Wildlife Society Bulletin; DOI: 10.1002/wsb.382

Tools and Technology

Likelihood-Based Photograph Identification: Application With Photographs of Free-Ranging Bison

JEROD A. MERKLE,[1] *Département de Biologie, Centre d'Étude de la Forêt, Université Laval, Pavillon Alexandre Vachon 1045, Avenue de la Médecine, Québec, QC, Canada G1V 0A6*

DANIEL FORTIN, *Département de Biologie, Centre d'Étude de la Forêt, Université Laval, Pavillon Alexandre Vachon 1045, Avenue de la Médecine, Québec, QC, Canada G1V 0A6*

ABSTRACT Using photographs to identify individual animals and monitor populations is becoming more common. However, photographic identification methods where measurements of morphological traits (e.g., horn length) are compared have received little attention. We present an approach for aiding with the identification of individual animals from photographs. The approach incorporates measurement data, metadata from photographs, and multiple sources of error, and calculates a matching score between pairs of photographs using a likelihood-based algorithm. We tested and identified the false-rejection error rate using 91 photographs, representing 33 known free-ranging bison (*Bison bison*), and 117 simulated data sets with varying numbers of unique individuals, morphological measurements, and photograph error. We then used the approach to estimate the adult population size of bison in Prince Albert National Park, Canada, in 2011. For bison, the false-rejection rate of our approach was 0.055. Using a Huggins closed population model with misidentification, we estimated 103 (95% CI = 82–130) and 46 (95% CI = 37–58) adult female and male bison, respectively. After incorporating field-based calf- and juvenile-to-female ratios, we estimated 202 (95% CI = 171.6–231.4) bison. We found this estimate to be plausible using 2 minimum-count aerial surveys conducted in March 2011 and 2012 for comparison. With our approach, researchers and managers can build capture histories of individuals, which can be used for studies of population dynamics and habitat selection. This approach can incorporate any morphological measurements extracted from photographs (e.g., coat color), making it robust to a variety of species and study systems. © 2013 The Wildlife Society.

KEY WORDS *Bison bison*, capture–mark–recapture, computer-assisted photographic-identification, likelihood, misidentification, photogrammetry, photographic mark–recapture.

The ability to uniquely identify individuals of a population is often the basis for estimating ecological parameters (e.g., population size, habitat selection coeff.; Nichols 1992, Nietfeld et al. 1994, Wilson et al. 2003). Many methods are available to mark individuals, and the feasibility of each method for a particular species can vary (Nietfeld et al. 1994). Traditional means of marking individuals (e.g., permanent tags, radiocollars) can be expensive and logistically challenging (Powell and Proulx 2003), and in some cases have negative effects on the marked individual (Gordon et al. 1988, Krausman et al. 2004, Cattet et al. 2008).

Researchers and managers are therefore turning to less invasive methods for marking individuals, such as non-invasively sampled DNA (Waits and Paetkau 2005), and tracking natural markings through the use of photography using photogrammetry (Cutler and Don 1999, Berger 2012). Photograph technology advancements (e.g., equipment, photograph quality, photograph manipulation software) and lower costs have permitted widespread use of photography to mark individual animals by identifying unique phenotypic characteristics of individual animals (Kelly 2001, Goswami et al. 2007, Sacchi et al. 2010). Because of phenotypic complexity, photograph quality variability, and database management, a variety of different computer-assisted photograph-identification (CAPI) software have been developed to aid in deciding whether the animals in 2 photographs are, or are not, a match (Araabi et al. 2000, Speed et al. 2007, Van Tienhoven et al. 2007, Gamble et al. 2008, Bolger et al. 2012).

Most types of software were developed for identifying individuals by the spatial and dimensional distribution of recognizable shapes (e.g., spots) and patterns (e.g., stripes). They are based on the assumption that the phenotypic markings in question are highly variable such that each individual within a population has its own unique phenotype or combination of phenotypes. For example, the distribution and size of spots on whale sharks (*Rhincodon typus*) and the stripe patterns on wildebeest (*Connochaetes taurinus*) flanks can be used to identify individuals (Arzoumanian et al. 2005,

Received: 25 April 2013; Accepted: 7 August 2013

[1]*E-mail: jerod.merkle.1@ulaval.ca*

Meekan et al. 2006, Speed et al. 2007, Bolger et al. 2012, Morrison and Bolger 2012). There are many types of available software, but most are species- (or genera-) specific with very little robust applicability.

Because most phenotypic traits have measurable features (e.g., shapes have area, perimeter, and diam), an approach that can take into account any type of measurement would be more beneficial than existing CAPI software that focuses on species-specific traits. This type of approach requires a quantitative basis, for example through likelihood theory, which has robust applicability and is the basis of a number of statistical processes. Each phenotypic trait of an individual can be thought of as a point along a distribution of phenotypic possibilities within a population, allowing comparison between 2 individuals based on a probability density function. This framework has received almost no attention within the literature, but would be useful to bridge and streamline methods across species and study areas (even for species where shape-based and pattern-based CAPI software are already developed). Furthermore, a more broad CAPI approach would allow for new species to be monitored by photographs for which CAPI methods are not yet developed.

Flinn (2010) recently showed that measurements of antler characteristics (e.g., width of antler spread, diam of antlers) of white-tailed deer (*Odocoileus virginianus*) can allow for individual recognition of males. Other information could also be incorporated into this measurement-based CAPI approach, such as qualitative characteristics (e.g., field observation of age class of animals; Goswami et al. 2007), as well as aspects of trait color (e.g., hue), trait area (area of shapes), and morphological curves (e.g., curve characteristics of an African elephant's [*Loxodonta* spp.] ear, or a bottlenose dolphin's [*Tursiops truncatus*] fin; Araabi et al. 2000). Once morphological characteristics are extracted from a given photograph, these measurement data can be used to develop a probability function of the similarity between 2 photographs. As with other software (Speed et al. 2007, Bolger et al. 2012), this similarity score could then be used to rank pairs of photographs based on their similarity for user specification of a match, or even for identifying thresholds where individuals can be matched.

