Hierarchical analysis of multiple noisy abundance indices

Paul B. Conn

PW8-RD08

Received: 7/19/2021

Hierarchical analysis of multiple noisy abundance indices

Paul B. Conn

Abstract: Fisheries analysts often rely on relative abundance indices for assessing stock status. However, trends in abundance can be difficult to discern when there are multiple indices or when correlation among indices is weak or negative. In this paper, I present a hierarchical framework for analyzing multiple, noisy indices with the goal of estimating a single time series of relative abundance. An implicit assumption is that each index is measuring the same quantity (relative abundance) but that each is subject to process error (attributable to variation in catchability, spatial distribution, etc.) in addition to an estimable level of within-survey variance (i.e., sampling error). I use simulation to explore estimator performance under a number of scenarios, including several that violate underlying assumptions. In general, the hierarchical approach produced estimators with reasonable properties. I illustrate the method with an analysis of seven fishery-dependent catch-per-unit-effort indices of Spanish mackerel (*Scomberomorus maculatus*) off the US Atlantic coast and provide several suggestions for how this approach can be used in practice.

Résumé : Les analystes des pêches se fient souvent sur les indices d'abondance relative pour évaluer le statut des stocks. Il peut cependant être difficile de discerner les tendances dans l'abondance lorsqu'il y a plusieurs indices ou lorsque la corrélation entre les indices est faible ou négative. Ce travail présente un cadre hiérarchique pour l'analyse d'indices multiples et contenant beaucoup de bruit dans le but d'estimer une seule série chronologique d'abondances relatives. Une présupposition implicite veut que chaque indice mesure la même entité quantifiable (l'abondance relative), bien que chacun soit sujet à une erreur de processus (attribuable à la variation de la capturabilité, la répartition spatiale, etc.) en plus d'avoir un niveau évaluable de variation dans les inventaires (c'est-à-dire une erreur d'échantillonnage). La simulation permet d'explorer la performance de l'estimateur dans divers scénarios, dont plusieurs qui violent les présuppositions sous-jacentes. En général, l'approche hiérarchique produit des estimateurs avec des propriétés acceptables. Une analyse de sept indices de CPUE dépendant de la pêche chez des thazards atlantiques (*Scomberomorus maculatus*) de la côte est des É.-U. illustre la démarche; plusieurs suggestions facilitent l'utilisation concrète de la méthode.

[Traduit par la Rédaction]

Introduction

Relative abundance indices are a critical component of most fisheries stock assessment models. A common but key assumption of these models is that each index is proportional to some measure of abundance or biomass. For instance, if abundance is of interest, an analyst might assume that $I_t = qN_t$, where I_t is the index value in year t, N_t is abundance in year t, and q is a time constant catchability coefficient. Unfortunately, catch-per-unit-effort (CPUE) time series are often deficient for making inferences about relative abundance because of fluctuations in q. For instance, catch rates can vary as a function of fishing vessel, wind speed, population density, spatial distribution of the population, and other biotic or abiotic factors (Byrne et al. 1981; Collie and Sissenwine 1983; Poulard and Trenkel 2007). The situation can be worse for fishery-dependent CPUE indices, where catch rates often remain high even when populations are declining (cf. Harley et al. 2001).

Given the importance of catchability changes for interpretation of indices, a focus of recent research has been to develop methods of index construction that relax the constant catchability assumption. One common approach is to use generalized linear models (GLMs) to control for covariates thought to influence catchability (e.g., Lo et al. 1992; Stefánsson 1996; Maunder and Punt 2004). Although this approach can control for measured quantities such as stochastic changes in environment, fishing vessels, or gear, etc., it cannot diagnose changes in q due to changes in spatial distribution of the target population, to a nonlinear relationship between catch rates and abundance, or even when there are trends in environmental covariates (cf. Nichols et al. 2009). Comparisons of interannual variation of indices with estimates of sampling variance (as computed from the design of a fishery-independent survey or as a by-product of model fitting) often indicate that there is overdispersion in index values that cannot adequately be explained by sample size alone (Byrne et al. 1981; Pennington 1985). When

Received 28 January 2009. Accepted 16 October 2009. Published on the NRC Research Press Web site at cjfas.nrc.ca on 11 December 2009.
J21027

Paper handled by Associate Editor Ray Hilborn.

