron, G., Lebreton, J., Grenier, P., Oudenhove, L. Van, 2016. Partitioning
734 prediction uncertainty in climate-dependent population models. *Proc. R. Soc. B* 283,
735 20162353.

- 736 Gerrodette, T., 1987. A power analysis for detecting trends. *Ecology* 68, 1364–1372.
- 737 Gibert, P., Appolinaire, J., ONCFS SD65, 2004. Intoxication d'isards au Lindane dans les Hautes-
738 Pyrénées. *Faune Sauvage*. 261, 42–47. In French.
- 739 Harris, R.B., 1986. Reliability of trend lines obtained from variable counts. *J. Wildl. Manage.* 50,
740 165–171.
- 741 Hostetler, J.A., Chandler, R.B., 2015. Improved state-space models for inference about spatial
742 and temporal variation in abundance from count data. *Ecology* 96, 1713–1723.
- 743 Kendall, W.L., Nichols, J.D., Hines, J.E., 1997. Estimating temporary emigration using capture-
744 recapture data with Pollock's robust design. *Ecology* 78, 563–578.
- 745 Link, W., Sauer, J., 1998. Estimating population change from count data: application to the North
746 American Breeding Bird Survey. *Ecol. Appl.* 8, 258–268.
- 747 Loison, A., Appolinaire, J., Jullien, J.M., Dubray, D., 2006. How reliable are total counts to detect
748 trends in population size of chamois *Rupicapra rupicapra* and *R. pyrenaica*? *Wildlife Biol.* 1,
749 77–88.
- 750 Marques, T.A., Buckland, S.T., Bispo, R., Howland, B., 2013. Accounting for animal density
751 gradients using independent information in distance sampling surveys. *Stat. Methods Appl.*
752 22, 67–80.
- 753 Mebane, W.R.J., Sekhon, J.S., 2011. Genetic Optimization Using Derivatives: The rgenoud
754 package for R. *J. Stat. Softw.* 42, 1–26.
- 755 Miller, D.L., 2015. Distance Sampling detection function and abundance Estimation [WWW
756 Document]. <http://github.com/DistanceDevelopment/Distance/>
- 757 Nichols, J., Hines, J., Sauer, J., Fallon, F., 2000. A double-observer approach for estimating
758 detection probability and abundance from point counts. *Auk* 117, 393–408.
- 759 Pépin, D., Gerard, J.-F., 2008. Group dynamics and local population density dependence of group

- 760 size in the Pyrenean chamois, *Rupicapra pyrenaica*. *Anim. Behav.* 75, 361–369.
- 761 Richard, Q., Toïgo, C., Appolinaire, J., Loison, A., Garel, M., 2017. From gestation to weaning:
762 Combining robust design and multi-event models unveils cost of lactation in a large
763 herbivore. *J. Anim. Ecol.* 86, 1497–1509.
- 764 Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated
765 counts. *Biometrics* 60, 108–115.
- 766 Schwarz, C.J., Seber, G.A.F., 1999. Estimating Animal Abundance: Review III. *Stat. Sci.* 14,
767 427–456.
- 768 Sillett, T.S., Chandler, R.B., Royle, J.A., Kery, M., Morrison, S. a, 2012. Hierarchical distance-
769 sampling models to estimate population size and habitat-specific abundance of an island
770 endemic. *Ecol. Appl.* 22, 1997–2006.
- 771 Sollmann, R., Gardner, B., Chandler, R.B., Royle, J.A., 2015. An open-population hierarchical
772 distance sampling model. *Ecology* 96, 325–331.
- 773 Toïgo, C., Gaillard, J.M., Michallet, J., 1996. La taille des groupes : un bioindicateur de l'effectif
774 des populations de bouquetin des Alpes (*Capra ibex ibex*) ? *Mammalia* 60, 463–472. In
775 French.
- 776 Veech, J.A., Ott, J.R., Troy, J.R., 2016. Intrinsic heterogeneity in detection probability and its
777 effect on *N*-mixture models. *Methods Ecol. Evol.* 7, 1019–1028.
- 778 Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson vs. negative binomial regression: how should
779 we model overdispersed count data? *Ecology* 88, 2766–72.
- 780 Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and management of animal
781 populations: modeling, estimation, and decision making. Academic Press, San Diego, CA.
- 782 Zhao, Q., Royle, J.A., 2019. Dynamic N-mixture models with temporal variability in detection
783 probability. *Ecol. Modell.* 393, 20–24.

784 **TABLES**

785 Table 1: Notation for the ‘chamois’ class of models

Notation	Meaning
$C_{k,t,u}$	Total number of animal groups observed during the u^{th} visit to site k in year (or other time unit) t
$N_{k,t}$	Total number of animal groups using site k in year t
$\varphi_{k,t,u}$	Probability that a group is available for detection during the u^{th} visit, following the “open distance” parameterization of (Chandler et al., 2011; Sollmann et al., 2015).
$p_{k,t,o,u,v}(g, d)$	Detection probability by observer o during secondary session v of visit u , for a group of size g at distance d from the observer. In practice we use either using the half-normal function with spread parameter $D_{k,t,o,u,v}$ (half-detection distance) or a histogram-like piecewise function.
$U_{k,t}$	Number of visits to site k in year t
$V_{k,t,u}$	Number of secondary sessions during visit u
$O_{k,t,u}$	Number of observers during visit u
$Pr(d k)$	Distribution of distances to the observer, including both the animals that eventually are detected, and the animals that are not detected, within site k . In our framework, this term is meant to accommodate the typically irregular shape of the survey sites and the potential offset of the observers’ position relative to the centroid of the sites. It is thus directly informed by the user rather than estimated. More generally, this term could be used to introduce variation in the population density among the sites.

