Research Notes

ESTIMATING POPULATION SIZE FROM DNA-BASED CLOSED CAPTURE–RECAPTURE DATA INCORPORATING GENOTYPING ERROR

PAUL M. LUKACS,[1] Colorado Cooperative Fish and Wildlife Research Unit, Department of Fishery and Wildlife Biology, 1484 Campus Delivery, Colorado State University, Fort Collins, CO 80523, USA
KENNETH P. BURNHAM, U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, 1484 Campus Delivery, Colorado State University, Fort Collins, CO 80523, USA

JOURNAL OF WILDLIFE MANAGEMENT 69(1):396–403; 2005

Key words: capture–recapture, closed population models, DNA markers, microsatellite, non-invasive sampling, population size estimation, tag-misread.

Animal identification based on DNA samples and microsatellite genotypes is widely used for capture–recapture studies (Woods et al. 1999, Boulanger et al. 2003, Eggert et al. 2003). The method shows promise in field protocols (Woods et al. 1999) and potentially minimal error rates in the DNA analysis (Paetkau 2003). Some studies show much higher error rates in individual identification (Creel et al. 2003). There will be some level of uncertainty, although in some situations the uncertainty level is small, in the identification of individuals from microsatellite genotypes.

Closed-population capture–recapture analysis has received substantial attention over the past century. More recently, it has been extended to conditional likelihood parameterizations that allow individual covariates to better estimate capture probability (Huggins 1989, 1991) and mixture models to estimate population size in the presence of individual heterogeneity in capture probability (Norris and Pollock 1996, Pledger 2000). The major focus of research has been developing methods to handle varying capture probability. Any methods developed in the future will also have to account for varying capture probability to obtain robust estimates of population size.

While DNA-based capture–recapture studies and standard tagging studies share several common characteristics, they differ in others. In a standard tagging study, the researcher attaches a unique tag to the animal and keeps a list of tags that have been used. In a DNA-based study, the genotype of the individual acts as the tag. Therefore, all individuals are tagged prior to the beginning of the study. Unfortunately, the researcher does not know what genotypes exist in the population and must obtain samples from the animals to extract DNA. In a standard tagging study, if a tag is read that does not match one known to be in the population, the researcher knows that the tag was incorrectly read and then either rereads the tag or ignores the observation. In DNA-based studies, the researcher does not have the luxury of immediately knowing which genotypes may be incorrect. Thus, a new form of sampling uncertainty is introduced. For both standard tagging and DNA-based studies, capture probability is <1.0. This necessitates a way to infer what portion of the population is not captured in order to determine the total population size. For a DNA-based study, capture probability is a combination of the probability of encountering a sample (hair, scat, feather, etc.) and the probability that the sample yields a sufficient quantity and quality DNA to amplify and genotype.

Current closed-population capture–recapture analysis for estimating population size assumes an animal's mark is permanent and read correctly when the animal is captured (Otis et al. 1978). The use of genotype based identification can meet these assumptions in some situations, but the cost may be high. The cost comes in 2 pieces that clearly interact: (1) the monetary cost of analyzing the DNA and (2) the information loss when discarding samples that contain some degree of uncertainty in their identification. For example, the protocol described by Paetkau (2003) places a high emphasis on certainty of the genotype of the sample. In doing so, a large number of samples may have to be culled during the analysis. It may be beneficial to allow a small degree of uncertainty in the identification of a sample, perhaps 1–5%, if such a tradeoff would

[1] E-mail: plukacs@cnr.colostate.edu

allow enough additional samples to be used in the estimation of population size to make up for the addition of a parameter to the model.

When genotyping error exists, it has been shown that population size estimates derived from capture–recapture assume no error was biased (Waits and Leberg 2000, Mills et al. 2000, Creel et al. 2003). Creel et al. (2003) demonstrated that the error level they have in their wolf (*Canis lupus*) data could produce population size estimates that were biased by 5.5 times the true population size. Waits and Leberg (2000) also found large overestimation when genotyping errors were present. In addition, the authors showed underestimation in population size when multiple individuals share the same genotype. Mills et al. (2000) show in detail the underestimation effects of individuals sharing genotypes. All of these conclusions are logical when one considers that the statistical inference being made in a capture–recapture study is to the number of genotypes in the population. If errors are being made, there will be more genotypes observed than individuals in the population. If multiple individuals share the same genotype, there will be more individuals in the population than genotypes.

