

# STATISTICAL REPORTS

*Ecology*, 88(2), 2007, pp. 282–288  
© 2007 by the Ecological Society of America

## ESTIMATING SPECIES-SPECIFIC SURVIVAL AND MOVEMENT WHEN SPECIES IDENTIFICATION IS UNCERTAIN

JONATHAN P. RUNGE,<sup>1,3</sup> JAMES E. HINES,<sup>2</sup> AND JAMES D. NICHOLS<sup>2</sup>

<sup>1</sup>Montana Cooperative Wildlife Research Unit, Wildlife Biology Program, The University of Montana,  
Missoula, Montana 59812 USA

<sup>2</sup>United States Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA

**Abstract.** Incorporating uncertainty in the investigation of ecological studies has been the topic of an increasing body of research. In particular, mark–recapture methodology has shown that incorporating uncertainty in the probability of detecting individuals in populations enables accurate estimation of population-level processes such as survival, reproduction, and dispersal. Recent advances in mark–recapture methodology have included estimating population-level processes for biologically important groups despite the misassignment of individuals to those groups. Examples include estimating rates of apparent survival despite less than perfect accuracy when identifying individuals to gender or breeding state. Here we introduce a method for estimating apparent survival and dispersal in species that co-occur but that are difficult to distinguish. We use data from co-occurring populations of meadow voles (*Microtus pennsylvanicus*) and montane voles (*M. montanus*) in addition to simulated data to show that ignoring species uncertainty can lead to biased estimates of population processes. The incorporation of species uncertainty in mark–recapture studies should aid future research investigating ecological concepts such as interspecific competition, niche differentiation, and spatial population dynamics in sibling species.

**Key words:** *apparent survival; competition; dispersal; mark–recapture methodology; misclassification; sibling species; spatial population dynamics; survival; vital rates.*

### INTRODUCTION

Uncertainty plagues ecological investigations. Quantifying and incorporating uncertainty into inference procedures has become an increasingly important focus of many research efforts in the past four decades (Williams et al. 2002). The field of statistical ecology has recently begun to recognize that one form of uncertainty, the ability (or lack thereof) to correctly classify individuals to groups such as gender or breeding class, can lead to uncertainty and bias in estimates of vital rates. When this uncertainty is incorporated in statistical methodology for mark–recapture data, vital rates such as survival and reproduction can be accurately estimated for species with cryptic external differences between individuals of different gender or breeding status (Conroy et al. 1999, Fujiwara and Caswell 2002,

Lebreton and Pradel 2002, Kendall et al. 2003, Nichols et al. 2004, Pradel 2005). A similar problem occurs when species themselves are difficult to differentiate. Here, we introduce a method for estimating species-specific rates of apparent survival and dispersal when individuals are difficult to identify to species.

Cryptic species co-inhabiting the same area (hereafter “sibling species,” sensu Futuyma 1998) occur throughout the natural world. Many small mammals are sibling species including white-footed mouse (*Peromyscus leucopus*) and deer mouse (*P. maniculatus*), meadow vole (*Microtus pennsylvanicus*) and montane vole (*M. montanus*), and various shrew (*Sorex*) species. Other vertebrate examples include female Blue-winged Teal (*Anas discors*) and Cinnamon Teal (*A. cyanoptera*), rainbow trout (*Oncorhynchus mykiss*) and cutthroat trout (*O. clarki*), *Cnemidophorus* lizards, *Anolis* lizards, and torrent salamanders (*Rhyacotriton*).

Rigorous estimation of species-specific vital rates for sibling species has yet to be investigated. Two elements are crucial for such research. First, individuals must have some morphological characteristics upon which to base field identification of species. Examples include shape and size of bill in female teal (LeMaster 1986),

Manuscript received 5 April 2006; revised 28 June 2006; accepted 24 July 2006; final version received 22 August 2006. Corresponding Editor: S. T. Buckland.

<sup>3</sup> Present address: Georgia Cooperative Fish and Wildlife Research Unit, D. B. Warnell School of Forestry and Natural Resources, University of Georgia, Athens, Georgia 30602 USA. E-mail: jprunge@uga.edu

pelage color in mice (Bruseo et al. 1999, Foresman 2001), and basibranchial teeth in trout (Leary et al. 1996). Second, an identifier of true species identification must be obtained from a subset of the sampled individuals. Possibly, species-specific vital rates can be estimated using mixture models even when a true species identity cannot be obtained for any animals (e.g., using the models of Pradel 2005), but such models may not be identifiable in at least some situations. At a minimum, knowledge of true species identity for some individuals permits more precise estimation of focal quantities.

