Chapter 26
Model Selection Uncertainty and Bayesian Model Averaging in Fisheries Recruitment Modeling

Yan Jiao, Kevin Reid, and Eric Smith

Abstract Stock recruitment (SR) modeling is the central part of fisheries population dynamics analysis. Models and modeling techniques on SR relationships have been evolving for decades, and have moved from traditional SR models, such as the Ricker and Beverton-Holt models to measurement error models, and Kalman filter time series models. Though SR models are evolving, people still typically select a specific model and then proceed as if the selected model had generated the data. This approach ignores the uncertainty in the model selection, leading to overconfident inferences and decisions with higher risk than expected. Bayesian model averaging (BMA) provides a coherent mechanism for accounting for this model uncertainty. In this study, Lake Erie walleye (Sander vitreus) fishery was used as an example. Six mathematical models were developed, which included a Ricker model, a hierarchical Ricker model, a residual auto-regressive model, a Kalman filter random walk model, a Kalman filter autoregressive model, and a Ricker measurement error model. The posterior distributions of estimated productivity and recruitment from these models were weighted based on the Deviance Information Criterion (DIC) to provide our predictive posterior distribution of population productivity and recruitment over time. To test the efficiency of the Bayesian averaging approach and the uncertainty from model selection, a further simulation study was done based on the example fishery. Our results showed that model selection uncertainty is high and BMA explained the data reasonably well. We suggest that BMA is more appropriate in simulating SR models. The framework developed here can be used for other species population SR analysis. We also suggest that the model selection uncertainty be considered and the

Y. Jiao(✉)
Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute & State University, Blacksburg, VA, 24061-0321, USA
e-mail: yjiao@vt.edu

K. Reid
Ontario Commercial Fisheries Association, Box 2129, 45 James Street, Blenheim, ON, N0P 1A0, Canada

E. Smith
Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, USA

R.J. Beamish and B.J. Rothschild (eds.), *The Future of Fisheries Science in North America*, Fish & Fisheries Series,

BMA be applied to other stock assessment models and even in the fisheries management decision making in the future.

Keywords Bayesian model averaging · Deviance Information Criterion · recruitment · time series · walleye

26.1 Introduction

Stock recruitment (SR) modeling is the central part of fisheries population dynamics analysis. However, it is also the bottleneck of many fisheries population dynamics and stock assessment. Concerns about SR modeling have never decreased. Models and modeling techniques have been evolving for decades, moving from traditional SR models, such as Ricker model and Beverton-Holt (BH) model, generalized Ricker model with environmental variables, to measurement error models, to autoregressive error models, and then extended to other Kalman-filter time series models (Walters and Ludwig 1981; Quinn and Deriso 1999; Peterman et al. 2003). Though SR models are evolving, the typical analysis is to select a single model from some class of models and then analysis proceeds as if the selected model had generated the data. This approach ignores the uncertainty in the model selection, leading to overconfident inferences and decisions that are more risky than one thinks (Draper 1995). Bayesian model averaging (BMA) provides a coherent mechanism for accounting for this model uncertainty. Framework development and application of BMA in SR modeling is necessary and will help SR modeling and fishery stock assessment and management.

In this study, a walleye (*Sander vitreus*) fishery from Lake Erie is used as an example. Its dynamic has been observed to be high and may be heavily influenced by environmental changes (Walleye Task Group 2004). In some years, the productivity can be very high while in some other years it can be very low in spite of the spawner stock size. A commonly used stock recruitment model obviously cannot satisfy the mission in modeling the dynamic variations of productivity and/or the environmental noise.

Traditional SR models and model selection for analyzing the SR relationship and the productivity of a stock can be inadequate because of the measurement errors in the SR data and/or the noisy signals transferred to the SR data. Besides a commonly used SR regression model, we included five other mathematical models developed based on available time series data for the example fishery, which includes a measurement error model, an autoregressive residual model, a random walk model, a Kalman filter autoregressive productivity model, and a hierarchical Ricker model (see the method section). Measurement error can be very important in analyzing the SR relationship because of the possible measurement error in the spawner stock size (Walters and Ludwig 1981). Recent research found that noise in the nature may not be white, but are colored in many cases (Caswell and Cohen 1995; Halley 1996; Vasseur and Yodzis 2004). The classification of noise by spectral density is given "color" terminology, with different types named after different colors. It can also be classified as how the noise signal is temporally autocorrelated. Spectral density is constant for white noise and the noise signal is independent; while spectral density for colored noise is changing with changing frequency and the noise signal is autocorrelated (Halley 1996; Petchey 2000). These colored noises can be more dangerous to species when population size is low (Halley and Kunin 1999; Morales 1999; Schwager et al. 2006). We developed these four models to simulate the possible colored noises in the example fishery. The autoregressive residual and population productivity models were used to simulate the colored noise for the population itself and for population productivity (Morales 1999; Schwager et al. 2006). The random walk model is a special case of the autoregressive population productivity model and is reasonable to investigate because it has fewer parameters and can simulate the nonwhite noise at the same time (Peterman et al. 2003). The hierarchical Ricker model was used to simulate the hierarchy of possible productivities, which has been discussed related to regime shifts, changes of productivity regimes, etc. (Beamish et al. 1999; Glantz 1992).

We used a Bayesian approach to analyze the time series models for population productivity and recruitment. WinBUGS was used for this purpose (Spiegelhalter et al. 2004). A Bayesian approach has been very useful in dealing with time series model, which uses observations to update prior models of process noise, measurement noise, and state variables. Deviance Information Criterion (DIC) was used to compare different models. It is difficult to say what constitutes an important difference in DIC (Spiegelhalter et al. 2004). After the Bayesian analysis of each model, a Bayesian model averaging was used to balance model goodness of fit and model selection uncertainty. The posterior distributions of estimated productivity and recruitment from these models were weighted based on their DIC to provide our predictive posterior distribution of population productivity and recruitment over time.

To test the efficiency of the Bayesian averaging approach in modeling SR data and the uncertainty from model selection, a further simulation study was done based on the example walleye fishery. Our goal in this study is to assess the model selection uncertainty, and to find an appropriate method for understanding the productivity and the SR relationship of different fisheries. The framework developed here can be used for other species productivity and SR analysis.

Although six models were used in this study, they have included most of the types of the recruitment models. Models can be exchanged or modified to other similar models, e.g., BH model can replace Ricker model, hierarchical BH model can replace the hierarchical Ricker model, etc. The BMA framework can also extend from six to seven or, however, many models are of interest. The new recruitment modeling approach developed in this study improved our understanding of recruitment dynamics as well as raised the quality of stock assessment and management of the fisheries.

26.2 Materials and Methods

26.2.1 Example Fishery Used in This Study

The recruitment of walleye is the 2-year-old walleye in Lake Erie. Walleye older than age 3 are regarded as the spawning stock. The walleye recruitment (2-year-old fish) and spawning stock (3 + fish) are estimated using a statistical catch-at-age method (Fig. 26.1, Walleye Task Group 2004).

