# Estimating Population Size with Noninvasive Capture-Mark-Recapture Data

ERIC PETIT* AND NATHANIEL VALIERE†

*Laboratoire Ethologie Evolution Ecologie, UMR CNRS 6552, Université Rennes I, Station Biologique, 35380 Paimpont, France, email eric.petit@univ-rennes1.fr
†Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Lyon I, Villeurbanne, France

**Abstract:** *Estimating population size of elusive and rare species is challenging. The difficulties in catching such species has triggered the use of samples collected noninvasively, such as feces or hair, from which genetic analysis yields data similar to capture-mark-recapture (CMR) data. There are, however, two differences between classical CMR and noninvasive CMR. First, capture and recapture data are gathered over multiple sampling sessions in classical CMR, whereas in noninvasive CMR they can be obtained from a single sampling session. Second, because of genotyping errors and unlike classical CMR, there is no simple relationship between (genetic) marks and individuals in noninvasive CMR. We evaluated, through simulations, the reliability of population size estimates based on noninvasive CMR. For equal sampling efforts, we compared estimates of population size* N *obtained from accumulation curves, a maximum likelihood, and a Bayesian estimator. For a closed population and without sampling heterogeneity, estimates obtained from noninvasive CMR were as reliable as estimates from classical CMR. The sampling structure (single or multiple session) did not alter the results, the Bayesian estimator in the case of a single sampling session presented the best compromise between low mean squared error and a 95% confidence interval encompassing the parametric value of* N *in most simulations. Finally, when suitable field and lab protocols were used, genotyping errors did not substantially bias population size estimates (bias < 3.5% in all simulations). The ability to reliably estimate population size from noninvasive samples taken during a single session offers a new and useful technique for the management and conservation of elusive and rare species.*

**Keywords:** Bayesian estimator, CMR, genetic tag, noninvasive sampling, population size

Estimación del Tamaño Poblacional con Datos de Captura-Marca-Recaptura No Invasivos

**Resumen:** *La estimación del tamaño poblacional de especies elusivas y raras es un reto, Las dificultades para capturar a esas especies ha llevado al uso de la recolección de muestras no invasivas como heces o pelo, cuyos análisis genéticos proporcionan datos similares a los de captura-marca-recaptura (CMR). Sin embargo, hay dos diferencias entre CMR clásica y CMR no invasiva. Primero, los datos de captura y recaptura son recolectados en sesiones múltiples de muestreo en CMR clásico, mientras que en CMR no invasivo se pueden obtener en una sola sesión de muestreo. Segundo, debido a errores en la determinación del genotipo y a diferencia de CMR clásico, no hay una relación simple entre marcas (genéticas) e individuos en CMR no invasiva. Por medio de simulaciones evaluamos la confiabilidad de las estimaciones de tamaño poblacional basadas en CMR no invasiva. Para esfuerzos de muestreo comparamos estimaciones del tamaño poblacional* N *obtenidas de curvas de acumulación, una probabilidad máxima y un estimador Bayesiano. Para una población cerrada y sin heterogeneidad de muestreo, las estimaciones obtenidas de CMR no invasivo eran tan confiables como las estimaciones de CMR clásico. La estructura del muestreo (sesión única o múltiple) no alteró los resultados, en el caso de una sola sesión de muestreo el estimador Bayesiano presentó el mejor arreglo entre el menor error promedio y un intervalo de 95% de confianza en el valor paramétrico de* N *en la mayoría de las simulaciones.*

*The two authors contributed equally to this work.*
*Paper submitted May 24, 2005; revised manuscript accepted September 21, 2005.*



Conservation Biology Volume 20, No. 4, 1062–1073

DOI: 10.1111/j.1523-1739.2006.00417.x

*Finalmente, cuando se utilizaron protocolos de campo y laboratorio adecuados, los errores en la determinación de genotipos no sesgaron sustancialmente a las estimaciones de tamaño poblacional (sesgo < 3.5% en todas las simulaciones). La habilidad para estimar el tamaño poblacional confiablemente a partir de muestras no invasivas recolectadas en una sola sesión ofrece una técnica nueva y útil para la gestión y conservación de especies elusivas y raras.*

**Palabras Clave:** CMR, estimador Bayesiano, marcador genético, muestreo no invasivo, tamaño poblacional

## Introduction

Capture-mark-recapture (CMR) experiments were developed to tackle the difficulties associated with the estimation of population size in mobile animals. The general principle of CMR experiments is to mark individuals in a first capture session and then to record the proportion of marked individuals in subsequent recapture sessions (Williams et al. 2001). In the simplest model, population size $N$ is then estimated from the ratio of marked to unmarked individuals in recapture sessions (e.g., Seber 1973), assuming that all individuals (marked and unmarked) randomly mixed after the first capture event and are thus all equally catchable during the recapture sessions. This simple principle has led to the statistical development of a variety of estimators of the population size $N$ (e.g., Darroch 1958; Seber 1973; Otis et al. 1978; Gazey & Staley 1986). However, it remains extremely difficult to obtain reliable estimates of population size in species that are difficult to catch, such as elusive or rare species, or to handle.

The need for abundance estimates in management programs of species with overall low catchability has triggered the use of molecular tags (Palsbøll et al. 1997; Kohn et al. 1999). Multilocus genotypes obtained, for instance, from microsatellites, can be used to discriminate samples on the basis of their allelic composition (Palsbøll 1999; Taberlet & Luikart 1999). When samples consist of noninvasively collected hairs or feces, molecular tags obtained without handling an animal can be used to study individual home ranges (Taberlet et al. 1997), dispersal (Gerloff et al. 1999; Lucchini et al. 2002), or paternities (Gerloff et al. 1999; Constable et al. 2001), or to estimate population size (e.g., Kohn et al. 1999; Mowat & Paetkau 2002; Eggert et al. 2003; Bellemain et al. 2005). Estimates of population size based on genetic tags can be derived following the principle described above, as long as samples are taken in a multisession sampling experiment that mimics the capture and recapture sessions of CMR experiments (Palsbøll et al. 1997; Bellemain et al. 2005; Prugh et al. 2005). But genetic tags also allow estimating population sizes from single sampling sessions. In a set of $n$ noninvasive samples collected in a single session, a subsample of $m$ different samples yields $m^{-1}$ "recaptures" if they bear the same genetic tag (and thus, presumably come from a single individual). Classical CMR estimators of population size rely on multisession sampling and therefore cannot be used to estimate $N$ from single-session sampling. Rather, when samples come from a single session, molecular ecologists have simply estimated $N$ from the asymptote of accumulation curves, which are plots of the number of unique molecular tags against the number of analyzed samples (Kohn et al. 1999; Eggert et al. 2003), a problem similar to the estimation of species diversity in an area (Colwell & Coddington 1994).

The potential for estimating population size from single-session sampling experiments is encouraging, especially in species for which field work is expensive or time consuming. However, it remains to be shown that single-session sampling is as reliable as multisession sampling to estimate $N$. Classical CMR estimates of population size rely on explicit modeling of the capture–recapture process (Otis et al. 1978). This approach presents the great advantage that it can take into account various sources of capture heterogeneity (i.e., when all stage/age classes do not have the same capture probability or when capture probability is not constant over time). In contrast, accumulation-curve estimates of $N$ simply rely on fitting capture data to the equation of an asymptotic curve, without explicit reference to the capture–recapture process. It is thus impossible to explicitly take capture heterogeneity into account with accumulation-curve methods. The single parameter with biological relevance in these equations is the value of the asymptote, which is assumed to be an estimate of $N$ (Kohn et al. 1999; Eggert et al. 2003). However, Eggert et al. (2003) show that all equations are not equal, with the equation proposed by Kohn et al. (1999) being strongly biased upward (equations in Methods).

