



A Simulation-Based Method to Determine Model Misspecification: Examples using Natural Mortality and Population Dynamics Models

Author: Piner, Kevin R.

Source: Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science, 3(1) : 336-343

Published By: American Fisheries Society

URL: <https://doi.org/10.1080/19425120.2011.611005>

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.

ARTICLE

A Simulation-Based Method to Determine Model Misspecification: Examples Using Natural Mortality and Population Dynamics Models

Kevin R. Piner*

National Oceanic and Atmospheric Administration–Fisheries, Pacific Islands Fisheries Science Center, 2570 Dole Street, Honolulu, Hawaii 96822, USA

Hui-Hua Lee

Joint Institute for Marine and Atmospheric Research, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA

Mark N. Maunder

Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, California 92037, USA

Richard D. Methot

National Oceanic and Atmospheric Administration–Fisheries, Office of Science and Technology, 2725 Montlake Boulevard East, Seattle, Washington 98112, USA

Abstract

Recent developments in the models used in wildlife and fisheries science have allowed the inclusion of a wider range of data than previously. However, the diagnostics of such complex models have not kept pace. We describe a new diagnostic technique based on simulation analysis. Model misspecification was identified through simulation methods that created a distribution of likely parameter values for a model that was correctly specified. If the actual estimate of that parameter is outside the bounds of the simulated distribution, then the model is probably misspecified. We tested the reliability of the new diagnostic by introducing known-model misspecification into complex fisheries stock assessment models. We then compared the results from this new diagnostic with those of a more traditional fisheries diagnostic, namely, retrospective analysis. The simulation-based diagnostic was shown to identify misspecification affecting the estimated dynamics more reliably than retrospective analysis.

In the last few decades, an explosion has occurred in the development of models used in wildlife and fisheries science. This has been facilitated by advances in computers and mathematical and statistical algorithms (Maunder et al. 2009). These models have allowed for a more comprehensive inclusion of various types of information and a wider range of hypotheses to be investigated. The inclusion of additional information re-

sults in models that are parameter rich, nonlinear, and dynamic, which greatly complicates parameter estimation. New parameter estimation algorithms, such as those based on automatic differentiation (Skaug and Fournier 2006; Fournier et al., in press) and Markov Chain Monte Carlo (Hastings 1970), make the parameter estimation possible (Schnute et al. 2007). However, the well-studied properties of simpler traditional models

Subject editor: Ana Parma, Centro Nacional Patagonica, Puerto Madryn, Chubut, Argentina

*Corresponding author: kevin.piner@noaa.gov

Received December 6, 2010; accepted March 27, 2011

are generally not understood for these relatively more complicated analyses. The diagnostics of complex model performance (Lynch and Western 2004) has not kept pace with their development.

Model misspecification results from incorrect assumptions about parameters (e.g., fixing them at incorrect values) and governing processes (e.g., the stock–recruit relationship) or, perhaps more rarely, the statistical properties (e.g., variance and error distributions) used to compare model expectations with observations. There are no universally accepted measures of model misspecification, standard goodness of fit tests, or even common interpretation of residuals among data sets. Determination of model performance is often based on the experiences of the analyst. Diagnostics that are used are subject to misinterpretation and difficult to explain to target audiences. It is therefore important to develop a standard set of diagnostics for these models that will improve their performance and acceptance.

Analysis of residuals is perhaps the most common method of evaluating model performance (Cox and Snell 1968). Residuals are examined to ensure that they comply with the model assumptions (e.g., Wang et al. 2009) and are uncorrelated. Violation of these assumptions indicates that the model is misspecified. Statistical tests are available to evaluate the residuals, but visual examination is often used. Complicating the situation, analysis of residuals can be difficult for some likelihood functions and can be further complicated if there are also a large number of residuals across many data sources.

Examining the correlation among parameter estimates is also used to diagnose problems within the model. Correlation estimates use a multivariate normal assumption; therefore, bivariate profile likelihoods or bivariate plots from Bayesian Markov chain–Monte Carlo analysis (Hastings 1970; Punt and Hilborn 1997) may provide a more accurate representation of the correlation and uncertainty.

A diagnostic method—posterior predictive checks—is often recommended (Anderson et al. 2001; Lynch and Western 2004) for checking the performance of Bayesian models (Rubin 1984; Gelman et al. 1995). Where a test statistic of the observed data are compared with a test statistic of predicted data based on the posterior distribution of the parameters, a bias between the observed and simulated test statistic can be interpreted as evidence of model misspecification. The Bayesian method allows for formalizing the description of the probability (*P*-values) that the data match the model's assumptions (Meng 1994; Lynch and Western 2004).

