This is to certify that the
dissertation entitled

IMPROVING STATISTICAL CATCH-AT-AGE STOCK ASSESSMENTS

presented by

MICHAEL J. WILBERG

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Fisheries and Wildlife

James R. Bence
Major Professor's Signature

April 22, 2005
Date

IMPROVING STATISTICAL CATCH-AT-AGE STOCK ASSESSMENTS

By

Michael J. Wilberg

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Fisheries and Wildlife

2005

ABSTRACT

IMPROVING STATISTICAL CATCH-AT-AGE STOCK ASSESSMENTS

By

Michael J. Wilberg

My dissertation addresses three objectives: 1) to estimate fishing mortality rates and abundance of yellow perch in southwestern Lake Michigan during 1986-2002 to determine the contribution of fishing to the collapse of yellow perch in southwestern Lake Michigan, 2) to determine robust methods of dealing with time-varying fishery catchability within a statistical catch-at-age analysis (SCA) framework, and 3) to determine whether using Bayesian model selection, specifically Deviance Information Criterion (DIC) and an approximation of Bayes factors, results in using accurate models for prediction of important fisheries management quantities.

In chapter 1, I conducted an age-, size-, and sex-structured stock assessment of yellow perch to estimate population size and examine historical trends in fishing mortality in Illinois and Wisconsin waters of southwestern Lake Michigan. Model estimates indicated that yellow perch abundance in 2002 was less than 10% of 1986 abundance in Wisconsin and about 20% in Illinois. Annual mortality rates for females age 4 and older averaged 69% during 1986-1996 in Wisconsin and 60% in Illinois during 1986-1997, which are quite high for a species like yellow perch that can live longer than 10 years. Estimated fishing mortality rates on adult females during 1986-1996 exceeded widely used reference points, suggesting that overfishing may have occurred. I believe unsustainably high fishing mortality rates were a substantial contributing cause of the rapid decline of mature females in the mid-1990s.

The relationship between fishing mortality and fishery effort (catchability) may change over time through either density dependent or density independent processes. I used Monte Carlo simulations in chapter 2 to evaluate how different methods of estimating fishery catchability within an SCA model performed when models were confronted with different data generating scenarios. I evaluated performance of the estimation models by their accuracy and precision in determining quantities of interest such as biomass in the last year. In many cases, including fishery effort data in the estimation model and allowing catchability to follow a random walk performed as well or better than other methods. Exceptions were cases where fishing mortality was low and catchability trended over time. The estimation model that ignored fishery effort data performed well in cases with a good survey, but performance degraded as survey precision decreased. White noise and density dependent estimation models performed poorly in situations where catchability trended over time. No estimation model was best for all underlying models of catchability, hence I recommend fitting multiple SCA models with alternative assumptions.

Structural flaws in SCA models may cause considerable bias in model estimates of mortality rates, abundance, and recruitment. I used simulations to evaluate whether using Deviance Information Criterion (DIC) or approximate Bayes factors to select the best SCA model provided more accurate estimates of quantities important for management than using a single model in all cases. Using the model selected by DIC or approximate Bayes factors resulted in estimates with lower mean square errors than using any single model.

ACKNOWLEDGMENTS

I am indebted to many people for helping me during my doctoral program. I especially thank my major advisor, Dr. Jim Bence, who was always available and helpful whenever I had questions or needed guidance. Jim has been a superb mentor and it has truly been a pleasure to work with him. I thank the rest of my graduate committee members, Drs. Dan Hayes, Mike Jones, and Rob Tempelman, for providing invaluable assistance. Emily Szalai, Mike Rutter, Norine Dobiesz, Brain Linton, and Wenjing Dai provided programming support. Genny Nesslage provided valuable editing help, encouragement, and general support that helped keep me relatively sane. Members of the Lake Michigan Yellow Perch Task Group, particularly Brad Eggold, Dave Clapp, and Dan Makauskas, and three anonymous reviewers provided helpful comments for chapter 1. Data used in chapter 1 were collected and provided by Ball State University, the Illinois Department of Natural Resources, the Illinois Natural History Survey, the Indiana Department of Natural Resources, and the Wisconsin Department of Natural Resources. Dave Glover and John Dettmers provided information on yellow perch movement in southern Lake Michigan. This research was supported in part by Michigan Sea Grant, the U.S. Fish and Wildlife Service Great Lakes Fish and Wildlife Restoration program, the Great Lakes Fishery Commission, Michigan Sea Grant, the International Association for Great Lakes Research, and the Michigan State University College of Agriculture and Natural Resources. The Bence/Jones lab provided encouragement, assistance, and much needed distractions. Finally, I thank my family for unwavering and unconditional support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Fishery managers need realistic predictions of future population dynamics of individual fish stocks and predictions of how these populations will respond to management actions. Most major fisheries are managed by a process where scientists estimate population size and other parameters of a fish population (and uncertainty of these estimates) and provide this information to fishery managers who then make decisions regarding which fishery policies to implement (e.g., catch quotas, bag limits, season or area closures). The process of estimating these quantities is called stock assessment.

Relatively recent advances in fisheries science have allowed researchers to estimate total abundance from fishery harvest and age or length composition data, and other diverse data sources, with a method known as statistical catch-at-age analysis (SCA; Fournier and Archibald 1982; Deriso et al. 1985; Megrey 1989; Methot 1990). This approach is preferable, in many cases, to other stock assessment methods because it can incorporate many diverse data sources and allows for a rigorous statistical approach (i.e., promotes explicit modeling of measurement and process error). Hence, SCA can allow estimation of uncertainty associated with parameter estimates and other model quantities. SCA methods are being applied worldwide for many fisheries and predominates applications to major marine fisheries in the northwestern U.S., New Zealand, Australia, and South Africa (Radomski et al. in press).

The basic idea behind SCA is that one can infer the effect of fishing on a population by estimating how absolute removals (e.g., fishery harvest or yield) affect

relative abundance. A model is created that describes the population and the process of removals, and this model is statistically fit to data from a fishery. Usually, one of the key assumptions of these models, called the separability assumption, is that fishing mortality can be described by an overall year effect (how a certain amount of fishing effort affects a population) and an age effect (the relative vulnerability of different aged fish to a fishery). This basic approach has also been extended to species for which there is no directed fishery (Szalai 2003).

In southern Lake Michigan, yellow perch abundance has declined substantially since the mid-1980s (Marsden and Robillard 2004). As the abundance of yellow perch declined during the mid to late 1990s, commercial fisheries in Indiana, Illinois, and southern Wisconsin were restricted to smaller quotas, and were eventually closed during 1996-1997 (Francis et al. 1996). Stricter regulations were also imposed on the recreational fishery with reductions in daily bag limits implemented in all states during 1996-1998, the incorporation of a slot size limit (i.e., only fish between 8 and 10 in could be kept) in Illinois during 1997-2000, and seasonal closures of the fishery (Marsden and Robillard 2004). Reproductive failure has been implicated as the primary cause of the population collapse, but the role of fishing in the collapse has not previously been rigorously investigated. My research investigates the role of fishing in the decline of yellow perch in southwestern Lake Michigan by using SCA models.

A frequent (but somewhat outdated) criticism of SCAs is that they do not allow for the flexibility to accurately model time-varying fishing mortality at age (NRC 1998). Specifically, the relationship between fishery effort and fishing mortality or the age-based vulnerability to the fishery may change over time (Butterworth et al. 2002; Radomski et

al. in press). Many methods have been developed to account for these changes over time (e.g., Fournier and Archibald 1982; Fournier 1983; Methot 1990; Hampton and Fournier 2001; Butterworth et al. 2003), but there is not consensus on which methods are best when faced with different underlying mechanisms for change. My research also aims to evaluate performance of several SCA methods under many situations and to develop guidelines to help researchers decide among several SCA model structures.

My dissertation addresses three objectives: 1) to estimate fishing mortality rates and abundance of yellow perch in southwestern Lake Michigan during 1986-2002 to determine the contribution of fishing to the collapse of yellow perch in southwestern Lake Michigan, 2) to determine robust methods of modeling time-varying catchability within an SCA framework, and 3) to determine whether using Bayesian model selection, specifically Deviance Information Criterion (DIC) and an approximation of Bayes factors, results in choosing models with accurate estimates of fishing mortality rates and abundance. These objectives arose out of questions that formed during my research program, and each chapter of my dissertation addresses an objective. Chapter 1 developed from a management need to evaluate the importance of fishing in the population declines of yellow perch in southern Lake Michigan. In working on chapter 1, I found that many of the model parameters were likely time-varying and wanted to determine whether the approaches I used (or alternatives) were robust methods for modeling these processes. This led to the evaluation of several methods for incorporating time-varying catchability into SCA models detailed in chapter 2. Based on the results of chapter 2 (differential performance of SCA models under different data-generating scenarios), the question arose as to whether statistical model selection techniques could

be used to select "good" (i.e., accurate) models when one does not know the true underlying pattern or mechanism of change.

Objective 1

In chapter 1, I detail my assessment model and describe results from the modeling efforts. I developed a length-, age-, and sex-based SCA model to estimate fishing mortality rates and abundance, to determine if fishing mortality rates exceeded the maximum that could be supported, and to integrate diverse sources of data to get the best estimates of recruitment and population size. My model allowed fishing mortality rates at age and sex to change over time 1) in response to changes in fishery effort, 2) by allowing fishery catchability to change according to random walk models, and 3) by modeling fishery selectivity as a function of length and allowing growth to change over time to match observed changes in size at age and sex in southern Lake Michigan.

Model estimates of catchability of the recreational fishery changed approximately five-fold during 1986-2002, and commercial fishery catchability changed approximately four- and eight-fold in Illinois and Wisconsin respectively. However, fishing mortality rates changed approximately 15-fold for females and eight-fold for males during 1986-2002, indicating that changes in effort and catchability were both important to changes in fishing mortality. This leads to questions of whether modeling fishery catchability as a random walk, as in this application, is the best approach and whether fishery effort data should be used at all (because of the extreme changes in fishery catchability).

Objective 2

Many SCAs of fish stocks assume that fishing mortality is directly proportional to fishing effort (i.e., constant catchability). However, fishery catchability has often

changed in response to changes in population abundance (e.g., Peterman and Steer 1981), environmental conditions (e.g., Ziegler et al. 2003), or changes in fishing gear or fisherman experience (e.g., Hilborn and Walters 1992 pp. 126, 130). Likewise, catchability in yellow perch fisheries in southwestern Lake Michigan has changed substantially over time, perhaps due to a combination of the factors listed above. Therefore, my second chapter describes an evaluation of several methods of modeling time-varying catchability within SCA models. I used Monte Carlo simulations to compare how four different methods of estimating fishery catchability within an SCA model performed when models were confronted with five different data generating scenarios.

In many cases, including fishery effort data in the estimation model and allowing catchability to follow a random walk performed better than ignoring fishery effort data. Exceptions were cases where fishing mortality was low and catchability trended over time. The estimation model that ignored fishery effort data performed well in cases with a good survey, but performance degraded as survey precision decreased. White noise and density dependent estimation models performed poorly in situations where catchability trended over time. No estimation model was best for all underlying models of catchability.

Objective 3

Structural flaws in SCA models may cause considerable bias in model estimates of mortality rates, abundance, and recruitment (McAllister and Kirchner 2002). Often researchers will make *ad hoc* decisions about model structure that may cause substantial biases in their ensuing model estimates (Burnham and Anderson 2002; Gavaris and

Ianelli 2002). Given that a wide variety of models can potentially describe dynamics of a given stock, methods to decide among several SCA models are needed. Helu et al. (2000) evaluated performance of Akaike's Information Criterion (AIC; Akaike 1973) and Schwartz's Bayesian Information Criterion (BIC; Schwartz 1978) in SCA models and found that AIC and BIC both performed well by selecting the candidate model that was the same as the data-generating model in most of their scenarios. Unfortunately, although AIC or BIC may perform well in some cases, their implementation is problematic when models differ in their random effects or hierarchical structures because the number of parameters in these models is not easy to determine (Burnham and Anderson 2002). Therefore, to be able to compare structurally complex SCA models requires alternative model selection approaches that can account for random effects and priors on parameters. DIC and Bayes factors are generally considered practical methods to choose the best model from a set of candidate models and do not require the user to specify the number of model parameters. However, performance of model selection using DIC and Bayes factors has not been evaluated for SCA models.

My third chapter evaluates whether using DIC or an approximation of Bayes factors results in choosing accurate models from the set of candidate models. Specifically, I was interested in whether Bayesian model selection could determine an appropriate model structure for time varying catchability because catchability is one of the most important parameters in SCA models (scales abundance relative to catch), catchability varied widely over time in yellow perch fisheries in southwestern Lake Michigan, and the accuracy of different structural forms of SCA models differs depending on the underlying true changes in catchability, quality of data, and average

fishing mortality rate. To achieve these objectives, I designed a simulation study and challenged the model selection criteria with three estimation models, which differed in how catchability was allowed to vary over time, and three scenarios of data accuracy and time-varying catchability. I evaluated whether using DIC and approximate Bayes factors to select among SCA model variants provided more accurate estimates of quantities used for management than an approach of using a single model structure in all cases.

References

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki eds. Second International Symposium on Information Theory 267-81. Budapest: Akademiai Kiado.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Springer-Verlag, New York.

Butterworth, D. S., J. N. Ianelli, and R. Hilborn. 2003. A statistical model for stock assessment of southern bluefin tuna with temporal changes in selectivity. African Journal of Marine Science 25:331-361.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Aquatic Sciences 42: 815-824.

Fournier, D. A. 1983. An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density dependent effects. Canadian Journal of Fisheries and Aquatic Sciences 40: 1233-1243.

Fournier, D. A., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195-1207.

Fournier, D. A., J. Hampton, and J. R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences 55: 2105-2116.

Francis, J. T., S. R. Robillard, and J. E. Marsden. 1996. Yellow perch management in Lake Michigan: a multi-jurisdictional challenge. Fisheries 21:18-20.

Gavaris, S., and J. N. Ianelli. 2002. Statistical issues in fisheries' stock assessments. Scandinavian Journal of Statistics 29:245-271.

Hampton, J., and D. A. Fournier. 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Marine and Freshwater Research 52:937-963.

Helu S. L., D. B. Sampson, and Y. Yin. 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Canadian Journal of Fisheries and Aquatic Sciences 57:1784-1793.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment. Chapman and Hall, New York.

Marsden, J. E., and S. R. Robillard. 2004. Decline of yellow perch in southwestern Lake Michigan, 1987-1997. North American Journal of Fisheries Management 24:952-966.

McAllister, M, and C. Kirchner. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. Bulletin of Marine Science 70: 499-540.

Megrey, B. A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. In Mathematical analysis of fish stock dynamics. Edited by E. F. Edwards and B. A. Megrey. American Fisheries Society Symposium 6:8-48.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Pages 259-277 in L. Lowe ed. Proceedings of the symposium on applications of stock assessment techniques to gadids. International North Pacific Fisheries Commission Bulletin 50.

National Research Council (NRC). 1998. Improving fish stock assessments. National Academy Press. Washington D.C.

Peterman, R. M., and G. J. Steer. 1981. Relation between sport-fishing catchability coefficients and salmon abundance. Transactions of the American Fisheries Society 110: 585-593.

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

Radomski, P. A., J. R. Bence, and T. J. Quinn II. In press. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational kill dominated fishery. Canadian Journal of Fisheries and Aquatic Sciences.

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

Szalai, E. B. 2003. Uncertainty in the population dynamics of alewife (*Alosa psuedoharengus*) and bloater (*Coregonus hoyi*) and its effects on salmonine stocking strategies in Lake Michigan. Ph.D. Dissertation, Michigan State University.

Ziegler, P. E., S. D. Frusher, and C. R. Johnson. 2003. Space-time variation in catchability of southern rock lobster Janus edwardsii in Tasmania explained by environmental, physiological and density-dependent processes. Fisheries Research 61: 107-123.

CHAPTER 1

YELLOW PERCH DYNAMICS IN SOUTHWESTERN LAKE MICHIGAN DURING 1986-2002

Introduction

Yellow perch *Perca flavescens* is an ecologically and economically important species in Lake Michigan (Wells and McLain 1972). Yellow perch are native to Lake Michigan, play an important role in near-shore energy cycling and transfer (Evans 1986), and have provided a fishery on Lake Michigan since the late 1800s (Wells and McLain 1972; Wells 1977). Yellow perch is the only native species in Lake Michigan that has supported a commercial fishery continuously during the last century (Baldwin et al. 1979), although the fishery has only continued in Green Bay since 1998. During the 1980s and 1990s, the recreational fishery harvested more yellow perch than any other species in Lake Michigan (Bence and Smith 1999).

In southern Lake Michigan, yellow perch abundance underwent periodic fluctuations during 1934-1964, and declined greatly during the 1960s (Francis et al. 1996). The decline in yellow perch abundance in the 1960s coincided with a large increase in alewife *Alosa pseudoharengus* abundance, and therefore alewife interference with yellow perch reproduction (either through competition or predation) was considered the primary cause of the decline (Wells 1977). However, exploitation was also considered a contributing factor to the overall decline and the primary cause of the decline of adults (Wells 1977). Prior to 1969, all the states bordering Lake Michigan (Indiana, Illinois, Michigan, and Wisconsin) had commercial fisheries for yellow perch

(Baldwin et al. 1979). In 1969, the state of Michigan was the first to close their commercial fishery (Wells 1977). During the 1970s, yellow perch populations in southern Lake Michigan began to recover (Wells and Jorgenson 1983), and abundance was high during the 1980s with strong year-classes in 1980 and 1983-1988 (Jude and Tesar 1985; Makauskas and Clapp 2000). Abundance declined to low levels during the 1990s with a series of weak year-classes during 1989-1997 and 1999-2000. As yellow perch abundance declined, the sex ratio became skewed toward males, which may have been caused by intense fishing mortality targeted on large females (Madenjian et al. 2002). The selective removal of large females may have led to further declines in yellow perch recruitment. As the abundance of yellow perch declined in southern Lake Michigan during the mid to late 1990s, commercial fisheries in Indiana, Illinois, and southern Wisconsin were restricted to smaller quotas (Francis et al. 1996), and were eventually closed during 1996-1997; these fisheries remain closed. Stricter regulations were also imposed on the recreational fishery with reductions in daily bag limits implemented in all states during 1996-1998, the incorporation of a slot size limit in Illinois during 1997-2000, and seasonal closures of the fishery (Francis et al. 1996). Reproductive failure has been implicated as the primary cause of the population collapse (Francis et al. 1996; Heyer et al. 2001; Marsden and Robillard 2004), but the role of fishing in the collapse has not been rigorously investigated.

Our objectives were to estimate fishing mortality rates and abundance of yellow perch in Wisconsin and Illinois waters of southwestern Lake Michigan during 1986-2002 to determine the contribution of fishing to the collapse of yellow perch in southern Lake Michigan. We also wanted to determine if fishing mortality rates exceeded the maximum

that could be supported, and to integrate diverse sources of data to get the best estimates of recruitment and population size. Our approach was to fit age-, size-, and sex-structured population models to fishery and survey data. No previous population model-based stock assessments have been conducted for yellow perch in southern Lake Michigan. Similar age-structured assessments have been applied to lake trout *Salvelinus namaycush* (Sitar et al. 1999), lake whitefish *Coregonus clupeaformis* (Ebener et al. in press), walleye *Sander vitreus* (Deriso et al. 1988), and yellow perch (Lake Erie Yellow Perch Task Group 2001) in other areas of the Great Lakes.

Methods

We implemented statistical catch-at-age models (detailed description in Appendix A) for yellow perch in southwestern Lake Michigan (Figure 1.1). Statistical catch-at-age models are age-structured models that follow cohorts of fish over time and consider the catch-at-age data to be measured with error (Megrey 1989). Such models consist of population and observation submodels, where the model parameters are estimated by fitting the models to data (Megrey 1989). Our assessment models contained annual time intervals and considered the period from 1986 to 2002, and ages 2 through 9 (age 9 was an aggregate age class that included all fish age 9 and older). We began our models in 1986 because recreational fishery data were not available for earlier years. During model development, we tested the effect of sequentially changing the aggregate age class lower (down to age 6) and results were similar to those we report. Our models also contained two fisheries, recreational and commercial, and a fishery independent gillnet survey. Our models produced estimates of fishing mortality rates, abundance, biomass, and spawning stock biomass (SSB). We defined SSB as the biomass of mature females in the

population and calculated this based on a length-based maturation curve derived outside our model fitting process (see Appendix A).

Our assessment model was age-, size-, and sex-structured. In statistical catch-at-age models, relative vulnerability (i.e., selectivity) to the fisheries is usually modeled as a time-invariant function of age (Quinn and Deriso 1999). However, this assumption does not appear to be reasonable for yellow perch, because these fisheries are highly size selective (Kraft and Johnson 1992) and yellow perch size-at-age has changed substantially over time (Marsden and Robillard 2004). Also, yellow perch show sexually dimorphic growth, with females growing faster and to larger sizes than males, which is suspected to cause higher fishing mortality rates for females (Wells and Jorgenson 1983; Madenjian et al. 2002). We modeled selectivity of the fisheries and surveys as functions of length and allowed growth to change over time with a time-varying von Bertalanffy growth model (Szalai et al. 2003). We accounted for temporal variations in growth by allowing the von Bertalanffy parameters to change in accord with random walk submodels (see Appendix A). Our approach allowed the relative vulnerability of different age-sex categories of yellow perch to change over time as their mean length-at-age changed, even though relative vulnerability was a constant function of length that did not differ between the sexes (Methot 1990; Hampton and Fournier 2001). We also included a different selectivity pattern to capture changes in recreational fishery selectivity during 1997-2000 when a slot size limit was implemented in Illinois. We assumed a time-, sex-, and age-invariant natural mortality rate, M, of 0.37, which was consistent with estimates of M for yellow perch in Indiana waters of southern Lake

Michigan (Allen 2000) and with values used for stock assessments of yellow perch in Lake Erie (Lake Erie Yellow Perch Task Group 2001).

As well as allowing for changes in the relative vulnerability of different ages in response to changes in growth, our model allowed for temporal changes in the vulnerability of the most selected size of yellow perch, so that the fishing mortality imposed by a given amount of fishing effort could change over time. As for the growth model this was done by having fishery catchability parameters vary according to random walk models (see Appendix A).

Genetic analyses have found that yellow perch in the southern basin of Lake Michigan form a single genetic stock (Miller 2003). However, our approach implicitly assumed that there was no net migration for either of the model areas (Illinois, and Wisconsin WM-4 to WM-6; Figure 1.1). We believe this assumption is a reasonable approximation because preliminary tagging data suggest that the median dispersal distance for adult yellow perch in southwestern Lake Michigan was relatively low (< 30 km; D. Glover, University of Illinois at Urbana-Champaign, personal communication). Also, Horns (2001) attributed differences in growth patterns among yellow perch stocks in southern Lake Michigan to geographic segregation. Evidence from physical current modeling studies suggests that genetic structure of the yellow perch population of southern Lake Michigan may be caused by mixing during the larval stage (Beletsky et al. 2004).

Model Fitting

We took a Bayesian approach to obtain posterior probability estimates for the parameter values and quantities of interest such as fishing mortality rates, abundance,

biomass, and SSB. We fitted our models to commercial yield, recreational harvest, commercial length frequency, recreational length frequency, commercial effort, recreational effort, mean length-at-age in the survey, age composition of the survey by sex, total survey CPE by sex, and survey length composition by sex. The objective function contained 11 additive components for the Wisconsin model and 12 additive components for the Illinois model (Appendix A). Each component represented a type of data or a specified informative distribution (i.e., prior distribution) for parameters. Variations in catchability and growth model parameters according to random walks were included as components. We estimated 149 parameters for the Wisconsin model and 151 parameters for the Illinois model. We used Markov Chain Monte Carlo (MCMC) simulations with a Metropolis-Hastings algorithm to estimate posterior probability intervals (the Bayesian analog of confidence intervals) of several model parameters and estimates (Otter Research Limited 2000). We ran the MCMC chain for 2,000,000 steps, sampling every 250 steps, and discarded samples from the initial 250,000 steps as a burn in period, which reduces the effect of starting values on the MCMC results (Gelman et al. 2004). We determined that the length of our burn in period was long enough by separating the MCMC chains (of the objective function) into several smaller chains and comparing the distributions of these blocks (Gelman et al. 2004); the distribution of each block was nearly identical to the other blocks.

