

# Importance of Accounting for Detection Heterogeneity When Estimating Abundance: the Case of French Wolves

SARAH CUBAYNES,\*†§§ ROGER PRADEL,\* RÉMI CHOQUET,\* CHRISTOPHE DUCHAMP,††  
 JEAN-MICHEL GAILLARD,‡ JEAN-DOMINIQUE LEBRETON,\* ERIC MARBOUTIN,§  
 CHRISTIAN MIQUEL¶ ANNE-MARIE REBOULET,\* CAROLE POILLOT,¶ PIERRE TABERLET,¶  
 AND OLIVIER GIMENEZ\*

\*Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Campus CNRS, 1919 Route de Mende, F-34293 Montpellier Cedex 5, France

†Institut de Mathématiques et Modélisation de Montpellier, UMR 5149, place E. Bataillon, F-34095 Montpellier Cedex 5, France

††Office National de la Chasse et de la Faune Sauvage, CNERA Prédateurs et Animaux déprédateurs, Parc Micropolis, F-05000 Gap, France

‡Université de Lyon, F-69000, Lyon, Université Lyon 1, Centre National de la Recherche Scientifique, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

§Office National de la Chasse et de la Faune Sauvage, ZI Mayencin, 5 Allée de Béthléem, F-38610 Gières, France

¶Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université J. Fourier, BP 53, F-38041 Grenoble Cedex 9, France

**Abstract:** *Assessing conservation strategies requires reliable estimates of abundance. Because detecting all individuals is most often impossible in free-ranging populations, estimation procedures have to account for a <1 detection probability. Capture-recapture methods allow biologists to cope with this issue of detectability. Nevertheless, capture-recapture models for open populations are built on the assumption that all individuals share the same detection probability, although detection heterogeneity among individuals has led to underestimating abundance of closed populations. We developed multievent capture-recapture models for an open population and proposed an associated estimator of population size that both account for individual detection heterogeneity (IDH). We considered a two-class mixture model with weakly and highly detectable individuals to account for IDH. In a noninvasive capture-recapture study of wolves we based on genotypes identified in feces and hairs, we found a large underestimation of population size (27% on average) occurred when IDH was ignored.*

**Keywords:** *Canis lupus*, E-SURGE, individual heterogeneity, mark-recapture, mixture models, multievent model, population size

Importancia de Considerar la Heterogeneidad de Detección al Estimar la Abundancia: el Caso de Lobos Franceses

**Resumen:** *La evaluación de estrategias de conservación requiere de estimaciones confiables de la abundancia. Debido a que a menudo es imposible la detección de todos los individuos en poblaciones de libre movimiento, los procedimientos de estimación necesitan considerar una probabilidad de detección <1. Los métodos de captura-recaptura permiten que los biólogos hagan frente al tema de la detectabilidad. Sin embargo, los modelos de captura-recaptura para poblaciones abiertas se elaboran sobre la suposición de que todos los individuos comparten la misma probabilidad de detección, aunque la heterogeneidad de detección entre individuos ha llevado a la subestimación de la abundancia de poblaciones cerradas. Desarrollamos modelos multievento de captura-recaptura para una población abierta y proponemos un estimador del tamaño poblacional asociada que considera la heterogeneidad de detección individual (HDI). Consideramos*

§§email sarab.cubaynes@cefe.cnrs.fr

Paper submitted February 9, 2009; revised manuscript accepted September 8, 2009.

*un modelo mezclado de dos clases con individuos difícil y fácilmente detectables para considerar HDI. En un estudio de captura-recaptura no invasivo de lobos nos basamos en los genotipos identificados en heces y pelos, encontramos una fuerte subestimación del tamaño poblacional (27% en promedio) cuando se ignoró la HDI.*

**Palabras Clave:** *canis lupus*, E-SURGE, heterogeneidad individual, marca-recaptura, modelos mezclados, modelo multievento, tamaño poblacional

## Introduction

Assessing the success of conservation and management actions performed on focal populations relies on accurate estimates of population size. Obtaining reliable abundance estimates, however, is difficult for most populations because exhaustive counting of free-ranging animals is usually impossible. Capture–recapture (CR) protocols have been developed to solve that problem and are increasingly being used to monitor populations of vertebrates (Williams et al. 2002).

