BioOne COMPLETE

Model evaluation in statistical population reconstruction

Authors: Skalski, John R., Clawson, Michael V., and Millspaugh, Joshua J.

Source: Wildlife Biology, 18(3) : 225-234

Published By: Nordic Board for Wildlife Research

URL: https://doi.org/10.2981/11-080

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Wildl. Biol. 18: 225-234 (2012)
DOI: 10.2981/11-080
© Wildlife Biology, NKV
www.wildlifebiology.com

Original article

Model evaluation in statistical population reconstruction

John R. Skalski, Michael V. Clawson & Joshua J. Millspaugh

Statistical population reconstruction can be a valuable tool for monitoring the status and trends of game populations at large spatial scales using age-at-harvest data. Despite their utility and increasing use in demographic studies, it is necessary that these models be evaluated before their results are applied. We recommend practitioners evaluate their fitted population models using a variety of approaches, including residual analyses, point-deletion techniques and sensitivity analyses, and we illustrate these techniques using several case studies. Although we stress the value of these quantitative procedures, the final evaluation criterion should be the biological realism of the estimated demographic parameters and trends. Auxiliary field data should be used whenever possible in this final model check. After investigators are satisfied that the selected model(s) is/are adequate, this auxiliary data can be incorporated in a final stage of the analyses to further improve accuracy and precision of the population projections. The procedures we outline and recommendations we make will improve the credibility and utility of results of population reconstruction modeling.

Key words: abundance estimation, age-at-harvest data, model evaluation, model selection, numerical optimization, population reconstruction, residual analyses

John R. Skalski, School of Aquatic and Fishery Sciences, University of Washington, 1325 Fourth Avenue, Suite 1820, Seattle, Washington 98101-2509, USA - e-mail: skalski@uw.edu
Michael V. Clawson, School of Forest Resources, University of Washington, 1325 Fourth Avenue, Suite 1820, Seattle, Washington 98101-2509, USA - e-mail: michael.v.clawson@gmail.com
Joshua J. Millspaugh, Department of Fisheries and Wildlife Sciences, University of Missouri, 302 Natural Resources Building, Columbia, Missouri 65211, USA - e-mail: millspaughj@missour.edu

Corresponding author: John R. Skalski

Received 25 August 2011, accepted 18 January 2012

Associate Editor: Olivier Gimenez

Within the last decade, statistical population reconstruction (SPR) has emerged as a useful tool for inventorying game populations based on the analysis of age-at-harvest data (Gove et al. 2002, Skalski et al. 2007, Broms et al. 2010). The method formulates recruitment, survival and harvest processes in order to reconstruct cohort and historical annual abundance levels. The attraction of population reconstruction is that it can be used to inventory populations over large geographic scales using the extensively collected harvest information commonly recorded by state wildlife agencies. The method provides a flexible statistical framework that can incorporate auxiliary field data from abundance surveys and tagging studies of survival and harvest to improve the accuracy and precision of the population reconstruction.

However, like all statistical models, population reconstructions are necessarily simplifications of reality and, as such, subject to systematic biases. Therefore, before implementing management actions based on the survey results, it behooves game managers to evaluate the reliability of the models and subsequent abundance estimates. The complexity of these models and their potential impact on population management suggest all possible avenues of model evaluation should be considered. A strategic approach for evaluation and checking the fit of SPR models is therefore needed.

Although the process of model evaluation is different from model selection, the two concepts are nevertheless intertwined by the common goal of finding the most appropriate model(s) for the data. Most recently, model selection has been based on

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

either Bayesian information criterion (BIC; Schwartz 1978) or the information-theoretic approach using Akaike information criterion (AIC; Burnham & Anderson 1998, Anderson 2008). Likelihood ratio tests (LRTs) still provide a conventional alternative when the models being compared are nested (Hogg & Craig 1995:413-422).

In contrast, one need look no further than regression analysis to see a very different perspective on model evaluation. Residual analyses are at the heart of such methods, whether ordinary least squares (Neter et al. 1996:97-110, 368-384, Draper & Smith 1998:59-68) or generalized linear models (GLM; McCullagh & Nelder 1989:37-40). These approaches use information outside the likelihood foundation to evaluate model lack-of-fit, identify outliers and assess robustness.

Merger of the different philosophies of model evaluation can be found in Johnson & Omland (2004) where they suggest using goodness-of-fit criteria in the early stages of analyses to identify a candidate set of models for subsequent model selection and weighting. They argue that model selection and model weighting will not help if the set of candidate models is inherently flawed from the beginning.

The purpose of our paper is to illustrate how SPR may be evaluated using available information and techniques. Examples from several different population reconstructions are used to illustrate the uses of residual analyses, model predictions and sensitivity analyses to evaluate model results. We hope that investigators will incorporate these model evaluation procedures into population reconstruction efforts in order to improve the reliability of management decisions based on the results of SPR models.

