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

Guillaume Péron, Mathieu Garel

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

Guillaume Péron (1)*
Mathieu Garel (2)

1. Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, F-69622 Villeurbanne, France
2. Office National de la Chasse et de la Faune Sauvage, Direction de la Recherche et de l'Expertise, Unité Ongulés Sauvages, 5 allée de Bethléem, Z.I. Mayencin, 38610 Gières, France.

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

Running title: Repeated population surveys

Highlights

- **We generalize distance sampling to analyze multi-year multi-site population surveys when sampling protocols vary across years and sites, using multiple observers and time to detection in addition to distance information.**
- **Standard distance sampling may misrepresent population trends in the presence of temporal variation in the availability to detection.**
- **The estimation of availability to detection is improved by combining multiple data types.**
- **The new framework is costly in terms of number of parameters to estimate and computing time, but compatible with the logistics of typical ungulate population surveys.**

ABSTRACT

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

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

INTRODUCTION

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

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

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

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

The Index of population size methodology (IPS)

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

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

Population reconstruction from individual-based data

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

Unmarked methods

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

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

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

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

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

OUR MODEL

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

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

Three types of detection data for unmarked animals

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

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

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

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

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

Group size

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

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

Model likelihood

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

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

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

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

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

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

Eq. 2

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

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

Eq. 3

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

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

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

Our model is a generalization of distance sampling because if we remove the multiple 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, and if we remove all the dependencies on g, we arrive at a likelihood of the form explained by Buckland, Rexstad, Marques, & Oedekoven (2015). By contrast, our model does not belong to the N-mixture class of models because the binomial error structure applies *within*, not *across* sites and visits.

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

Post-hoc manipulations

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

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

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

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

SIMULATION STUDIES

Demonstrating bias in simpler methods

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

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

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

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

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

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

Quantifying the loss of precision

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

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

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

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

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

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

REAL STUDY CASES

Application case #1: Pyrenean chamois

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

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

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

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

Application case #2: Mediterranean mouflon

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

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

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

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

Application case #3: Feral cat

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

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

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

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

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

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

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

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

DISCUSSION

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

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

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

ACKNOWLEDGEMENTS

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

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

AUTHOR'S CONTRIBUTIONS

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

SUPPLEMENTARY FILES

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

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

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

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

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

Supplementary Data file: beta-version R-package, tested on Windows operating systems only. This includes a user's manual with installation instructions.

LITERATURE CITED

Alldredge, M.W., Pacifici, K., Simons, T.R., Pollock, K.H., 2008. A novel field evaluation of the effectiveness of distance and independent observer sampling to estimate aural avian detection probabilities. J. Appl. Ecol. 45, 1349–1356.

Alldredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A., Shriner, S.A., 2007. Time-of-detection method for estimating abundance from point-count surveys. Auk 124, 653–664.

Amundson, C.L., Royle, J.A., Handel, C.M., 2014. A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. Auk 131, 476–494.

Anderson, D.R., 2003. Response to Engeman: Index values rarely constitute reliable information. Wildl. Soc. Bull. 31, 288–291.

Arnason, N.A., Schwarz, C.J., Gerrard, J.M., 1991. Estimating Closed Population-Size and Number of Marked Animals From Sighting Data. J. Wildl. Manage. 55, 716–730.

Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D., Mills Flemming, J., 2016. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. Sci. Rep. 6, 26677.

Barker, R.J., Schofield, M.R., Link, W.A., Sauer, J.R., 2018. On the reliability of N-mixture models for count data. Biometrics 74, 369–377.

Besbeas, P., Freeman, S.N., Morgan, B.J.T., Catchpole, E.A., 2002. Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. Biometrics 58, 540–547.

Borchers, D.L., Cox, M.J., 2017. Distance sampling detection functions: 2D or not 2D? Biometrics 73, 593–602.

Borchers, D.L., Laake, J.L., Southwell, C., Paxton, C.G.M., 2006. Accommodating Unmodeled

Heterogeneity in Double-Observer Distance Sampling Surveys. Biometrics 62, 372–378.