Working with molecular tags adds another type of problem, namely that, unlike physical marks, multilocus genotypes cannot always be assigned to a unique individual with 100% confidence. First, two different individuals can share the same tag. This kind of error, called the "shadow effect" (Mills et al. 2000), is likely when populations consist of highly related individuals and/or when the loci chosen to build the molecular tags lack variability. There is thus some risk of considering samples that originate from different individuals as identical. This leads to underestimates of population size (Mills et al. 2000). Second, DNA extracted from noninvasively collected samples is particularly prone to genotyping errors (Taberlet et al. 1999).

These typically lead to overestimates of population size (Waits & Leberg 2000) because incorrect genotypes are wrongly recognized as genuine individuals. Genotyping error rates can, however, be considerably lowered when following suitable protocols at field, laboratory, or analysis level (see Paetkau 2003 for a review in the context of population size estimation).

Our primary aim in this simulation study was to investigate the reliability of population size estimates based on single-session sampling of noninvasively collected samples. To achieve this, we compared population size estimation methods based on single-session and multisession sampling schemes. We were particularly interested in testing whether single-session sampling leads to estimates that are as reliable as multisession sampling for similar sampling effort. We extended our comparison to a number of methods that were not yet used to estimate population sizes from molecular tags. In particular, we evaluated a Bayesian estimator of $N$ (Gazey & Staley 1986) that is usable for data obtained from single-session sampling because it explicitly models the capture-recapture process. Finally, we quantified the bias introduced with the use of molecular tags (i.e., taking into account the correction of genotyping errors as performed in realistic laboratory conditions).

## Methods

### Simulations

Simulations were performed using GEMINI 1.4.1, a software designed to simulate all steps of a noninvasive genetic survey with multilocus genotypes to identify individuals (Valière et al. 2002). There are four simulation steps in the program. First, a population is built in which all individuals have genotypes determined using specified allele frequencies. Allele frequencies were from one population of noctule bats (*Nyctalus noctula*) (Petit & Mayer 1999). This population was chosen because it has the lowest probability of identity (i.e., the lowest probability that two distinct individuals share an identical multilocus genotype; Waits et al. 2001) among the 13 studied populations of noctule bats. The data set is composed of eight loci (alleles, heterozygosity, and probability of identity in Table 1). Probability of identity is low enough ($4.38 \times 10^{-13}$ for the unbiased probability of identity and $3.23 \times 10^{-4}$ for the probability of identity of siblings; see Waits et al. 2001) to correctly discriminate individuals so that the shadow effect should not be a serious problem. The size of the simulated populations was set to 100, but additional simulations were run with populations of 1000 individuals and yielded similar results (see also Waits & Leberg 2000).

Second, a given number of samples ($n$) were taken in this population over a given number of sampling sessions. Random sampling from the population was performed with replacement to simulate noninvasive sampling, with $n = 50$, 100, 150, and 300 samples to test several sampling-effort scenarios. Five sampling occasions were performed for multiple-session methods. Sample sizes were kept constant when comparing single- and multiple-estimation-session methods (see below). For instance, for a sample size of 50, there were either 50 samples from a single sampling session or 10 samples from each of 5 sessions. All individuals had the same probability of being sampled (no capture heterogeneity).

Third, all samples were genotyped, which means genotypes were modified given error rates. Genotyping errors can be of two main kinds (Taberlet et al. 1996). Stochastic sampling of alleles in a diluted DNA extract can lead to the amplification of only one allele in heterozygotes. This is called allelic dropout (ADO). False alleles (FA) is a second type of error that arises when alleles are amplified in addition to or instead of the true alleles (for instance due to slippage of the polymerase during PCR). Error rates (ADO/FA) were set to 0.3/0.1 for each locus. These are realistic rates derived from several noninvasive

**Table 1. Characteristics of the set of loci used for simulating noninvasive surveys of population size (data from Petit & Mayer 1999).**

| | *Statistic*[a] | | | | | |
|---|---|---|---|---|---|---|
| *Locus* | *Na* | *Na < 0.05* | $H_o$ | $H_e$ | *PIunb* | *PIsib* |
| P2 | 8 | 5 | 0.58 | 0.74 | $7.77 \times 10^{-2}$ | $4.03 \times 10^{-1}$ |
| P8 | 8 | 4 | 0.75 | 0.76 | $6.32 \times 10^{-2}$ | $3.92 \times 10^{-1}$ |
| P13 | 12 | 8 | 0.88 | 0.89 | $1.07 \times 10^{-2}$ | $3.13 \times 10^{-1}$ |
| P18 | 5 | 4 | 0.42 | 0.65 | $1.30 \times 10^{-1}$ | $4.67 \times 10^{-1}$ |
| P20 | 19 | 13 | 0.92 | 0.93 | $1.59 \times 10^{-3}$ | $2.88 \times 10^{-1}$ |
| P217 | 11 | 8 | 0.79 | 0.88 | $1.12 \times 10^{-2}$ | $3.14 \times 10^{-1}$ |
| P219 | 5 | 4 | 0.54 | 0.69 | $1.28 \times 10^{-1}$ | $4.44 \times 10^{-1}$ |
| P223 | 10 | 8 | 0.75 | 0.83 | $2.81 \times 10^{-2}$ | $3.48 \times 10^{-1}$ |
| Mean | 9.75 | 6.75 | 0.7 | 0.8 | $4.38 \times 10^{-13}$[b] | $3.23 \times 10^{-4}$[b] |

[a] *Statistics: Na, number of allele; Na < 0.05, number of alleles with frequency <0.05;* $H_o$*, observed heterozygosity;* $H_e$*, expected heterozygosity; PIunb, unbiased probability of identity; PIsibs, probability of identity for sibs.*

[b] *Product of all single loci values.*

studies (e.g., Gagneux et al. 1997; Lathuillière et al. 2001; Lucchini et al. 2002).

Fourth, for each sample, a consensus genotype was built according to a given consensus genotype rule. Following Taberlet et al. (1996), eight PCR amplifications per locus and sample were simulated and consensus genotypes were constructed using the threshold method implemented in GEMINI. Alleles were saved in the consensus genotypes if they appeared at least three times among the eight amplifications because the probability of observing a particular false allele is low (around 5%; Taberlet et al. 1996). These rules theoretically reduce genotyping errors due to allelic dropout and false allele to low rates (Taberlet et al. 1996). Errors that remain after multiple amplifications and the application of the consensus rule are called residual errors.

We built two data files from the simulations. The first one contained the identities of individual samples obtained after sampling the population and was constructed from step 2 of the simulation process (this is the file called SPL in GEMINI). The second file contained consensus genotypes and was constructed after step 4 of the simulation process (this is the file called CONS in GEMINI). The difference between the two types of data files is thus simply the possibility given by the genotyping process to introduce residual errors in sample identities. This allowed us to evaluate the bias introduced by these residual errors in the population size estimate.

The main assumptions of the simulations were a stationary (i.e., no birth nor death) and closed (i.e., no migration) population, so population size did not change over the study period; an equal capture probability for all individuals; and a recapture probability that was identical to the capture probability and that did not vary during the study. We ran simulations under different degrees of capture effort and replicated each condition 200 times.