We assumed that total catch for all fisheries was median-unbiased, and that the coefficient of variation (CV) of the catches was constant for each fishery (i.e., we assumed lognormal errors). We set the CV for the commercial fishery by assuming that recorded yield was accurate to within approximately 10% in Illinois and 20% Wisconsin

95% of the time. The CV for the recreational fishery was set to approximately 10% based on estimates of the CV from the Wisconsin recreational fishery during 1998-2001 (Wisconsin Department of Natural Resources [WDNR], unpublished data). Independent estimates for the CV of the Illinois recreational fishery were not available. The CVs of survey CPEs and effective sample sizes of the age and length compositions of the surveys and recreational and commercial fisheries were estimated using an iterative approach where we adjusted the assumed initial CVs and effective sample sizes of the objective function components to match the residual variance (McAllister and Ianelli 1997). Effective sample sizes for survey age composition determined by otoliths or anal fin spines were weighted five times higher than those determined by scales because scale aging is thought to be a less accurate method of aging yellow perch (Baker and McComish 1998; Robillard and Marsden 1996; Wisconsin Department of Natural Resources, unpublished data). For the Illinois model, we set the CVs of the random walk deviations for commercial and recreational catchability to about 25%. For the Wisconsin model, we used the same CV for recreational fishery catchability, but used a higher CV of about 40% for commercial catchability because, based on the large amounts of unreported catch, we thought the commercial effort data were less accurate for Wisconsin than for Illinois. For the Wisconsin model, we set the CVs to about 5% for the random walk deviations for the L_∞ and K parameters of the growth model because mean length-at-age of the older age groups rarely changed rapidly from year to year. In contrast, we set the CV of the random walk deviations for mean length-at-age 2 to 10% because mean length-at-age 2 showed more variation from year to year than older ages. Using the same CV values for L_∞ and K in the Illinois model as in the Wisconsin model resulted in poor

convergence. Therefore, we set the CVs on L_∞ and K to about 2.5% to further constrain the growth model for Illinois, but the CV for deviations in mean length-at-age 2 was the same as the Wisconsin model.

Sensitivity Analyses

We performed sensitivity analyses to determine the effects of some of our assumptions on the results of the analysis. To test the sensitivity of the model estimates to the weighting factors for each data source, we increased and decreased the weighting factors for each data source five-fold and refit the models. We also tested the sensitivity of our estimates to our assumed value of M by increasing and decreasing M by 20% and refitting our models. We then evaluated sensitivity of the model estimates to the change by comparing model estimates of abundance, biomass, and mean fishing mortality rates for females and males age 4 and older in 2002 to those obtained with the baseline weighting factors and natural mortality rate. Also, because of large suspected amounts of unreported commercial harvest in Wisconsin during 1989-1992, we tested the effects of three levels (one to three times the reported amount) of commercial harvest during those years on our results.

Data

Commercial yield and effort were estimated from mandatory bimonthly reports submitted by commercial fishermen. In some cases, these reports were validated by law enforcement officials, but underreporting may have been a large problem, especially in Wisconsin. The exact magnitude of underreporting is unknown, but during 1990-1992 commercial yield in Wisconsin was underreported by at least 44%, which law enforcement officials documented during a multi-year sting operation (WDNR,

unpublished data). Two commercial fishermen indicted for unreported harvest testified that unreported harvest was two to three times reported harvest. Wisconsin implemented a commercial quota for yellow perch in the summer of 1989, so there was less incentive for commercial fishermen to underreport prior to 1989. For observed commercial yield in Wisconsin during 1989-1992, we added the reported commercial yield and the verified illegal yield and multiplied the number by two. In Illinois, unreported commercial harvest was thought to be relatively low (Illinois Department of Natural Resources [IDNR], unpublished data). Length frequency estimates of the commercial catch were collected by dockside monitoring. Sampling did not occur for most lifts.

Creel surveys were conducted by the Wisconsin DNR and the Illinois DNR to estimate recreational fishery harvest, effort, and composition of the harvest (details in Austen et al. 1995). Creel clerks visited access points and interviewed anglers to determine target species and angler effort. Anglers' catches were examined for species composition and length frequency.

Graded-mesh gillnet surveys were conducted in Wisconsin (2.54-7.62 cm stretch-measure with 0.64 cm increments) in the winter and in Illinois (2.54-8.89 cm stretch-measure with 1.27 cm increments) in June of each year to obtain fishery independent relative abundance data. Nets were set overnight in the same locations each year at multiple depths. CPE was measured as the number of yellow perch per 30.5 m gillnet.

The length of each fish was measured, and the age composition of the catch was estimated by estimating ages for a randomly chosen subsample and applying the subsequent age-length key to the length frequency. Ages were estimated by counting the annuli on scales during 1986-1999 in Wisconsin and 1986-1993 in Illinois. However,

this method was found to be fairly unreliable (Robillard and Marsden 1996; Baker and McComish 1998; WDNR, unpublished data). Therefore, Illinois estimated ages of fish by counting annuli in otoliths during 1994-2002, and Wisconsin estimated ages of fish by counting the annuli in anal fin spines during 2000-2002. Ages estimated by different readers of spines and otoliths agreed 86% of the time (WDNR, unpublished data).

Results

Model Fits

Most of our data sources contained relatively large amounts of contrast and our models produced reasonable fits to all data sources. Fishery and survey catch was relatively high in the beginning of our time series and decreased to low levels during the mid 1990s. Our models predicted observed commercial yield and recreational harvest within 5% of observed values in most years (Figure 1.2). For total survey CPE, our models produced the same declining trend as was observed, but produced lower predictions of survey CPE than was observed in most years prior to 1991 (Figure 1.2). This may be due to decreases in survey catchability caused by increases in water clarity since the colonization of Lake Michigan by zebra mussels *Dreissena polymorpha*. Relative differences between observed and predicted survey CPE tended to be larger than fishery catch residuals (especially for the Wisconsin survey); this result is not surprising given that survey CPE had relatively high CVs and that CVs were higher for the Wisconsin survey than for the Illinois survey. Mean age in the survey was relatively stable during 1986-1992, increased during 1992-1997, and decreased thereafter (Figure 1.3). Deviations between model predictions and observations of mean age in the survey were usually less than 15%. Mean length in the recreational fishery and surveys

increased during 1986-2002, but did not show a trend for commercial fisheries (Figure 1.4). Predicted mean length was usually within 10% of the observed value for the commercial fishery and surveys and within 5% of observed values for the recreational fishery (Figure 1.4). Predicted mean length of females in the Illinois survey during 1986-1992 was lower than observed values and may be low because the survey mainly targets mature fish; after 1990, a smaller proportion of females were immature.

Model Estimates

Model estimates of mortality rates were generally higher for females than males, and were higher during the mid-1980s through the mid-1990s than in the late 1990s and after (Figure 1.5). In Wisconsin, the commercial fishery was the predominant source of fishing mortality until the commercial fishery was closed, and in Illinois, the recreational fishery was the predominant source of fishing mortality. Estimated instantaneous fishing mortality rates for females age 4 and older exceeded 1.0 in most modeled years prior to 1996 in Wisconsin waters and averaged 1.16, which corresponds to an annual mortality rate of about 69%. In Illinois, estimated fishing mortality rates were not as high as in Wisconsin, although total mortality rates averaged about 0.92 (annual mortality rate of about 60%) for females age 4 and older during 1986-1997. In Wisconsin during 1986-1996, instantaneous total mortality rates for males age 4 and older averaged 0.67 (annual mortality rate of about 49%), and in Illinois during 1986-1997, instantaneous total mortality rates averaged 0.57 (annual mortality rate of about 44%). Until severe restrictions were placed on commercial and recreational fisheries (1996-1997), fishing was the predominant source of mortality for female yellow perch age 4 and older in Wisconsin and Illinois. After the fisheries were considerably restricted in 1996 in

Wisconsin and 1997 in Illinois, fishing mortality rates declined substantially and natural mortality was the predominant source of mortality.

Model estimates of recruitment in Illinois and Wisconsin showed similar patterns, with recruitment generally higher in Illinois than in Wisconsin (Table 1.1; Figure 1.6). Recruitment was relatively high during 1984-1989 and was substantially lower than 1980s levels thereafter, except for the 1998 year-class. The largest year-class during the 1980s was in 1988 and the largest year-class during the 1990s was in 1998. Model estimates of average recruitment of the 1984-1989 year-classes were 13 times higher in Illinois and 23 times higher in Wisconsin than the estimated average recruitment of the 1990-1997 year-classes. Recruitment was not strongly related to stock size and yellow perch produced weak year-classes across a wide range of stock size (Figure 1.6).

Estimated abundance of yellow perch in Wisconsin waters of southwestern Lake Michigan increased from 1986 to 1990, and then decreased from 1991 to 2002 except for a small increase in 2000 (Figure 1.7). Estimated abundance of yellow perch in Illinois waters declined from 1986 to 2002, except during 1990 and 2000. In 2002, yellow perch abundance was approximately 8% of 1986 abundance in Wisconsin and approximately 20% of 1986 abundance in Illinois. Model estimates of relatively high abundance throughout the 1980s resulted from high estimated recruitment during that period. Abundance decreased drastically during the 1990s because recruitment declined and fishing mortality rates were relatively high.

Changes in estimated biomass were smaller than changes in abundance; estimated biomass in 2002 was approximately 74% of 1986 biomass in Wisconsin and 123% of 1986 biomass in Illinois (Figure 1.7). Estimated biomass showed somewhat different

trends over time than abundance because the age structure of the population changed and growth rates increased. In 1986, the population was composed of mostly age-2 and 3 yellow perch. In 2002, the majority of the population was age-4 and substantially larger at a given age due to faster growth.

Patterns of estimated SSB were similar to patterns of biomass (Figure 1.7). Model estimates of SSB increased during 1986-1992 in Illinois and during 1986-1991 in Wisconsin, and decreased until the late 1990s. Estimated SSB increased greatly during 1997-2002 in Illinois and during 1999-2002 in Wisconsin. In 2002, SSB was at its highest level since the early 1990s and was 346% and 854% of 1986 levels in Illinois and Wisconsin, respectively. The large increase in SSB during 1999-2002 was due to the relatively good recruitment of the 1998 year-class, low fishing mortality rates, and rapid growth and maturity of females. We estimated that spawning stock biomass per recruit (SSB/R) was approximately 0.46 kg in Wisconsin and 0.44 kg in Illinois in 2002. We compared these SSB/R values to scenarios without fishing mortality, and estimated that 2002 SSB/R was approximately 84% of the unexploited scenario in Wisconsin and 87% of the unexploited scenario in Illinois. In contrast, SSB/R during 1986-1995 was approximately 0.03 kg (18% of the unexploited scenario) in Wisconsin and 0.06 kg (33% of the unexploited scenario) in Illinois. These dramatic differences in SSB/R occurred because fishing mortality rates were much lower during 2002 than during 1986-1995 and yellow perch were growing faster, and therefore maturing at younger ages, during 2002 than during 1986-1995.

Females grew faster and to larger sizes than males (Figure 1.8); the mean length-at-age of females at all ages older than age-2 were higher than males of the same age.

Estimated mean length-at-age remained relatively stable during 1986-1994 and increased substantially during 1994-2000. During 2000-2002, mean length-at-age decreased slightly, but was still higher than during the 1980s and early 1990s. In Wisconsin, yellow perch were generally smaller at a given age than in Illinois.

Selectivity patterns of the recreational fisheries in Wisconsin and Illinois were quite similar to one another when no length-based regulations were in effect (Figure 1.9). Commercial selectivity patterns were also similar. This latter result was not surprising because the scarcity of biological data for the Illinois commercial catch had led us to assume an informative prior for the selectivity parameters, based on the results of the Wisconsin assessment (see Appendix A). Due to differences in selectivity of the commercial and recreational fisheries, yellow perch recruited to the recreational fishery at smaller sizes than to the commercial fishery. Selectivity of the Illinois recreational fishery changed substantially when a slot size limit was implemented during 1997-2000. In Illinois during 1997-2000, average mortality rates for males age-4 and older were slightly higher than for females due to the selectivity pattern of the recreational fishery. Selectivity patterns in the survey were substantially different between Illinois and Wisconsin. Differences in selectivity patterns are likely attributable to differences in the surveys such as mesh sizes of assessment gillnets and time of year of the survey.

Sensitivity Analyses

The models were somewhat sensitive to changes in the assumed CVs and effective sample sizes for the different data sources (Table A.4). The Illinois model was slightly less sensitive to these assumptions than the Wisconsin model. In general, five-fold changes in the weights for each data source usually resulted in less than 15%

changes in mean fishing mortality rates, abundance, and biomass. Weights that resulted in increased estimates of mean fishing mortality rates usually resulted in decreased estimates of abundance and biomass. The Illinois model was most sensitive to changes in the CV and effective sample sizes associated with females caught in the survey and the effective sample size of the length composition from the recreational fishery. The Wisconsin model was most sensitive to CV and effective sample size associated with males caught in the survey and the CV for catchability of the commercial fishery. Increasing M by 20% resulted in higher model estimates of average fishing mortality rates and lower estimates of abundance and biomass. The Illinois model was less sensitive to our assumed value of M than the Wisconsin model; Illinois model estimates changed approximately 12% and Wisconsin model estimates changed approximately 47%.

The Wisconsin model estimates of abundance, biomass, and mean fishing mortality rates were also somewhat sensitive to the different levels of commercial harvest (Table 1.2). When we fit the model using only reported yield, model estimates of abundance and biomass in 2002 were more than 20% lower than the baseline (2x reported during 1989-1992) scenario, and estimates of mean fishing mortality rates were about 27% higher than baseline estimates. Under the 3x reported yield scenario, abundance and biomass were about 20% greater than the baseline scenario, but mean fishing mortality rates were about 17% lower than the baseline.

Discussion

The decline in abundance of yellow perch in southwestern Lake Michigan during the 1990s was likely caused by a combination of recruitment failure and relatively high

fishing mortality rates, and our results are consistent with other authors' descriptions of the decline. During 1989-1994, yellow perch larvae were abundant shortly after hatching, but recruitment to age 0 in the fall was poor, which has led some researchers to propose that at least the initial decline in recruitment was not due to fishing (Francis et al. 1996; Robillard et al. 1999; Marsden and Robillard 2004). Our results also indicated that several successive year-classes failed despite relatively high SSB. However, after 1994, the relative abundance of yellow perch larvae was less than 10% of the relative abundance during the early 1990s, which may indicate that SSB had decreased to low enough levels to limit recruitment (Francis et al. 1996; Marsden and Robillard 2004). We estimated that between 1991 and 1996 yellow perch SSB in Wisconsin declined almost 94% and between 1992 and 1997 yellow perch SSB in Illinois declined almost 90%. The resultant low SSB may have prolonged the period of poor reproduction.

The decline of yellow perch SSB in southern Lake Michigan would probably not have occurred at such a rapid pace if fishing mortality rates had been lower. We projected dynamics for the 1986 through 1996 period using our estimated recruitment time series and age-based selectivity estimates, while changing the overall level of *F*. Our projections indicated that SSB in 1996 would have been more than five times higher than our model estimates in Wisconsin and nearly twice as high in Illinois if fishing mortality rates for fully selected ages and sexes had been equal to the natural mortality rate (0.37) during 1986-1997. While our simple projections do not account for compensatory changes that might have occurred if fishing mortality had been lower, we believe they do illustrate that high fishing mortality rates on adult females were a substantial contributor to the rapid decline in SSB that occurred. An alternative

hypothesis to the effect of fishing yellow perch population dynamics is that natural mortality decreased concurrently with restrictions on the fisheries. In a supplemental analysis (detailed results not reported), we explored this possibility by adding one more estimated parameter to each model that allowed natural mortality to change from one level for the 1986-1996 period to another for 1997 and after. The estimated changes in M were opposite in sign for the Wisconsin and Illinois models and were much less than the estimated changes in fishing mortality for these periods.

The declines of yellow perch abundance in southern Lake Michigan were similar in the 1960s and 1990s, and recruitment failures of several successive year-classes may be likely in the future. In the early 1960s, yellow perch suffered a recruitment failure (Wells 1977) similar to the recruitment failure observed in the early 1990s (Robillard et al. 1999; Marsden and Robillard 2004). The recruitment failure in the 1960s was preceded by an increase in abundance during the late 1950s (Wells 1977), which was similar to the increase in abundance during the late 1980s (Francis et al. 1996). Adult abundance had decreased rapidly by the mid-1960s due to intense fisheries (Wells 1977). Yellow perch growth was slow during the 1950s (Wells 1977) and the 1980s (Marsden and Robillard 2004). Extremely high fishery catches preceded both declines in abundance. However two major differences in the Lake Michigan community exist regarding exotic species: alewife abundance in Lake Michigan was extremely high during the 1960s compared to relatively low alewife abundance in the 1980s and 1990s, and zebra mussels were absent from Lake Michigan in the 1960s, but their abundance was high in the 1990s (Madenjian et al. 2002). Because the reproduction failure in the 1960s was associated with extremely high levels of alewife abundance, the decline in

recruitment was blamed on alewife (Eck and Wells 1987). Schroyer and McComish (2000) found a negative correlation between alewife abundance and yellow perch recruitment in Indiana waters of Lake Michigan during 1988-1997, but little direct evidence of alewife preying upon yellow perch larvae has been observed in southern Lake Michigan (Dettmers et al. 2003). Also, alewife abundance during the 1990s was substantially lower (perhaps more than 20 times lower) than during the mid-1960s (Madenjian et al. 2002), the period when alewife interference with yellow perch recruitment was originally proposed as a cause for yellow perch reproduction failure. Marsden and Robillard (2004) suggested that declines in yellow perch recruitment may be exacerbated by changes in the ecosystem due to zebra mussel colonization, and Janssen and Leubke (2004) found that poor recruitment was correlated with the presence of zebra mussels in Indiana waters of Lake Michigan. Indeed, zebra mussels can alter the composition of the zooplankton community (MacIsaac et al. 1992), which may decrease food supplies for larval yellow perch. However, yellow perch recruitment did not collapse after invasion of zebra mussels in Oneida Lake (Mayer et al. 2000) or the western basin of Lake Erie (Tyson and Knight 2001).

Based on several reference points, yellow perch likely experienced overfishing in southwestern Lake Michigan during 1986-1996. Beverton (1998) recommended the use of the F_{95} reference point (F at which yield is 95% of maximum sustainable yield) to sustainably manage fisheries. A rough estimate of F_{95} is usually around M for medium-lived species (Beverton 1998), which would be approximately 0.37 for yellow perch in southern Lake Michigan. Others have argued that M should be an upper bound on fishing mortality rates that maximize yield (Deriso 1982, Quinn and Deriso 1999). Fishing

mortality rates for adult females were well above M in Illinois (1-2 times M) and Wisconsin (2-4 times M). A number of U.S. marine commercial fisheries are managed to keep fishing mortality below levels that would reduce SSB/R below a set percentage of the unfished situation ($F_{X\%}$), and typical percentages have been in the 35% to 45% range (Quinn and Deriso 1999). In Wisconsin and Illinois, F was higher than $F_{35\%}$ during 1986-1996.

Regulation changes likely helped to substantially reduce fishing mortality rates. In 1996 in Wisconsin, the commercial quota was set to zero and a daily bag limit of five yellow perch per angler was implemented for the recreational fishery (reduced from 50 to 25 in 1995). When these policies were introduced, fishing mortality decreased noticeably. Recreational effort decreased, but may not have been a direct consequence of the implemented bag limit. When stricter bag limits were implemented in some inland Wisconsin lakes for walleye, anglers preferred to fish in lakes that had less restrictive bag limits (Beard et al. 2003). In Illinois in 1995, the recreational daily bag limit was reduced from no limit to 25 yellow perch per angler. In 1997, the commercial quota was reduced to zero and a daily bag limit of 15 yellow perch per angler and a slot size limit of 8-10 in (fish within this range could be kept) were implemented for the recreational fishery. Mortality rates also declined substantially in Illinois, as they did in Wisconsin; commercial effort was reduced to zero, and recreational fishing effort decreased noticeably. Also, the slot size limit caused the recreational fishery selectivity to change so that average fishing mortality rates were higher for age-4 and older males than for age-4 and older females.

We did not incorporate age-estimation error into our model and this may bias our estimates of recruitment and mortality rates. Our results likely underestimate the amount of variability in recruitment because age-estimation error tends to blend strong and weak year classes together (Richards and Schnute 1998). Specifically, our estimates of recruitment of the 1989 and 1990 year-classes are probably high because of age-estimation error associated with the 1988 year-class. However, our estimates of recruitment are consistent with external estimates of year-class strength from age-0 assessments (Pientka et al. 2003). Our mortality rate estimates are likely biased low for the beginning of the time series when ages of yellow perch were estimated from scales. Younger yellow perch tended to be aged as older when ages were estimated from scales (Robillard and Mardsen 1996; Baker and McComish 1998; Wisconsin DNR, unpublished data) and the overrepresentation of older fish in the data is most likely interpreted by the model as an indication that older fish were more abundant. Annual mortality rates in the late 1970s in Indiana and Illinois were estimated to be about 70% for males age-3 and older and substantially higher for females age-3 and older (Wells and Jorgenson 1983). These mortality rate estimates are similar to our estimates for Wisconsin in the late 1980s and for Illinois in the mid-1980s.

Yellow perch growth may be density dependent and may also have increased due to zebra mussel colonization. Patterns of growth during 1986-1998 resembled growth during 1954-1979 for yellow perch in southern Lake Michigan. Yellow perch growth may have been density dependent during 1986-2002 and 1954-1975 (Wells 1977). We found similar growth patterns in Wisconsin and Illinois; growth was relatively slow when yellow perch were at high abundance and growth was fastest at low abundance.

However, growth during 1999-2002 (low abundance) was the fastest observed for yellow perch in southern Lake Michigan during the last five decades. This increased growth coincided with substantial changes in yellow perch habitat due to colonization by zebra mussels. Thayer et al. (1997) found increased adult yellow perch growth associated with zebra mussels in ponds enclosures and Tyson and Knight (2001) found increased growth of age-2 and age-3 yellow perch in the western basin of Lake Erie after zebra mussel colonization; these increases in growth were attributed to increased food availability. However, Mayer et al. (2000) found no increase in adult yellow perch growth associated with zebra mussel colonization in Oneida Lake.

Management Implications

Since 1998, recruitment has continued to be poor in southern Lake Michigan except for the 2002 year-class (Pientka et al. 2003; Clapp and Dettmers 2004; Fitzgerald et al. 2004). Success of the 1998 year-class has renewed pressure on the agencies to implement less restrictive regulations. Based partially on development of the models described here, the Lake Michigan Yellow Perch Task Group recommended that regulations remain unchanged for the time being. The models we developed will continue to be used to monitor changes in the population and to advise managers.

Overexploitation of yellow perch has not previously been considered a likely hypothesis for the decline of yellow perch in southern Lake Michigan (Francis et al. 1996). However, we found that SSB had reached very low levels by the mid-1990s and intense fishing likely compounded the rapidity of the decline in SSB. Although exotic species or climatic changes may have affected recruitment, fishing mortality rates during the late 1980s and early 1990s probably were above levels that would be sustainable over

the long term. Therefore, management of yellow perch in Lake Michigan should focus on limiting fishing mortality and be flexible to adjust to future recruitment failures. Despite poor recruitment, SSB has increased to its highest point since the early 1990s in Wisconsin and Illinois. This is partly a response to extensive management actions taken by Wisconsin and Illinois, which have reduced fishing mortality rates. However, relatively few year-classes are represented in the population and future increases in biomass and SSB will depend upon relatively strong recruitment of future cohorts to the adult population.

References

Allen, P. J. 2000. A computer simulation model for the yellow perch population in the Indiana waters of Lake Michigan. Master's Thesis. Ball State University.

Austen, D., W. Brofka, J. E. Marsden, J. Francis, J. Palla, J. R. Bence, R. Lockwood, and B. Eggold. 1995. Lake Michigan creel survey methods. Report to the Lake Michigan Technical Committee.

Baker, E. A., and T. S. McComish. 1998. Precision of ages determined from scales and opercles for yellow perch *Perca flavescens*. Journal of Great Lake Research 24:658-665.

Baldwin, N. S., R. W. Saalfeld, M. A. Roos, and H. J. Buettner. 1979. Commercial fish production in the Great Lakes. Great Lakes Fishery Commission Technical Report 3. Ann Arbor.

Beard, T. D., S. P. Cox, and S. R. Carpenter. 2003. Impacts of daily bag limit reductions on angler effort in Wisconsin walleye lakes. North American Journal of Fisheries Management 23:1283-1293.

Beletskym D., D. Schwab, D. Mason, E. Rutherford, M. McCormick, H. Vanderploeg, and J. Janssen. 2004. Modeling the transport of larval yellow perch in Lake Michigan. Estuarine and Coastal Modeling, Proceedings of the 8th International Conference of the American Society of Civil Engineers, held November 3-5, 2003, p.439-454.

Bence, J. R., and K. D. Smith. 1999. An overview of recreational fisheries of the Great Lakes. *In* W. W. Taylor and P. A. Ferreri ed. Great Lakes Fishery Policy and Management: a Binational Perspective. Michigan State University Press, East Lansing, Michigan.

Beverton, R. 1998. Fish, fact and fantasy: a long view. Reviews in Fish Biology and Fisheries 8:229-249.

Bowker, D. W. 1995. Modelling patterns of dispersion of length at age in teleost fishes. Journal of Fish Biology 46:469-484.

Clapp, D. F., and J. M. Dettmers. 2004. Yellow perch research and management in Lake Michigan: evaluating progress in a cooperative effort, 1997-2001.

Deriso, R. B. 1982. Relationship of fishing mortality and growth and the level of maximum sustainable yield. Canadian Journal of Fisheries and Aquatic Sciences 39:1054-1058.

Deriso, R. B., S. J. Nepszy, and M. R. Rawson. 1988. Age structured stock assessment of Lake Erie walleye. Great Lakes Fishery Commission Special Publication 88-3. Ann Arbor.

Dettmers, J. M., M. J. Raffenberg, and A. K. Weis. 2003. Exploring zooplankton changes in southern Lake Michigan: implications for yellow perch recruitment. Journal of Great Lakes Research 29:355-364.