The true identity of sibling species can be determined by a variety of methods, and our modeling approach can incorporate the many types of research that investigate this problem. Examples include dental pattern or skull morphology of individuals removed due to pathology research or handling mortality, isozymes from blood samples taken from a subset of animals in a combined demographic and immunological study, and DNA samples taken from blood or body tissue. One specific example: the USGS Bird Banding Lab currently prohibits banding of individuals that cannot be identified to species (e.g., some *Empidonax* flycatchers). However, if banding was possible, then researchers using the statistical methodology described below could remove feathers from cryptic individuals, identify them to species with a DNA-based method, and obtain species-specific information regarding both demography and identification.

In this paper, we present a method to estimate species-specific apparent survival and dispersal despite uncertainty in species identification. We apply the method to populations of meadow and montane voles occurring in two habitats. Additionally, we conduct a simulation analysis to investigate properties of the new estimators and show that the incorporation of a classification parameter for species identification can lead to different results than those obtained from “naïve” estimates derived from species identification unadjusted for misclassification.

METHODS

Statistical model

The model we present is an extension to the mark-recapture, multistate, Arnason-Schwarz model (Arnason 1972, Brownie et al. 1993, Schwarz et al. 1993). The extension concerns the incorporation of uncertain assignment of individuals to species when estimating relevant parameters. In this model, animals that are not released (e.g., that die on capture) are positively identified to species, and animals that are released receive only a judgment-based assignment (hereafter termed “assignment”) for species identification. We define state as a habitat, although one could also define state as a stage class based upon morphological characters of individuals (e.g., mass) or as a certain stage class in a certain habitat. Five types of parameters are required by our model:

1)  $\phi_{i(u)}^{(v)rs}$  = probability of survival from period  $i$  to  $i + 1$  for individuals of species  $u$ ,  $u \in \{A, B\}$ , age  $v$  ( $y$  = young,  $a$  = adult), in state  $r$ ,  $r \in \{1, 2\}$ , at period  $i$  and in state  $s$ ,  $s \in \{1, 2\}$ , at period  $i + 1$ ;

2)  $p_{i(u)}^r$  = probability of recapture in period  $i$  for adults of species  $u$  in state  $r$ ;

3)  $\pi_{i(A)}^{(v)r}$  = probability that an individual of age  $v$  that is first captured in period  $i$  and in state  $r$  is a member of species A;

4)  $\delta_{i(u)}^{(v)r}$  = probability that an individual of species  $u$ ,  $u \in \{A, B\}$ , is correctly assigned to species  $u$ , for period  $i$ , age  $v$ , state  $r$  (assigned  $u | u$ );

5)  $\eta_{i(u)}^{(v)r}$  = probability that an individual of species  $u$ , age  $v$ , in state  $r$  that is captured in period  $i$  survives capture at period  $i$  to be released.

Also note that  $1 - \delta_{i(u)}^{(v)r}$  = probability that an individual of species  $u$ ,  $u \in \{A, B\}$ , is incorrectly assigned to species  $z$ ,  $z \in \{A, B\}$ , for period  $i$ , age  $v$ , state  $r$  (assigned  $z | u$ ). Recapture probability is undefined for young animals because we consider conditional (on release) models, and we assume that all young animals become adults after one time step (animals mature in the interval separating successive sampling occasions). We note that the probability of surviving and dispersing from one location to the other is expressed as  $\phi_{i(u)}^{(v)12}$  or  $\phi_{i(u)}^{(v)21}$ . If survival between periods  $i$  and  $i + 1$  depends only on the state at  $i$ , then we can separately estimate the survival and transition probabilities based on the following relationship:

$$\phi_{i(u)}^{(v)rs} = S_{i(u)}^{(v)r} \psi_{i(u)}^{(v)rs}$$

where  $S_{i(u)}^{(v)r}$  is the probability that an individual of species  $u$  and age  $v$  released in state  $r$  at sampling period  $i$  survives and remains in the study system until just before sampling period  $i + 1$ , and  $\psi_{i(u)}^{(v)rs}$  is the probability of moving from state  $r$  to state  $s$ . The probability of staying within a state (e.g.,  $\psi_i^{11}$ ) can be estimated by subtraction because  $\sum_s \psi_i^{rs} = 1$ . For instance, in the two-state system presented here,  $\psi_i^{11} = 1 - \psi_i^{12}$ . With the software developed to implement this model (MSSRV\_UNK\_SPECIES), the parameters  $\delta$ ,  $\eta$ ,  $\pi$ ,  $p$ ,  $S$ , and  $\psi$  are all estimable from mark-recapture data.