## Population Size Estimation

We compared six methods of population size estimation that all use information on recapture of marked animals. Population sizes were estimated with a maximum-likelihood method (MLM), three different equations of accumulation-curve methods (ACM), and two versions of a Bayesian method (BM). All methods were originally individual-based methods, but in our study we used the genotype-based approach (e.g., Schwartz et al. 1998), in which unique and identical genotypes detected during a sampling occasion are pooled and considered a single individual. We then used unique genotypes as individual marks in the mathematical models to estimate population size.

### MAXIMUM-LIKELIHOOD METHOD (MLM)

Otis et al. (1978) proposed various maximum-likelihood models to estimate population sizes with capture-mark-recapture data. These models differ in their underlying assumptions. The simplest one, M0, is a model for a closed population and without any sampling heterogeneity. We used M0 because it fits the hypotheses of our simulations. We used CAPTURE (Otis et al. 1978) to estimate population size from data from the five sampling sessions. In such a case, the maximum-likelihood estimator $\hat{N}$ of $N$ satisfies:

$$\ln L(\hat{N}\,|X) = \max_{N_i \in \{M, M+1, M+2, \ldots\}} \left[ \ln \frac{N_i!}{(N_i - M)!} + n \ln n + (5N_i - n)\ln(5N_i - n) - 5N_i \ln 5N_i \right],$$

where $X$ is for the data, $M$ is the total number of distinct individuals caught (here, genotypes sampled) during the experiment, and $n$ is the total number of captures (or samples taken) during the study (see pp. 14–16 and Appendix B in Otis et al. [1978]).

### ACCUMULATION (OR RAREFACTION) CURVE METHODS (ACM)

Accumulation-curve methods were originally used for the estimation of species diversity in an area (Colwell & Coddington 1994). The estimation of the number of individuals of a population is similar to this problem, so an ACM approach could also be used for the estimation of the number of individuals in an area (Kohn et al. 1999). The principle of the ACM approach is to fit the cumulative number of different species/individuals/genotypes to the number of newly discovered species/individuals/genotypes. The asymptote of the curve is an estimation of the total number of entities present in the area. Accumulation data can be fitted to various equations, of which three have been used by molecular ecologists. We evaluated the three equations here.

Kohn et al. (1999) proposed to use the equation $y = \frac{ax}{b+x}$ (where $y$ is the number of unique genotypes, $x$ is the number of samples analyzed, $a$ is the value of the asymptotic number of unique genotypes, and $b$, which has no obvious biological interpretation, is related to the rate of decrease of the slope of the asymptote) to estimate the number of coyotes (*Canis latrans*) in the Santa Monica Mountains (California, U.S.A.). Eggert et al. (2003) used a second equation to estimate population size of forest elephant (*Loxodonta cyclotis*): $y = a\,(1 - e^{-bx})$. A third equation was proposed from the classical occupancy problem (D. Chessel, personal communication) and has been applied to the estimation of population size in the Eurasian badger (*Meles meles*, Frantz et al. 2004). This equation represents the expectation of the number of full boxes (unique genotypes, $y$) when balls (samples, $x$) are randomly distributed into boxes (individuals, $a$): $y = a - a\,(1 - \frac{1}{a})^x$.

We applied these three approaches by pooling the data from the five sampling sessions into a single session. We

used script files provided by GEMINI and a nonlinear model available in the *nls* function of R software (Ihaka & Gentleman 1996) to fit data to each equation. Because the order in which samples are analyzed has an influence on the estimation of the population size (Kohn et al. 1999), samples were randomly permuted 50 times for each replicate. We then calculated estimates of population size as the mean of the asymptotes over the 50 curves. This permutation procedure also provided approximate confidence intervals estimated from the standard deviation over the 50 curves (Kohn et al. 1999; Eggert et al. 2003).

#### BAYESIAN METHODS (BM)

We used the sequential Bayesian method proposed by Gazey and Staley (1986) to estimate $N$ based on a noninformative prior distribution (all possible population sizes have the same probability to be the true one) and on three attributes obtained from the capture–recapture history of each individual (i.e., the number of marked individuals at the start of each sampling occasion, the total number of individuals sampled during each sampling occasion, and the number of recaptures appearing during each sampling occasion). This method was originally designed for multisession sampling (later on referred to as BM-multiple) but can readily be adapted to single-session sampling if one considers each capture a sampling session (later on referred to as BM-single). For this Bayesian algorithm, the order in which the samples are analyzed does not influence the population size estimation because the method is an iterative process that always gives the same final posterior distribution. As a Bayesian point estimator $\hat{N}$ of $N$, we used the mode of the posterior distribution of $N$, which satisfies:

$$\Pr(\hat{N}\,|X) = \max_{N_i \in \{M, M+1, M+2, \ldots\}} \prod_{t=1}^{T} \left(\frac{M}{N_i}\right)^{r_t} \left(1 - \frac{M}{N_i}\right)^{n_t - r_t},$$

where $X$ is for the data, $M$ is the total number of distinct individuals caught (here, genotypes sampled) during the experiment, $n_t$ is the number of individuals caught (samples taken) during capture (sampling) session $t$, $r_t$ is the number of recaptures during capture (sampling) session $t$, and $T$ is the total number of capture (sampling) sessions (Gazey & Staley 1986). This estimation method can be implemented as a sequential Bayesian algorithm, the sequence being implemented over the $T$ capture sessions. In this algorithm, the uniform prior distribution is thus only used to compute the posterior distribution for $t = 1$. From $t = 2$ onward, each estimation step uses as a prior distribution for $N$ the posterior distribution computed at the previous step, until $t$ reaches $T$. An estimate for $N$ is obtained from the mode of the posterior distribution at $t = T$ (for more details about the algorithm, see Gazey & Staley 1986). This procedure uses some information (the number of recaptures) not used in model M0, so the Bayesian and the maximum-likelihood estimates are likely to differ.

The Bayesian method needs three initial parameters to estimate population size: the minimal and maximal population sizes tested and the incremental interval between two tested sizes. These parameters were set for each condition: the minimal population size $N_{\min}$ was set as the number of unique genotypes detected in the complete sample; the maximal population size $N_{\max}$ was set high enough to include the highest probability density interval (Gazey & Staley 1986); the incremental parameter $N_k$ was set so that all integer population sizes between and including $N_{\min}$ and $N_{\max}$ were tested. The method was implemented in the R software and script files were provided by GEMINI.

### Statistical Comparison of Estimates

For each of the four sampling-effort conditions, we obtained six different estimates from the six methods. Additionally, estimates were produced based on data from either sample identities or consensus genotypes. Sample identities allowed comparisons of the six different methods irrespective of problems linked to the use of genetic techniques. We used consensus genotypes to investigate how laboratory procedures, and in particular problems inherent to the genetic analysis of noninvasive samples, can interfere with population size estimation. Our aim here was not to quantify the bias introduced by different types of genotyping errors but rather to check whether residual errors in genotypes substantially bias population size estimates.

In each simulation case, we calculated the mean squared error (MSE, also called quadratic mean error), which is equal to the squared bias plus the variance. A low MSE is characteristic of a good trade-off between low bias and low variance, and we considered a method "better" than the others if it had a low MSE. For two methods with similar MSE, the trade-off between bias and variance was considered.