Retrospective analysis is another common approach widely used in fisheries modeling to evaluate the reliability of current estimates of key derived quantities and parameters (Cadigan and Farrell 2005; Cadrin and Vaughn 1997). Retrospective analysis involves rerunning the model by consecutively removing successive years of data. It is assumed that as more data are available, the estimates from prior years converge towards the true value. Comparisons of estimates obtained with a reduced number of years with those based on all years of data are used to indicate bias in the recent estimates. If retrospective patterns exist it is

diagnostic of model misspecification. Often, the magnitude and direction of the bias indicated by the retrospective pattern are used to adjust the estimates in an attempt to better represent reality (Northeast Fisheries Science Center 2008).

In this paper, we introduce another approach to identifying model misspecification by using simulation methods. The premise of the diagnostic is that simulation analysis can build a distribution of likely parameter values for a model that is correctly specified, but if in the actual application the parameter estimate is outside the bounds of the simulated distribution, then the model is probably misspecified. We test the proposed model diagnostic via simulations based on examples using parameter-rich, complex, nonlinear stock assessment models and the parameter natural mortality.

METHODS

Diagnostic method.—The general steps for performing the simulation-based diagnostic described below are summarized in Figure 1.

Step 1: Choose a fixed (not estimated) parameter from the model to be the test parameter (*p*). Fit the original model to the original data set (*D_o*), and use the resulting parameter estimates as the true values (we use “true” in this paper to mean representing the actual population dynamics) for the simulation. Simulate *k* new data sets (*D_k*) based on the original model (all parameters either fixed or estimated the same as the original model) and parametric sampling of the appropriate error. Because each *D_k* is simulated from the model's expectations and statistical assumptions, this eliminates any model misspecification.

Step 2: Fit the model to each *D_k* via the same model parameterization as the original model, except that parameter (*p*) is estimated. The resulting estimates of *p_k* are subsequently used to describe the distribution of *p* when the data conform to the model's assumptions.

Step 3: Fit the model to *D_o* using the same model parameterization as the original model, except that *p* is again freely estimated (*p_o*).

Step 4: Determine the evidence of model misspecification using a two-sided test. Reject the null hypothesis of correctly specified model at significance level α if $p_o < p_{\alpha/2}$ or if $p_o > p_{1 - \alpha/2}$, where p_x is the *x*th percentile of the ordered values of *p_k*.

Validation of diagnostic via known-model misspecification.—To verify that the simulation diagnostic is capable of detecting model misspecification in complex models, we evaluated our method using two real-world applications of fisheries stock assessment models with controlled model misspecification and $\alpha = 0.1$. In the first application, we introduced misspecification of the steepness (*h*) parameter of the Beverton–Holt spawner–recruit function (Mace and Doonan 1988), using as an example the stock assessment (Hamel 2008) of darkblotched rockfish *Sebastes cramerii*. In the second application, we evaluated the misspecification of fishery selectivity pattern using a stock assessment (Field 2008) of chilipepper *Sebastes goodei*. Both assessments are relatively complex with many data types

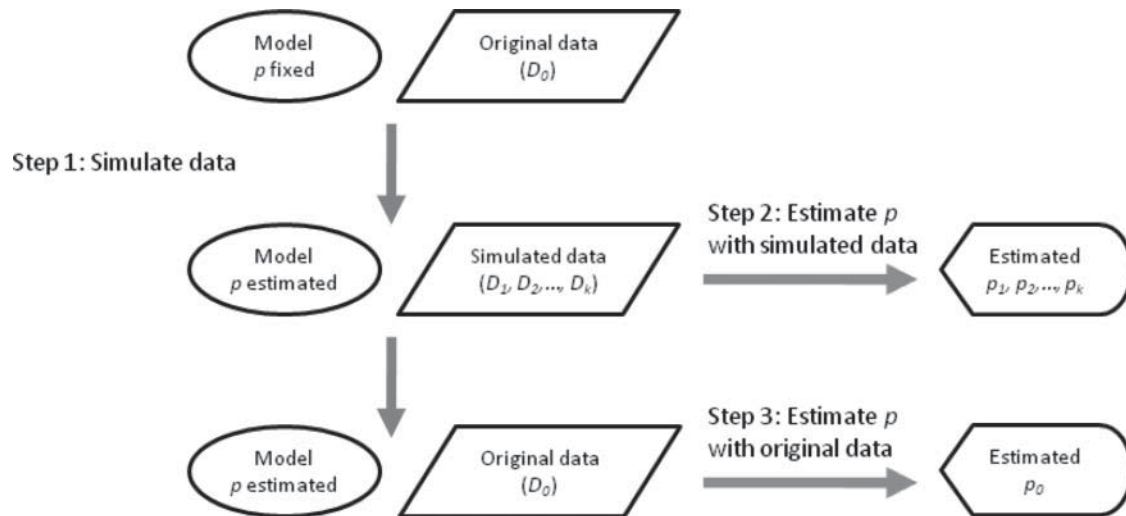


FIGURE 1. Flow chart depicting the general steps of the simulation-based diagnostic model. Step 4 comparing p_o (the estimate of parameter p using original data set) and the distribution of p_k (the estimate of parameter p using simulated data sets) is omitted.

and model processes, and both were completed using the stock synthesis (SS) population dynamics model (Table 1). In both examples we chose natural mortality as the test parameter p , which was a fixed parameter based on life history and maximum age methods.