Overview of SPR

Gove et al. (2002) developed stochastic models for wildlife reconstruction using age-at-harvest data and called the approach 'statistical age-at-harvest', or more generally, SPR (Skalski et al. 2005). These models offer a general statistical approach to analyzing age-at-harvest data and are hand-tailored for the data at hand. Age-at-harvest data can take several forms, including the number of harvested animals in individual age classes or biologically relevant groups (e.g. juvenile and adult; Broms et al. 2010). SPR allows researchers to estimate recruitment, abundance, harvest rates over time and age-specific survival rates with confidence intervals.

The first step in SPR involves the development of a joint likelihood model, which is comprised of three elements. First, an age-at-harvest likelihood describes the probability of observing the age-at-harvest data as a function of recruitment, natural survival, harvest probabilities and reporting rates. The age-at-harvest likelihood is a product of likelihoods, one for each cohort (i.e. age or stage) represented in the age-at-harvest matrix. Second, a reporting likelihood describes how harvest data were sampled to obtain the age-structure data. Last, one or more auxiliary likelihoods are included that describe the probability of observing ancillary field data used in estimating one or more of the demographic parameters or a function of the demographic parameters in the age-at-harvest likelihood. Commonly, hunter effort or radio-telemetry or both are used as the auxiliary information.

$$L = L_{\text{Age-at-harvest}} \cdot L_{\text{Reporting}} \cdot \prod_{i=1}^{k} L_{\text{Auxiliary}_i} \qquad (1).$$

The likelihood model (1) provides a convenient approach to incorporate the necessary data in order for the joint likelihood to be estimable and to propagate the error variances associated with all aspects of the reconstruction. Because of the complexity of the likelihood model (1), there are no closed form estimators and investigators must use computer algorithms to numerically solve for the maximum likelihood estimates and associated variance estimates. These process-based models necessitate careful model evaluation.

When the only data available are the age-at-harvest matrix, even the simplest of population reconstruction models is overparameterized (Gove et al. 2002). Estimability is obtainable only after auxiliary data such as harvest effort (Skalski et al. 2007), radio-telemetry (Broms et al. 2010) or other survey data become available. Even then, both age- and time-specific survival and harvest parameters will be unavailable, requiring simplification of the survival, recruitment and harvest processes. Model evaluation must therefore be central to any game inventory based on SPR.

SPR has become a popular alternative to traditional reconstruction methods. Gove et al. (2002) used the SPR to estimate elk *Cervus elaphus* abundance in Idaho, USA. Skalski et al. (2005) used the approach on the Washington State black bear *Ursus*

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

americanus population and a Washington State mule deer *Odocoileus hemionus* population. Recently, Broms et al. (2010) adapted the method to estimate small game populations including the sage grouse *Centrocerus urophasianus* population in Oregon, USA. Fieberg et al. (2010) used SPR to estimate black bear demographics in Minnesota, USA. SPR has been applied within a Bayesian context as well (Conn et al. 2008).

Residual plots

The numbers in an age-at-harvest matrix can be compared to their expected values based on the fitted SPR model where

h_{ij} = number of animals harvested and aged in year i (i = 1, ... , Y) for age class j (j = 1, ... , A),
$\hat{h}_{ij}$ = expected number of animals harvested and aged in year i (i = 1, ... , Y) for age class j (j = 1, ... , A).

The most familiar form of standardized residuals,

$$Z_{ij} = \frac{(h_{ij} - \hat{h}_{ij})}{\sqrt{\hat{h}_{ij}}} \qquad (2),$$

are the values that contribute to the Pearson goodness-of-fit statistic (Kotz et al. 1985:518), where

$$\chi^2_{AY\text{-}P} = \sum_{i=1}^{Y} \sum_{j=1}^{A} \frac{(h_{ij} - \hat{h}_{ij})^2}{h_{ij}} \qquad (3),$$

which is χ^2 distributed with degrees of freedom AY-P and where P is the number of parameters estimated in the SPR model. Assuming the counts are Poisson distributed, the individual Z_{ij} are asymptotically distributed as a standard normal random variable (i.e. N(0,1)). The advantage in using Equation (2) is that individual contributions to the overall goodness-of-fit statistic (3) can be examined.