We also assessed the precision of the estimation using 95% confidence intervals (95% CI). In particular, we evaluated the percentage of simulations in which 95% CI included the true value of $N$. For MLM, confidence intervals corresponded to the approximate 95% CI given in the output file (see Otis et al. 1978, p. 17 and Appendix O). For BM, confidence intervals corresponded to the highest probability density (more accurate than the 2.5% and 97.5% quantiles, see Gazey & Staley 1986). For ACM, the confidence intervals were constructed from the standard deviation of the distribution of estimates over the 50 permutations with the classical equation 95% CI $= 2 \times 1.96 \times \frac{\mathrm{SD}}{\sqrt{N_{\mathrm{it}}}}$, where SD is the standard deviation of estimates over the 50 iterations and $N_{\mathrm{it}}$ is the number of iterations ($N_{\mathrm{it}} = 50$).

## Results

### Estimates from the Six Methods Based on Sample Identities

The accumulation-curve method used in Kohn et al. (1999) gave the most biased estimates in all sampling-effort conditions. Mean bias ranged from 136% for sampling effort $n = 300$ to 266% for sampling effort $n = 50$. The MSE for this method was at least 8 (and up to 200) times greater than for the other methods, whatever the sampling effort. The ACM-Kohn was thus removed from further analyses.

For all methods, MSE, bias, and variance decreased as sampling effort increased (Fig. 1). All methods overestimated population size for sample sizes up to $n = 150$ (bias $> 0$, Fig. 1). Overall, the bias became negligible at $n = 150$, and the variance became negligible at $n = 300$. Difference in MSE between methods seemed to become negligible when sampling effort was higher than $n = 150$. However, for $n = 50$, the lowest sampling effort, ACM-Eggert gave the worst results and BM-single and ACM-Chessel yielded the best results (with small difference with BM-multiple and MLM-M0). The poor performance of ACM-Eggert at small sample sizes was explained by both a higher bias and a higher variance than other methods. This method was consistently more biased than the others up to $n = 150$.

As expected, the width of the confidence intervals tended to decrease as sampling effort increased (Fig. 2). For $n = 50$, ACM-Eggert and ACM-Chessel had narrower confidence intervals than other methods. However, and perhaps as a consequence, for these two methods, the true population size was included in confidence intervals in only 9.5–42% of simulations, whatever the sampling size (Fig. 3). For the three other methods (MLM-M0, BM-multiple, and BM-single), this proportion was at least 82%.

### Estimates Obtained from Consensus Genotypes

With initial allelic dropout and false allele error rates respectively set to 0.3 and 0.1, the mean percentage of

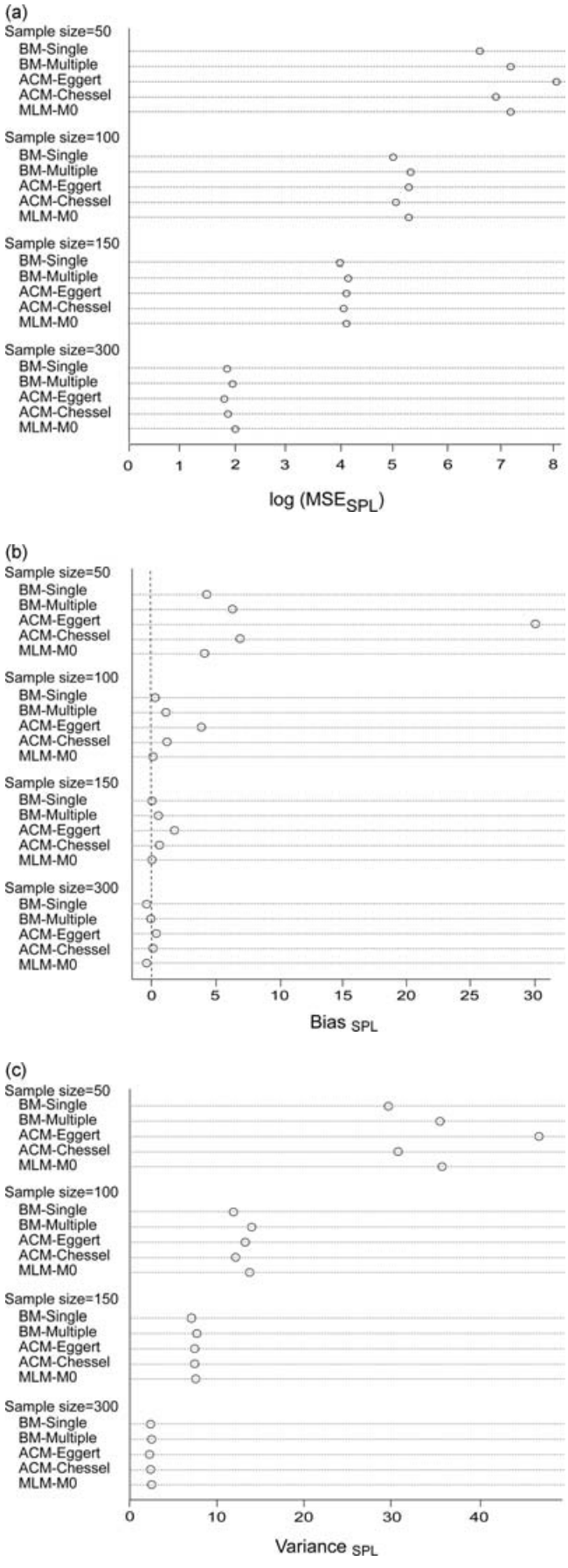


*Figure 1. Minimum squared error (MSE), bias, and variance for each method of population size estimation (MLM-M0, maximum-likelihood method, model M0 of Otis et al. [1978]; ACM-Chessel and ACM-Eggert, accumulation-curve methods, equations given by D. Chessel [personal communication] and Eggert et al. [2003]; BM-multiple and BM-single, Bayesian methods developed for multiple- or single-session sampling schemes respectively [Gazey & Staley 1986]) and sampling-effort modality. Population size estimates based on sample identities (SPL).*

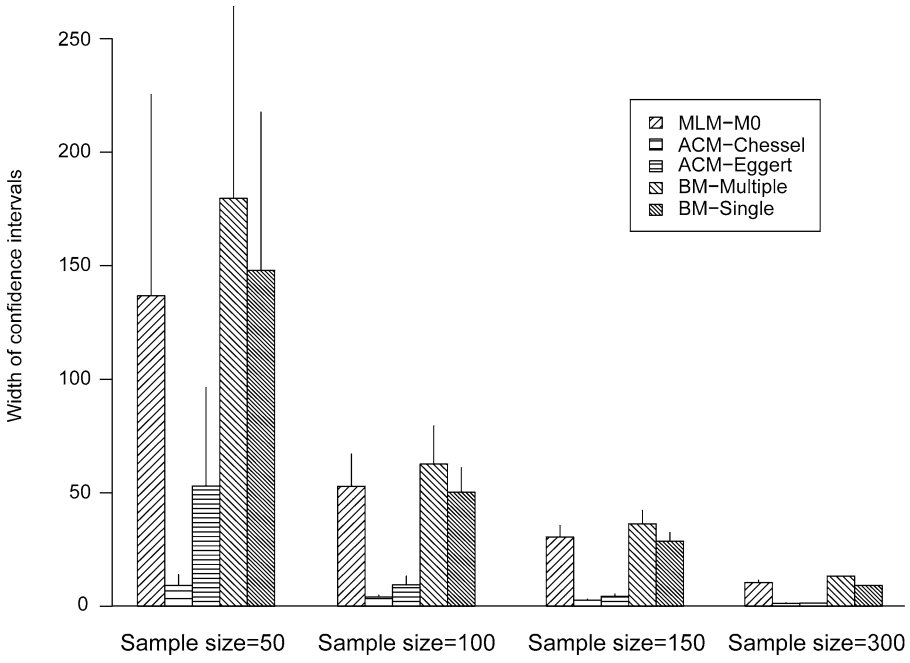


*Figure 2. Width of confidence intervals (mean over all iterations ± SD) for each method of population size estimation (MLM-MO, ACM-Chessel, ACM-Effert, BM-multiple, BM-single) in each sampling-effort modality. See Fig.1 legend for definitions of abbreviations of methods of population size estimation.*