The capture history data from which multistate, mark-recapture statistics are estimated have a specific format. For instance, the capture history 102 would indicate that an animal is captured in state 1 during sample period 1, not captured during sample period 2, and captured in state 2 during sample period 3. As additional modifiers to this standard multistate capture history, we use yBAN to indicate that the individual was young when first caught, was found to belong to species B, was assigned to species A at its first capture, and was not released upon its final capture. Note that logistical considerations caused us to assign species only upon initial capture of each animal. If independent assignments are made at multiple captures of an animal, then our model can readily be extended to include this

additional information on the classification probability parameters. The probability associated with the above capture history (conditional on release in period 1) would be

$$\begin{aligned}
 &P(102\_yBAN) \\
 &= [1 - \pi_{1(A)}^{(y)1}] \eta_{1(B)}^{(y)1} [1 - \delta_{1(B)}^{(y)1}] \\
 &\quad \times \left\{ \phi_{1(B)}^{(y)11} [1 - p_{2(B)}^1] \phi_{2(B)}^{(a)12} + \phi_{1(B)}^{(y)12} [1 - p_{2(B)}^2] \phi_{2(B)}^{(a)22} \right\} \\
 &\quad \times p_{3(B)}^2 [1 - \eta_{3(B)}^2]. \quad (1)
 \end{aligned}$$

The first expression,  $1 - \pi_{1(A)}^{(y)1}$ , is the probability that a young animal caught in state 1 during sample period 1 is a member of species B; because there are only two species in this model, we can simply define this probability as the complement of the probability of being species A. The second expression,  $\eta_{1(B)}^{(y)1}$ , is the probability of release for individuals of species B, age  $y$ , captured in state 1 during sample period 1. The next expression,  $1 - \delta_{1(B)}^{(y)1}$ , is the probability that a young animal of species B was incorrectly assigned to species A during sample period 1 in state 1. The term within the braces in Eq. 1 accounts for the fact that we know the animal's species but have incomplete knowledge regarding its location between sample periods 1 and 3. Either the animal stayed in state 1 where it was not recaptured in sample period 2, then moved to state 2 prior to sample period 3 ( $\phi_{1(B)}^{(y)11} [1 - p_{2(B)}^1] \phi_{2(B)}^{(a)12}$ ), or it moved to state 2 prior to sample period 2, was not recaptured there in period 2, and remained there until period 3 ( $\phi_{1(B)}^{(y)12} [1 - p_{2(B)}^2] \phi_{2(B)}^{(a)22}$ ). Finally, the term  $p_{3(B)}^2 [1 - \eta_{3(B)}^2]$  is the probability the animal was recaptured but not released in state 2 during sample period 3.

If we modify the above capture history so that the animal is released at its last capture (denoted by a "Y"), and therefore true species is never known (denoted by a "U"), i.e., capture history 102\_yUAY, then we must adjust the above probability structure by incorporating the possibility that the species was correctly assigned to species A:

$$\begin{aligned}
 &P(102\_yUAY) \\
 &= \left[ [1 - \pi_{1(A)}^{(y)1}] \eta_{1(B)}^{(y)1} [1 - \delta_{1(B)}^{(y)1}] \right. \\
 &\quad \times \left\{ \phi_{1(B)}^{(y)11} [1 - p_{2(B)}^1] \phi_{2(B)}^{(a)12} + \phi_{1(B)}^{(y)12} [1 - p_{2(B)}^2] \phi_{2(B)}^{(a)22} \right\} \\
 &\quad \times p_{3(B)}^2 \eta_{3(B)}^{(a)2} \left. + \left( \pi_{1(A)}^{(y)1} \eta_{1(A)}^{(y)1} \delta_{1(A)}^{(y)1} \right) \right. \\
 &\quad \times \left\{ \phi_{1(A)}^{(y)11} [1 - p_{2(A)}^1] \phi_{2(A)}^{(a)12} + \phi_{1(A)}^{(y)12} [1 - p_{2(A)}^2] \phi_{2(A)}^{(a)22} \right\} \\
 &\quad \left. \times p_{3(A)}^2 \eta_{3(A)}^{(a)2} \right].
 \end{aligned}$$

The addition of the probability statements pertaining to species A is necessary to properly incorporate the uncertainty associated with species assignment and to

admit the possibility that the animal could belong to either species A or B.

Two other types of histories are possible. An individual may never receive a species assignment but may be identified to species with certainty. For example, capture history 102\_yBUN would have a similar probability statement to Eq. 1 but with no  $\delta$  terms because it never received a species assignment. Another possibility is an animal that received neither a species assignment nor a positive species identification, e.g., capture history 102\_yUUY. The associated probability statement would be similar to Eq. 2 but would have no  $\delta$  terms.

Probabilities such as Eqs. 1 and 2 can be viewed as multinomial cell probabilities. Together with the observed numbers of individuals exhibiting each capture history, they form a likelihood function from which estimates can be obtained. In this case, the likelihood was coded into program SURVIV (White 1983).