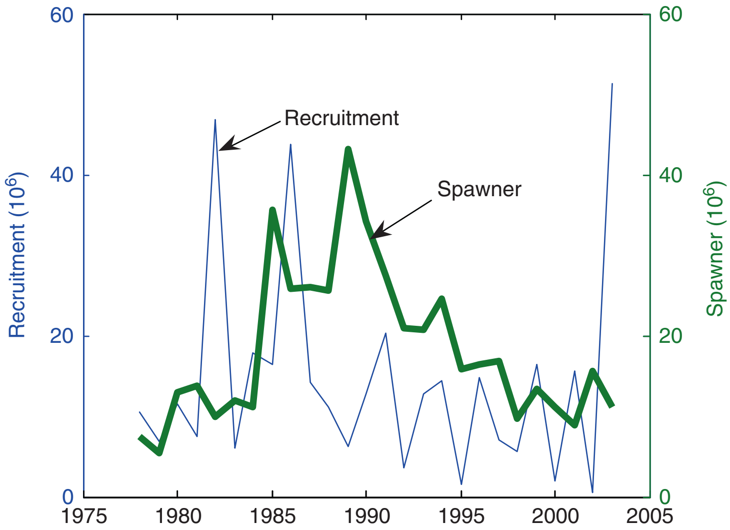

Fig. 26.1 Stock and recruitment of walleye fisheries over time

26.2.2 Models Used

According to the stock recruitment pattern (Fig. 26.2), Ricker model (RM) was selected to fit the walleye stock recruitment data. The Ricker model can be written as:

$$R_t = \alpha S_{t-2} e^{-\beta S_{t-2}^{\varepsilon_1}} \text{ or}$$
$$Ln(R_t) = Ln(\alpha) + Ln(S_{t-2}) - \beta S_{t-2} + \varepsilon_1 \quad (26.1)$$

where R_t is the recruitment numbers and S_t is the spawner abundance at year t, error ε_1 is independent and normally distributed with mean 0 and variance $\sigma^2_{\varepsilon_1}$. t ranges from 1978 to 2003.

The second model that we used was

$$Ln(R_t) = Ln(\alpha) + Ln(S'_{t-2}) - \beta S'_{t\text{-}2} + \varepsilon_2$$
$$Ln(S) = Ln(S') + \varepsilon_3 \quad (26.2)$$

In this model, measurement error in spawner abundance S is considered; S' is the true spawner abundance, and S is the measurement of S' with error ε_3. Errors ε_2 and ε_3 are independent and normally distributed with mean 0 and variance $\sigma_{\varepsilon_2}{}^2$ and $\sigma_{\varepsilon_3}{}^2$. We called this model the measurement error model (MEM).

The third model that we used was

$$Ln(R_t) = Ln(\alpha) + Ln(S_{t-2}) - \beta S_{t-2} + u_t$$
$$u_{t+1} = \phi u_t + \varepsilon_4 \quad (26.3)$$

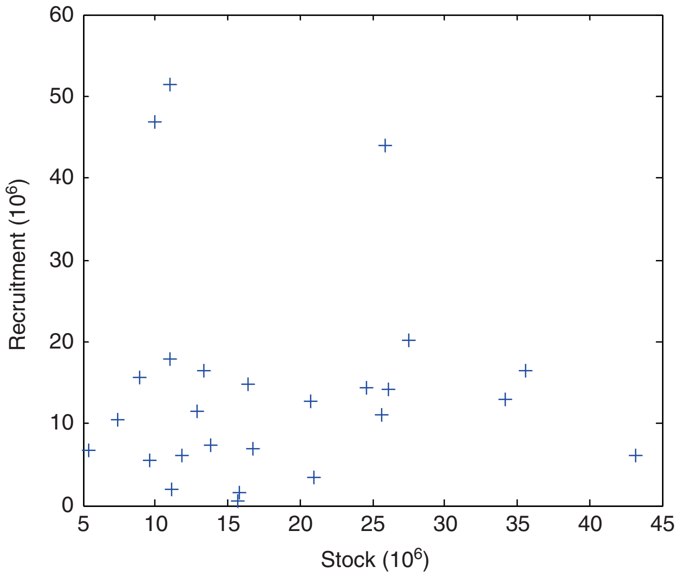

Fig. 26.2 Spawner stock and the corresponding recruitment

In this model, the residual error u_t is modeled as a first-order autoregressive process. ϕ is the autocorrelation coefficient, and the error ε_4 is independent and normally distributed with mean 0 and variance $\sigma_{\varepsilon_4}^2$. We called this model the residual autoregressive model (RAM).

The forth model that we used was

$$\begin{aligned} Ln(R_t) &= Ln(\alpha_t) + Ln(S_{t-2}) - \beta S_{t-2} + \varepsilon_5 \\ Ln(\alpha_{t+1}) &= Ln(\alpha_t) + \varepsilon_6 \end{aligned} \quad (26.4)$$

In this model, productivity α_t is modeled as a random walk process; and errors ε_5 and ε_6 are independent and normally distributed with mean 0 and variances $\sigma_{\varepsilon_5}^2$ and $\sigma_{\varepsilon_6}^2$. We called this model the random walk model (RWM).

The fifth model that we used was

$$\begin{aligned} Ln(R_t) &= Ln(\alpha_t) + Ln(S_{t-2}) - \beta S_{t-2} + \varepsilon_7 \\ Ln(\alpha_{t+1}) &= Ln(\bar{\alpha}) + \varphi[\mathrm{Ln}(\alpha_t) - Ln(\bar{\alpha}) + \varepsilon_8 \end{aligned} \quad (26.5)$$

where productivity α_t is modeled as a first-order autoregressive process, and ϕ is the autocorrelation coefficient. We called this model the Kalman filter autoregressive model (KFAM).

The sixth model that we used was

$$\begin{aligned} Ln(R_t) &= Ln(\alpha_t) + Ln(S_{t-2}) - \beta S_{t-2} + \varepsilon_9 \\ \alpha_t &\in N(a,b) \\ a &\in U(c,d) \end{aligned} \quad (26.6)$$

Table 26.1 Notations used in this paper

Symbols	Meaning
t	Year when recruitment data is measured
R_t	The observed recruitment in year t
S_t	The observed spawner stock size in year t
α_t	Productivity parameter in the SR models
β	Density-dependent parameter in the SR models
$\sigma_{\varepsilon_1}^2$ to $\sigma_{\varepsilon_9}^2$	Variance of residual errors in different models
ϕ	Autocorrelation coefficient in the residual autoregressive model
φ	Autocorrelation coefficient in the K-F autoregressive model
a	Mean of α in the hierarchical Bayesian model
b	Variance of α in the hierarchical Bayesian model
c	Lower bound of the uniform distribution of a
d	Upper bound of the uniform distribution of a

where error ε_9 is independent and normally distributed with mean 0 and variance $\sigma_{\varepsilon_9}^2$. α_t is modeled to follow a hierarchical distribution, i.e., α_t follows a normal distribution $N(a, b)$ with mean a and variance b; a, the mean of α follows a uniform distribution between c and d. The $N(a, b)$ distribution is truncated to make sure that α has positive values. This is a Bayesian hierarchical model, so we called it the hierarchical Ricker model (HRM).