Buckland, S.T., Anderson, D., Burnham, K., Laake, J., 1993. Distance sampling: estimating abundance of biological populations. Chapman and Hall, London.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2007. Advanced distance sampling. Oxford University Press, Oxford UK.

Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S., 2015. Model-Based Distance Sampling: Full Likelihood Methods, in: Distance Sampling: Methods and Applications. Springer, Cham, pp. 141–163.

Burnham, K.P., 1993. A theory for combined analysis of ring recovery and recapture data, in: Lebreton, J., North, P. (Eds.), Marked Individuals in the Study of Bird Populations. Birkhauser Verlag, Basel, Switzerland.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer, New York.

Caswell, H., 2001. Matrix Population Models: Construction, Analysis, and Interpretation. Sinauer Associates, Sunderland, MA.

Chandler, R., Royle, J., King, D., 2011. Inference about density and temporary emigration in unmarked populations. Ecology 92, 1429–1435.

Choquet, R., Rouan, L., Pradel, R., 2009. Program E-SURGE: a software application for fitting multievent models, in: Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.), Modeling Demographic Processes in Marked Populations. Springer US, Environmental and Ecological Statistics, Springer, New York, pp. 845–865.

Clement, M.J., Converse, S.J., Royle, J.A., 2017. Accounting for imperfect detection of groups and individuals when estimating abundance. Ecol. Evol. 7, 7304–7310.

Conn, P.B., Laake, J.L., Johnson, D.S., 2012. A hierarchical modeling framework for multiple

observer transect surveys. PLoS One 7, e42294.

Couturier, T., Cheylan, M., Bertolero, A., Astruc, G., Besnard, A., 2013. Estimating abundance and population trends when detection is low and highly variable: A comparison of three methods for the Hermann's tortoise. J. Wildl. Manage. 77, 454–462.

Dennis, E.B., Morgan, B.J.T., Ridout, M.S., 2015. Computational aspects of N-mixture models. Biometrics 71, 237–246.

Engeman, R.M., 2005. Indexing principles and a widely applicable paradigm for indexing animal populations. Wildl. Res. 32, 203–210.

Fan, J., Liu, H., Wang, Z., Yang, Z., 2018. Curse of Heterogeneity: Computational Barriers in Sparse Mixture Models and Phase Retrieval. http://arxiv.org/abs/1808.06996.

Fiske, I.J., Chandler, R.B., 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. J. Stat. Softw. 43, 1–23.

Garrett, E.S., Zeger, S.L., 2000. Latent Class Model Diagnosis. Biometrics 56, 1055–1067.

Gauthier, G., Péron, G., Lebreton, J., Grenier, P., Oudenhove, L. Van, 2016. Partitioning prediction uncertainty in climate-dependent population models. Proc. R. Soc. B 283, 20162353.

Gerrodette, T., 1987. A power analysis for detecting trends. Ecology 68, 1364–1372.

Gibert, P., Appolinaire, J., ONCFS SD65, 2004. Intoxication d'isards au Lindane dans les Hautes-Pyrénées. Faune Sauvag. 261, 42-47. In French.

Harris, R.B., 1986. Reliability of trend lines obtained from variable counts. J. Wildl. Manage. 50, 165–171.

Kendall, W.L., Nichols, J.D., Hines, J.E., 1997. Estimating temporary emigration using capture-recapture data with Pollock's robust design. Ecology 78, 563–578.

Link, W., Sauer, J., 1998. Estimating population change from count data: application to the North

American Breeding Bird Survey. Ecol. Appl. 8, 258–268.

Loison, A., Appolinaire, J., Jullien, J.M., Dubray, D., 2006. How reliable are total counts to detect trends in population size of chamois Rupicapra rupicapra and R. pyrenaica? Wildlife Biol. 1, 77–88.

Marques, T.A., Buckland, S.T., Bispo, R., Howland, B., 2013. Accounting for animal density gradients using independent information in distance sampling surveys. Stat. Methods Appl. 22, 67–80.

