NEGATIVE BINOMIAL MODELS FOR ABUNDANCE ESTIMATION OF MULTIPLE CLOSED POPULATIONS

MARK S. BOYCE,[1] Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada
DARRYL I. MACKENZIE, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA
BRYAN F. J. MANLY, Western EcoSystems Technology Inc., 2003 Central Avenue, Cheyenne, WY 82001, USA
MARK A. HAROLDSON, Interagency Grizzly Bear Study Team, Forestry Sciences Laboratory, Montana State University, Bozeman, MT 59717, USA
DAVID MOODY, Wyoming Game and Fish Department, 260 Buena Vista, Lander, WY 82520, USA

Abstract: Counts of uniquely identified individuals in a population offer opportunities to estimate abundance. However, for various reasons such counts may be burdened by heterogeneity in the probability of being detected. Theoretical arguments and empirical evidence demonstrate that the negative binomial distribution (NBD) is a useful characterization for counts from biological populations with heterogeneity. We propose a method that focuses on estimating multiple populations by simultaneously using a suite of models derived from the NBD. We used this approach to estimate the number of female grizzly bears (*Ursus arctos*) with cubs-of-the-year in the Yellowstone ecosystem, for each year, 1986–1998. Akaike's Information Criteria (AIC) indicated that a negative binomial model with a constant level of heterogeneity across all years was best for characterizing the sighting frequencies of female grizzly bears. A lack-of-fit test indicated the model adequately described the collected data. Bootstrap techniques were used to estimate standard errors and 95% confidence intervals. We provide a Monte Carlo technique, which confirms that the Yellowstone ecosystem grizzly bear population increased during the period 1986–1998.

JOURNAL OF WILDLIFE MANAGEMENT 65(3):498–509

Key words: AIC, bootstrap, grizzly bear, heterogeneity, maximum likelihood, model selection, Monte Carlo simulations, negative binomial distribution, Poisson distribution, population estimation, *Ursus arctos*, Yellowstone National Park.

Estimating population size continues to be a problem that challenges ecological studies. Many animals are cryptic or elusive, and thus require indirect means of enumeration. New techniques using DNA microsatellites offer exciting promise for population estimation because individuals can be identified from hair, blood, feces, or tissue samples (Paetkau and Strobeck 1994, Palsbøll et al. 1997, Mills et al. 2000). Other methods for identifying individuals include horn and ear shape and scars in the great Indian one-horned rhinoceros (*Rhinoceros unicornis*; Laurie 1978); coloration patterns in the fur of lions (*Panthera leo*; Bertram 1978), African wild dogs (*Lycaon pictus*; Frame et al. 1979), and cheetahs (*Acinonyx jubatus*; Caro and Durant 1991); variation in pigment and scars in the dorsal fins and backs of killer whales (*Orcinus orca*; Bigg 1982, Hammond et al. 1990); and geographically separated female grizzly bears with a variable number of cubs-of-the-year (Knight et al. 1995). Still, identifying each individual in a population typically requires an extraordinary sampling effort.

We focus on situations where sightings of individuals can occur at any time during data collection rather than at discrete trapping occasions and where several population sizes are to be estimated (e.g., the size of a population during each year for several years). A method is available to estimate a single population size from sightings of unique individuals (Craig 1953, Lewontin and Prout 1956, Darroch 1958, Boyce et al. 2001). However, this model assumes that all individuals are equally likely to be counted, and thus it does not allow for heterogeneity in the probability of detection among animals, which is a major consideration in many studies (Eberhardt 1969). Other methods for estimation are also available that do allow for heterogeneity (Burnham and Overton 1978, Otis et al. 1978, Chao 1987, Chao and Lee 1992, Pledger 2000), but again they were developed for discrete sampling occasions and only for estimating 1 population size at a time.

When heterogeneity exists in the sighting probabilities, it is reasonable to consider the negative binomial distribution (NBD) as a model for the counts of the number of sightings of individual animals. A justification for using NBD comes from the fact that if the number of sightings per animal has a Poisson distribution and the mean

[1] E-mail: boyce@ualberta.ca

of the Poisson distribution for each animal is a random value from a gamma distribution, then the resulting distribution for the frequency of sightings will be a NBD (Kendall and Stuart 1969). The Poisson assumption is reasonable if the probability of a sighting during any small time interval is small and the numbers of sightings in any 2 disjoint time intervals are independent, while the gamma distribution is a very flexible unimodal distribution that enables the values of the Poisson means to be either very skewed, symmetric, or constant. These assumptions seem plausible and reasonable for many sampling situations. Furthermore, the NBD often is an adequate description of the statistical distribution for heterogeneous (aggregated) ecological data (Tanton 1965, Wilbur and Landwehr 1974, Hilborn and Mangel 1997, Krebs 1999, Southwood and Henderson 2000).

Applying the NBD to single populations has been shown to result in unreasonable estimates (Wilbur and Landwehr 1974, Seber 1982:496). However, the situation is different if the estimation of several population sizes is considered at the same time because then it is possible for the estimation of 1 population size to be aided by data that were primarily collected for the estimation of the other population sizes.

We assume that the aim of an analysis is to estimate the size of several closed populations simultaneously. We propose a suite of 6 models from the NBD family for this purpose, some of which allow for heterogeneity, and use the Akaike Information Criteria (AIC) as an objective basis to select 1 of these models. We use bootstrap methods for estimating standard errors and constructing confidence intervals; simulation is used to evaluate the properties of these procedures. We envisage that these methods might be appropriate where distinct populations or age-sex groups have been sampled in different years or at different locations.

We offer an example estimating population sizes of female grizzly bears with cubs-of-the-year from 1986 to 1998 using the sighting frequencies of unique family groups (Knight et al. 1995) and propose a Monte Carlo test for detecting trend in population size. Observations of female grizzly bears with cubs-of-the-year are compiled during each field season in the Yellowstone ecosystem with unique individuals identified by the number of cubs, geographic location, and characteristics of family groups. However, variation in weather conditions for conducting surveys, funding for flight time, and availability of field personnel create variation in search effort (Mattson 1997). Furthermore, variation occurs from year to year in the detectability of bears because bears are more active during years when food is scarce, and therefore they are more likely to be detected (Picton et al. 1986, Noyce and Garshelis 1997). We suggest models that accommodate these among-year influences on counts. These sources of heterogeneity plus potential heterogeneity in sighting probabilities due to variation in the relative accessibility of individual animal's home ranges (Wilbur and Landwehr 1974, Otis et al. 1978) and the opportunistic nature of the sampling justify our application of the NBD for estimating the population size of grizzly bears.