In standard CR models, all individuals are assumed to have identical detection probabilities. Nevertheless, individual attributes (e.g., age, body mass, social status) and habitat features (home-range location and composition) generate heterogeneity in the detection process (individual detection heterogeneity [IDH]) that may lead to biased estimates of abundance in closed populations (Carothers 1973; Otis et al. 1978; Hwang & Huggins 2005) when it is ignored. In closed populations, IDH can be handled by considering classes of individuals with distinct detection probabilities in so-called mixture models (Agresti 1994; Norris & Pollock 1996; Pledger 2000). Although some factors can be incorporated in CR models as individual covariates (Huggins 1989, 1991; McDonald & Amstrup 2001), most of them are generally not measured. Moreover, in the case of endangered and small-sized populations for which individuals are not physically captured but instead are monitored through DNA collected in the field (e.g., Waits & Paetkau 2005), no covariate is available.

We sought to demonstrate the risks of flawed inference in estimation of the size of open populations when IDH is ignored. We combined estimators of detection probabilities in heterogeneous open populations (Pledger et al. 2003) with an estimator of population size when individuals are sampled with unequal probabilities. We used multievent CR models (Pradel 2005, 2009) that allow inference about hidden states through observations. We included two classes of individuals (i.e., highly vs weakly detectable) to model IDH and to estimate abundance.

To illustrate our approach, we used CR data on wolves (*Canis lupus*). After their extirpation from France 2 centuries ago, wolves entered the French Alps in 1992 as a consequence of their natural recovery in adjacent Italy (Valière et al. 2003). Since then, the population has been increasing, and individuals have spread across the Alps. Because the agricultural landscape is extensive, a large

number of depredations on livestock occurred, which created political and economic problems (Mech 1995). Reliable estimates of annual abundance are thus needed as the first step of a wolf action plan for population management and for evaluation of the species conservation status in France. The large territories, low population densities and high mobility of wolves prevented us from using traditional survey methods (i.e., observation or capture). We used noninvasive monitoring via genotyping of biological material to identify individual wolves (Valière 2002). We applied models incorporating IDH to this individual monitoring and evaluated the resulting bias in abundance when IDH was ignored. Finally, we considered the benefits of accounting for IDH when estimating population size and how it could affect the reliability of conservation actions.

## Methods

### Data Collection

From 1995 to 2003 we collected feces samples from wolves in the French Alps. A network of about 1000 trained people conducted opportunistic-sign surveys over a large area (72,000 km<sup>2</sup>). Additionally, we used standardized snow-tracking methods to find and sample all pack territories. We developed microsatellite-based identification of wolves to identify individuals (Valière 2002). For each wolf sample, we amplified seven microsatellite loci with a polymerase chain reaction (PCR) multitube approach with eight replicates per sample (Taberlet & Luikart 1999). To assess genotype reliability and minimize errors, we used a quality index ([QI] corresponding to the mean frequency of the consensus genotype among the eight replicates at each locus (Miquel et al. 2006)). We discarded genotypes with average QI < 0.4. We constructed a “genetic capture history” for each detected genotype by grouping the observations within 3-month period from 1 January. Overall, we genotyped 1181 wolf samples among which 840 were classified as reliable. The number of detections varied from one to 91 depending on the genotype. The data set included the capture history of 160 different genotypes.

### Multievent CR Modeling

To account for IDH, we used mixture models (Pledger 2000; Pledger et al. 2003; Pradel 2009) that incorporated

hidden groups of individuals with contrasting detection probability. We did not have a priori information on the detection rate for any genotyped wolf. Thus, our model looked like a usual CR model in which the state of each individual was imperfectly known. Therefore, it described a particular case of multievent CR models that specifically dealt with uncertainty in state assignment (Pradel 2005).

A multievent CR model includes both states and observations generated from the underlying state of an individual. We distinguished living wolves with high detection probability ( $H$ ), living wolves with low detection probability ( $L$ ), and dead wolves ( $D$ ). The observations were coded 0 (previously genotyped wolf was not detected) and 1 (previously genotyped wolf was detected). For example, a capture history of 101 denotes an individual who was encountered on the first and third sampling occasions, but missed on the second. The probability (Pr) of this history is

$$\begin{aligned} \text{Pr}(101) = & \pi \cdot \phi_1^L \cdot (1 - p_2^L) \cdot \phi_2^L \cdot p_3^L \\ & + (1 - \pi) \cdot \phi_1^H \cdot (1 - p_2^H) \cdot \phi_2^H \cdot p_3^H, \quad (1) \end{aligned}$$

where  $\pi$  is the probability the individual belongs to state  $L$ ,  $p_i^L$  and  $p_i^H$  are the recapture probabilities in states  $L$  and  $H$  at occasion  $t$ ,  $\phi_i^L$  and  $\phi_i^H$  are the survival probabilities in states  $L$  and  $H$  between occasions  $t$  and  $t + 1$ . Assuming all individuals have equal survival and detection probabilities (homogeneity model), the results are the standard Cormack-Jolly-Seber model (Lebreton et al. 1992). We also tested for differences in survival of wolves between states  $L$  and  $H$  (Supporting Information) because detection rate might be related to biological factors such as rank within the pack or age classes.