correct identifications was 8.9 ± 2.83 (0–18%) for the uncorrected genotypes (i.e., genotypes read from step 3 of the simulation process) and 99.3 ± 0.85 (94–100%) for the consensus genotypes (i.e., genotypes built after step 4 of the simulation process for each locus from the comparison of the eight amplifications simulated for each sample). The multilocus residual error rates after correction was always below 6% and averaged 0.7%. The effect of residual errors after eight PCR amplifications on population size estimation was between 1.5% and 3.5%, depending on the method and sample size (Fig. 4).

The average population size estimates from sample identities and consensus genotypes were significantly different from each other only in the case of the highest

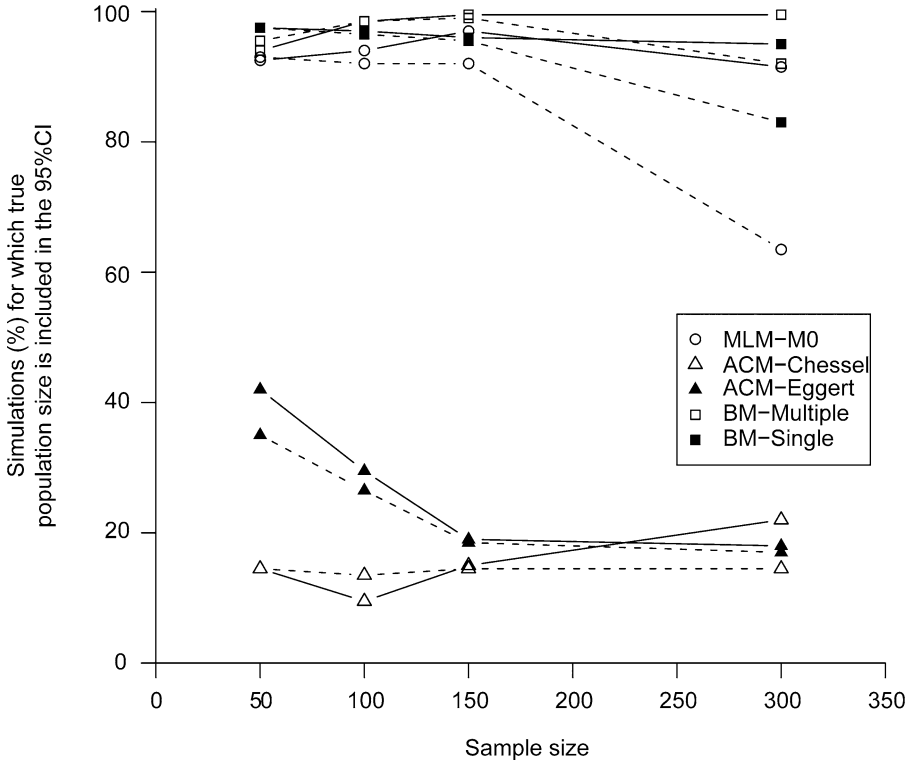


*Figure 3. Proportion of simulations of surveys of population size for which the true population size is included in the 95% CI for each estimation method (MLM-MO, ACM-Chessel, ACM-Effert, BM-multiple, BM-single; sample identities, solid lines; consensus genotypes, dashed lines). See Fig.1 legend for definitions of abbreviations of methods of population size estimation.*

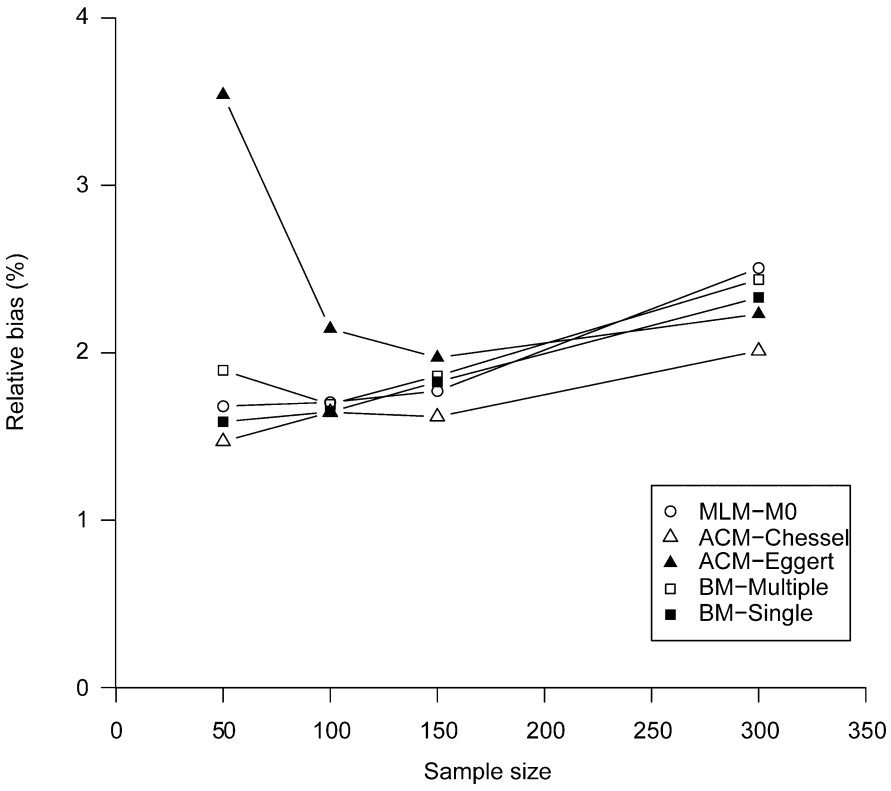


*Figure 4. Bias introduced by genotyping errors relative to estimates of population size obtained from sample identities. Relative bias was computed as 100 × (CONS-SPL)/SPL, where SPL is the estimate of population size obtained from sample identities and CONS is the estimate of population size obtained from consensus genotypes. See Fig. 1 legend for definitions of abbreviations of methods of population size estimation (MLM-MO, ACM-Chessel, ACM-Effert, BM-multiple, BM-single). Data are from Table 2.*

sampling efforts for all methods (Table 2). The MSE for estimates from consensus genotypes were slightly higher than estimates from sample identities, and this was mainly due to a higher bias for consensus genotypes compared with sample identities, especially for the highest sampling effort ($n = 150$ and $n = 300$, Fig. 5). Nevertheless, differences were always below 3.5% (Fig. 4), even in the cases where both estimates were significantly different from each other. For sample sizes 50, 100, and 150, these biases resulted in a slight decrease of the percentage of simulations for which the 95% CI of the estimators included the parametric value of $N$ (Fig. 3). For $n = 300$, this decrease was more pronounced and reached 28% for the maximum-likelihood estimator.