Currently, many studies collect far more hair, feathers, or scat samples than they can afford to have genotyped. Therefore, the limiting factor for sample size often is funding rather than a lack of sampled DNA. During the analysis, some samples are culled due to lack of confidence in genotyping results. It would be beneficial to be able to cull fewer samples and therefore increase usable sample size while taking into account the possibility of errors in genotyping.

We present a class of models for estimating the size of a demographically and geographically closed population when there exists some probability of misidentifying individuals. The misidentification occurs in such a way that it is unknown on a case-by-case basis if the sample is correctly identified or not. The method extends the full likelihood models of Otis et al. (1978) and the conditional likelihood models of Huggins (1991). In addition, mixture models similar to those of Pledger (2000) can be built that estimate both heterogeneity in capture probability and misidentification.

Methods

The notation presented here follows Otis et al. (1978) and Pledger (2000) where applicable. The model assumes there are t sampling occasions, the population of interest is well defined, and the population is demographically and geographically closed during sampling. The following notation is used to describe the models:

p_{ia} probability of initially observing a genotype at time i and from mixture component a. The second subscript was omitted for models not including mixtures.

c_{ia} probability of subsequently observing a genotype at time i and from mixture component a. The second subscript was omitted for models not including mixtures.

π_a probability of a genotype belonging to an animal in mixture a

α probability that a genotype is identified correctly given it is observed for the first time

f_0 the number of genotypes in the population that are never observed

N population size

$\mathbf{h} = \{h_1, h_2, ..., h_t\}$ encounter history vector; $h_i = 1$ if the genotype is observed, 0 otherwise

M_{t+1} number of distinct genotypes observed

$n_{\mathbf{h}}$ a count of the number of times encounter history $\mathbf{h}$ is observed

A the number of mixtures in a given model, usually 1 or 2. $A = 1$ corresponds to the case of no mixtures.

We assume a set of loci, currently microsatellites, are being used that have enough loci and enough alleles per locus to ensure with high probability that all individuals within the population are unique if correctly genotyped. We further assume that a genotyping error will lead to a genotype that is not identical to that of any member of the population. In addition, 2 errors made at different trapping occasions are assumed to never produce identical genotypes. These assumptions were asserted as reasonable by Paetkau (2003) and are further addressed in the Discussion.

Given the above assumptions, we computed the probability of each encounter history. If a genotype is first observed at time k and subsequently observed in the future, the probability of the encounter history is.

$$\Pr[\mathbf{h}] = \left[\prod_{i=1}^{k-1}(1-p_i)\right]\left[p_k\alpha\right]\left[\prod_{j=k+1}^{t} c_j^{h_j}(1-c_j)^{1-h_j}\right].$$

For a genotype that is only observed at occasion k and never seen again, the probability of the encounter history is:

$$\Pr[\mathbf{h}] = \left[\prod_{i=1}^{t-1}(1-p_i)\right]\left[p_t\alpha\left(\prod_{j=t+1}^{t}(1-c_j)\right)+p_t(1-\alpha)\right].$$

Heuristically, the probability expression states that the genotype was not observed from occasions 1 to $k-1$ with probability $(1-p_i)$. It is observed with probability p_k. Then it is either correctly genotyped with probability α and not seen again from occasions $k+1$ to k with probability $(1-c_i)$, or it was incorrectly genotyped with probability $(1-\alpha)$ and, by assumption, never seen again.

The full multinomial likelihood function can be constructed given the probabilities of each capture history. The likelihood is:

$$L(f_0, p_i, c_i, \alpha \mid n_{\mathbf{h}}, M_{t+1}) \propto \frac{(f_0 + M_{t+1})!}{f_0!} \prod_{\mathbf{h}} \Pr[\mathbf{h}]^{n_{\mathbf{h}}}.$$

We obtained parameter estimates by numerically optimizing the log–likelihood function. We used a quasi-Newton optimization function in SAS PROC IML (SAS Institute 2002). The variance–covariance matrix can be obtained by numerically estimating the information matrix, inverting and taking its negative.