P.B. Conn. National Marine Fisheries Service, Southeast Fisheries Science Center, NOAA Center for Coastal Fisheries and Habitat Research, 101 Pivers Island Road, Beaufort, NC 28516, USA (e-mail: Paul.Conn@noaa.gov).

Can. J. Fish. Aquat. Sci. **67**: 108–120 (2010) doi:10.1139/F09-175 Published by NRC Research Press

the magnitude of interannual changes in index values is consistently larger than that expected from knowledge of life history and theoretical variances, this suggests a residual, unexplained source of variation in index values (Stockhausen and Fogarty 2007). Several approaches have been advanced to deal with this unexplained variance. For instance, Pennington (1985) and Stockhausen and Fogarty (2007) suggested using time series methods to smooth indices. Several authors have extended this framework to incorporate relevant life history information or other data sources as auxiliary information within state–space models (Besbeas et al. 2002; Mazzetta et al. 2007). However, each of these approaches assumed that inference was based on one index.

In contrast, many assessments include multiple indices, which may provide additional information with which to estimate error components. For instance, in a recent assessment of Spanish mackerel (*Scomberomorus maculatus*) off the US Atlantic coast, a total of seven indices of abundance were computed from commercial and recreational fishery databases, few of which correlated well with each other (SouthEast Data, Assessment, and Review (SEDAR) 2008*a*). Although time series methods could be used to smooth each individual index, this would do little to resolve differences in population trend if there were shifts in migration patterns that involved “trade-offs” between fisheries. Indeed, commercial fishermen recently observed increased catch rates in some Spanish mackerel fisheries (e.g., commercial gillnet) and decreases in others (e.g., commercial handlines), which they attributed to shifts in wintering distributions (B. Hartig, South Atlantic Fisheries Management Council, 9277 SE Saron Street, Hobe Sound, FL 33455, personal communication, 2008).

In this paper, I introduce a hierarchical modeling approach for estimating a single time series of relative abundance values from multiple noisy indices. The method works by assuming that each CPUE index is attempting to sample relative abundance but is subject to both sampling and process errors, with the latter due to temporal variation in index-specific catchability (and possibly to differences in selectivity between gear types). I start by providing a description of model structure, demonstrating how Markov chain Monte Carlo (MCMC) can be used to fit the model to data. I then explore a number of different simulation scenarios to evaluate estimator performance and illustrate an example application to Spanish mackerel off the US Atlantic coast. Finally, I discuss some of the benefits and limitations of the method as currently proposed and make several suggestions about how to use resultant estimates in stock assessments. I also propose several future research directions.

Materials and methods

Model development

Consider the case in which an investigator has access to I different time series of relative abundance (each of which is referred to individually as “an index”). Each index $\boldsymbol{U_i}$ is composed of $\{U_{it}\}$ for $t \in \{t_{i1},\ldots,t_{iT}\}$, where i distinguishes the index and t is a time subscript (T gives the ending year of the study). Each index may be computed in any number of ways, including CPUE sample means from fishery-independent surveys, as well as estimates from GLMs that control for a large number of covariates. Whatever indices are available, the fundamental issues with conducting joint inference are that (*i*) indices are measured with error, (*ii*) the indices may be on different scales, and (*iii*) true abundance is rarely known. Each index is related to absolute abundance through the relationship $U_{it} = q_{it}N_t + \varepsilon_{it}$, where q_{it} gives catchability of index i in year t, N_t is vulnerable biomass or abundance in year t, and ε_{it} is an error term (hopefully random!). The problem is that given multiple index time series, it is difficult to determine whether differences in indices are due to differences in q_{it}, differences in ε_{it}, or actual changes in abundance (e.g., Fig. 1). This problem is reduced, but not eliminated, when constant catchability is assumed (i.e., $q_{it} = q_i$).