TABLE 1. Summary of the data types included (yes, no), parameters estimated (est.) or fixed, and version of the stock synthesis model used in the stock assessments that formed the basis of the simulations for darkblotched rockfish and chilipepper.

Variable	Darkblotched rockfish	Chilipepper
Data		
Catch	Yes	Yes
Indices (survey)	Yes	Yes
Age data	No	Yes
Length data	Yes	Yes
Age at length	Yes	No
Mean length at age	No	No
Mean body weight	No	No
Parameterization		
Unfished recruitment (R_0)	Est.	Est.
Recruitment variability (σ_r)	0.8	1.0
Natural mortality (M /year)	0.07	0.16
Steepness (h)	0.6	0.57
Maximum size (L_∞ ; cm)	Est.	47.3
Brody growth coefficient (k)	Est.	0.19
Length CV	Est.	0.06
Model		
SS version	v2.00g	v2.00c
Number of estimated parameters	69	78

In the application using the darkblotched rockfish assessment, before applying the simulation steps outlined above we first simulated three data sets to act as replicate D_o from the original model's parameter estimates (Figure 2). Each D_o was based on the assumption that $h = 0.6$, and each data set varied based on the parametric sampling of the error. However, because each D_o was simulated from the original model's parameter estimates and assumptions, the created data sets were free of model misspecification. Elimination of any potential misspecification allowed us to later introduce a controlled misspecification to test the diagnostic method. Thus, the original assessment model fit to each of the three D_o acted as our replicate true values for this validation test.

Subsequently, we refit each replicate D_o with a model where h was specified as 0.3, 0.9 and 0.6 and generated 500 D_k as in step 1. In addition, to avoid potential process-error bias, we added random recruitment deviations in the fitting that had the same variability (σ_r) assumed in the original assessment model. This produced nine sets of 500 D_k ($3D_o \times 3$ levels of h). We then followed step 2 described above by fitting to each D_k , via a model parameterized with the same alternative h used to generate that D_k . The estimates of p_k based on data set D_k were used to determine how well the estimator performs if the true dynamics had alternative assumptions about h . In step 3, we estimated p_o via each replicate D_o and a model with each alternative h . Thus, in step 3 the models contained a single known misspecification of $h = 0.3$ or $h = 0.9$, but models with $h = 0.6$ were correctly specified and acts as our controls. Prior to step 4, all model runs with gradients greater than 1 were assumed to have not converged and were removed from further analysis. Step 4 was then performed as described above for each combination of replicate and alternative h .

We applied the same procedure using the chilipepper stock assessment model that assumed a dome-shaped selectivity