Note that N is not in this likelihood. We estimated N as a derived parameter. The closed capture–recapture models of Otis et al. (1978) were written equivalently with N or f_0 in the likelihood. The f_0 parameterization was chosen in modern software to easily enforce the constraint that abundance was greater than or equal to the total number of individuals captured, such as is done in program MARK (White and Burnham 1999). This constraint was necessary if marks were assumed to be correctly read, but it does not hold if marks can be read incorrectly. It is possible to observe more genotypes than are actually in the population. Therefore, we estimate N as:

$$\hat{N} = \hat{\alpha}\left(\hat{f}_0 + M_{t+1}\right).$$

The variance of $\hat{N}$ is estimated as:

$$\begin{aligned}\widehat{\mathrm{var}}[\hat{N}] = {} & \hat{\alpha}^2\, \widehat{\mathrm{var}}[\hat{f}_o] + \left(\hat{f}_0 + M_{t+1}\right)^2 \widehat{\mathrm{var}}[\hat{\alpha}] \\ & + 2\hat{\alpha} M_{t+1}\, \widehat{\mathrm{cov}}[\hat{f}_0, \hat{\alpha}] + \widehat{\mathrm{var}}[M_{t+1}].\end{aligned}$$

Having M_{t+1}^2, a potentially large positive number, in the variance was worrisome to us, but $\widehat{\mathrm{var}}[\hat{\alpha}]$ is typically small because $\hat{\alpha}$ often is near 1 and the multinomial variance is therefore small. In addition, the $\widehat{\mathrm{cov}}[\hat{f}_0, \hat{\alpha}]$ is typically small and sometimes negative keeping $\widehat{\mathrm{var}}[\hat{N}]$ on the same order of magnitude as the closed capture–recapture models not incorporating recaptures.

The parameters p, c, and α can be modeled as functions of group covariates as is commonly done in generalized linear models (McCullagh and Nelder 1989) and in program MARK (White and Burnham 1999). The α parameter should almost always be modeled with a sine link because it will be very near the boundary of 1.0 in many studies. The sine link allows for better estimation of the number of estimable parameters and of the shape of the log–likelihood function at its maximum, while constraining the parameter to be within [0–1] than a logit function (White and Burnham 1999).

We construct a conditional multinomial likelihood function by conditioning on the probability that an animal is never captured, similar to the models of Huggins (1989, 1991). Therefore, we remove f_0 from the likelihood. The probability of an encounter history is now:

$$\Pr[\mathbf{h}_c] = \frac{\Pr[\mathbf{h}]}{1 - \left[\prod_{i=1}^{t}(1-p_i)\right]}.$$

The likelihood function is:

$$L(p_i, c_i, \alpha \mid n_{\mathbf{h}}) \propto \prod_{\mathbf{h}_c} \Pr[h_c]^{n_{\mathbf{h}}}.$$

Again N is a derived parameter. It is estimated as:

$$\hat{N} = \hat{\alpha} \sum_{i=1}^{M_{t+1}} \left[\frac{1}{1 - \prod_{j=1}^{t}(1-\hat{p}_{ij})} \right].$$

The large sample estimated variance is:

$$\widehat{\mathrm{var}}[\hat{N}] = \begin{bmatrix} \frac{\partial N}{\partial \alpha} & \frac{\partial N}{\partial p_1} & \cdots & \frac{\partial N}{\partial p_t} \end{bmatrix} \times$$

$$\begin{bmatrix} \widehat{\mathrm{var}}[\hat{\alpha}] & \widehat{\mathrm{cov}}[\hat{\alpha}, \hat{p}_1] & \cdots & \widehat{\mathrm{cov}}[\hat{\alpha}, \hat{p}_t] \\ \widehat{\mathrm{cov}}[\hat{p}_1, \hat{\alpha}] & \widehat{\mathrm{var}}[\hat{p}_1] & & \widehat{\mathrm{cov}}[\hat{p}_1, \hat{p}_t] \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathrm{cov}}[\hat{p}_t, \hat{\alpha}] & \widehat{\mathrm{cov}}[\hat{p}_t, \hat{p}_1] & \cdots & \widehat{\mathrm{var}}[\hat{p}_t] \end{bmatrix} \begin{bmatrix} \frac{\partial N}{\partial \alpha} \\ \frac{\partial N}{\partial p_1} \\ \vdots \\ \frac{\partial N}{\partial p_t} \end{bmatrix}.$$

The variance can be computed numerically or analytically.