Our objectives were to 1) develop a CAPI method capable of incorporating phenotypic measurement data and metadata, and multiple sources of error, while allowing users to make the final decision about a potential match; 2) test and identify the error rate (i.e., false-rejection rate) of such an approach; 3) use the approach to analyze and quantify error rates from simulated data sets of morphological measurements with varying population sizes, varying number of measurements and photograph error; and 4) test the approach's ability to develop capture histories for estimating population size of free-ranging bison (*Bison bison*) in Prince Albert National Park, Canada. We also offer discussion and suggestions for implementation, and provide a package to implement our approach in the statistical and data management program R, version 2.15.2 (R Core Team 2012).

METHODS

Explanation of Likelihood-Based Matching

The expression of individual traits is based on interactions between genetics and the environment, leading to phenotypic plasticity and animals with slightly different traits within a population (Via and Lande 1985). The expression of each trait can be thought of as an individual characteristic along a gradient of population characteristics for the species in question. This gradient can be described by a distribution (e.g., Gaussian), which represents the frequency of different trait expressions in the population. Assuming that there are some traits where the gradient of characteristics is variable enough to identify some individuals as different from others, and that each species has multiple traits, it is possible to identify individual animals by measuring a combination of their phenotypic traits.

Variation and error in measuring (from photographs of animals) true trait characteristics, and subsequent matching, comes in 5 forms. First, there is the variation within a population for each trait (population variation). Some traits are more variable than others, allowing for individuals to be more easily identified. Second, there is error in capturing a 3-dimensional animal into a photograph that is 2-dimensional (photograph error). Measurements of traits can become skewed for example, if the photograph is not always taken at the same exact angle. Third, there is error when a human extracts a measurement of a trait from a photograph (measurer error). This error is derived from experience of the person measuring the trait, the pixel size of the photograph, and the complexity of the measurement (e.g., straight line or circle). Fourth, population size (i.e., the no. of unique individuals) can affect the ability to successfully match 2 photographs of the same individual, where the higher the population size, the higher the probability that 2 individuals have the same (or at least indistinguishable) suite of morphological measurements. Finally, the number of morphological measurements that can be extracted from a photograph can affect successful matching, where fewer morphological measurements will lead to more errors during matching.

Based on the aforementioned framework and error sources, we devised the following approach to compare 2 photographs to help decide whether they are the same individual or not. For each pair of photographs (the unknown photograph and a known potential match), we calculated a similarity score by summing the Gaussian likelihood estimates between the measurements of the same traits between the 2 photographs. The similarity score was based on the following equation:

$$\text{Similarity score} = \sum_{i=1}^{n} f(x_n|\mu_n, \sigma_n) \times \omega_n$$

where

$$f(x_n|\mu_n, \sigma_n) = \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left(-\frac{(x_n - \mu_n)^2}{2\sigma_n^2}\right)$$

where n corresponds to the number of measurements the user has specified to compare photographs, f is the probability density function that x_n (measurement value of the known photograph) is the same value as μ_n (measurement value of the focal, unknown photograph of interest) given a Gaussian distribution with standard deviation (SD), σ_n. The SD for each measurement (i.e., photograph error) can be calculated by measuring a sample of multiple photographs of the same individual, and calculating SD of the repeated measurements. We also incorporated a weight, ω_n, for each measurement, where the measurement's likelihood can be increased or decreased by a specified multiplicative amount. In our case, we estimated the measurer error and used this information to weigh the similarity score (see below for estimation details).

For implementation, the user must have a database containing numeric measurements of traits (e.g., ear length, spot diam) extracted from the photographs in question. Then, the user can calculate the similarity score between each unknown photograph and all known photographs ($\geq$1 photograph must be specified as known) in the known database. We used a step-by-step process to find matches, which ensures that each photograph is compared unidirectionally (i.e., photograph A is compared with photograph B, but photograph B is not compared again with photograph A). The approach then provides the user with the best potential matches from which to choose. In our case, we reviewed the unknown photograph and up to 5 of the closest potential matches to the unknown photograph (Fig. 1). The user then specifies whether or not there is a correct match (for more information and further explanation, download package *MatchImage* for use in Program R (R Core Team 2012), available at https://r-forge.r-project.org/R/?group_id=1628).

Identifying Misidentification Error Rate

We estimated the error rate (i.e., false-rejection rate) of our approach using a set of 91 photographs, representing 33 known individual bison. Photographs of the faces of ear-tagged female bison ($n = 28$) and female bison with other distinguishing marks ($n = 5$; e.g., large and identifiable chips in the horn) were taken in the southwest corner of Prince Albert National Park (approx. 600 km^2; 53°44′N, 106°39′W), Canada, during summers of 2008, 2011, and