An additional source of information for  $\delta$  in many studies may be animals from other areas that are not part of the capture–recapture data set undergoing analysis. Examples may include small mammals from other trapping arrays or amphibians sampled in separate areas. If no differences in rates of correct classification are found among areas sampled, then individuals from different areas can be incorporated in the estimation of  $\delta$  with a binomial likelihood expression. This likelihood can be multiplied by the more general capture–recapture likelihood, and estimation can be based on the combined data.

The uncertain species model contains assumptions standard for multiage, multigroup, multistate, mark–recapture models, and violations are discussed in depth elsewhere (Williams et al. 2002: 422–423, 434–436, 442, 458–459). The model and associated software were developed for a specific sampling situation that includes two species, two states, and two ages with age progression being deterministic rather than stochastic. The modeling assumes that transitions between states are Markovian, i.e., the probability of an animal occurring in state  $r$  during period  $i + 1$  is determined solely by the state of the animal during period  $i$ , but no earlier. Our field data included only one individual that was not released and not identified to species, so we did not account for such capture histories in the modeling, although this possibility could be readily added. We note that the incorporation of time-specific covariates affecting groups of individuals (e.g., weather conditions) is possible using the current software, but the incorporation of individual covariates (e.g., body mass) would require additional modeling and software development. The software (MSSRV\_UNK\_SPECIES) providing estimates and variances of the parameters listed above is available from the software archive at the Patuxent Wildlife Research Center (*available online*).<sup>4</sup>

<sup>4</sup> <http://www.mbr-pwrc.usgs.gov/software/>

### Microtus data

The field data analyzed here were collected between June and August 2002 from trapping grids located near Charlo, Montana. The subjects of the capture–recapture data were voles (*Microtus*) on a single grid, although data collected on voles from other grids were additionally used to estimate classification probabilities,  $\delta$ . Grids were  $100 \times 160$  m in size with traps spaced every 10 m. Each grid was bisected by a livestock fence, and one half of the grid was grazed while the other half was ungrazed. Five grids formed the study. The capture data we analyze here come from a grid that did not receive the anticipated treatment of grazing during the trapping season. The “grazed” side of the grid had last contained livestock in November 2001, and the “ungrazed” side had last contained a small number of livestock in October 2000. Nonetheless, significant differences in vegetation structure were recorded between the habitats with the ungrazed side containing higher, denser vegetation and deeper, more extensive vegetative litter (Runge 2005).

Trapping was conducted according to the robust design (Pollock 1982), in which several consecutive secondary trapping periods (or trap nights) compose a single primary trapping period. Traps were set from 20:00 to 21:00 hours, and trap checking and animal release began at 06:00 hours, typically ending by 09:00 hours. Intervals between primary trapping periods were 3 weeks. Four primary periods composed the total study length for the grid analyzed here, and the third primary period consisted of five secondary periods while the first, second, and fourth primary periods consisted of four secondary periods. For this data set, we combine the data across secondary periods within a single primary period so that if an animal was captured in any one secondary period, it is considered captured for the primary period. Thus, the integers composing the capture history (0,1,2) denote whether the animal was not captured (0), or captured in ungrazed (1) or grazed (2) habitat for a given primary period.

Once caught, animals were marked both with ear tags and clipped toes to ensure that marks were not lost or misread. The majority of animals also received an assignment for species based upon dorsal pelage color, with meadow voles having darker overall coloration than montane voles; the first author was responsible for all species assignments. Animals that died in traps were identified to species based upon upper molar ( $M^2$ ) pattern; meadow voles have a posterior loop in the  $M^2$  that is absent in montane voles (Foresman 2001). Age was determined by weight, and the criterion separating young from adults was 14.75 g for animals known or judged to be montane voles and 17.25 g for animals known or judged to be meadow voles. All animals identified as young in one primary sampling period gained sufficient weight to be classified as adults in the subsequent period.

*Model selection.*—We used a data set consisting only of females for model selection and estimation. We did this because program MSSRV\_UNK\_SPECIES currently does not incorporate the variable “gender” for model selection, and previous work shows that gender-specific differences in both apparent survival and dispersal occur in vole populations (Coffman et al. 2001), which precludes combining males and females in one data set for estimation purposes. The data set consisted of 143 and 73 females (total for both species) first caught in ungrazed and grazed habitat, respectively.

The parameters  $\delta$ ,  $\eta$ , and  $S$  could all be modeled to examine whether they varied across time, state, age, or species; whereas  $\pi$  could be modeled only on time, age, and state. As mentioned above, we required a  $p$  for adults only in our modeling. Potential predictor variables for the parameter  $\psi$  included time, age, species, and direction (e.g., from ungrazed to grazed habitat or vice versa). Insufficient data precluded examining how  $\psi$  varied with time.