We then defined a set of candidate models incorporating biologically relevant combinations of several temporal and individual effects on both survival and detection probabilities. Regarding temporal effects on survival, we considered year and seasonal effects to test for the influence of possible random fluctuations in unpredictable external factors such as poaching or diseases in juveniles. To investigate temporal variations in sampling effort we considered year and seasonal effects on detection. Moreover, we divided the year in two ways. First, we expected detection to remain lower for some time after reproduction (July–December) because the young, and to some extent the pack itself, are relatively sedentary (hereafter biological effect). Second, we expected a higher detection probability when it was cold (October–March) because cold and snow are favorable to DNA sample detection and preservation (hereafter climate effect) (Lucchini et al. 2002).

To choose among models and assess the effect of detection heterogeneity on abundance estimation, we selected the model with lowest Akaike information criterion corrected for small sample size (AIC<sub>c</sub>) (Burnham & Anderson 2002). We used program E-SURGE (Choquet

et al. 2009a) to perform CR analyses (Supporting Information). Although goodness-of-fit tests are well developed for standard open CR models (Lebreton et al. 1992), there is no test available for multievent CR models (Pradel 2009). Nevertheless, an ad hoc procedure had been proposed to test the goodness of fit of a model with heterogeneity in detection probabilities (Péron et al. 2009). Using program U-CARE (Choquet et al. 2009b), we rejected the CJS model ( $\chi_{115}^2 = 180.73$ ,  $p < 0.01$ ), but the corresponding model accounting for detection heterogeneity fitted the data appropriately ( $\chi_{113}^2 = 118.996$ ,  $p = 0.331$ ). Consequently, we considered all candidate models accounted for IDH.

### Population Size Estimation

In a model with homogeneous detection probabilities, the number of individuals  $N_t$  alive at time  $t$  is estimated from the ratio of the number captured  $n_t$  over an estimate  $\hat{p}_t$  of the detection probability at time  $t$ ,  $\hat{N}_t = \frac{n_t}{\hat{p}_t}$ . The sample  $n$  consists of  $u$  newly detected individuals and  $m$  previously detected individuals. In the same way,  $N_t$  consists of the sum of  $U_t$  new individuals and  $M_t$  individuals already present and still alive at time  $t$ . In the heterogeneity model, the newly detected are made up of  $\pi \cdot U_t$  individuals in state  $L$  and  $(1 - \pi) \cdot U_t$  individuals in state  $H$ . Using a Horvitz-Thompson type estimator that specifically accounts for unequal detection probabilities, we obtained an estimate  $\hat{U}_t$  of the expected number of new individuals in the population

$$\hat{U}_t = \frac{\hat{\pi} \cdot u_t}{\hat{p}_t^L} + \frac{(1 - \hat{\pi}) \cdot u_t}{\hat{p}_t^H}.$$

Applying the survival estimates to already detected individuals, we obtained an estimate  $\hat{M}_t$  of the expected number of already detected individuals still alive at time  $t$

$$\hat{M}_t = \sum_{j=1}^{t-1} u_j \cdot \left( \hat{\pi} \cdot \prod_{i=j}^{t-1} \hat{\phi}_i^L + (1 - \hat{\pi}) \cdot \prod_{i=j}^{t-1} \hat{\phi}_i^H \right).$$

Finally, we obtained an estimator of population size accounting for IDH (Supporting Information):

$$\hat{N}_{het}_t = \hat{U}_t + \hat{M}_t.$$

For example, consider a population with  $\hat{\pi} = 30\%$  of individuals in state  $L$ ,  $\hat{p}_t^L = 0.2$ ,  $\hat{\phi}_t^L = 0.6$ , and  $1 - \hat{\pi} = 70\%$  of individuals in state  $H$ , with  $\hat{p}_t^H = 0.5$  and  $\hat{\phi}_t^H = 0.9$ . If  $u_1 = 1$  individual newly detected at the first occasion,  $u_2 = 2$  individuals newly detected at the second occasion,  $\hat{U}_2 = \frac{0.3 \times 2}{0.2} + \frac{0.7 \times 2}{0.5} = 5.8$ , and  $\hat{M}_2 = 1 \times (0.3 \times 0.6 + 0.7 \times 0.9) = 0.81$ , an estimate of the size of the population at the second occasion is obtained:  $\hat{N}_{het}_2 = 5.8 + 0.81 = 6.61$ .