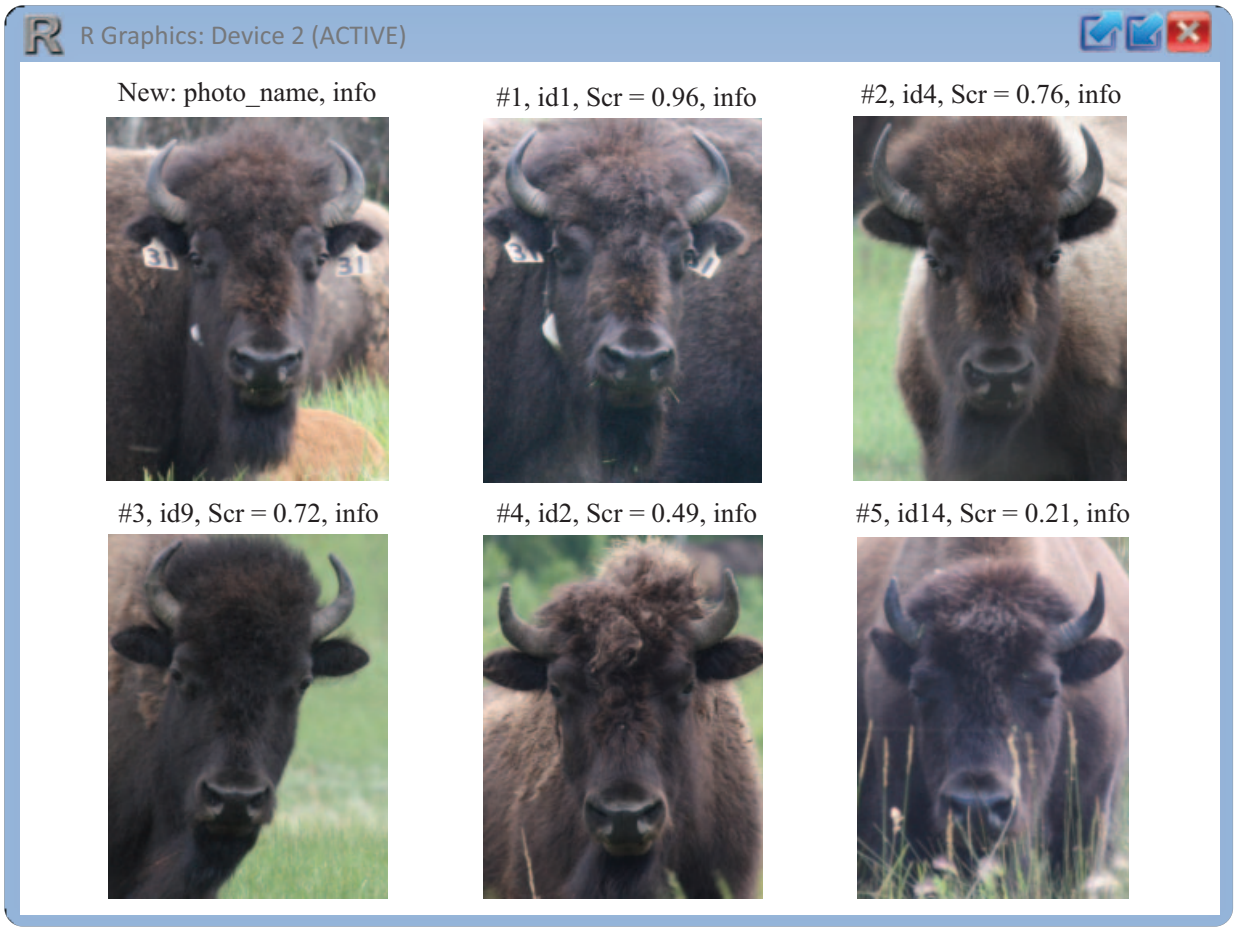

Figure 1. User interface in Program R for selecting a match from 5 ranked potential matches based on the similarity-score calculated between photographs of plains bison in Prince Albert National Park, Canada, 2008, 2011, and 2012. Photograph in upper left-hand corner represents the unknown photograph, with information including file name and user-specified information (i.e., metadata). The other 5 photographs represent ranked potential matches. Information provided for the potential matches includes rank, known id number, similarity score (on a scale from 0 to 1), and the user-provided qualitative information (i.e., metadata). Based on horn shape and texture, Photograph no. 1 is clearly a match with the unknown photograph. This individual also provides an example of an ear-tagged bison (no. 31) used to estimate the misidentification error rate.

2012. The majority of the photographs were taken from a distance of 50–300 m from bison, using an EOS Canon Rebel XS 10.1 megapixel camera with a fixed Canon EF 400-mm lens, and a 2× Canon extender (Canon Canada, Inc., Mississauga, ON, Canada) mounted on a tripod. Some photographs were also taken opportunistically using a handheld digital point-and-shoot camera.

In 1969, plains bison from Elk Island National Park were translocated northeast of Prince Albert National Park. Shortly thereafter, a small group of bison moved into the park and established the current population, which does not intermix with other adjacent bison populations (Bergeson 1992). Prince Albert National Park was located within the mixed-wood boreal forest (Rowe 1972), and had a cool sub-humid continental climate. Prince Albert National Park was characterized by long cold winters (Jan daily $\bar{x} = -19^{\circ}$C) and short warm summers (Jul $\bar{x} = +16^{\circ}$C), with the majority (approx. 250 mm) of the 450 mm of annual precipitation falling as rain between June and September. The landscape of Prince Albert National Park consisted of forest (85%), meadows (10%), and water bodies (5%; Fortin et al. 2003).

We took photographs of bison from the edge of the forest when we observed them in meadows. We only used photographs of high quality, where the individual bison was facing the camera (i.e., both eyes of the bison were visible), both eyes were open (i.e., the eyeball was apparent), and the photograph was crisp (i.e., crisp enough to see texture on the horns). We then cropped out everything except the face (including the horns and beard) of the bison, and manually estimated (in pixels) the following 9 measurements on the face of bison (Fig. 2) using the free photograph management software ImageJ (Abramoff et al. 2004).

M1: Distance between the very tip (the part of the horn that constitutes the longest distance from the base) of each horn.
M2: Distance between the horns at the point where the horns are widest from left to right.
M3 and M4: Diameter of the bison's right (and left) horn when the growth is vertical. This value will be zero if the horn is not long enough to grow vertically. If there is a long section of the horn that is vertical, the measurement is the largest diameter.
M5 and M6: Diameter of the horn at base of the bison's right (and left) horn.
M7 and M8: Distance from the outer base of the bison's right (and left) horn to the tip.
M9: Distance between the eyes, measured from the visible inner edge of the eye socket (i.e., at the point where the eyeball touches skin; Fig. 2).

To minimize the effects of photograph error through varying distances between the photographer and the bison and of the angle of the bison's face (i.e., bison can be looking down or up, or from side to side), we used a ratio method to calculate 9 derived ratios from our 9 measurements. Thus, ratios 1–8 correspond to measurements M1–M8 divided by M9 (the facial measurement with the least variability among measurements), and ratio 9 is based on the ratio between measurement M1 and M2 (Table 1; Fig. 2). These 9 ratios were used to test our database of 91 known photographs of bison.