The above models were used to predict the recruitment and to determine the productivity. A summary of the notation used in this chapter can also be found in Table 26.1.

26.2.3 Bayesian Method and Priors

A Bayesian method was used to estimate the parameters in these models. WinBUGS software was used. WinBUGS is numerically intensive software package that implements general Bayesian models using "Metropolis-Hasting within Gibbs sampling" (Gilks 1996; Spiegelhalter et al. 2004). Bayesian implementation of these models requires specification of prior distributions on all unobserved quantities. In general, noninformative priors (here, wide uniform distribution) were used for variances $\sigma_{\varepsilon_1}^2$, $\sigma_{\varepsilon_2}^2$, and so on.

A uniform distribution was used for the prior of $Ln(\alpha)$, i.e., $U(Ln(\alpha_{\min}), Ln(\alpha_{\max}))$, where $\alpha_{\min}$ was determined as the $Min(R_t / S_{t-2})$ and $\alpha_{\max}$ was determined as the $Max(R_t / S_{t-2})$.

Equation (26.1) can be written as $\beta = [Ln(\alpha) + Ln(S_{t-2}) - Ln(R_t)]/S_{t-2}$. A uniform distribution was used for the prior of β, i.e., $U(0.0001, \beta_{\max})$, where $\beta_{\max}$ was determined as the $Max([Ln(\alpha) + Ln(S_{t-2}) - Ln(R_t)]/S_{t-2}) = Max([Ln(\alpha_{\max}) + Ln(S_{t-2}) - Ln(R_t)]/S_{t-2})$.

Priors for ϕ and φ were assumed to be between −1 and 1 with uniform distributions. A lognormal distribution was used for the prior of the spawner stock size with measurement error, i.e., $Ln(S) \sim N(Ln(\overline{S}_t), \text{var}(Ln(S_t)))$, where $\text{var}(Ln(S_t))$ is the variance of the $Ln(S_t)$.

Table 26.2 Priors used for the Bayesian time series models and their posterior median, and standard deviation (Std) and the Deviance Information Criteria (DIC)

Model	Parameters	Prior	Median	Std	DIC
Ricker model	$Ln(\alpha)$	$U(-3.32, 1.55)$	0.29	0.45	198.45
	β	$U(0.0001, 0.31)$	0.04	0.02	
	$\sigma_{\varepsilon_1}^2$	$U(0.0001, 10)$	1.24	0.43	
Measurement error model	$Ln(\alpha)$	$U(-3.32, 1.55)$	0.22	0.46	327.70
	β	$U(0.0001, 0.31)$	0.04	0.02	
	$Ln(S_t)$	$N(2.78, 0.27)$	2.81	0.21	
	$\sigma_{\varepsilon_2}^2$	$U(0.0001, 10)$	1.93	1.71	
	$\sigma_{\varepsilon_3}^2$	$U(0.0001, 10)$	1.20	0.42	
Residual auto-regressive model	$Ln(\alpha)$	$U(-3.32, 1.55)$	0.10	0.35	197.35
	β	$U(0.0001, 0.31)$	0.04	0.02	
	ϕ	$U(-1, 1)$	−0.42	0.24	
	$\sigma_{\varepsilon_4}^2$	$U(0.0001, 10)$	1.17	0.41	
Random walk model	$Ln(\alpha_t)$	$U(-3.32, 1.55)$	−0.1–0.51*	0.47–0.59	200.38
	β	$U(0.0001, 0.31)$	0.04	0.02	
	$\sigma_{\varepsilon_5}^2$	$U(0.0001, 10)$	1.16	0.43	
	$\sigma_{\varepsilon_6}^2$	$U(0.0001, 10)$	0.06	0.11	
Kalman filter auto-regressive model	$Ln(\bar{\alpha})$	$U(-3.32, 1.55)$	0.24	0.24	178.35
	β	$U(0.0001, 0.31)$	0.04	0.01	
	φ	$U(-1, 1)$	−0.51	0.43	
	$\sigma_{\varepsilon_7}^2$	$U(0.0001, 10)$	0.54	0.46	
	$\sigma_{\varepsilon_8}^2$	$U(0.0001, 10)$	0.43	0.37	
Hierarchical Ricker model	$Ln(\alpha_t)$	$N(a, b)$	−0.64–0.62[a]	0.53–0.77	192.99
	β	$U(0.0001, 0.31)$	0.03	0.02	
	$\sigma_{\varepsilon_9}^2$	$U(0.0001, 10)$	0.90	0.48	
	a	$U(-3.32, 1.55)$	0.10	3.68	
	b	$U(0.0001, 10)$	0.31	0.36	

[a]See Fig. 26.4 for the mean values

Besides the above priors, noninformative uniform distributions were used for other variance parameters. Although a common choice of inverse-gamma distribution for variance parameters we used uniform distributions as these have better properties with multilevel models (Gelman 2005). A summary of the priors used in the models above can be found in Table 26.2.

26.2.4 Convergence Diagnostics

A critical issue in using Markov Chain Monte Carlo (MCMC) methods is how to determine when random draws have converged to the posterior distribution. Here, three methods were considered: monitoring the trace, diagnosing the autocorrelation plot, and Gelman and Rubin statistics (Spiegelhalter et al. 2004). In this study, three chains were used. After several sets of analysis, for each chain, the first 20,000 iterations with a thinning interval of 5 were discarded, and another 50,000 iterations were used in the Bayesian analysis.

26.2.5 Model Selection and Bayesian Model Averaging

The DIC is used as a model selection criterion in this study.

$$\begin{aligned} DIC &= 2\overline{D} - \hat{D} \, or \, \overline{D} + p_D \\ D(y,\theta) &= -2\log \text{Likelihood}(\text{y}|\theta) \\ p_D &= \overline{D} - \hat{D} \end{aligned} \tag{26.7}$$

D is the deviance, a measurement of prediction goodness for our models. p_D is called the effective number of parameters in a Bayesian model. The DIC is a hierarchical modeling generalization of the Akaike information criterion (AIC) and Bayesian information criterion (BIC), also known as the Schwarz criterion. It is particularly useful in Bayesian model selection problems, where the posterior distributions of the models have been obtained by MCMC simulation. Like AIC and BIC it is an asymptotic approximation as the sample size becomes large. It is only valid when the posterior distribution is approximately multivariate normal (Spiegelhalter et al. 2002, 2004).