The conditional likelihood models allow parameters to be modeled as functions of individual

covariates. Given the observations generally are collected through noninvasive methods, many standard individual covariates that can affect capture probability will not be collected, for example, length of fish captured with electrofishing equipment. An interesting covariate for α may be a measure of the quality of the DNA sample collected.

The full likelihood models can be extended to incorporate heterogeneity in capture probability with a mixture distribution (Pledger 2000). The full likelihood probability of an encounter history with mixtures for a genotype that is encountered more than 1 time is

$$\Pr[\mathbf{h}] = \sum_{a=1}^{A} \pi_a \left[\prod_{i=1}^{k-1} (1 - p_{ia}) \right] [p_{ka}\alpha] \left[\prod_{j=k+1}^{t} c_{ja}^{h_j} (1 - c_{ja})^{1-h_j} \right].$$

For a genotype that is only observed once, the probability of the encounter history is

$$\Pr[\mathbf{h}] = \sum_{a=1}^{A} \pi_a \left[\prod_{i=1}^{t-1} (1 - p_{ia}) \right] \times \left[p_{ta}\alpha \left(\prod_{j=t+1}^{t} (1 - c_{ja}) \right) + p_{ka}(1 - \alpha) \right].$$

The likelihood estimation of N and the variance of N follow the full likelihood results presented above. We used a trust region optimization in SAS PROC IML (SAS Institute 2002) to fit the model.

We compared the genotype misidentification models presented here with standard closed capture–recapture models for a simulated closed population experiencing genotyping error. We used models representing several forms of variation in detection probability including changes in behavior due to previous encounter, changes across time, and constant capture probability (Table 1). The data were simulated in a factorial assignment with 5 levels of α ranging from 0.95–0.99 and 5 levels of constant capture probability ranging from 0.1–0.5. Five sampling occasions were used for each population. Capture probabilities ranging from 0.1–0.5 for each of 5 sampling occasions cover a large part of the range of possible levels of encounters from very few animals encountered to nearly all animals encountered. Each design point was replicated 200 times. SAS code used for simulation and estimation is available from the authors.

We did not include a constant capture probability form of the genotype misread model in the analysis because it was not reasonable to expect the initial capture probability to equal the recapture probability when initial captures included both correct and incorrect genotypes while recaptures were only correct genotypes. It is important to note that p is the probability of observing a genotype correctly or incorrectly, whereas c is the probability of correctly observing a genotype that has been seen at a previous observation. The time varying capture probability model has the same logical flaw as the constant capture probability model for the genotype misread models, but it was included. One could argue that the time varying capture probability model allows enough flexibility to be reasonable if true p is nearly equal to c.

Results

In most situations, the full likelihood misidentification models had lower bias in estimating population size than did the standard capture–recapture models (Fig. 1). The percent bias in estimated population size often was half as large for the misidentification models as it was the standard capture–recapture models. Bias was worst for the genotyping error models when capture probability was 0.1. This results from a lack of recaptures due to the low capture probability. As capture probability increased, bias quickly decreased.

Confidence interval coverage of estimated population size was near the nominal 95% level for the misidentification models. Confidence interval coverage was well below the nominal 95% when misidentification was present but not estimated (Fig. 2). Confidence interval coverage broke down for both the standard capture–recapture models and the misidentification capture–recapture models when capture probability was very high and misidentification was present.

The estimation of genotyping error performed poorly when capture probability was 0.1. This result would also hold for capture probabilities less than 0.1 given 5 sampling occasions were used. When

Table 1. Models representing different forms of variation in capture probability and whether they are used in capture–recapture models incorporating genotyping error. Sources of variation include changes in capture probability due to a behavioral response to first encounter, changes across time, and constant capture probability.

Model	Variation in capture probability	Including variation attributable to genotype: Yes	Including variation attributable to genotype: No
M_0	Constant		X
M_b	behavior	X	X
M_t	time	X	X
M_{t+b}	time + behavior	X	

