Discriminating alternative stock–recruitment models and evaluating uncertainty in model structure

Shijie Zhou *

CSIRO Marine and Atmospheric Research, GPO Box 120, Cleveland, Qld 4163, Australia

Abstract

Fish stock–recruitment (S–R) assessment is one of the most essential keystones for fisheries management. Yet the analysis involves a variety of uncertainties. Amidst these difficulties, uncertainty in model structure is perhaps the most problematical to investigate because no rigorous statistical techniques can be used to explore the fundamental biological processes in S–R relationships. In this paper, I used computer simulations to investigate: (1) the differences between the estimated parameters of alternative S–R models as a function of stock characteristics: population growth rate, data range, fishing mortality, and process noise; and (2) the probability of selecting a correct model using information criteria. Two popular S–R functions, the Ricker and the Beverton–Holt models, were used as examples. Time series data were generated from a known S–R model and fitted by alternative models. The results show that when the two models fit the data similarly well, significant differences in parameters existed between the alternative models. The Ricker model tended to underestimate the population growth rate (initial slope) and the carrying capacity parameter, whereas the Beverton–Holt model overestimated these parameters. The management quantities (e.g., optimal virgin stock size) produced by one model were more conservative (i.e., larger optimal stock size or lower optimal harvest rate) under some conditions but became less conservative under other conditions. The differences between the alternative models were functions of the population growth rate, long-term fishing mortality, and data range of the stock size. The correct and incorrect models were statistically indistinguishable. For typical fishery data the probability of selecting the correct model based on information criteria was approximately 0.70 for the Ricker model and 0.61 for the Beverton–Holt model.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Stock–recruitment; Model uncertainty; Ricker; Beverton–Holt

1. Introduction

Fish stock–recruitment (S–R) analysis is among the most important assessments of fish population dynamics. Generally, fisheries scientists believe that there is a fundamental relationship between spawner abundance and subsequent recruitment (Myers and Barrowman, 1996). Because of its importance in fishery management, a great deal of effort has been devoted to studies of S–R relationships (Hilborn and Walters, 1992; Quinn and Deriso, 1999; Haddon, 2001). Many fishery management reference points and policies are based on such a relationship (Kimura, 1988; Clark, 1991; Myers et al., 1994; Bradford et al., 2000; Brodziak, 2002). For example, escapement goals for many Pacific salmon stocks are established based on S–R analysis

(CTC, 1999). Scientists routinely conduct stock assessments for a variety of ground fish species to develop annual harvest guidelines and rebuilding plans for stocks along the West Coast of the continental United States (PFMC, 2002; Ralston, 2002). Several mathematical models have been built to describe the relationship between spawning stock and recruitment at various life stages. These models include Beverton and Holt (1957), Ricker (1954), Cushing (1971, 1973), Deriso (1980), Schnute (1985), Shepherd (1982), Gamma (Reish et al., 1985), and recently the hockey-stick models (Barrowman and Myer, 2000; Bradford et al., 2000).

Fishery research and management includes a variety of uncertainties, ranging from a lack of basic data to institutional inefficiency (Rice and Richards, 1996; Francis and Shotton, 1997; Flaaten et al., 1998; Weeks and Berkeley, 2000). There are at least three major uncertainties involved in S–R modeling: process noise, measurement error, and model uncertainty (Charles, 1998; Schnute and Richards, 2001). The first two types of uncertainties have been extensively studied (Ludwig

* Correspondence address: 233 Middle Street, P.O. Box 120, Cleveland, Qld 4163, Australia. Tel.: +61 7 3826 7263; fax: +61 7 3826 7222.

E-mail address: shijie.zhou@csiro.au.

and Walters, 1981; Walters and Ludwig, 1981; Hilborn and Walters, 1992; Ehrhardt and Legault, 1997; Quinn and Deriso, 1999; Valpine and Hastings, 2002). These two types of uncertainties can be quantitatively captured and evaluated by modern statistical methods. The third uncertainty, model uncertainty, is also called structural uncertainty or model misspecification. This uncertainty reflects a fundamental lack of knowledge about underlying fish biological processes. There are two basic types of model uncertainties. First, the assumed structural relationships between variables may be incorrect. Second, potentially influential relationships and dynamics may have been totally excluded from the model (Mace and Sissenwine, 2002). Because quantitative techniques cannot deal with model uncertainties, this topic is particularly challenging to fishery researchers. The appropriate form of the S–R relationship remains a significant issue of debate for many stocks (Patterson et al., 2001). Kimura (1988) showed that the particular family of S–R curves that is used can strongly affect the conclusions of a stock assessment. Levels of exploitation that appear near optimum under one family of recruitment curves may force a stock to extinction under another.

The literature on S–R relationships is substantial. However, we lack a systematic evaluation of the differences in the estimated parameters from different models. Fisheries management often depends on these estimated parameters and if alternative models produce inconsistent results the impact on management decisions could be profound. A typical S–R model is a mathematical equation in which recruitment is a function of spawner abundance. Most models were originally based on certain biological assumptions (Ricker, 1954; Beverton and Holt, 1957). However, scientists commonly fit a mathematical equation to observed stock size and recruitment data and estimate the parameters in the equation without considering the validity of the underlying biological assumptions. In practice, any mathematical model may be used for spawner–recruitment relationships as long as it goes through the origin, increases monotonically at low spawning levels, and shows some level of density dependence at high spawning levels (Quinn and Deriso, 1999). We may not understand the underlying biological processes in S–R relationships, or the relationships may be too complex to be described with a simple mathematical function. Statistical criteria, rather than biological considerations, have become the primary arbitrator to determine which model is suited for a particular species. This dependence on statistical fit could be misleading, particularly if the better fit of one model over others is spurious because of measurement and process error.

While some scientists continue to recommend using statistical criteria for selecting a best S–R model, others consider statistical criteria to be potentially misleading. Alternative models may fit a data set equally well and yet produce very different parameter estimates. For example, when investigating different models to estimate the threshold biomass required to ensure protection of recruitment, Myers et al. (1994) found dramatically different results for the Ricker model versus the Beverton–Holt model, even though the goodness-of-fit was rarely different (Barrowman and Myer, 2000). However, empirical analysis of 72 real datasets from a variety of fish stocks using the Ricker and Beverton–Holt models resulted in a mixed picture on

which S–R model was more conservative (Myers et al., 1994). When comparing spawner–recruitment curves for coho salmon, Barrowman and Myer (2000) concluded that the Beverton–Holt model always overestimated the initial slope and carrying capacity. In an extensive simulation study that compared biological reference points, Williams and Shertzer (2003) concluded that very different biological reference points emerged from the Beverton–Holt and Ricker functions. They found that with typical stock–recruitment data it was very difficult to determine which function best represented the behavior of a particular stock, although they recommended that the Beverton–Holt function was a better choice for management because of its more conservative biological reference point values. Wang and Liu (2006) compared two criteria that may be used to select among S–R functions and found that both Akaike information criterion (AIC) and Bayesian information criterion (BIC) were valid. However, in another simulation study, Valpine and Hastings (2002) found that the Ricker model systematically fit the data better than the Beverton–Holt model even when the latter was used as the underlying model to generate the data. They cautioned that using information criteria (such as AIC) can be misleading. These inconsistent results can cause disagreement over which models should be used for a particular stock or management decision (Patterson, 1999; Hammond and O'Brien, 2001).

The objectives of this paper were two-fold. It is possible that one model may be more conservative than other models under certain circumstances. Therefore, the first objective was to systematically compare the differences between the estimated parameters from alternative S–R models under a variety of scenarios. This objective relied on selecting a time series dataset that could be fit by two or more competing models equally well. The second objective was to test whether the correct and incorrect models were statistically distinguishable. The correct model referred to the one that best represented the underlying function that produced S–R relationship, whereas the incorrect model was the alternative one. I used an approach reported in the literature of simulating a series of data from a known S–R function and then fitting alternative models to these data. Of the models mentioned above, the classical Beverton–Holt and Ricker models are most popularly used in fisheries research. Therefore, I chose to investigate these two models and compare their results. Obviously, other models could also be examined by this method. Both models have two parameters to be estimated. The Ricker model assumes, biologically, that the mortality rate of the recruitment is proportional to the stock abundance (Ricker, 1954, Hilborn and Walters, 1992). Consequentially, the curve is dome-shaped with a declining limb at high spawner abundance. In contrast, the Beverton–Holt model is based on the assumption that the mortality rate of the recruitment is linearly dependent upon the number of fish alive in the cohort at any time. As a result the curve is asymptotic in shape (Beverton and Holt, 1957; Hilborn and Walters, 1992). In this paper, spawner abundance or stock size refers to either the number of spawners, biomass, or index of either, while recruitment or production refers to the number, biomass, or index of offspring. This study is primarily pertinent to semelparous species such as Pacific salmon, but it should also

be applicable to iteroparous species since the underlying process is similar.