Given the discrete and positive nature of the data and the relatively small sample size, we used a nonparametric

**Table 1.** Parameter estimates of wolf survival and detection probabilities of the best-fitting model assuming a two-class mixture of individuals on survival and detection probabilities, with an additive biology effect on the later (two distinct parameters for July–December vs. January–June).\*

Parameter	Class of individuals	
	weakly detectable (95% CI)	highly detectable (95% CI)
Proportion of newly marked individuals	0.76 (0.65–0.85)	0.24 (0.15–0.34)
Survival probability	0.75 (0.54–0.94)	0.90 (0.71–0.98)
Detection probability January–June	0.22 (0.11–0.37)	0.86 (0.74–0.96)
Detection probability July–December	0.10 (0.05–0.18)	0.64 (0.51–0.82)

\*The detection probability over a 6-month period was calculated as the probability of being detected at least once during the period (e.g., detection probability January–June was equal to the detection probability January–March plus the detection probability April–June minus the detection probability January–March times the detection probability April–June).

bootstrap (Davison & Hinkley 1997) to obtain confidence intervals of abundance and functions of demographic parameters.

## Results

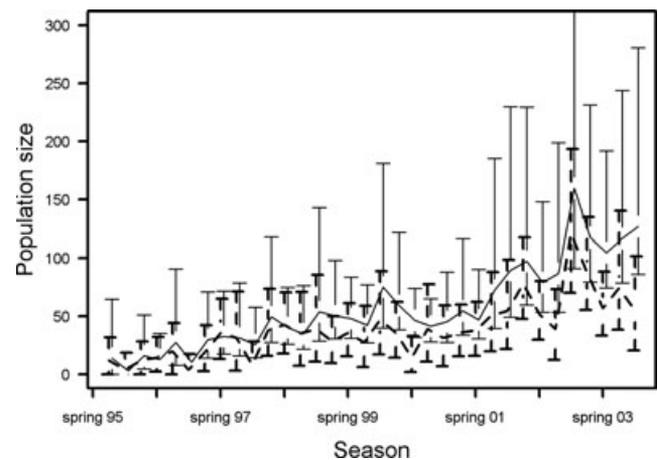
The model with individual heterogeneity in both detection and survival probabilities was best supported by the data (Supporting Information). Temporal variation in survival received little support from the data and did not affect detection parameter estimates (results not shown). We therefore retained models with constant survival only. Detection probabilities exhibited a biological effect that acted on an additive scale with heterogeneity. In support of the goodness-of-fit tests, models with a two-class mixture on detection always had lower AIC<sub>c</sub> values than their homogenous counterpart (Supporting Information). Although there was some uncertainty about whether heterogeneity should also be incorporated with survival, estimates of detection parameters given by the four top models were very similar.

Detection probability was higher from January to June than from July to December in both *L* and *H* states (Table 1). Detection probability strongly differed in the two classes of detectability. Depending on which of the two 6-month periods we considered, the detection probability of wolves in state *H* was four to six times greater than that of wolves in state *L*. Annual survival (product of all 3-month survival probabilities) of wolves in state *H* was 90% (95% CI: 0.71–0.98) versus 75% (95% CI: 0.54–0.94) for wolves in state *L*. Overall, survival and detection probabilities were positively correlated, and weakly detectable individuals had a lower survival probability than highly detectable individuals.

The total population size of wolves in the French Alps increased from three (95% CI: 0–7) in winter 1995 to 126 (95% CI: 85–280) in winter 2003 (Fig. 1). Marked seasonal variations were observed. When IDH was ignored (Fig. 1), abundance was underestimated by 27% on average (ranging from 0% to 70%).