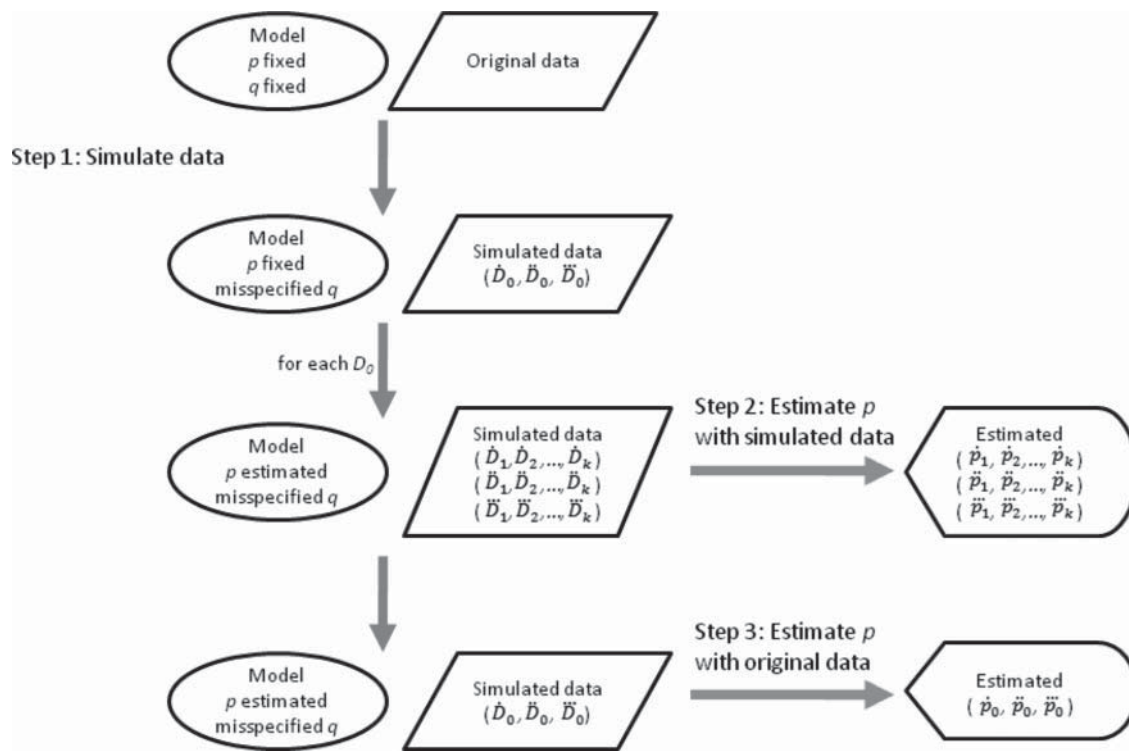


FIGURE 2. Flow chart depicting the steps of the validation of the model diagnostic. Parameter p is the test parameter, and q represents the controlled misspecification in the validation. Step 4 comparing p_o and the distribution of p_k is omitted.

pattern to produce three replicate D_o . In step 1 we fit models by assuming domed and asymptotic fishery selectivity patterns and generated 500 D_k . This produced six sets of D_k ($3D_o \times 2$ selectivity patterns). We then performed the rest of the validation the same as in the darkblotched rockfish example, where the asymptotic selectivity pattern representing the misspecification and the domed pattern as the control.

Comparison to retrospective analysis.—We compared our detection of model misspecification by using the simulation-based diagnostic with the commonly used retrospective analysis. In this case, a 10-year retrospective analysis was conducted on the same assessment models (three levels of h or two levels of selectivity pattern) fit to each D_o by sequentially eliminating 2 years of data. Misspecification was defined as at least four out of five retrospective runs estimating consistently higher or lower terminal year biomass (relative to the original model) with each successive removal of 2 years of data.

RESULTS

Known-Model Misspecification

Steepness.—Given the model structure and data of the darkblotched rockfish assessment, there appeared to be information on p (natural mortality in our illustration). The distribution of p_k was relatively unbiased (bias of about 5%) when D_k was generated using the alternative levels of h (Figure 3). The distribution of p_k was slightly wider for $h = 0.9$ (CV, about 18%; Figure 3C)

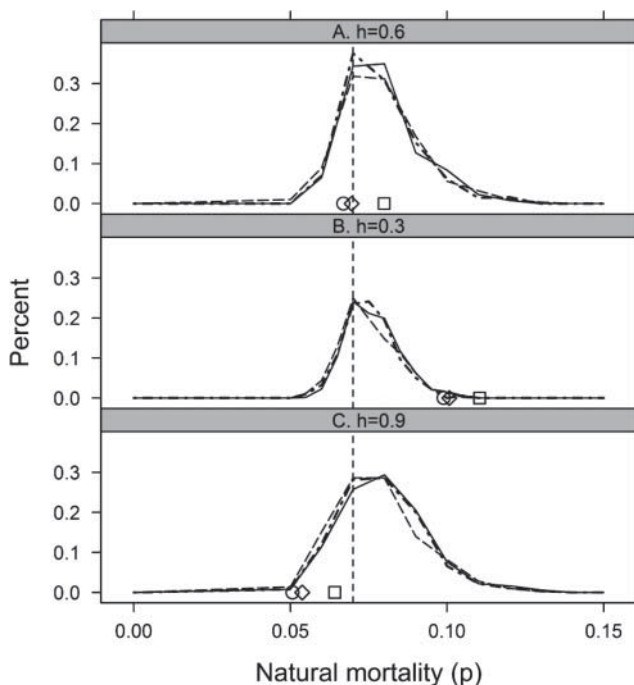


FIGURE 3. Distributions of natural mortality (p_k) estimated from three replicate models with spawner–recruit steepness (h) fixed at (A) 0.6, (B) 0.3, and (C) 0.9. Each replicate distribution is the result of 500 simulated data sets (D_k). The vertical dotted line is the assumed value of p , which was used to generate the data files. The circles, squares, and triangles are the estimates of p from the three original replicate data files (p_o).