Mebane, W.R.J., Sekhon, J.S., 2011. Genetic Optimization Using Derivatives: The rgenoud package for R. J. Stat. Softw. 42, 1–26.

Miller, D.L., 2015. Distance Sampling detection function and abundance Estimation [WWW Document]. http://github.com/DistanceDevelopment/Distance/

Nichols, J., Hines, J., Sauer, J., Fallon, F., 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117, 393–408.

Pépin, D., Gerard, J.-F., 2008. Group dynamics and local population density dependence of group size in the Pyrenean chamois, Rupicapra pyrenaica. Anim. Behav. 75, 361–369.

Richard, Q., Toïgo, C., Appolinaire, J., Loison, A., Garel, M., 2017. From gestation to weaning: Combining robust design and multi-event models unveils cost of lactation in a large herbivore. J. Anim. Ecol. 86, 1497–1509.

Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60, 108–15.

Schwarz, C.J., Seber, G.A.F., 1999. Estimating Animal Abundance: Review III. Stat. Sci. 14, 427–456.

Sollmann, R., Gardner, B., Chandler, R.B., Royle, J.A., 2015. An open-population hierarchical distance sampling model. Ecology 96, 325–331.

Toïgo, C., Gaillard, J.M., Michallet, J., 1996. La taille des groupes : un bioindicateur de l'effectif des populations de bouquetin des Alpes (Capra ibex ibex) ? Mammalia 60, 463-472. In French.

Veech, J.A., Ott, J.R., Troy, J.R., 2016. Intrinsic heterogeneity in detection probability and its effect on *N* -mixture models. Methods Ecol. Evol. 7, 1019–1028.

Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88, 2766–72.

Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and management of animal populations: modeling, estimation, and decision making. Academic Press, San Diego, CA.

Zhao, Q., Royle, J.A., 2019. Dynamic N-mixture models with temporal variability in detection probability. Ecol. Modell. 393, 20–24.

Alldredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A., Shriner, S.A., 2007. Time-of-detection method for estimating abundance from point-count surveys. Auk 124, 653–664.

Anderson, D.R., 2003. Response to Engeman: Index values rarely constitute reliable information. Wildl. Soc. Bull. 31, 288–291.

Arnason, N.A., Schwarz, C.J., Gerrard, J.M., 1991. Estimating Closed Population-Size and Number of Marked Animals From Sighting Data. J. Wildl. Manage. 55, 716–730.

Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D., Mills Flemming, J., 2016. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. Sci. Rep. 6, 26677.

Barker, R.J., Schofield, M.R., Link, W.A., Sauer, J.R., 2018. On the reliability of N-mixture models for count data. Biometrics 74, 369–377.

Besbeas, P., Freeman, S.N., Morgan, B.J.T., Catchpole, E.A., 2002. Integrating Mark-Recapture-

Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. Biometrics 58, 540–547.

Borchers, D.L., Cox, M.J., 2017. Distance sampling detection functions: 2D or not 2D? Biometrics 73, 593–602.

Borchers, D.L., Laake, J.L., Southwell, C., Paxton, C.G.M., 2006. Accommodating Unmodeled Heterogeneity in Double-Observer Distance Sampling Surveys. Biometrics 62, 372–378.

Buckland, S.T., Anderson, D., Burnham, K., Laake, J., 1993. Distance sampling: estimating abundance of biological populations. Chapman and Hall, London.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2007. Advanced distance sampling. Oxford University Press, Oxford UK.

Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S., 2015. Model-Based Distance Sampling: Full Likelihood Methods, in: Distance Sampling: Methods and Applications. Springer, Cham, pp. 141–163.

Burnham, K.P., 1993. A theory for combined analysis of ring recovery and recapture data, in: Lebreton, J., North, P. (Eds.), Marked Individuals in the Study of Bird Populations. Birkhauser Verlag, Basel, Switzerland.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer, New York.

Caswell, H., 2001. Matrix Population Models: Construction, Analysis, and Interpretation. Sinauer Associates, Sunderland, MA.