However, because the typical likelihood in SPR models describes the harvest numbers for a cohort as a multinomial distribution, we recommend using the Anscombe residuals for a binomial distribution (Anscombe 1953, Cox & Snell 1968) because the Pearson residuals can be skewed. The Anscombe residual Z_i for the ith cell of the multinomial is also asymptotically distributed as (N(0,1)) where

$$Z_i = \sqrt{R}\,\frac{\beta\left(\frac{A_i}{R}, \frac{2}{3}, \frac{2}{3}\right) - \beta\left(\hat{p}_i, \frac{2}{3}, \frac{2}{3}\right)}{\left[\hat{p}_i(1 - \hat{p}_i)\right]^{1/6}} \qquad (4)$$

(Collett 1991:330-331), and where

R = sample size (i.e. $\sum_i A_i$),
A_i = count in the ith cell of the multinomial distribution,
$\hat{p}_i$ = estimated probability of occurrence based on the fitted model, and

$$\beta\text{eta}(x, a, b) = \int_0^x t^{a-1}\,(1 - t)^{b-1}\,dt \text{ for } 0 \le x \le 1.$$

We recommend plotting these standardized residuals either by age classes within a year or across years within an age class (Fig. 1). Such residual plots are helpful in identifying any systemic lack-of-fit for a particular age class or year of reconstruction. These two types of residual plots for a black-tailed deer *Odocoileus hemionus* SPR model of 13 age classes × 24 years of harvest data performed by Skalski et al. (2007) suggest no serious systematic lack-of-fit (see Fig. 1). In expectation, 95% of all the standardized residuals should have values within ± 2. Five of 24 standardized residuals for age class 12 are outside ± 2. However, because the harvest numbers are very low (i.e. 0, 1 or 2) in this older age class, any lack-of-fit here is likely to have a very small effect on the reconstructed abundance.

In contrast, misspecification of the youngest age classes can have major consequences for the reliability of estimates produced from SPR models. For example, in an SPR model of a cougar *Puma concolor* population in the Blue Mountains of Oregon, USA (Clawson 2010), residuals showed no lack of model fit when plotted by year (Fig. 2). However, when plotted by age class, the residuals showed a substantial lack of model fit to the first few age classes (see Fig. 2). In order to address the lack of model fit, the harvest and survival probabilities were modeled as age-class-specific processes. The result was a better fitting model with lower juvenile natural survival and harvest probabilities than those of adults. The effect was also a lower and more realistic estimate of population abundance (Clawson 2010).

Deletion techniques

Classical regression analyses use point-deletion techniques to identify influential data points and the sensitivity of the regression results to outliers (Belsley et al. 2004:6-39). All possible single-point deletions are used in such analyses. Considering SPR, all possible single-point (i.e. annual) detections are not

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

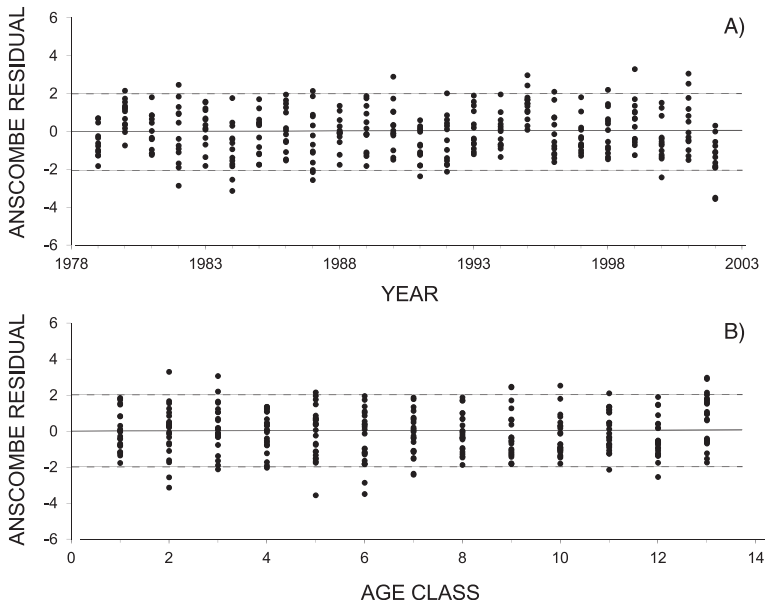


Figure 1. Anscombe residual plots from a black-tailed deer population reconstruction (2007) plotted by year of harvest (A) and age class (B). In expectation, 95% of these standardized residuals should be within ± 2 (i.e. 1.96) standard deviations from the mean zero when the SPR model fits adequately.

possible because of the cohort structure of the age-at-harvest matrix. Instead, in this context, one or more consecutive years of harvest data can be deleted at either the beginning or end of the reconstruction period to determine the stability of the estimated population trend.

For managers using SPR to inventory game populations, stability of the most recent years of the reconstruction is most important. In these circumstances, sensitivity analyses should focus on deleting some of the earliest years of harvest data to assess robustness of the current population esti-