Ebener, M. P., J. R. Bence, K. Newman, and P. Schneeberger. In press. An overview of the application of statistical catch-at-age models t assess lake whitefish stocks in the 1836 treaty-ceded waters of the upper Great Lakes. Great Lakes Fishery Commission Technical Report.

Eck, G. W., and L. Wells. 1987. Recent changes in Lake Michigan's fish community and their probable causes, with emphasis on the role of alewife (*Alsoa psuedoharengus*). Canadian Journal of Fisheries and Aquatic Sciences 44(supplement 2): 53-60.

Evans, M. S. 1986. Recent major declines in zooplankton populations in the inshore region of Lake Michigan: probable causes and implications. Canadian Journal of Fisheries and Aquatic Sciences 43:154-159.

Fitzgeralt, D. G., D. F. Clapp, and B. J. Belonger. 2004. Characterization of growth and winter survival of age-0 yellow perch in southeastern Lake Michigan. Journal of Great Lakes Research 30:227-240.

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195-1207.

Francis, J. T., S. R. Robillard, and J. E. Marsden. 1996. Yellow perch management in Lake Michigan: a multi-jurisdictional challenge. Fisheries 21:18-20.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis, 2nd edition. Chapman and Hall, Boca Raton.

Hampton, J., and D. A. Fournier. 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Marine and Freshwater Research 52:937-963.

Heyer, C. J., T. J. Miller, F. P. Binkowski, E. M. Caldrone, and J. A. Rice. 2001. Maternal effects as a recruitment mechanism in Lake Michigan yellow perch (*Perca flavescens*). Canadian Journal of Fisheries and Aquatic Sciences 58:1477-1487.

Horns, W. H. 2001. Spatial and temporal variation in length at age and condition of yellow perch in southern Lake Michigan during 1986-1988. North American Journal of Fisheries Management 21:580-591.

Janssen, J., and M. A. Leubke. 2004. Preference for rocky habitat by age-0 yellow perch and alewives. Journal of Great Lakes Research 30:93-99.

Jude, D. J., and F. J. Tesar 1985. Recent changes in the inshore forage fish of Lake Michigan. Canadian Journal of Fisheries and Aquatic Sciences 42:1154-1157.

Kraft, C. E., and B. L. Johnson. 1992. Fyke-net and gill-net size selectivities for yellow perch in Green Bay, Lake Michigan. North American Journal of Fisheries Management 12:230-236.

Lake Erie Yellow Perch Task Group. 2001. Report of the Lake Erie Yellow Perch Task Group to the Lake Erie Committee.

MacIsaac, H. J. W. G. Sprules, O. E. Johannsson, and J. H. Leach. 1992. Filtering impacts of larval and sessile zebra mussels (*Dreissena polymorpha*) in western Lake Erie. Oecologia 92:30-39.

Madenjian, C. P., G. L. Fahnenstiel, T. H. Johengen, T. F. Nalepa, H. A. Vanderploeg, G. W. Fleischer, P. J. Schneeberger, D. M. Benjamin, E. B. Smith, J. R. Bence, E. S. Rutherford, D. S. Lavis, D. M. Robertson, D. J. Jude, and M. P. Ebener. 2003. Dynamics of the Lake Michigan food web, 1970-2000. Canadian Journal of Fisheries and Aquatic Sciences 59:736-753.

Makauskas, D., and D. Clapp. 2000. Status of yellow parch in Lake Michigan and yellow perch task group progress report. In meeting minutes of the 2000 Annual Meeting of the Lake Michigan Committee. Great Lakes Fishery Commission, Ann Arbor, Michigan.

Marsden, J. E., and S. R. Robillard. 2004. Decline of yellow perch in southwestern Lake Michigan, 1987-1997. North American Journal of Fisheries Management 24:952-966.

Mayer, C. M., A. VanDeValk, J. L. Forney, L. G. Rudstam, and E. L. Mills. 2000. Response of yellow perch (Perca flavescens) in Oneida Lake, New York, to the establishment of zebra mussels (Dreissena polymorpha). Canadian Journal of Fisheries and Aquatic Sciences 57:742-754.

McAllister, M. K., and J. N. Ianelli. 1997. Bayesian stock assessment using catch-at-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences 54:284-300.

Megrey, B. A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. In Mathematical analysis of fish stock dynamics. *Edited* by E. F. Edwards and B. A. Megrey. American Fisheries Society Symposium 6:8-48.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Pages 259-277 in L. Lowe ed. Proceedings of the symposium on applications of stock assessment techniques to gadids. International North Pacific Fisheries Commission Bulletin 50.

Miller, L. M. 2003. Microsatellite DNA loci reveal genetic structure of yellow perch in Lake Michigan. Transactions of the American Fisheries Society 132:503-513.

Otter Research Ltd. 2000. An introduction to AD Model Builder version 6.0.2 for use in nonlinear modelling and statistics. Otter Research Ltd: Nanaimo, Canada.

Pientka, B., S. J. Czesny, and J. M. Dettmers. 2003. Yellow perch population assessment in southwestern Lake Michigan, including the identification of factors that determine yellow perch year-class strength. Annual Performance Report to the Illinois Department of Natural Resources. Aquatic Ecology Technical Report 03/06.

Quinn, T. J., II, and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press. New York.

Richards, L. J., and J. T. Schnute. 1998. Model complexity and catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences 55:949-957.

Robillard, S. R., and J. E. Mardsen. 1996. Comparison of otolith and scale ages for yellow perch from Lake Michigan. Journal of Great Lakes Research 22:429-435.

Robillard, S. R., A. Weis, and J. M. Dettmers. 1999. Yellow perch population assessment in southwestern Lake Michigan, including evaluation of samplingtechniques and the identification of factors that determine yellow perch year-class strength. Annual Report to the Illinois Department of Natural Resources. Aquatic Ecology Technical Report 99/5.

Schroyer, S. R., and T. S. McComish. 2000. Relationship between alewife abundance and yellow perch recruitment in southern Lake Michigan. North American Journal of Fisheries Management 20:220-225.

Sitar, S., J. R. Bence, J. E. Johnson, M. P. Ebener, and W. W. Taylor. 1999. Lake trout mortality and abundance in southern Lake Huron. North American Journal of Fisheries Management 19:881-900.

Szalai, E. B., G. W. Fleischer, and J. R. Bence. 2003. Modelling time-varying growth using a generalized von Bertalanffy model with application to bloater (*Coregonus hoyi*) growth dynamics in Lake Michigan. Canadian Journal of Fisheries and Aquatic Sciences 60: 55-66.

Thayer, S. A., R. C. Haas, R. D. Hunter, and R. H. Kushler. 1997. Zebra mussel (*Dreissena polymorpha*) effects on the sediment, other zoobenthos, and the diet and growth of adult yellow perch (*Perca flavescens*) in pond enclosures. Canadian Journal of Fisheries and Aquatic Sciences 54:1903-1915.

Tyson, J. T., and R. L. Knight. 2001. Response of yellow perch to changes in the benthic invertebrate community of western Lake Erie. Transactions of the American Fisheries Society 130:766-782.

Wells, L. 1977. Changes in yellow perch (*Perca flavescens*) population of Lake Michigan, 1954-1975. Journal of the Fisheries Research Board of Canada 34:1821-1829.

Wells, L., and S. C. Jorgenson. 1983. Population biology of yellow perch in southern Lake Michigan, 1971-79. U.S. Fish and Wildlife Service Technical Paper No. 109. Washington D. C.

Wells, L., and A. L. McLain. 1972. Lake Michigan: effects of exploitation, introductions, and eutrophication on the salmonid community. Journal of the Fisheries Research Board of Canada 29:889-898.

Table 1.1. Model estimates of yellow perch abundance-at-age (in thousands) during 1986-2002 in Illinois and Wisconsin waters of southwestern Lake Michigan.

	Age							
Year	2	3	4	5	6	7	8	9+
Illinois								
1986	9,674	11,417	2,082	769	284	105	39	14
1987	9,598	6,682	7,518	862	146	33	10	4
1988	6,807	6,629	4,518	3,715	247	32	6	2
1989	7,255	4,701	4,457	2,567	1,245	67	8	2
1990	17,535	5,011	3,180	2,759	1,370	545	30	4
1991	5,432	12,110	3,322	1,837	1,383	633	231	15
1992	2,521	3,726	8,070	1,949	969	694	311	116
1993	444	1,718	2,401	4,930	1,090	526	371	224
1994	22	302	1,070	1,370	2,691	573	270	295
1995	190	15	188	598	727	1,434	298	284
1996	325	127	9	102	310	378	747	292
1997	1,153	216	70	4	48	144	175	476
1998	130	787	143	45	3	31	93	420
1999	879	89	529	96	30	2	21	342
2000	8,911	599	59	349	63	20	1	240
2001	38	6,144	404	40	235	43	14	163
2002	38	26	4,139	265	26	153	28	115

Table 1.1. Continued.

Wisconsin								
1986	5,863	11,702	1,490	300	101	40	16	6
1987	5,083	4,050	7,922	741	106	21	5	2
1988	8,757	3,510	2,776	4,246	195	20	2	0
1989	8,438	6,045	2,353	1,676	1,718	71	5	0
1990	11,935	5,826	3,994	1,250	694	578	19	1
1991	5,221	8,230	3,786	2,020	530	287	204	6
1992	1,237	3,603	4,964	1,665	682	161	78	47
1993	310	853	2,315	2,227	627	225	48	34
1994	102	214	572	1,327	1,043	285	96	33
1995	83	71	144	336	579	408	108	44
1996	60	57	49	85	158	231	153	53
1997	289	42	39	33	54	96	139	122
1998	128	200	29	27	22	35	63	170
1999	373	88	138	20	18	14	22	147
2000	3,115	258	61	93	13	11	9	107
2001	29	2,147	175	41	62	9	8	77
2002	29	20	1,405	111	26	39	5	53

Table 1.2. Model estimates of abundance (N; 1000s), biomass (B; 1000 kg), mean rate of fishing mortality for females age-4 and older ($\overline{F}_{4+}$ females), and mean rate of fishing mortality for males age-4 and older ($\overline{F}_{4+}$ males) for 2002 under three scenarios of unreported commercial harvest in Wisconsin waters of southwestern Lake Michigan during 1989-1992.

	N	B	$\overline{F}_{4+}$ females	$\overline{F}_{4+}$ males
Reported	1,331	280	0.095	0.077
2 X Reported	1,690	356	0.075	0.060
3 X Reported	2,020	427	0.062	0.050

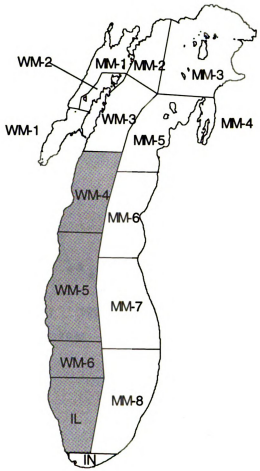

Figure 1.1. Map of Lake Michigan statistical districts with modeled areas shaded. WM indicates Wisconsin waters, IL indicates Illinois waters, IN indicates Indiana waters, and MM indicates Michigan waters.

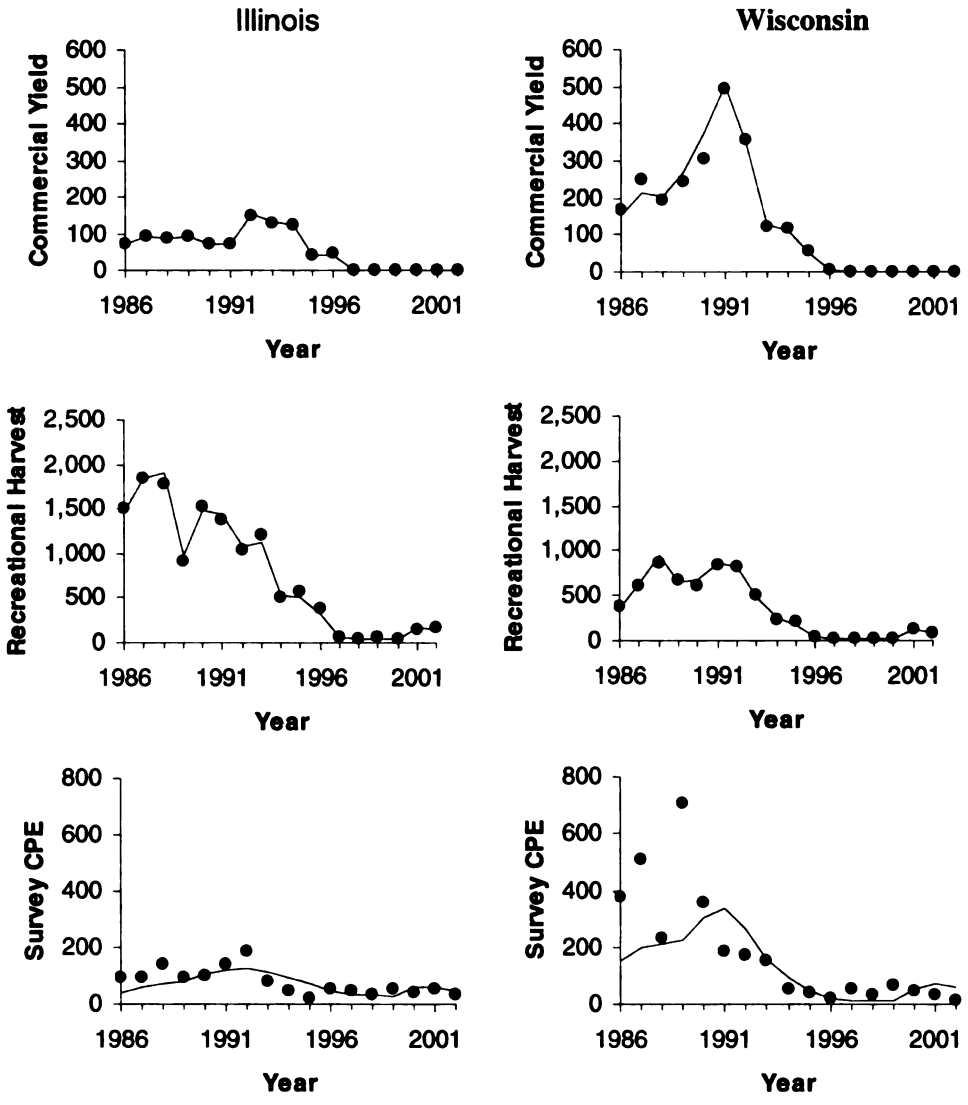

Figure 1.2. Model fits to commercial yield (1000 kg), recreational harvest (1000s), and gill net survey catch-per-effort (CPE; number per 30.5 m) in Illinois and Wisconsin waters of southwestern Lake Michigan during 1986-2002. Model predictions are represented by solid lines and observed values are represented by dots.

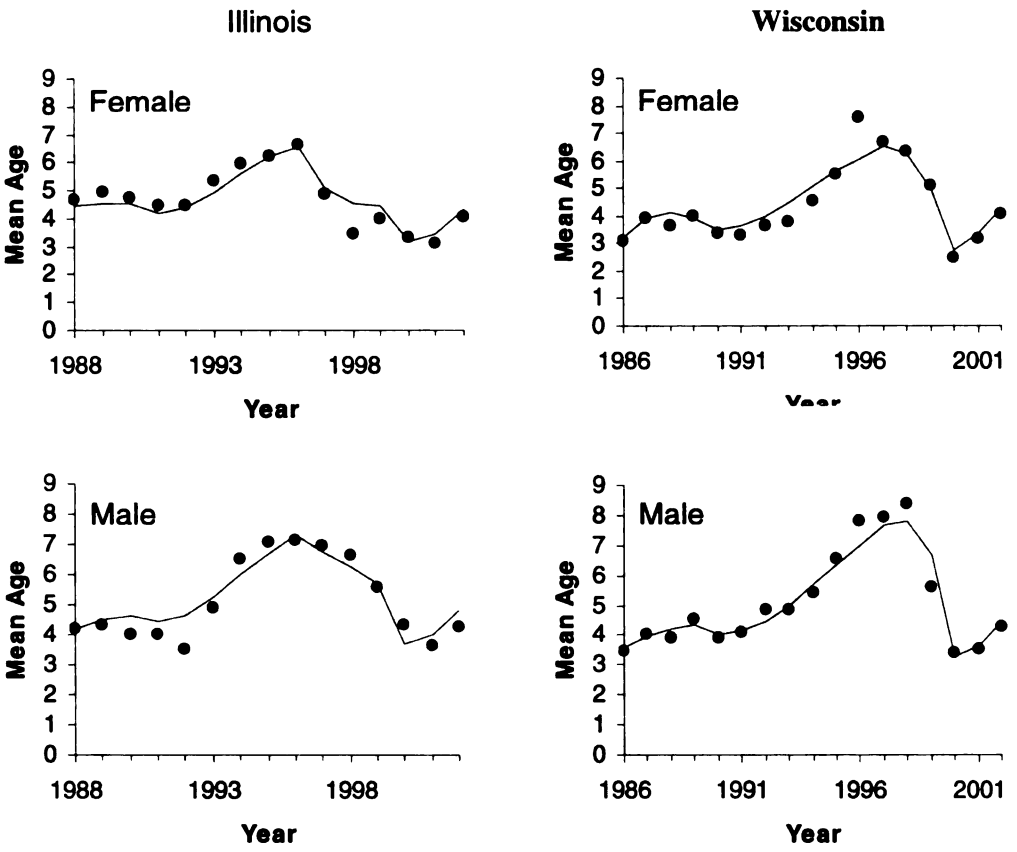

Figure 1.3. Mean age of yellow perch caught in gill net surveys in Illinois and Wisconsin waters of southwestern Lake Michigan during 1986-2002. Lines represent model predictions and dots represent observed values.

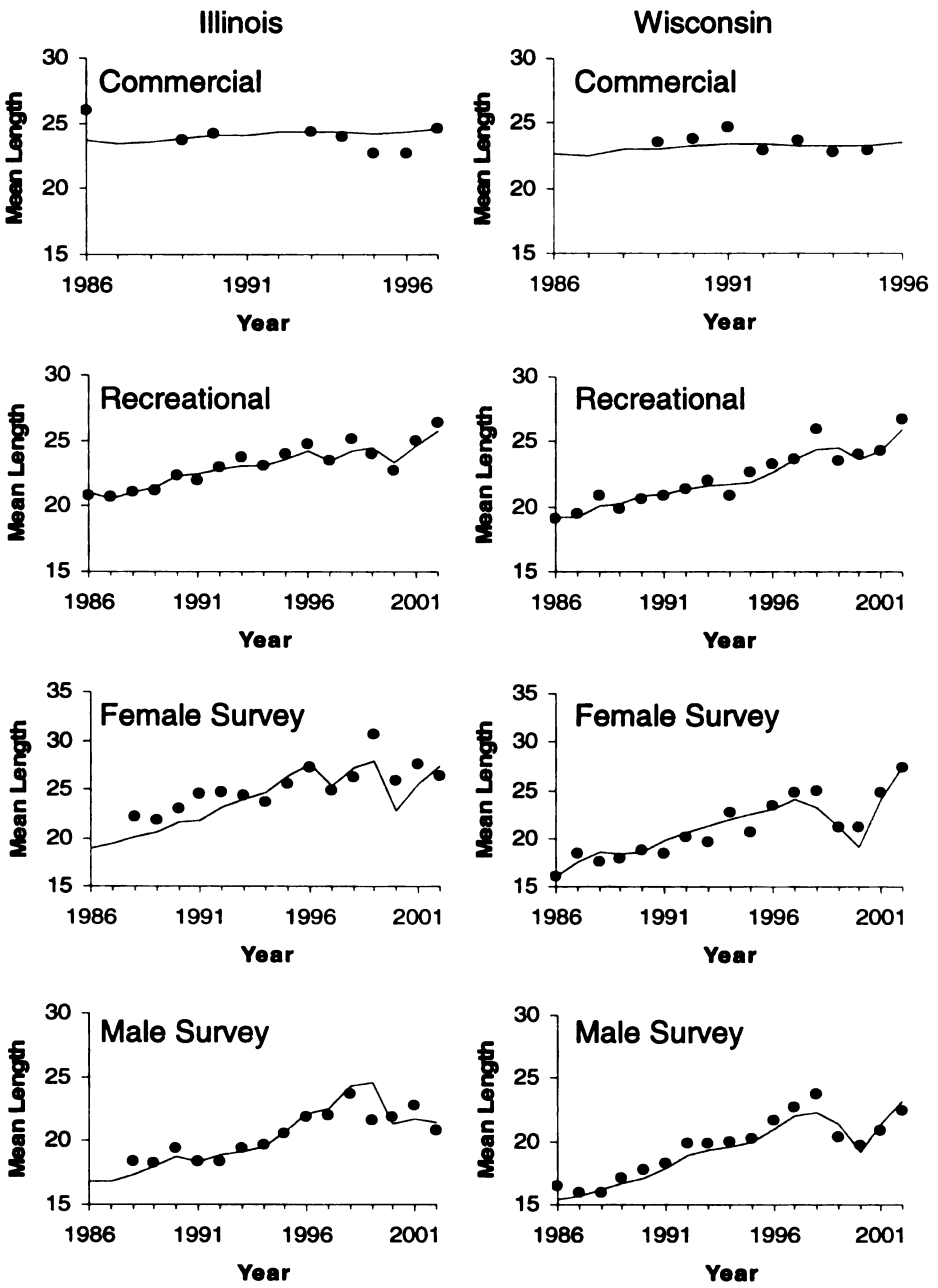

Figure 1.4. Mean length of yellow perch caught in the commercial and recreational fisheries and gill net surveys in Illinois and Wisconsin waters of southwestern Lake Michigan during 1986-2002. Lines represent model predictions and dots represent observed values.

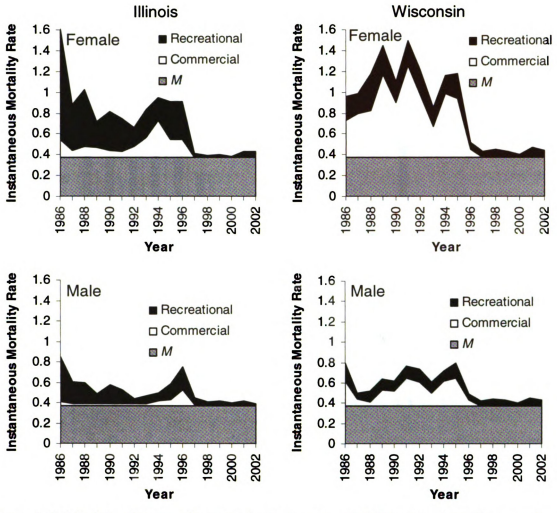

Figure 1.5. Model estimates of average instantaneous mortality rates for yellow perch age-4 and older in Illinois and Wisconsin waters of southwestern Lake Michigan during 1986-2002.

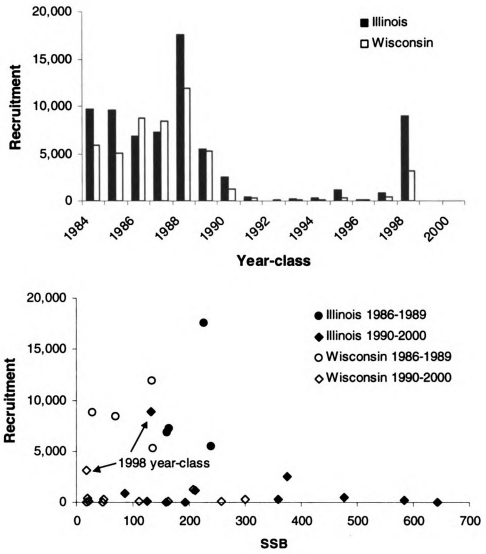

Figure 1.6. Model estimates of yellow perch recruitment (1000s) in Illinois and Wisconsin waters of southwestern Lake Michigan for the 1984-2000 year-classes and estimates of recruitment plotted against yellow perch spawning stock biomass (SSB; 1000 kg).