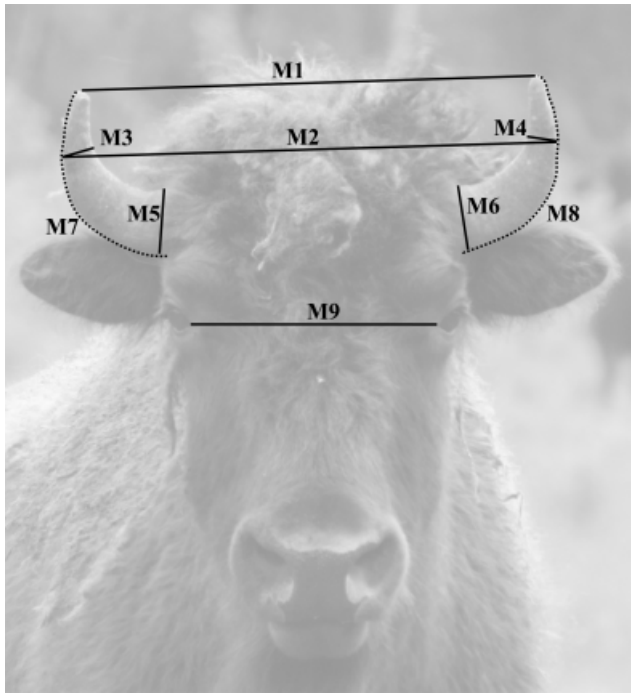

Figure 2. Example photograph of an adult female plains bison taken during summer 2011 in Prince Albert National Park, Canada. Photograph was used to estimate population size through photographic mark–recapture via the likelihood-based photograph identification approach. Black lines (solid and dashed) portray measurements 1 through 9 manually measured in pixels for use in calculating the likelihood-based similarity score between 2 photographs.

We calculated a standardized SD for each ratio, representing photograph error, using a set of 31 known pictures representing 5 different animals (between 3 and 6 photographs for each individual). We calculated means for each of the 9 ratios by individual. We then subtracted the individual's mean ratio from each ratio (i.e., centering each

Table 1. Standard deviations and weights used to calculate the similarity score to identify matches among facial photographs of plains bison (*Bison bison*) in Prince Albert National Park, Canada, in 2011. Standard deviations (photograph error) were calculated from standardized ratios (i.e., ratio minus the individual's ratio mean) from 31 known photographs representing 5 different animals. Weights (measurer error) were calculated by repeatedly measuring 10 photographs of individual bison 5 times each, estimating the standardized standard deviation (i.e., ratio minus the individual's ratio mean), and then adjusting the standard deviations so the smallest SD would have a weight of 1.

Ratio	Measurement	SD	Weight
r1	M1/M9	0.122	1.000
r2	M2/M9	0.097	0.968
r3	M3/M9	0.018	0.206
r4	M4/M9	0.029	0.283
r5	M5/M9	0.034	0.268
r6	M6/M9	0.029	0.302
r7	M7/M9	0.068	0.638
r8	M8/M9	0.117	0.526
r9	M2–M1/M1	0.039	0.999

animal's ratio on zero to standardize), and took the SD of all 31 ratios. We calculated weights for each ratio by repeatedly measuring the same exact photograph of an individual 5 times (i.e., measurer error). We did this for 10 random photographs. We again standardized each ratio around zero and calculated the SD of all ratios. To turn these SDs into a weight, we scaled them so that the ratio with the smallest SD would have a weight of 1, and the other SDs would have a smaller weight based on their size.

When we made a decision whether or not there was a match between an unknown photograph and 1 of the potential known photographs, we assumed that we always correctly assessed whether or not there was a match (i.e., false acceptance rate was zero). For bison this assumption was met, particularly when the texture of individual horns was clearly visible in the photograph (as with our high-quality photograph selection). Each individual has variation in the color, growth rings, chips, and angles of the horn that are all easily distinguishable when comparing high-quality photographs. Therefore, we only estimated the false-rejection rate (i.e., the probability of failing to match 2 photographs of the same individual) of the approach, where there was a match within the known database but the approach failed to provide the option in the top 5 potential matches. We calculated the false-rejection rate by dividing the number of photographs that were incorrectly matched by the number of photographs within the test database.

Data Simulation

To quantify how misidentification error rates vary based on sources of error and variation, we simulated 117 data sets with varying numbers of unique individuals (i.e., from 100 to 2,600 in increments of 500), morphological measurements, and photograph error. The morphological measurements were a theoretically derived linear measurement that could be extracted from a photograph, and each data set contained 2 sets of measurements for each individual. Each simulated individual was given 5, 10, or 15 measurements, with values drawn from an arbitrary mean and SD. Then, a second set of traits for the same individual was generated from a random deviate of the animal's measurement value based on a SD of 10% (low), 30% (medium), or 50% (high) of the arbitrary population SD. We chose these SD values because they reflected the ratio between population variation and photograph error in our bison measurements.

To calculate error rates, we calculated similarity scores using our approach, and ranked potential matches. We specified no weights for each measurement for the simulation. Finally, for each of the 117 data sets, we calculated the false-rejection rate as the percentage of photographs that did not fall within the top 5 ranked potential photographs, but should have.

Application to Estimate Population Size

The bison range in Prince Albert National Park during summer was a subset of the winter range and encompassed an area of approximately 200 km^2 (J. A. Merkle and D. Fortin, unpublished data). During summer 2011, we developed 4 circuits (i.e., trail networks containing meadows frequently used by bison) within this area where we searched for bison. Each circuit was completed 2 times/week for 10 consecutive weeks between 30 May and 1 August. While surveying the circuits we searched for bison groups, and when found, we opportunistically took photographs of the faces of as many adult bison ($\geq$3 yr old) as possible. We extracted measurements from each photograph as explained above and in Figure 2. We then identified matches using our approach within the photograph database for males and females separately, using the SDs and weights in Table 1.

We developed capture histories based on 10 sampling events, corresponding to each week during the summer. Using Program MARK (White and Burnham 1999), we estimated adult population size for males and females with a Huggins closed-capture model (Huggins 1989, 1991). To approximate individual heterogeneity (e.g., some individuals are always more likely to be captured because they scan their surroundings more often; hence, they are easier to photograph), we extended our model to include 2 mixtures for probability of capture, where the model allowed individual animals to have 1 of 2 different capture probabilities (Norris and Pollock 1996, Pledger 2000). Furthermore, we incorporated misidentification error by fixing a parameter corresponding to the probability of correctly identifying the individual (Lukacs and Burnham 2005). Using this method to incorporate misidentification error, we assumed that individuals in the final database with $\geq$1 matches were correctly identified; thus, only animals without a match were subject to misidentification.