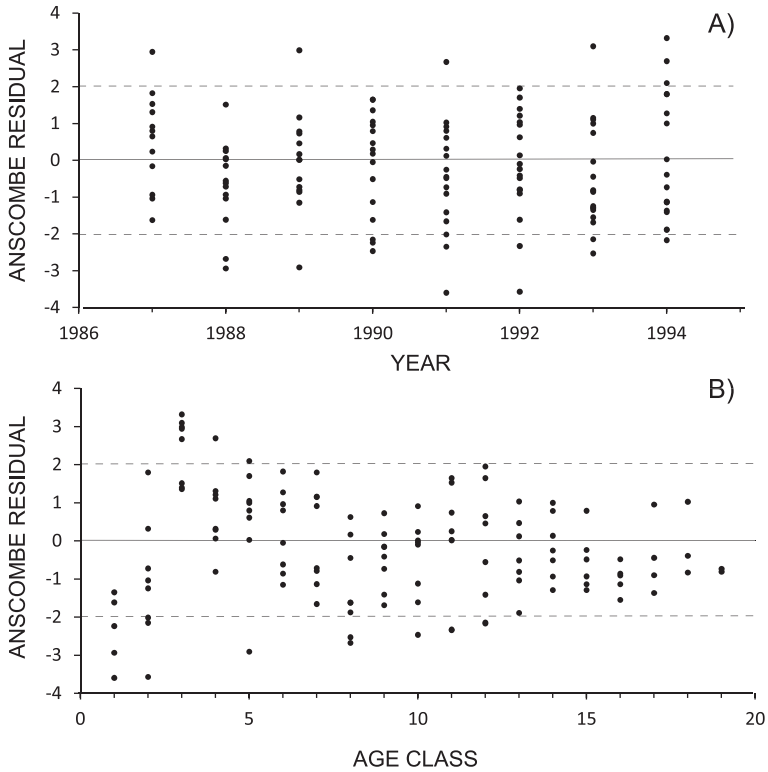


Figure 2. Anscombe residual plots from a cougar population reconstruction illustrating reasonable dispersion of residuals by year of harvest (A), but with systematic problems when plotted by age class (B).

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

mates. Hilborn & Walters (1992) further warn against using a long series of harvest data in reconstructing current population levels. They argue that historical data may not represent current harvest conditions and, as such, could have an undue influence on the population reconstruction. In fisheries, the recommended deletion techniques are called retrospective analyses (Parma 1993, NRC 1998).

In an SPR model of an American marten *Martes americana* population in the Upper Peninsula of Michigan, USA, with 11 age classes and 12 years of data (Skalski et al. 2011), deletions of two and four of the most recent years of harvest data have little effect on historical population trends (Fig. 3). However, in the population reconstruction for black-tailed deer (1979-2002) performed by Skalski et al. (2007), the current year's abundance estimate was robust to two years of historical harvest data being removed, but estimates became more sensitive when four or more of the earliest years of harvest data were deleted (Fig. 4).

Sensitivity of current population estimates to point-deletion techniques should be viewed as a warning. Sensitivity that goes beyond the uncertainty captured by standard confidence intervals needs to be accounted for in management decisions. Currently, there are no guidelines concerning how many years of data are reasonable to delete before estimation inherently suffers. Nevertheless, a high degree of sensitivity, with small amounts of data removed, is an indicator of insufficient data for population reconstruction, and the need to collect additional auxiliary data should be a part of future management plans. More research is needed to assess these critical issues in SPR models and we raise this issue so others do not blindly follow model results.

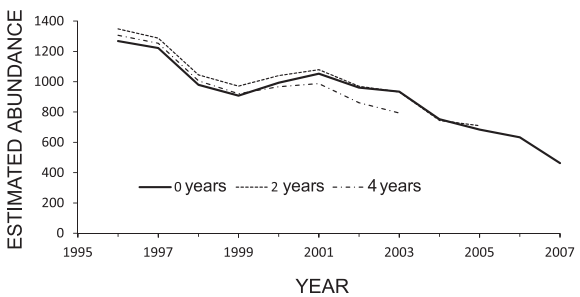


Figure 3. Annual abundance trends from SPR of American martens in Michigan, USA, with varying numbers of current years of age-at-harvest data removed. The numbers of years removed is indicated by the different lines.

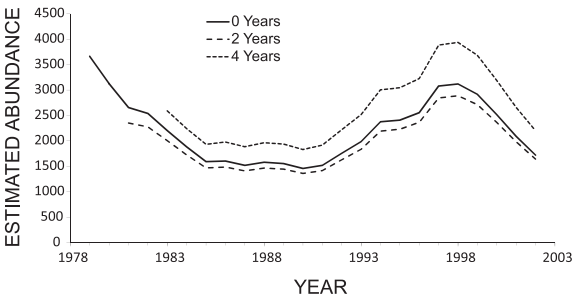


Figure 4. Annual abundance trends from SPR of female black-tailed deer, with varying numbers of historical age-at-harvest data removed, 1978-2003. The numbers of years removed is indicated by the different lines.

Realism of model output

The processes of recruitment, survival and harvest estimated by the SPR should always be compared to an investigator's best understanding of the population dynamics of the species of focus from other relevant data sources. Appreciable departures of model estimates from independent sources of information or data trends should be investigated and resolved before the results of an SPR are included in management plans. Estimates of natural survival and rates of recruitment should be compared to values in the literature as a check of model realism. Reconstructed abundance and population trends should be compared to independent population estimates and/or other indices of abundance for consistency.