After the above models were solved using Bayesian approaches. DIC values were used to compare different models and to weight different model outcomes when averaging the posterior distribution of the averaged model result (Hoeting et al. 1999; Spiegelhalter et al. 2004). Both MATLAB and WinBUGS languages were used for this purpose. Equation (26.8) showed how the weight of each model is decided, where *k* is the *k*th model.

$$\begin{aligned} \Delta_{\text{DIC}_k} &= \text{DIC}_k - \min(\text{DIC}) \\ weight_k &= e^{-2\Delta_{DIC_k}} \Big/ \sum_i e^{-2\Delta_{\text{DIC}_k}} \end{aligned} \tag{26.8}$$

26.3 Simulation Study

A simulation study was designed to test the performance of the proposed Bayesian model averaging approach and evaluate model selection uncertainty. The following simulation algorithm was used: (1) estimate population productivity parameter $Ln(\alpha)$ and the other parameters from these models using the example species; treat these estimates as true parameters; (2) generate data with uncertainties using a Monte Carlo simulation approach with the uncertainty levels the same as the "true" uncertainties (variances here) estimated from the example population; (3) analyze the generated data set by using different models (the six above, always pick the best model based on model selection criteria and a model averaging approach); (4) evaluate the uncertainty arising from model selection and the performance of Bayesian model averaging (see Table 26.3 for the simulation results).

Procedures (2)–(4) described above were repeated for 500 times to yield 500 sets of estimated population productivity and recruitment over time from each of the models. The sum of square errors (SSE) for log-transformed population productivity in the ith simulation $Ln(\hat{\alpha}_t)$ can be calculated as

$$\mathrm{SSE}(Ln(\hat{\alpha}_i)) = \sum_{t=1}^{k} [Ln(\hat{\alpha}_{t,i}) - Ln(\alpha_t')]^2 \tag{26.9}$$

where i indicates the ith simulation run and k is the numbers of years. The $\hat{\alpha}_{t,i}$ is the estimated α in year t and in the ith simulation; α_t' is the "true" α in year t. The relative estimation error (REE) for estimated log-transformed population productivity in the ith simulation, $\mathrm{REE}(Ln(\hat{\alpha}_i))$, was used here, which was calculated as

$$\mathrm{REE}(Ln(\hat{\alpha}_i)) = \sum_{t=1}^{k} \{[Ln(\hat{\alpha}_{t,i}) - Ln(\alpha_t')] / Ln(\alpha_t')\}^2 \tag{26.10}$$

The REE calculated in Equation (26.10) measures the overall estimation errors including both estimation biases and variations in estimates among the 500 simulation runs. A boxplot was used to summarize the REE_s derived in the 500 simulation runs. An estimation procedure with small REE suggests that it performs well and tends to have smaller estimation error in estimating population productivity and recruitment. The same method was used to calculate $\mathrm{REE}(Ln(\hat{\mathrm{R}}_i))$, relative estimation error for estimated log-transformed recruitment in the ith simulation.

Model selection uncertainty was evaluated through the probability of choosing the true model. For example, when the Ricker model was used as the true model, in each of these 500 runs, the simulation algorithm would pick up the best model based on the DIC values (smallest DIC means the best model); the best model would be recorded in each of the simulation runs. After the 500 runs, the probability of each model chosen as the best model was counted. For example, if the Ricker model was chosen as the best model in 100 of 500 runs, then the probability is 20%.

After the simulation study with standard deviation $\sigma_{\varepsilon_{1:9}}$ used to define the "true" uncertainty levels, a further simulation study with uncertainty levels of 30% of the true uncertainty level were developed, i.e., the 30% of the $\sigma_{\varepsilon_{1:9}}$. The purpose of this study was to investigate whether the model selection uncertainty was controlled by the uncertainty levels.

26.4 Results

The posterior means of the fitted recruitment were different when different models were used (Fig. 26.3). The KFAM, the HRM, and the RWM tended to describe the data better than the other models. The posterior mean and medians

Table 26.3 Relative estimation error (REE) of population productivity estimates and the probability of being the best model among 500 simulation runs when the "true" error values were used. CI is the posterior credible interval

True model used to generate data with uncertainty	models used to fit the simulated data	Estimates						
		REE			DIC			
		Median	5% CI	95% CI	Median	5% CI	95% CI	P^*
Ricker model (RM)	RM	**3.58**	**0.29**	**8.54**	**200.33**	**176.63**	**214.17**	**0.024**
	MEM	3.71	0.38	8.66	325.08	301.98	339.08	0.000
	RAM	3.53	0.28	8.83	202.94	177.97	216.61	0.008
	RWM	5.30	2.20	10.83	201.93	177.59	215.96	0.006
	KFAM	6.19	4.73	9.37	194.80	164.24	210.67	0.636
	HRM	5.75	3.11	10.92	196.53	165.72	211.99	0.326
Measurement error model (MEM)	RM	5.58	0.53	12.62	201.93	179.80	214.96	0.068
	MEM	**5.20**	**0.57**	**11.54**	**333.98**	**306.40**	**352.30**	**0.000**
	RAM	5.23	0.59	12.63	203.79	180.45	217.44	0.012
	RWM	6.30	2.89	13.28	203.59	179.84	218.15	0.000
	KFAM	8.30	6.51	12.22	196.41	166.22	212.43	0.638
	HRM	6.49	4.24	13.63	198.02	166.01	213.58	0.282
Residual autoregressive model (RAM)	RM	7.05	0.76	19.04	211.84	193.35	221.37	0.000
	MEM	6.79	0.71	18.81	336.24	316.71	345.85	0.000
	RAM	**6.55**	**0.62**	**14.89**	**197.99**	**182.91**	**207.35**	**0.292**
	RWM	6.62	2.53	14.30	215.59	197.09	225.04	0.000
	KFAM	29.73	17.95	43.22	193.05	161.90	207.96	0.698
	HRM	12.52	7.99	20.49	210.02	187.35	220.81	0.010

Random walk model (RWM)	RM	11.60	2.81	95.23	213.09	139.61	254.60	0.100
	MEM	11.55	2.81	95.41	338.01	264.26	376.72	0.000
	RAM	10.67	2.72	83.92	214.62	139.33	252.96	0.056
	RWM	**7.10**	**2.34**	**39.66**	**212.39**	**131.73**	**254.03**	**0.068**
	KFAM	10.01	2.28	58.82	207.87	126.68	250.59	0.612
	HRM	8.91	2.62	56.00	211.19	129.51	255.67	0.164
K-F autoregressive model (KFAM)	RM	12.46	5.09	70.69	201.02	178.88	213.09	0.008
	MEM	11.79	5.09	69.76	325.63	302.34	337.75	0.000
	RAM	12.56	5.07	68.56	202.01	180.91	214.11	0.004
	RWM	10.16	5.15	47.60	203.69	180.56	216.17	0.000
	KFAM	**13.45**	**5.26**	**57.97**	**193.27**	**165.73**	**208.13**	**0.740**
	HRM	10.26	4.66	45.35	196.27	166.85	210.91	0.248
Hierarchical Ricker model (HRM)	RM	24.52	8.51	108.88	251.45	227.92	262.94	0.522
	MEM	23.87	8.20	105.28	375.70	350.27	387.97	0.000
	RAM	21.78	7.80	98.66	253.05	227.69	264.36	0.066
	RWM	13.94	5.38	59.08	254.47	229.20	266.02	0.004
	KFAM	15.24	5.63	70.86	251.60	222.43	263.41	0.390
	HRM	**13.05**	**4.99**	**57.86**	**252.30**	**227.04**	**264.22**	**0.018**