Chandler, R., Royle, J., King, D., 2011. Inference about density and temporary emigration in unmarked populations. Ecology 92, 1429–1435.

Choquet, R., Rouan, L., Pradel, R., 2009. Program E-SURGE: a software application for fitting multievent models, in: Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.), Modeling

Demographic Processes in Marked Populations. Springer US, Environmental and Ecological Statistics, Springer, New York, pp. 845–865.

Clement, M.J., Converse, S.J., Royle, J.A., 2017. Accounting for imperfect detection of groups and individuals when estimating abundance. Ecol. Evol. 7, 7304–7310.

Conn, P.B., Laake, J.L., Johnson, D.S., 2012. A hierarchical modeling framework for multiple observer transect surveys. PLoS One 7, e42294.

Couturier, T., Cheylan, M., Bertolero, A., Astruc, G., Besnard, A., 2013. Estimating abundance and population trends when detection is low and highly variable: A comparison of three methods for the Hermann's tortoise. J. Wildl. Manage. 77, 454–462.

Dénes, F. V., Silveira, L.F., Beissinger, S.R., 2015. Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. Methods Ecol. Evol. 6, 543–556.

Dennis, E.B., Morgan, B.J.T., Ridout, M.S., 2015. Computational aspects of N-mixture models. Biometrics 71, 237–246.

Engeman, R.M., 2005. Indexing principles and a widely applicable paradigm for indexing animal populations. Wildl. Res. 32, 203–210.

Fan, J., Liu, H., Wang, Z., Yang, Z., 2018. Curse of Heterogeneity: Computational Barriers in Sparse Mixture Models and Phase Retrieval. http://arxiv.org/abs/1808.06996.

Fiske, I.J., Chandler, R.B., 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. J. Stat. Softw. 43, 1–23.

Garrett, E.S., Zeger, S.L., 2000. Latent Class Model Diagnosis. Biometrics 56, 1055–1067.

Gauthier, G., Péron, G., Lebreton, J., Grenier, P., Oudenhove, L. Van, 2016. Partitioning prediction uncertainty in climate-dependent population models. Proc. R. Soc. B 283, 20162353.

Gerrodette, T., 1987. A power analysis for detecting trends. Ecology 68, 1364–1372.

Gibert, P., Appolinaire, J., ONCFS SD65, 2004. Intoxication d'isards au Lindane dans les Hautes-Pyrénées. Faune Sauvag. 261, 42–47. In French.

Harris, R.B., 1986. Reliability of trend lines obtained from variable counts. J. Wildl. Manage. 50, 165–171.

Hostetler, J.A., Chandler, R.B., 2015. Improved state-space models for inference about spatial and temporal variation in abundance from count data. Ecology 96, 1713–1723.

Kendall, W.L., Nichols, J.D., Hines, J.E., 1997. Estimating temporary emigration using capture-recapture data with Pollock's robust design. Ecology 78, 563–578.

Link, W., Sauer, J., 1998. Estimating population change from count data: application to the North American Breeding Bird Survey. Ecol. Appl. 8, 258–268.

Loison, A., Appolinaire, J., Jullien, J.M., Dubray, D., 2006. How reliable are total counts to detect trends in population size of chamois Rupicapra rupicapra and R. pyrenaica? Wildlife Biol. 1, 77–88.

Marques, T.A., Buckland, S.T., Bispo, R., Howland, B., 2013. Accounting for animal density gradients using independent information in distance sampling surveys. Stat. Methods Appl. 22, 67–80.

Mebane, W.R.J., Sekhon, J.S., 2011. Genetic Optimization Using Derivatives: The rgenoud package for R. J. Stat. Softw. 42, 1–26.

Miller, D.L., 2015. Distance Sampling detection function and abundance Estimation [WWW Document]. http://github.com/DistanceDevelopment/Distance/

Nichols, J., Hines, J., Sauer, J., Fallon, F., 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117, 393–408.