## Discussion

Wolf populations are increasing in Europe (Salvatori & Linnell 2005), which often translates into increasing conflicts due to depredation on livestock. Although large carnivores are protected by law (Bern Convention, European Fauna-Flora-Habitat Directive), the long-term persistence and reestablishment of populations is based on their acceptance by stakeholders (Bath 2000). Toward this aim, if the conservation status of the species is favorable, governments may implement management strategies involving some form of lethal control (Treves & Karanth 2003). Assessing a sustainable level for lethal control requires a reliable estimate of abundance, which may be hampered if IDH is ignored.



**Figure 1.** Population size estimates of wolves in the French Alps from 1995 to 2003. Results were obtained from the best-fitting model (solid line) that assumed two classes of individuals regarding survival and detection probabilities, with an additive biological effect on the later (two distinct parameters for July–December vs. January–June), and from a model with a structure similar to this best-fitting model but that ignored individual heterogeneity in the detection (dashed line). Bars are 95% confidence intervals.

Marucco et al. (2009) used noninvasive CR data and an intensive sampling protocol to estimate wolf population size in a 4000-km<sup>2</sup> area at the core of wolf presence in the Italian Alps. In their smaller data set, no heterogeneity was detected, and they used a standard model that assumed homogeneity. Our sampling effort, within the constraints of a national management plan, was spread over 72,000 km<sup>2</sup> and goodness-of-fit tests and model-selection procedures indicated IDH occurred among wolves. Detection heterogeneity could result from genotyping errors (Lukacs & Burnham 2005), heterogeneities in the sampling effort (Devineau et al. 2006), or be a direct consequence of the species' biology (Crespin et al. 2008). We used multitube PCR and discarded unreliable genotypes to minimize errors. Among reliable samples, genotypes detected only once did not have a lower QI than genotypes detected more than once ( $\chi_1^2 = 1.854$ ,  $p = 0.17$ ), which suggests no evidence of false genotypes.

To disentangle the effects of sampling heterogeneity from the biology of the species, we compared two different temporal patterns in the detection process. The biological effect was more plausible than the climate effect, and detectability was higher before than after the breeding period. The lower detection probability after the breeding period might reflect the impossibility of detecting cubs before they start to follow adults at 6 months old. Under the biological scenario, highly detected individuals would correspond to dominant individuals that are more mobile within the pack territory and more likely to mark territory with feces and urine (Vila et al. 1994), whereas weakly detectable individuals would be young and subordinate individuals that are difficult to detect because they are not mobile or have moved outside the pack territory (Mech & Boitani 2003). This theory is reinforced by the estimated survival of highly detectable individuals, which matched the survival of dominant individuals reported in other studies (Mech & Boitani 2003). Weakly detectable individuals exhibited lower survival because this group may be a mixture of young, subordinates, and migrants that have lower survival than dominant individuals. Nevertheless, the model specifying a transition of individuals from low to highly detectable—allowing subordinate individuals to access the dominant status—was not supported by the data ( $\Delta AIC_c = -56.29$  when compared with a model without transition). The limited length of the study period combined with a small data set may be insufficient to detect transitions.

Link (2003) established that the same capture-history data set may arise from a wide range of heterogeneity models, which in turn yield entirely different estimates of population size. On the basis of biological considerations and following Pledger (2000)'s suggestion, we used a simple dichotomy of individuals (i.e., low versus high detection). Moreover, the limited size of our data set prevented us from using a larger number of classes.

Problems of unidentifiability of population size may also arise when some individuals have encounter probabilities close to zero (Link 2003). In wolves the chance of detection of subordinates is not very low and probably similar among them because packs occupy a territory and, especially in winter, it is possible to follow the tracks and find the droppings of all individuals along the way. The really problematic individuals are the dispersing individuals that do not belong to a pack. They are probably secretive, do not stay very long in the same area, and may escape detection entirely. Our estimate of population size is, therefore, likely to be biased low. Nevertheless, our estimate is still reliable because the dispersers represent a minority of individuals during a restricted period because the dispersers will join a pack and become detectable, will leave the study area permanently, or will die.

Overall, our results show that ignoring IDH can lead to a severe underestimation of population size, and the procedure we developed accounted for IDH in free-ranging populations.

## Acknowledgments

The authors are indebted to all the fieldworkers who gathered the biological samples. This work was partly supported by a grant from the French Research National Agency (ANR), reference ANR-08-JCJC-0028-01.