METHODS

Grizzly Bear Counts

We compiled observations of adult female grizzly bears with cubs-of-the-year according to field protocols described by Knight et al. (1995). Identification of unique family groups was based on 3 criteria: distance between sightings, timing of sightings, and physical characteristics of individuals in each group. Unique family groups were identifiable by the size and number of cubs in the litter, as well as geographic and temporal separation.

Field surveys were conducted by selected personnel who were well trained in the identification of bears. Observations of family groups from both ground and aerial surveys were compiled throughout each field season, with most observations being recorded during May through September. Multiple observations of the same bear by different observers on the same day were counted as a single observation. This rule was developed to reduce the high frequency of sightings that might occur when a bear was located in an accessible area, such as near a road, where multiple observers might see the same bear. When multiple observations were made of a bear at the same location by the same observer, we used only the first observation. Thus, if a bear was revisited by the same observer at the same location, as might occur if a bear was feeding on an ungulate carcass, we recorded only the first observation at that location. Again, this recording rule was developed to reduce heterogeneity in sighting frequency.

In addition to accumulating ground-based observations of female grizzly bears with cubs-of-the-year throughout the entire season, these counts were supplemented by aerial surveys. Aerial

surveys were coordinated to minimize double-counting caused by movement of bears between areas surveyed by different observers. All bear management units in the Yellowstone ecosystem were surveyed during each field season with no overlap between field survey areas. Areas not covered by the Wyoming Game and Fish Department or National Park Service personnel were surveyed by the Interagency Grizzly Bear Study Team (IGBST). When weather conditions and funds permitted, as many as 3 flights over each of 18 bear management units were flown June through August each year.

Occasionally, 2 sightings of females with the same number of cubs in the same year might have adjacent home ranges within 30 km. Whenever there was a question regarding the uniqueness of these individuals, they were not considered distinct bears. Unique identification was not possible unless they were sighted at approximately the same time on the same day (Knight et al. 1995). These sightings were assigned to the most proximal bear and considered a duplicate sighting. Such errors were more likely in areas of high bear density where infrequent observations occurred making estimates conservative. In much of Yellowstone National Park, uniqueness was more accurately determined by high frequency of observations.

Statistical Background for the Abundance Estimator

Assuming that the number of sightings of an individual comes from a Poisson distribution allows sightings to occur at any time during data collection and does not require a discrete number of trapping occasions as in many mark–recapture studies. In such studies, the number of sightings per animal would follow a binomial distribution; however, the Poisson distribution could be used as an approximation if the number of trapping occasions is large, with a small probability of a sighting occurring within each trapping period.

The properties of the NBD are well known and a variety of formulations exist for the probability function. We follow Kotz et al. (1988) in using

$$\Pr(X = x) = \frac{\Gamma(k + x)P^x(1 + P)^{-(k+x)}}{\Gamma(k)x!} \quad (1)$$

for $x = 0, 1, 2, \ldots$ where $k, P > 0$ with mean kP and variance $kP(1 + P)$, where k and P are the shape and scale parameters of the gamma distribution for the Poisson means, and X is the number of times an animal is sighted. We refer to this distribution with the notation NBD(k, P). The level of heterogeneity can be measured by the coefficient of variation (CV) for the Poisson means, which is $1/\sqrt{(k)}$. The Poisson distribution is the limiting form of the NBD when k tends to infinity and P tends to zero, as the CV tends to zero while the mean and variance tend to equality. Thus, this represents the case where there is no heterogeneity. The probability function of the Poisson distribution with mean λ is

$$\Pr(X = x) = e^{-\lambda}.\lambda^x/x! \quad (2)$$

for $x = 0, 1, 2, \ldots$, where $\lambda > 0$. We use the notation Po(λ) to refer to this distribution.

We consider using these distributions truncated at zero to allow for the fact that the number of animals not seen is unknown. These have the form

$$\Pr(X = x)^* = \frac{\Pr(X = x)}{1 - \Pr(X = 0)} \quad (3)$$

for $x = 1, 2, \ldots$, where $\Pr(X = 0) = (1 + P)^{-k}$ and $e^{-\lambda}$ for the truncated negative binomial distribution (TNBD) and truncated Poisson distribution (TPD), respectively.

An intuitive moment estimator $\hat{N}$ of the population size can be obtained from the expected relationship $N[1 - \Pr(X = 0)] = n$, where n is the number of unique animals seen. Thus, the estimator

$$\hat{N} = \frac{n}{1 - \hat{P}_0} \quad (4)$$

where $\hat{P}_0$ is the estimate of $\Pr(X = 0)$ using the maximum likelihood estimates of the required parameters (Tanton 1965, Bunge and Fitzpatrick 1993).

Model Descriptions and Selection

Because we wish to estimate abundance for multiple populations simultaneously, it is reasonable to postulate a number of models that are derived from the TPD and TNBD by constraining some parameters to be constant for all populations. We focus our application to the example of 13 years of grizzly bear data where the populations of interest are the number of female bears with cubs-of-the-year.

A variety of models can be fit to the data, ranging from 1 assuming no heterogeneity and the fraction of the population expected to be observed is constant, to 1 assuming different levels of heterogeneity and the expected observable

fraction of the population differs in each year. We fit 6 models to the data:

Model 1.—No heterogeneity and a constant fraction of the populations expected to be observed in all years implying that counts come from the same Poisson distribution for all years. This has only 1 unknown: the Poisson mean.

Model 2.—No heterogeneity and the fraction of the populations expected to be observed differs among years. This suggests a different Poisson distribution each year, $\mathrm{Po}(\lambda_j)$. For the grizzly bear data there are 13 unknown parameters, 1 for each year.

Model 3.—For all years, there is a constant level of heterogeneity and constant fraction of the populations expected to be observed. The sighting distribution every year is thus the same, and is equivalent to the $\mathrm{NBD}(k, P)$. There are 2 unknown parameters: k and P.