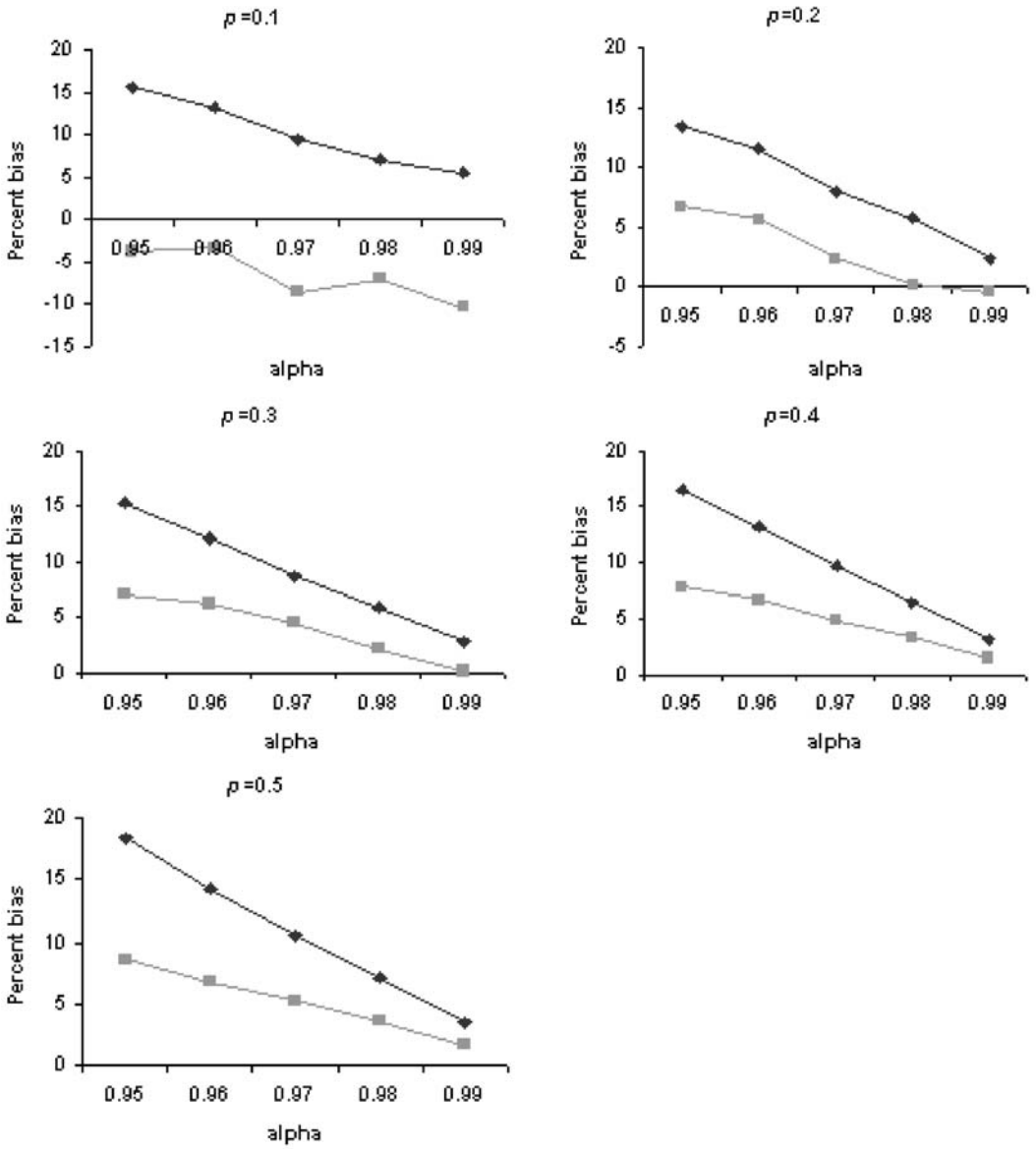

Fig. 1. Percent bias in estimated abundance for closed capture–recapture models including an estimate of genotyping error (gray squares) and not accounting for genotyping error (black diamonds) across 5 levels of capture probability (p) and 5 levels of the probability of correctly genotyping a sample (alpha). Percent bias is averaged across models containing different forms of variation in capture probability. Each design point was replicated 200 times. The coefficient of variation for all bias estimates is <1%.

capture probability was 0.1 an animal only had a 0.08 probability of being caught more than once. Recaptures are required to estimate the probability of correctly identifying a genotype. The low number of recaptures caused the genotyping error model containing both time variation and a behavioral response in capture probability to occasionally fail to converge to a reasonable estimate of population size when capture probability was 0.1.