## Supporting Information

Mathematical formulation and instructions for implementing in program E-SURGE a model incorporating a two-class mixture in survival and detection (Appendix S1); general formulation of the population size estimator accounting for IDH (Appendix S2), results of model selection for modeling survival and detection of wolves from the French Alps (1995–2003) (Appendix S3) are available as part of the on-line article. The authors are responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

## Literature Cited

- Agresti, A. 1994. Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics* 50:494–500.
- Bath, A. 2000. Human dimensions in wolf management in Savoie and Des Alpes Maritimes, France. Report for France LIFE-Nature Project and the Large Carnivore Initiative Project for Europe. Memorial University of Newfoundland, St. John's, Canada.
- Burnham, K. P., and D. R. Anderson 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition. Springer-Verlag, New-York.
- Carothers, A. D. 1973. The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29:79–100.

- Choquet, R., L. Rouan, and R. Pradel. 2009a. Program E-SURGE: a software application for fitting multievent models. Pages 845–865 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. Modeling demographic processes in marked populations. Environmental and ecological statistics series. Volume 3. Springer-Verlag, New York.
- Choquet, R., Lebreton, J. D., Gimenez, O., Reboulet, A.M. and Pradel R. 2009b. U-CARE: utilities for performing goodness of fit tests and manipulating CAPTURE-REcapture data. *Ecography*: **32**:1071–1074.
- Crespin, L., R. Choquet, M. Lima, J. Merritt, and R. Pradel. 2008. Is heterogeneity of catchability in capture–recapture studies a mere sampling artifact or a biologically relevant feature of the population? *Population Ecology* **50**:247–256.
- Davison, A. C., and D. V. Hinkley. 1997. Bootstrap methods and their application. Cambridge University Press, Cambridge, United Kingdom.
- Devineau, O., R. Choquet, and J.-D. Lebreton. 2006. Planning capture–recapture studies: straightforward precision, bias, and power calculations. *Wildlife Society Bulletin* **34**:1028–1035.
- Huggins, R. M. 1989. On the statistical analysis of capture–recapture experiments. *Biometrika* **76**:133–140.
- Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* **47**:725–732.
- Hwang, W. H., and R. Huggins. 2005. An examination of the effect of heterogeneity on the estimation of population size using capture–recapture data. *Biometrika* **92**:229–233.
- 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 Monographs* **62**:67–118.
- Link, W. 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.
- Lukacs, P. M., and K. P. Burnham. 2005. Estimating population size from DNA-based closed capture–recapture data incorporating genotyping error. *Journal of Wildlife Management* **69**:396–403.
- Marucco, F., D. H. Pletscher, L. Boitani, M. K. Schwartz, K. L. Pilgrim, and J.-D. Lebreton. 2009. Wolf survival and population trend using noninvasive capture–recapture techniques in the Western Alps. *Journal of Applied Ecology* DOI: 10.1111/j.1365-2664.2009.01696.x.
- McDonald, T. L., and S. C. Amstrup. 2001. Estimation of population size using open capture–recapture models. *Journal of Agricultural Biological and Environmental Statistics* **6**:206–220.
- Mech, L. D. 1995. The challenge and opportunity of recovering wolf populations. *Conservation Biology* **9**:270–278.
- Mech, L. D., and L. Boitani. 2003. Wolves: behavior, ecology, and conservation. University of Chicago Press, Chicago, Illinois.
- Miquel, C., E. Bellemain, C. Poillot, J. Bessiere, A. Durand, and P. Taberlet. 2006. Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes* **6**:985–988.
- Norris, III, J. L., and Pollock, K. H. 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**:1–135.
- Péron, G., P.-A. Crochet, R. Choquet, R. Pradel, J.-D. Lebreton, and O. Gimenez. 2009. Capture–recapture models with heterogeneity to study survival senescence in the wild. *Oikos* DOI: 10.1111/j.1600-0706.2009.017882.x.
- Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* **56**:434–442.
- Pledger, S., K. H. Pollock, and J. L. Norris. 2003. Open capture–recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics* **59**:786–794.
- Pradel, R. 2005. Multievent: An extension of multistate capture–recapture models to uncertain states. *Biometrics* **61**:442–447.
- Pradel, R. 2009. The stakes of capture–recapture models with state uncertainty. Pages 781–795 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. Modeling demographic processes in marked populations. Environmental and ecological statistics series. Volume 3. Springer-Verlag, New York.
- Salvatori, V., and J. Linnell. 2005. Report on the conservation status and threats for wolf (*Canis lupus*) in Europe. PVS/Inf(2005) 16. Council of Europe, Strasbourg.
- Taberlet, P., and G. Luikart. 1999. Noninvasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**:41–55.
- Treves, A., and K. U. Karanth. 2003. Human-carnivore conflict and perspectives on carnivore management worldwide. *Conservation Biology* **17**:1491–1499.
- Valière, N. 2002. Amélioration et optimisation des méthodes non invasives et des marqueurs microsatellites en biologie des populations et de la conservation. PhD thesis. University of Lyon I, Lyon, France (in French).
- Valière, N., L. Fumagalli, L. Gielly, C. Miquel, B. Lequette, M. L. Poulle, J. M. Weber, R. Arlettaz, and P. Taberlet. 2003. Wolf recolonization of France and Switzerland inferred from noninvasive genetic sampling over a period of 10 years. *Animal Conservation* **6**:83–92.
- Vila, C., V. Urios, and J. Castroviejo. 1994. Use of faeces for scent marking in Iberian wolves (*Canis lupus*). *Canadian Journal of Zoology* **72**:374–377.
- Waits, L. P., and D. Paetkau. 2005. Non invasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* **69**:1419–1433.
- Williams, B. K., J. D. Nichols, and M. J. Conroy 2002. Analysis and management of animal populations. Academic Press, San Diego, California.