P^*: probability of being the best model

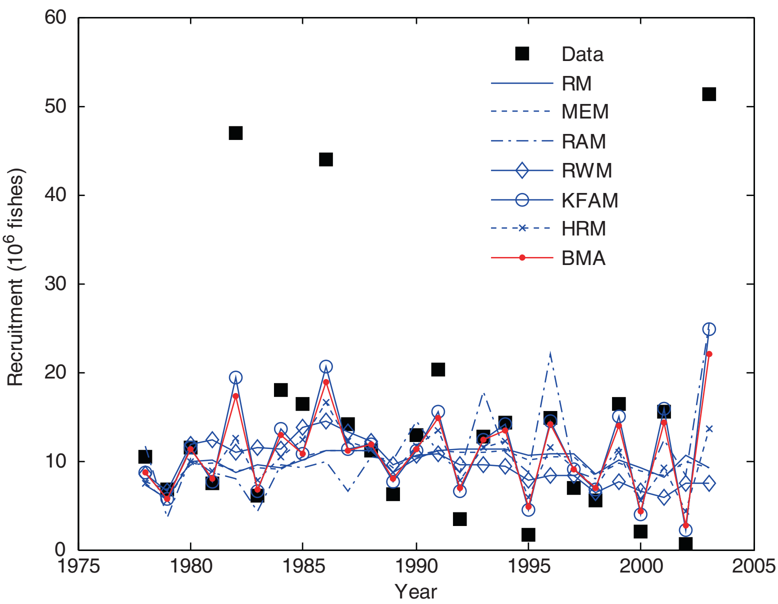

Fig. 26.3 Recruitment data and the estimated recruitment using different models and the Bayesian model averaging results

of the population productivity and the other parameters were different when different models were used (Fig. 26.4 and Table 26.2). The estimated productivity over time from the RWM model showed shift in two regimes. Before mid-1980s, productivity was high but decreased entering into the 1990s. The estimated productivity over time from the KFAM and HRM showed shift in two regimes. Before mid-1980s, productivity was high but decreased entering into the 1990s. This approach also showed that the variation in the productivity is increasing (Fig. 26.4). Among the six models, the KFAM resulted in the lowest DIC value, and the resulting posterior means of the $Ln(\alpha)$ values over time were very different. The HRM and RAM, and the Ricker model resulted in DIC values relatively lower than models other than KFAM (Table 26.2).

From the simulation study based on the example walleye fishery data, we can see that the model selection uncertainty is high because in only 2.4% of the cases the true model was found when the true model was the Ricker model; and it is 0%, 29.2%, 6.8%, 74%, and 1.8% when the true models were the Ricker model, MEM, RAM, RWM, KFAM, and HRM separately (Table 26.3). In general, KFAM tended to be selected as the best model no matter what the true models were. MEM had no chance to be selected as the best model in this study. The Ricker model and RWM have very limited chance to be selected as the best model in the simulation study. The same pattern of model selection uncertainty could be observed from the simulation study when the variance of the errors used in the

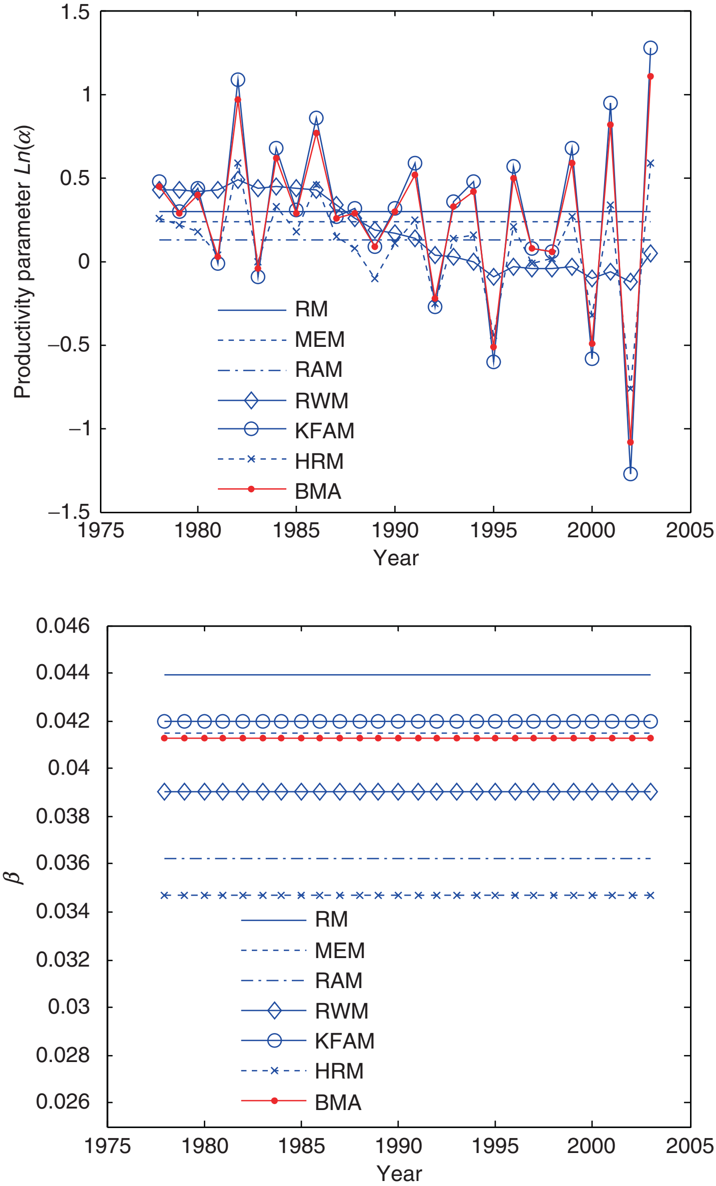

Fig. 26.4 Productivity ($Ln(\alpha)$ here) estimation and parameter β estimation using different models and the Bayesian model averaging results

data simulation is 30% of the "true" variance (Table not shown). KFAM has an extremely high chance to be selected as the best model even when the magnitudes of the noises are low.