2. Methods

2.1. Generating time-series S–R datasets

I investigated model-structure uncertainty by first generating data from known models with known parameters and statistical errors. I used the two most common S–R models to generate data: the Ricker model and the Beverton–Holt model. If R_i is the recruitment at time i and S_i is the stock size that produces R_i, the Ricker model can be expressed as:

$$R_i = a^{Rk} S_i e^{-a^{Rk} S_i / (R_p^{Rk} e) + u_i} \tag{1}$$

and the Beverton–Holt model as

$$R_i = \frac{a^{BH} S_i}{1 + a^{BH} S_i / R_p^{BH}} e^{u_i}. \tag{2}$$

Both models include two parameters: a, the population growth rate when the population is small and R_p, the maximum production that describes the carrying capacity of the system. In the models all natural mortalities were included in the resulting recruitment, and the stock size S_i and recruitment R_i were assumed to be measured in the same units. The initial slope a and the maximum production R_p were set equal to the same values such that $a^{Rk} = a^{BH}$ and $R_p^{Rk} = R_p^{BH}$. A random process noise e^{u_i} was applied to each recruitment R_i. This noise was assumed to be multiplicative because this type of error structure is by far the most common assumption used in fitting S–R models (Peterman, 1981; Hilborn and Walters, 1992; Myers et al., 1999; Quinn and Deriso, 1999). Variable u_i followed a normal distribution with mean 0 and variance σ_u^2. A proportion of the recruits survived from fishing and became the spawners for the next generation:

$$S_{i+1} = R_i e^{-F + \varepsilon_i} \tag{3}$$

where F is the average long-term instantaneous fishing mortality and ε_i is a normally distributed random variable with mean 0 and variance σ_ε^2. In addition to process noise, the "observed" stock size s_i and "observed" recruitment r_i were both assumed to have a same level of measurement error e^{v_i}:

$$s_i = S_i e^{v_i} \quad \text{and}$$

$$r_i = R_i e^{v_i}. \tag{4}$$

The random variable v_i followed a normal distribution with mean 0 and variance σ_v^2. I assumed that both process and measurement errors had the same distribution so that they could be additive for approximate comparison with errors in real fishery data (Myers et al., 1995a,b; Quinn and Deriso, 1999; Berkson et al., 2002; Valpine and Hastings, 2002).

The stock size and recruitment were measured on an arbitrary but fixed scale. In this paper, the maximum production R_p was set to 2000 and a starting stock size of 1000 was used. For the Ricker model, there is a stock size S_p that produces a peak recruitment

R_p, but the spawners for the maximum production is infinite in the Beverton–Holt model. The model outputs are sensitive to the range of the data (Myers et al., 1994) so the range of stock size with regard to S_p may be an important factor affecting the estimated parameters. Therefore, I chose three population ranges where the ratio of $S_{upper}:S_p$ is 1, 2, and 3; or in other words, the upper limit of the observed stock size was 1–3 times that of the spawners that produced the maximum recruitment in the Ricker model. Fishery data typically falls within $S_{upper} \leq 3S_p$ (Myers et al., 1995a,b; CTC, 1999). The population growth rate parameter a used in the simulation ranged from 2 to 32 (log(a) ranged from 0.69 to 3.47) and increased exponentially by a factor of 2^x, where the increment step $x = 1, 2, \ldots$. These values are within the typical productivity range for Pacific salmon and many other fish species (Myers et al., 1995a,b, 1999; CTC, 1999; Chen and Holtby, 2002). Fishing mortality F controls the stock size and may be a factor that affects model outputs. I used F from 0.05 to 1.6 in increments that increased exponentially by a factor of 2^x. To create a spread of stock size over a reasonable range, I applied a relatively large random variable ε_i with $\sigma_\varepsilon = 0.5$. The random process noise expressed as standard error σ_u varied from 0.02 to 0.64 and also increased exponentially by a factor of 2^x. A small measurement error was set at $\sigma_v = 0.01$.

Initially this study design resulted in 450 simulation scenarios with different combinations of the data range, population growth rate, long-term fishing mortality, and level of process noise. For each scenario both the Ricker and Beverton–Holt models were run for 500 life cycles (iterations) before taking observations. The "observed" S–R data were drawn from a stationary distribution under each specific condition. One time series containing 20 pairs of S–R data points was taken from each model. This process was repeated 500 times to create 500 time series of S–R datasets for each simulation scenario. Since one of the objectives was to investigate the impact of the range of the data only those data that fell within the range ($s_i \leq S_{upper}$) under that specific scenario were retained for further analysis.

By incorporating the fishing mortality F into this simulation, it was possible to examine a wide range of stocks sizes with various growth rates. Some simulation scenarios did not produce valid S–R data because stocks become "extinct" before 500 generations. This happened in two circumstances: stocks with low growth rates exposed to high long-term fishing mortalities and stocks with high growth rates but exposed to low fishing pressures. The latter situation only occurred in the Ricker model because of its dome-shaped S–R relationship. If the Ricker model did not generate valid data under a specific scenario then the data from the Beverton–Holt model under the same scenario was also excluded from further analysis. This extinction behavior reduced the simulation scenarios to 322 and a total of about 236,500 time series of S–R datasets.

The artificial S–R data generated in this step were further treated in two ways for two purposes: to compare the estimated parameters of alternative models by combining time series data from two models, and to examine whether the correct and incorrect models are distinguishable by using time series data from one single model at a time, as described below.

2.2. Comparison of model parameters and management parameters of alternative models

A critical step in the comparison of alternative models is to create data sets that are fit by the competing models equally well. When generating data sets by the two models described in the previous section, the growth rate parameter a and the maximum production R_p were set to equal for the Ricker and Beverton–Holt models. In this next step of the analysis, the two data sets generated from the two models with the same combinations of known parameters, process and measurement errors were combined to form a mixed time-series data set. The new combined data set was then fit by both Ricker and Beverton–Holt models with unknown model parameters. The purpose of this step was to provide as equal a goodness-of-fit for the two models as possible. Although the mixed data did not represent either the Ricker model or the Beverton–Holt model, since approximately one half of the data were from each model, intuitively, the two models should fit the data equally well. Indeed, the mean sum squared error was essentially the same for each model when the data range S_{upper} was within $2S_p$. However, the Ricker model tended to be more robust and had a better fit when $S_{upper} = 3S_p$.

The maximum likelihood method was used to estimate three parameters in each model: the population growth rate a, the maximum recruitment R_p, and the standard deviation σ between the "observed" and estimated recruitment. Three management parameters were chosen to evaluate the discrepancy between the two models: the optimal spawner population S_{msy} for the maximum sustained yield, the optimal harvest rate U_{msy} for the maximum sustained yield, and the stock size $S_{20\%B_0}$ corresponding to the steepness (recruitment produced by 20% of virgin biomass B_0). S_{msy} and U_{msy} are often used as reference points for the management of Pacific salmon (*Oncorhynchus* spp.), while a reference point based on the percentage of virgin biomass is one of the most commonly adopted overfishing thresholds in fisheries literature (Myers et al., 1994).

Since I assumed the stock size and recruits were measured in the same units and I used a semelparous species as an example, the median management parameter $\hat{S}_{msy}$ (the hat indicates an estimate) for the Ricker model can be obtained taking the partial derivative of R with respective to S from Eq. (1), setting this equal to 1, and numerically solving the equation:

$$\left(1 - \frac{\hat{a}^{Rk}}{\hat{R}_p^{Rk}e}\hat{S}_{msy}^{Rk}\right)\hat{a}^{Rk}e^{-\hat{a}^{Rk}/(\hat{R}_p^{Rk}e)\hat{S}_{msy}^{Rk}} = 1. \tag{5}$$

The median optimum harvest rate for the maximum sustainable yield is then:

$$\hat{U}_{msy}^{Rk} = 1 - (\hat{a}^{Rk})^{-1}e^{\hat{a}^{Rk}/(\hat{R}_p^{Rk}e)\hat{S}_{msy}^{Rk}}. \tag{6}$$

Similarly, for the Beverton–Holt model,

$$\hat{S}_{msy}^{BH} = \frac{\hat{R}_p^{BH}}{\hat{a}^{BH}}(\sqrt{\hat{a}^{BH}} - 1), \tag{7}$$

$$\hat{U}_{msy}^{BH} = 1 - \sqrt{\frac{1}{\hat{a}^{BH}}}. \tag{8}$$

Note that these quantities are the estimated medians rather than their means. The estimate of the virgin biomass B_0 was determined from the intersection of the fitted S–R relationships with the $F = 0$ replacement line. Again, because the stock size and recruitment were measured in the same units and the species was assumed to be semelparous, the slope of the replacement line was one and $S_{20\%B_0}$ was simply 20% of B_0.