Ultimately, inference based on relative abundance indices is concerned with proportional changes in abundance over time rather than absolute abundance. That is, if we assume that relative abundance is known at a fixed reference point, all we really desire is to be able to estimate proportional change from that point in any other year of the study. Alternatively, we might set mean relative abundance to a fixed value and estimate annual deviations. Given these types of constraints (which are needed to make parameters identifiable), multiple indices provide exactly the sort of replication that is needed to estimate both the error term and an additional scaling term for each index, as will be shown below.

Let $\boldsymbol{\mu} = \mu_1, \mu_2, \ldots, \mu_T$ represent a scaled abundance time series in which annual changes in μ_t are reflective of changes of abundance at the population scale (that is, $\mu_t = cN_t$ for some unknown constant c). A lognormal error structure is often assumed for indices of abundance. Adopting this convention and assuming that indices are subject to independent and multiplicative process and sampling errors, we may write

$$\log(U_{it}) \sim \text{Normal}(\log(\mu_t) + \log(q'_{it}),\ (\sigma^p_{it})^2 + (\sigma^s_{it})^2)$$

Here, σ^p_{it} and σ^s_{it} give the standard deviation associated with process and sampling errors, respectively, and q'_{it} is a scaling factor for index i in year t (e.g., Fig. 1). I assume that the investigator has access to estimates of survey sampling error as could be computed from knowledge of the sampling design (in the case of fishery-independent surveys) or from standard output of model fitting exercises (e.g., from the Hessian in delta-GLM analysis). Assuming a lognormal error structure, the precision attributable to sampling error may be written as a function of the estimated coefficient of variation on the absolute scale, so that $\sigma^s_{it} = \sqrt{\log([\text{CV}(U_{it})]^2 + 1)}$. Making the substitutions $\chi_{it} = \log(q'_{it})$ and $\nu_t = \log(\mu_t)$ results in the model

$$(1) \qquad \log(U_{it}) \sim \text{Normal}(\nu_t + \chi_{it},\ (\sigma^p_{it})^2 + (\sigma^s_{it})^2)$$

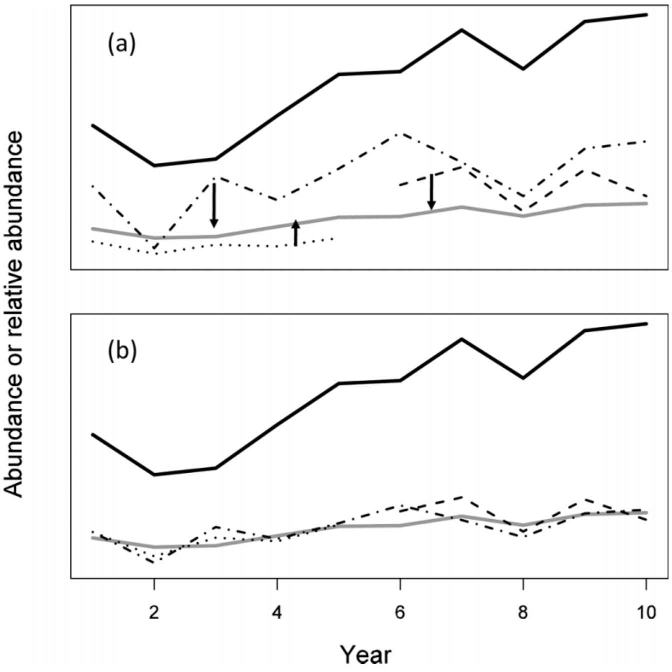
If catchability is assumed to be stationary (e.g., no trends in catchability), the model may be further simplified to

$$(2) \qquad \log(U_{it}) \sim \text{Normal}(\nu_t + \chi_i,\ (\sigma^p_{it})^2 + (\sigma^s_{it})^2)$$

Likelihood and constraints

The model in eq. 1 implies a joint likelihood function for the index data that is conditional on known values of sampling error, namely