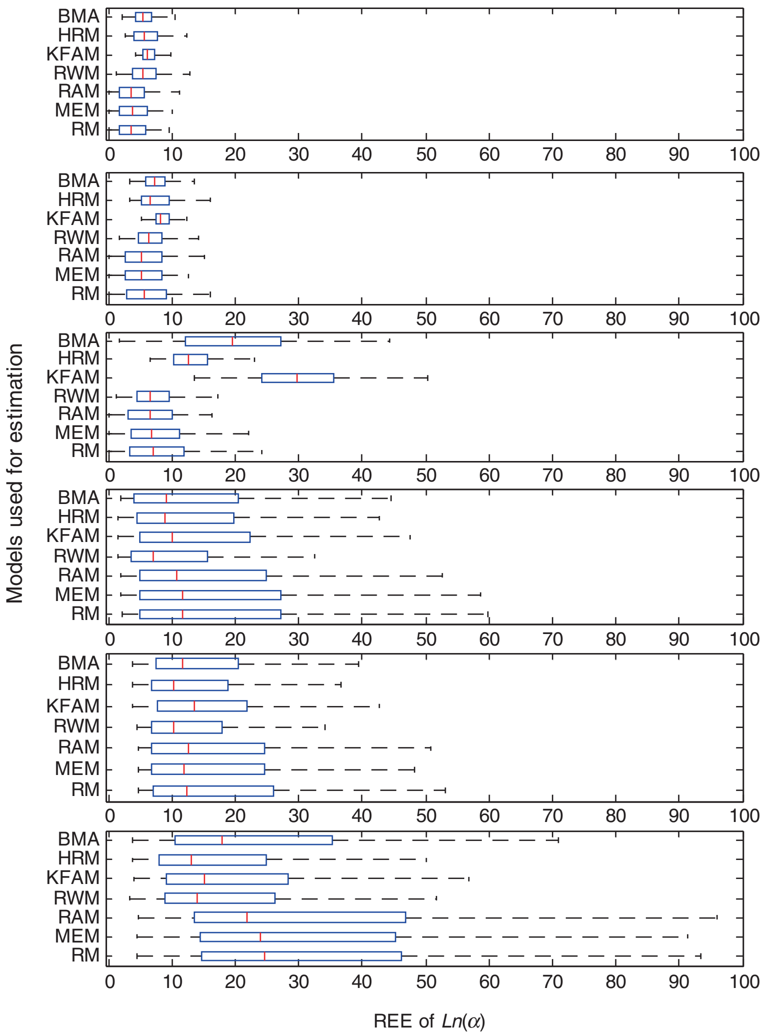

Fig. 26.5 Boxplot of the relative estimation error (REE) of population productivity *Ln*(*α*) estimation in the simulation study when the "true" error values were used. From top to the bottom, the true models used in the simulation are RM, MEM, ARM, RWM, KFAM, and HRM

The REE$_s$ of *Ln*(*α*) were higher when the true models were KFAM, RWM, and HRM (Table 26.3, Fig. 26.5). The magnitude differences of REE$_s$ were caused by the errors used in the simulation study. Though the error values were from the same example fishery, the colored-noise models tended to generate higher noises in magnitude over time (Halley and Kunin 1999). *Ln*(*α*) tended to be highly influenced by the colored noise or the multilevel productivity. This is an important implication in fisheries because the "noise" might not be white and there have been multilevel productivity observed in many fisheries under different environmental regimes (Beamish et al. 1999).

The simulation study also showed that the "true" model tended to give estimates with lower or lowest REE of $Ln(\alpha)$ and $Ln(R)$ (Table 26.3 and Figs. 26.5 and 26.7). MEM tended not to work well, having higher REE values regardless of the true model except when the Ricker model was used as the true model in the simulation study. The REE estimates when the BMA approach was used were low and were very close to the REE estimates when the KFAM was used (Figs. 26.5–26.7). This is because of the low DIC values for KFAM, which also resulted in an extremely high probability of it to be selected as the best model.

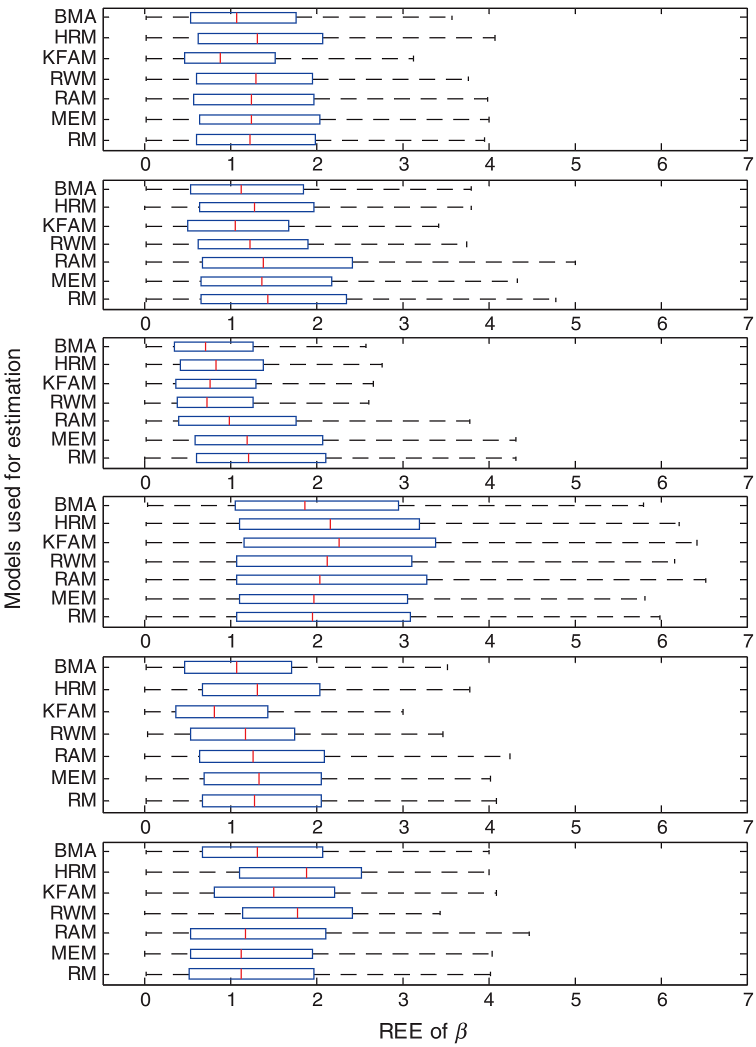

Fig. 26.6 Boxplot of the relative estimation error (REE) of β estimation in the simulation study when the "true" error values were used. From top to the bottom, the true models used in the simulation are RM, MEM, ARM, RWM, KFAM, and HRM