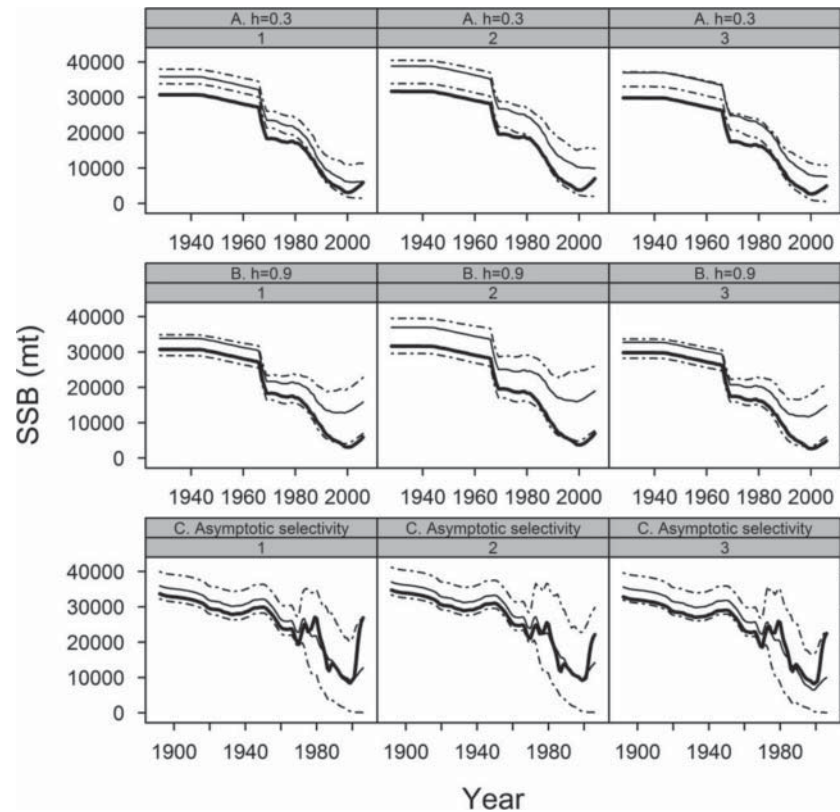


FIGURE 4. Average spawning biomass (SSB [metric tons]) and the 10th and 90th percentiles from three replicates of 500 data sets fit with steepness (h) fixed at (A) 0.3 and (B) 0.9 and (C) and the asymptotic selectivity pattern. The true spawning biomass is denoted by the heavy line.

than $h = 0.3$ (CV, about 12%; Figure 3B), while $h = 0.6$ was intermediate (CV, about 16%; Figure 3A). The estimate of p_o , based on a misspecification of $h = 0.3$, was located in the tails of the distribution of p_k in all three replicates (Figure 3B). The estimate of p_o based on a misspecification of $h = 0.9$ was located in the tails of the distribution of p_k in two of the three replicates (Figure 3C). Thus, in nearly all cases we would have correctly diagnosed the model estimating p_o as misspecified. In the control without misspecification ($h = 0.6$) the estimate of p_o was near the center of the distribution of p_k in all cases (Figure 3A). Thus, we would have correctly inferred that all three replicates with $h = 0.6$ were not misspecified.

Misspecification of h altered our perception of the population dynamics. Generation of data sets reflecting a less resilient stock ($h = 0.3$) had significant impact on the population dynamics, as measured by the time series of spawning biomass relative to the true model. The true value of spawning biomass (based on $h = 0.6$) often lay outside the 90th percentiles of the simulated distribution and resulted in a large bias between true and the average simulated estimates (Figure 4A). The generation of data sets reflecting a more resilient stock ($h = 0.9$) also had a significant impact on the dynamics, a large and consistent bias occurring between the true and the average simulated estimates of spawning biomass. However, unlike the less resilient sce-

nario, only the most recent estimates of true spawning biomass were outside the 90th percentiles (Figure 4B).

Selectivity pattern.—Given the model structure and the data of the chilipepper assessment, there appeared to be strong information on the magnitude of p . The distribution of p_k was unbiased (bias < 0.1%) when D_k was generated using the asymptotic fishery selectivity pattern. This result was similar to that produced with the domed fishery selectivity pattern. The distribution of p_k was very precise (CV of about 5%) when estimated from D_k via the asymptotic fishery selectivity pattern, as well as from the dome-shaped pattern. The estimate of p_o appeared in the tails of the distribution of p_k in none of the three replicates, failing to diagnose the model estimating p_o as misspecified (Figure 5B). In one replication however, the estimate of p_o was greater than 0.94% of the distribution of p_k , and we might interpret this as borderline evidence of misspecification. In the control without misspecification the estimate of p_o was near the center of the distribution in two of the three replicates (Figure 5A). Thus, we would generally have correctly inferred the models were not misspecified. However, in one case we would have committed a type II error and incorrectly inferred misspecification.