Fig. 1. A hypothetical, pictorial depiction of the inferential problem with trying to estimate a common population trend from multiple CPUE indices. Unknown true abundance is represented by the continuous black line, relative abundance (denoted by μ_{it} in modeling efforts) by the continuous shaded line, and three hypothetical CPUE index time series with broken and dotted lines. In practice, there is an infinite number of possible relative abundance time series (any time series that preserves proportional changes in abundance), so one must "anchor" it by setting its mean or the value in a reference year to a fixed value. Inference can then focus on process errors and scaling parameters ($q_i^{'}$; represented by arrows in part *a*) needed to best explain relative abundance (e.g., part *b*).



$$(3)\qquad [\boldsymbol{U}|\,\boldsymbol{\chi},\,\boldsymbol{\nu},\,\boldsymbol{\sigma^p},\,\boldsymbol{\sigma^s}] = \prod_{i=1}^{I}\prod_{t=1}^{T} f\left(\log(U_{it})|\,\nu_t + \chi_{it},\ (\sigma_{it}^p)^2 + (\sigma_{it}^s)^2\right)$$

where bold–italic symbols denote vectors, and $f(x|\theta_1, \theta_2)$ denotes a Gaussian probability density function with mean θ_1 and variance θ_2. In this form, the constraint needed to scale the problem can be made in a number of ways, including setting ν_k to a fixed value for an arbitrary k, or by setting $\bar{\nu}_k$ to a fixed value. For this paper, I implemented a variant of the second approach, based on hierarchical centering (Chen et al. 2000; see Bayesian analysis section). Although analysis using classical likelihood approaches is certainly possible with additional constraints (e.g., constant process errors among indices), a Bayesian analysis via MCMC provides increased flexibility. At present, it appears necessary to have some of the the σ_{it}^p constant over time. However, it may be possible to consider autoregressive models for process error in certain circumstances, or to consider mixture models that allow for the possibility that individual indices are noninformative (e.g., Schnute and Hilborn 1993). Further consideration of possible extensions is deferred to the discussion; subsequent analysis proceeds by making the constraint that process errors are time-invariant (i.e., using eq. 2).

Bayesian analysis

Performing a Bayesian analysis requires that one specify prior distributions for $[\nu_t]$, $[\chi_{it}]$, and $[\sigma_i^p]$. The following priors were chosen for use throughout this paper as they conformed to scenarios chosen for simulation analyses, as well as perceived dynamics of Spanish mackerel (Fig. 2); however, they may deserve revisiting in other circumstances. First, I specified a Normal(log(100), 1) distribution for $[\nu_t]$. The mean of this distribution, log(100), was chosen so that the mean of the relative abundance time series ($\boldsymbol{\mu}$) would be approximately 100. This number is completely arbitrary, but some such constant is needed to scale the problem. In the end, it is not this number that we care about; rather, it is the deviations from this mean value that are of interest.

Next, recalling that χ_{it} specifies the natural logarithm of catchability relative to the latent abundance time series $\boldsymbol{\mu}$ (where the μ_k are centered near 100) and assuming that individual indices are standardized to their mean (and thus have a mean value of 1.0), we might expect an "average" value for the scaling parameter $q_i^{'}$ of 0.01. However, estimation is performed in $\chi_i = \log(q_i^{'})$ space, so I chose to set the prior for χ_i as $[\chi_i]$ = Normal(log(0.01), 0.5), which gives reasonable support to plausible parameter values. Finally, for σ_i^p, Gelman (2006) suggests that a Uniform(0, *m*) distribution may outperform other choices when there is a small number of group effects. For purposes of this paper, I specified a Uniform(0, 5) prior distribution for $[\sigma_i^p]$, which gives equal weight to all plausible precision values.