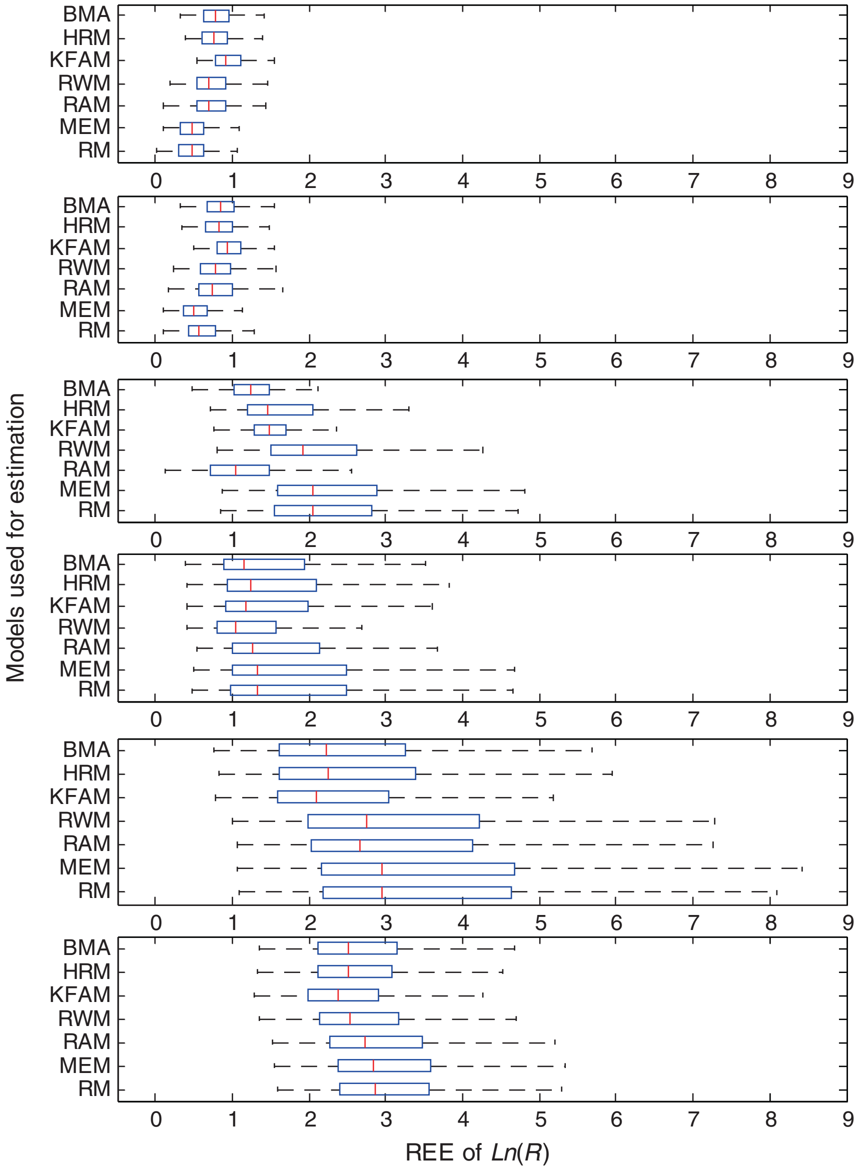

Fig. 26.7 Boxplot of the relative estimation error (REE) of recruitment ($Ln(R)$) estimation in the simulation study when the "true" error values were used. From top to the bottom, the true models used in the simulation are RM, MEM, ARM, RWM, KFAM, and HRM

In this specific example fishery, it looks like that BMA framework was dominated by KFAM. The result also showed that the REE_s did not have to be the lowest even when the model averaging approach was used. However, we need to realize that the simulation study was based on comparing the estimates with the "true" values, such as the "true" $Ln(\alpha)$ and the "true" recruitment, but the real fisheries fact in the last 100 years indicated that noise may be as important as the default "true" models. REE_s for $Ln(R)$ was obviously lower when simulated R was used instead of "true" R and when KFAM was used for estimation. We also need to realize that the corresponding risks of extinction are different when different models are used (Halley and Kunin 1999; Morales 1999; Schwager et al. 2006).

Parameters $Ln(\alpha)$ and β are correlated, which explains some of the REE_s patterns. For example, when the true models were MEM in generating data, the REE_s of $Ln(\alpha)$ were low when RM, MEM, and RAM were used for estimation, and the REE_s of $Ln(\alpha)$ were high when RWM, KFAM, and HRM were used for estimation. However, the REE_s of β were high when RM, MEM, and RAM were used for estimation, and the REE_s of β were low when RWM, KFAM, and HRM were used for estimation. When the true models were KFAM and HRM, the REE_s of $Ln(\alpha)$ were high when RM, MEM, and RAM were used for estimation, and the REE_s of $Ln(\alpha)$ were low when RWM, KFAM, and HRM were used for estimation. However, the REE_s of β were low when RM, MEM, and RAM were used for estimation, and the REE_s of β were high when RWM, KFAM, and HRM were used for estimation.

The analysis indicated that estimated $Ln(\alpha)$ and other parameter values and their creditable intervals (CI) were robust to priors for variances of $\sigma_{\varepsilon 1}{}^2$ to $\sigma_{\varepsilon 7}{}^2$ in the Bayesian time series models, e.g., when uniform distribution or inverse-gamma distribution was used. Little difference was observed, so we did not show the results here because of their robustness to the noninformative priors. However, the informative priors of $Ln(\alpha)$, β, and others do influence the results, and we think that it is necessary to use these informative priors, such as the prior of $Ln(\alpha)$, β, and the spawner stock size in the measurement error model. Information is based on the 26 years of observations and helped to constraint the parameter estimates to a degree.

26.5 Discussion

In this example fishery, the top model was KFAM, according to the model selection criterion DIC. In another population dynamics analysis on an endangered freshwater mussel we found that the top models were the hierarchical Bayesian model, the random walk model, and the exponential growth model through real data analysis and a simulation study (Jiao et al. 2008). However, the simulation study we have discussed showed that for this specific example fishery the KFAM tended to be selected as the best model regardless of the default true model, after adding noise to these models. A further study on other SR data set is suggested to look at the goodness of fit of the KFAM and the other models, which will help us to understand whether this is caused by the data pattern of this walleye fishery or a general phenomenon for all the fisheries. This simulation study also suggests that DIC may not be efficient enough as a model selection criterion in model selection and model averaging. The performance of DIC as a model selection criterion is an ongoing topic in statistics. It is beyond the scope of this study. Considering the fact that no other criterion is currently in use other than DIC, we used this criterion.

In this study, we chose the class of model for BMA based on a good model supported by the data, here the Ricker model, and then averaged over an expanded class of models centered around the good model, such as the residual autoregressive Ricker model, RWM based on the structure of the Ricker model. How to choose the class of models for BMA is discussed by many scientists (Draper 1995; Madigan and Raftery 1994; Hoeting et al. 1999), which incorporate the approach we used here and over the entire class of models or over models with different error structures. We would support all these approaches if they can explain the model from the biological data, i.e., any models supported by the data. The BMA approach provides a very flexible framework to incorporate different models and it balances the model selection uncertainty and the goodness of fit of different models incorporated in the BMA framework. It is important to incorporate good models in the BMA framework and at the same time weak models do not need to be worried about because they will be less weighted if they are less supported by the data.