Pépin, D., Gerard, J.-F., 2008. Group dynamics and local population density dependence of group

size in the Pyrenean chamois, Rupicapra pyrenaica. Anim. Behav. 75, 361–369.

Richard, Q., Toïgo, C., Appolinaire, J., Loison, A., Garel, M., 2017. From gestation to weaning: Combining robust design and multi-event models unveils cost of lactation in a large herbivore. J. Anim. Ecol. 86, 1497–1509.

Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60, 108–15.

Schwarz, C.J., Seber, G.A.F., 1999. Estimating Animal Abundance: Review III. Stat. Sci. 14, 427–456.

Sillett, T.S., Chandler, R.B., Royle, J.A., Kery, M., Morrison, S. a, 2012. Hierarchical distance-sampling models to estimate population size and habitat-specific abundance of an island endemic. Ecol. Appl. 22, 1997–2006.

Sollmann, R., Gardner, B., Chandler, R.B., Royle, J.A., 2015. An open-population hierarchical distance sampling model. Ecology 96, 325–331.

Toïgo, C., Gaillard, J.M., Michallet, J., 1996. La taille des groupes : un bioindicateur de l'effectif des populations de bouquetin des Alpes (Capra ibex ibex) ? Mammalia 60, 463–472. In French.

Veech, J.A., Ott, J.R., Troy, J.R., 2016. Intrinsic heterogeneity in detection probability and its effect on *N* -mixture models. Methods Ecol. Evol. 7, 1019–1028.

Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88, 2766–72.

Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and management of animal populations: modeling, estimation, and decision making. Academic Press, San Diego, CA.

Zhao, Q., Royle, J.A., 2019. Dynamic N-mixture models with temporal variability in detection probability. Ecol. Modell. 393, 20–24.

TABLES

Table 1: Notation for the 'chamois' class of models

Notation	Meaning
$C_{k,t,u}$	Total number of animal groups observed during the uth 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 uth 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.

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

Table 3: Results of the feral cat case study. φ and p denote availability and detection probabilities, respectively, a dot denotes a time-constant model, *r* denotes a random effect (acting at the primary occasion scale and implemented as described in Appendix B). 'Distance' denotes the standard distance methodology: function ds in R-package Distance (Miller, 2015) applied to each sampling occasion separately. 'IPS' denotes the Poisson regression of population count against time. ΔAIC is the difference in Akaike points between the focal and preferred model. A hyphen indicates a quantity that could not be estimated. ΔAIC was not computed for Distance and IPS because the different treatment of the constant terms in the likelihoods prevented the 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)
$\varphi(.)p(r)$	1.72	-0.045 (CI -0.050, -0.038)
$\varphi(.)p(.)$	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)