**Appendix 1** Mathematical formulation and instructions for implementing in program E-SURGE a model incorporating a two-class mixture in survival and detection.

At a given sampling occasion, an animal may be alive in class L, alive in class H, or may be dead. The following observations may be made: ‘1’ (if detected) and ‘0’ (if non detected). Following Pradel (2005), we define the initial state vector  $\Pi$ , the transition matrix  $\Phi$  and the event matrix  $\Omega$ . If  $\pi$  (resp.  $1 - \pi$ ) denotes the proportion of newly marked in state L (resp. state H),  $\phi^L$  the probability that an individual survives and stays in state L ( $L = L$  or H) and  $\phi^H$  the detection probability of an individual in state H ( $L = L$  or H), we have:

$$\Pi = [\pi \quad 1 - \pi \quad 0],$$

$$\Phi = \begin{bmatrix} \phi^L & 0 & 1 - \phi^L \\ 0 & \phi^H & 1 - \phi^H \\ 0 & 0 & 1 \end{bmatrix},$$

$$\Omega = \begin{bmatrix} 1 - \phi^L & \phi^L \\ 1 - \phi^H & \phi^H \\ 1 & 0 \end{bmatrix}.$$

Columns of the matrix  $\Phi$  correspond respectively to state L, H and dead, columns of the matrix  $\Omega$  correspond to the observations ‘non detected’ and ‘detected’ while rows of the matrices  $\Phi$  and  $\Omega$  correspond respectively to state L, H and dead.

Multievent models can be implemented in Program E-SURGE (Choquet et al. 2009) which is freely downloadable at <http://www.cefe.cnrs.fr/biom/En/software.htm>. The first step is to load the data into the program and specify the number of groups (1), states (3 here), events (2 here), age classes (1), and covariates (none).

Then, the model specification procedure is decomposed into i) implementing the basic structural form of the matrices using the GEPAT interface, ii) setting linear model of each parameter using the GEMACO interface and iii) fixing initial parameters using IVFV interface .

In the GEPAT module in E-SURGE, ‘\*’ entries denote the complement of the sum of positive row entries, and ‘-’ entries denote zeroes. For the initial states vector, the transition and event matrices introduced above, we have:

$$\Pi = [\pi \quad *],$$

$$\Phi = \begin{bmatrix} \psi & - & * \\ - & \psi & * \\ - & - & * \end{bmatrix},$$

$$\square = \begin{bmatrix} * & \beta \\ * & \beta \\ * & - \end{bmatrix}$$

In the GEMACO interface (Choquet 2008), predefined shortcuts are used to specify which parameters are to be time-constant, time-specific or state-specific (e.g. ‘i’ denotes constancy, ‘t’ means time effect, ‘from’ means that parameters are not equal in each matrix row. In the model considered here, we used ‘i’ for initial parameters and ‘from’ for transitions which depend on the individual state. For the events, the detection probability at the first encounter is 1 because the encounter history is conditional on being caught in the first period and the following detection probabilities depend on the state and the time occasion. Hence, the formulation for the event is ‘first+nexte.from’.

In the IVFV interface, when specifying initial parameter values, one should then constrain the first beta value associated with the first detection to 1 before running the model.