Model 4.—For all years, there is a constant level of heterogeneity and the fraction of the populations expected to be observed differs between years. The sighting distribution in year j is then $\mathrm{NBD}(k, P_j)$. There are 14 unknown parameters: k and P_1 to P_{13}.

Model 5.—The level of heterogeneity differs among years, and the relative difference in heterogeneity between years determines the relative fraction of the populations expected to be observed. The sighting distribution is then $\mathrm{NBD}(k_j, P)$ for year j. There are 14 unknown parameters: P and k_1 to k_{13}.

Model 6.—The level of heterogeneity differs among years, as does the expected fraction of the population that is observed. The sighting distribution is then $\mathrm{NBD}(k_j, P_j)$ for year j. There are 26 unknown parameters: k_1 to k_{13} and P_1 to P_{13}.

Although models 1, 3, and 5 do not seem biologically realistic with respect to the Yellowstone grizzly bear data, they should not be rejected without suitable consideration. If there is little variation in the fraction observed, then the simpler model may have advantages over the more complex models by requiring the estimation of fewer parameters.

We use the approach of selecting 1 model from a range of possible alternative models (Hilborn and Mangel 1997, Burnham and Anderson 1998). A best model then can be chosen by selecting the model with the smallest AIC value (2·[number of parameters – log likelihood]; Akaike 1973). An alternative approach would be to use model averaging (Burnham and Anderson 1998), but this method is not explored here.

We stress that a model is only an approximation of ecological reality, and the purpose of AIC is to choose a model that captures notable features of the sample data with as few parameters as possible (Burnham and Anderson 1998). A selected model is not necessarily the true or correct model that generates the observed data.

Simulation Study 1

A simulation study was undertaken to determine the effectiveness of the AIC procedure for model selection. We investigated the effect of the following conditions: Population size (N) was made constant for all years and set at 3 levels with (i) $N = 20$, (ii) $N = 30$, and (iii) $N = 40$. Likewise, the level of heterogeneity was made constant for all years with (i) CV = 0.0, (ii) CV = 0.3, (iii) CV = 0.6, (iv) CV = 0.9, or alternatively (v) each year the CV was randomly chosen from a uniform distribution between 0.1 and 0.9. These situations led to data being generated from Model 2 when the CV = 0, Model 4, when the CV = 0.3, 0.6 or 0.9, and Model 6 when the CV was varied randomly. These factor values seem reasonable given the results of the analysis on the grizzly bear data. For direct comparability with the 1986–1998 grizzly bear data, 13 years of simulated data were generated from NBDs such that the expected total number of sightings during each year was equal to that observed (i.e., parameters for the simulated NBD in year j were: $k_j = 1/\mathrm{CV}_j^2$ and $P_j = S_j/N$, where S_j is the observed total number of sightings in year j).

For each of the 15 factor combinations, 13 years of data were generated by drawing N observations from the $\mathrm{NBD}(k_j, P_j)$ for year j. The 6 models were then fit to the data using maximum likelihood, and the estimated population size was recorded for each model. One model was chosen on the basis of AIC, and the results recorded separately. This procedure was repeated 5,000 times.

To find the maximum likelihood estimates of the parameters, a downhill simplex method was used to find the minimum of the negative log-likelihood (Press et al. 1992:408–412). Occasionally, this did not converge to finite estimates and in these situations the models that did not converge were not considered for selection on the basis of AIC. Because k, P, and λ must be greater than zero, the downhill simplex method was implemented using the natural logarithm of these parameters. Also, because the distribution of population estimates is skewed to the right, the average and root mean square error (RMSE)

were calculated on the natural logarithm of the estimates for each model. The model selected on the basis of AIC and the number of times convergence was not achieved were also recorded.

Lack-of-Fit

For the Yellowstone grizzly bear data, we calculated the Pearson χ^2 statistic to measure the lack-of-fit of the AIC-selected model to the observed data. This was done in the usual manner by comparing the observed frequency of bears seen X times to the expected frequency from the AIC-selected model. Cells with expected frequencies < 2 were pooled, starting with cells corresponding to the higher X values. The degrees of freedom (df) were calculated as the number of cells used, minus the number of parameters in the model, minus the number of populations being modeled. The Pearson χ^2 statistic was then compared to the χ^2 distribution with appropriate df to determine the lack-of-fit.

Bootstrap Standard Errors and Confidence Interval

Bootstrap techniques were used to estimate the standard error and for constructing 95% confidence intervals on the estimated number of unseen individuals (Manly 1997). A semi-parametric bootstrap approach was used where the "population" from which the bootstrap samples were drawn for year j consisted of the sightings for the observed n_j individuals, plus the estimated number of individuals with zero sightings ($\hat{U}_j$), rounded up to the nearest integer (U_j^*). N_j^* (= $U_j^* + n_j$) individuals were then randomly resampled, with replacement, from the "population" to create a bootstrap sample of n'_j individuals with non-zero sightings. Using this bootstrap method, the number of sightings and observed individuals will vary with each bootstrap sample. The only constraint used in the bootstrap process was that there must be at least 1 resighting. This restriction was necessary to prevent some estimators from tending to infinity and seems reasonable here because the minimum number of bears resighted ≥1 time in any year was 8 for the real grizzly bear data.

The first 4 models were fit to the 13 bootstrap samples and the best model selected on the basis of AIC. The estimated number of unseen individuals from the bootstrap data was recorded. This was repeated 1,000 times so that the expected value and standard error of the bootstrapped estimates of U^* could be calculated. The standard error of the bootstrap estimates can be used as an approximation of the standard error for $\hat{U}_j$ and $\hat{N}_j$ (Manly 1997). Because the distribution of population size estimates was skewed to the right, we assumed the estimates had a log-normal distribution. Models 5 and 6 were not fitted in this procedure because the simulation study on AIC model selection showed that these models often yielded population-size estimates that are unrealistically high and generally have poor properties.

Consider the following example of this semi-parametric bootstrap method. Suppose in 1 year 15 individuals were observed (n), of which 8 were sighted once, 5 were sighted twice, 1 sighted 4 times, and 1 sighted 8 times. From the AIC-selected model, the population was estimated to be 22.3 in which case $N^* = 23$ and $U^* = 8$. A bootstrap resample of size 23 is then taken from the population of 23 (= 8 + 15) animals, which resulted in 10 individuals being drawn that were not sighted, 9 that were sighted once, 3 sighted twice, and 1 sighted 8 times. This gives a total of 13 animals with at least 1 sighting (n') and would be used as the bootstrap sample to which the AIC selection procedure is applied.