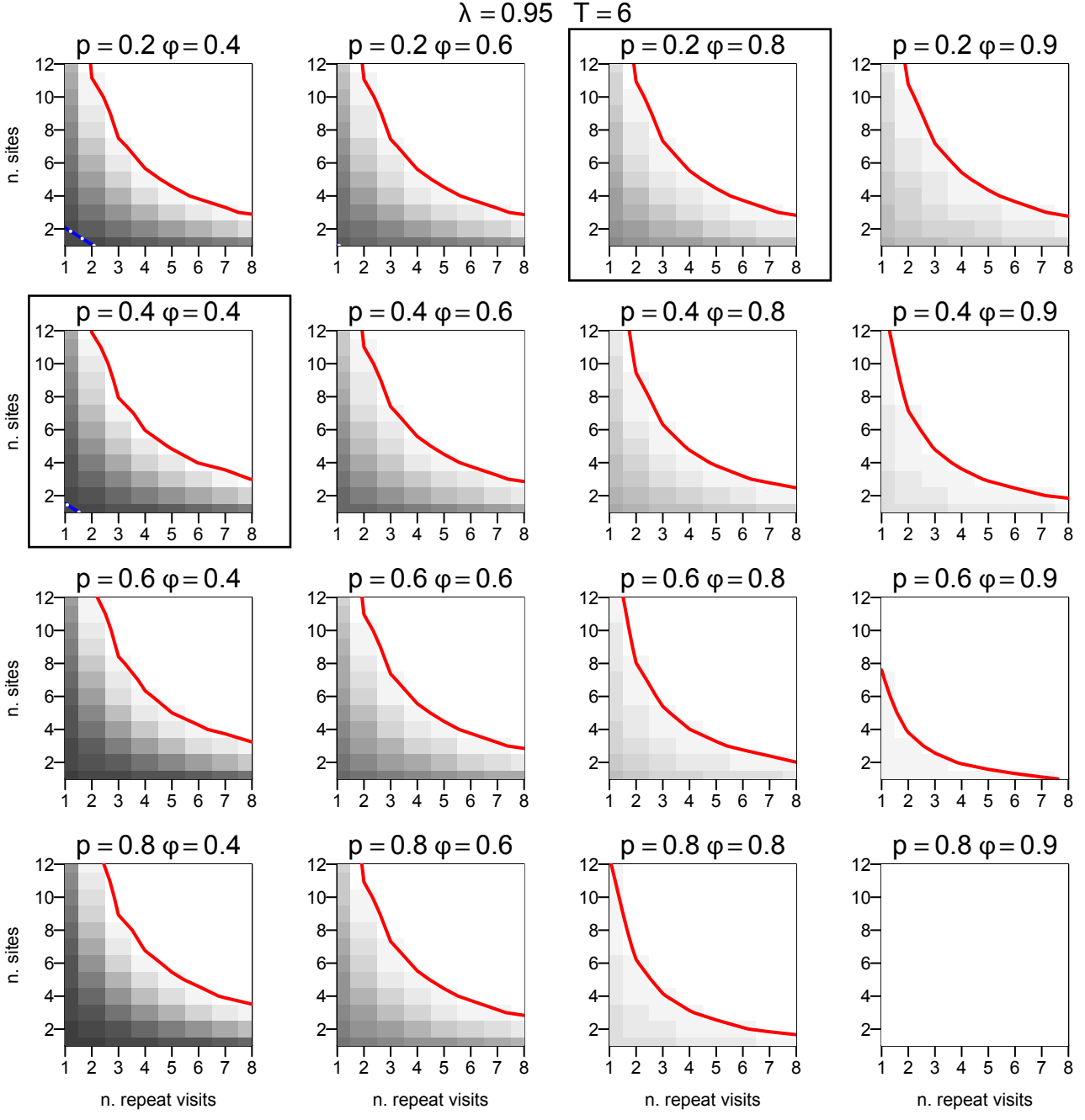
FIGURE LEGENDS

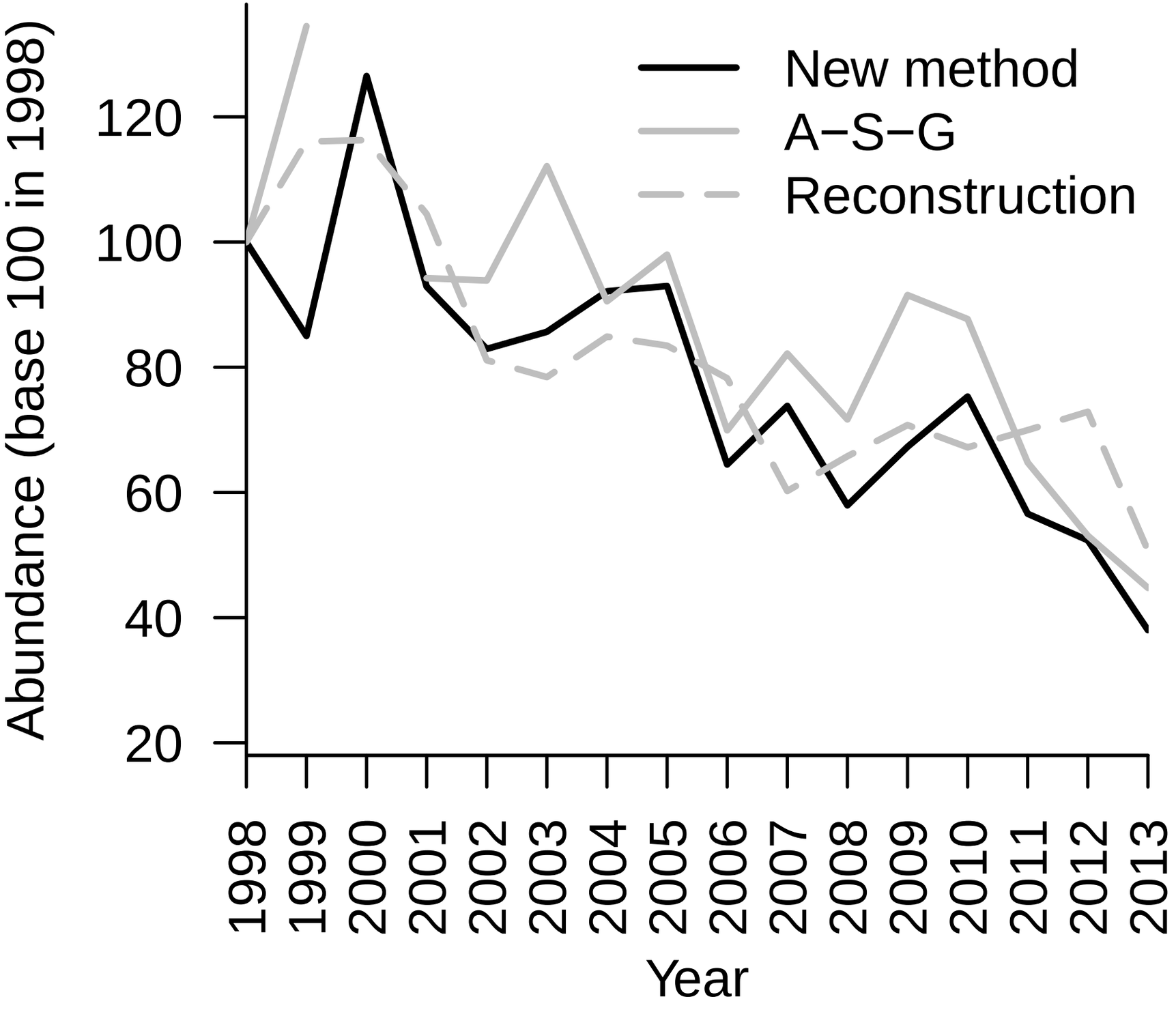
Fig. 1: Quantification of the loss of precision in a scenario without any variance in nuisance parameters. Probability of not detecting an annual rate of change (increase or decrease, at random) of 5% over 6 years, for various scenarios of detection probability p and availability probability φ using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