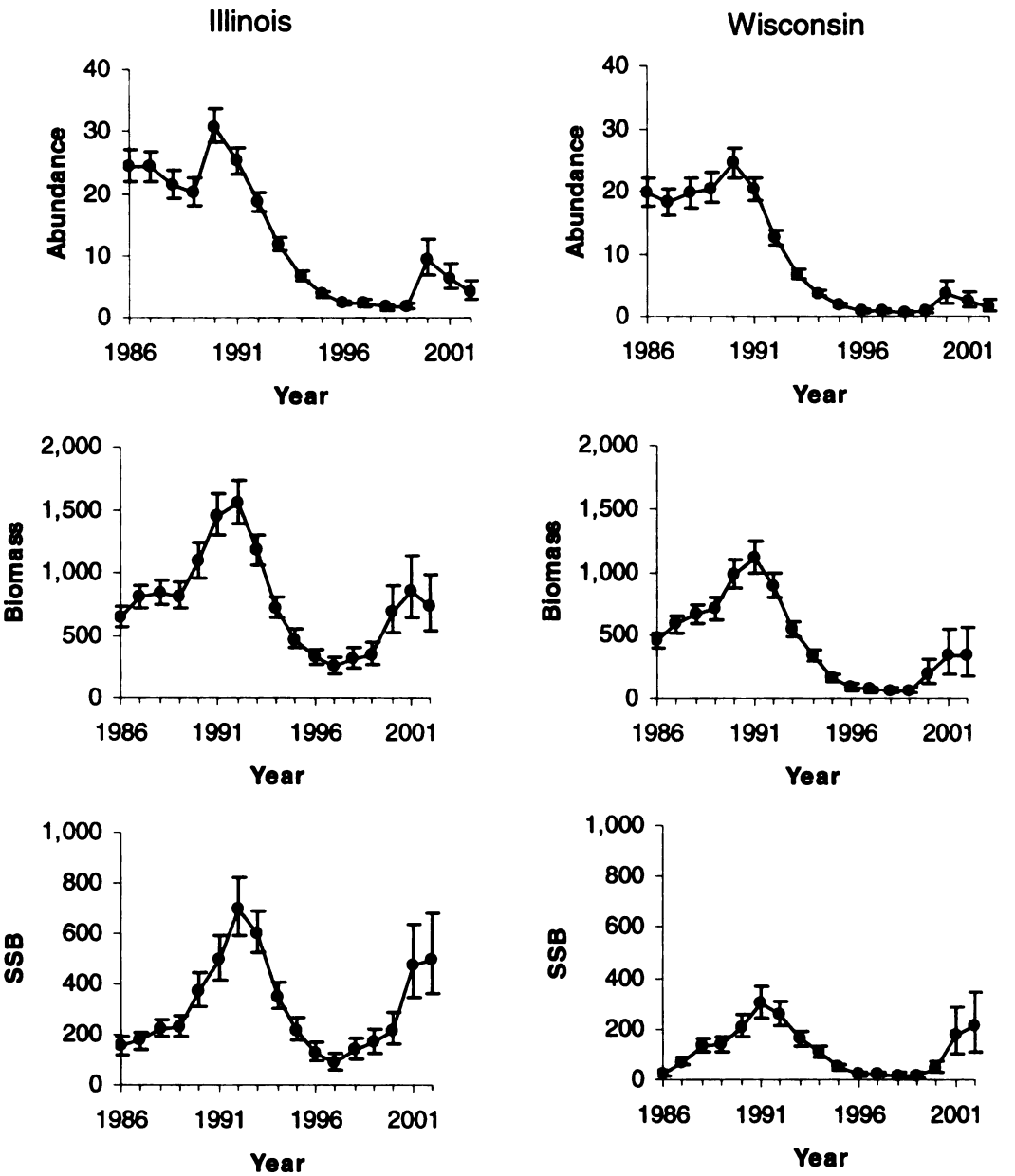

Figure 1.7. Estimated abundance (1000s), biomass (1000 kg), and spawning stock biomass (SSB; 1000 kg) of yellow perch age-2 and older in Illinois and Wisconsin waters of southern Lake Michigan during 1986-2002. Error bars represent 95% probability intervals (the Bayesian analog of confidence intervals).

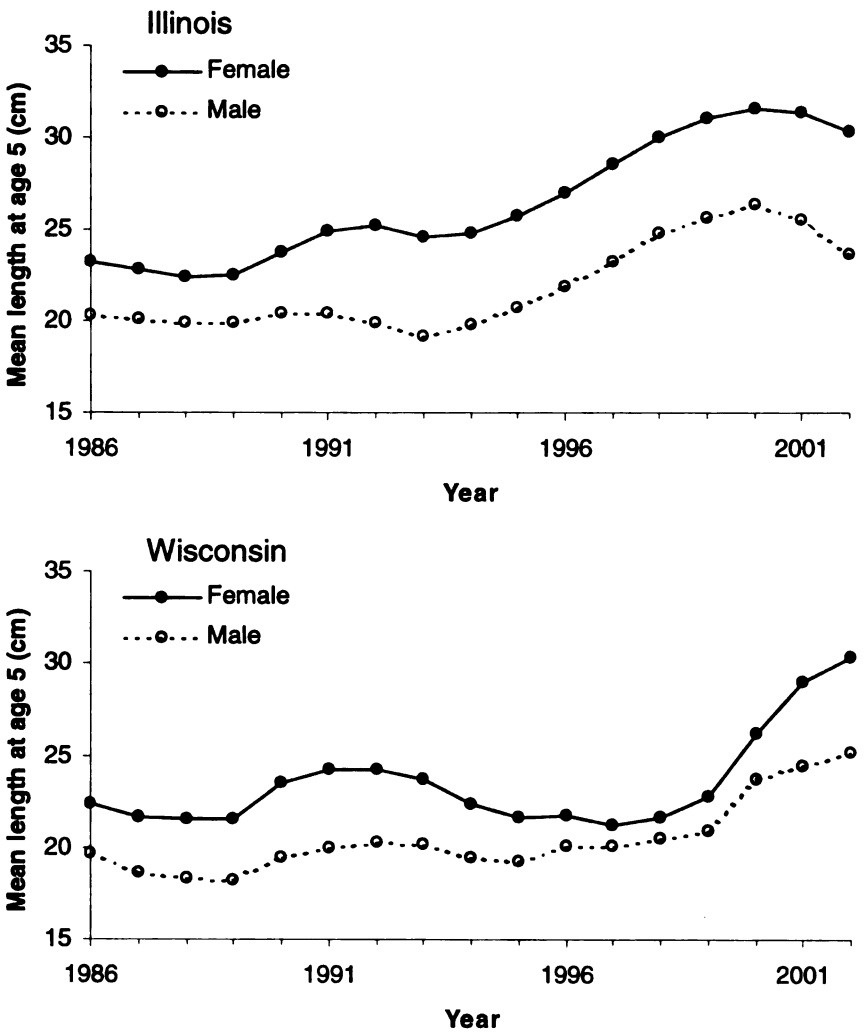

Figure 1.8. Model estimates of mean length at age 5 Illinois and Wisconsin waters of southwestern Lake Michigan during 1986-2002.

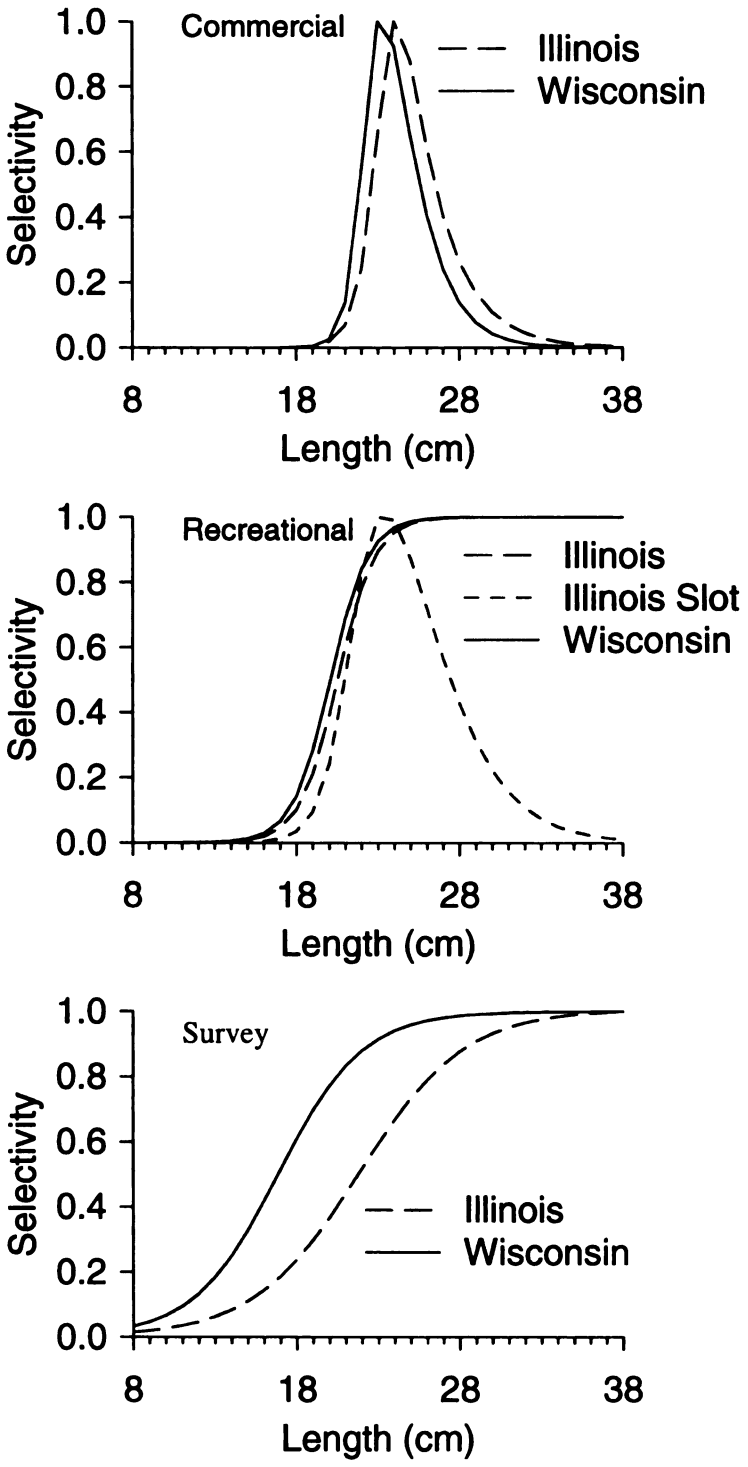

Figure 1.9. Model estimates of selectivity of the commercial fishery, recreational fishery, and survey in Illinois and Wisconsin during 1986-2002.

CHAPTER 2

PERFORMANCE OF TIME-VARYING CATCHABILITY ESTIMATORS IN STATISTICAL CATCH-AT-AGE ANALYSIS

Introduction

Statistical catch-at-age analysis (SCA) is used to provide estimates of absolute abundance, recruitment, and fishing mortality for many fisheries throughout the U.S. and the rest of the world (National Research Council [NRC] 1998; Quinn and Deriso 1999). In contrast to virtual population analysis and its variants, SCAs generally assume that fishing mortality rate-at-age can be modeled as a function of a year effect and an age effect (selectivity). This approach allows statistical estimation where fishery catch-at-age data are assumed to contain some amount of measurement error (Megrey 1989; Fournier and Archibald 1982). These models require catch-at-age data as well as an index of abundance; other data sources can also be included in the model (Deriso et al. 1985). Under many conditions, SCA provides more accurate estimates of stock size and other important management quantities than other stock assessment techniques (NRC 1998, Punt et al. 2001, Radomski et al. in press)

Many SCAs use fishery catch per effort (CPE) as an index of relative abundance, and thus assume that fishery CPE is proportional to abundance (Quinn and Deriso 1999). However, violations of this assumption can cause SCA models (and other stock assessment models) to produce biased estimates (NRC 1998). Time-varying catchability has been documented in a wide range of fisheries, spanning commercial and recreational fisheries and freshwater and marine systems. In some cases, catchability may change with abundance or the area inhabited by a stock (e.g., Peterman and Steer 1981; Winters

and Wheeler 1985; Crecco and Overholtz 1990; Harley et al. 2001), environmental effects (Ziegler et al. 2003), or due to changes in fisherman behavior or gear (Hilborn and Walters 1992). Interactions between population size, stock area, and fisher behavior can lead to hyperstable fishery CPE, where CPE remains high despite decreases in abundance (Hilborn and Walters 1992; Harley et al. 2001). Hyperstable CPE in combination with a stock assessment model that does not account for this can lead to overestimated stock size and catch limits (NRC 1998).

Methods have been developed to account for time-varying fishery catchability, but there is little consensus about best practices in this area (e.g., Fournier and Archibald 1982, Fournier 1983; Methot 1990, Fournier et al. 1998, NRC 1998). Generally, fishery effort or CPE data are ignored if an independent survey is available for a stock (NRC 1998). However, in many fisheries, survey data are not available and ignoring fishery effort data is not an option (NRC 1998). Likewise, ignoring fishery effort data may decrease the accuracy and precision of SCA estimates in some cases because fishery CPE may be informative about changes in relative population size or survey data may be poor. Our objective was to determine how well different methods of estimating time-varying catchability performed within an SCA framework. Specifically, we tested four estimation models to determine how well they performed in scenarios where catchability changed over time.

Methods

We used Monte Carlo simulations to compare how four different methods of estimating fishery catchability within an SCA model performed when models were confronted with different data generating scenarios. Our data generating models included

five cases where catchability changed abruptly or gradually over time and where catchability was explicitly a function of population abundance. Our data generating models also contained three levels of fishing mortality and three levels of survey measurement error. While the general influence of fishing mortality level and survey measurement error on the performance of SCA methods is well understood (e.g., Bence et al. 1993), we included these factors to determine whether they act to change the relative performance of different approaches of modeling time-varying catchability. We generated 1000 datasets for each scenario (45 total scenarios). For data sets that included survey data, we fit each data set with four different models that made different assumptions regarding fishery catchability; catchability was modeled as white noise, a random walk, density dependent, or catchability was effectively estimated as a free parameter for each year. This last method ignores any information contained in fishery CPE or effort. For data sets that did not include survey data we used the first three of these estimation methods. Each scenario used the same sets of random numbers.

All models contained 15 years of data and eight age classes with the last age class representing all fish that age and older. Data generating models were based on commercial fisheries for lake whitefish (*Coregonus clupeaformis*) in the upper Great Lakes. Symbols and equations defining the data generating models and estimation models are presented in Tables 2.1 and 2.2. Equations are referred to in the text as eq. Tx.y, where x is the table number and y is the equation number within Table x. To avoid redundancy, equivalent quantities and parameters in estimation and data generating models are not differentiated except when they both appear in the same equation, in which case estimated quantities are denoted with a caret above the symbol.

Data Generating Model

The data generating model described the population dynamics and created data sets of total fishery catch, the age composition of the fishery catch, and in some scenarios total survey CPE and the age composition of the survey. For the population dynamics, we used an age-structured model that followed cohorts over time. Recruitment (abundance at age 1) was generated from a lognormal distribution with a coefficient of variation (CV) of 100%. Numbers-at-age in the first year were calculated assuming a stable age distribution with lognormal errors, where recruitment and mortality rates prior to the first year of the simulation were on average the same as in the first year (eq. T2.2.1). Cohorts were tracked over time by applying a simple exponential mortality model (eq. T2.2.2a); the last age class was treated as representing all fish age 8 and older (eq. T2.2.2b). Biomass each year was the sum of age-specific abundance and mean weight-at-age (eq. T2.2.3).

We used a separable model to generate fishing mortality rates. The total mortality rates were determined by the natural mortality rate and age-specific fishing mortality rates (eq. T2.2.4). M was held constant across ages and years at 0.25. The instantaneous fishing mortality rate was a function of catchability, fishing effort, and age-specific selectivity (eq. T2.2.5). We used three levels of fishing mortality where F at fully selected ages was approximately $2M$ (high), M (medium), and $0.5\ M$ (low). We allowed fishing mortality to change over time by allowing effort to change (Figure 2.1) and by incorporating several processes of time-varying catchability (see below). For a given level of fishing mortality, each of the models used the same effort series and each effort

series had the same amount of contrast in absolute terms. The selectivity pattern for the fishery was dome shaped to simulate a gill net fishery (Figure 2.2).

We included five models for time-varying catchability, which incorporated a range of possible ways that catchability could vary over time. The $\log_e$ of catchability was modeled as white noise to simulate a fishery where catchability varied from year to year about a constant mean (eq. T2.2.6), perhaps due to environmental effects, but where year-to-year deviations were not correlated. We also included four treatments that had varying amounts of autocorrelation: first order autoregressive (AR1), density dependent, linear increase, and abrupt change. The AR1 process was also on the $\log_e$-scale and could mimic catchability changes from many sources (eqs. T2.2.7a, T2.2.7b), such as density dependent catchability or correlated environmental effects. We set the correlation (ρ) of the AR1 process to 0.9 and the CV (σ_ε) to 0.16. Density dependent catchability followed a power relationship where catchability declined with increasing abundance (eq. T2.2.8; Paloheimo and Dickie 1964). Because each of the different levels of fishing mortality had different average levels of abundance, we used three sets of parameters (α and β) to define the density dependent power function, one for each level of fishing mortality. In the linear increase scenario, catchability increased linearly over time (eq. T2.2.9), which could represent learning by fishers or increases in gear efficiency. In the abrupt change scenario, catchability was constant until year eight of the time series and increased to a higher level where it remained for the rest of the time series (eq. T2.2.10). This scenario simulated the adoption of a more efficient technology by the fishery. All models were parameterized to have the same expected catchability (over the time series) and similar variances of $\log_e q_f$. We achieved this by simulating data sets and adjusting

the catchability parameters until the mean and variance of catchability were the same as in the white noise case. We used a value of 0.2 for the standard deviation of the log of catchability as the standard for all other catchability models. This value is similar to estimates of the CV of catchability for commercial fisheries in New Zealand (Francis et al. 2003), but was less than median values of the CV of fishery CPE estimated by Harley et al. (2001) for International Council for the Exploration of the Sea fisheries of 0.4-0.8, which should be an upper bound.

Fishery catch was calculated with the Baranov catch equation (eq. T2.2.13; Quinn and Deriso 1999). We multiplied total catch by a lognormal measurement error to calculate observed fishery catch (eq. T2.2.14); the measurement error CV for fishery catch was 10%. Observed age compositions were generated by drawing a sample from a multinomial distribution of size n (100 for the fishery) with proportions equal to the expected catch-at-age in the fishery. Survey CPE-at-age was calculated as the product of survey catchability, abundance, and survey selectivity (eq. T2.2.15), and observed survey CPE was the product of total survey CPE and a lognormal measurement error (eq. T2.2.16). Our simulation model contained three levels of survey quality with differing levels of measurement error: good CV=0.25, poor CV=1.0, and no survey. Catchability of the survey was constant over time. Observed survey age compositions were generated by drawing a random sample from a multinomial distribution of size 75 with proportions equal to the expected CPE at age in the survey.

Estimation Model

The estimation models were largely the same as the simulation models except for how catchability was estimated and how numbers-at-age in the first year and recruitments

were handled. Common parameters among models included $N_{1,1}...N_{15,1}$ (Recruitment), $N_{1,2}...N_{1,8}$ (numbers-at-age in the first year), and $s_{1,f}...s_{7,f}$ (fishery selectivity); models with surveys also included $s_{1,s}...s_{7,s}$ (survey selectivity) and q_s (survey catchability). Numbers-at-age in the first year and recruitment for each year were estimated as parameters during the model fitting process. After the first year and age, abundance-at-age followed a standard exponential mortality model with the last age representing all fish that age and older (eqs. T2.2.2a, T2.2.2b).

The total mortality rate ($Z_{y,a}$) was the sum of M and $F_{y,a}$ (eq. T2.4); M was assumed known at 0.25 (the true value from the simulation models). Fishing mortality followed a separable model for all of our estimation models. Fishery and survey selectivities were estimated as individual parameters by setting selectivity at the oldest age-class to one. Estimation models contained four methods of estimating catchability: white noise, random walk, density dependent, and no catchability (directly estimating fishing mortality) with survey data. The first estimation model allowed fishery catchability to vary with white noise about a constant mean (eq. T2.2.6). The second estimation model allowed fishery catchability to vary according to a random walk (eq. T2.2.16). The third estimation model allowed catchability to be a density dependent function (eq. T2.2.8). The density dependent model did not contain any random deviations. In our fourth estimation model, we estimated the fishing mortality rate for fully selected age classes as a parameter, and then applied the estimated fishery selectivity to calculate age-specific fishing mortality rates (eq. T2.2.17). This method does not use fishery effort as a data source. The estimation models also predicted proportions of fishery and survey catch-at-age.

Model Fitting and Convergence

We fit the models using a likelihood-based approach where we used a numerical search to find parameter values that minimized our objective function. The objective function was the sum of the likelihood components and each component was the negative of the log-likelihood for a single data source or a penalty related to time-varying catchability (eq. T2.3.1).

Our estimation models assumed lognormal distributions of errors for total catch for the fishery (eq. T2.3.2) and survey CPE (eq. T2.3.3) and multinomial distributions for age compositions of the fishery (eq. T2.3.4) and the survey (eq. T2.3.5; Fournier and Archibald 1982). Effective sample sizes and CVs of the fishery and survey catch and age compositions were set to their true values from the generating models. The likelihood components for survey CPE and age composition were only included in models that included survey data.

For estimation models that used fishery effort as a data source, fishery CPE was not explicitly modeled. Instead, fishing mortality was an explicit function of effort, and catch was linked to abundance and fishery effort by estimating the catchability coefficient. We assumed lognormal deviations for catchability in the white noise (eq. T2.3.6) and random walk (eq. T2.3.7) estimation models. The CV for the white noise catchability was set to the true expected value, which was 0.2 for all data generating models. For the random walk model, we set the CV to 0.165, the CV that on average created a time series with a sample CV of 0.2. This component in the objective function can be thought of as a penalty that produces a shrinkage estimator (in the Frequentist case) or as a Bayesian prior and penalizes large deviations from mean catchability (for

the white noise model) or large year-to-year deviations (in the random walk model). Estimation models that contained density dependent catchability or ignored effort data did not contain likelihood component ℓ_5.

We minimized the objective function iteratively using an efficient quasi-Newton implementation in AD Model Builder software that takes advantage of automatic differentiation (Otter Research Limited 2000). We minimized the objective function in stages, where the initial stages were penalized if the model estimates deviated from the expected average fishing mortality rates under each scenario (early stages can be viewed as providing starting values for subsequent stages). This constraint was removed for the final stage of fitting and therefore did not penalize final model estimates. Iterative adjustment of the parameters terminated when the maximum gradient of parameters with respect to the objective function was less than 0.0001, or more than 1000 function evaluations had occurred. We denoted any terminated parameter estimates where the maximum gradient component was less than 0.0005 as converged, based on trial investigations after the completion of the simulations that used different parameter starting values.

Evaluation of Estimation Model Performance

In stock assessments, estimated quantities in the last year are often most important for forecasting and management. Therefore, we evaluated estimation model performance by calculating the relative error (RE) of estimated biomass in the last year.

$$RE = \frac{estimated - true}{true} \tag{2.1}$$

We report only results for stock size measured in biomass. Other common measures of stock size (e.g., measures of exploitable abundance) showed similar patterns

and estimates of fully selected F or exploitation rate reflected similar but inverse patterns (i.e., if estimated biomass was higher than the true value, estimated F was usually lower than the true value and vice versa). We evaluated systematic over or under estimation using the median of the relative error (MRE). If MRE equals zero, half of the estimates are higher than the true value and half are lower than the true value. Throughout the rest of the paper we use the term unbiased as meaning median unbiased (i.e., MREs near zero). We also compared estimation model performance using the median of the absolute values of relative error (MARE), which indicates the width of the distribution of REs if the median is zero. In situations where the REs are either all (or mostly) positive or negative, the MRE will equal the MARE. We compared relative performance of the estimation models by calculating the difference of their MAREs and report these differences as percentages because the units of MARE are percent. We used MRE and MARE instead of mean relative error and root mean squared error because mean values were heavily influenced by several cases with large relative errors (>100). We checked whether these outliers represented false convergence by restarting the estimation with different starting values. Convergence was verified and obtained the same parameter estimates.

Results

All estimation models performed best in situations with high fishing mortality and low survey CV and worst in cases with low fishing mortality and no survey (Table 2.4, Table 2.5). The performance of a given estimation model depended on the level of fishing mortality, survey quality, and data generating model. In almost all cases, estimation models that made use of both survey CPE and fishery effort outperformed

models that used only fishery effort or survey CPE. Performance of the estimation model that ignored fishery effort data was independent of the underlying catchability model that generated the data and was only a function of survey quality and fishing mortality. The estimation model that ignored effort data was relatively unbiased (MRE near zero) in all cases, but the MARE was often significantly higher than for estimation models assuming white noise and random walk catchability and the relative performance of this method was highly dependent on survey quality. For the other estimation models, the results can be separated into two categories: ones where all estimation models were relatively unbiased (white noise, autoregressive, and density dependent) and ones where some estimation models had substantial bias (linear increase and abrupt change). Although the density dependent estimation model was relatively unbiased in many cases, it performed relatively poorly overall because it did not converge for 15-35% of the simulated data sets that did not contain density dependent catchability; the other estimation models usually failed to converge less than 1% of the time. This lack of convergence likely occurred because the two parameters describing density dependent catchability were confounded with one another (i.e., many combinations of α and β could produce equally good fits) for many data sets, and thus the optimization procedure could not find a unique best solution. Because of problems with convergence in most cases, we did not believe that the density dependent estimation model was a viable candidate for most situations.

White Noise, First Order Autoregressive, and Density Dependent

In cases where the data generating models contained white noise catchability, first order autoregressive catchability, or density dependent catchability, all estimation models produced relatively unbiased estimates of biomass in the last year (i.e., MREs near zero;

Table 2.4), with the most biased estimation model in these scenarios having an MRE of only -6.1% (random walk estimation model fitting density dependent generation model with low mortality and no survey). There were larger differences in precision among the estimators and this was reflected in MARE and the tightness of the distributions of relative errors (Table 2.4). For cases where the estimation model was the same as the generating model (white noise and density dependent), the estimation model that matched the generating model performed best (i.e., had the lowest MARE and tighter distributions). In the case of the AR1 data generating model, the random walk model performed best in most cases. Differences in MARE among estimation models that modeled catchability as white noise, a random walk, or ignored fishery effort were usually less than 5% for cases with good surveys (Figure 2.3; Figure 2.4). However, MAREs of random walk and white noise estimation models were 7-30% lower than estimation models that ignored fishery effort in cases with a poor survey. Differences in estimation model relative performance were largely accounted for by differing performance of random walk and white noise catchability models because the performance of the estimation model that ignored fishery effort data was relatively constant for a given level of fishing mortality and survey quality. White noise and random walk models were most accurate in cases with white noise catchability, somewhat less accurate for cases with density dependent catchability, and least accurate in cases with AR1 catchability.