## Discussion

As Pollock et al. (1990) stated, the first constraint when planning a survey of population size is the study design and the sampling method. Two solutions could be used: collecting samples on a single occasion or collecting samples on several occasions. These two sampling methods imply very different constraints in the field. Whereas the multisession method requires spatial and temporal plans to be set up (grid size, capture probability, number of sessions, frequencies of sampling [Otis et al. 1978; Mowat & Strobeck 2000]), the single-session method could be performed without such constraint. The single-session

**Table 2. Means of the population size estimates (true $N$ is 100) for the six methods we evaluated and for the different sampling-effort modalities for estimates based on sample identities (SPL) and estimates based on consensus genotypes (CONS).**

| *Method*[a] | *Sampling effort* | *Mean SPL* | *Mean CONS* | p[b] |
|---|---|---|---|---|
| MLM-M0 | 50 | 104.26 | 106.00 | 0.631 |
| | 100 | 100.21 | 101.90 | 0.228 |
| | 150 | 100.07 | 101.83 | 0.0281 |
| | 300 | 99.67 | 102.155 | <0.0001 |
| ACM-Chessel | 50 | 107.06 | 108.63 | 0.618 |
| | 100 | 101.26 | 102.91 | 0.182 |
| | 150 | 100.66 | 102.28 | 0.037 |
| | 300 | 100.18 | 102.19 | <0.0001 |
| ACM-Eggert | 50 | 130.32 | 134.93 | 0.336 |
| | 100 | 103.95 | 106.17 | 0.111 |
| | 150 | 101.84 | 103.84 | 0.001 |
| | 300 | 100.45 | 102.68 | <0.0001 |
| BM-multiple | 50 | 106.48 | 108.49 | 0.584 |
| | 100 | 101.20 | 102.91 | 0.233 |
| | 150 | 100.58 | 102.44 | 0.0214 |
| | 300 | 100.01 | 102.44 | <0.0001 |
| BM-single | 50 | 104.40 | 106.06 | 0.585 |
| | 100 | 100.33 | 101.98 | 0.180 |
| | 150 | 100.06 | 101.88 | 0.0153 |
| | 300 | 99.67 | 101.99 | <0.0001 |

[a] *Abbreviations: MLM-M0, maximum-likelihood method, model M0 of Otis et al. (1978); ACM-Chessel and ACM-Eggert, accumulation-curve methods, equations given by D. Chessel (personal communication) and Eggert et al. (2003); BM-multiple and BM-single, Bayesian methods developed for multiple- or single-session sampling schemes (Gazey & Staley 1986).*

[b] *Results of* t *tests between mean estimates from sample identities (SPL) and consensus genotypes (CONS) are given.*

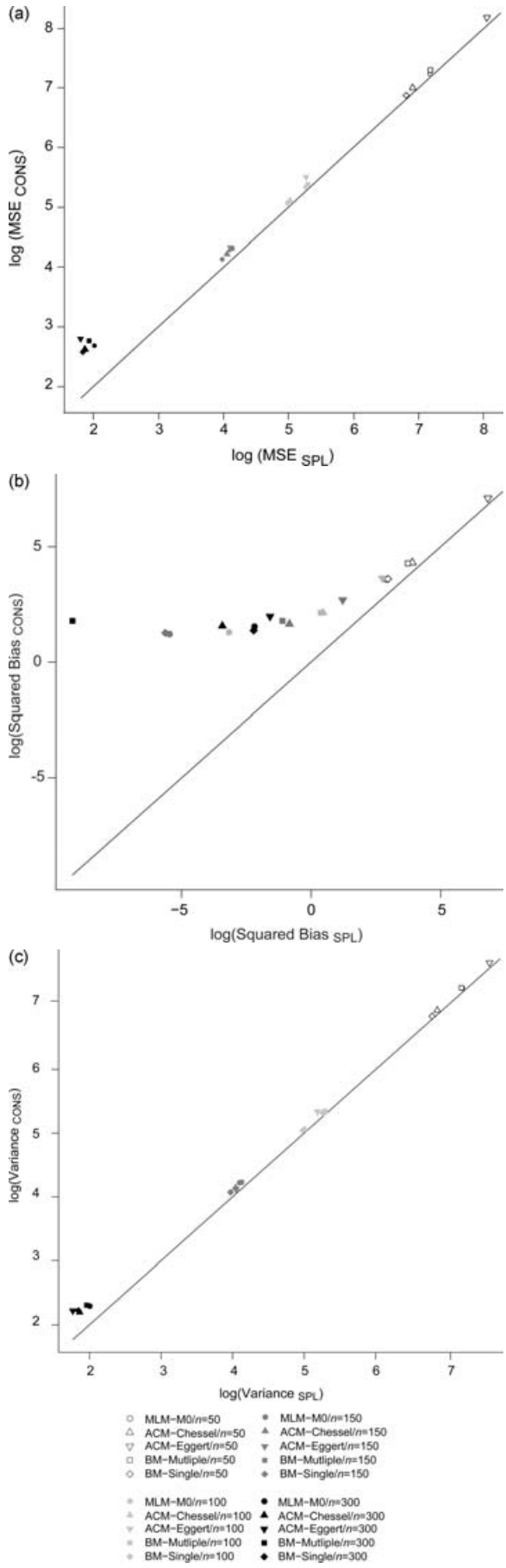


*Figure 5. Minimum squared error (MSE), bias, and variance for population size estimates from sample identities (SPL) and consensus genotypes (CONS). Solid lines show SPL = CONS. See Fig. 1 legend for definitions of abbreviations of methods of population size estimation (MLM-MO, ACM-Chessel, ACM-Effert, BM-multiple, BM-single). The* n *values in the key correspond to the different sampling-effort modalities.*

method thus has a practical advantage over the multisession method because the user will select the protocol that involves less difficulty in the field. However, it is worth noting that even though the single-session method has a great advantage in the field, this sampling method is only dedicated to population size estimation. Indeed, this method could not be used to assess survival and recruitment parameters as for the multisession method (Lebreton et al. 1992; Pradel & Lebreton 1999). The choice of the sampling method will thus also be constrained by the purpose of the study.

In practice, sampling during only one session versus sampling over two or more sessions can be seen as an arbitrary classification. Sometimes, the number of sessions is defined a posteriori, without any reference to a sampling scheme. In noninvasive CMR studies so far, samples have been collected without interruption over days (e.g., Kohn et al. 1999; Banks et al. 2003; Wilson et al. 2003) or tens of days (e.g., Mowat & Strobeck 2000; Mowat & Paetkau 2002; Prugh et al. 2005) and have then been analyzed with accumulation-curve methods, the total sampling period corresponding to a single sampling session (Kohn et al. 1999; Wilson et al. 2003), or with classical CMR models (Otis et al. 1978), the sampling period being arbitrarily split into several sampling sessions corresponding to days (Banks et al. 2003; Wilson et al. 2003), weeks (Mowat & Strobeck 2000; Mowat & Paetkau 2002), or months (Prugh et al. 2005). There thus seems to be little justification to distinguish between single-session and multisession methods.

There are however three reasons why such a distinction is useful, especially in the context of noninvasive CMR. First, analytical models available for the analysis of multisession CMR experiments (such as MLM-M0) allow each individual to be captured only once during a given capture session. In contrast, the Bayesian procedure we adapted to single-session experiments takes into account all recaptures. Because noninvasive sampling frequently yields multiple captures for a single individual during a sampling session, not taking these recaptures into account would result in a loss of information.