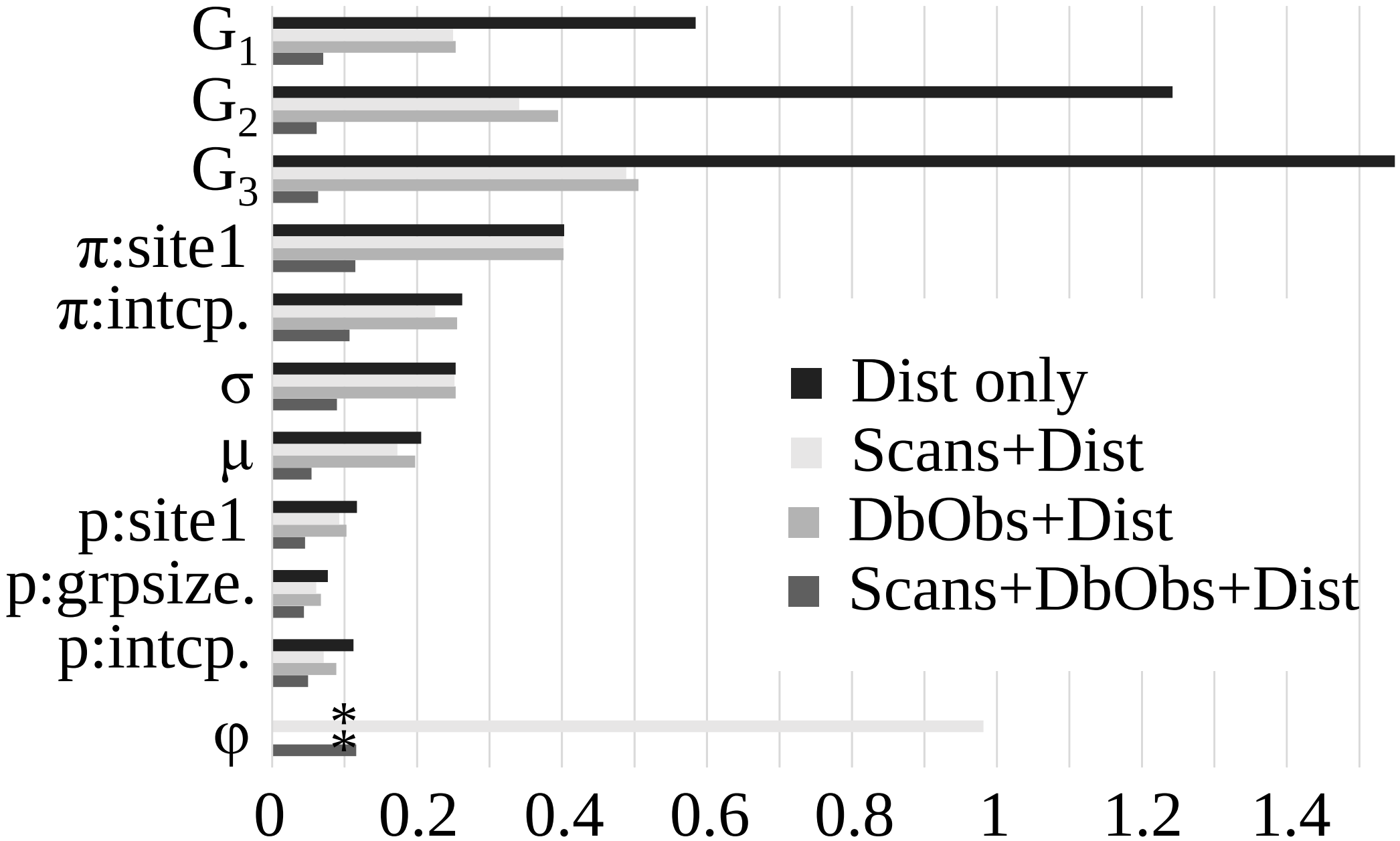
Fig. 2: Comparison of estimated chamois abundance in the Bazès study area, with the Arnason-Schwarz-Gerard model fitted to resighting data from marked animals ('A-S-G'), with a 10 age-class population model with demographic rates estimated from individual-based data ('reconstruction'), and with our new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

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

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


λ = 0.95 T = 6
p = 0.2 φ = 0.4
p = 0.2 φ = 0.6
p = 0.2 φ = 0.8
p = 0.2 φ = 0.9
p = 0.4 φ = 0.4
p = 0.4 φ = 0.6
p = 0.4 φ = 0.8
p = 0.4 φ = 0.9
p = 0.6 φ = 0.4
p = 0.6 φ = 0.6
p = 0.6 φ = 0.8
p = 0.6 φ = 0.9
p = 0.8 φ = 0.4
p = 0.8 φ = 0.6
p = 0.8 φ = 0.8
p = 0.8 φ = 0.9
n. sites
n. repeat visits


New method
A−S−G
Reconstruction
Abundance (base 100 in 1998)
120
100
80
60
40
20
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Year


G1
G2
G3
π:site1
π:intcp.
σ
μ
p:site1
p:grpsize.
p:intcp.
φ
*
*
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Dist only
Scans+Dist
DbObs+Dist
Scans+DbObs+Dist