To assess whether our estimates of the number of adult bison were plausible, we compared our results to minimum-count aerial surveys of bison in March 2011 (i.e., prior to the spring birth pulse) and March 2012 (S. G. Cherry, Parks Canada, unpublished data). Surveys were conducted by helicopter along linear transects (between 1 km and 1.5 km wide covering approx. 900 km^2) across the bison range, including the area where photographs were taken during summer. We did not estimate the number of calves and juveniles (i.e., only adults) using photographs; therefore, we also noted the number of calves, juveniles (i.e., yearlings and 2-yr-olds), and adult females for each group seen during summer 2011 ($n = 46$). Using these counts, we estimated calf-, juvenile-, and combined calf- and juvenile-to-female ratios using methods for estimating age ratios for repeated surveys in Skalski et al. (2005). Then, using a simple bootstrapping procedure (Efron and Tibshirani 1986) where we resampled each estimate (male estimate, female estimate, and juvenile- and calf-to-female ratios) 1,000 times based on a normal distribution, we calculated total population size of bison to compare to the March 2012 survey, and a population size excluding calves to compare to the March 2011 survey.

We expected the aerial survey minimum counts to be >70% of the population size estimates. We used the value of 70% because it is the mean sightability of another large ungulate, moose (*Alces alces*; Timmermann 1993), which is often found in similar land-cover types as are bison in Prince Albert National Park. However, we assumed higher sightability

than moose because bison in our study area spend the majority of their time in meadows (Fortin et al. 2003) where they are more easily sighted from a helicopter.

RESULTS

Using our photograph matching approach, photograph error, or the standardized SD for bison facial measurements, was between 0.018 and 0.122 (Table 1). Weights, or the measurer error, varied from 1 to 0.206 (Table 1) corresponding to the least variable ratios, r1 and r9, to the most variable, r3, respectively (Table 1).

Of the 91 known photographs in our test data set, 5 were not matched correctly, and were given their own separate identifications. As a result, the approach estimated 38 unique individuals instead of 33 in the test data set of photographs. The false-rejection rate for the known bison data set was 5.5%.

Based on simulated data sets and subsequent matching, false-rejection rates varied by the number of unique individuals in the population, the number of morphological measurements, and the amount of photograph error (Fig. 3). In general, lower false-rejection rates were observed in smaller populations, where there were more traits extracted from photographs, and those traits were extracted with lower photograph error. False-rejection rates were near zero for all population sizes when there were $\geq$10 traits to measure, and the photograph error was low (i.e., an individual SD that is 10% of the population SD). False-rejection error rates were >20% for populations consisting of >500 individuals, when there were $\leq$5 traits to measure, and when the photograph error was medium or high (SD that is 30% or 50% that of the population SD; Fig. 3).

We took 287 high-quality photographs (176 F, 111 M) during summer of 2011, with an average of 17.6 (SD = 10.6)/week. Using our matching approach, we identified 74 female and 39 male unique adult bison. Estimated population size from mark–recapture models was 153 adult bison, with a 95% confidence interval of 83.5–133.7 adult females, and 36.2–62.2 adult males (Table 2).

Mean (±SE) calf-, juvenile-, and combined calf- and juvenile-to-female ratios for summer 2011 were 0.24 (±0.01), 0.22 (±0.01), and 0.46 (±0.03), respectively. Using bootstrapping, the total population size of bison, including juveniles and calves, was 202 (95% CI = 171.6–231.4). For comparison to 2011 minimum-count aerial surveys, the population size of bison, excluding calves, was 176 (95%

Table 2. Huggins closed-capture model parameter estimates with heterogeneity and misidentification. Parameters include probability of initial capture (p_i), 2 recapture probabilities based on 2 mixtures of heterogeneity (p_1 and p_2), and estimated 2011 population size (N) for the number of adult (≥3 yr old) male and female bison in Prince Albert National Park, Canada.

	Females		Males	
Parameter	**Estimate**	**SE**	**Estimate**	**SE**
p_i	0.019	0.020	0.033	0.748
p_1	0.628	0.221	0.262	1.294
p_2	0.101	0.018	0.137	0.071
N	105.642	12.751	47.415	6.588