This semi-parametric bootstrapping approach is justified based on the following grounds. Bootstrapping generally is used to avoid distributional assumptions about the data collected, hence the resampling of the observed data. However, a standard error should reflect the variability in the abundance estimates obtained by repeating the study and collecting alternative sets of sighting data from the entire population. Therefore, we need to estimate the number of individuals that were not encountered. There is also uncertainty in the AIC-selected model, because a different set of data collected from the same population may result in a different model being selected. This also needs to be reflected by the standard error.

Assuming $\ln(\hat{U}_j)$ has a normal distribution, a 95% confidence limit for $\ln(\hat{U}_j)$ is $\ln(\hat{U}_j) \pm 1.96 * \hat{\sigma}_{\ln(\hat{U}_j)}$. This leads to a lower limit for $\hat{N}_j$ of

$$n_j + \exp\left[\ln(\hat{U}_j) - 1.96 * \hat{\sigma}_{\ln(\hat{U}_j)}\right] \quad (5)$$

and an upper limit of

$$n_j + \exp\left[\ln(\hat{U}_j) + 1.96 * \hat{\sigma}_{\ln(\hat{U}_j)}\right] \quad (6)$$

where the standard error of $\ln(\hat{U}_j)$, $\hat{\sigma}_{\ln(\hat{U}_j)}$, is estimated by bootstrap resampling. Using the log-normal properties of $\ln(\hat{U}_j)$ the standard error for $\hat{U}_j$ and $\hat{N}_j$ will be

$$\sqrt{\exp\left(2 * \ln(\hat{U}_j) + \hat{\sigma}^2_{\ln(\hat{U}_j)}\right) * \left[\exp\left(\hat{\sigma}^2_{\ln(\hat{U}_j)}\right) - 1\right]} \quad (7)$$

Simulation Study 2

A simulation study was used to assess the bootstrap estimates of standard errors and 95% confidence intervals. Here, population size and level of heterogeneity were fixed for all years and set at 2 levels each with (i) $N = 30$ or (ii) $N = 40$, and (i) CV = 0.6 or (ii) CV = 0.9. Once again, the expected number of sightings for each of the 13 years was set equal to that observed. For the 4 factor combinations, 100 sets of data were generated, and the above bootstrap method was applied with 1,000 bootstrap samples drawn for each year. For the real data we obtained 5,000 bootstrap samples.

Trend Assessment

When estimates of population size are made over several years, managers often want to know whether there is a trend in abundance estimates over time. Monte Carlo simulations can be used to determine the significance of an apparent trend. Using the model selected by AIC, simulated sets of data were generated such that the expected number of sightings was the same as that observed, with the population size fixed and constant for all years. For each set of generated data, the AIC procedure was then used to select 1 of the first 4 models described earlier. The Spearman-rank correlation coefficient (r) was calculated to measure the degree of correlation between the AIC abundance estimates and year. Thus, each simulated set of data gave a rank correlation (r'). The significance of the observed r was determined by the proportion of the 5,000 simulations that provided an absolute value of r' greater than or equal to the absolute value of the observed r. The simulations were performed with population size and level of heterogeneity fixed at 3 levels each ($3 \times 3 = 9$ combinations) to investigate the sensitivity of the test to changes in these 2 factors. Also, N and CV were set at levels near the estimates of the original AIC-selected model.

Simulation Study 3

Simulations also were performed to explore the power of this test to detect a linear annual change in population size. For the grizzly bear example, population size was changed over time, either increasing at a rate of 2 bears per year or decreasing by 2 bears per year. Heterogeneity was set at 3 levels to examine the sensitivity of the test's power. From the above Monte Carlo simulations with constant population size, the critical values for the $\alpha = 5\%$ level were determined. These were compared to the absolute values of r' obtained from the 5,000 simulations with a trend in population size. Power was defined to be the proportion of the 5,000 simulations with the absolute value of r' greater than the critical value.

Software to perform the AIC model selection, bootstrap, and Monte Carlo procedures may be obtained from the senior author.

RESULTS

Simulation Study on Model Selection

Non-convergence of the likelihood function was a major problem only when Model 6, NBD(k_j, P_j), was fit to any of the simulated data sets, but particularly to those that were generated with little or no heterogeneity (Table 1). Results not presented here suggest that a reasonable explanation for this is the TNBD is tending to the TPD for 1 or more populations, hence the TNBD is an inappropriate model for that population. All other models converged for all simulated sets of data.

When data were generated from Model 2 (CV = 0), AIC most often (≥68.1%) selected the correct model and selected Model 4 incorrectly <27.5% of the time. Model 4 was correctly selected in most (>65.3%) of the simulations when CV was fixed at 0.6 or 0.9, but Model 2 was incorrectly selected most often (>52.8%) when CV = 0.3. Model 6 was rarely correctly identified as the true model when the level of heterogeneity varied randomly among years; instead, Model 4 was usually identified as the best (Table 2).

We also found a shift in which model was selected as population size increased. For larger values of N, and as the data become more sparse (fewer sightings per individual), the simpler, more restrictive models were selected on the basis of AIC, particularly with no, little, or random heterogeneity.

Table 1. Percentage of simulations for which the log-likelihood of Model 6, where both heterogeneity and sightability vary among years, did not converge. Results from simulation studies for $n = 20$ and $n = 40$ are presented here. The same conclusions can be made from the results for $n = 30$.

CV	$n = 20$	$n = 40$
0.0	14.3%	13.3%
0.3	5.0%	7.2%
0.6	0.1%	1.2%
0.9	0.0%	0.1%
Random	0.8%	2.9%

Table 2. Percentage of simulations for which model was selected on the basis of AIC. Entries in italics indicate where the true models have been correctly selected while values in bold text identify the model that was chosen most often. Results from the simulation studies for $n = 20$ and $n = 40$ are presented here. The same conclusions can be made from the results for $n = 30$.