To compare the Ricker and Beverton–Holt models, a relative deviation between an estimated parameter $\hat{\theta}$ (including $\hat{a}$, $\hat{R}_p$, $\hat{\sigma}$, $\hat{S}_{msy}$, $\hat{U}_{msy}$ and $\hat{S}_{20\%B_0}$) and a known parameter θ (parameter a and R_p) or between the two estimates was obtained as:

$$RD_\theta\,(\%) = \frac{\hat{\theta}^{BH} - \hat{\theta}^{Rk}}{\hat{\theta}^{RK}} \times 100. \tag{9}$$

2.3. Distinguishing between the correct and incorrect models

S–R data generated from one known model, either the Ricker model (Eq. (1)) or the Beverton–Holt (Eq. (2)) model, were fit by the two models with unknown parameters using the maximum likelihood method. The standard error of the residuals $\hat{\sigma}$ was estimated along with other model parameters. This standard error was the basic quantity for evaluating how close a model fit to the data and was the basis for deriving other statistical criteria such as R^2, AIC, BIC, etc. Because the Ricker model and the Beverton–Holt model contain the same number of parameters, the standard error was sufficient for model comparison and was equivalent to using AIC or BIC. Therefore, the standard error was used as a criterion of goodness-of-fit in this part of study and was examined in two ways. First, the distributions of $\hat{\sigma}$ were compared between the two models; and secondly, $\hat{\sigma}$ was used to calculate the probability of selecting the correct model, which was the ratio between the number of times when the correct model had a smaller $\hat{\sigma}$ and the total number of simulations for that specific scenario.

3. Results

3.1. Comparison of estimated parameters of alternative models

3.1.1. Population growth rate a

The Ricker model consistently produced a smaller estimate of a than the Beverton–Holt model. This deviation between the two models increased as the population growth rate and the data range increased. When compared with the known growth rates used for generating the data, the Ricker model generally underestimated this parameter while the Beverton–Holt model overestimated it. The Beverton–Holt model had a small bias when the true a was small (e.g., $a = 2$) or when the data range was narrow (i.e., $S_{upper} = S_p$). Overall, the bias for the Ricker model was less variable: -15%, -18%, and -20% for the spawning population range $S_{upper} = S_p$, $2S_p$, and $3S_p$, respectively. For the Beverton–Holt model the bias between the estimated $\hat{a}$ and the true a increased rapidly as the data

range increased: 0%, 27%, and 71% at the three spawner data ranges.

3.1.2. Maximum recruitment R_p

In the majority of cases the Ricker model underestimated R_p while the Beverton–Holt model overestimated R_p. The deviation between the two models tended to diminish as the growth rate a or the data range S_{upper} increased. On average, the Ricker model underestimated the true R_p by −12%, −9%, and −7%, while the Beverton–Holt model overestimated R_p by 49%, 22%, and 14% when the spawning population range $S_{upper} = S_p$, $2S_p$, and $3S_p$, respectively.

3.1.3. Optimum stock size for the maximum sustained yield S_{msy}

The results of S_{msy} from the two models were not always consistent. However, the differences between the two models were clearly a function of the growth rate, fishing mortality, and data range (Fig. 1). The optimum stock size S_{msy} was smaller for more productive stocks (larger growth rate a) and this was consistent for both models. For stocks that had been under a long-term low fishing pressure ($F \leq 0.4$), the Ricker model was more conservative (produced a larger S_{msy}) than the Beverton–Holt model. The difference between the two models generally increased as the data range widened. In some cases the value of S_{msy} from

the Ricker model was more than twice as high as that from the Beverton–Holt model. When the average long-term fishing mortality was high ($F \geq 0.8$), the pattern sometimes reversed and the Ricker model became less conservative than the competing model. The noise level in the S–R data did not have much effect on the S_{msy} difference between the two models.

3.1.4. Optimum harvest rate for the maximum sustained yield U_{msy}

The results for the optimal harvest rate U_{msy} were almost exactly the same as for the S_{msy}. The growth rate, fishing mortality, and data range all influenced the difference between the estimates from the two models (Fig. 2). More productive stocks could sustain a higher harvest rate. For stocks that had been under a long-term low fishing pressure ($F \leq 0.4$), the Ricker model was more conservative, resulting in a smaller U_{msy} than the Beverton–Holt model. This pattern also reversed in some cases if the long-term fishing mortality was high ($F \geq 1.6$) and the Beverton–Holt model became more conservative, especially when the available data range was not very wide ($S_{upper} < 3S_p$).

3.1.5. Spawning population at 20% of virgin stock size $\hat{S}_{20\%B_0}$

The disparity of this reference point estimated by the two models also depended on the growth rate, fishing mortality, and
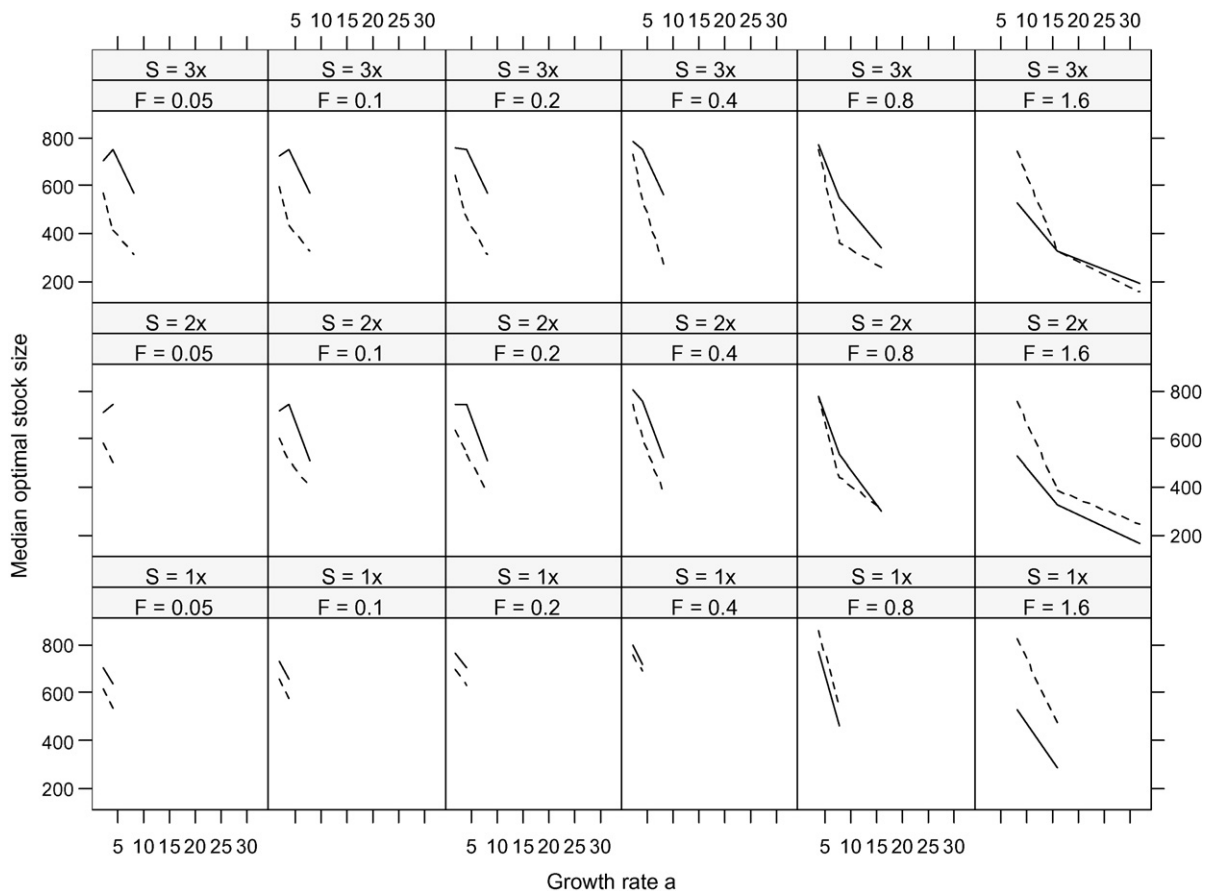