When capture probability was 0.5 and sampling occurred on 5 occasions, approximately 97% of animals were detected. Thus, nearly all individuals were expected to be captured at least once and 81% were expected to be caught more than once. Therefore, there were few encounter histories with only a single observation, so it was difficult to estimate genotyping error rate effectively. This is a minor issue because such a situation is rarely

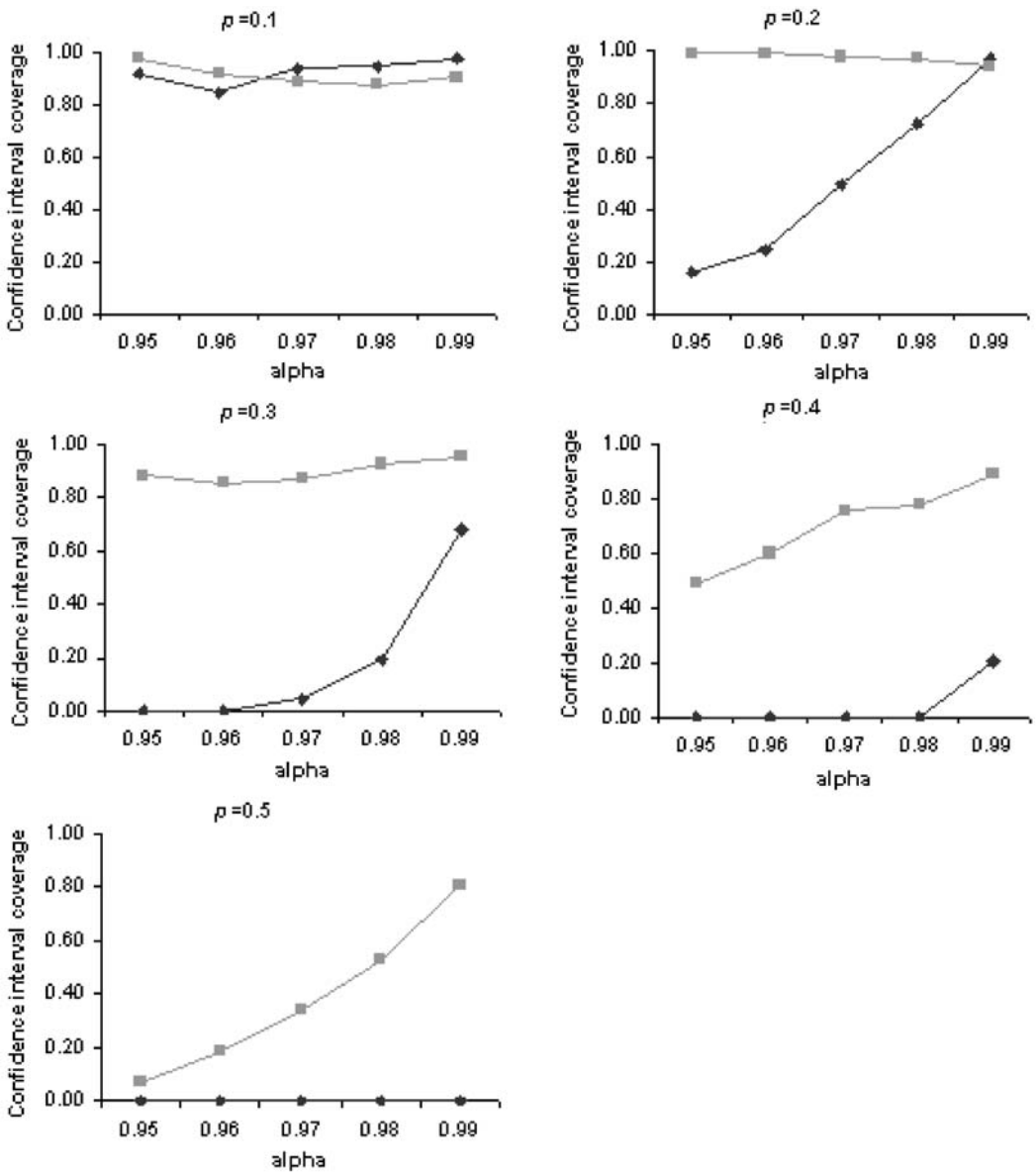

Fig. 2. Confidence interval coverage of estimated abundance for closed capture–recapture models including an estimate of genotyping error (gray squares) and not accounting for genotyping error (black diamonds) across 5 levels of capture probability (p) and 5 levels of the probability of correctly genotyping a sample (alpha). Confidence interval coverage is averaged across models containing different forms of variation in capture probability. Each design point was replicated 200 times.

feasible in the field, and when it occurs the confidence interval width is very small, and bias is trivial (1–4%). The estimate typically is only off of the true value by a few animals.