The misspecification of an asymptotic fishery selectivity pattern had minimal effects on the population dynamics, as

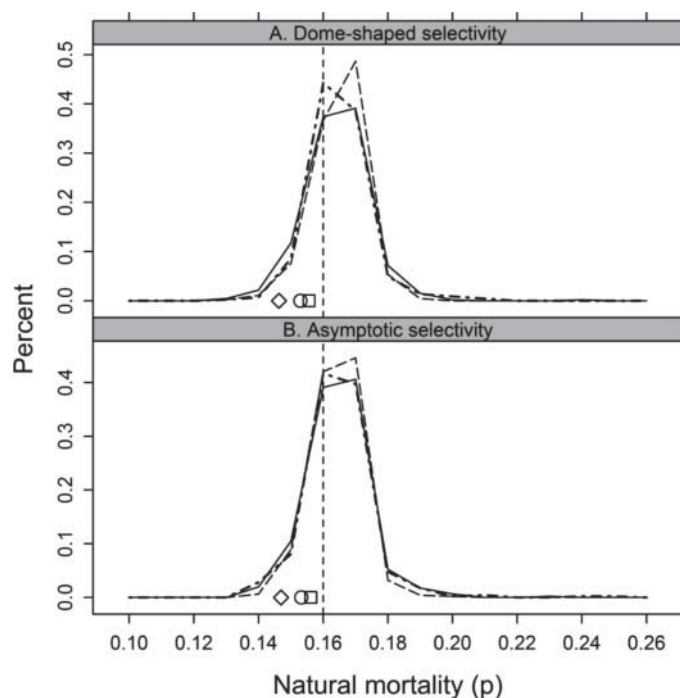


FIGURE 5. Distribution of natural mortality (p_k) estimated with three replicates, where the fishery selectivity pattern is assumed to be (A) domed shaped or (B) asymptotic. Each replicate distribution is the result of 500 simulated data sets (D_k). The vertical dotted line is the assumed value of p used to generate the data files. The circles, squares, and triangles are the estimates of p from the three original replicate data files (p_o).

measured by the time series of spawning biomass relative to the true model. The true value of spawning biomass was contained within the 90th percentile of the simulated spawning biomass distribution. The bias between the average of the simulated estimates and true spawning biomass was generally quite small and not consistent (Figure 4C).

Comparison with Retrospective Analysis

Retrospective analysis was not as good a diagnostic as the simulation-based method in the darkblotched rockfish example. Retrospective analysis correctly showed no misspecification when h was correctly specified in two of the three replicates; however, we would have committed a type II error by incorrectly inferring model misspecification in one replicate (Figure 6A, replicate 3). Retrospective analysis incorrectly indicated model misspecification in none of the three replicates when p_o was estimated based on misspecification of $h = 0.3$ (Figure 6B). Retrospective analysis correctly indicated misspecification in one of the three replicates when p_o was estimated based on a misspecification of $h = 0.9$ (Figure 6C, replicate 3).

Retrospective analysis also fared poorly in the chilipepper example. The retrospective analysis correctly indicated no misspecification when the selectivity pattern was correctly specified as domed (Figure 7A). However, retrospective analysis indicated model misspecification in none of the three replicates when p_o

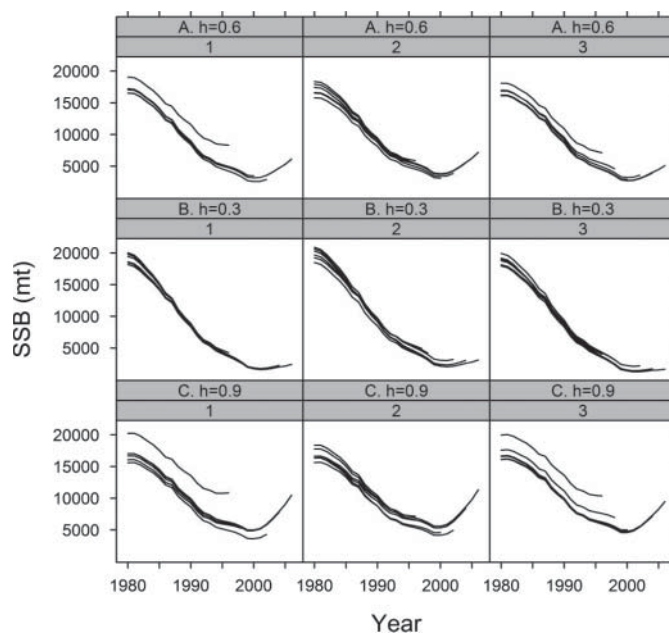


FIGURE 6. Plots of the retrospective analysis for the three replicate trials with spawner–recruit steepness (A) correctly specified as 0.6, (B) incorrectly specified as 0.3, and (C) incorrectly specified as 0.9.

was estimated based on a misspecification of an asymptotic selectivity pattern (Figure 7B).