Skalski et al. (2007) compared reconstructed black-tailed deer abundance (1979-2002) against an independent annual browse index (Fig. 5). They found an independent browse index of deer abundance to be more useful as a post-reconstruction model check than as an auxiliary data source in the population reconstruction analysis. Although an index can be used to evaluate the reconstructed population trend, it cannot be used to confirm the absolute scale of abundance estimates from the SPR. In a population reconstruction of American martens in the Upper Peninsula of Michigan, reconstruction abundance was favorably compared to an independent estimate of density derived from a DNA-based mark-recapture study (Skalski et al. 2011). The mark-recapture data could have been formally incorporated into the population reconstruction model but were used instead as an independent source of confirmatory data. Conversely, Fieberg et al. (2010) incorporated three years of independent

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

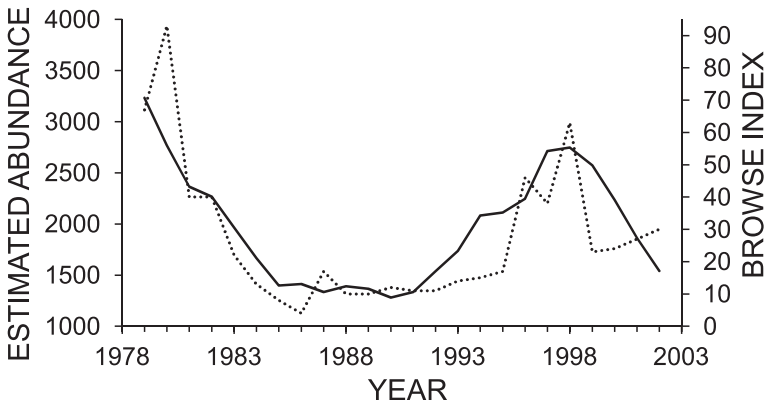


Figure 5. Reconstruction abundance (N) of black-tailed deer in Pierce County, Washington, USA, 1979-2002 (solid line) and an annual browse index (dotted line). Data from Skalski et al. (2007). This type of model check can be used to determine whether the population reconstruction has properly captured the time trend but cannot be used to assure whether absolute abundance has been properly estimated.

mark-recapture data into their reconstruction of a Minnesota black bear population.

Survival estimates and age-specific survival trends should be realistic and within their nominal range. We do not advocate reparameterizing survival parameters so they are automatically set to $0 \leq S \leq 1$. A common reparameterization is the logistic transformation

$$S_i = \frac{e^{\beta_i}}{1 + e^{\beta_i}} \qquad (5).$$

However, using Equation (5) can mask the degree of model misspecification. For example, estimates of $\hat{S} = 1.03$ and $\hat{S} = 2.03$ are both outside the nominal range for a probability. In the case of $\hat{S} = 1.03$ with a standard error (SE) of 0.05, the problem may be due to sampling error when attempting to estimate a true survival value near 1. In which case, a logistic transformation might be the appropriate analytical solution. On the other hand, $\hat{S} = 2.03$ with an $SE = 0.30$ may suggest serious model misspecification that must be resolved before the population reconstruction can be trusted.

The estimated trend in age-specific natural survivals should also be checked for realism when sufficient data are available for their estimation. In the case of a population reconstruction of American martens in the Upper Peninsula of Michigan (Skalski et al. 2011), a quadratic trend in age-specific survivals was estimated with young-of-year having lower survival than adults and with some indication of senescence with increased age (Fig. 6). The survival trend seemed plausible and was supported by an age-specific model. Conversely, an initial attempt at modeling natural survivals of Michigan elk resulted in an unrealistic, trough-shaped survivorship curve when a harvest model with a common vulnerability coefficient was initially employed (Fig. 7). The trough-shaped model unrealistically estimated natural survival to be higher for young-of-year, yearlings and the very old rather than prime-aged adults, suggesting possible model misspecification. The problem was remedied by including age-specific vulnerability coefficients in the catch-effort relationship (see Fig. 7). Because of the martingale relationship between natural and harvest survival in demographic models, they can have a strong negative covariance where one parameter may be overestimated and the other is subsequently underestimated.

The cohort-based models of Gove et al. (2002) and Skalski et al. (2007) directly estimate the abundance of entering cohorts over time. The pattern and level of recruitment can therefore also be checked for realism. First, recruit-to-adult ratios should be calculated and compared to reported values in the literature or field observations. Secondly, estimated trends in recruitment abundance should be compared to what is known about the population's health and the age-at-harvest data. For example,

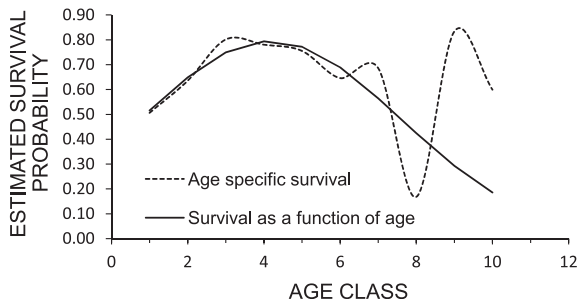


Figure 6. Fitted age-specific survival function (solid line) for American martens in the Upper Peninsula of Michigan and associated age-specific survival estimates (dashed line). Data from Skalski et al. (2011).