A preferred method for model selection involves calculating AIC for a full set of pre-defined biologically reasonable models (Burnham and Anderson 2002). We hypothesized five models to be reasonable for modeling  $\delta$ , 23 for  $\eta$ , five for  $\pi$ , eight for  $p$ , 10 for  $S$ , and four for  $\psi$ . We were thus forced to use a sequential model selection approach, although we note that modeling every possible combination of such models has been conducted (King and Brooks 2002).

Because  $\delta$  could strongly influence species-specific estimates of other parameters, it was modeled first. Next,  $\eta$ ,  $\pi$ ,  $p$ ,  $S$ , and  $\psi$  were sequentially modeled. For  $\eta$  and  $\pi$ , the lowest AIC model from fitting one parameter was used to fit the a priori variable structures for the subsequent parameter. Then all variable structures within 1.00 AIC unit of the best model for the previous parameter were considered in combination with all structures yielding AIC values within 1.00 AIC of the best model for the subsequent parameter. For  $p$  and  $S$ , the same process was repeated, but additional combinations of  $p$  and  $S$  were considered because the manner in which  $p$  was modeled strongly affected the estimation of  $S$ . We make no claims of optimality of this approach to model selection and note that it simply represents one pragmatic possibility.

*Analyses.*—We conducted two analyses to obtain parameter estimates; one formally incorporated uncertainty with the model and software discussed above, and the other (labeled the “naïve” analysis) treated species assignments for released animals as actual species identifications using program MARK (White and Burnham 1999). We compare results obtained from the uncertain species analysis with corresponding results from the naïve analysis both for field-based and simulated data.

For the simulated data, we generated capture histories, treated them as data, and then observed the correspondence of parameter “estimates” with true

TABLE 1. Model selection results.

$\pi$	Model			$K$	$\Delta AIC$
	$p$	$S$	$\psi$		
st + age	...	sp	...	13	0.00
st + age	...	st	...	13	0.21
st + age	sp	sp	...	14	0.77
st + age	...	sp $\times$ seas	...	15	0.80
st + age	T	sp	...	14	0.96
st + age	...	sp + st	...	14	1.07
st $\times$ age	...	st	...	14	1.17
st + age	T	st	...	14	1.30
st + age	...	sp	sp	14	1.31
st + age	st	sp	...	14	1.49
st + age	st	sp $\times$ seas	...	14	1.69
st + age	...	sp	dir	14	1.73
st + age	sp	sp	sp	15	1.78
st + age	sp	sp $\times$ seas	...	16	1.90
st $\times$ age	...	sp	...	14	1.90
st + age	...	st	sp $\times$ dir	16	1.92
st + age	T	sp $\times$ seas	...	16	1.95

Notes: Each model also contains the terms  $\delta(\text{age})$  and  $\eta(\text{state} + \text{time}^2)$ . Models shown here are within 2.0 AIC units of the model with the lowest AIC. Parameters are as defined in *Methods*:  $\delta$  is the probability of being correctly identified to species;  $\eta$  is the probability of release;  $\pi$  is the probability of being a montane vole;  $p$  is the probability of recapture;  $S$  is the probability of surviving and staying in the study system;  $\psi$  is the probability of inter-habitat dispersal;  $K$  is the number of parameters in a given model; and  $\Delta AIC$  is the difference between the model of interest and the model with the lowest AIC. Abbreviations: sp, species; st, state (habitat); dir, direction-specific movement ( $\psi_{i(u)}^{(v)12} \neq \psi_{i(u)}^{(v)21}$ ), seas, season (the interval between the first and second primary period took place mainly during spring, the other two intervals during summer); T indicates that the logit of the parameter was modeled as a linear function of time (with periods numbered consecutively beginning with 0).

parameter values. For both uncertain and naïve analyses, we computed bias ( $\text{bias} = [\sum_{j=1}^n \hat{\theta}_j - \theta]/n$ ), where  $\hat{\theta}$  is the estimator being evaluated and  $n$  is the number of simulations) in addition to 95% confidence interval coverage and root mean squared error (full details are in the Appendix).