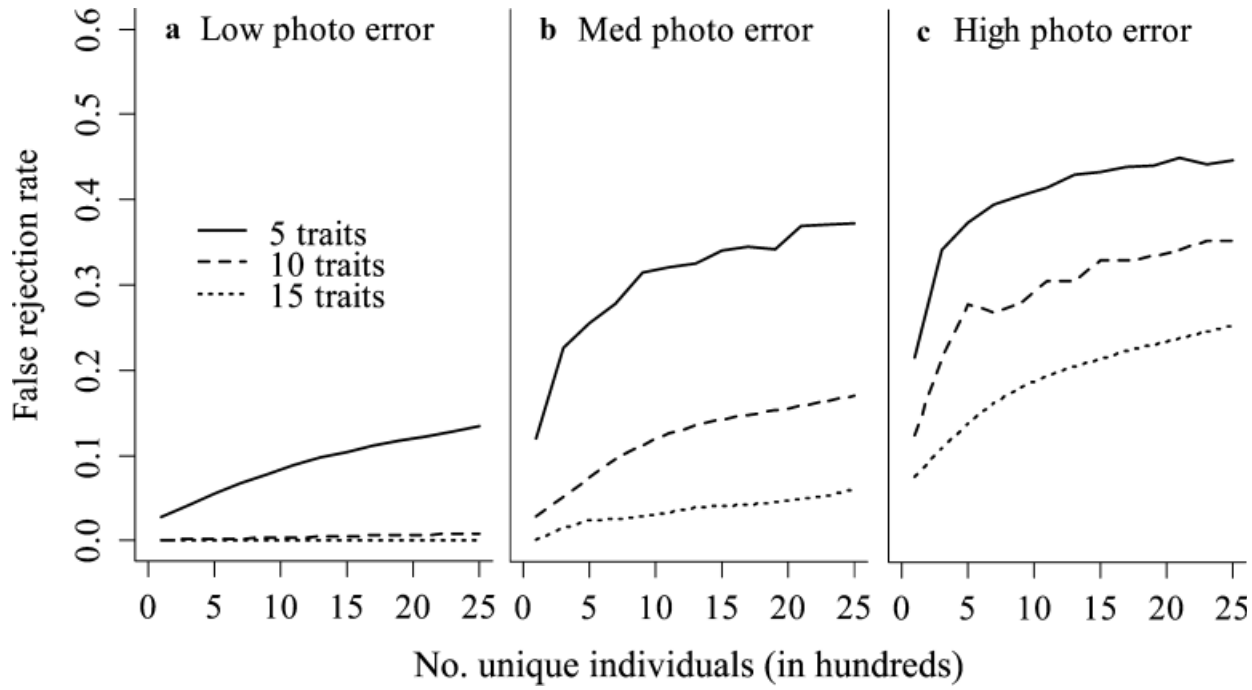

Figure 3. Variation in misidentification error rates (false-rejection rate) obtained when identifying individual matches based on the likelihood-based similarity score, for simulated population sizes (from 100 to 2,600 simulated individuals) based on the number of traits measured (5, 10, or 15) and the photograph error within those traits (low, medium, high). Each simulated population contained 2 sets of measurements for each individual, where the second individual's measurements were generated based on varying photograph error. For a given range of measured values of a trait, low, medium, and high photograph error corresponds to a standard deviation of 10%, 30%, and 50% of the population SD.

CI = 147.3–205.3). Minimum counts from aerial surveys for bison were 142 for March 2011, and 177 for March 2012.

DISCUSSION

We outline a widely applicable CAPI approach that can be used to find matching photographs of individual animals based on measurements of user-specified morphological traits. We found that our bison data set and most simulated data sets had similar, and sometimes lower, false-rejection rates than those reported by other CAPI methods (see Bolger et al. 2012 for a review of error rates). For example, using a curve-matching method for large marine mammals (e.g., bottlenose dolphin), users must inspect the top 23–30 images to obtain an error rate near 0.25 (Gope et al. 2005). Additionally, using fingerprint identification technology for zebra (*Equus quagga burchelli*) stripes, users can expect a 0.20 false-rejection rate (Foster et al. 2006). In contrast, CAPI methods such as the Scale Invariant Feature Transform operator for matching the patterns of the Masai giraffe (*Giraffa camelopardalis*), have error rates <0.01 (Bolger et al. 2012). Decreasing the error rate using our approach will be based on minimizing photograph error and measurer error, while maximizing the number of highly variable phenotypic traits used.

Although the comparison with aerial surveys was not a true validation, our approach provided the first plausible estimate of the bison population in Prince Albert National Park. As expected, the counts from the aerial surveys were >70% (i.e., 80% and 87% for 2011 and 2012, respectively) of estimated population sizes. Further, the minimum counts from the aerial surveys were both within 6 individuals of the lower 95% CI of population estimates. The implications of this population estimate may be substantial. The viability of small populations can be uncertain because they are more susceptible to demographic, environmental, and genetic stochasticity, along with natural catastrophes (Shaffer 1981) such as anthrax outbreaks for bison (Shury et al. 2009). Further, bison populations of <400 individuals provide only a small contribution to the ecological restoration of bison in North America (Sanderson et al. 2008). Given that this population is the only free-ranging population of plains bison within their historical range in Canada, a total population of <231 individuals including only 86–134 breeding females may not be sustainable.

Although we used horn measurements to identify individuals, our approach can incorporate measurements of phenotypic traits from other species. For example, other ungulate species have unique horns (e.g., bighorn sheep [*Ovis canadensis*]), and species such as black bears (*Ursus americanus*) and wolverines (*Gulo gulo*) have chest patches that can be measured. Indeed, Higashide et al. (2012) found that most Asiatic black bears (*U. thibetanus*) can be individually identified based on their chest marks. Although that study focused on captive bears, data extracted from chest mark photographs could be easily incorporated into our approach, and used to generate capture histories of wild individuals. Further, although we took photographs by approaching bison, creative methods to take remote photographs of other species (e.g., Bergeron 2007, Ngoprasert et al. 2012) will likely provide photographs with the correct angle. For example, Ngoprasert et al. (2012) designed a baited, 3-camera arrangement, which attracted Asiatic black bears and sun bears (*Helarctos malayanus*) to stand on 2 legs while facing a camera. Phenotypic measurements extracted from strategically angled photographs, combined with our approach to find matches, will likely provide novel non-invasive monitoring techniques. We do acknowledge, however, that difficult-to-photograph species will never be candidates for photographic monitoring in some circumstances.

Based on our simulation (Fig. 3), researchers can conduct small pilot studies to test whether there are enough traits and low enough photograph error in the species in question to identify individuals. The false-rejection rates estimated were conservative, because we did not specify a weight (i.e., other possible sources of error); furthermore, the databases we generated consisted of 2 photographs of each individual. For mark–recapture studies, it is not necessary to recapture every individual; thus, the number of photographs in a database could be smaller than what we simulated. A careful pilot study along with a power analysis could provide an idea of the effort needed to achieve parameter estimates of desired precision. Researchers and conservationists should try to maximize the number of measurable traits per animal and minimize measurer and photograph error to minimize false-rejection rates, thereby narrowing confidence intervals on ecological parameter estimates.

Our approach is, in some cases, advantageous over other CAPI methods for 3 reasons. First, our approach is based on likelihood theory, which provides well-accepted probability functions to identify the probability that one measurement is the same as another. In its current form, the similarity score is completely based on a Gaussian distribution. However, with slight changes, other distributions could be incorporated as well. For example, if a type of phenotypic angle is incorporated into our approach (e.g., curve characteristics of an antler or a fin), a circular distribution may be more suitable (e.g., wrapped Cauchy distribution). Second, because our approach can evaluate any type of phenotypic measurement, users may transform measurement data to suit certain needs. For example, many phenotypic traits change over time (e.g., horns grow; Yoshizaki et al. 2009), which can be problematic for static matching methods such as explained in Ardovini et al. (2008) for identifying African elephants by irregularities in their ears. Slight modifications prior to using our approach could provide a measurement-based transformation to meet the needs of changing traits. Third, most methods do not allow for multiple types of error to be incorporated into the matching algorithm (Ardovini et al. 2008, Anderson et al. 2010). With our framework, the user can not only adjust based on photograph error (error inherent to photograph use), but also other error sources, in the form of weights. For bison, we weighted the measurements based on measurer error. Other weights, such as difficult-to-obtain measurements and qualitative-based measurements (e.g., age class classification) can be incorpo-

rated, decreasing overall misidentification error rates of the approach.