		Selected model					
True model		1	2	3	4	5	6
CV = 0.0	$n = 20$	0.0%	***68.1%***	0.0%	27.5%	3.3%	1.0%
	$n = 40$	0.1%	***95.5%***	0.0%	3.9%	0.2%	0.3%
CV = 0.3	$n = 20$	0.0%	**52.8%**	0.1%	*37.5%*	8.9%	0.7%
	$n = 40$	0.2%	**75.9%**	0.3%	*22.1%*	1.4%	0.1%
CV = 0.6	$n = 20$	0.0%	3.2%	2.2%	***77.2%***	17.4%	0.1%
	$n = 40$	0.0%	8.8%	2.2%	***75.0%***	13.9%	0.0%
CV = 0.9	$n = 20$	0.0%	0.0%	16.7%	***65.3%***	18.0%	0.0%
	$n = 40$	0.0%	0.1%	9.8%	***73.0%***	17.1%	0.0%
CV is random	$n = 20$	0.0%	14.3%	1.1%	**67.9%**	15.8%	*0.9%*
	$n = 40$	0.0%	22.2%	1.5%	**68.7%**	7.5%	*0.1%*

Estimates of abundance from Models 5 and 6 were highly unstable with undesirable properties for years with a small number of sightings (<50), even when selected as the best model by the AIC criterion. This suggests that Models 5 and 6 are unsuitable and should not be considered in situations where the numbers of sightings are similar to those we observed. Estimates from these models have been removed from all subsequent results.

Figs. 1–3 show the back-transformed values of the average ln-transformed estimates of population size. In all situations, the AIC-selected model compared favorably to the true model, and often the AIC model was less biased. The results of Model 2 have been presented in Figs. 2 and 3 to illustrate the effect of assuming no heterogeneity when heterogeneity is present. For all models, the bias decreases as the number of sightings increases and the AIC model appears to be approximately unbiased, on the natural logarithm scale, for >30 sightings.

The residual mean square error (RMSE) of the AIC-selected model is similar to that of the true model for all fixed levels of heterogeneity (Figs. 4 and 5) and approximately equal to that of Model 2 when the level of heterogeneity varies randomly (Fig. 6). As would be expected, the RMSE is smaller for years with a greater number of sightings.

Bootstrap Standard Errors and Confidence Intervals

Simulations suggest that coverage of the 95% confidence intervals for abundance estimates are

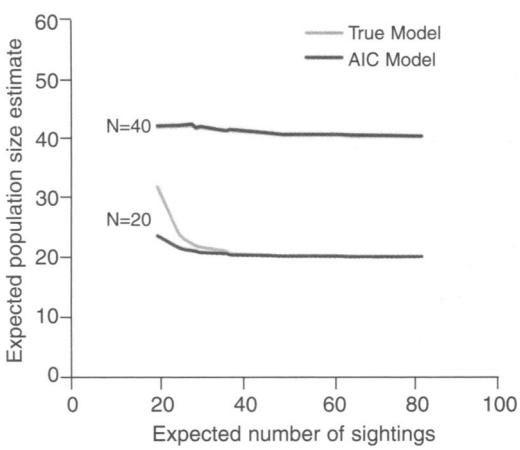

Fig. 1. Expected population size estimates for 13 years of data with expected number of sightings the same as those observed and CV = 0.

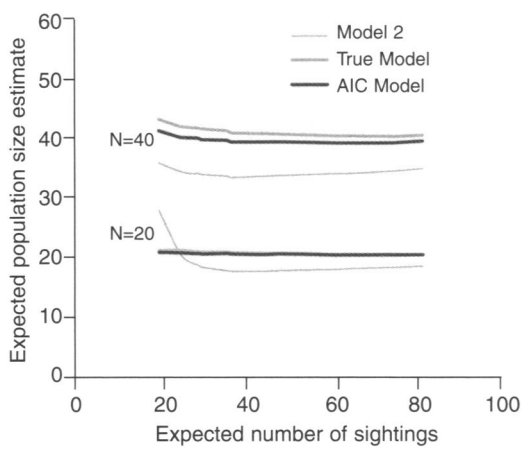

Fig. 2. Expected population size estimates for 13 years of data with expected number of sightings the same as those observed and CV = 0.6.

slightly low at approximately 90%, with the true population level being outside the confidence intervals on too many occasions (Table 3). Results not presented here indicate the bootstrap standard errors are in reasonable agreement with the true standard errors for the estimates of abundance, with a tendency to be slightly high.

Yellowstone Grizzly Bear Example

The sighting frequencies of female grizzly bears with cubs-of-the-year in the Yellowstone ecosystem from 1986–1998, as used for our analysis, is reproduced in Table 4. These differ from the data used by earlier authors (Knight et al. 1995, Boyce et al. 2001) due to recently identified discrepancies. Applying the AIC criteria, Model 4 was selected (Table 5), which appears to be a reasonable fit of the observed data (χ^2 = 23.03, df = 21, p = 0.170). The AIC weights (Burnham and Anderson 1998) provide further overwhelming evidence that Model 4 is the best of the models considered. Estimates of model parameters, population size, standard errors and 95% bootstrapped confidence intervals for abundance are presented in Table 6. Abundance estimates appear reasonable with an apparent upward trend. The estimate of k implies a CV for the Poisson mean of 0.74.

These results suggest the population size be fixed at 25, 35, and 45 using our Monte Carlo sim-

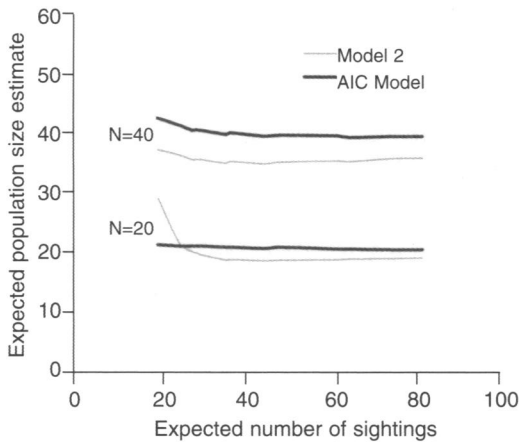

Fig. 3. Expected population size estimates for 13 years of data with expected number of sightings the same as those observed and random CV.