## RESULTS

The model with the lowest AIC for the *Microtus* data was  $\delta(\text{age}) \eta(\text{state} + \text{time}^2) \pi(\text{state} + \text{age}) p(\cdot) S(\text{species}) \psi(\cdot)$ . The next closest model was 0.21 AIC units higher and differed only in  $S$  being fit by state. Fifteen additional models were between 0.77 and 1.95 AIC units higher (Table 1). For the low-AIC model,

estimates of correct classification probability ( $\hat{\delta}$ ) were 0.586 ( $\hat{SE} = 0.152$ ) for young and 0.870 (0.038) for adults. As mentioned above, ancillary data from both sexes were combined for  $\hat{\delta}$ , whereas all other parameter estimates are based on capture histories for females only. The estimate of recapture probability ( $\hat{p}$ ) from this model was 0.518 (0.141). The estimates of the probability of release ( $\hat{\eta}$ ) for the four primary periods were 0.496 (0.078), 0.794 (0.061), 0.869 (0.042), and 0.833 (0.051) in ungrazed habitat and 0.778 (0.090), 0.932 (0.039), 0.959 (0.024), and 0.947 (0.029) in grazed habitat. Cold weather in time period 1 resulted in high trap mortality, with wetter conditions in ungrazed habitat exacerbating these conditions.

Estimates of apparent survival ( $\hat{S}$ ) for the uncertain species analysis differed from those of the naïve analysis by 0.089–0.182 (11.0–51.3%) when  $S$  was fit by species and 0.016–0.031 (3.7–3.8%) when  $S$  was fit by state (Table 2). Estimates of inter-habitat dispersal ( $\hat{\psi}$ ), however, were identical, and indeed this was expected for models without species-specific differences in movement. For the simulated data, differences between the two types of analysis were also apparent. Bias of the estimated parameter values in simulations were generally near 0 for the uncertain species analysis but increasingly diverged from 0 for the naïve analysis as the difference between  $S$  of species A and B increased (Fig. 1). Further information regarding simulation results is found in the Appendix (Tables A1–A4, Figs. A1 and A2).

Incorporating uncertainty in species classification caused higher coefficients of variation for parameter estimates of  $S$  and  $\psi$ , providing a more accurate indication of uncertainty. For the two best models, the coefficients of variation (CV) for the naïve analysis were less than those of the uncertain species analysis for both  $\hat{S}$  (average difference in CV = 10.3%) and  $\hat{\psi}$  (average difference in CV = 127.4%). This pattern was consistent with simulated data suggesting that the uncertain species analysis overestimates standard error (Appendix: Tables A1–A3), whereas the naïve analysis does not account for the uncertainty in species classification.

## DISCUSSION

Species identification problems occur in various taxa, and here we have presented a method for incorporating species uncertainty when estimating probabilities of survival and dispersal in mark–recapture studies. This

TABLE 2. Parameter estimates (with SE in parentheses) for the two best statistical models for meadow voles (MiPe) and montane voles (MiMo) using both uncertain species (Unsp) and naïve analyses.

Analysis	$S$ (species)			$S$ (state)		
	$\hat{\psi}$	$\hat{S}$ MiPe	$\hat{S}$ MiMo	$\hat{\psi}$	$\hat{S}$ ungrazed	$\hat{S}$ grazed
Unsp	0.041 (0.076)	0.812 (0.133)	0.355 (0.150)	0.043 (0.078)	0.810 (0.133)	0.427 (0.136)
Naïve	0.041 (0.023)	0.723 (0.079)	0.537 (0.107)	0.043 (0.024)	0.779 (0.092)	0.411 (0.094)

Note: Parameters are apparent survival ( $\hat{S}$ ) and estimates of inter-habitat dispersal ( $\hat{\psi}$ ).