We assumed that the user always correctly assesses whether or not there was a match (i.e., false acceptance rate was zero). For high-quality photographs of bison this assumption was met, but for some species it may more difficult to always correctly match 2 photographs of the same individual. For example, when the only phenotypic characteristics to use for distinguishing 2 photographs include color hue or saturation, these differences can be difficult for humans to see. As noted by other authors, using trained and qualified users may decrease the false acceptance rate (Huffard et al. 2008, Schofield et al. 2008, Waite and Mellish 2009), but we suggest that users use a test data set of known individuals to estimate false acceptance rate before using results from our approach to estimate population parameters. Users could test their false acceptance error by taking $\geq$2 photographs of the same individual, removing the background or other distinguishing characteristics within the photographs using photograph manipulation software, and testing their ability to match them among photographs of different individuals.

Based on our analysis of known photographs, we were able to incorporate misidentification error into our mark–recapture model of population size. As our approach is used for other applications, we assume that a variety of different misidentification errors will arise. In some cases there will be false-positive and false-negative error rates, as well as complicated misidentification patterns. For example, not all false-positive errors will result in another, single unique identification, but may sometimes be matched with different individuals. In other words, 2 identifications may represent the same animal, and both identifications may be associated with $>$1 photograph. These misidentification scenarios will require more sophisticated methods to modify models for estimating population dynamics. For example, Link et al. (2010) provides a Bayesian framework to handle a range of misidentification errors, where mark–recapture models incorporate a latent multinomial random variable to take into account the errors. The methods available are still developing however (Yoshizaki et al. 2009, Link et al. 2010, Morrison et al. 2011), and do not account for the wide range of misidentification possibilities in mark–recapture models. We foresee an important avenue of research that must be conducted to clarify and simplify how misidentification is taken into account in photographic mark–recapture studies.

MANAGEMENT IMPLICATIONS

The approach outlined in this study provides a new opportunity for managers and researchers in applied wildlife ecology. Non-invasive population monitoring is likely possible for species and in study areas where previous CAPI methods are not adequate. We show that identifying error and incorporating it into our approach produces plausible population estimates of free-ranging bison using mark–recapture models with misidentification. Over time, these estimates can be used to build more integrated population models in which population parameters such as age-specific survival and population growth can be estimated. Furthermore, analyzing photographs using our approach provides capture histories for use in answering other ecological questions about space use, behavior, and phenotypic diversity. If researchers and managers want to non-invasively monitor a wildlife population, or have a database of photographs with potential for photograph matching, our simple-yet-broad approach provides a basis for development of such projects.

ACKNOWLEDGMENTS

Funding for this study was provided by Parks Canada Species at Risks Recovery Action and Education Fund (a program supported by the National Strategy for the Protection of Species at Risk), the Natural Sciences and Engineering Research Council of Canada, the Canadian Foundation for Innovation, and Université Laval. We thank Parks Canada for use of minimum count survey data for bison in Prince Albert National Park. We are grateful for the insight provided by S. Cherry and C. Ricaud during the development of the manuscript.

LITERATURE CITED

Abramoff, M. D., P. J. Magalhaes, and S. J. Ram. 2004. Image processing with ImageJ. Biophotonics International 11:36–42.

Anderson, C. J. R., N. D. Lobo, J. D. Roth, and J. M. Waterman. 2010. Computer-aided photo-identification system with an application to polar bears based on whisker spot patterns. Journal of Mammalogy 91:1350–1359.

Araabi, B. N., N. Kehtarnavaz, T. McKinney, G. Hillman, and B. Wursig. 2000. A string matching computer-assisted system for dolphin photo-identification. Annals of Biomedical Engineering 28:1269–1279.

Ardovini, A., L. Cinque, and E. Sangineto. 2008. Identifying elephant photos by multi-curve matching. Pattern Recognition 41:1867–1877.

Arzoumanian, Z., J. Holmberg, and B. Norman. 2005. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks Rhincodon typus. Journal of Applied Ecology 42:999–1011.

Berger, J. 2012. Estimation of body-size traits by photogrammetry in large mammals to inform conservation. Conservation Biology 26:769–777.

Bergeron, P. 2007. Parallel lasers for remote measurements of morphological traits. Journal of Wildlife Management 71:289–292.

Bergeson, D. 1992. A comparative assessment of management problems associated with the free-roaming bison in Prince Albert National Park. Thesis, University of Manitoba, Winnipeg, Canada.

Bolger, D. T., T. A. Morrison, B. Vance, D. Lee, and H. Farid. 2012. A computer-assisted system for photographic mark–recapture analysis. Methods in Ecology and Evolution 3:813–822.

Cattet, M., J. Boulanger, G. Stenhouse, R. A. Powell, and M. L. Reynolds-Hogland. 2008. An evaluation of long-term capture effects in ursids: implications for wildlife welfare and research. Journal of Mammalogy 89:973–990.

Cutler, T. L., and E. S. Don. 1999. Using remote photography in wildlife ecology: a review. Wildlife Society Bulletin 27:571–581.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1:54–75.

Flinn, J. J. 2010. Accuracy of estimating age and antler size of photographed deer. Mississippi State University, Mississippi State, USA.

Fortin, D., J. M. Fryxell, L. O'Brodovich, and D. Frandsen. 2003. Foraging ecology of bison at the landscape and plant community levels: the applicability of energy maximization principles. Oecologia 134:219–227.

Foster, G., H. Krijger, and S. Bangay. 2006. Zebra fingerprints: towards a computer- aided identification system for individual zebra, African Journal of Ecology 45, 225–227.

Gamble, L., S. Ravela, and K. McGarigal. 2008. Multi-scale features for identifying individuals in large biological databases: an application of pattern recognition technology to the marbled salamander *Ambystoma opacum*. Journal of Applied Ecology 45:170–180.