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

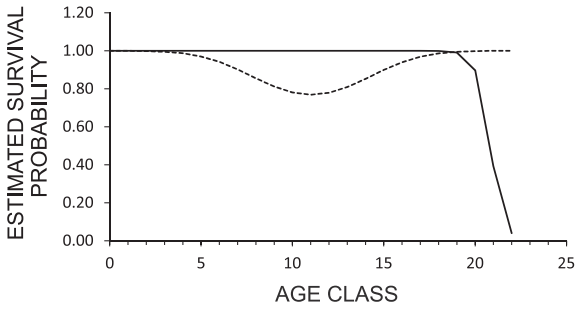


Figure 7. Unnatural, trough-shaped natural survival function for Michigan elk when the harvest model was misspecified using a common vulnerability model (dashed line) and a more realistic age-specific vulnerability model (solid line).

estimates of soaring recruitment while indices of abundance are on the decline might suggest incompatible results. Or, for another example, an estimated sudden suspension of recruitment when the harvest data are indicating the proportion of young-at-year is stable might also suggest incompatible results (Fig. 8). This latter example occurred when the need for age-specific harvest parameters was ignored during initial stages in a population reconstruction of an elk herd (see Fig. 8).

Proper optimization

Because closed-form estimators of abundance in population reconstruction do not exist, investigators must rely on computer algorithms to numerically solve for the maximum likelihood estimates and associated variance estimates. Therefore, numerical aspects of population reconstruction are a very important component in the overall analysis of age-at-harvest data. The choice of initial parameter estimates, the selection of the optimization routine, as well as convergence criteria and step size used in the analysis, are all important considerations. Diligence is required to assure that the numerical methods find the global maximum, and that they do not produce a false convergence.

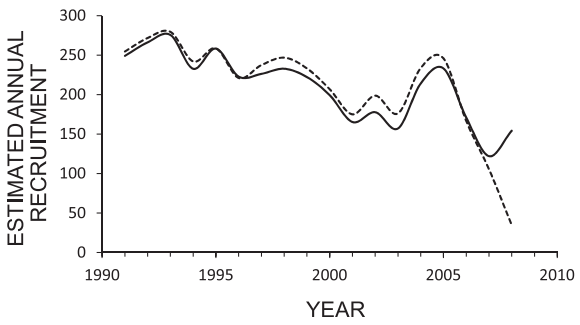


Figure 8. Recruitment trends for a Michigan elk herd under two model conditions: a common hunting vulnerability (dotted line), causing recruitment to falsely plummet and a separate hunting vulnerability for elk 1.5 and $\geq$ 2.5 years old (solid line).

There can be huge differences in the performance of alternative numerical optimizations or solver programs. The task of selecting a numerical program is made more difficult because the best choice can vary with the nature of the problem. Most optimization algorithms are variations on the Newton-Raphson method (Seber 1982:16-18). In its original form, the Newton-Raphson method requires the investigator to provide the first and second derivatives of the log-likelihood model with regard to each model parameter. Thankfully, modern optimizers can approximate these derivatives using numerical techniques. One exception is AD Model Builder (Otter Research Ltd., available at: http://otter-rsch.com/admodel.htm) that algebraically derives the first derivatives of the log-likelihood model. The various numerical software programs thus differ in the way they use derivatives. There are non-derivative methods as well as first derivative and both first and second derivative optimization routines. Generally, when more of the derivative information is used, the optimization method is faster and more reliable. The analytical approach used in AD Model Builder is one reason it tends to produce reliable point and variance estimates, but the syntax of the software can be a barrier for some users.

Program USER, which we have used in population reconstruction, has several alternative optimizers available. We recommend a multistage approach to optimization. For instance, if reliable initial parameter guesstimates are unavailable, a non-derivative grid search routine such as SIMPLEX may be used (Press et al. 2007:502-507). The resulting parametric estimates from this stage would then be used in a more sophisticated optimization routine that uses numerical first and/or second derivatives such as a quasi-Newton method (Press et al. 2007:521-526). The results should be run repeatedly through multiple optimizers, varying both step size and convergence criteria, and monitoring parameter estimates throughout the process. It is often necessary to re-seed portions of the model, especially when using logistically transformed survival parameters. Convergence criteria should be very strenuous with

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

12 to 16 decimal places in the log-likelihood value in common between iterations.