Discussion

We make 3 assumptions beyond those needed for a standard capture–recapture study to estimate genotyping error rate from capture–recapture data. The assumptions are relatively easily met in real world problems. We assume that a set of loci are used that contain enough genetic information such that each individual is uniquely identified if the genotype is correctly read. For a wide range of species these systems exist, such as bears (*Ursus* spp.; Paetkau 2003), elephants (*Lox-*

odonta cyclotis; Eggert et al. 2003), Canada lynx (*Lynx canadensis*; Schwartz et al. 2003), and sage-grouse (*Centrocercus urophasianus*; Taylor et al. 2003), and more are being developed continuously. Resolution power can be assessed prior to beginning a study by computing the probability of identity for the marker set (Waits et al. 2001). Therefore, it is not difficult to obtain the power necessary to discriminate among individuals when no errors are present.

We assume that any error in genotyping will result in a genotype that does not match that of another individual in the population of interest. Far more genotypes are possible than individuals that exist in many wildlife species. For example, a set of 6 loci each with 3 alleles has 46,656 possible genotypes. Therefore, the chance of an error resulting in an existing genotype rather than some other genotype is quite low. If this assumption is violated and an error results in an existing genotype, there is only a trivial effect on the estimation of population size from a closed capture–recapture model. If the error results in an animal that has never been seen before, then that animal is no longer at risk of capture and the animal that was truly caught remains at risk of capture. Thus, M_{t+1} and $\hat{p}$ are virtually unchanged, hence $\hat{N}$ is also nearly unchanged. For example, consider the case of the simplest form of a capture–recapture model with time varying capture probability, the Lincoln–Petersen estimator. Three quantities are needed to estimate abundance from this model: (1) the number caught in the first sample (n_1), (2) the number caught in the second sample (n_2), and (3) the number of marked animals caught in the second sample (m_2). If 50 animals are caught at each sample and 25 of the animals caught during the second sample are recaptures, then the estimated abundance is $50 \times 50/25 = 100$. Now if 1 of the animals caught on the first occasion is misidentified as an animal that is not caught on the first occasion, but the animal is caught on the second occasion, then none of the statistics change and the population estimate remains 100. Thus, a violation of this assumption is trivial.

We assume that errors were never repeated in exactly the same way to produce an identical, incorrect genotype. This assumption can be violated in 2 ways. First, an individual can be sampled twice and incorrectly genotyped in the same way twice. Paetkau (2003) showed that occurred in about 15% of his samples that were incorrectly genotyped. For capture–recapture this is a minor issue because the individual is still correctly matched across samples even though the genotype is not correct. Second, 2 different individuals could both be incorrectly genotyped and coincidentally produce a matching incorrect genotype. Many factors would all have to happen, each with low probability to generate the same genotype incorrectly twice from independent samples (Paetkau 2003). Therefore, this assumption may be the weakest of the 3 assumptions made here, but the consequences of a violation of the assumption is minor.

Estimation of the model parameters, and most importantly $\hat{N}$, is good when α is high ($\geq$0.95). This is reasonable performance for the estimator. Lab protocols can easily keep error rates within that range (Paetkau 2003). Yet, even with an error rate of only 1% per sample, substantial bias in $\hat{N}$ can occur if it is not taken into account. The misidentification models perform well when error rates exceed 5%, but the variance on estimated population size becomes large and quickly makes the results relatively uninformative about population size. Despite the large variance on population size, one would learn about the high error rate. The cutoff for an allowable level of genotyping error depends on the precision required for the study.

The models presented here rely on recaptures to estimate the probability of correctly genotyping a sample. Heuristically, this quantity is estimated by the imbalance in the number of genotypes observed only once to those observed more than once. If error is present in genotyping, there will be an excess of genotypes observed only once. A portion of these (α) are truly seen only once, and the remainder ($1-\alpha$) are seen once because the genotypes are mistakes. There are 2 ways to ensure obtaining recaptures: 1 is to sample intensively to get capture probability high, and the other way is to sample on more occasions.