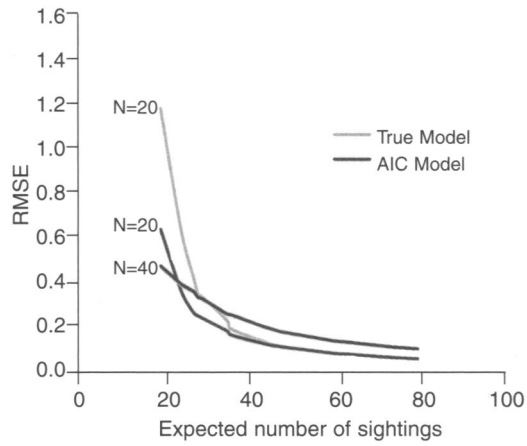

Fig. 4. RMSE of population size estimates for 13 years of data with expected number of sightings the same as those observed and CV = 0.

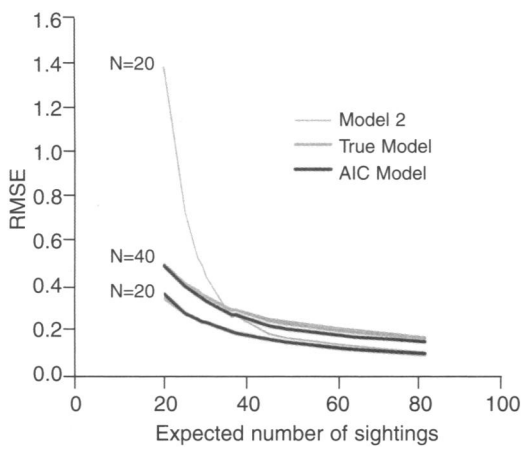

Fig. 5. RMSE of population size estimates for 13 years of data with expected number of sightings the same as those observed and CV = 0.6.

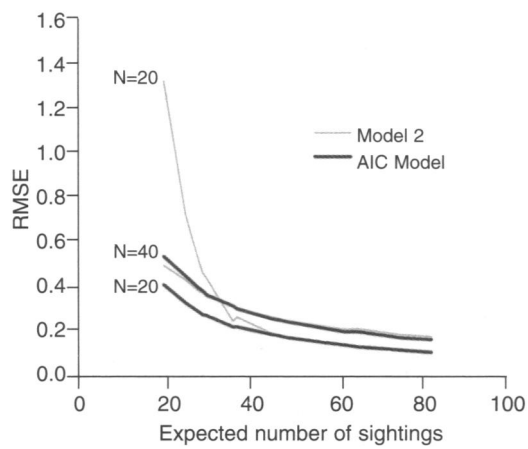

Fig. 6. RMSE of population size estimates for 13 years of data with expected number of sightings the same as those observed and random CV.

Table 3. Coverage of the semi-parametric bootstrap 95% confidence intervals as determined from 1,300 simulated confidence intervals. In theory, limits should be too extreme on less than 2.5% of occasions.

True model		Lower limit too high	Upper limit too low
CV = 0.6	$n = 30$	6.5%	4.9%
	$n = 40$	4.5%	4.8%
CV = 0.9	$n = 30$	6.4%	3.9%
	$n = 40$	4.6%	2.2%

ulation test for trend, with a CV of 0.55, 0.75, and 0.95. The observed value of the Spearman-rank correlation coefficient is 0.868 for which we obtain *P*-values of <0.001 in all 9 sets of simulations. This indicates very strong evidence of an increasing population. The average critical value at the $\alpha = 5\%$ level from these simulations is 0.515, which gives power estimates between 80–90% for both increasing and decreasing trends.

DISCUSSION

A variety of other methods are available for estimating abundance of heterogeneous populations including those of Burnham and Overton (1978), Chao (1987), and Chao and Lee (1992). However, as noted previously these methods are appropriate for situations where sightings occur at discrete occasions unlike the situation we are considering here. One could make data of sightings in continuous time discrete by considering whether an individual was sighted at least once within a relatively large time period (i.e., sighted at least once within a month), and then make comparison of abundance estimates with these other methods. A comparison of our proposed method with existing alternatives would be of value.

We acknowledge that our simulations do not address situations where the type of heterogeneity present in the Poisson means has a form that cannot be described by the gamma distribution, such as a bimodal distribution. However, we also note that the approach to estimation suggested here easily can be used with other distributions for the count data. The robustness of the negative binomial model and possible alternative models is an area that deserves to be explored in the future.

Our simulation results show that Models 5 and 6 provide highly unstable estimates of population size in data sets similar to the Yellowstone grizzly bear data. This may be due to the poor estimation of the heterogeneity parameter k from only 1 year of information. The estimates from Model 5 appeared to stabilize with greater than 60 sightings and that may indicate a minimum number of sightings required for this model to be used in practice. However, the estimates from Model 6 still had not stabilized with 80 sightings. This is similar to the findings of Wilbur and Landwehr (1974) who reported that the TNBD provided unrealistic estimates of the number of unseen animals in a single population, even when the TNBD was an adequate description of the observed frequencies.

On the basis of these findings and our other results, it appears that when analyzing multiple closed populations with potential sighting heterogeneity, using AIC to select 1 of the first 4

Table 4. Number of female grizzly bears with cubs of the year sighted X times, from 1986–1998. Total number of sightings is m_j, while n_j is the number of unique individuals.

	Year												
X	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
1	7	7	7	7	7	11	15	8	9	13	15	13	11
2	5	3	4	5	6	3	5	8	7	2	10	7	13
3	6	1	4		7	3	1	2	2	1	2	4	5
4	1	1	1	1	1	3	1				1	1	1
5	1		1		1	1	1			1		3	1
6						2							
7	1			1		1						1	2
8	2												
...													
15	1												
m_j	82	20	36	28	49	62	37	30	29	25	45	65	75
n_j	24	12	17	14	22	24	23	18	18	17	28	29	33

Table 5. Number of parameters, differences in AIC values relative to Model 4 (constant heterogeneity but with sightability varying among years), and AIC weights for models fitted to the Yellowstone grizzly bear data, 1986–1998.