Choquet R. 2008. Automatic generation of multistate capture recapture models. *The Canadian Journal of Statistics* □ □: 43-57.

Choquet, R., L. Rouan, and R. Pradel. 2009a. Program E-SURGE: a software application for fitting multievent models. Pages 845–865 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. *Modeling Demographic Processes in Marked Populations*, Vol. 3. Springer Series: Environmental and Ecological Statistics, New York.

Pradel, R. 2005. Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics* □ 1: 442-447.

## A Appendix 2 General formulation of the population size estimator accounting for IDH.

Once demographic parameters are estimated, we turn to population size estimation.

We consider  $C$  classes alive. Let  $\mathbf{1}_C$  denotes a column vector of ones of size  $C$ . We

introduce two matrices, respectively the matrix of initial states and the matrix of survival

associated to the  $C$  classes  $\boldsymbol{\pi}_C^* = [\pi_C^1, \dots, \pi_C^C]$ , with  $\pi_C^i$  the estimated proportion of newly

marked  $i$  in state  $C$  at time  $t$ ,  $\Phi_C^* = \begin{bmatrix} \phi_C^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_C^C \end{bmatrix}$ ,  $\phi_C^i$  the estimated survival

probability in state  $C$  between time  $t$  and  $t+1$ , and the column vector

$\boldsymbol{\omega}_C^* = [1/\omega_C^1, \dots, 1/\omega_C^C]$ ,  $\omega_C^i$  the estimated detection probability in state  $C$  at

time  $t$ . Then a matrix formulation of the population size estimator at time  $t$  ( $t \geq 1$ ) is:

$$\hat{N}_C^t = \boldsymbol{\omega}_C^* \boldsymbol{\pi}_C^* \prod_{s=1}^{t-1} \boldsymbol{\pi}_C^* \boldsymbol{\Phi}_C^* \mathbf{1}_C.$$

This formula also allows to consider transitions of individuals between classes: if  $\boldsymbol{\Phi}$  is the

matrix gathering the transition probabilities between classes, then one just needs to replace

$\boldsymbol{\Phi}_C^*$  by the matrix product  $\boldsymbol{\Phi}$ . The formula we used in the main text is obtained as a

particular application with  $\boldsymbol{\pi}^* = [\pi \quad 1-\pi]$ ,  $\boldsymbol{\Phi}^* = \begin{bmatrix} \phi & 0 \\ 0 & \phi \end{bmatrix}$  and  $\boldsymbol{\omega}^* = \begin{bmatrix} 1/\omega \\ 1/\omega \end{bmatrix}$ .

**A** **d** **Results of model selection for modeling survival and detection of wolves from the French Alps (1995-2003).**

The optimal model, its neighborhood, and some further relevant models are shown, sorted by increasing AICc. The notation are:  $n_p$ , number of estimated parameters,  $\Delta AICc$ , difference between the current model and the initial model,  $\Delta AICc$ , difference in AICc between the current model and the same model without detection heterogeneity.

Abbreviations used:  $\square\square$ , heterogeneity effect;  $\square\square$ , homogeneity effect;  $\square\square\square\square$ , semester effect, 'before reproduction' analogy to the 'after reproduction' analogy

$\square\square\square\square$ , semester effect, 'cold' period (before to March) vs 'hot' period (from

September);  $\square\square\square\square$ , seasonal effect;  $\square\square\square r$ , year effect; +, additive effect of several variables;

$\square\square\square\square$ , interaction effect of several variables.

Model		K	AICc	$\Delta AICc(1)$	$\Delta AICc(2)$
$r$	c				
		6	1239.95	0.00	83.25
		5	1240.29	0.34	127.46
		7	1240.85	0.90	82.35
		6	1241.04	1.09	126.71
		8	1242.81	2.86	83.45
		7	1243.18	3.23	127.81
		11	1247.42	7.47	123.58
		10	1247.68	7.73	123.31
		38	1249.76	9.81	89.29
		37	1250.21	10.81	130.72
		5	1268.25	28.30	78.50
		4	1268.43	28.49	121.67
		21	1268.56	28.61	86.05
		6	1269.05	29.10	78.78
		5	1269.28	29.33	122.25
		7	1270.85	30.90	76.98
		6	1271.15	31.20	120.38
		20	1273.10	33.15	124.03
		12	1277.67	35.72	121.46
		13	1275.86	35.91	78.74
		67	1287.40	47.45	94.08
		68	1292.17	52.22	46.88