Fig. 1. Comparison of the estimated median S_{msy} for the Ricker (solid line) and the Beverton–Holt (dashed line) models as a function of the population growth rate a, long term fishing mortality F, and data range S ($S_{upper} = 1$, 2, and 3 times of the stock size for the maximum production S_p). The process noise in recruitment expressed as standard deviation set as S.D. = 0.16.
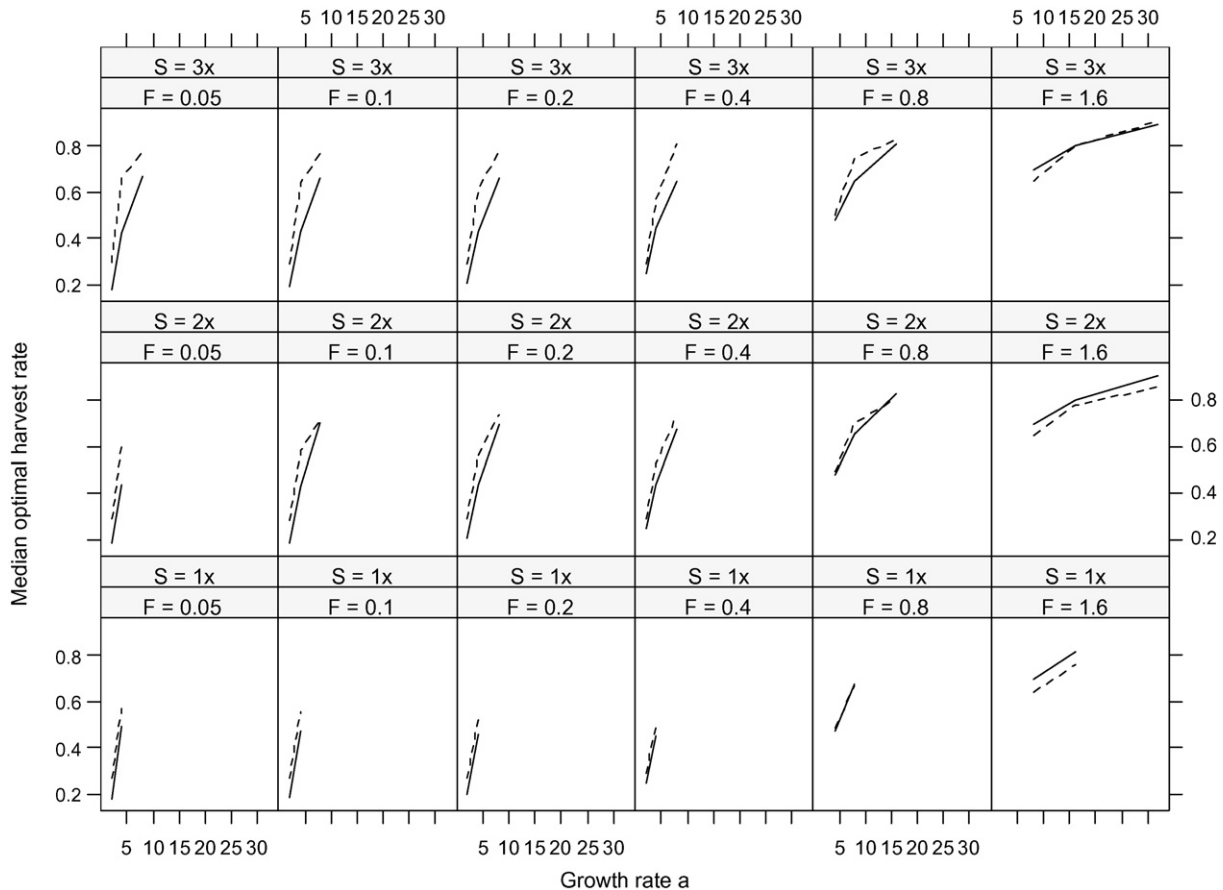
Fig. 2. Comparison of the estimated median U_{msy} for the Ricker (solid line) and the Beverton–Holt (dashed line) models as a function of the population growth rate a, long term fishing mortality F, and data range S ($S_{upper} = 1$, 2, and 3 times of the stock size for the maximum production S_p). The process noise in recruitment expressed as standard deviation set as S.D. $= 0.16$.

data range of the spawning population (Fig. 3). If the stock had a low growth rate and had been under a low long-term fishing pressure ($F \leq 0.2$), the Ricker model was more conservative and resulted in a higher $\hat{S}_{20\%B_0}$ than the Beverton–Holt model. However, the Beverton–Holt model became more conservative for stocks with higher growth rates. When the average long-term fishing mortality was high ($F \geq 0.8$), the trends shifted and the Ricker model was less conservative (produced a lower $\hat{S}_{20\%B_0}$). The estimates differed several-fold between the two models.

3.2. Distinguishing between the correct and incorrect models

In the first test, the data generated by the Ricker model with known parameters were fit by both the Ricker and Beverton–Holt models. Therefore, the Ricker model was the correct model while the Beverton–Holt model was the incorrect one. The residuals between the observed and the estimated recruitments, expressed as standard deviation $\hat{\sigma}$, were examined and compared between the two models. As expected, the estimated $\hat{\sigma}$ increased linearly with the process noise SD used in generating the data. An example is given in Fig. 4 where the data range was set at $S_{upper} = 2S_p$ and fishing mortality at $F = 0.4$. The correct (Ricker) model had a smaller $\hat{\sigma}$ than the incorrect (Beverton–Holt) one

when the process noise level was low and the stocks were more productive. As the process noise increased, it was difficult to distinguish the difference between the two models, especially when the stock had a low growth rate. The differences in the distribution of $\hat{\sigma}$ between the correct and incorrect models were more noticeable when the data range increased.

The second test was to use the data generated by the Beverton–Holt model and fit them by both the Ricker and Beverton–Holt models. Now the Beverton–Holt model was the correct model while the Ricker model was the incorrect one. The results were similar to those in the first test but the difference in $\hat{\sigma}$ between the two models became less clear (Fig. 5). The standard deviation of residuals also increased with the process noise used in generating the data. The distributions of $\hat{\sigma}$ from the correct (Beverton–Holt) and incorrect (Ricker) models overlapped each other under nearly all simulation scenarios. In this case, the correct and incorrect models could not be distinguished simply by the residuals.