Optimizers based on numerical derivatives or modified grid search routines can be sensitive to step-size settings. For this reason, both point and variance estimates should be examined for robustness to the step-size specifications in the optimizer. The robustness of the point estimates should also be examined with regard to the initial values used in starting the optimization procedure and to assure that the global optimum, and not a local optimum, has been found. A global optimum should be verified by evaluating the likelihood model at values other than at the supposed maximum likelihood estimates to assure no other set of parameter values produces a larger likelihood.

When the likelihood model includes one or more vulnerability coefficients for modeling the catch-effort relationships, e.g.

$$p_i = 1 - e^{-cf_i}$$

where p_i = probability of harvest in year i, f_i = total hunter effort and c = vulnerability coefficient, it is often advisable to rescale hunter effort. Numerical optimizers can have difficulty in locating the global maximum if parameter values differ by many orders of magnitude. For instance, animal abundance may be in the 1,000s or 10,000s while the vulnerability coefficient(s) may be in the range of 0.0001s if hunter effort is large. One way of reducing this range in parameter values is to reexpress hunter efforts in terms of 100s or 1000s of hunters rather than the number of hunters. In so doing, the usually small vulnerability of coefficients can be increased several orders of magnitude, making the disparity in parameter values less and location of the optimum easier and more reliable.

Recommendations

We recommend using residual and sensitivity analyses combined with checking the realism of the model output as an integral step in population reconstruction. These steps need to occur before any model averaging is attempted to assure the set of models being considered is realistic. While the purpose of model averaging is to be a robust method of estimation and provide a realistic estimate of variance, neither will be achieved if the underlying set of models is inherently flawed. It is possible that only a few of the candidate models may successfully pass through this screening process. Even fewer might rank high enough to be meaningful in model averaging. We caution investigators against artificially inflating the number of candidate models to permit model averaging. Redundant models can have the serious effect of underestimating model uncertainty and the supposed benefits of model averaging. Faced with too few candidate models, we recommend reporting, as an alternative, the range in abundance predictions from the different models, until sufficient auxiliary data have been collected to construct a reasonable set of candidate models for model averaging.

In light of our discussion above, we make the following recommendations to those evaluating SPR models:

1) Plot the Anscombe residuals for a binomial distribution (Anscombe 1953, Cox & Snell 1968). Be sure to plot these standardized residuals both by age classes within a year and across years within an age class. By doing so, one can identify systemic lack-of-fit for a particular age class or year of reconstruction.
2) One should assess the stability of the population estimate and determine whether it is robust to the number of years of data used in the reconstruction. To do this, we recommend deleting one or more consecutive years of harvest data at the beginning or end of the reconstruction period and assess stability in the estimated trend and the estimated demographic rates. Should the results be affected using point-deletion techniques, it should be viewed as a warning and might indicate either that enough data are not yet available or that early and later years of reconstruction may be incompatible.

   Further, we recommend holding back some of the auxiliary data in order to use it specifically as an independent check of model realism. Once investigators are convinced the most appropriate models have been constructed and that these models provide realistic parameter estimates and population projections, the remaining auxiliary data might be incorporated into a final modeling step to further help improve accuracy and precision. In this manner, all possible analytical steps have been taken to help assure model reliability and the subsequent precision and accuracy of the demographic parameters being estimated. No less should be expected when providing stewardship of our wild resources.

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

3) Each parameter in the reconstruction analysis, including survival, recruitment, abundance and harvest rates should be evaluated for realism. SPR models have one important distinction from typical regression models that also must be taken into account. The SPR models are inherently process-based models, and the subsequent estimates of recruitment, survival and harvest must be realistic and sensible. Models, regardless of their ranking or weight, that fail to predict known patterns of response or produce implausible parameter estimates must be considered suspect or untenable (Lytle 2002). SPR models allow investigators to examine estimates of recruitment, abundance, harvest rates over time and age-specific survival processes for realism and consistency with the known data. This reality check in the final analysis may be more important than any goodness-of-fit test or model ranking. Should appreciable differences between model projections and independent sources of information exist, it is necessary to resolve these differences before applying reconstruction results. Estimates of natural survival and rates of recruitment should be compared to values in the literature as a check on model realism.
4) We do not recommend routinely reparameterizing survival parameters so they are inherently $0 \leq S \leq 1$. Substantial departures (e.g. $\hat{S} = 2.03$) may suggest serious model misspecification issues. Parameterizing so they are forced to be $0 \leq S \leq 1$ can mask these issues.
5) Proper optimization is critically important. There are often tremendous differences in the performance of alternative numerical optimization routines and researchers must carefully consider which approach is most appropriate for their analysis. We find that such differences are often not well appreciated by those simply interested in obtaining an answer. Unfortunately, if due diligence is not used, software programs may produce erroneous results.

References

Anderson, D.R. 2008: Model based inferences in the life sciences: A primer on evidence. - Springer, New York, New York, USA, 184 pp.