Model	Number of parameters	ΔAIC	AIC weight
4	14	0.0	99.99%
3	2	9.0	0.01%
2	13	29.0	0.00%
1	1	60.0	0.00%

models that we described seems to be a valid approach to obtain reasonable estimates of population size if the selected model provides an adequate fit of the observed data. Again, we stress that the purpose of using AIC is not to identify the correct model that may be generating the data, but to select 1 model from a range of alternatives that is adequate in describing the collected data with as few parameters as possible.

Using bootstrap resampling to obtain estimates of the standard errors seems appropriate in this situation. Other, fully parametric approaches could have been used, but that would require the assumption that a particular model was true. By using a semi-parametric bootstrap approach, we assume that the AIC-selected model provides a good estimate of the number of unseen individuals. By incorporating the real data via a bootstrap procedure we avoid assuming a parametric model for individuals with 1 or more sightings. One reason for the lack of coverage of our confidence intervals may be that the log-normality assumption of the U_j's is not strictly correct. Alternatively, as we estimate the standard error, a constant other than 1.96 might be required when calculating the upper and lower limits. This situation would be equivalent to constructing a 95% confidence interval on a sample mean when the population standard deviation is unknown, where the *z*-value of 1.96 should be replaced by the corresponding *t*-value with appropriate degrees of freedom.

Monte Carlo simulations provide a method to determine the significance of an apparent trend by creating data sets with similar properties to that observed. Because the simulated sets of data have the same expected number of sightings for any given year, we have automatically allowed for bias in the population size estimates caused by a small number of sightings. Our results clearly suggest that despite any bias that may be present in our estimates, the number of female grizzly bears with cubs-of-the-year was increasing.

Table 6. Estimates of the number of female grizzly bears with cubs of the year ($\hat{N}_j$), standard error (SE), along with lower and upper limits of the 95% confidence intervals for 1986–1998. Maximum likelihood estimates of $\hat{P}_j$ from Model 4 (constant heterogeneity but with sightability varying among years) are also included where $\hat{k}$ = 1.817.

Year	$\hat{N}_j$	SE	95% confidence intervals Lower limit	Upper limit	$\hat{P}_j$
1986	29.4	3.4	26.0	38.9	1.535
1987	24.5	10.4	15.7	54.1	0.449
1988	26.8	5.5	20.9	41.8	0.740
1989	23.2	7.5	16.8	44.4	0.666
1990	33.4	5.4	27.0	47.9	0.808
1991	33.2	5.8	27.4	49.2	1.025
1992	49.3	16.4	32.6	94.9	0.413
1993	36.6	8.4	26.5	58.7	0.451
1994	38.5	10.2	26.8	65.7	0.414
1995	42.8	27.5	23.3	122.7	0.322
1996	60.0	14.0	42.9	96.7	0.413
1997	43.8	7.6	35.2	64.2	0.818
1998	49.4	7.5	40.4	69.1	0.836

MANAGEMENT IMPLICATIONS

Several authors have used counts of females with cubs to estimate the number of grizzly bears in the Yellowstone ecosystem (Dennis et al. 1991, Foley 1994, Boyce et al. 2001). However, these counts clearly are underestimates of the number of females with cubs; the counts vary among years because of variation in sampling effort (Mattson 1997, Boyce et al. 2001). The method that we propose removes the sampling-effort bias and will provide managers with better estimates of the true number of female grizzly bears with cubs and a statistically defensible method for detecting trends in the number of females.

For the application of this method to grizzly bears, we note that our estimates are only for the total number of adult females with cubs-of-the-year. Based on this value, and independent estimates of the proportion of the population composed of adult females with cubs-of-the-year, estimates of total population size can be obtained (Eberhardt et al. 1986). Because much of the year-to-year variation in counts of females with cubs may be attributed to annual fluctuations in reproductive output, a 3-year moving sum of the

number of adult females with cubs-of-the-year will be a more reliable basis for estimation of total population size (Eberhardt and Knight 1996). The 3-year moving sum is used for management purposes and is justified because grizzly bears in the Yellowstone ecosystem typically give birth to cubs every third year (U.S. Fish and Wildlife Service 1993). Even though some females may breed early if cubs are lost, other bears have intervals longer than 3 years such that the average interbirth interval in the Yellowstone ecosystem is close to 3 years (Eberhardt et al. 1994).

We caution that use of abundance estimates for grizzly bear populations should not be considered sufficient for monitoring trends, and we believe that continued monitoring of survival and reproduction of the bears is essential for this purpose (Eberhardt and Knight 1996). To extrapolate abundance estimates of females with cubs of the year to estimate total population size, demographic data are necessary to estimate the proportion of the population composed of females with cubs-of-the-year. Demographic details not only permit the estimation of trend to corroborate estimates based on other methods but also provide details to help understand mechanisms behind changes in the population (Knight and Eberhardt 1984, Eberhardt et al. 1994).

ACKNOWLEDGMENTS

We thank field personnel who helped collect data, including B. Blanchard, J. Emmerich, K. Gunther, R. Knight, L. Roop, D. Scholly, D. Stradley, and R. Swalley. Thanks to S. Cherry, L. Eberhardt, D. Mattson, K. Noyce, D. Otis, C. Servheen, S. Stringham, G. White, and anonymous reviewers for comments on the manuscript. K. Keating challenged our thinking on heterogeneity and furnished several useful references. We received support from the Alberta Conservation Association, Biological Resources Division of the U.S. Geological Survey, Boone and Crockett Club, Natural Sciences and Engineering Research Council of Canada, U.S. Forest Service, U.S. Fish and Wildlife Service, U.S. National Park Service and the Wyoming Game and Fish Department. Thanks to the National Fish and Wildlife Foundation for funding and administrative assistance.

LITERATURE CITED

AKAIKE, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 268–281 *in* B. N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory, Publishing House of the Hungarian Academy of Science, Budapest. Reprinted 1992, pages 610–624 *in* S. Kotz and N. Johnson, editors. Breakthroughs in statistics. Springer-Verlag, New York, USA.

BERTRAM, B. C. R. 1978. Pride of lions. Dent, London, United Kingdom.

BIGG, M. A. 1982. An assessment of killer whale (*Orcinus orca*) stocks off Vancouver Island, British Columbia. Report of the International Whaling Commission 32:655–666.

BOYCE, M. S., B. M. BLANCHARD, R. R. KNIGHT, AND C. SERVHEEN. 2001. Population viability for grizzly bears: a critical review. International Association for Bear Research and Management, Monographs Series, number 4.