Finally, the pair-wise comparison of $\hat{\sigma}$ between the two models for each time series data was tested. The probability of selecting the correct model was calculated as the ratio between the number of times when the correct model had a smaller $\hat{\sigma}$ than the incorrect model and the total number of time series for that specific simulation scenario. A probability of 0.5 meant
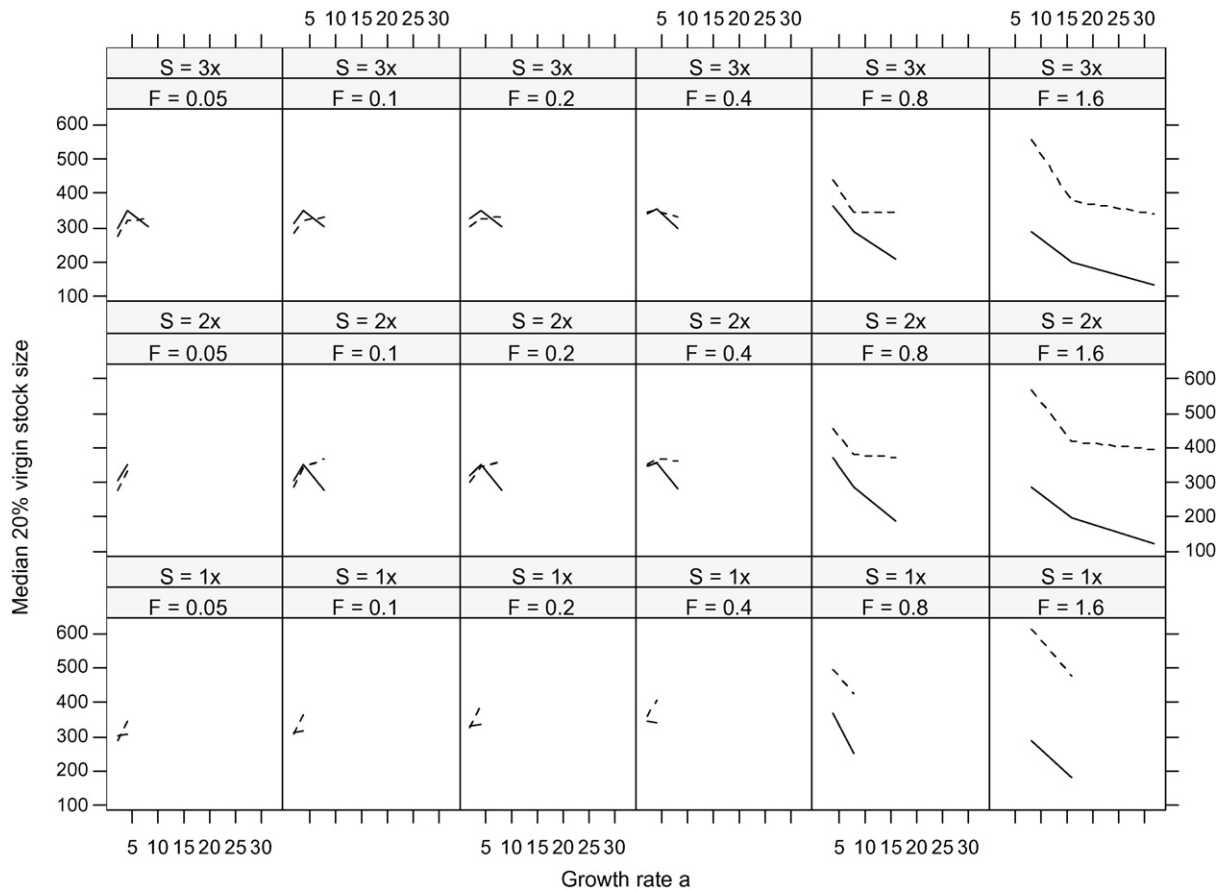
Fig. 3. Comparison of the estimated median 20% of virgin stock size $\hat{S}_{20\%B_0}$ for the Ricker (solid line) and the Beverton–Holt (dashed line) models as a function of the population growth rate a, long term fishing mortality F, and data range S ($S_{upper} = 1$, 2, and 3 times of the stock size for the maximum production S_p). The process noise in recruitment expressed as standard deviation set as S.D. = 0.16.

that the incorrect model had the same chance of being selected as the correct one. When the process noise level was extremely low (S.D. = 0.02), the probability of selecting the correct model was high (over 0.8, Fig. 6). However, this probability decreased rapidly as the process noise increased. When data were generated by the Ricker model, it was more difficult to detect the correct model when the growth rate was lower or the data range was narrower. When data were generated by the Beverton–Holt model, the pattern was less clear. When the data range was narrow the probability of selecting the correct model sometimes decreased slightly as the growth rate increased. Fishing mortality did not have significant impact on the probability of selecting the correct model. Interestingly, the Ricker model tended to be more flexible than the Beverton–Holt model and had a higher chance of being selected when it was the correct model. Moreover, when the data range was narrow (i.e., $S_{upper} = S_p$) and the growth rate was relatively high ($a \geq 8$) the correct Beverton–Holt model (the source of the data) had an even lower probability of being selected than the incorrect Ricker model. For example, when the growth rate $a = 8$, data range $S_{upper} = S_p$, fishing mortality $F = 0.4$ and the process noise S.D. = 0.16, the probability of selecting the correct Beverton–Holt model was approximately 0.38 while the probability of selecting the incorrect Ricker model was 0.62. For all simulation scenarios where the pro-

cess noise S.D. ≥ 0.16, the probability of selecting the correct model was approximately 0.70 (S.D. = 0.16) for data generated by the Ricker model and 0.61 (S.D. = 0.11) for data generated by the Beverton–Holt model.

4. Discussion

demonstrates that very different model parameters and management reference values may result from alternative S–R functions even when they all fit the data similarly well. The study shows that the differences between the Ricker and Beverton–Holt models are functions of stock characteristics including the population growth rate, the range of the stock size, and the long-term average fishing mortality. It also shows that while one management parameter from a particular S–R model may be more conservative under specific conditions than it is from the alternative model, other management parameters may be opposite under the same conditions. For example, while the Ricker model may conservatively recommend a higher optimal stock size and a lower optimal harvest rate, it may also recommend a lower stock size when the management reference is based on the percentage of the virgin biomass.

Fishing mortality may completely reverse the model disparity even when the stocks have the same growth rate and data range.
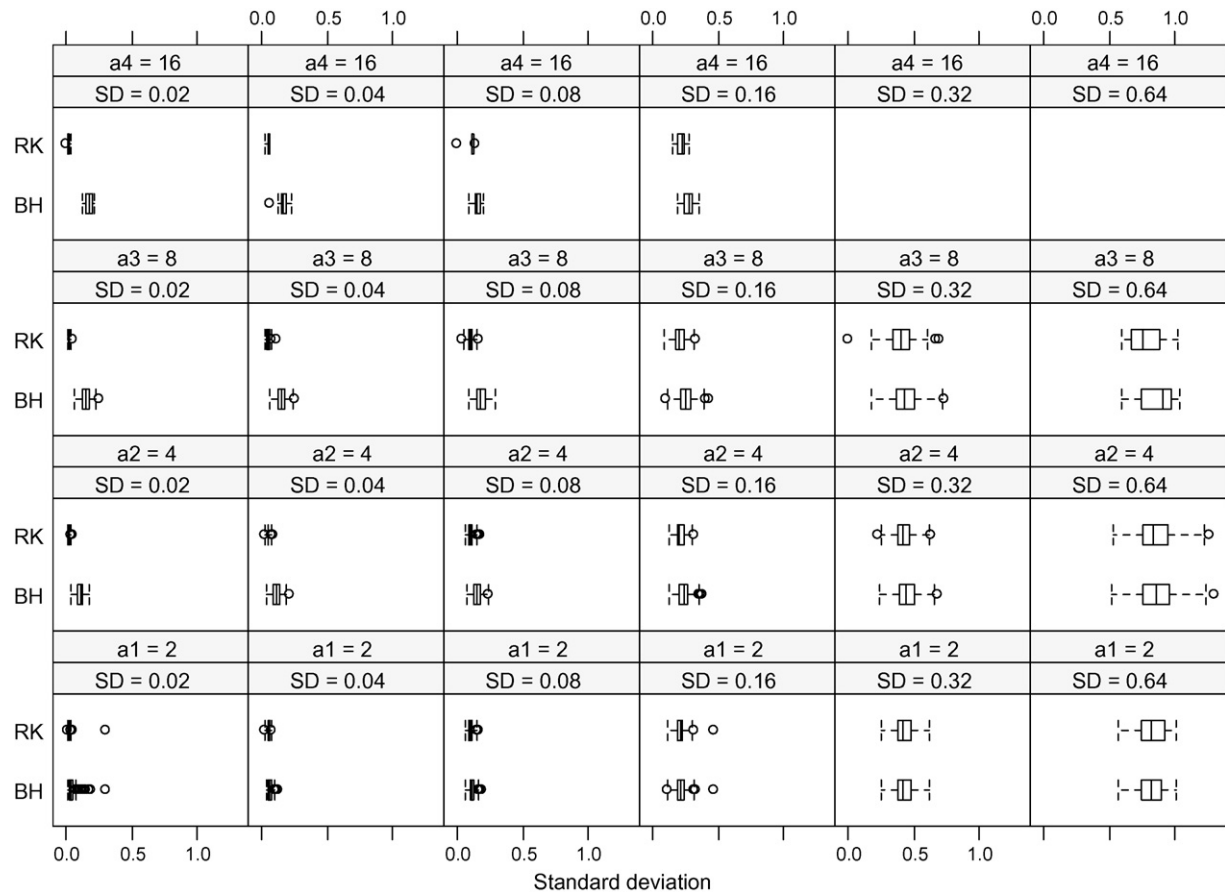
Fig. 4. Comparison of the standard deviations of residuals estimated from the Ricker (RK, correct) and Beverton–Holt (BH, incorrect) models. The data are generated from the Ricker model with known growth rate $a = 2, 4, 8,$ or 16, process noise in recruitment expressed as standard deviation S.D., long-term fishing mortality $F = 0.4$, and data range $S_{upper} = 2S_p$.