DISCUSSION

This simulation-based diagnostic is similar to posterior predictive checks (Gelman et al. 1995) in accounting for the parametric uncertainty of the model (Lynch and Western 2004). It differs from posterior predictive checks in that instead of testing the equivalence of observed and implied data we tested the equivalence of parameters. This simulation-based method is also applicable to both Bayesian and non-Bayesian models. The proposed diagnostic also shares similarity to randomization methods (Solow 1993; Adams and Anthony 1996; Helser 1996), except that instead of randomizing the observed data we have parametrically simulated data based on the models structure and estimated parameters. All three approaches are similar in that rejection of the null hypothesis is based on how far into the tails of the distribution the observed statistic lies. The choice of α level is dependent on the risk of type II error and the sample size needed to adequately describe the tails (Manly 1991). Although we have presented the results as an acceptance or rejection of the null hypothesis of a correctly specified model, the results could also be interpreted as the strength of evidence (Manly 1991) for rejecting the null hypothesis.

In addition to identifying misspecification, the simulation results may also hold clues as to the direction of the misspecification. In the case of misspecifying h to be more resilient, it resulted in an estimate of p_o that was less productive (lower

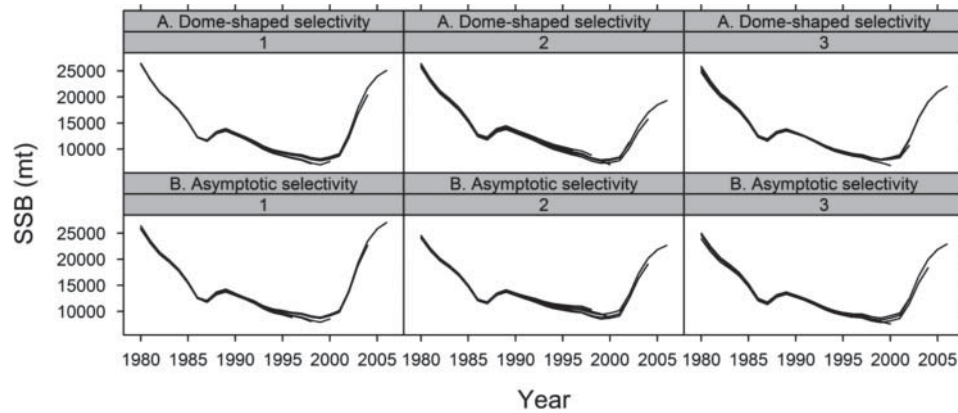


FIGURE 7. Plots of the retrospective analysis for the three replicate trials with the fishery selectivity pattern (A) correctly specified as domed shaped and (B) incorrectly specified as asymptotic. The plots are from a 10-year retrospective analysis that sequentially excluded the last 2 years of data.

than assumed in the true model and vice versa. It appears that the model is simply trading off productivity in one process (recruitment) with productivity in another (mortality). We also note that plotting the distribution of each component's likelihood from the simulation and from p_o estimation model may also give clues to the location of structural issues by identifying which components are fitting much better when p is allowed to change freely.

The greater the effect of the misspecification on model results, the more reliably the proposed diagnostic detected the misspecification. In the case of misspecifying the fishery-selectivity pattern model, dynamics were not greatly influenced. The simulations were based on a stock-assessment model that included multiple fisheries and surveys, and the misspecification of the chosen selectivity pattern had only minor effects on the dynamics. Resulting estimates of p_o from the misspecified and correct models were correspondingly almost identical and, thus, neither the retrospective nor the simulation-based diagnostic appeared to be reliable at determining model performance. Misspecifications that have little effect on model expectations will not be easily detected with these or any diagnostic methods.

In our examples, natural mortality was an ideal choice for the test parameter p because it is correlated with nearly all aspects of the estimated population dynamics. Thus, any misspecification in the parameterization of the model that affects the estimated dynamics can be at least partially offset by changes in natural mortality. In our validation examples, it did not matter how p was originally specified because we controlled the introduction of the misspecification. However, parameters are not appropriate choices for p if the parameters are freely estimated or fixed at values close to the maximum likelihood estimate via an iterative process because the process of setting them has already accounted for the model misspecification. Researchers would also need to keep in mind that any diagnosed misspecification could be the result of p itself having been misspecified. If the model is shown to be misspecified and the estimate of p_o is unrealistic based on prior knowledge of a likely range of that

parameter, then it is likely that some other aspect of the model beyond the assumption of p was misspecified. However even in these cases, researchers could not conclude that p was not also misspecified.