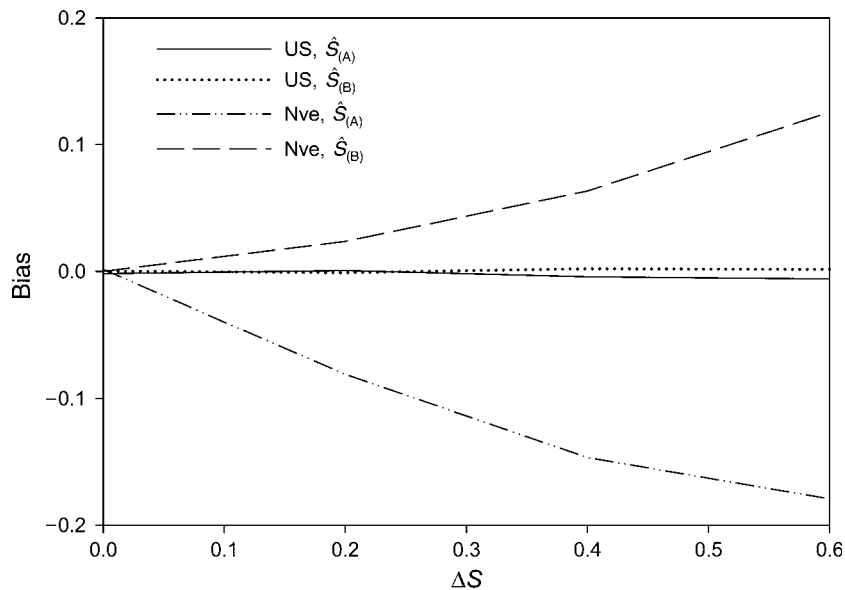


FIG. 1. Bias of estimated apparent survival rates ( $\hat{S}_{(A)}$ ,  $\hat{S}_{(B)}$ ) for uncertain species (US) and naïve (Nve) analyses as a function of the true difference in apparent survival rates between the two species ( $\Delta S = S_{(A)} - S_{(B)}$ ) for  $S_{(A)} = 0.8$ . Correct classification rates were 0.9 for adults and 0.6 for young.

adds to the growing literature investigating group uncertainty that includes misassignment of individuals to gender (Conroy et al. 1999, Nichols et al. 2004), genotype (Lukacs and Burnham 2005), and state (Fujiwara and Caswell 2002, Lebreton and Pradel 2002, Kendall et al. 2003, Pradel 2005). The inclusion of group classification parameters in mark–recapture methodology provides additional realism and properly incorporates uncertainty in species identification when estimating population level processes.

Correct inference for parameters such as apparent survival rests upon the proper incorporation of relevant sources of uncertainty. Naïve analysis of data characterized by uncertain species identification would be expected to underestimate species differences, because misassignment incorrectly classifies members of one species as another, therefore diluting species-specific differences. This was evident in the simulation study, in which the naïve analysis overestimated  $S$  for the species with lower true rates of apparent survival and underestimated  $S$  for the species with higher rates (Fig. 1). Similarly, the naïve analysis underestimated the difference in  $S$  ( $\hat{\Delta S}$ ) between the two species for both field-based and simulation analyses (Table 2; Appendix: Table A4). In summary, the simulated examples show that when apparent survival rates between species diverge, naïve analyses can produce estimates with high bias and low confidence interval coverage (Fig. 1; Appendix: Figs. A1, A2, Tables A1–A4). Thus, use of naïve analyses in systems with species uncertainty can dilute the ability to accurately estimate species-specific survival rates.

The uncertain species approach may overestimate standard error (Appendix: Tables A1–A3). We are

uncertain of the reason for this but note that it is not uncommon for asymptotic, model-based variance estimates to be biased in the case of small sample sizes. We emphasize that the uncertain species analysis includes the true quantity of interest in an estimated 95% confidence interval rather than excludes it as the naïve analysis often does (Appendix: Fig. A2) and is superior at estimating the true value of  $\Delta S$  (Appendix: Table A4), which is often a quantity of primary interest in the study of sibling species. Our simulation results in the Appendix do lead us to recommend a bootstrap approach to variance estimation when these models are used with small sample sizes.

We also note that the uncertain species analysis likely requires more data for convergence than a multi-state analysis without the species classification parameter. For the simulation study, when sample sizes were 100 individuals per state and age class, only 46% of the simulations converged (Appendix: Table A3). Similarly, convergence was achieved for the field data only by entering initial parameter values close to those provided by the maximum likelihood routine. However, given that sample size in grazed habitat over the four periods of the field study was 73 individuals, that state-specific models in the uncertain species analysis converged when initial parameter values were entered, and that the naïve analysis poorly estimates  $\Delta S$ , we submit that incorporating species uncertainty is advisable for studies investigating survival and dispersal processes in sibling species.

The current model contains rather specific assumptions about the system of interest, e.g., there are only two states, two species, and two age classes, and all individuals that are not released should be unequivocally

identified to species. The details provided here should enable future research to relax these assumptions. We foresee that studies using genetic methods to identify individuals to species may identify a sub-sample of individuals that are also released. If so, an extra parameter incorporating the probability of an individual being identified to species, whether it was released or not, would need to be included in the likelihood.

The coexistence of sibling species is a specific, yet interesting, phenomenon. In some situations, sibling species may represent a recent evolutionary divergence in a state of non-equilibrium, but sibling species also have the capacity to coexist in perpetuity (Amarasekare 2003, Zhang et al. 2004). Competition between sibling species may have implications for management and conservation of native species (Griffith 1988, Geller 1999), and incorporating uncertainty in species classification should aid researchers investigating this phenomenon. Ignoring species uncertainty can lead to both inaccuracy in parameter estimation and spurious conclusions in tests of ecological hypotheses and thus may advance unwarranted conservation and management actions. Therefore, the incorporation of species uncertainty should aid both effective implementation of ecological management and advancement of ecological theory for sibling species.