Given the choice of prior distributions (here assumed independent), the posterior distribution is specified up to a proportionality constant:

$$(4)\qquad [\boldsymbol{\chi},\ \boldsymbol{\nu},\ \boldsymbol{\sigma^p}\,|\,\boldsymbol{U},\ \boldsymbol{\sigma^s}] \propto [\boldsymbol{U}\,|\,\boldsymbol{\chi},\ \boldsymbol{\nu},\ \boldsymbol{\sigma^p},\ \boldsymbol{\sigma^s}][\boldsymbol{\chi}][\boldsymbol{\nu}][\boldsymbol{\sigma^p}]$$

Provided an index data set and estimates of sampling error, eq. 4 can be sampled with a technique such as Markov chain Monte Carlo (MCMC). I performed all posterior simulation using the software package WinBUGS (Lunn et al. 2000), with the package R2WinBUGS (Sturtz et al. 2005) used to pass data sets between WinBUGS and the R programming environment (R Development Core Team 2007) (see Computation section).

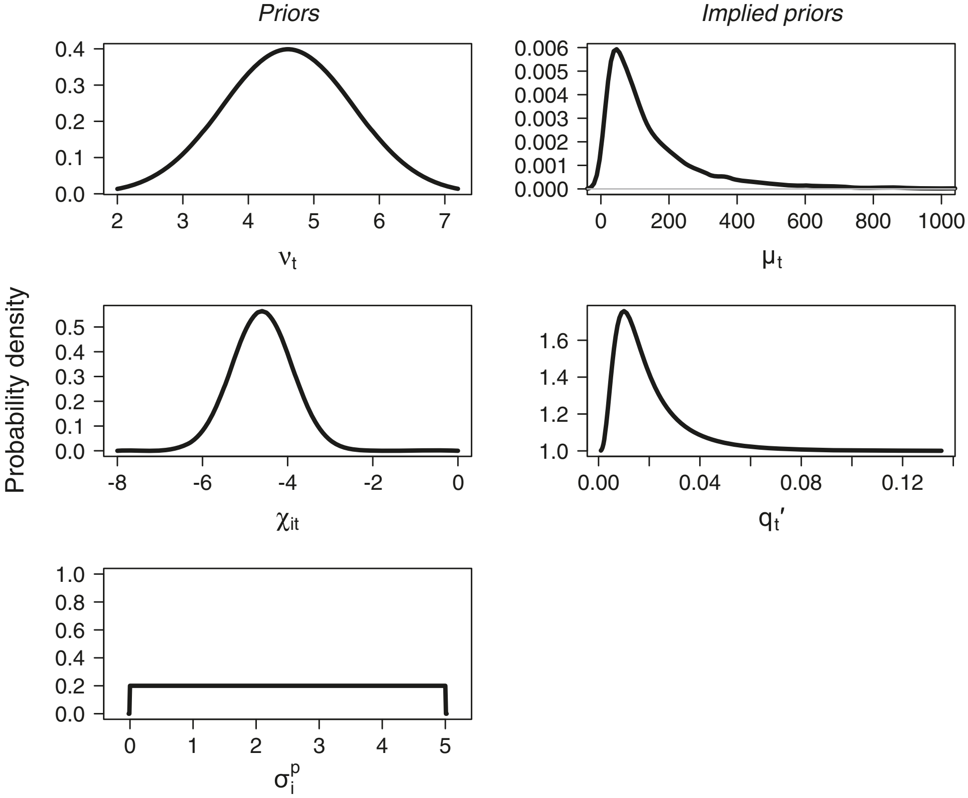
Simulation testing

To demonstrate that the hierarchical approach produces a combined index with reasonable properties, I conducted simulations to quantify performance under a number of scenarios.

Simulations scenarios 1 and 2: all assumptions met

Simulation scenarios 1 and 2 measured estimator performance when index data were simulated according to the same probabilistic structure assumed for analysis (that is, lognormally distributed process and sampling errors where the sampling errors are assumed known). To generate abundance time series, I set abundance in year 1 of each simulation to N_1 = 100 and then allowed it to follow a random walk subject to a boundary condition. More specifically, I let $N_t = \max(20, N_{t-1} + \varepsilon_{t-1})$, where $\varepsilon_t \sim \text{Normal}(0, 15^2)$. The value of 15 for the standard deviation of the random walk was subjective but permitted abundance to vary quite sub-