Gope, C., N. Kehtarnavaz, G. Hillman, and B. Würsig. 2005. An affine invariant curve matching method for photo-identification of marine mammals. Pattern Recognition 38:125–132.
Gordon, H. R., B. C. Bock, M. B. Gordon, and A. S. Rand. 1988. Techniques for identifying individual lizards at a distance reveal influences of handling. Copeia 1988:905–913.
Goswami, V. R., M. D. Madhusudan, and K. U. Karanth. 2007. Application of photographic capture–recapture modelling to estimate demographic parameters for male Asian elephants. Animal Conservation 10:391–399.
Higashide, D., S. Miura, and H. Miguchi. 2012. Are chest marks unique to Asiatic black bear individuals? Journal of Zoology 288:199–206.
Huffard, C. L., R. L. Caldwell, N. DeLoach, D. W. Gentry, P. Humann, B. MacDonald, B. Moore, R. Ross, T. Uno, and S. Wong. 2008. Individually unique body color patterns in octopus (*Wunderpus photogenicus*) allow for photoidentification. PLoS ONE 3:e3732.
Huggins, R. M. 1989. On the statistical analysis of capture experiments. Biometrika 76:133–140.
Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47:725–732.
Kelly, M. J. 2001. Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs. Journal of Mammalogy 82:440–449.
Krausman, P. R., V. C. Bleich, J. W. Cain, T. R. Stephenson, D. W. DeYoung, P. W. McGrath, P. K. Swift, B. M. Pierce, and B. D. Jansen. 2004. From the field: neck lesions in ungulates from collars incorporating satellite technology. Wildlife Society Bulletin 32:987–991.
Link, W. A., J. Yoshizaki, L. L. Bailey, and K. H. Pollock. 2010. Uncovering a latent multinomial: analysis of mark–recapture data with misidentification. Biometrics 66:178–185.
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.
Meekan, M. G., C. J. A. Bradshaw, M. Press, C. McLean, A. Richards, S. Quasnichka, and J. G. Taylor. 2006. Population size and structure of whale sharks *Rhincodon typus* at Ningaloo Reef, western Australia. Marine Ecology—Progress Series 319:275–285.
Morrison, T. A., and D. T. Bolger. 2012. Wet season range fidelity in a tropical migratory ungulate. Journal of Animal Ecology 81:543–552.
Morrison, T., J. Yoshizaki, J. D. Nichols, and D. Bolger. 2011. Estimating survival in photographic capture–recapture studies: overcoming misidentification error. Methods in Ecology and Evolution 2:454–463.
Ngoprasert, D., D. H. Reed, R. Steinmetz, and G. A. Gale. 2012. Density estimation of Asian bears using photographic capture–recapture sampling based on chest marks. Ursus 23:117–133.
Nichols, J. D. 1992. Capture–recapture models. BioScience 42:94–102.
Nietfeld, M. T., M. W. Barret, and N. Silvy. 1994. Wildlife marking techniques. Pages 140–168 *in* T. A. Bookhout, editor. Research and management techniques for wildlife and habitats. The Wildlife Society, Bethesda, Maryland, USA.
Norris, J. L., and K. H. Pollock. 1996. Nonparametric MLE under two closed capture recapture models with heterogeneity. Biometrics 52:639–649.
Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56:434–442.
Powell, R. A., and G. Proulx. 2003. Trapping and marking terrestrial mammals for research: integrating ethics, performance criteria, techniques, and common sense. ILAR Journal 44:259–276.
R Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rowe, J. S. 1972. Forest regions of Canada. Canadian Forestry Service, Ottawa, Ontario, Canada.
Sacchi, R., S. Scali, D. Pellitteri-Rosa, F. Pupin, A. Gentilli, S. Tettamanti, L. Cavigioli, L. Racina, V. Maiocchi, P. Galeotti, and M. Fasola. 2010. Photographic identification in reptiles: a matter of scales. Amphibia-Reptilia 31:489–502.
Sanderson, E. W., K. H. Redforda, B. Weber, K. Aune, D. Baldes, J. Berger, D. Carter, C. Curtin, J. Derr, S. Dobrott, E. Fearn, C. Fleener, S. Forrest, C. Gerlach, C. Gates, J. E. Gross, P. Gogan, S. Grassel, J. A. Hilty, M. Jensen, K. Kunkel, D. Lammers, R. List, K. Minkowski, T. Olson, C. Pague, P. B. Robertson, and B. Stephensom. 2008. The ecological future of the North American bison: conceiving long-term, large-scale conservation of wildlife. Conservation Biology 22:252–266.
Schofield, G., K. A. Katselidis, P. Dimopoulos, and J. D. Pantis. 2008. Investigating the viability of photo-identification as an objective tool to study endangered sea turtle populations. Journal of Experimental Marine Biology and Ecology 360:103–108.
Shaffer, M. L. 1981. Minimum population sizes for species conservation. BioScience 31:131–134.
Shury, T. K., D. Frandsen, and L. O'Brodovich. 2009. Anthrax in free-ranging bison in the Prince Albert National Park area of Saskatchewan in 2008. Canadian Veterinary Journal 50:152.
Skalski, J. R., K. E. Ryding, and J. J. Millspaugh. 2005. Wildlife demography: analysis of sex, age, and count data. Elsevier Academic Press, Burlington, Massachusetts, USA.
Speed, C. W., M. G. Meekan, and C. J. A. Bradshaw. 2007. Spot the match —wildlife photo-identification using information theory. Frontiers in Zoology 4:2.
Timmermann, H. 1993. Use of aerial surveys for estimating and monitoring moose populations—a review. Alces 29:35–46.
Van Tienhoven, A. M., J. E. Den Hartog, R. A. Reijns, and V. M. Peddemors. 2007. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark *Carcharias taurus*. Journal of Applied Ecology 44:273–280.
Via, S., and R. Lande. 1985. Genotype–environment interaction and the evolution of phenotypic plasticity. Evolution 39:505–522.
Waite, J. N., and J.-A. E. Mellish. 2009. Inter- and intra-researcher variation in measurement of morphometrics on Steller sea lions (*Eumetopias jubatus*). Polar Biology 32:1221–1225.
Waits, L. P., and D. Paetkau. 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. Journal of Wildlife Management 69:1419–1433.
White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46:S120–S139.
Wilson, B., P. M. Thompson, P. S. Hammond, K. Grellier, and C. A. Sanders-Reed. 2003. Use of photo-identification data to quantify mother–calf association patterns in bottlenose dolphins. Canadian Journal of Zoology 81:1421–1427.
Yoshizaki, J., K. H. Pollock, C. Brownie, and R. A. Webster. 2009. Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. Ecology 90:3–9.

Associate Editor: Fulbright.