For example, for a stock with a medium growth rate, the Ricker model yields a higher optimal stock size than the Beverton–Holt model when that stock has been under a low to medium long-term fishing pressure. However, the same Ricker model yields a lower optimal stock size than the Beverton–Holt model for a stock with the same growth rate but experiencing high fishing mortality. This result is due to the nature of the S–R data: when a stock is continuously exposed to high fishing pressures, we will observe more data in the area of smaller stock sizes (the increasing limb of the Ricker curve), whereas when the fishing mortality is low we will observe more data in the higher stock size region (the declining limb of the Ricker curve).

This analysis explains certain results of other studies. For example, in cases where one S–R model consistently provides more conservative reference points for real fishery data sets, the stocks tend to have similar traits such as low population growth rates (Barrowman and Myer, 2000). However, when one compares across many stocks and finds that the results are mixed and a conservative model cannot be identified, it is more likely these stocks have very different growth rates, data ranges, and/or fishing mortalities (Myers et al., 1994).

The simulations in this study demonstrate that the correct and incorrect models may not be distinguishable for most fish stocks based on statistical methods or information criteria such as AIC.

An attempt to select the correct or best S–R model using only these criteria could be very risky. For most simulation scenarios in this study, the total errors (process noise plus measurement error) used were relatively small (standard deviations were less than 0.2) compared to real fishery data where standard deviations often exceed 0.5 (Myers et al., 1995a,b; CTC, 1999). For typical fishery data with a total error of greater than 0.2, the chance of selecting the correct model over the incorrect model by information criteria would be only 60%. Although the probability of selecting a correct model increases when the data range is wide ($S_{upper} = 3S_p$), for most real fishery data we only have a narrow range of spawner data (Myers et al., 1995a,b). The smaller the spawner range, the closer the goodness-of-fit for alternative models.

This study also found a systematic bias toward better fits with the Ricker model over the Beverton–Holt model for some simulation scenarios (Valpine and Hastings, 2002), even when the Ricker model was the incorrect one. The artificial data used in this study included a random fishing mortality which was necessary for preventing the simulated S–R data from surrounding an undisturbed equilibrium point. Some authors have recommended that it is necessary to substantially disturb a stock from its equilibrium point in order to uncover its true S–R relationship (Hilborn and Walters, 1992). The same
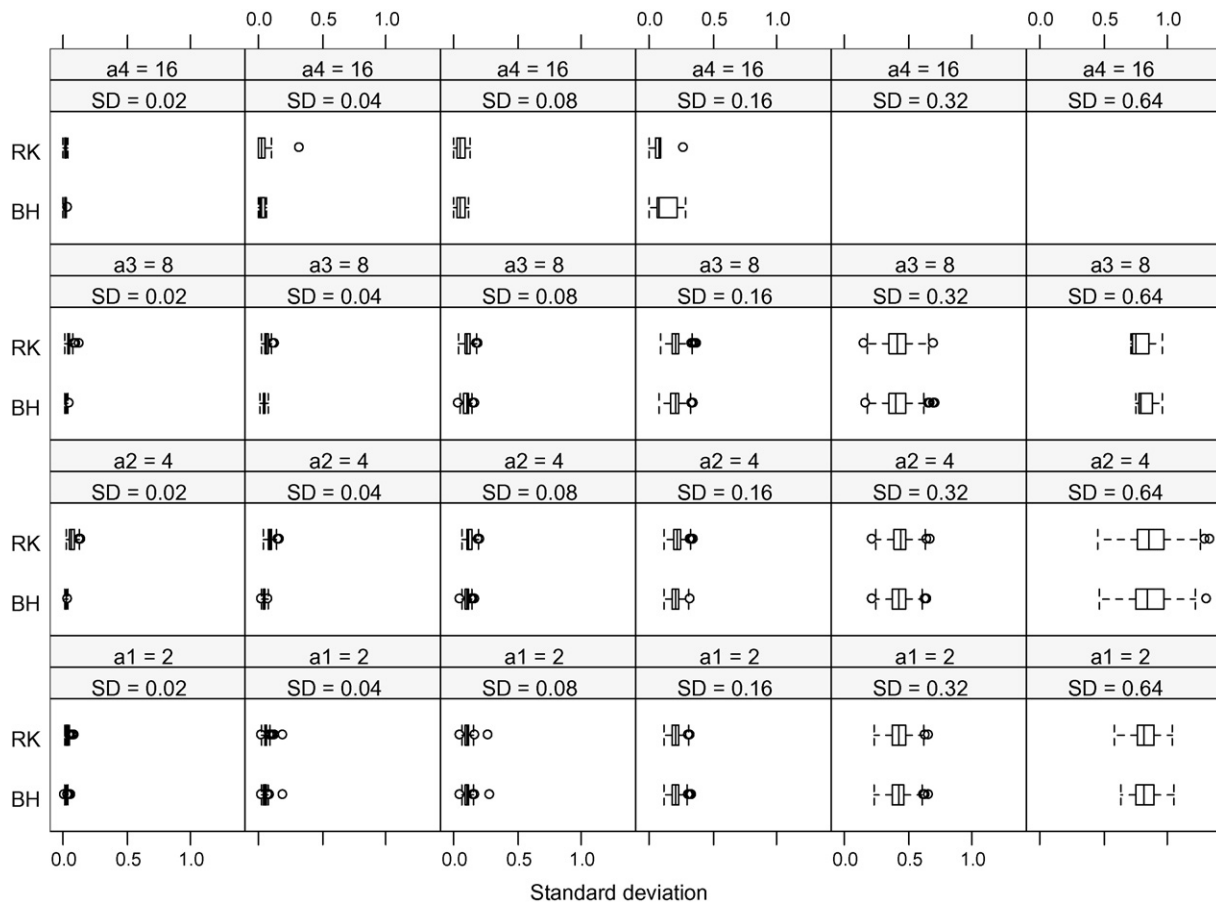
Fig. 5. Comparison of the standard deviations of residuals estimated from the Ricker (RK, incorrect) and Beverton–Holt (BH, correct) models. The data are generated from the Beverton–Holt model with known growth rate $a = 2, 4, 8$, or 16, process noise in recruitment expressed as standard deviation S.D., long-term fishing mortality $F = 0.4$, and data range $S_{upper} = 2S_p$.

systematic bias resulted from this study and an early simulation study that did not include a random fishing mortality (Valpine and Hastings, 2002), which further undermines the approach of using information criteria for selecting a correct S–R model.

Model selection has inspired a wide range of interests from pure statisticians to researchers in applied sciences. The narrow goal of model selection is to identify the one model, from a set of competing models, that best captures the relationship underlying the process of interest (Buckland et al., 1997; Kieseppa, 2001a, 2001b; White and Lubow, 2002). In fisheries, model selection is of great importance as management decisions frequently depend on one particular model selected from the range of models examined. Most often, model selection is done by application of one or several of the criteria that have been developed for this purpose. The majority of these criteria assess models using a function of the usual estimate of residual variance and the model dimension, such as the well-known criteria like Mallow's C_p, R^2, AIC, and BIC. If alternative models result in similar management parameters, it is not critical to just use one best model. However, the simulation results in this study and other studies show that different models may reach very different conclusions even though they fit the data similarly well. Increasingly more researchers have recognized that the isolation of a unique "true" model may

be undesirable. Biological processes in an ecological system are very complicated. It is hard to believe that the biological data we observed are generated from a "true" mathematical model. A real fishery time series dataset is just one random realization of many potential outcomes of a natural process. A model that based on such one sample of data cannot reveal full reality of the nature. Although selecting one best model may be useful, especially among hierarchical models, this practice should be avoided in stock assessment using different families of S–R functions.