The proposed simulation-based diagnostic performed as well as or better than the retrospective analysis at discovering model misspecification in these fisheries stock-assessment examples. We do not propose that this simulation-based approach replace the more traditional diagnostics (residual analysis, correlation, etc.) or that in all cases this is a better diagnostic than retrospective analysis. This simply is one more alternative that may be most useful when dealing with models containing many likelihood components that make interpretation of residuals across likelihood components problematic. It should be considered that this diagnostic method is more computer intensive than some of the more traditional diagnostic tools. However, that cost is balanced by its ability to diagnose misspecifications and the relative simplicity of the characterization of model performance.

ACKNOWLEDGMENTS

We thank our colleagues at the Northwest Fisheries Science Center and especially Stacey Miller for collating and providing these stock assessments. The authors would also like to thank the Southwest Fisheries Science Center that supported some of the authors during early stages of the work. In addition, we want to thank the scientists of the International Scientific Committee for providing early critiques of this paper. We also thank Ana Parma and Andre Punt for comments on the manuscript.

REFERENCES

- Adams, D. C., and C. D. Anthony. 1996. Using randomization techniques to analyze behavioral data. *Animal Behavior* 51:733–738.
- Anderson, D. R., W. A. Link, D. H. Johnson, and K. P. Burnham. 2001. Suggestions for presenting the results of data analyses. *Journal of Wildlife Management* 65:373–378.

- Cadigan, N. G., and P. J. Farrell. 2005. Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES Journal of Marine Science* 62:256–265.
- Cadrin, S. X., and D. S. Vaughn. 1997. Retrospective analysis of virtual population estimates for Atlantic menhaden stock assessment. *U.S. National Marine Fisheries Service Fishery Bulletin* 95:445–455.
- Cox, D. R., and E. J. Snell. 1968. A general definition of residuals. *Journal of the Royal Statistical Society Series B* 30:248–275.
- Field, J. 2008. Status of the chilipepper rockfish, *Sebastes goodei*, in 2007. In *Status of the Pacific Coast groundfish fishery through 2008, stock assessment and fishery evaluation: stock assessments, STAR panel reports, and rebuilding analyses*. Pacific Fishery Management Council, Portland, Oregon.
- Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, and J. Sibert. In press. AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian data analysis*. Chapman and Hall, London.
- Hamel, O. S. 2008. Status and future prospects for the darkblotched rockfish resource in waters off Washington, Oregon, and California as assessed in 2007. In *Status of the Pacific coast groundfish fishery through 2008: stock assessment and fishery evaluation—stock assessments, STAR panel reports, and rebuilding analyses*. Pacific Fishery Management Council, Portland, Oregon.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
- Helser, T. E. 1996. Growth of silver hake within the U.S. continental shelf ecosystem of the northwest Atlantic Ocean. *Journal of Fish Biology* 48:1059–1073.
- Lynch, S. M., and B. Western. 2004. Bayesian posterior predictive checks for complex models. *Sociological Methods and Research* 32:301–335.
- Mace, P. M., and I. J. Doonan. 1988. A generalized bio-economic simulation model for fish population dynamics. New Zealand Fisheries Research Centre, Fisheries Assessment Research Document 88/4, Wellington.
- Manly, G. F. J. 1991. *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London.
- Maunder, M. N., J. T. Schnute, and J. Ianelli. 2009. Computers in fisheries population dynamics. Pages 337–372 in B. A. Megrey and E. Moksness, editors. *Computers in fisheries research*. Springer-Verlag, New York.
- Meng, X. L. 1994. Posterior predictive p -values. *Annals of Statistics* 22:1142–1160.
- Northeast Fisheries Science Center. 2008. Assessment of 19 Northeast groundfish stocks through 2007: report of the 3rd groundfish assessment review meeting (GARM III). National Oceanic and Atmospheric Administration Fisheries, Northeast Fisheries Science Center, Reference Document 08-15, Woods Hole, Massachusetts.
- Punt, A. E., and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* 7:35–63.
- Rubin, D. B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* 12:1151–1172.
- Schnute, J. T., M. N. Maunder, and J. Ianelli. 2007. Designing tools to evaluate fishery management strategies: can the scientific community deliver? *ICES Journal of Marine Science* 64:1077–1084.
- Solow, A. R. 1993. A simple test for change in community structure. *Journal of Animal Ecology* 62:191–193.
- Skaug, H., and D. Fournier. 2006. Automatic approximation of the marginal likelihood in nonlinear hierarchical models. *Computational Statistics and Data Analysis* 51:699–709.
- Wang, S. P., M. N. Maunder, and A. Aires-da-Silva. 2009. Implications of model and data assumptions: an illustration including data for the Taiwanese longline fishery into the eastern Pacific Ocean bigeye tuna (*Thunnus obesus*) stock assessment. *Fisheries Research* 97:118–126.