Fig. 2. Prior distributions used in MCMC analysis (log space; left panels) together with implied priors (real space; right panels).



stantially throughout the time series. In simulation scenario 1, all indices had the same number of years and error levels, but these could vary from simulation to simulation. For each simulation in this scenario, the number of years T was drawn from a Uniform$(10, 20)$ distribution, the number of indices I was drawn from a Uniform$(4, 10)$ distribution, and the CVs for sampling and process errors were drawn from Uniform$(0.1, 0.3)$ and Uniform$(0.0, 0.3)$ distributions, respectively. Each index was assumed to have a catchability coefficient of $q = 0.01$, was sampled with lognormally distributed error, and was scaled to its mean prior to analysis.

In simulation scenario 2, the number of indices was set to five for all simulations, and there was increased heterogeneity among the indices. In particular, the first index was assumed to span a 20-year time period, with the remaining indices spanning anywhere from 5 to 15 years. Once the duration of an index was sampled (according to a uniform distribution), its placement within the 20-year time series was randomly sampled. Each index had a process error CV sampled from a Uniform$(0, 0.3)$ distribution and a sampling error CV sampled from a Uniform$(0.1, 0.3)$ distribution. This scenario included the increased realism of different levels of precision and indices that spanned different time periods. One thousand data sets were simulated for each scenario.

Simulation scenario 3: spatial dynamics

One possible source of process variation in index values is a difference in spatial coverage among indices. This is particularly the case when the proportion of fish using a given area varies annually. To investigate consequences of this dynamic on estimator performance, I simulated data for the extreme case in which the range of the stock was broken into multiple "regions”, and an index was computed for each. Note that this is a highly idealized case; the entire population is still sampled in its entirety.

To simulate dynamics, I first generated abundance values for the entire population using the same random walk model as in the previous section. I then let the proportion of fish that use spatial area i at time t be given by π_{it}, where the distribution of π_{it} is defined by a multinomial logit link function:

$$\pi_{it} = \begin{cases} \dfrac{\exp(\beta_{it})}{1 + \sum_{i=1}^{I-1} \exp(\beta_{it})}, & i < I \\[2ex] 1 - \dfrac{\sum_{i=1}^{I-1} \exp(\beta_{it})}{1 + \sum_{i=1}^{I-1} \exp(\beta_{it})}, & i = I \end{cases}$$

Values for β_{i1} were initialized by drawing from a Normal$(0, 0.04)$ distribution at the start of the time series, while later values were determined according to a bounded random walk:

$$\beta_{i,t+1} = \beta_{it} + Z_{it}$$

where $Z_{it} \sim$ Normal$(0, \sigma^2_{\text{spatial}})$. This simple generating model constrains the proportion of fish using different areas to sum to 1.0 and allows for serial autocorrelation in these proportions. The level of autocorrelation is defined by

σ_{spatial}; variation in this quantity implies very different spatial dynamics (e.g., Fig. 3). In Fig. 3 and throughout simulations, the β_{it} were resampled any time that a π_{it} value was less than 0.1 to keep mixing proportions at reasonably high values.

For each of 1000 simulations, σ_{spatial} was drawn from a Uniform(0.0, 0.3) distribution. Each simulation assumed five indices, 20 years of data, and a sampling error CV of 0.2 for each index. No process error was included in the model apart from that generated by annual changes in mixing proportions. Individual indices were calculated as

$$U_{it} = \exp(Z_{it})$$

where $Z_{it} \sim \text{Normal}(\log(N_t \pi_{it} q_i), 0.2)$ and q_i was set to 0.01. Each index was scaled by its mean prior to analysis.