Hilborn and Walters (1992) discussed several desirable properties of S–R models including that the curve should pass through the origin and not fall to the abscissa at higher levels of stock and that the rate of recruitment should decrease with increases in parental stock. Quinn and Deriso (1999) asserted: "In principle, any mathematical model can be used for spawner–recruit relationships as long as it goes through the origin, increases monotonically at low spawning levels, and shows some level of density dependence at high spawning levels." Since the underlying biological relationship between spawner abundance and recruits is rarely known in practice (Kimura, 1988; Clark, 1991; Myers et al., 1994; Barrowman and Myer, 2000), one could conceivably build an unlimited number of models of S–R relationships. Unless one finds a mathematical model that
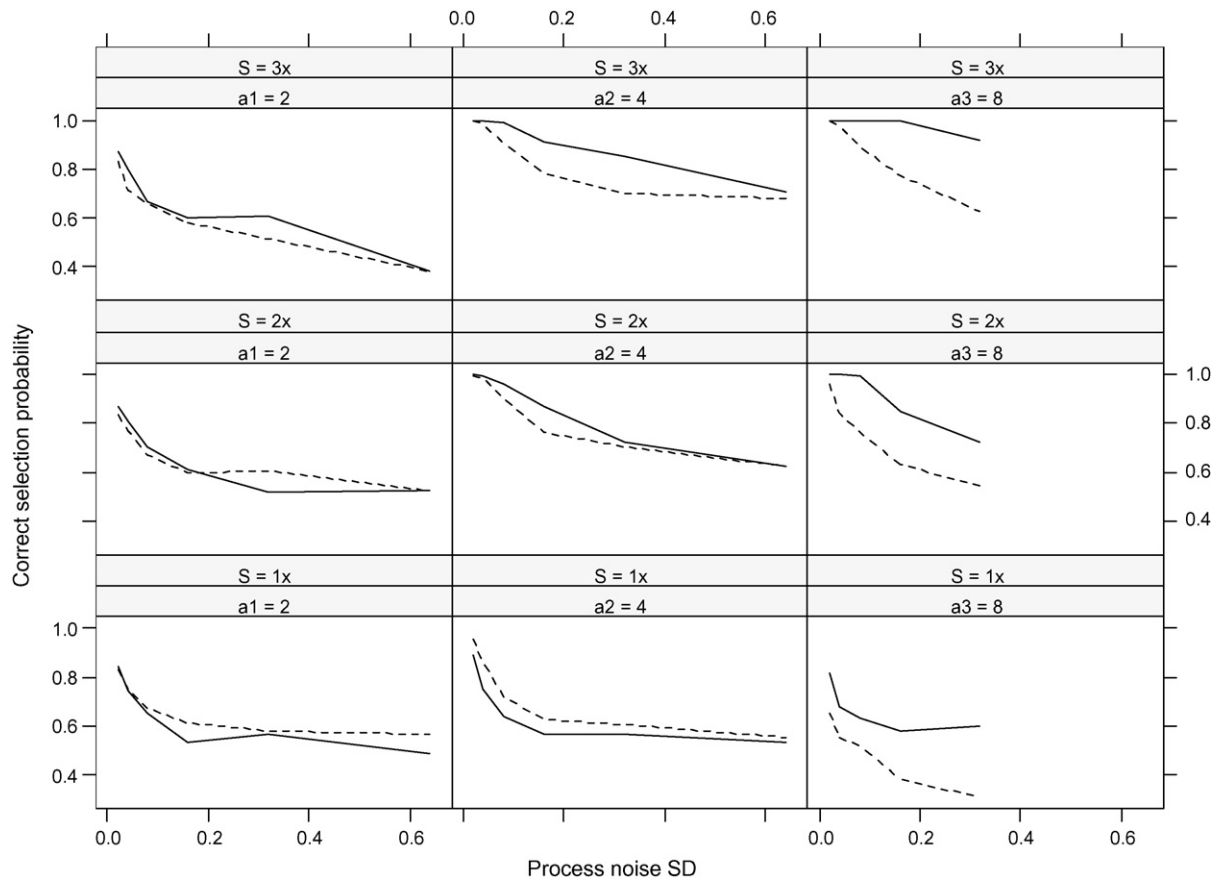
Fig. 6. Probability of selecting the correct model based on the estimated standard deviation of residuals. Data are generated either by the Ricker model (solid line) or Beverton–Holt model (dashed line) with known process noise in recruitment expressed as standard deviation S.D., growth rate $a = 2$, 4, or 8, data range $S = 1S_p$ to $3S_p$, and long-term fishing mortality $F = 0.4$.

perfectly fits the S–R data with residual $\varepsilon_i = 0$ for all i, one can hardly claim that any single model is the only correct model for the data set. Knowledge of the biology and behavior of specific species may provide scientific insight into which model may be more appropriate. However, such knowledge is typically unavailable or unclear for many fish species or specific stocks. Further, one may argue that it is not necessary to force the curve to go through the origin since one fish cannot produce any offspring. Chen et al. (2002) demonstrated an "Allee effect" in salmon S–R relationships where the curve did not go through the origin.

The fundamental dilemma is that we lack an understanding of the underlying biological processes in fish S–R relationships in the natural environment. Much discussion focuses on whether there is depensation at low stock abundance (Hilborn and Walters, 1992; Myers et al., 1995a,b; Liermann and Hilborn, 1997; Quinn and Deriso, 1999; Walters and Kitchell, 2001; Chen and Holtby, 2002). Some researchers even argued that recruitment is independent of spawner abundance (Gilbert, 1997). To overcome such a dilemma, fisheries scientists must recommend robust, adaptive, and precautionary management policies (Kimura, 1988; Hilborn and Walters, 1992; Punt and Hilborn, 1997; Charles, 1998; Weeks and Berkeley, 2000; Hilborn et al., 2001; Schnute and Richards, 2001).

Fisheries management will continue to rely on mathematical modeling, in spite of inherent shortcomings. Since no rigorous statistical methods are available to identify a single best model in most fishery cases, and since the underlying biological relationship between spawner abundance and recruits is usually uncertain, it would be prudent to use several alternative models to fit the data and provide a decision table listing management parameters from different models (Punt and Hilborn, 1997; Marjomaki, 2004). If conservation has high priority, one may choose a combination of reference points from multiple structurally different models to safe guard against the risk due to the direction of potential biases may differ for different parameters. For example, one model may overestimate optimal stock size while underestimating virgin spawning biomass.

Recently, some new methods have been proposed to study stock–recruitment relationship and forecasting stock recruitment. Artificial neural network (Chen and Ware, 1999; Huse and Ottersen, 2003; Zhou, 2003; Megrey et al., 2005), fuzzy logic (Mackinson et al., 1999; Chen et al., 2000; Chen, 2001), and nonparametric analysis (Jacobson and MacCall, 1995; Chen and Irvine, 2001) provide novel approaches for S–R studies. Studies showed that these new approaches improve the accuracy of fishery stock assessment over the traditional methods. We may see more research in these new techniques in the near future.

Acknowledgments

References

Barrowman, N.J., Myer, R.A., 2000. Still more spawner–recruitment curves: the hockey stick and its generalizations. Can. J. Fish. Aquat. Sci. 57, 665–676.

Berkson, J.M., Kline, L.L., Orth, D.J. (Eds.), 2002. American Fisheries Society, Symposium 27. Bethesda, Maryland.

Beverton, R.J.H., Holt, S.J., 1957. On the Dynamics of Exploited Fish Populations. Chapman and Hall, London (facsimile reprint, 1993).

Bradford, M.J., Myers, R.A., Irvine, J.R., 2000. Reference points for coho salmon (Oncorhynchus kisutch) harvest rates and escapement goals based on freshwater production. Can. J. Fish. Aquat. Sci. 57, 677–686.

Brodziak, J., 2002. In search of optimal harvest rates for west coast groundfish. N. Am. J. Fish. Manage. 22, 258–271.

Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model selection: an integral part of inference. Biometrics 53, 603–618.

Charles, A.T., 1998. Living with uncertainty in fisheries: analysis methods, management priorities and the Canadian groundfishery experience. Fish. Res. 37, 37–50.

Chen, D.G., 2001. Detecting environmental regimes in fish stock–recruitment relationships by fuzzy logic. Can. J. Fish. Aquat. Sci. 58, 2139–2148.

Chen, D.G., Holtby, L.B., 2002. A regional meta-model for stock–recruitment analysis using an empirical Bayesian approach. Can. J. Fish. Aquat. Sci. 59, 1503–1514.

Chen, D.G., Irvine, J.R., 2001. A semiparametric model to examine stock–recruitment relationships incorporating environmental data. Can. J. Fish. Aquat. Sci. 58, 1178–1186.

Chen, D.G., Ware, D.M., 1999. A neural network model for forecasting fish stock recruitment. Can. J. Fish. Aquat. Sci. 56, 2385–2396.