$Pr(g|k, t, u)$ Distribution of group sizes in site k , during visit u . This includes both detected and undetected groups. In practice, we used a one-inflated negative-binomial distribution with parameters $\pi_{k,t,u}$, $\mu_{k,t,u}$, and $\sigma_{k,t,u}$ respectively for the proportion of groups of size 1 (solitary animals), the average size of groups of size > 1 , and the shape parameter of the negative-binomial distribution of groups of size > 1 .

786

787

788 Table 2: Simulation study of the bias in simpler methods over 20 replicates. ‘IPS’ stands for the
 789 Poisson regression of population counts. ‘N-mixture’ means that we fit a separate N-mixture
 790 model for each year using function pcount in R-package unmarked (Fiske and Chandler, 2011).
 791 ‘Distance’ denotes the standard distance methodology: function ds in R-package Distance (Miller,
 792 2015) applied to each year \times site combination separately. ‘Non-expected trend’ means that the
 793 estimated population trend was positive (whereas the true simulated one was negative). ‘Type I
 794 error’ means that the positive trend was statistically significant. ‘Type II error’ means that the P-
 795 value of the population trend was above 0.05, meaning that no definitive conclusion about
 796 population trend would have been reached. ‘Trend RMSE’ is the % root mean squared error of
 797 the estimated rate of population decline (log scale).

	IPS	N-mixture	Distance	New method
Non-expected trend	98%	42%	0%	2%
Type I error	54%	20%	0%	0%
Type II error	46%	42%	6%	16%
Trend RMSE	>100%	85%	35%	15%

798

799

800 Table 3: Results of the feral cat case study. φ and p denote availability and detection
 801 probabilities, respectively, a dot denotes a time-constant model, r denotes a random effect (acting
 802 at the primary occasion scale and implemented as described in Appendix B). ‘Distance’ denotes
 803 the standard distance methodology: function `ds` in R-package `Distance` (Miller, 2015) applied to
 804 each sampling occasion separately. ‘IPS’ denotes the Poisson regression of population count
 805 against time. ΔAIC is the difference in Akaike points between the focal and preferred model. A
 806 hyphen indicates a quantity that could not be estimated. ΔAIC was not computed for `Distance`
 807 and `IPS` because the different treatment of the constant terms in the likelihoods prevented the
 808 comparison of AIC values.

Model	ΔAIC	Log-scale temporal trend in abundance (month⁻¹)
$\varphi(r)p(r)$	0	-0.045 (CI -0.056, -0.021)
$\varphi(\cdot)p(r)$	1.72	-0.045 (CI -0.050, -0.038)
$\varphi(\cdot)p(\cdot)$	15.98	-0.020 (CI -0.025, -0.014)
Distance	-	-0.052 (CI -0.071, -0.033)
IPS	-	-0.026 (CI -0.034, -0.018)

809

810

811 **FIGURE LEGENDS**

812 Fig. 1: Quantification of the loss of precision in a scenario without any variance in nuisance
 813 parameters. Probability of not detecting an annual rate of change (increase or decrease, at
 814 random) of 5% over 6 years, for various scenarios of detection probability p and availability
 815 probability φ using our new method. The grey shading darkens when the probability of type II
 816 error increases. The bold line is the 5% contour (right of the line, the probability of type II error is
 817 lower than 5%). The white-dashed lines correspond to the 5% contour for the population index
 818 methodology (if these white-dashed lines are absent then the probability to detect the trend was
 819 always >95% using the index). X-axis: number of repetitions. Y-axis: number of survey sites. The
 820 framed plots indicate situations that correspond to a 40% coefficient of variation, typical of
 821 mountain ungulate monitoring, even if the CV tends to get smaller than that with more replicates
 822 (Loison et al., 2006). The same figure for the probability that a 10% annual rate of change over
 823 three years was detected with a 5% risk threshold are provided in Figs. A1-3.

824
 825 Fig. 2: Comparison of estimated chamois abundance in the Bazès study area, with the Arnason-
 826 Schwarz-Gerard model fitted to resighting data from marked animals ('A-S-G'), with a 10 age-
 827 class population model with demographic rates estimated from individual-based data
 828 ('reconstruction'), and with our new method.

829
 830 Fig. 3: Comparison of the standard errors from our new method using various combinations of
 831 distance sampling ('Dist'), time-to-detection ('Scans') and double-observer ('DbObs'), in the
 832 mouflon case study. 'G1', 'G2', 'G3' stands for the log of the number of undetected animal
 833 groups in each of three survey sites, ' π ' is the proportion of groups of size 1 (logit-scale intercept
 834 and effect of site 1). ' σ ' is the shape parameter of the negative-binomial distribution of group

835 sizes >1 , ' μ ' is its mean, ' p ' is the group detection probability (logit scale intercept, effect of log-
836 transformed group size and of site 1), and ' ϕ ' is the availability probability. Asterisks indicate
837 missing standard errors because the estimate was at boundary 1, i.e., weak identifiability.
838

Abundance (base 100 in 1998)