Second, sampling and analyzing data over multiple sessions implicitly implies that samples taken during the different sampling sessions are independent of each other. This is particularly problematic when sampling feces in

the field. Depending on species and environmental conditions, feces can persist a very long time (100 days for forest elephants, Eggert et al. 2003; years for European bats, E.P., personal observation), so samples taken during a given sampling session may have been laid down at a time corresponding to another sampling session. Possible precautions include sampling only fresh feces, which reduces nonindependence between sampling sessions but can reduce dramatically the number of available samples, or sampling over sessions separated by great time periods, which dramatically reduces the probability of population closure. Alternatively, not having to split sampling periods into different sessions avoids the problem of nonindependence between sampling sessions when collecting feces in the field.

Third, for species in which feces are accumulated in latrines, such as badgers (Wilson et al. 2003), or roosts, such as bats, a representative sample of a population can be obtained if feces are collected over a sampling period that allows all individuals of the population an equal chance of having defecated in the site that is sampled. If $N$ is to be estimated with classical methods, this work has to be repeated one or more times. Again, this decreases the probability of population closure. Within one sampling period, however, enough material is gathered to yield an estimate of $N$ with analytical tools suitable for single-session method.

Using simulations, we showed that population size estimates based on single-session sampling are as reliable as estimates based on multisession sampling. Indeed, two of the three accumulation-curve equations and the Bayesian algorithm for single-session sampling estimated population size with a MSE comparable to the MSE of the classical M0 maximum-likelihood model or the Bayesian algorithm for multisession sampling. The Bayesian estimator that was adapted to both single-session and multisession sampling schemes even showed less bias and less variance in the case of single-session sampling. Among the three equations for accumulation curves we evaluated, the method proposed by Kohn et al. (1999) proved to be highly biased under the conditions tested. This supports results published by Eggert et al. (2003), who proposed a second equation that performed well only if sample size was large. In all sampling situations, the equation of ACM that worked best was that proposed by D. Chessel, which had almost the same bias and variance as the Bayesian estimator for single-session sampling but a narrower 95% CI under the conditions tested. This confidence interval, however, did not encompass the parameter's true value in 60–80% of the simulations. Overall, Bayesian estimators behaved better than other methods. This behavior may be due to the fact that the Bayesian method explicitly models the capture–recapture process, which is not the case of accumulation-curve methods, and it uses more information from the data than the maximum-likelihood estimator of $N$.

The need for the correction of genotyping error is essential in all studies in which noninvasive genetic sampling is used (Taberlet et al. 1999; Paetkau 2003). The most important step is the pilot study, which should be conducted before any large-scale study (Taberlet et al. 1999). The aim of the pilot study is to investigate feasibility and to estimate genotyping error rates (Broquet & Petit 2004). These estimates are then used to define the minimum number of PCR amplifications per sample and locus to be performed to correct for genotyping errors (Valière 2002). Our results showed that despite the large number of PCR amplifications (eight) used, residual errors remained and could marginally bias population size estimates. In correlation with its higher bias for small sample sizes, the ACM-Eggert was the most sensitive to genotyping errors with small sample sizes (Fig. 4). As expected, the bias introduced by residual errors increased with sample size because the more samples that are analyzed, the more likely is it to create spurious genotypes due to residual genotyping errors (see also Waits & Leberg 2000). Nevertheless, even with a sampling effort as high as three times the population size, the bias hardly reached 2.5% and inclusion of the true value of $N$ in 95% CI was only slightly altered by residual errors (except for MLM-M0 under conditions of high sampling). The bias was always positive, showing that, as expected, genotyping errors had more influence than the shadow effect in our simulations. This is likely to be the case in most realistic situations in which enough polymorphic markers are used to discriminate between related individuals.

Even if the error rates used in the simulations (allelic dropout = 0.3 and false alleles = 0.1) were in the upper part of the range of the error rates published in noninvasive studies, the bias introduced by using molecular tags was negligible. Again, cautious estimation of error rates and determination of the minimum number of PCR amplifications are required to lower genotyping errors (Valière 2002) and will be of prime importance to minimize the associated bias (see also Waits & Leberg 2000; Paetkau 2003). The figures we provide, however, are conservative because residual genotypic error rates reported in published noninvasive CMR surveys amount to a maximum of 0.39% (Kohn et al. 1999; Eggert et al. 2003; Paetkau 2003; Frantz et al. 2004; Prugh et al. 2005), whereas it was on average 0.7% in our simulation study.

The use of noninvasive sampling methods offers the possibility to estimate population size when individuals are difficult to catch. This approach could be extended to any species because, besides yielding reliable estimates of the parameter of interest, it has other interesting properties that are directly linked to the total absence of animal disturbance. All the estimation methods we used rely on a number of assumptions: (1) the population is stable and closed, (2) capture probability does not vary among individuals, (3) capture probability does not vary with time, (4) marks are not lost, and (5) marks do not alter behavior.

By definition, molecular tags are marks that cannot be lost or alter the behavior of animals. Furthermore, the absence of handling removes the problem of the effect of capture history on subsequent catchability (Cormack 1966). In a closed population, the sources of heterogeneity that are thus likely to affect population size estimates based on noninvasive sampling are the time and individual components. We already showed that single-session sampling reduces the probability of violating the population closure assumption in noninvasive studies because periods of sampling can be limited when compared with multisession sampling. For the same reason, the time component of heterogeneity is also reduced with single-session sampling compared with multisession sampling.

Individual heterogeneity however remains unchanged. For example, scent-marking behavior may vary according to sex or age, making it more likely to find samples belonging to the individuals with the more conspicuous behaviors in comparison with others. Models have been built to take heterogeneity into account (Otis et al. 1978). There are however two main problems that remain. First, available analytical tools do not allow the identification of $N$ when heterogeneity is present (Link 2003). Various distributions of capture heterogeneity can most often fit equally well a given empirical data set, but these different distributions can yield population size estimates that significantly differ from each other (Link 2003). Second, these models of capture heterogeneity were developed for multisession sampling, and we are aware of only one recent attempt to model capture heterogeneity for single-session sampling, but only for small population sizes (Miller et al. 2005). A step forward in the analysis of CMR data from single-session sampling experiments would thus be to incorporate individual heterogeneity in capture probability, for example in the sequential Bayesian framework developed by Gazey and Staley (1986).

Our preliminary results showed that individual heterogeneity leads to biased population size estimates in single-session experiments, just as it does in multisession CMR (Miller et al. 2005; N.V. and E.P., unpublished data). Sampling schemes should thus be designed carefully to minimize known sources of individual sampling heterogeneity, which are most often related to differences in sex, age, size, or reproductive status. Noninvasive sampling allows sex to be recorded using molecular techniques (e.g., Bradley et al. 2001). When feces are the source of DNA, they can also be used to infer the reproductive status of the individual that defecated (Garnier et al. 2001). Other means to increase the homogeneity of the sampled population require the researcher to have some knowledge of the biology of the species being investigated so that he or she can choose a period and a place where individuals of similar status are likely to gather. For instance, in most European bat species, before females give birth nurseries are closed entities that consist mainly of adult females, whereas individuals of different ages and sexes mix in hibernacula or swarming sites. Finally, before running analyses of population size estimates, data sets should be checked for heterogeneity (Miller et al. 2005; S. Puechmaille and E.P., unpublished data).