Traditional SR models such as the Ricker model and the BH model are widely used because of their simplicity while models such as RAM, RWM, KFAM, and HRM are rarely used in current fisheries because of the difficulty in estimating parameters. MCMC provides a very convenient way to solve these models and Bayesian approaches are the only computational, possible, or convenient method to solve above models (De Valpine and Hasting 2002). Models such as RAM, RWM, KFAM, and HRM are highly recommended to be incorporated into such a model averaging framework because they are much better in capturing the noise or stochastic component than the traditional SR models. The RWM and KFAM were said to be better at describing the trend of the population productivity changes over time (Peterman et al. 2000, 2003).

Random noise caused by environmental variation can be more important than the "true" population dynamic pattern. This is especially true and crucial for populations with small stock size (Halley and Kunin 1999; Morales 1999; Petchey 2000; Schwager et al. 2006). A BMA approach is suggested to incorporate these models describing various possible noise components. A simulation study is recommended in studying SR modeling and risk analyses, which would help to understand the model selection uncertainty. The approach should also help lower the risk of misunderstanding the population status, and help to improve the problems involved in our modeling approach and better manage our natural resources.

High model selection uncertainty implies that estimating population parameters and evaluating the population status based on one model or one best model is dangerous. In theory, BMA provides better average predictive performance than any single model that could be selected, and this theory has been supported by many examples (Hoeting et al. 1999). The application of BMA in fisheries is still very limited. Based on the fact that the model selection uncertainty is extremely high in fisheries and the model averaging approach works well, a model averaging approach is recommended in dealing with these SR relationship and the other models in fisheries, such as the catchability model used in age-aggregated models and the age- or size-structured models, the productivity function used for the surplus production models, and the error structure used in both the age-aggregated models and the age- or size-structured models (Polacheck et al. 1993; Harley et al. 2001; De Valpine and Hasting 2002; Jiao and Chen 2004; Jiao et al. 2006). A lot of discussions are currently going on among scientists as to which model should be used. The answer may be different for different fisheries. However, it is very difficult to assess because of the uncertainties in the data and in the model selection. The BMA approach balances the needs for best models and model selection uncertainty, takes into account these important sources of uncertainty and provides better inferences about parameters. It eventually lowers the risk of mis-managing our fisheries through better stock assessment.

Acknowledgement This research was supported by Virginia Polytechnic Institute & State University's new faculty start-up funding to Y. Jiao, the USDA Cooperative State Research, Education and Extension Service, Hatch project #0210510 to Dr. Y. Jiao, and grant for Decision Analysis and Adaptive Management of Lake Erie Percid Fisheries awarded to Y. Jiao by the Ontario Commercial Fisheries Association.

References

Beamish RJ, Noakes DJ, McFarlane GA, Klyashtorin L, Ivanov VV, Kurashov V (1999) The regime concept and natural trends in the production of Pacific salmon. Can J Fish Aquat Sci 56:516–526

Caswell H, Cohen JE (1995) Red, white and blue: environmental variance spectra and coexistence in metapopulations. J Theor Biol 176:301–316

De Valpine P, Hasting A (2002) Fitting population models incorporating process noise and observation error. Ecol Monogr. 72(1):57–76

Draper D (1995) Assessment and propagation of model uncertainty (with discussion). J Roy Statist Soc Ser B 57:45–97

Gelman A (2005) Prior distributions for variance parameters in hierarchical models. Bayesian Anal 2:1–19

Gilks WR (1996) Full conditional distributions. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov Chain Monte Carlo in practice. Chapman & Hall, London, pp 75–78

Glantz MH (1992) Climate variability, climate change and fisheries. Cambridge University Press, Cambridge

Halley JM (1996) Ecology, evolution and 1/f-noise. Trends Ecol Evol 11:33–37

Halley JM, Kunin WE (1999) Extinction risk and the 1/f family of noise models. Theor Pop Biol 56:215–230

Harley SJ, Myers RA, Dunn A (2001) Is catch-per-unit-effort proportional to abundance? Can J Fish Aquat Sci 58:1760–1772

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Statist Sci 14:382–417

Jiao Y, Chen Y (2004) An application of generalized linear model in production model and sequential population analysis. Fish Res 70:367–376

Jiao Y, Reid K, Nudds T (2006) Variation in the catchability of yellow perch (*Perca flavescens*) in the fisheries of Lake Erie using a Bayesian error-in-variable approach. ICES J Mar Sci 63:1695–1704

Jiao Y, Neves R, Jones J (2008) Models and model selection uncertainty in estimating growth rates of endangered freshwater mussel populations. Can J Fish Aquat Sci 65:0000–0000 (in press)

Madigan D, Raftery AE (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. J Am Statist Assoc 89:1535–1546

Morales JM (1999) Viability in a pink environment: why "white noise" models can be dangerous. Ecol Lett 2:228–232

Petchey OL (2000) Environmental colour affects aspects of single-species population dynamics. Proc Roy Soc Lond B 267:747–754

Peterman RM, Pyper BJ, Grout JA (2000) Comparison of parameter estimation methods for detecting climate-induced changes in productivity of Pacific salmon (*Onchorhynchus* spp.). Can J Fish Aquat Sci 57:181–191

Peterman RM, Pyper BJ, MacGregor BW (2003) Use of the Kalman filter to reconstruct historical trends in productivity of Bristol Bay sockeye salmon (*Oncorhynchus nerka*). Can J Fish Aquat Sci 60(7):809–824

Polacheck T, Hilborn R, Punt AE (1993) Fitting surplus production models: comparing methods and measuring uncertainty. Can J Fish Aquat Sci 50:2597–2607

Quinn II TJ, Deriso RB (1999) Quantitative fish dynamics. Oxford University Press, Oxford, 542 pp

Schwager M, Johst K, Jeltsch F (2006) Does red noise increase or decrease extinction risk? Single extreme events versus Series of unfavorable conditions. Am Nat 167:879–888

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). J Roy Statist Soc B 64:583–640

Spiegelhalter DJ, Thomas A, Best N, Lunn D (2004) WinBUGS user manual (version 1.4.1). MRC Biostatistics Unit, Cambridge, UK

Vasseur DA, Yodzis P (2004) The color of environmental noise. Ecol 85:1146–1152

Walleye Task Group (WTG) (2004) Report of the walleye task group to the standing technical committee, Lake Erie Committee of the Great Lakes Fishery Commission. 26 pp

Walters CJ, Ludwig D (1981) Effects of measurement errors on the assessment of stock–recruit relationship. Can J Fish Aquat Sci 38:704–710