Linear Increase and Abrupt Change

The white noise and random walk estimation models were biased in cases where catchability increased linearly or changed abruptly, but the amount of bias depended on

survey quality, fishing mortality rate, and data generating model. The MREs of biomass in the last year for estimation models with white noise and random walk catchability were above zero in all cases, indicating a positive bias (Table 2.5). The positive bias seen in our simulations undoubtedly reflects the direction of change in catchability built into our simulations, where the estimation models did not fully account for the increase in fishery catchability. Neither the white noise nor the random walk estimation models performed well in cases with no survey, trending catchability, and low mortality. The amount of bias was highest in cases where fishery catchability changed abruptly and fishing mortality rates were low and decreased as the level of fishing mortality increased and as survey quality improved.

Although the random walk estimation model was biased, it usually had a lower MARE than our other estimation models, but performance relative to the other estimation models depended on the treatment. In cases with a good survey, the MARE of the estimation model that ignored fishery effort and the MARE of the random walk estimation model were within 5% of one another (Figure 2.3). However, in cases with a poor survey, the random walk model usually had MAREs 10-20% lower than the estimation model that ignored fishery effort. The estimation model that ignored fishery effort data only outperformed the random walk model in the scenario with an abrupt change in catchability and low fishing mortality. The estimation model that ignored fishery effort and the random walk estimation model clearly outperformed the white noise estimation model in these cases and had MAREs 12-50% lower than the white noise estimation model (Figure 2.4).

Discussion

Often stock assessment scientists will not use or will substantially downweight (i.e., specify an arbitrarily large CV) fishery effort or CPE data in an SCA if a fishery independent index of abundance is available for a given stock. Indeed, the NRC (1998) recommended that fishery dependent indices of abundance should be ignored if an independent index of abundance is available based on the results of their simulations. However, our results argue against automatically ruling out the use of fishery dependent indices of abundance when a survey is present. In cases where the survey CV is large, we believe that use of fishery dependent indices is justified if they are believed to contain information on stock size. Of course fishery effort should be adjusted for known changes in fishing efficiency, and the estimation model should allow for flexible changes in catchability over time, as was the case for our random walk estimator. The reliability of fishery effort data may be suspect in some fisheries and, in these cases, it may make sense to ignore fishery effort. Using methods that do not allow for trends in catchability can lead to severely biased SCA estimates, and modeling fishery catchability as white noise (which is often done) may not provide the necessary flexibility for models to accurately depict system dynamics. Also, there may be a tendency to overstate the precision of survey data and understate the precision of fishery data in SCAs, which is what Francis et al. (2003) found for assessments of many New Zealand commercial fisheries.

Our recommendations are contrary to NRC (1998), because we evaluated a wider range of structural models for time-varying fishery catchability within SCAs, but our results yield similar insights for the cases they explored. In the NRC (1998) study,

fishery catchability increased over time combined with density dependence; their survey had a CV of 30% (near the level of our "good" survey). Also, the NRC (1998) study mainly included SCA estimation models that contained white noise models for catchability or ignored fishery effort data (see Restrepo (1998) for details of models used in the NRC (1998) study). The exception was one estimation model where fishery catchability was modeled as a mixture of random walk and white noise processes (Ianelli and Fournier 1998). However, the CV of the white noise term was large relative to the CV of the random walk term (Ianelli and Fournier 1998), which likely caused the model to perform similarly to a white noise model. Similar to the results of NRC (1998), we also found that that SCA models that ignored fishery effort data outperformed SCA models that modeled fishery catchability as white noise in cases with trending catchability.

Independent survey indices of abundance or relative abundance are extremely important for obtaining accurate SCA estimates, especially in situations with low fishing mortality. Our results agree with the NRC (1998) recommendation to use survey data if they are available. In our study, estimation models that utilized fishery effort data and survey data (even with a CV of 100%) outperformed models that used only fishery effort data, especially in cases where catchability trended over time and fishing mortality was not high.

It is important to standardize effort series to remove catchability trends to as large an extent as possible. Our experiments showed that SCA estimates were most biased when trends or abrupt changes in fishery catchability occurred and that all our estimation models performed reasonably well in cases where catchability did not trend over time.

Trending fishery catchability is probably common. Many mechanisms could lead to trends in fishery catchability, such as increasing power of the fishery, increasing aggregation of fish stocks and fishers, or trending recruitment dynamics and density dependent catchability. Salthaug and Aanes (2003) presented a method to correct CPE for the spatial distribution of fishing effort, which has been shown to affect catchability (Winters and Wheeler 1985; Rose and Kulka 1999). Also, improvements in vessels, and other fisher behaviors can be accounted for either by preprocessing (e.g., analyzing CPE data to estimate mean CPE by accounting for vessel characteristics and spatial and temporal patterns of fishing) fishery data or by integrating the standardization process into the stock assessment model (e.g. Maunder 2001; Maunder and Starr 2003; Maunder and Punt 2004). The procedure of simultaneously standardizing catch and effort data and fitting the stock assessment model can lead to improved estimates over the two-step approach of standardizing catch and effort data and then fitting the assessment model with the standardized values (Maunder 2001; Maunder and Langley 2004).

Our results probably provide a best-case view of the performance of SCAs when faced with time-varying catchability and may exaggerate the accuracy of all estimation models used in our study. Except for the catchability aspect, the structure of the estimation models was correct (i.e., the same as the data generating model). In reality, it is likely that M may vary among years and ages and that the data analyst will not know the true M. Fishery selectivity may vary over time, which can cause biased estimates from SCA models if it is not accounted for (Radomski et al. in press). Likewise, our models did not contain trends in survey catchability over time or correlation with changes in fishery catchability, which could cause models that used survey indices of abundance

to generate less accurate estimates. Lastly, our data generating models contained a survey with an asymptotic selectivity pattern, which allows SCA models to produce more accurate estimates than other survey selectivity patterns (Bence et al. 1993).

While our results favored the random walk model in general, this was not true under all circumstances. We recommend that data analysts fit multiple stock assessment models with different assumptions about time-varying catchability. One may be able to then determine the best catchability model using a Bayesian framework, where each of the catchability models we fit is a special case of a "full" model (McAllister and Kirchner 2002; Gelman et al. 2004). For instance, the estimation models that ignore effort data, use white noise, or a random walk are all special cases of a first order autoregressive process (eq. T2.8). In the case of white noise, the correlation coefficient (ρ) equals 0. In the case of random walk, ρ equals 1. And in the case of ignoring effort data, the CV of the random deviations (σ) is infinity. Thus, one possible procedure would be to allow catchability to follow a first order autoregressive process and estimate the ρ and σ parameters. If the CVs of the other likelihood components are specified, these parameters (ρ and σ) may be estimable and this method could lead to better SCA estimates of parameters and uncertainty. Alternative approaches would be to select among our special case models using the deviance information criterion (Spiegelhalter et al. 2002) or other measures that account for both goodness of fit and model complexity, or to average over the alternative models using Bayesian Model Averaging (McAllister and Kirchner 2002). These are topics warranting future research.

References

Bence, J. R., A. Gordoa, and J. E. Hightower. 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. Can. J. Fish. Aquat. Sci. 50: 827-840.

Crecco V. A., and W. Overholtz. 1990. Causes of density-dependent catchability for Georges Bank haddock *Melanogrammus aeglefinus*. Can. J. Fish. Aquat. Sci. 47: 385-394.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42: 815-824.

Fournier, D. A. 1983. An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density dependent effects. Can. J. Fish. Aquat. Sci. 40: 1233-1243.

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39: 1195-1207.

Fournier, D. A., J. Hampton, and J. R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55: 2105-2116.

Francis, R. I. C. C., R. J. Hurst, and J. A. Renwick. 2003. Quantifying annual variation in catchability for commercial and research fishing. Fish. Bull. 101: 293-304.

Harley, S. J., R. A. Myers, and A. Dunn. 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. 58: 1760-1772.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment. Chapman and Hall, New York.

Maunder, M. N. 2001. A general framework for integrating the standardization of catch per unit effort into stock assessment models. Can. J. Fish. Aquat. Sci. 58: 795-803.

Maunder, M. N., and A. D. Langley. 2004. Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fish. Res. 70: 389-395.

Maunder, M. N., and A. E. Punt. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70: 141-159.

Maunder, M. N., and P. J. Starr. 2003. Fitting fisheries models to standardized CPUE abundance indices. Fish. Res. 63: 43-50.

McAllister, M, and C. Kirchner. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. Bull. Mar. Sci. 70: 499-540.

Megrey, B. A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. *In* Mathematical analysis of fish stock dynamics. *Edited by* E. F. Edwards, and B. A. Megrey. American Fisheries Society Symposium 6: 8-48.

Methot, R. D. 1990. Synthesis Model: an adaptable framework for analysis of diverse stock assessment data, p. 259-277. *In* Proceedings of the Symposium on Application of Stock Assessment Techniques to Gadids. *Edited by* L. Low. International North Pacific Fisheries Commission Bulletin 50.

National Research Council (NRC). 1998. Improving fish stock assessments. National Academy Press. Washington D.C.

Otter Research Limited. 2000. An introduction to AD Model Builder version 4 for use in nonlinear modeling and statistics. Otter Research Ltd., Sidney B.C.

Peterman, R. M., and G. J. Steer. 1981. Relation between sport-fishing catchability coefficients and salmon abundance. Trans. Am. Fish. Soc. 110: 585-593.

Punt, A. E., A. D. M. Smith, and G. Cui. 2002. Evaluation of management tools for Australia's south east fishery 2. how well can management quantities be estimated? Mar. Freshw. Res. 53: 631-644.

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

Radomski, P. A., J. R. Bence, and T. J. Quinn II. In press. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational kill dominated fishery. Can. J. Fish. Aquat. Sci.

Rose, G. A., and D. W. Kulka. 1999. Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus horhua*) declined. Can. J. Fish. Aquat. Sci. 56(Suppl. 1): 118-127.

Salthaug, A., and S. Aanes. 2003. Catchability and the spatial distribution of fishing vessels. Can. J. Fish. Aquat. Sci. 60: 259-268.

Spiegelhalter, D.J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 64:583-639.

Winters, G. H., and J. P. Wheeler. 1985. Interactions between stock area, stock abundance, and catchability coefficient. Can. J. Fish. Aquat. Sci. 42: 989-998.

Ziegler, P. E., S. D. Frusher, and C. R. Johnson. 2003. Space-time variation in catchability of southern rock lobster *Janus edwardsii* in Tasmania explained by environmental, physiological and density-dependent processes. Fish. Res. 61: 107-123.

Table 2.1. Symbols and descriptions of variables for data generating and estimation models.

Symbol	Description	Value (if needed in the data generating model)
$\overline{R}$	Average recruitment	1,000,000
$N_{y,a}$	Abundance by age and year	
B_y	Biomass	
$Z_{y,a}$	Total instantaneous mortality rate by age and year	
$F_{y,a}$	Instantaneous fishing mortality rate by age and year	
M	Instantaneous natural mortality rate	0.25
$s_{a,f}$	Fishery age-specific selectivity	See figure 2.2
$s_{a,s}$	Survey age-specific selectivity	See figure 2.2
E_y	Fishery effort	See figure 2.1
$q_{y,f}$	Fishery catchability	
q_s	Survey catchability	0.0001
$\overline{q}_f$	Mean fishery catchability	0.05
$C_{y,a}$	Expected fishery catch-at-age	

$I_{y,a}$	Expected survey catch-at-age	
$\tilde{C}_y$	Observed total fishery catch	
$\tilde{I}_y$	Observed total survey catch	
$u_{y,a,f}$	Proportion of catch-at-age in fishery	
$u_{y,a,s}$	Proportion of catch-at-age in survey	
w_a	Mean weight at age	0.16, 0.45, 0.82, 1.2, 1.55, 1.86, 2.11, 2.3
δ_y	Deviations for white noise catchability	
ε_y	Deviations for first order autoregressive catchability	
ω_y	Deviations for random walk catchability	
α, β	Parameters for density dependent catchability (low, medium, high)	175, 0.53; 90, 0.49; 35, 0.42
a, b	Parameters for linear increase in catchability	0.032, 0.00225
q_1 , q_2	Parameters for abrupt change in catchability	0.0402, 0.0598
f_y	Fishing intensity by year	
ρ	Correlation parameter for autoregressive catchability	0.9
σ_γ	CV for recruitment variation	1.0

σ_τ	Fishery measurement error CV	0.1
σ_υ	Survey measurement error CV	0.25; 1.0
σ_δ	CV for white noise catchability deviations	0.2
σ_ε	CV for autoregressive catchability deviations	0.16
σ_ω	CV for random walk catchability deviations	0.165

Table 2.2. Data generating and estimation model equations.

	Population model equations	Application
(T2.2.1)	$N_{1,a} = \bar{R}e^{-\sum_{a=1}^{a-1} Z_{1,a} + \gamma_a}; \gamma \sim N\left(0, \sigma_\gamma^2\right)$	Generation
(T2.2.2a)	$N_{y+1,a+1} = N_{y,a}e^{-Z_{y,a}}$	Both
(T2.2.2b)	$N_{y+1,8} = N_{y,7}e^{-Z_{y,7}} + N_{y,8}e^{-Z_{y,8}}$	Both
(T2.2.3)	$B_y = \sum_a N_{y,a} w_a$	Both
(T2.2.4)	$Z_{y,a} = M + F_{y,a}$	Both
(T2.2.5)	$F_{y,a} = q_y E_y s_a$	Both
	Catchability model equations	
(T2.2.6)	*White noise*	Both
	$\log_e q_{y,f} = \log_e \bar{q}_f + \delta_y; \delta_y \sim N\left(0, \sigma_\delta^2\right)$	
(T2.2.7a)	*First order autoregressive*	Generation
	$\log_e q_{1,f} \sim N\left(\log_e \bar{q}_f, \frac{\sigma_\varepsilon^2}{1-\rho^2}\right)$	
(T2.2.7b)	$\log_e q_{y+1,f} = \log_e \bar{q}_f + \rho\left(\log_e q_{y,f} - \log_e \bar{q}\right) + \varepsilon_y$ $\varepsilon_y \sim N\left(0, \sigma_\varepsilon^2\right)$	Generation

(T2.2.8) *Density dependent* Both

$$q_{y,f} = \alpha N_y^{-\beta}$$

(T2.2.9) *Linear increase* Generation

$$q_{y,f} = a + b(y)$$

(T2.2.10) *Abrupt change* Generation

$$q_{y,f} = \begin{matrix} q_1 \text{ if } y < 8 \\ q_2 \text{ if } y \geq 8 \end{matrix}$$

(T2.2.11) *Random walk* Estimation

$$\log_e q_{y+1,f} = \log_e q_{y,f} + \omega_y ; \omega_y \sim N\left(0, \sigma_\omega^2\right)$$

(T2.2.12) *Freely estimate f_y (ignore fishery effort)* Estimation

$$F_{y,a} = f_y s_{a,f}$$

Observation model equations

(T2.2.13) Both

$$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}}(1 - e^{-Z_{y,a}})N_{y,a}$$

(T2.2.14) Both

$$\tilde{C}_y = e^{\tau_y} \sum_a C_{y,a} ; \tau_y \sim N(0, \sigma_\tau)$$

(T2.2.15) $I_{y,a} = q_s s_a N_{y,a}$ Both

(T2.2.16) Both

$$\tilde{I}_y = e^{\upsilon_y} \sum_a I_{y,a} ; \upsilon_y \sim N(0, \sigma_\upsilon)$$

Table 2.3. Objective function equations for statistical catch-at-age analysis simulation study. Equations T2.3.3 and T2.3.5 were only used in estimation models that considered survey data. Equations T2.3.6 and T2.3.7 were only used in estimation models that modeled fishery catchability as white noise or a random walk respectively.

(T2.3.1)	$L = \sum_i \ell_i$	Objective function
(T2.3.2)	$\ell_1 = \frac{1}{2\sigma_\tau^2} \sum_y \left(\log_e(\tilde{C}_y) - \log_e(\hat{C}_y) \right)^2$	Fishery catch
(T2.3.3)	$\ell_2 = \frac{1}{2\sigma_\upsilon^2} \sum_y \left(\log_e(\tilde{I}_y) - \log_e(\hat{I}_y) \right)^2$	Survey catch-per-effort
(T2.3.4)	$\ell_3 = -n_f \sum_y \sum_a u_{y,a,f} \log_e(\hat{u}_{y,a,f})$	Proportion at age in the fishery catch
(T2.3.5)	$\ell_4 = -n_s \sum_y \sum_a u_{y,a,s} \log_e(\hat{u}_{y,a,s})$	Proportion at age in the survey catch
(T2.3.6)	$\ell_5 = \frac{1}{2\sigma_q^2} \sum_y \left(\hat{\delta}_y \right)^2$	White noise catchability
(T2.3.7)	$\ell_5 = \frac{1}{2\sigma_q^2} \sum_y \left(\hat{\omega}_y \right)^2$	Random walk catchability

Table 2.4. Simulation results for statistical catch-at-age estimation model performance in cases where data generating models included white noise catchability (WN), first order autoregressive catchability (AR), and density dependent catchability (DD). Shown are median relative error (MRE) and median of the absolute values of relative error (MARE) for estimated biomass in the last year (year 15) from four statistical catch-at-age estimation models: white noise (WN), random walk (RW), density dependent (power relationship; DD), and freely estimated F at maximum selectivity (i.e., not fitted to fishery effort data; FF). Data generating models included three levels of fishing mortality (high [$F=2M$], medium [$F=M$], and low [$F=0.5M$]), and 3 levels of survey precision (good [CV=25%], poor [CV=100%], and no survey). Estimation models with the lowest MARE for each treatment are indicated in bold.

			Estimation Model							
			MRE				MARE			
q-Model	Mortality	Survey	WN	RW	DD	FF	WN	RW	DD	FF
WN	low	good	-0.005	0.009	0.007	0.017	**0.196**	0.212	0.259	0.216
WN	low	poor	-0.006	-0.010	0.016	0.018	**0.214**	0.258	0.284	0.493
WN	low	none	0.006	-0.019	0.000		**0.231**	0.255	0.319	
WN	medium	good	0.001	0.009	0.019	0.024	**0.133**	0.155	0.191	0.170
WN	medium	poor	0.012	0.002	0.031	0.036	**0.156**	0.213	0.219	0.400
WN	medium	none	0.018	0.006	0.035		**0.183**	0.234	0.233	
WN	high	good	0.012	0.015	0.015	0.024	**0.103**	0.119	0.143	0.146
WN	high	poor	0.011	0.001	0.025	0.046	**0.124**	0.166	0.149	0.308
WN	high	none	0.020	0.017	0.041		**0.141**	0.185	0.179	
AR	low	good	-0.004	0.004	0.022	0.000	0.271	**0.222**	0.356	0.226
AR	low	poor	0.009	-0.025	0.030	-0.001	0.355	**0.335**	0.434	0.498
AR	low	none	0.020	-0.015	0.011		**0.409**	0.419	0.456	
AR	medium	good	0.008	0.006	0.006	0.031	0.211	**0.164**	0.302	0.169

AR	medium	poor	0.002	-0.045	0.034	0.014	0.297	**0.278**	0.365	0.412
AR	medium	none	0.031	-0.031	0.028		**0.332**	0.340	0.373	
AR	high	good	0.007	0.009	0.029	0.030	0.174	**0.131**	0.251	0.148
AR	high	poor	0.004	-0.022	0.029	0.019	0.236	**0.196**	0.301	0.314
AR	high	none	0.018	-0.037	0.025		0.271	**0.236**	0.311	
DD	low	good	-0.019	0.022	0.006	0.013	0.244	0.233	**0.195**	0.243
DD	low	poor	-0.025	-0.006	0.009	0.046	0.342	0.324	**0.234**	0.514
DD	low	none	-0.025	-0.061	0.007		0.401	0.384	**0.273**	
DD	medium	good	-0.026	0.009	0.024	0.027	0.177	0.163	**0.139**	0.185
DD	medium	poor	-0.023	-0.031	0.023	0.039	0.252	0.249	**0.177**	0.418
DD	medium	none	-0.020	-0.047	0.018		0.295	0.292	**0.199**	
DD	high	good	-0.010	0.005	0.027	0.031	0.137	0.121	**0.114**	0.146
DD	high	poor	-0.012	-0.028	0.020	0.037	0.193	0.178	**0.130**	0.329
DD	high	none	0.004	-0.030	0.022		0.218	0.213	**0.155**	

Table 2.5. Simulation results for statistical catch-at-age estimation model performance in cases where data generating models included linearly increasing catchability (LI) and an abrupt increase in catchability (AC). Shown are median relative error (MRE) and median of the absolute values of relative error (MARE) for estimated biomass in the last year (year 15) from four statistical catch-at-age estimation models: white noise (WN), random walk (RW), density dependent (power relationship; DD), and freely estimated F at maximum selectivity (i.e., not fitted to fishery effort data; FF). Data generating models included three levels of fishing mortality (high [$F=2M$], medium [$F=M$], and low [$F=0.5M$]), and 3 levels of survey precision (good [CV=25%], poor [CV=100%], and no survey). Estimation models with the lowest MARE for each treatment are indicated in bold.

			Estimation Model							
			MRE				MARE			
q-Model	Mortality	Survey	WN	RW	DD	FF	WN	RW	DD	FF
LI	low	good	0.353	0.050	0.437	0.013	0.353	**0.187**	0.437	0.234
LI	low	poor	0.762	0.419	0.647	0.044	0.762	**0.419**	0.647	0.501
LI	low	none	0.751	0.635	0.642		0.751	**0.635**	0.642	
LI	medium	good	0.345	0.074	0.488	0.029	0.345	**0.150**	0.488	0.174
LI	medium	poor	0.671	0.299	0.653	0.032	0.671	**0.302**	0.653	0.390
LI	medium	none	0.716	0.429	0.647		0.716	**0.429**	0.648	
LI	high	good	0.322	0.072	0.482	0.031	0.322	**0.121**	0.482	0.140
LI	high	poor	0.551	0.172	0.598	0.033	0.551	**0.189**	0.598	0.280
LI	high	none	0.605	0.211	0.611		0.605	**0.213**	0.611	
AC	low	good	0.571	0.165	1.386	0.018	0.571	0.255	1.386	**0.242**
AC	low	poor	1.154	0.655	2.145	0.055	1.154	0.655	2.145	**0.491**
AC	low	none	1.755	1.413	2.462		1.755	**1.413**	2.462	
AC	medium	good	0.371	0.088	0.730	0.029	0.371	**0.168**	0.730	0.171

AC	medium	poor	0.699	0.282	1.107	0.039	0.699	**0.300**	1.107	0.390
AC	medium	none	0.842	0.472	1.207		0.842	**0.472**	1.207	
AC	high	good	0.291	0.038	0.488	0.030	0.291	**0.114**	0.488	0.139
AC	high	poor	0.497	0.084	0.641	0.019	0.497	**0.144**	0.641	0.280
AC	high	none	0.550	0.103	0.620		0.550	**0.154**	0.620	

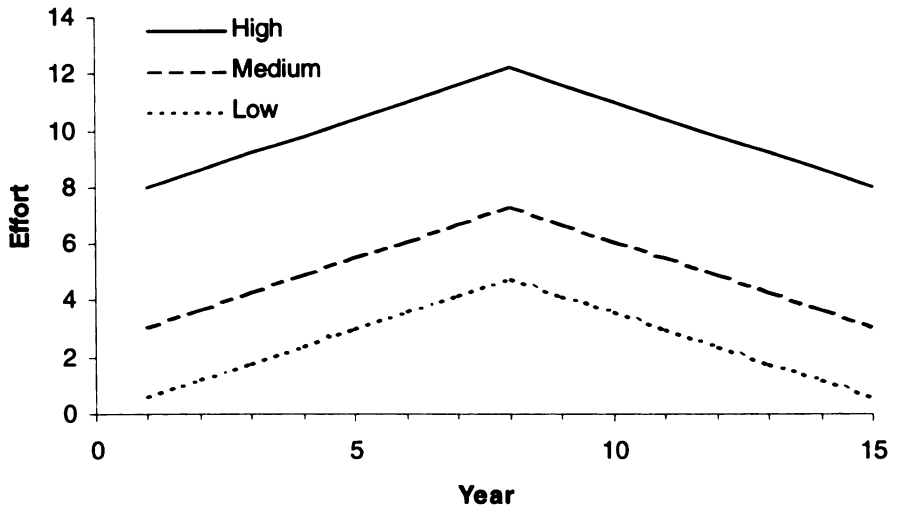

Figure 2.1. Effort series used for high, medium, and low fishing mortality rate scenarios in the data generating models. The average fishing mortality rates for fully selected age classes were approximately $2M$ for the high scenario, M for the medium scenario, and $0.5M$ for the low scenario.