Chen, D.G., Hargreaves, N.B., Ware, D.M., Liu, Y., 2000. A fuzzy logic model with genetic algorithm for analyzing fish stock–recruitment relationships. Can. J. Fish. Aquat. Sci. 57, 1878–1887.

Chen, D.G., Irvine, J.R., Cass, A.J., 2002. Incorporation Allee effects in fish stock–recruitment models and applications for determining reference points. Can. J. Fish. Aquat. Sci. 59, 242–249.

Clark, W.G., 1991. Groundfish exploitation rates based on life history parameters. Can. J. Fish. Aquat. Sci. 48, 734–750.

CTC (Chinook Technical Committee), 1999. Maximum sustained yield or biologically based escapement goals for selected Chinook salmon stocks used by the Pacific Salmon Commission's Chinook Technical Committee for escapement assessment. Pacific Salmon Commission Joint Chinook Technical Committee Report TCCHINOOK (99)-3. Vancouver, Canada.

Cushing, D.H., 1971. Dependence of recruitment on parent stock in different groups of fishes. J. Con. Int. Expl. Mer. 33, 340–362.

Cushing, D.H., 1973. Dependence of recruitment on parent stock. J. Fish. Res. Board Can. 30, 1965–1976.

Deriso, R.B., 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37, 268–282.

Ehrhardt, N.M., Legault, C.M., 1997. The role of uncertainty in fish stock assessment and management: a case study of the Spanish mackerel, Scomberomorus maculatus, in the US Gulf of Mexico. Fish. Res. 29, 145–158.

Flaaten, O., Salvanes, A.G.V., Schweder, T., Ulltang, O., 1998. Fisheries management under uncertainty – an overview. Fish. Res. 37, 1–6.

Francis, R.I.C.C., Shotton, R., 1997. "Risk" in fisheries management: a review. Can. J. Fish. Aquat. Sci. 54, 1699–1715.

Gilbert, D.J., 1997. Towards a new recruitment paradigm for fish stocks. Can. J. Fish. Aquat. Sci. 54, 969–977.

Haddon, M., 2001. Modeling and Quantitative Methods in Fisheries. Chapman & Hall/CRC, Boca Raton.

Hammond, T.R., O'Brien, C.M., 2001. An application of the Bayesian approach to stock assessment model uncertainty. ICES J. Mar. Sci. 58, 648–656.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. Chapman & Hall, New York.

Hilborn, R., Maguire, J.J., Parma, A.M., Rosenberg, A.A., 2001. The precautionary approach and risk management: can they increase the probability of successes in fishery management? Can. J. Fish. Aquat. Sci. 58, 99–107.

Huse, G., Ottersen, G., 2003. Forecasting recruitment and stock biomass of Northeast Arctic cod using neural networks. Sci. Mar. 67, 325–335.

Jacobson, L.D., MacCall, A.D., 1995. Stock–recruitment models for Pacific sardine (Sardinops sagax). Can. J. Fish. Aquat. Sci. 52, 566–577.

Kieseppa, I.A., 2001a. Statistical model selection criteria and the philosophical problem of under determination. Br. J. Phil. Sci. 52, 761–794.

Kieseppa, I.A., 2001b. Statistical model selection criteria and Bayesianism. Phil. Sci. 68, S141–S152.

Kimura, D.K., 1988. Stock–recruitment curves as used in the stock-reduction analysis model. J. Con. Int. l'Expl. Mer. 44, 253–258.

Liermann, M., Hilborn, R., 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. Can. J. Fish. Aquat. Sci. 54, 1976–1984.

Ludwig, D., Walters, C.J., 1981. Measurement errors and uncertainty in parameter estimates for stock and recruitment. Can. J. Fish. Aquat. Sci. 38, 711–720.

Mace, P.M., Sissenwine, M.P., 2002. Coping with uncertainty: evolution of the relationship between science and management. In: Berkson, J.M., Kline, L.L., Orth, D.J. (Eds.), Incorporating Uncertainty into Fishery Models. American Fisheries Society, Bethesda, Maryland, pp. 9–28 (symposium 27).

Mackinson, S., Vasconcellos, M., Newlands, N., 1999. A new approach to the analysis of stock–recruitment relationships: "model-free estimation" using fuzzy logic. Can. J. Fish. Aquat. Sci. 56, 686–699.

Marjomaki, T.J., 2004. Analysis of the spawning stock–recruitment relationship of vendace (Coregonus albula (L.)) with evaluation of alternative models, additional variables, biases and errors. Ecol. Freshwater Fish. 13, 46–60.

Megrey, B.A., Lee, Y.-W., Macklin, S.A., 2005. Comparative analysis of statistical tools to identify recruitment–environment relationships and forecast recruitment strength. ICES J. Mar. Sci. 62, 125–1269.

Myers, R.A., Barrowman, N.J., 1996. Is fish recruitment related to spawner abundance? Fish. Bull. 94, 707–724.

Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N.J., Restrepo, V.R., 1994. In search of thresholds for recruitment overfishing. ICES J. Mar. Sci. 51, 191–205.

Myers, R.A., Bridson, J., Barrowman, N.J., 1995a. Summary of worldwide spawner and recruitment data. Can. Tech. Rep. Fish. Aquat. Sci. 2020, 327.

Myers, R.A., Barrowman, N.J., Hutchings, J.A., Rosenberg, A.A., 1995b. Population dynamics of exploited fish stocks at low population levels. Science 269, 1106–1108.

Myers, R.A., Bowen, K.G., Barrowman, N.J., 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56, 2404–2419.

Patterson, K.R., 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. Can. J. Fish. Aquat. Sci. 56, 208–221.

Patterson, K., Cook, R., Barby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D.W., Stefansson, G., 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish Fish. 2, 125–157.

Peterman, R.M., 1981. Form of random variation in salmon smolt-to-adult relations and its influence on production estimates. Can. J. Fish. Aquat. Sci. 38, 1113–1119.

PFMC, 2002. Pacific Fishery Management Council September 2002 briefing book. Portland.

Punt, A.E., Hilborn, R., 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. Fish. 7, 35–63.

Quinn, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

Ralston, S., 2002. West coast groundfish harvest policy. N. Am. J. Fish. Manage. 22, 249–250.

Reish, R.L., Deriso, R.B., Ruppert, D., Carroll, R.J., 1985. An investigation of the population dynamics of Atlantic menhaden (Brevoortia tyrannus). Can. J. Fish. Aquat. Sci. 42, 147–157.

Rice, J.C., Richards, L.J., 1996. A framework for reducing implementation uncertainty in fisheries management. N. Am. J. Fish. Manage. 16, 488–492.

Ricker, W.E., 1954. Stock and recruitment. J. Fish. Res. Board Can. 11, 559–623.

Schnute, J., 1985. A general theory for the analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42, 414, 429.

Schnute, J.T., Richards, L.J., 2001. Use and abuse of fishery models. Can. J. Fish. Aquat. Sci. 58, 10–17.

Shepherd, J.G., 1982. A versatile new stock–recruitment relationship for fisheries, and the construction of sustainable yield curves. J. Con. Int. l'Expl. Mer 40, 67–75.

Valpine, P.D., Hastings, A., 2002. Fitting population models incorporating process noise and observation error. Ecol. Monogr. 72, 57–76.

Walters, C., Kitchell, J.F., 2001. Cultivation/depensation effects on juvenile survival and recruitment: implications for the theory of fishing. Can. J. Fish. Aquat. Sci. 58, 39–50.

Walters, C.J., Ludwig, D., 1981. Effects of measurement errors on the assessment of stock–recruitment relationships. Can. J. Fish. Aquat. Sci. 38, 704–710.

Wang, Y., Liu, Q., 2006. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationship. Fish. Res. 77, 220–225.

Weeks, H., Berkeley, S., 2000. Uncertainty and precautionary management of marine fisheries: can the old methods fit the new mandates? Fisheries 25, 6–15.

White, G.C., Lubow, B.C., 2002. Fitting population models to multiple sources of observed data. J. Wildlife Manage. 66, 300–309.

Williams, E.H., Shertzer, K.W., 2003. Implications of life-history invariants for biological reference points used in fishery management. Can. J. Fish. Aquat. Sci. 60, 710–720.

Zhou, S., 2003. Application of artificial neural network for forecasting salmon escapements. N. Am. J. Fish. Manage. 23, 48–59.