BUNGE, J., AND M. FITZPATRICK. 1993. Estimating the number of species: a review. Journal of the American Statistical Association 88:364–373.

BURNHAM, K. P., AND D. R. ANDERSON. 1998. Model selection and inference. Springer-Verlag, New York, USA.

———, AND W. S. OVERTON. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65:625–633.

CARO, T. M., AND S. M. DURANT. 1991. Use of quantitative analyses of pelage characteristics to reveal family resemblances in genetically monomorphic cheetahs. Journal of Heredity 82:8–14.

CHAO, A. 1987. Estimating the population size for capture–recapture data with unequal catchability. Biometrics 43:783–791.

———, AND S.-M. LEE. 1992. Estimating the number of classes via sample coverage. Journal of the American Statistical Association 87:210–217.

CRAIG, C. C. 1953. On the utilisation of marked specimens in estimating populations of flying insects. Biometrika 42:170–176.

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

DENNIS, B., P. L. MUNHOLLAND, AND J. M. SCOTT. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs 61:115–143.

EBERHARDT, L. L. 1969. Population estimates from recapture frequencies. Journal of Wildlife Management 33:28–39.

———, B. M. BLANCHARD, AND R. R. KNIGHT. 1994. Population trend of the Yellowstone grizzly bear as estimated from reproductive and survival rates. Canadian Journal of Zoology 72:360–363.

———, AND R. R. KNIGHT. 1996. How many grizzlies in Yellowstone? Journal of Wildlife Management 60:416–421.

———, ———, AND B. M. BLANCHARD. 1986. Monitoring grizzly bear population trends. Journal of Wildlife Management 50:613–618.

FOLEY, P. 1994. Predicting extinction times from environmental stochasticity and carrying capacity. Conservation Biology 8:124–137.

FRAME, L. H., J. R. MALCOLM, G. W. FRAME, AND H. VAN LAWICK. 1979. Social organization of African wild dogs (*Lycaon pictus*) on the Serengeti plains, Tanzania, 1967–1978. Zeitschrift für Tierpsychology 50:225–249.

HAMMOND, P. S., S. A. MIZROCH, AND G. P. DONOVAN, editors. 1990. Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. Special Issue 12. International Whaling Commission, Cambridge, United Kingdom.

HILBORN, R., AND M. MANGEL. 1997. The ecological detective. Princeton University Press, New Jersey, USA.

KENDALL, M. G., AND A. STUART. 1969. The advanced theory of statistics. Volume 1. Third edition. Griffin, London, United Kingdom.

KNIGHT, R. R., B. M. BLANCHARD, AND L. L. EBERHARDT. 1995. Appraising status of the Yellowstone grizzly bear population by counting females with cubs-of-the-year. Wildlife Society Bulletin 23:245–248.

———, AND L. L. EBERHARDT. 1984. Projected future abundance of the Yellowstone grizzly bear. Journal of Wildlife Management 48:1434–1438.

KOTZ, S., N. L. JOHNSON, AND C. B. READ, editors. 1988. Encyclopedia of statistical sciences. John Wiley & Sons, New York, USA.

KREBS, C. J. 1999. Ecological methodology. Benjamin Cummings, Menlo Park, California, USA.

LAURIE, W. A. 1978. The ecology of the greater one-horned rhinoceros. Dissertation, Cambridge University, United Kingdom.

LEWONTIN, R. C., AND T. PROUT. 1956. Estimation of the number of different classes in a population. Biometrics 12:211–223.

MANLY, B. F. J. 1997. Randomization and Monte Carlo methods in biology. Second edition. Chapman & Hall, London, United Kingdom.

MATTSON, D. J. 1997. Sustainable grizzly bear mortality calculated from counts of females with cubs-of-the-year: an evaluation. Biological Conservation 81 103–111.

MILLS, L. S., J. J. CITTA, K. P. LAIR, M. K. SCHWARTZ, AND D. A. TALLMAN. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. Ecological Applications 10:283–294.

NOYCE, K. V., AND D. L. GARSHELIS. 1997. Influence of natural food abundance on black bear harvests in Minnesota. Journal of Wildlife Management 61:1067–1074.

OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D. R. ANDERSON. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62.

PAETKAU, D., AND C. STROBECK. 1994. Microsatellite analysis of genetic variation in black bear populations. Molecular Ecology 3:489–495.

PALSBØLL, P. J., J. ALLEN, M. BERUBE, P. J. CLAPHAM, T. P. FEDDERSEN, P. S. HAMMOND, R. R. HUDSON, H. JORGENSEN, S. KATONA, A. H. LARSEN, F. LARSEN, J. LIEN, D. K. MATTILA, J. SIGURJONSSON, R. SEARS, T. SMITH, R. SPONER, P. STEVICK, AND N. OIEN. 1997. Genetic tagging of humpback whales. Nature 388:767–769.

PICTON, H. L., D. M. MATTSON, B. M. BLANCHARD, AND R. R. KNIGHT. 1986. Climate, carrying capacity, and the Yellowstone grizzly bear. Pages 129–135 *in* G. Contreras and K. Evans, compilers. Proceedings of the grizzly bear habitat symposium. U.S. Forest Service General Technical Report INT-207.

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

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY. 1992. Numerical recipes in C: the art of scientific computing. Second edition. Cambridge University Press, United Kingdom.

SEBER, G. A. F. 1982. The estimation of animal abundance and related parameters. Second edition. Charles Griffin, London, United Kingdom.

SOUTHWOOD, T. R. E., AND P. A. HENDERSON. 2000. Ecological methods. Third edition. Blackwell Science, Oxford, United Kingdom.

TANTON, M. T. 1965. Problems of live-trapping and population estimation for the wood mouse, *Apodemus sylvaticus* (L.). Journal of Animal Ecology 34:1–22.

U.S. FISH AND WILDLIFE SERVICE. 1993. Grizzly bear recovery plan. U.S. Department of the Interior, Missoula, Montana, USA.

WILBUR, H. M., AND J. M. LANDWEHR. 1974. The estimation of population size with equal and unequal risks of capture. Ecology 55:1339–1348.

Received 29 December 1999.
Accepted 10 February 2001.
Associate Editor: Udevitz.