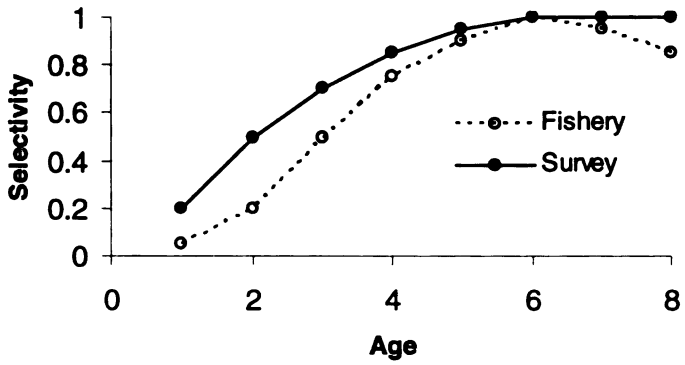

Figure 2.2. Fishery and survey selectivity patterns used in the data generating model.

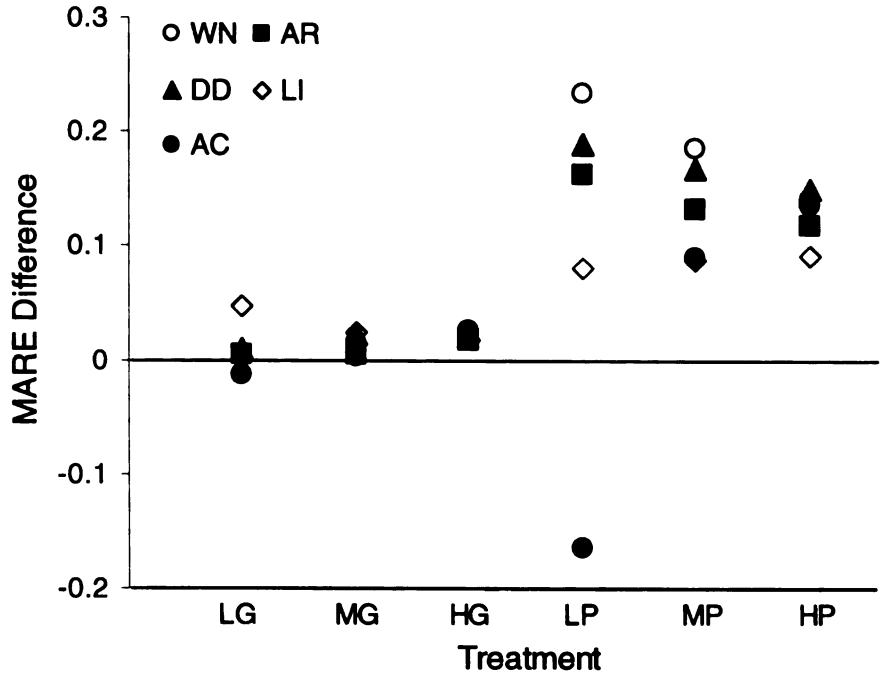

Figure 2.3. Relative performance of the estimation model that ignores fishery effort versus the random walk estimation model measured by the difference of median of the absolute value of the relative errors (MARE). Positive values indicate that the estimation model that ignored fishery effort data had a larger MARE than the random walk estimation model and vice versa. Data generating models are indicated by the symbol shape: WN – white noise, AR – autoregressive, DD – density dependent, LI – linear increase, and AC – abrupt change. Two letters identify each treatment: the first letter for level of fishing mortality (L – low, M – medium, H – high) and the second letter for level of survey quality (G – good, P – poor).

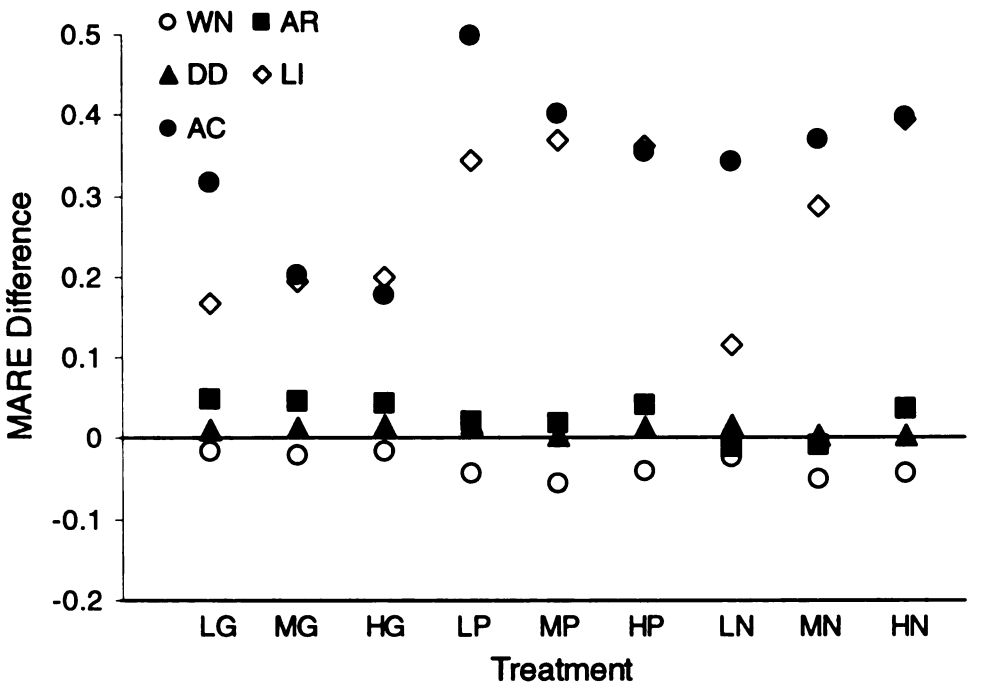

Figure 2.4. Relative performance of white noise versus random walk estimation model measured by the difference of median of the absolute value of the relative errors (MARE). Positive values indicate that the white noise estimation model had a larger MARE than the random walk estimation model and vice versa. Data generating models are indicated by the symbol shape: WN – white noise, AR – autoregressive, DD – density dependent, LI – linear increase, and AC – abrupt change. Two letters identify each treatment: the first letter for level of fishing mortality (L – low, M – medium, H – high) and the second letter for level of survey quality (G – good, P – poor, N – none).

CHAPTER 3

PERFORMANCE OF BAYESIAN MODEL SELECTION IN STATISTICAL CATCH-AT-AGE ANALYSIS

Introduction

Development of a fishery stock assessment often involves fitting alternative models and using what is thought to be the best among them to provide management advice. The "best" model is often selected by *ad hoc* criteria with unknown performance characteristics. Model selection is an area of importance because estimated quantities important for management, such as exploitable biomass, can be extremely sensitive to model structure (McAllister and Kirchner 2002). Common uncertainties in statistical catch-at-age (SCA) model structure include stock-recruitment relationships, selectivity functions, and assumptions linking fishery catch with abundance and effort (McAllister and Kirchner 2002). In some cases, results from several models will be reported to managers, but quantitative estimates of the relative likelihood a particular model being most "correct" are typically not provided (McAllister and Kirchner 2002).

Model selection has been applied to SCA models, but previous applications have been limited in the types of models that could be compared. Helu et al. (2000) evaluated performance of Akaike's Information Criterion (AIC; Akaike 1973) and Schwartz's Bayesian Information Criterion (BIC; Schwartz 1978) to assess model selection in SCA models and found that AIC and BIC both performed well by selecting the candidate model that was the same as the data-generating model in most of their scenarios. Unfortunately, although AIC or BIC may perform well in some cases, their implementation is problematic when models differ in their random effects or hierarchical

structures because the number of parameters in these models is not easy to determine (Burnham and Anderson 2002). Therefore, to be able to compare structurally complex SCA models requires alternative model selection approaches that can account for random effects and priors on parameters.

The Deviance Information Criterion (DIC) has been developed relatively recently to select among complex hierarchical models where the number of effective parameters is not readily apparent (Spiegelhalter et al. 2002). Much like AIC and BIC, DIC selects among models by trading off goodness of fit and model complexity. DIC is a generalization of AIC and reduces to AIC in the case of a model with diffuse priors (Spiegelhalter et al. 2002). DIC is particularly applicable to models with random effects or hierarchical structure because it estimates the effective number of parameters rather than requiring the user to provide this. Unlike BIC, DIC does not depend on the number of data points directly in its calculation.

Although DIC has been applied in many studies (e.g., Zhu and Carlin 2000; Barry et al. 2003), relatively few studies have evaluated the performance of DIC model selection (Spiegelhalter et al. 2002; Cardoso and Tempelman 2003; Kizilkaya and Tempelman 2003; Berg et al. 2004; Kizilkaya and Tempelman 2005; van der Linde 2005). In general, these studies found that DIC usually selected the correct model (i.e., the model that generated the data) from the set of candidate models and that the estimated number of effective parameters seemed reasonable for their given models.

Bayes factors are another method to compare models that can account for random effects and hierarchical structure (Gelman et al. 2004). Fournier et al. (1998) used posterior Bayes factors (an approximation to Bayes factors; Aitkin 1991) to estimate

"weight of evidence" of one model over another (Lavine and Schervish (1999) showed that weight of evidence is not quite an accurate description of Bayes factors). However, like AIC and BIC, posterior Bayes factors also require the number of parameters as an input to the calculation. McAllister and Kirchner (2002) estimated Bayes factors for several competing assessment models of Namibian orange roughy (*Hoplostethus atlanticus*) using the sampling-importance resampling algorithm. To date, there are no published studies of fishery stock assessments that have evaluated the performance of model selection or model averaging based on Bayes factors. However, in complex models, such as SCA models, Bayes factors can be difficult to calculate and sensitive to priors (Kass and Raftery 1995; Lavine and Schervish 1999; Han and Carlin 2001).

My objectives were to determine if using DIC or an approximation of Bayes factors as model selection criteria resulted in choosing an appropriate model structure and level of complexity. Also, I wanted to evaluate whether using formal model selection methods provided more accurate estimates of important fishery management quantities, such as fishing mortality rate and biomass in the last year. To achieve these objectives, I designed a simulation study and challenged the model selection criteria with three estimation models and three scenarios of data accuracy and time-varying catchability.

Methods

I evaluated whether using DIC and approximate Bayes factors to select among SCA model variants provided more accurate estimates of quantities used for management than an approach of using a single model structure in all cases. My data-generating models contained three basic scenarios, which differed in their relationship between fishing mortality and observed effort. These scenarios included (1) modeling fishery

catchability as white noise, (2) modeling fishery catchability as increasing a constant amount each year, and (3) treating fishing mortality as unrelated to observed effort. I chose these data-generating scenarios because previous results indicated that the relative performance of different estimation models was likely to change over this range of conditions. Three different estimation models were fitted to each of the 30 datasets (ten from each scenario). These estimation models contained different assumptions regarding fishery catchability; (1) catchability was modeled as white noise, (2) as a random walk, and (3) where catchability was effectively estimated as a free parameter for each year. This last method ignores any information contained in fishery effort data.

All models contained 15 years of data and eight age classes with the last age class representing all fish that age and older. Data-generating models were based on commercial fisheries for lake whitefish (*Coregonus clupeaformis*) in the upper Great Lakes. Symbols and equations defining the data-generating models and estimation models are presented in Tables 3.1 and 3.2. Equations are referred to in the text as eq. T*x.y*, where *x* is the table number and *y* is the equation number within Table *x*. To avoid redundancy, equivalent quantities and parameters in estimation and data-generating models are not differentiated except when they both appear in the same equation, in which case estimated quantities are denoted with a caret above the symbol.

Data-generating Model

The data-generating model described the population dynamics and created data sets of total fishery catch, the age composition of the fishery catch, total survey CPE, the age composition of the survey, and fishery effort. To model population dynamics, I used an age-structured model that followed cohorts over time. Recruitment (abundance at age

1) was generated from a lognormal distribution with a coefficient of variation (CV) of 100%. Numbers-at-age in the first year were calculated assuming a stable age distribution with lognormal errors, where recruitment and mortality rates prior to the first year of the simulation were on average the same as in the first year (eq. T3.2.1). Cohorts were tracked over time by applying a simple exponential mortality model (eq. T3.2.2a); the last age class was treated as representing all fish age 8 and older (eq. T3.2.2b). Biomass each year was the sum of age-specific abundance and mean weight-at-age (eq. T3.2.3).

I used a separable (i.e., fishing mortality was the product of an age effect and a year effect) model to generate fishing mortality rates. The total mortality rates were determined by the natural mortality rate and age-specific fishing mortality rates (eq. T3.2.4). M was held constant across ages and years at 0.25. The instantaneous fishing mortality rate was a function of catchability, fishing effort, and age-specific selectivity (eq. T3.2.5). I allowed fishing mortality to change over time by allowing fishery effort to change and by incorporating two processes of time-varying catchability (see below).

The overall level of fishing mortality varied among simulations. This was accomplished by multiplying the baseline effort (Figure 3.1) by a Uniform(1,2) number selected for each simulation. The baseline effort series was designed to produce an average level of F for fully selected ages approximately equal to M. Thus, this procedure led to F for fully selected ages varying among simulations between M and $2M$. For the white noise catchability and linearly increasing catchability scenarios observed effort equaled true effort. For the scenario with uninformative effort, the observed effort series was drawn as uniform random numbers between the minimum true effort (effort in year

1) and the maximum true effort (effort in year 8). The selectivity pattern for the fishery was dome shaped to simulate a gill net fishery (Figure 3.2).

I included two models for time-varying catchability, which caused SCA models to have variable performance (chapter 2). The $\log_e$ of catchability was modeled as white noise to simulate a fishery where catchability varied from year to year about a constant mean (eq. T3.2.6), perhaps due to environmental effects. In the second scenario, catchability increased linearly over time with a small amount of white noise error (eq. T3.2.9), which could represent learning by fishers or increases in gear efficiency. Both models were parameterized to have the same expected catchability (over the time series) and similar variances of $\log_e q_f$. I achieved this by simulating data sets and adjusting the catchability parameters until the mean and variance of catchability were the same as in the white noise case. I used a value of 0.2 for the standard deviation of the $\log_e$ of catchability. This value is similar to estimates of the CV of catchability for commercial fisheries in New Zealand (Francis et al. 2003), but was less than median values of the CV of fishery CPE estimated by Harley et al. (2001) for International Council for the Exploration of the Sea fisheries of 0.4-0.8, which should be an upper bound on the CV of catchability.

Fishery catch was calculated with the Baranov catch equation (eq. T3.2.13; Quinn and Deriso 1999). I multiplied total catch by a lognormal measurement error to calculate observed fishery catch (eq. T3.2.14); the measurement error CV for fishery catch was about 0.1. Observed age compositions for the fishery catch were generated by drawing a random sample from a multinomial distribution of size 200 with proportions equal to the true proportions of catch-at-age in the fishery. Survey CPE-at-age was calculated as the

product of survey catchability, abundance, and survey selectivity (eq. T3.2.15), and observed survey CPE was the product of total survey CPE and a lognormal measurement error (eq. T3.2.16).

As was the case for average fishing mortality, survey quality varied randomly among simulated datasets. This was accomplished by selecting the measurement error CV for each simulation from a Uniform(0.2,0.8) distribution. These levels of survey CV were selected because they provided contrast in performance of several estimation models in chapter 2. Catchability of the survey was constant over time. Observed survey age compositions were generated by drawing a random sample from a multinomial distribution of size 150 with proportions equal to the true proportions of CPE at age calculated from eq. T3.2.15.

Estimation Model

The estimation models were largely the same as the simulation models except for how catchability was estimated and how numbers-at-age in the first year and recruitments were handled. Common parameters among models included $N_{1,1}...N_{15,1}$ (Recruitment), $N_{1,2}...N_{1,8}$ (numbers-at-age in the first year), and $s_{1,f}...s_{7,f}$ (fishery selectivity), $s_{1,s}...s_{7,s}$ (survey selectivity) and q_s (survey catchability). All models had 52 unique estimated parameters. Parameterization of the models to reduce correlations among parameters is described in Appendix B. Numbers-at-age in the first year and recruitment for each year were estimated as parameters during the model fitting process. After the first year and age, abundance-at-age followed a standard exponential mortality model with the last age representing all fish that age and older (eqs. T3.2.2a, T3.2.2b).

The total mortality rate ($Z_{y,a}$) was the sum of M and $F_{y,a}$ (eq. T2.4); M was assumed known at 0.25 (the true value from the simulation models). Fishing mortality followed a separable model for all of my estimation models. Fishery and survey selectivities were estimated as individual parameters by constraining the log of the age-specific selectivities to sum to zero. This method was used to reduce correlations among selectivity parameters. Estimation models contained three methods of estimating catchability: white noise, random walk, and no catchability (directly estimating fishing mortality). The first estimation model allowed $\log_e$ fishery catchability to vary with white noise about a constant mean (eq. T3.2.6). The second estimation model allowed $\log_e$ fishery catchability to vary according to a random walk (eq. T3.2.16). In my third estimation model, I estimated the fishing mortality rate for fully selected age classes as a parameter, and then applied the fishery selectivity to calculate age-specific fishing mortality rates (eq. T3.2.17). This method does not use fishery effort as a data source. The estimation models also predicted proportions of fishery and survey catch-at-age.

Model Fitting and Convergence

I fit the models using a Bayesian approach as implemented in AD Model Builder version 6.0.2 (Otter Research Ltd. 2000). The objective function was the sum of the likelihood components and priors. Each component was the negative of the log-likelihood for a single data source or an informative prior related to time-varying catchability (eq. T3.3.1). My estimation models assumed lognormal distributions of errors for total catch for the fishery (eq. T3.3.2) and survey CPE (eq. T3.3.3) and multinomial distributions for age compositions of the fishery (eq. T3.3.4) and the survey (eq. T3.3.5; Fournier and Archibald 1982). Effective sample sizes and CVs of the fishery

and survey catch and age compositions were set to their true values from the generating models.

For estimation models that used fishery effort as a data source, fishing mortality was an explicit function of effort and catch was linked to abundance and fishery effort by estimating the catchability coefficient. I assumed lognormal deviations for catchability in the white noise (eq. T3.3.6), and random walk (eq. T3.3.7) estimation models. The standard deviation for the white noise and random walk catchability deviations (on the $\log_e$ scale) was assumed known at 0.2, which was approximately equal to the expected standard deviation in the data-generating models. This component in the objective function is a prior and penalizes large deviations from mean catchability (for the white noise model) or large year-to-year deviations (in the random walk model). Note that priors ℓ_5 contained the constants for the likelihood function so that the priors were comparable to compare approximate Bayes factors. I placed uninformative uniform priors on common parameters among models and these priors were the same in each model.

The AD Model Builder implementation of Markov Chain Monte Carlo (MCMC) includes first estimating the maximum likelihood parameter estimates and asymptotic variance-covariance matrix, then using the estimated parameters as starting values for the MCMC chain. The Metropolis-Hastings algorithm sampled from a scaled multivariate normal distribution with variances and covariances proportional to the asymptotic variance covariance matrix. I ran the MCMC chain for each model for 5,000,000 cycles and saved values from every 100^{th} cycle. To estimate the precision of the DIC estimates, I estimated the variance of a shorter chain (as a minimum estimate for my cases) using

the "brute force" method of Zhu and Carlin (2000), which involves running many parallel MCMC chains, estimating DIC, and then estimating the variance of DIC from the parallel chain estimates. The MCMC chains were divided into subchains of 500,000 cycles to estimate the variance of DIC for a chain of that length. I dropped the initial 100,000 cycles of each chain as a burn in period, which reduces the effect of starting values on the MCMC estimates (Gelman et al. 2004). In some cases, the models did not converge to a stable mixing distribution for at least 1,000,000 cycles. In these cases, I used a burn in period of 1,500,000 cycles. I then estimated the variance of DIC estimates from the ten subsamples (seven in the cases with long burn in periods) for each chain. If the MCMC chain has converged to a stable mixing distribution, this method should provide the same result as running ten independent chains.

DIC Calculations

DIC, like other information-theoretic information criteria, trades off a measure of model fit (estimated deviance) and a measure of model complexity (effective number of parameters; Spiegelhalter et al. 2002).

$$DIC = \overline{D} + p_D$$

The average deviance, $\overline{D}$, for model j is an estimate of model adequacy and is estimated by

$$\overline{D}_j = \frac{1}{C} \sum_{c=1}^{C} -2\log_e p(data \mid \boldsymbol{\theta}_c)$$

where C is the number of MCMC cycles saved minus the burn in, and $\log_e p(data \mid \boldsymbol{\theta}_c)$ was the natural logarithm of the likelihood function (Spiegelhalter et al. 2002). Like with AIC and BIC, smaller DIC values indicate better models. I estimated the effective

number of parameters as the difference between the average deviance and the deviance evaluated at the maximum likelihood parameter estimates,

$$p_D = \overline{D} - D(\boldsymbol{\theta}_{ML}).$$

Normally, the effective number of parameters is estimated as the difference between the mean deviance and the deviance evaluated at the mean of the parameter vector, which is estimated by the mean parameters from the MCMC chain (Spiegelhalter et al. 2002). However, Spiegelhalter et al. (2002) noted that other measures of the central tendency, such as the mode or median of the parameters could be used. DIC differences calculated using the maximum likelihood estimates were usually within 0.1 DIC units of DIC differences calculated with the mean of the parameters from the MCMC chain. I also attempted a third method of estimating the effective number of parameters, which used ½ of the variance of the deviance chain values to approximate the effective number of parameters (Gelman et al. 2004),

$$p_D = \frac{1}{2(C-1)} \sum_{c=1}^{C} \left(D(\boldsymbol{\theta}_{c,j}) - \overline{D}_j \right)^2$$

This method performed poorly (large DIC variance) and almost always estimated more parameters for the model than the actual number of parameters in the models.

Approximate Bayes Factors

The probability that model M_i is the best in a set of candidate models can be approximated by

$$p(M_i) = \frac{p_i(data \mid \boldsymbol{\theta}_{ML}) p_i(\boldsymbol{\theta}_{ML} \mid M_i)}{\sum_i p_i(data \mid \boldsymbol{\theta}_{ML}) p_i(\boldsymbol{\theta}_{ML} \mid M_i)},$$

where $p_i(data \mid \boldsymbol{\theta}_{ML})$ is the likelihood evaluated at the maximum likelihood estimates of the parameters, and $p_i(\boldsymbol{\theta}_{ML} \mid M_i)$ is the prior for the parameters conditional on model i (Hilborn and Mangle 1997). Throughout the rest of the paper, this method for estimating the posterior probability that a model is the best in the set of candidate models is referred to as approximate Bayes factors. However, the term "approximate Bayes factors" in the model selection literature usually refers to using BIC differences to approximate Bayes factors (Kass and Raftery 1995).

Evaluation of Estimation Model Performance

I determined how often the correct structural model was selected, even though there was not a truly correct model in the scenario with a linear increase in catchability or in the uninformative effort scenario. In the white noise case, the white noise estimation model was correct. In the linear increase case, the random walk model was considered the correct model because it tended to perform better than other models in this scenario (chapter 2) and because it is designed to allow for gradual changes. In the case with uninformative effort data, the model that ignored fishery effort data was considered the correct model.

In stock assessments, estimated quantities in the last year are often most important for forecasting and management. Therefore, I evaluated estimation model performance by calculating the relative error (RE) of estimated biomass and average fishing mortality (for ages 4-8) in the last year.

$$RE = \frac{estimated - true}{true}$$

I evaluated systematic over or under estimation using the mean of the relative error (MRE). I also calculated the mean square relative error (MSE), which summarizes the variance and bias of model predictions. If the bias of the estimates is zero, MSE equals the variance of the estimator.

Results

Most of the MCMC chains appeared to have converged to their stable mixing distribution within 10,000 cycles. However, in several cases, the MCMC routine required nearly 1,500,000 cycles as a burn in period. The estimated standard deviation for DIC from a chain of 500,000 cycles was about 0.3. This indicates that, for a chain length of 500,000 cycles, DIC differences less than one are probably not important. For the 5,000,000 cycle chains, effective sample sizes were usually greater than 19,000 (from an actual sample size of 49,000 saved cycles) and DIC estimates should have lower standard deviations than from a substantially smaller chain.

Estimates of the effective number of parameters, p_D, were generally less than the actual number of estimated parameters, 52. The effective number of parameters for the estimation model with random walk catchability was the lowest with a mean of 47.7 and a range of 47.0-48.4. The estimation model with white noise catchability had the second fewest effective parameters with a mean of 48.8 (range 47.4-50.0). The estimation model that freely estimated fishing mortality for each year had the most effective parameters with a mean of 52.4 (range 51.4-54.1), which was quite close to the true number of estimated parameters.

DIC usually selected the correct model. However, DIC differences between the best model and the other models were usually less than seven, except in the

uninformative fishery effort scenario, indicating that the evidence was not overpoweringly in favor of the model with the lowest DIC (Figure 3.3). In the white noise catchability case, DIC selected the white noise model (correct model) eight out of ten times. In the two times when the white noise estimation model was not selected, the white noise catchability model and the random walk catchability model were quite close in terms of DIC scores (<0.7). In the linear increase catchability scenario, the random walk model was selected in nine out of ten cases. In the only case where DIC did not choose the correct model, the white noise estimation model was chosen and the difference in DIC scores was less than 1.0. In the uninformative effort scenario, the model that ignored fishery effort was always selected.