The capture–recapture models presented here are based on the same likelihood function as the standard capture–recapture models they generalize. Therefore, model selection criteria such as AIC_c can be used to compare models (Burnham and Anderson 2002). The researcher need not worry whether they should or should not use a model incorporating genotyping error. Estimation can be performed with both standard capture–recapture models and the genotyping error models presented here, and AIC_c will determine which model is better supported by the data. Resulting abundance estimates may be model averaged to further reduce the effects of model selection bias (Burnham and Anderson 2002).

This class of models can help reduce the cost or increase the sample size of a DNA-based capture–recapture study. Cost is reduced by having to extract fewer samples to achieve enough samples that meet a desired quality. Sample size is increased for a set cost because fewer samples will be culled.

Although these models were developed in the context of identification from microsatellite data, the concept extends to other types of analyses as well. Identification based on photographs is commonly attempted in marine mammal studies (Jefferson 2000). These models also could be used for that application. The models presented here do not fit well with misidentification of physical tags because the set of tags available for capture should be known. Therefore, if a tag is read that does not match a tag in the population, it is known to be incorrect.

The models presented here are most applicable to smaller populations up to several thousand individuals. In theory, the method applies to any size population. Populations that are very large are quite expensive and time consuming to sample with DNA-based methods due to the large number of samples that would need to be processed and large number of loci needed to resolve individuals.

Management Implications.—Ignoring genotyping error when it is in fact present will lead to overestimation of animal abundance. DNA-based capture–recapture typically is used on species with small population sizes that are difficult to observe. Overestimating the size of a small population could lead to potentially detrimental conclusions for an endangered or exploited species. The method we present allows genotyping error rate to be directly estimated and abundance appropriately corrected. Therefore, management decisions can be made based on an accurate assessment of the population size.

Acknowledgments.—The Colorado Division of Wildlife, Great Outdoors Colorado, and Dr. T. Shenk provided support for this research. Dr. G. White, Mr. E. Bergman, Dr. R. Steidl, and an anonymous referee provided comments on the manuscript.

Literature Cited

BOULANGER, J., G. C. WHITE, B. N. MCLELLEN, J. WOODS, M. PROCTOR, AND S. HIMMER. 2003. A meta-analysis of grizzly bear DNA mark–recapture projects in British Columbia, Canada. Ursus 13:137–152.

BURNHAM, K. P., AND D. R. ANDERSON. 2002. Model selection and multimodel inference: a practical information–theoretic approach. Springer-Verlag, New York, USA.

CREEL, S., G. SPONG, J. L. SANDS, J. ROTELLA, J ZEIGLE, L. JOE, K. M. MURPHY, AND D. SMITH. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Molecular Ecology 12:2003–2009.

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.

HUGGINS, R. M. 1989. On the statistical analysis of capture experiments. Biometrika 76:133–140.

———. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47:725–732.

JEFFERSON, T. A. 2000. Population biology of the Indo-Pacific hump-backed dolphin in Hong Kong waters. Wildlife Monographs 144.

MCCULLAGH, P., AND J. A. NELDER. 1989. Generalized linear models. Second edition. Chapman and Hall, New York, USA.

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.

NORRIS, J. L., AND K. H. POLLOCK. 1996. Nonparametric MLE under two closed capture–recapture models with heterogeneity. Biometrics 52:639–649.

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.

PLEDGER, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56:434–442.

SAS INSTITUTE. 2002. SAS version 9. SAS Institute, Cary, North Carolina, USA.

SCHWARTZ, M. K., L. S. MILLS, Y. ORTEGA, L. F. RUGGIERO, AND F. W. ALLENDORF. 2003. Landscape location affects genetic variation of Canada lynx (*Lynx canadensis*). Molecular Ecology 12:1807–1816.

TAYLOR, S. E., S. J. OYLER-MCCANCE, T. W. QUINN. 2003. Isolation and characterization of microsatellite loci in greater sage-grouse (*Centrocercus urophasianus*). Molecular Ecology Notes 3:262–264.

WAITS, J. L., AND P. L. 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.

WHITE, G. C., AND K. P. BURNHAM. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46:S120–S139.

WOODS, J. G., D. PAETKAU, D. LEWIS, B. N. MCLELLAN, M. PROCTOR, AND C. STROBECK. 1999. Genetic tagging of free-ranging black and brown bears. Wildlife Society Bulletin 27:616–627.

Associate Editor: Steidl.