Anscombe, F.J. 1953: Contribution to the discussion of H. Hotelling's paper. - Journal of the Royal Statistical Society, Series B 15: 165-173.

Belsley, D.A., Kuh, E. & Welsch, R.E. 2004: Regression diagnostics: Identifying influential data and sources of collinearity. - Wiley-Interscience, Hoboken, New Jersey, USA, 292 pp.

Broms, K.M., Skalski, J.R., Millspaugh, J.J., Hagen, C.A. & Schulz, J.H. 2010: Using statistical population reconstruction to estimate demographic trends in small game populations. - Journal of Wildlife Management 74: 310-317.

Burnham, K.P. & Anderson, D.R. 1998: Model selection and inference: A practical information-theoretic approach. - Springer, New York, New York, USA, 353 pp.

Clawson, M.V. 2010: Use of age-at-harvest information to inform wildlife management. - M.Sc. thesis, University of Washington, Seattle, USA, 188 pp.

Collett, D. 1991: Modelling binary data. - Chapman & Hall/CRC, New York, New York, USA, 369 pp.

Conn, P.B., Diefenbach, D.R., Laake, J.L., Ternent, M.A. & White, G.C. 2008: Bayesian analysis of wildlife age-at-harvest data. - Biometrics 64: 1170-1177.

Cox, D.R. & Snell, E.J. 1968: A general definition of residuals. - Journal of the Royal Statistical Society, Series B 30: 248-275.

Draper, N.R. & Smith, H. 1998: Applied regression analysis. - Wiley & Sons, New York, New York, USA, 706 pp.

Fieberg, J.R., Shertzer, K.W., Conn, P.B., Noyce, K.V. & Garshelis, D.L. 2010: Integrated population modeling of black bears in Minnesota: Implications for monitoring and management. - PLoS ONE 5(8): e12114.

Gove, N.E., Skalski, J.R., Zager, P. & Townsend, R.L. 2002: Statistical models for population reconstruction using age-at-harvest data. - Journal of Wildlife Management 66: 310-320.

Hilborn, R. & Walters, C.J. 1992: Quantitative fisheries stock assessment: Choice, dynamics, and uncertainty. - Chapman and Hall, New York, New York, USA, 570 pp.

Hogg, R.V. & Craig, A.T. 1995: Introduction to Mathematical Statistics. - Prentice Hall, Englewood Cliffs, New Jersey, USA, 564 pp.

Johnson, J.B. & Omland, K.S. 2004: Model selection in ecology and evolution. - Trends in Ecology and Evolution 19: 101-108.

Kotz, S., Johnson, N.L. & Read, C.B. 1985: Lindeberg conditions to multitrait-multimethod matrices, Volume 5, Encyclopedia of Statistical Sciences. - Wiley-Interscience, Malden, Massachusetts, USA, 752 pp.

Lytle, D.A. 2002: Flash floods and aquatic insect life-history evolution: Evaluation of multiple models. - Ecology 83: 370-385.

McCullagh, P. & Nelder, J.A. 1989: Generalised linear models. - Chapman & Hall, London, UK, 532 pp.

National Research Council Committee on Fish Stock Assessment Methods 1998: Improving fish stock assessments. - National Academy Press, Washington, D.C., USA, 177 pp.

Neter, J., Kutner, M.H., Nachtsheim, C.J. & Wasserman, W. 1996: Applied linear statistical models. - Irwin, Chicago, USA, 720 pp.

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Parma, A. 1993: Retrospective catch-at-age analysis of Pacific halibut: Implications on assessment of harvesting policies. - In: Kruse, G., Eggers, D.M., Marascom, R.J., Pautzke, C. & Quinn, T.J. (Eds.); Management strategies for exploited fish populations. Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, Fairbanks, Alaska, USA, Sea Grant Program Report 93-02: 247-289.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. 2007: Numerical recipes: The art of scientific computing. - Cambridge University Press, New York, New York, USA, 1235 pp.

Schwarz, G. 1978: Estimating the dimension of a model. - Annals of Statistics 6: 461-464.

Seber, G.A.F. 1982: The estimation of animal abundance. - MacMillan, New York, New York, USA, 654 pp.

Skalski, J.R., Millspaugh, J.J., Clawson, M.V., Belant, J.L., Etter, D.R., Frawley, B.J. & Friedrich, P.D. 2011: Abundance trends of American martens in Michigan based on statistical population reconstruction. - Journal of Wildlife Management 75: 1767-1773.

Skalski, J.R., Ryding, K.E. & Millspaugh, J.J. 2005: Wildlife demography: Analysis of sex, age, and count data. - Academic Press, San Diego, California, USA, 636 pp.

Skalski, J.R., Townsend, R.L. & Gilbert, B.A. 2007: Calibrating population reconstruction models using catch-effort and index data. - Journal of Wildlife Management 71: 1309-1316.

Downloaded From: https://complete.bioone.org/journals/Wildlife-Biology on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use