Model selection using approximate Bayes factors performed somewhat differently than DIC model selection and always selected the correct model in the case of white noise catchability (between 80-99% probability). However, approximate Bayes factors only selected the correct model six out of ten times for scenarios with a linear increase in catchability or uninformative fishery effort data. In the scenario where effort data were uninformative, approximate Bayes factors selected the white noise model twice and random walk model twice. In the scenario where catchability increased linearly, approximate Bayes factors did not choose any model strongly; posterior model probabilities were between 55 and 91% for the best model, and only the white noise model was selected in cases where the random walk model was not. The posterior model probabilities for the estimation model that ignored fishery effort were always less than 0.1% in both scenarios with informative effort data.

In general, using Bayesian model selection helped to choose relatively accurate models. Models selected using either DIC or approximate Bayes factors had smaller MSEs than always using any single model (table 3.3, figure 3.4). DIC model selection slightly outperformed approximate Bayes factors, but the difference was probably not significant because of the small sample size.

Discussion

In general, DIC and approximate Bayes factor model selection produced better point estimates of biomass and fishing mortality in the last year on average than relying on any single model. However, the best DIC or approximate Bayes factor model did not always produce the best estimates of biomass and fishing mortality rates in the last year. Indeed, DIC and approximate Bayes factors only selected the model with the lowest relative errors in fishing mortality or biomass in the last year between 7% (for DIC and fishing mortality) and 14% (approximate Bayes factors and biomass) of the time. Helu et al. (2000) also found that incorrect models often produced more accurate estimates of biomass in the last year than the structurally correct model in their study of AIC and BIC model selection for SCA models.

DIC model selection seems to perform well in cases where increased model complexity is warranted, but may not perform as well in determining when less complexity is warranted. Kizilkaya and Tempelman (2005) found that DIC strongly selected their model with heteroskedastic residual variances when residual variances were heteroskedastic, but did not strongly select the simpler model when variances were homoskedastic in linear mixed models and generalized linear mixed models. This is similar to my results where DIC fairly strongly selected the model that ignored fishery

effort data (the most complex model) in cases where fishery effort data were uninformative, but did not strongly select the models with fewer effective parameters when fishery effort data were informative. Indeed, Spiegelhalter et al. (2002) and van der Linde (2005) suggested that DIC does not provide a large enough penalty for model complexity for models with exponential family likelihoods. This family of distributions includes the normal and multinomial distributions that I used in the objective functions of my estimation models. However, increasing the penalty term for DIC would increase selection of simpler models in cases where a more complex model may be warranted.

Although MCMC methods can be quite time-consuming, calculating DIC should not be prohibitive in terms of time, given current levels of computer speed. In general, estimation models took about 1.5-2 hours to run 5,000,000 cycles on a computer with 2.8 gHz processors (Intel Xeon). These times are probably overestimates because I ran these models longer than was necessary (in most cases) to ensure convergence and to estimate the variance of DIC estimates for shorter chains. However, models that are structurally more complex or have more data (i.e., more years or age classes) will require longer run-times.

In most cases, model averaging provides superior predictive performance than using only the best model selected by DIC (or some other method) because estimates from a single model ignore uncertainty in model selection (Hoeting et al. 1999; Burnham and Anderson 2002; Burnham and Anderson 2004 and references therein). Therefore, I calculated model average estimates of biomass in the last year with the approximate Bayesian posterior model probabilities and posterior model probabilities derived from DIC differences (by adapting the method of Burnham and Anderson (2002) for AIC).

Both methods of model averaging had slightly larger MREs (about 1%) and slightly smaller MSEs (0.1-0.3%) than using only the "best" model. Differences in performance between the best model and the model average were probably slight because the best models were the same as or quite similar to the data-generating models. However, in real world applications it is unlikely that the estimation models will be as similar to the data-generating reality as was the case in this study. Therefore, model average estimates may provide a larger increase in performance than in this study. Interestingly, using DIC differences to estimate model probabilities and average model results seemed to perform reasonably well, although Spiegelhalter et al. (2002) describe this as an area requiring more research.

Certainly DIC and approximate Bayes factors are not exhaustive tools for model selection. Factors such as model plausibility, sensitivity, and examination of residual patterns should also be considered when choosing among models. However, DIC does show some promise for helping select among stock assessment models even when models are quite similar.

References

Aitkin, M. 1991. Posterior Bayes factors. Journal of the Royal Statistical Society Series B 53:111-142.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki eds. Second International Symposium on Information Theory 267-81. Budapest: Akademiai Kiado.

Barry, S. C., S. P. Brooks, E. A. Catchpole, and B. J. T. Morgan. 2003. The analysis of ring recovery data using random effects. Biometrics 59:54-65.

Berg, A., R. Meyer, and J. Yu. 2004. Deviance information criterion for comparing stochastic volatility models. Journal of Business & Economic Statistics 22:107-120.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Springer-Verlag, New York.

Burnham, K. P., and D. R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods & Research 33:261-304.

Cardoso, F. F., and R. J. Tempelman. 2003. Bayesian inference on genetic merit under uncertain paternity. Genetics Selection Evolution 35:469-487.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences Sci. 42: 815-824.

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39: 1195-1207.

Fournier, D. A., J. Hampton, and J. R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences 55: 2105-2116.

Francis, R. I. C. C., R. J. Hurst, and J. A. Renwick. 2003. Quantifying annual variation in catchability for commercial and research fishing. Fish. Bull. 101: 293-304.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis, 2nd edition. Chapman and Hall, Boca Raton.

Han, C., and B. P. Carlin. 2001. Markov Chain Monte Carlo methods for computing Bayes factors: a comparative review. Journal of the American Statistical Association 96:1122-1132.

Helu S. L., D. B. Sampson, and Y. Yin. 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Canadian Journal of Fisheries and Aquatic Sciences 57:1784-1793.

Hilborn, R., and M. Mangle. 1997. The ecological detective. Princeton University Press, Princeton.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. Bayesian model averaging: a tutorial. Statistical Science 4:382-417.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773-795.

Kizilkaya, K., and R. J. Tempelman. 2003. Cumulative *t*-link threshold models for genetic analysis of calving ease scores. Genetics Selection Evolution 35:489-512.

Kizilkaya, K., and R. J. Tempelman. 2005. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. Genetics Selection Evolution 37:31-56.

Lavine, M., and M. J. Schervish. Bayes factors: what they are and what they are not. The American Statistician 53:119-122.

McAllister, M, and C. Kirchner. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. Bulletin of Marine Science 70: 499-540.

Otter Research Limited. 2000. An introduction to AD Model Builder version 4 for use in nonlinear modeling and statistics. Otter Research Ltd., Sidney B.C.

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

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

Spiegelhalter, D.J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 64:583-639.

van der Linde, A. 2005. DIC in variable selection. Statistica Neerlandica 59:45-56.

Table 3.1. Symbols and descriptions of variables for data-generating and estimation models.

Symbol	Description	Value (if needed in the data-generating model)
$\overline{R}$	Average recruitment	1,000,000
$N_{y,a}$	Abundance by age and year	
B_y	Biomass	
$Z_{y,a}$	Total instantaneous mortality rate by age and year	
$F_{y,a}$	Instantaneous fishing mortality rate by age and year	
M	Instantaneous natural mortality rate	0.25
$s_{a,f}$	Fishery age-specific selectivity	See figure 3.2
$s_{a,s}$	Survey age-specific selectivity	See figure 3.2
E_y	Fishery effort	See figure 3.1
$q_{y,f}$	Fishery catchability	
$\tilde{E}_y$	Observed fishery effort	
q_s	Survey catchability	0.0001
$\overline{q}_f$	Mean fishery catchability	0.05

$C_{y,a}$	Expected fishery catch-at-age	
$I_{y,a}$	Expected survey catch-at-age	
$\tilde{C}_y$	Observed total fishery catch	
$\tilde{I}_y$	Observed total survey catch	
$u_{y,a,f}$	Proportion of catch-at-age in fishery	
$u_{y,a,s}$	Proportion of catch-at-age in survey	
w_a	Mean weight at age	0.16, 0.45, 0.82, 1.2, 1.55, 1.86, 2.11, 2.3
δ_y	Deviations for white noise catchability	
ε_y	Deviations for linear increase catchability	
ω_y	Deviations for random walk catchability	
a, b	Parameters for linear increase in catchability	0.032, 0.00225
f_y	Fishing intensity by year	
σ_γ	Standard deviation for $\log_e$ recruitment variation	1.0
σ_τ	Standard deviation for $\log_e$ fishery measurement error	0.1
σ_υ	Standard deviation for $\log_e$ of survey	0.2-0.8

	measurement error	
σ_δ	Standard deviation for $\log_e$ catchability deviations for white noise	0.2
σ_ε	Standard deviation for $\log_e$ catchability deviations	0.05
σ_δ	Standard deviation for $\log_e$ random walk catchability deviations	0.2

Table 3.2. Data-generating and estimation model equations.

	Population model equations	Application
(T3.2.1)	$N_{1,a} = \bar{R}e^{-\sum_{a=1}^{a-1} Z_{1,a} + \gamma_a}; \gamma \sim N\left(0, \sigma_\gamma^2\right)$	Generation
(T3.2.2a)	$N_{y+1,a+1} = N_{y,a}e^{-Z_{y,a}}$	Both
(T3.2.2b)	$N_{y+1,8} = N_{y,7}e^{-Z_{y,7}} + N_{y,8}e^{-Z_{y,8}}$	Both
(T3.2.3)	$B_y = \sum_a N_{y,a} w_a$	Both
(T3.2.4)	$Z_{y,a} = M + F_{y,a}$	Both
(T3.2.5)	$F_{y,a} = q_y E_y s_a$	Both
	Catchability model equations	
(T3.2.6)	*White noise*	Both
	$\log_e q_{y,f} = \log_e \bar{q}_f + \delta_y; \delta_y \sim N\left(0, \sigma_\delta^2\right)$	
(T3.2.9)	*Linear increase*	Generation
	$q_{y,f} = a + b(y) + \varepsilon_y; \varepsilon_y \sim N\left(0, \sigma_\varepsilon^2\right)$	
(T3.2.11)	*Random walk*	Estimation
	$\log_e q_{y+1,f} = \log_e q_{y,f} + \omega_y; \omega_y \sim N\left(0, \sigma_\omega^2\right)$	
(T3.2.12)	*Freely estimate f_y (ignore fishery effort)*	Estimation

$$F_{y,a} = f_y s_{a,f}$$

Observation model equations

(T3.2.13) $$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}}(1 - e^{-Z_{y,a}})N_{y,a}$$ Both

(T3.2.14) $$\tilde{C}_y = e^{\tau_y} \sum_a C_{y,a}; \tau_y \sim N(0, \sigma_\tau)$$ Both

(T3.2.15) $$I_{y,a} = q_s s_a N_{y,a}$$ Both

(T3.2.16) $$\tilde{I}_y = e^{\upsilon_y} \sum_a I_{y,a}; \upsilon_y \sim N(0, \sigma_\upsilon)$$ Both

Table 3.3. Objective function equations for statistical catch-at-age analysis simulation study.

(T3.3.1)	$L = \sum_i \ell_i$	Objective function
(T3.3.2)	$\ell_1 = \frac{1}{2\sigma_\tau^2} \sum_y \left(\log_e(\tilde{C}_y) - \log_e(\hat{C}_y) \right)^2$	Fishery catch
(T3.3.3)	$\ell_2 = \frac{1}{2\sigma_\upsilon^2} \sum_y \left(\log_e(\tilde{I}_y) - \log_e(\hat{I}_y) \right)^2$	Survey catch-per-effort
(T3.3.4)	$\ell_3 = -n_f \sum_y \sum_a u_{y,a,f} \log_e\left(\hat{u}_{y,a,f} \right)$	Proportion at age in the fishery catch
(T3.3.5)	$\ell_4 = -n_s \sum_y \sum_a u_{y,a,s} \log_e\left(\hat{u}_{y,a,s} \right)$	Proportion at age in the survey catch
(T3.3.6)	$\ell_5 = -n_y \log_e \sqrt{2\pi}\sigma_q + \frac{1}{2\sigma_q^2} \sum_y \left(\hat{\delta}_y \right)^2$	White noise catchability
(T3.3.7)	$\ell_5 = -n_y \log_e \sqrt{2\pi}\sigma_q + \frac{1}{2\sigma_q^2} \sum_y \left(\hat{\omega}_y \right)^2$	Random walk catchability
(T3.3.8)	$\ell_5 = -n_y \log_e\left(\frac{1}{20} \right)$	Freely estimate F

Table 3.4. Mean relative error (MRE) and mean square error (MSE) of models selected using deviance information criterion (DIC), approximate Bayes factors (ABF), only white noise catchability estimation model (WN), only random walk catchability estimation model (RW), and only using the estimation model that estimated fishing mortality for each year independent of effort (FF).

	DIC	ABF	WN	RW	FF
Biomass MRE	0.182	0.189	0.279	0.203	0.246
Biomass MSE	0.077	0.087	0.151	0.092	0.159
Fishing Mortality MRE	-0.028	-0.037	-0.111	-0.016	0.005
Fishing Mortality MSE	0.046	0.056	0.058	0.066	0.080

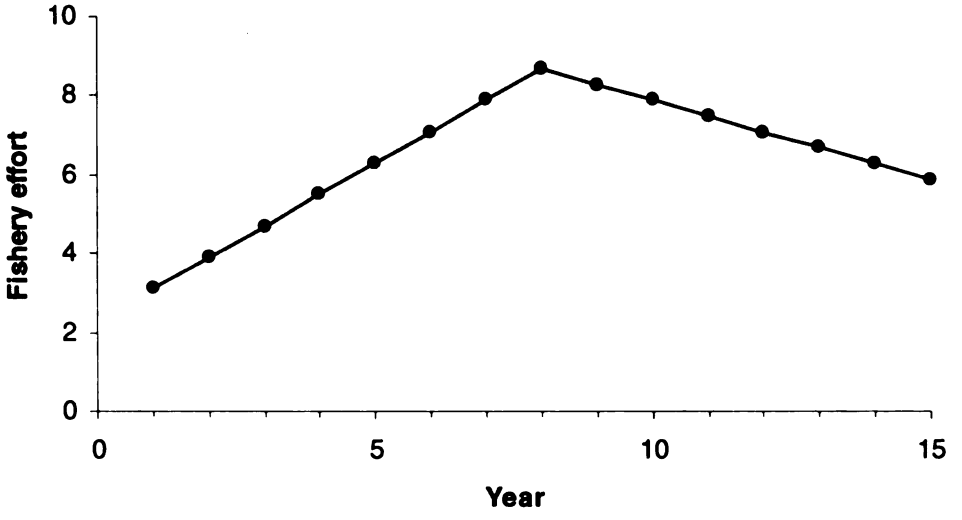

Figure 3.1. Baseline effort series used in data-generating models.

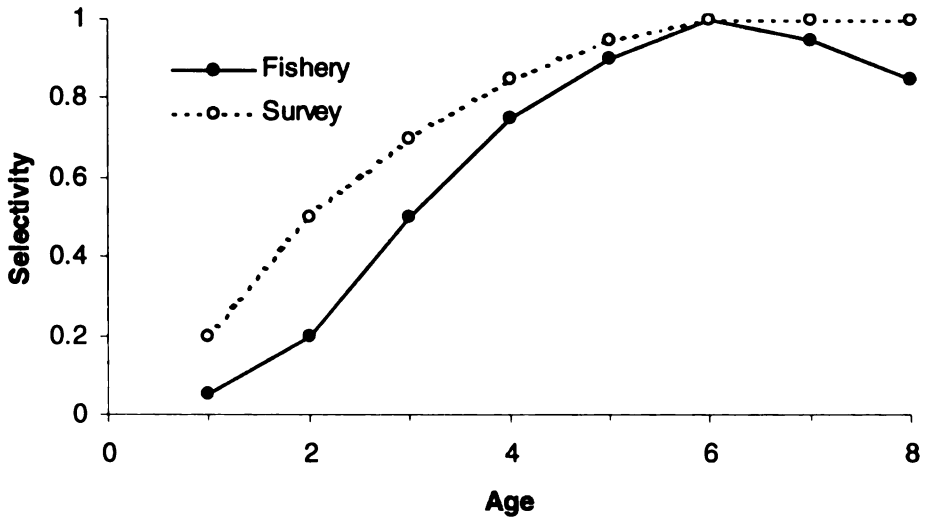

Figure 3.2. Fishery and survey selectivity patterns used in data-generating models.

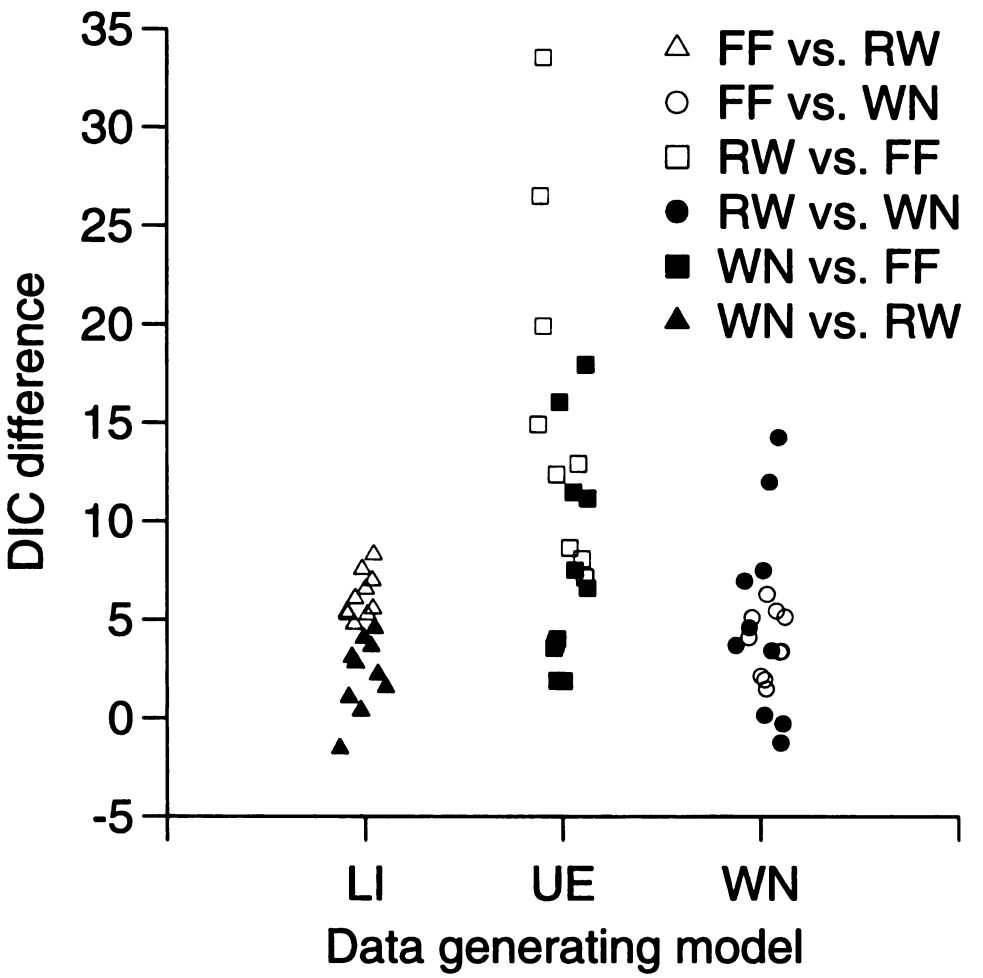

Figure 3.3. Deviance Information Criterion (DIC) differences among models. Differences from the best model for each data set are shown. Data-generating models are indicated by WN for white noise catchability, LI for linear increase in catchability, and UE for the case where observed effort data were uninformative. Estimation model comparisons are indicated by X vs. Y (legend), where Y is the hypothetical best estimation model for the scenario. Positive DIC differences indicate that the model Y is better than model X. Points are randomly jittered to reduce overlap.

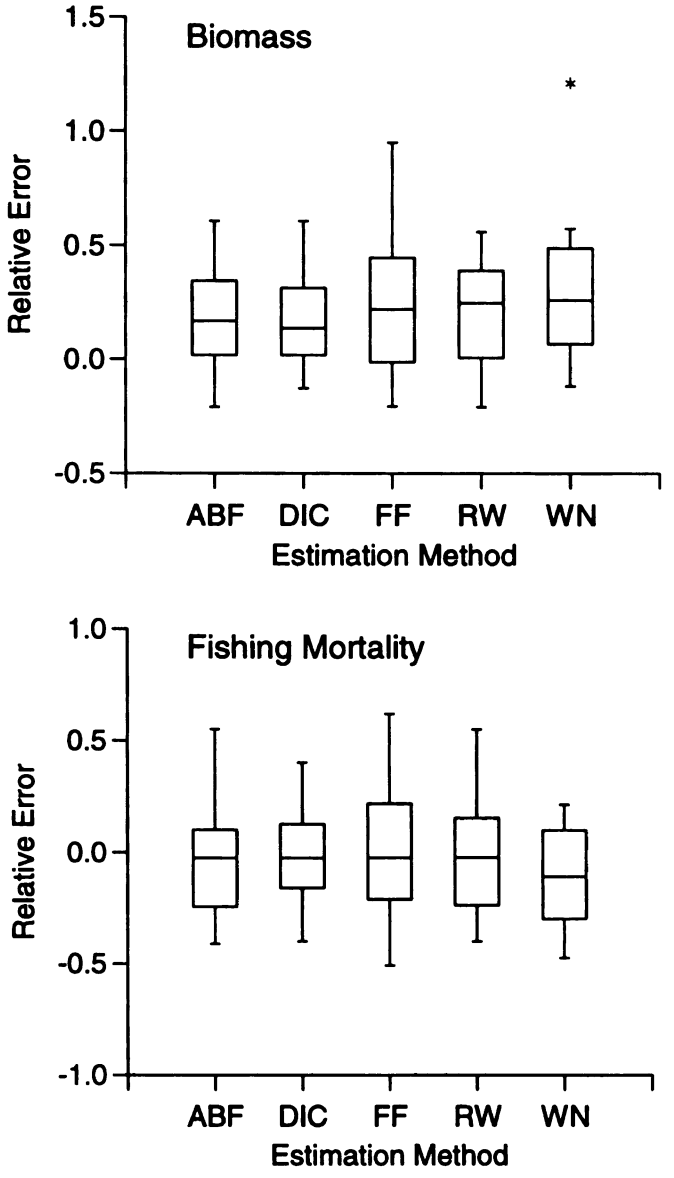

Figure 3.4. Box plots of relative error of estimates of biomass and average fishing mortality in year 15. The middle line indicates the median, the box indicates the interquartile range, and the whiskers indicate the 95% quantile range. Estimation methods are indicated by ABF for approximate Bayes factors, DIC for deviance information criterion, FF for the estimation model that freely estimated F, RW for the estimation model with random walk catchability, and WN for the estimation model with white noise catchability.

APPENDIX A

Appendix A describes the yellow perch assessment models and additional results for chapter 1.

Description of Yellow Perch Models

The population submodel predicted how yellow perch numbers-at-age and size-at-age changed over time, while the observation submodel predicted observed quantities given the predicted dynamics. Symbols used in the population and observation submodels are in Table A.1, and equations for these submodels are in Table A.2. We used the posterior likelihood to determine the best fit parameters.

Population Submodel

Total recruitment (defined as age-2 numbers) at the start of each year was estimated as a free parameter, and the sex ratio at recruitment was assumed to be 1:1 (eq. A.2.1). Numbers-at-ages 3 and 4 for each sex in the first year (1986) were also estimated as parameters. Numbers at ages 5-9+ in 1986 were calculated based on an assumption that each of those cohorts had the same abundance at age-4 as was estimated for age-4 in 1986 and suffered an estimated mortality rate that was sex specific (Wisconsin) or the same for both sexes (Illinois) (eq. A.2.2 and eq. A.2.3). We used this approach because sample sizes for ages five and above were low and these cohorts were not observed for many subsequent years. For Illinois we used a common mortality parameter for both sexes because sexes were aggregated in the Illinois survey data for 1986-1988. These assumptions about numbers-at-age in the first year have a relatively small effect on model estimates, because there were few old yellow perch in 1986.

Abundance-at-age of these cohorts were then tracked over time by applying age- and sex-specific mortality rates (eq A.2.10). Biomass was simply the product of the number of fish in a given length bin and their length-specific weight summed over sexes, ages, and lengths. Spawning stock biomass (SSB) was calculated using only females and a time-invariant maturity schedule based on length, which we estimated by fitting a logistic function to maturity-at-length data from Indiana waters of Lake Michigan (Ball State University, unpublished data) outside the model fitting process.

Total mortality rate for a given age and sex was the sum of the natural mortality rate and the age-, sex-, and year-specific fishing mortality rates for the two fisheries (recreational and commercial) (eq. A.2.5). Fishing mortality rates at age for a sex were calculated as a weighted average of the length specific fishing mortality rates, with weights equal to the proportion of fish that were a given age, sex, and length (eq. A.2.6).

The age specific rates were calculated from length specific ones. For each fishery, fishing mortality rates for a given length bin of yellow perch for the commercial and recreational fisheries was the product of catchability, effort, and selectivity, and the log of catchability followed a random walk (eq. A.2.7) and therefore was year-specific for each fishery. We modeled selectivity as constant functions of length, based on the midpoint for each length bin. Note that the fishing process influences fish in the same length in the same way, irrespective of their sex or age. We used a double logistic function to model the dome-shaped selectivity pattern (Quinn and Deriso 1999) for the commercial gill net fisheries (Kraft and Johnson 1992) and for the Illinois recreational fishery during 1997-2000 when a slot limit was in effect (eq. A.2.8). For the Illinois and

Wisconsin recreational fisheries (except for the Illinois fishery during 1997-2000), we modeled the selectivity pattern with an asymptotic logistic function (eq. A.2.9).