#### ACKNOWLEDGMENTS

We thank E. Rexstad and an anonymous referee for valuable suggestions. J. Ball, D. Christian, E. Crone, K. Foresman, and S. Mills provided useful suggestions for an earlier version of this manuscript. J. P. Runge received support from the Montana Cooperative Wildlife Research Unit and NSF-EPSCoR Grant #EPA 009-1995. This manuscript represents a portion J. P. Runge's doctoral dissertation at The University of Montana.

#### LITERATURE CITED

- Amarasekare, P. 2003. Competitive coexistence in spatially structured environments: a synthesis. *Ecology Letters* 6: 1109–1122.
- Arnason, A. N. 1972. Parameter estimates from mark-recapture experiments on two populations subject to migration and death. *Researches on Population Ecology* 13:97–113.
- Brownie, C., J. E. Hines, J. D. Nichols, K. H. Pollock, and J. B. Hestbeck. 1993. Capture-recapture studies for multiple strata including non-Markovian transitions. *Biometrics* 49:1173–1187.
- Bruseo, J. A., S. H. Vessey, and J. S. Graham. 1999. Discrimination between *Peromyscus leucopus noveboracensis* and *Peromyscus maniculatus nubiterrae* in the field. *Acta Theriologica* 44:151–160.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.
- Coffman, C. J., J. D. Nichols, and K. H. Pollock. 2001. Population dynamics of *Microtus pennsylvanicus* in corridor-linked patches. *Oikos* 93:3–21.
- Conroy, M. J., J. C. Senar, J. E. Hines, and J. Domenach. 1999. Development and application of a mark-recapture model incorporating predicted sex and transitory behaviour. *Bird Study* 46(supplement):62–73.
- Foresman, K. R. 2001. The wild mammals of Montana. Special Publication No. 12. American Society of Mammalogists, Lawrence, Kansas, USA.
- Fujiwara, M., and H. Caswell. 2002. Estimating population projection matrices from multi-stage mark-recapture data. *Ecology* 83:3257–3265.
- Futuyma, D. F. 1998. Evolutionary biology. Third edition. Sinauer, Sunderland, Massachusetts, USA.
- Geller, J. B. 1999. Decline of a native mussel masked by a sibling species invasion. *Conservation Biology* 13:661–664.
- Griffith, J. S. 1988. Review of competition between cutthroat trout and other salmonids. American Fisheries Society Symposium 4:134–140.
- Kendall, W. L., J. E. Hines, and J. D. Nichols. 2003. Adjusting multistate capture-recapture models for misclassification bias: Manatee breeding proportions. *Ecology* 84:1058–1066.
- King, R., and S. P. Brooks. 2002. Bayesian model discrimination for multiple strata capture-recapture data. *Biometrika* 89:785–806.
- Leary, R. F., W. R. Gould, and G. K. Sage. 1996. Success of basibrachial teeth in indicating pure populations of rainbow trout and failure to indicate pure populations of westslope cutthroat trout. *North American Journal of Fisheries Management* 16:210–213.
- Lebreton, J. D., and R. Pradel. 2002. Multistate recapture models: modelling incomplete individual histories. *Journal of Applied Statistics* 29:353–369.
- LeMaster, R. 1986. Waterfowl identification: the LeMaster method. Contemporary Books, Chicago, Illinois, USA.
- Lukacs, P. M., and K. P. Burnham. 2005. Estimating population size for DNA-based closed capture-recapture data incorporating genotyping error. *Journal of Wildlife Management* 69:396–403.
- Nichols, J. D., W. L. Kendall, J. E. Hines, and J. A. Spendelov. 2004. Estimation of sex-specific survival from capture-recapture data when sex is not always known. *Ecology* 85: 3192–3201.
- Pollock, K. H. 1982. A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management* 46:757–760.
- Pradel, R. 2005. Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics* 62:442–447.
- Runge, J. P. 2005. Spatial population dynamics of *Microtus* in grazed and ungrazed grasslands. Dissertation. The University of Montana, Missoula, Montana, USA.
- Schwarz, C. J., J. F. Schweigert, and A. N. Arnason. 1993. Estimating migration rates using tag recovery data. *Biometrics* 49:177–193.
- White, G. C. 1983. Numerical estimation of survival rates from band recovery and biotelemetry data. *Journal of Wildlife Management* 47:716–728.
- White, G. C., and K. P. Burnham. 1999. Program MARK: Survival estimation from populations of marked animals. *Bird Study* 46(supplement):120–138.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic, San Diego, California, USA.
- Zhang, D.-Y., K. Lin, and I. Hanski. 2004. Coexistence of cryptic species. *Ecology Letters* 7:165–169.

#### APPENDIX

Simulation study for uncertain species analysis (*Ecological Archives* E088-016-A1).