Population size is a parameter of paramount importance in both fundamental and applied population biology. The ability to reliably estimate population size from noninvasive samples taken during single-session sampling experiments is thus a promising step toward increased knowledge of elusive species and better management policies for endangered species.

## Acknowledgments

We thank P. Inchausti, J. Labonne, and M. Schaub, who made helpful comments on earlier drafts of this manuscript, E. Cam for sharing her experience about CMR, J.-S. Pierre for answering statistical questions, and two anonymous reviewers whose criticisms greatly enhanced the scope of this work. E.P. was supported by the Region Bretagne.

### Literature Cited

Banks, S. C., S. D. Hoyle, A. Horsup, P. Sunnucks, and A. B. Taylor. 2003. Demographic monitoring of an entire species (the northern hairy-nosed wombat, *Lasiorhinus krefftii*) by genetic analysis of noninvasively collected material. Animal Conservation **6:**1–10.

Bellemain, E., J. E. Swenson, D. A. Tallmon, S. Brunberg, and P. Taberlet. 2005. Estimating population size of elusive animals with DNA from hunter-collected feces: four methods for brown bears. Conservation Biology **19:**150–161.

Bradley B. J., K. E. Chambers, and L. Vigilant. 2001. Accurate DNA-based sex-identification of apes using non-invasive samples. Conservation Genetics **2:**179–181.

Broquet, T., and E. Petit. 2004. Quantifying genotyping errors in noninvasive population genetics. Molecular Ecology **13:**3601–3608.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London B Biological Sciences **345:**101–118.

Constable, J. L., M. V. Ashley, J. Goodall, and A. E. Pusey. 2001. Noninvasive paternity assignment in Gombe chimpanzees. Molecular Ecology **10:**1279–1300.

Cormack, R. M. 1966. A test for equal catchability. Biometrics **22:**330–342.

Darroch, J. N. 1958. The multiple-recapture census. I. Estimation of a closed population. Biometrika **45:**343–359.

Eggert, L. S., J. A. Eggert, and D. S. Woodruff. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology **12:**1389–1402.

Frantz, A. C., M. Schaul, L. C. Pope, F. Fack, L. Schley, C. P. Muller, and T. J. Roper. 2004. Estimating population size by genotyping remotely plucked hair: the Eurasian badger. Journal of Applied Ecology **41:**985–995.

Gagneux, P., C. Boesch, and D. S. Woodruff. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. Molecular Ecology **6:**861–868.

Garnier J. N., D. L. Green, A. R. Richard, H. I. Shaw, and W. V. Holt. 1998. Non-invasive diagnosis of pregnancy in wild black rhinoceros (*Diceros bicornis minor*) by faecal steroid analysis. Reproduction, Fertility and Development **10:**451–458.

Gazey, W. J., and M. J. Staley. 1986. Population estimation from mark-recapture experiments using a sequential Bayes algorithm. Ecology **67:**941–951.

Gerloff, U., B. Hartung, B. Fruth, G. Hohmann, and D. Tautz. 1999. Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (*Pan paniscus*) determined from DNA analysis of faecal samples. Proceedings of the Royal Society of London, B **266:**1189–1195.

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics **5:**299–314.

Kohn, M. H., E. York, D. A. Kamradt, G. Haught, R. Sauvajot, and R. K. Wayne. 1999. Estimating population size by genotyping faeces. Proceedings of the Royal Society of London, Series B **266:**657–663.

Lathuillière, M., N. Ménard, A. Gautier-Hion, and B. Crouau-Roy. 2001. Testing the reliability of noninvasive genetic sampling by comparing analyses of blood and fecal samples in Barbary Macaques (*Macaca sylvanus*). American Journal of Primatology **55:**151–158.

Lebreton, J. D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monograph **62:**67–118.

Link, W. A. 2003. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics **59:**1123–1130.

Lucchini, V., E. Fabbri, F. Marucco, S. Ricci, L. Boitani, and E. Randi. 2002. Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. Molecular Ecology **11:**857–868.

Miller, C. R., P. Joyce, and L. P. Waits. 2005. A new method for estimating the size of small populations from genetic mark-recapture data. Molecular Ecology **14:**1991- -2005.

Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, and D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. Ecological Applications **10:**283–294.

Mowat, G., and D. Paetkau. 2002. Estimating *Martes americana* population size using hair capture and genetic tagging. Wildlife Biology **8:**201–209.

Mowat, W., and C. Strobeck. 2000. Estimating population size of grizzly bears using hair capture, DNA profiling, and mark-recapture analysis. Journal of Wildlife Management **64:**183–193.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs **62**.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology **12:**1375–1387.

Palsbøll, P. J. 1999. Genetic tagging: contemporary molecular ecology. Biological Journal of the Linnean Society **68:**3–22.

Palsbøll, P. J., et al. 1997. Genetic tagging of humpback whales. Nature **388:**767–769.

Petit, E., and F. Mayer. 1999. Male dispersal in the noctule bat (*Nyctalus noctula*): where are the limits? Proceedings of the Royal Society of London B Biological Sciences **266:**1717–1722.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture-recapture experiments. Wildlife Society Monographs **107**.

Pradel, R., and J.-D. Lebreton. 1999. Comparison of different approaches to the study of local recruitment of breeders. Birds Study **46:** 74–81.

Prugh, L. R., C. E. Ritland, S. M. Arthur, and C. J. Krebs. 2005. Monitoring coyote population dynamics by genotyping faeces. Molecular Ecology **14:**1585–1596.

Schwartz, M. K., D. A. Tallmon, and G. Luikart. 1998. Review of DNA-based census and effective population size estimators. Animal Conservation **1:**293–299.

Seber, G. A. F. 1973. Estimating animal abundance and related parameters. Hafner, New York.

Taberlet, P., J.-J. Camarra, S. Griffin, E. Uhrès, O. Hanotte, L. P. Waits, C. Dubois-Paganon, T. Burke, and J. Bouvet. 1997. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Molecular Ecology **6:**869–876.

Taberlet, P., S. Griffin, B. Goossens, S. Questiau, V. Manceau, N. Escaravage, L. P. Waits, and J. Bouvet. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research **24:**3189–3194.

Taberlet, P., and G. Luikart. 1999. Non-invasive genetic sampling and individual identification. Biological Journal of the Linnean Society **68:**41–55.

Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: look before you leap. Trends in Ecology & Evolution **14:**323–327.

Valière, N. 2002. Amélioration et optimisation des méthodes non-invasives et des marqueurs microsatellites en biologie des populations et de la conservation. Ph.D. thesis. University of Lyon I, Lyon, France.

Valière, N., P. Berthier, D. Mouchiroud, and D. Pontier. 2002. GEMINI: software for testing the effects of genotyping errors and multitubes approach for individual identification. Molecular Ecology Notes **2:**83–86.

Waits, J., and P. Leberg. 2000. Biases associated with population estimation using molecular tagging. Animal Conservation **3:**191–199.

Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology **10:**249–256.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2001. Analysis and Management of Animal Populations. Academic Press, New York.

Wilson, G. J., A. C. Frantz, L. C. Pope, T. J. Roper, T. Burke, C. L. Cheeseman, and R. J. Delahay. 2003. Estimation of badger abundance using faecal DNA typing. Journal of Applied Ecology **40:**658–666.