Growth was modeled using a stochastic von Bertalanffy growth model, where the parameters were allowed to vary over time (Szalai et al. 2003). For 1986, mean length at age (for the beginning of the year) was calculated assuming these fish had lived under constant growth conditions with all cohorts starting with mean length-at-age 2 as in 1986, and experiencing constant L_∞ and K pre-1986 values (eq. A.2.4). Mean length-at-age 2 was equal for males and females, but changed over time with a random walk (eq. A.2.14). For years after 1986, mean length-at-ages 3-8 were equal to the mean from the previous age and year plus the increments from the von Bertalanffy model (eq. A.2.11). The same model was used to estimate the mean length for the aggregated age-9 and older group, but this was based on a weighted average of growth expected for age-8 and age-9 fish, with weights determined by the contribution of the two ages to this group in the next year (eq. A.2.12). To estimate mean length at age in the fall, fish were grown for $8/10^{th}$ of the year (eq. A.2.13). Like length at-age-2, asymptotic mean length and the Brody growth coefficient also changed over time with with a random walk (eq. A.2.14), which were modeled separately for males and females. The modeled length composition for a given age was normally distributed with a mean predicted by the von Bertalanffy equation. The proportion in each one cm length bin was calculated from the corresponding standard normal cumulative distribution function (Φ) (eq. A.2.15). The standard deviation of each normal distribution was the product of the mean length-at-age and an age and sex-specific coefficient of variation (CV). We used a hockey stick function to describe how the CV decreased with increasing age for ages 2 to 5, and then remained constant after

age 5. This pattern of decreasing variation in length-at-age with increasing age is common to many teleost fishes (Bowker 1995), and the CVs we used were based on observed variation of length-at-age (WDNR, unpublished data).

Observation Submodel

Catch-at-length (in numbers) for the commercial and recreational fisheries was calculated with the Baranov catch equation (eq. A.2.16 and eq A.2.17). Commercial catch calculations used numbers-at-length calculated from numbers-at-age reassigned to length categories based on the fall distribution of length-at-age whereas recreational catch calculations were based on spring length distributions. This is an approximation that is intended to account for the fact that the two fisheries are prosecuted at different times during the year (commercial fishery centered in the fall, recreational fishery in the spring and summer), that fish grow during the year, and that fishery selectivity is length-based. Total catch in numbers was simply the sum over length bins of catch-at-length. Commercial yield was calculated by multiplying catch-at-length by weight-at-length (from fall lengths) and summing over length categories.

Catch per effort (CPE) at-length and sex for the survey were calculated as the product of catchability, selectivity, and numbers-at-length (eq. A2.18). Catchability of the survey was sex-specific for Illinois, but the same for males and females in Wisconsin, because of differences in survey design between the two surveys. We modeled survey selectivity using the same logistic function of length used for recreational fishery selectivity (eq, A.2.9). Total CPE by sex for the survey was the sum over lengths of the length-specific survey CPEs. CPE at-age and sex for the survey was calculated as the product of the survey catchability, numbers at age and sex, and the age- and sex-specific

survey selectivity (given by a weighted sum of length specific selectivity values) (eq. A.2.19). For each year proportions of the catch for the fisheries and the survey falling into each length bin and proportions of the survey catch for each age were calculated for comparison with observed proportions.

Model predictions of mean length-at-age seen in the survey were calculated by taking the modeled population length distribution at age and adjusting it for the estimated survey selectivity (eq. A.2.20).

Likelihood Equations

Our objective function was the posterior negative log-likelihood, $\Delta = \sum_i \ell_i$, with individual negative log-likelihood components and priors (dropping some ignored constants) given by ℓ_i. Our point estimates minimized this function. One set of components had the general form:

$$\ell_i = \frac{1}{2}\sum_j X_j^2 \quad \text{(A1)}$$

Where X_j is an assumed standard normal variate and j is an index distinguishing the terms being summed for the i^{th} component. These likelihood components were based on an assumed independent normal (mean length-at-age) or lognormal distribution (fishery total catch or survey total catch per unit effort) for deviations between observed quantities and model predictions or an informative normal prior distribution for random walk errors (for mean length-at-age 2, L_∞, K, and catchability for the commercial and recreational fisheries) and for two parameters of the Illinois commercial fishery selectivity function (Table A.3). We used an informative prior for two of the four Illinois commercial fishery

selectivity parameters because the observed length composition of the Illinois commercial catch contained relatively few measurements, and we based these priors on the point estimates and standard errors for of the same parameters from the Wisconsin model. Small constants were added to observed and predicted values (for the lognormal distributions) to reduce the influence of very small values (Hampton and Fournier 2001).

An additional set of components took the general form:

$$\ell_i = -\sum_k n_k \sum_y \sum_T u_{T,k} \log_e(\hat{u}_{T,k} + c) \quad \text{(A2)}$$

based on our assumption that multinomial distributions led to the observed proportions at length and age for all data sources for which there were observations. This included a component for the fishery length compositions, and components for the survey length and age compositions. The outer sum is over categories of data (k), which were fisheries 1 and 2 (for the fishery length compositions) and sexes (for survey age and length compositions), and the inner sum was over types (T) of fish within a category and year (lengths bins or ages). Small constants (c = 0.0001 for length compositions and c = 0.001 for age compositions) were added to likelihood functions to reduce the effect of small proportions during model fitting (Fournier and Archibald 1982).

For completeness we note that for parameters other than those with the normal priors described above, we assumed uniformly distributed priors on the scale they were estimated. These priors did not enter explicitly into the objective function because they were implemented by placing bounds on the allowed parameter range during estimation.

Table A.1. Symbols representing parameters, data, and calculated quantities for assessment models.

Parameter	Definition
	Indicator Variables
a	Age-class; 2-9+
y	Year; 1986-2001
l	Midpoint of each length bin; 8-38 cm
G	Sex; male or female
f	Fishery; commercial = 1, recreational = 2 or, survey = 3
	Estimated Parameters
R_y	Recruitments for each year
$N_{1986,a,G}$	Numbers at age in 1986 for ages 3 and 4
Z_{init_G}	Mortality rate for the final five age classes in the first year
q_f	Catchability
λ_f	Parameters for logistic and double logistic selectivity functions
$L_{\infty}{}_{y,G}$	Asymptotic length
$K_{y,G}$	Brody growth coefficient
$L_{y,2}$	Mean length-at-age 2
M	Rate of natural mortality time-, sex-, and age-invariant

δ_y Random walk deviations for mean length-at-age 2

$\gamma_{y,G}$ Random walk deviations for $L_{\infty_{y,G}}$

$\varpi_{y,G}$ Random walk deviations for $K_{y,G}$

$\varepsilon_{y,f}$ Random walk deviations for catchability

Calculated Quantities

$Z_{y,a,l,G}$ Total instantaneous mortality rate

$F_{y,a,l,G,f}$ Instantaneous rate of fishing mortality

$p_{y,a,l,G}$ Proportions-at-length for each age

$N_{y,a,l,G}$ Numbers-at-age, length in the beginning of the year, and sex in year y

$\dot{N}_{y,a,l,G}$ Numbers-at-age, length in the fall of the year, and sex in year y

$L_{y,a,G}$ Mean length-at-age in population in beginning of year

$\hat{\bar{L}}_{y,a,G}$ Model predicted mean length-at-age measured by survey

$\dot{L}_{y,a,G}$ Mean length-at-age in population in fall

s_f Selectivity

$\hat{I}_{y,a,l,G}$ Survey index of abundance

$\hat{u}_{y,a,l,G,f}$ Model prediction of proportions of catch-at-age, length, and sex

$\hat{C}_{y,l,f}$	Model prediction of catch
$\hat{W}_{y,l}$	Model predicted commercial yield (kg)

Likelihood Weighting Components

$n_{y,a,l,G}$	Sample size of fish aged for the mean length-at-age likelihood function and effective sample size for age and length compositions
σ_f	CV for fishery catches
σ_δ	Standard deviation for mean length-at-age 2 random walk deviations
$\sigma_{\gamma,G}$	Standard deviation for $L_{\infty_{y,G}}$ random walk deviations
$\sigma_{\varpi,G}$	Standard deviation for $K_{y,G}$ random walk deviations
$\sigma_{\varepsilon,f}$	Standard deviation for fishery catchability random walk deviations
σ_{λ_i}	Standard deviation for commercial selectivity prior for Illinois

Data

$\tilde{L}_{y,a,G}$	Observed mean length-at-age in the survey
$I_{y,a,l,G}$	Observed CPE in the survey
$u_{y,a,l,G,f}$	Observed proportions at age and length in the fisheries
$C_{y,l,f=2}$	Harvest (numbers) in the recreational fishery
$W_{y,l,f=1}$	Yield (kg) in the commercial fishery

$E_{y,f}$	Fishery effort
w_l	Weight-at-length
$\omega_{y,a,G}$	Number of fish aged by age year and sex
$\tilde{\lambda}_{i,f=1}$	Mean parameter for the prior of commercial selectivity function for Illinois
M	Instantaneous rate of natural mortality (age- and sex-independent)

Table A.2. Equations for population and observation submodels.

Population submodel

Recruitment, initial abundances at age, initial mean length at age

$$N_{y,a=2,G} = \frac{R_y}{2}. \tag{A.2.1}$$

$$N_{y=1986,a,G} = N_{y=1986,a=4,G} e^{-(a-4)Z_{init}} \quad G \text{ ; } a\text{>4, Wisconsin} \tag{A.2.2}$$

$$N_{y=1986,a,G} = N_{y=1986,a=4,G} e^{-(a-4)Z_{init}} \text{ ; } a\text{>4, Illinois} \tag{A.2.3}$$

$$L_{y=1986,a+1,G} = L_{y=1986,a,G} + \left(L_{\infty\, y=pre1986,G} - L_{y=1986,a,G} \right) \left(1 - e^{-K_{y=pre1986,G}} \right) \tag{A.2.4}$$

Mortality rates

$$Z_{y,a,G} = M + \sum_{f=1}^{2} F_{y,a,G,f} \tag{A.2.5}$$

$$F_{y,a,G,f} = \sum_{l} p_{y,a,l,G} F_{y,l,f} \qquad \sum_{l} p_{y,a,l,G} = 1 \tag{A.2.6}$$

$$F_{y,l,f} = q_{y,f} E_{y,f} s_{l,f} \qquad q_{y+1,f} = q_{y,f} e^{\varepsilon_{y,f}} \tag{A.2.7}$$

$$s_{l,f} = \left(\frac{1}{1+e^{-\left(\lambda_{1,f}(l) - \lambda_{3,f}\right)}} \right) \left(1 - \frac{1}{1+e^{-\left(\lambda_{2,f}(l) - \lambda_{4,f}\right)}} \right) \tag{A.2.8}$$

$$s_{l,f} = \frac{1}{1+e^{-\left(\lambda_{1,f}(l) - \lambda_{2,f}\right)}} \tag{A.2.9}$$

Population and length-at-age dynamics

$$N_{y+1,a+1,G} = N_{y,a,G} e^{-Z_{y,a,G}} \tag{A.2.10}$$

$$L_{y+1,a+1,G} = L_{y,a,G} + (L_{\infty}{}_{y,G} - L_{y,a,G})(1 - e^{-K_{y,G}}) \tag{A.2.11}$$

$$L_{y+1,a=9,G} = \frac{N_{y,a=8,G}\left(L_{y,a=8,G} + (L_{\infty}{}_{y,G} - L_{y,a=8,G})\left(1 - e^{-K_{y,G}}\right)\right)}{N_{y,a=8,G} + N_{y,a=9,G}} + \frac{N_{y,a=9,G}\left(L_{y,a=9,G} + \left(L_{\infty}{}_{y,G} - L_{y,a=9,G}\right)\left(1 - e^{-K_{y,G}}\right)\right)}{N_{y,a=8,G} + N_{y,a=9,G}} \tag{A.2.12}$$

$$\dot{L}_{y,a,G} = L_{y,a,G} + (L_{\infty}{}_{y,G} - L_{y,a,G})(1 - e^{-0.8K_{y,G}}) \tag{A.2.13}$$

$$L_{y+1,2} = L_{y,2} e^{\delta_y} \quad L_{\infty}{}_{y+1,G} = L_{\infty}{}_{y,G} e^{\gamma_{y,G}}$$

$$K_{y+1,G} = K_{y,G} e^{\varpi_{y,G}} \tag{A.2.14}$$

$$p_{y,a,l,G} = \Phi\left(\frac{(l+1) - L_{y,a,G}}{\sigma_{l}{}_{a,G}}\right) - \Phi\left(\frac{l - L_{y,a,G}}{\sigma_{l}{}_{a,G}}\right) \tag{A.2.15}$$

Observation submodel

$$\hat{C}_{y,l,f=1} = \frac{F_{y,l,f=1}}{Z_{y,l}}(1-e^{-Z_{y,l}})\sum_{G}\dot{N}_{y,l,G} \tag{A.2.16}$$

$$\hat{C}_{y,l,f=2} = \frac{F_{y,l,f=2}}{Z_{y,l}}(1-e^{-Z_{y,l}})\sum_{G}N_{y,l,G} \tag{A.2.17}$$

$$\hat{I}_{y,l,G} = q_{G,f=3}{}^{s}l,f=3{}^{N}y,l,G \tag{A.2.18}$$

$$\hat{I}_{y,a,G} = q_{G,f=3}{}^{N}y,a,G\sum_{l}{}^{s}l,f=3{}^{p}y,a,l,G \tag{A.2.19}$$

$$\hat{\hat{L}}_{y,a,G} = \frac{\sum_{l}{}^{s}l,f=3{}^{p}y,a,l,G^{(l)}}{\sum_{l}{}^{s}l,f=3} \tag{A.2.20}$$

Table A.3. Specification of terms for normal and lognormal negative log-likelihood components (see equation A1).

Standard normal variate	Squared variates summed over these indices
$(\tilde{L}_{y,a,G} - \hat{\tilde{L}}_{y,a,G}) / (\sigma_{y,a,G} / \sqrt{\omega_{y,a,G}})$	y, a, G
$(\log_e W_{y,f=1} - \log_e \hat{W}_{y,f=1}) / \sigma_{f=1}$	y
$(\log_e (C_{y,f=2}) - \log_e (\hat{C}_{y,f=2})) / \sigma_{f=2}$	y
$(\log_e (I_{y,G}) - \log_e (\hat{I}_{y,G})) / \sigma_{f=3}$	y, G
δ_y / σ_δ	y
$\gamma_{y,G} / \sigma_{\gamma,G}$ and $\varpi_{y,G} / \sigma_{\varpi,G}$	y, G
$\varepsilon_{y,f} / \sigma_{\varepsilon_f}$	y, f
$\left(\lambda_{j,f=1} - \tilde{\lambda}_{j,f=1}\right) \Big/ \sigma_{\lambda_j}$	$j, j<3$

Table A.4. Results of sensitivity analyses of changes of weights of data sources in the objective function for yellow perch catch-at-age models for Illinois and Wisconsin waters of southwestern Lake Michigan. Differences from baseline estimates are displayed as percentages. Baseline model estimates of abundance (N; 1000s), biomass (B; 1000 kg), mean fishing mortality for females age-4 and older ($\overline{F}_{4+}$ females), and mean fishing mortality for males age-4 and older ($\overline{F}_{4+}$ males) for 2002 are displayed for comparison. In two cases the model's parameter estimates failed to converge to values that minimized the objective function and these are denoted by NC.

Illinois	Baseline value	Adjustment factors	N	B	$\overline{F}_{4+}$ females	$\overline{F}_{4+}$ males
Baseline			4,790	818	0.058	0.025
Commercial yield	0.0025	5	-3.1	-3.1	3.3	3.2
Commercial yield	0.0025	0.2	0.7	0.7	-0.8	-0.7
Commercial catchability	0.06	5	-16.1	-16.6	12.8	5.8
Commercial catchability	0.06	0.2	41.7	43.1	-30.5	-28.7
Commercial length	32	5	-10.2	-10.9	12.8	5.8
Commercial length	32	0.2	-2.0	-0.7	0.4	5.0
Recreational harvest	0.01	5	2.3	2.0	-5.4	-6.7
Recreational harvest	0.01	0.2	-0.7	-0.6	3.4	4.1
Recreational catchability	0.06	5	3.5	2.9	-0.5	-2.7
Recreational catchability	0.06	0.2	-7.6	-6.2	3.4	9.5
Recreational length	367	5	-24.1	-24.4	34.1	7.6
Recreational length	367	0.2	17.6	11.5	-15.1	-12.1

Survey CPE, females	0.19	5	30.7	29.3	-22.8	-25.8
Survey CPE, females	0.19	0.2	-32.6	-31.0	45.8	53.0
Survey female ages	27	5	17.2	11.0	-8.3	-30.7
Survey female ages	27	0.2	-10.1	-7.2	7.2	20.5
Survey female lengths	61	5	25.1	17.0	-20.4	-24.8
Survey female lengths	61	0.2	-21.8	-17.1	23.2	40.1
Survey CPE, males	0.22	5	4.3	4.2	-4.0	-4.5
Survey CPE, males	0.22	0.2	-5.5	-5.3	5.4	6.4
Survey male ages	53	5	-9.7	-6.8	7.3	21.5
Survey male ages	53	0.2	-8.9	-5.9	6.2	14.0
Survey male lengths	58	5	-4.9	2.7	-2.8	15.3
Survey male lengths	58	0.2	-17.0	-13.9	16.5	30.5
Female L_∞	0.0006	5	3.2	2.1	-3.8	-2.8
Female L_∞	0.0006	0.2	-0.5	-0.1	2.7	-1.9
Female K	0.0006	5	5.3	4.5	-4.4	-6.4
Female K	0.0006	0.2	-1.4	-1.2	1.8	0.9
Male L_∞	0.0006	5	-2.9	-4.7	5.4	-8.6
Male L_∞	0.0006	0.2	-19.1	-15.2	19.6	37.8
Male K	0.0006	5	0.0	-0.9	1.0	-1.8
Male K	0.0006	0.2	-0.2	0.2	-0.2	0.5
Length-at-age 2	0.01	5	-3.2	-3.1	3.2	5.0
Length-at-age 2	0.01	0.2	6.1	5.7	-5.4	-7.4
M	0.37	1.2	-9.3	-11.5	17.2	6.6
M	0.37	0.8	12.8	16.3	-15.8	-9.3

Wisconsin						
Baseline			1,690	356	0.075	0.060
Commercial yield	0.0125	5	-3.7	-3.8	4.2	3.8
Commercial yield	0.0125	0.2	3.9	4.0	-4.0	-3.8
Commercial catchability	0.16	5	25.8	25.2	-20.3	-21.2
Commercial catchability	0.16	0.2	-22.8	-22.1	28.2	31.4
Commercial length	43	5	-7.9	-7.7	8.3	9.0
Commercial length	43	0.2	5.8	2.4	-0.8	-8.3
Recreational harvest	0.01	5	4.0	3.7	6.0	5.5
Recreational harvest	0.01	0.2	-2.0	-1.9	0.6	0.7
Recreational catchability	0.06	5	6.6	6.3	-7.8	-8.2
Recreational catchability	0.06	0.2	-16.1	-15.4	27.1	28.7
Recreational length	141	5	-1.4	3.2	-3.3	5.7
Recreational length	141	0.2	-7.7	-8.5	9.4	7.8
Survey CPE, females	1.06	5	4.0	3.8	-3.8	-3.9
Survey CPE, females	1.06	0.2	-4.2	-3.8	4.0	4.6
Survey female ages	31	5	13.6	10.7	-10.0	-7.9
Survey female ages	31	0.2	-17.9	-17.2	21.8	20.1
Survey female lengths	45	5	NC	NC	NC	NC
Survey female lengths	45	0.2	0.2	-0.1	0.6	-0.1
Survey CPE, males	0.92	5	25.0	24.6	-20.1	-20.6
Survey CPE, males	0.92	0.2	-41.3	-40.9	72.0	74.1
Survey male ages	50	5	NC	NC	NC	NC
Survey male ages	50	0.2	-33.3	-33.3	50.6	52.5

Survey male lengths	63	5	20.0	14.4	-11.9	-19.6
Survey male lengths	63	0.2	6.6	8.9	-8.4	-5.3
Female L_∞	0.0025	5	-1.2	-1.0	0.2	1.5
Female L_∞	0.0025	0.2	-0.4	-1.0	1.5	0.1
Female K	0.0025	5	-1.3	-2.5	0.9	1.7
Female K	0.0025	0.2	-0.4	0.3	0.6	0.2
Male L_∞	0.0025	5	13.7	13.0	-11.8	-13.1
Male L_∞	0.0025	0.2	-18.7	-18.3	23.2	24.8
Male K	0.0025	5	-2.1	-2.0	2.0	2.0
Male K	0.0025	0.2	1.9	1.8	-1.7	-1.9
Length-at-age 2	0.01	5	0.7	0.3	-1.2	0.0
Length-at-age 2	0.01	0.2	-5.7	-5.6	7.4	6.0
M	0.37	1.2	-34.3	-35.0	59.0	58.3
M	0.37	0.8	49.4	50.8	-36.1	-35.5

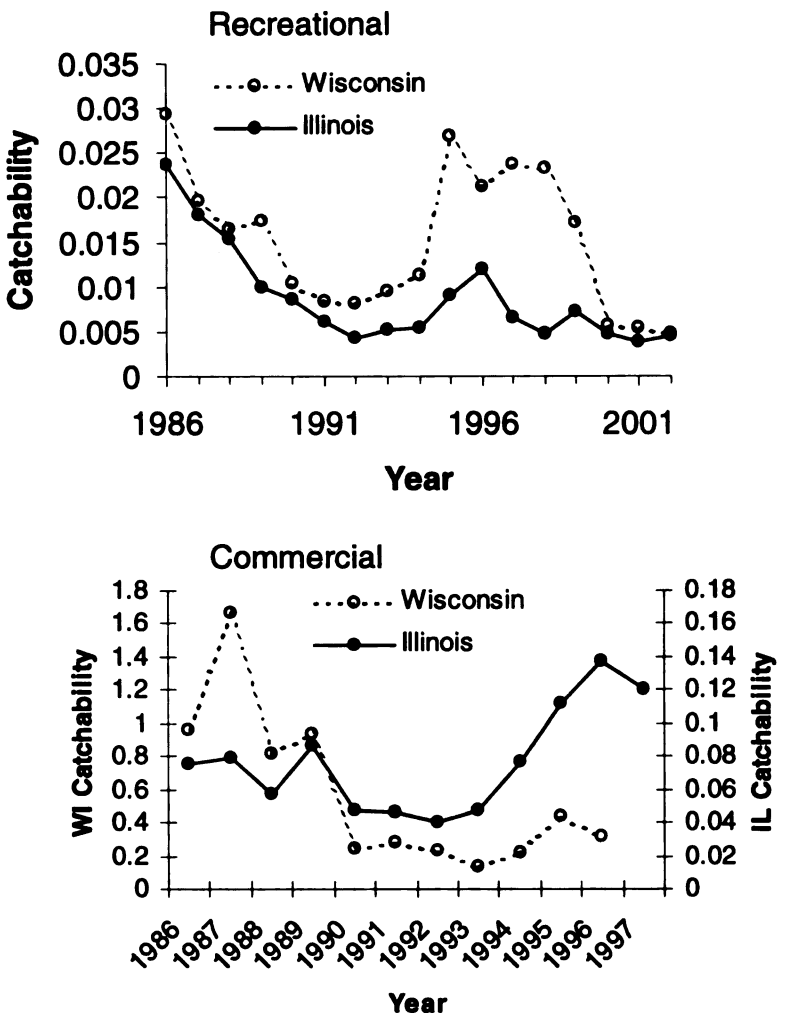

Figure A.1. Estimated catchability coefficients for Wisconsin and Illinois recreational and commercial fisheries in southwestern Lake Michigan during 1986-2002.

APPENDIX B

Appendix B describes the parameterization of estimation models used in chapter 3 to reduce correlations among parameters. The MCMC algorithm I used was very sensitive to parameter correlations greater than about 0.8. Under these conditions, the MCMC algorithm mixed very poorly and produced very "sticky" MCMC chains (i.e., chains with high autocorrelation). Therefore, I reparameterized aspects of the models to reduce these correlations. All parameters described below were estimated on the log scale. Two groups of parameters were highly correlated within each group: parameters that determined overall scale of population size, and selectivity parameters for the fishery and survey. Parameters that determine the overall scale of the population size included, in this case, mean recruitment, mean abundance at age in year 1, fishery catchability (or mean F in the model that ignored fishery effort data), and survey catchability. In order to minimize correlation among these parameters, I parameterized the model by estimating the $\log_e$ of mean recruitment and a deviation from this for each of these other "scale-setting" parameters. The other parameters that had high correlations were the selectivity at age for the fishery and the survey. To reduce these correlations, the models were parameterized to estimate deviations from a mean $\log_e$ selectivity that was forced to equal zero. This constraint serves to make the selectivity parameters identifiable and not confounded with the associated catchability (for fishery or survey), in the same way that the more usual approach of setting selectivity to 1.0 for a fully selected age (e.g., Fournier and Archibald 1982) does.