Simulation scenario 4: selectivity variation

To investigate whether the combined index approach is robust to differences in gear selectivity, I simulated data from an age-structured population model (Table 1) that differed by recruitment variability and by the selectivity pattern used to construct various indices. In all cases, index values were determined by

$$U_{it} = \exp(\varepsilon_{it}) \sum_a N_{ta}\, s_{ia}\, w_a$$

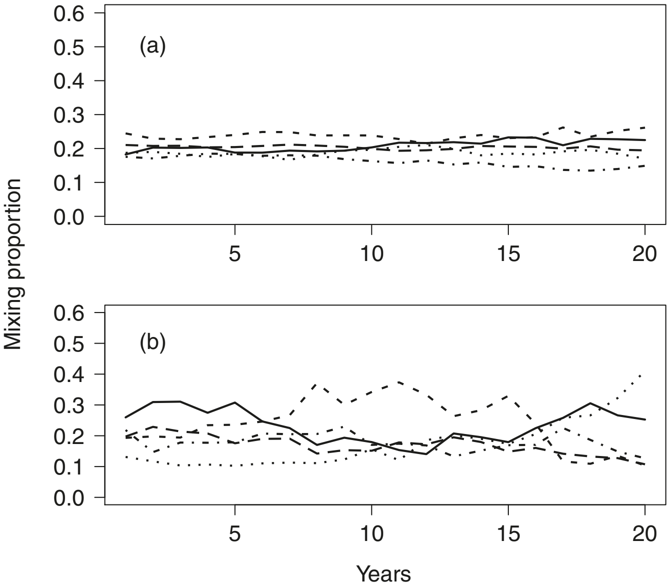
where N_{ta} is abundance of age class a in year t, s_{ia} is selectivity of gear i to individuals in age class a, w_a is weight at age, and $\varepsilon_t \sim \text{Normal}(0, 0.2)$ is Gaussian noise associated with the sampling process. Each index was standardized to its mean prior to analysis.

Four functional forms for selectivity were considered, including (A) 100% selection, (B) logistic, (C) dome-shaped (double logistic), and (D) recruits only (Table 1; Fig. 4). Interest focused on performance of the hierarchical approach when different combinations of these functional forms (and different slope parameters) were used to generate index data. In each simulation, the comprising functional form for each index was sampled with replacement from the above list; if logistic or dome-shaped functions were selected, the parameters of these distributions were sampled as follows: $\eta \sim \text{Uniform}(1, 3)$, $\alpha_1 \sim \text{Uniform}(1, 4)$, and $\alpha_2 \sim \text{Uniform}(4, 6)$. Random selection of parameters increased heterogeneity in indices.

Several options for population trajectory and recruitment variability were also considered. In particular, fishing mortality could be constant at the same level as natural mortality ("stable" population; $F_t = M$), could increase linearly from $F_1 = M$ to $F_T = 2M$ ("decreasing population"), or could decrease linearly from $F_1 = M$ to $F_T = 0.5M$ ("increasing population"). For recruitment variability, I considered two possibilities: $\sigma_R = 0.3$ or $\sigma_R = 0.8$.

Simulations in scenario 4 were conducted according to a factorial design, with three levels for population trajectory and two levels for recruitment. For each of 100 simulations at each design point, I initialized $\boldsymbol{N_1}$ (the abundance at age vector in year 1 of the study) by simulating the dynamics of an exploited population for 11 years (number of age classes plus 1) prior to the first year of data. The population was set to the stable age distribution at the beginning of the initialization phase (cf. Caswell 2001). All simulations assumed that total mortality in the initialization phase was $Z = 2M$, where $M = 0.2$. Ten indices were generated for each simulation, each of which was assumed to be sampled for 20 years.

Fig. 3. Mixing proportions for five indices simulated using a random walk on parameters subject to a multinomial logit transformation. (*a*) A small standard deviation for the random walk (e.g., $\sigma_{\text{spatial}} = 0.05$) resulted in relatively constant mixing proportions, whereas (*b*) larger annual variations are possible with $\sigma_{\text{spatial}} = 0.30$.



Post hoc scenario

Following analysis of scenarios 1–4, I conducted an additional simulation scenario to highlight differences between the hierarchical index and the arithmetic mean index (when calculable). This scenario assumed five indices, each of which was monitored for 20 years. In this case, each index was assumed to be subject to different levels of sampling and process errors. In particular, $\sigma_i^p \sim \text{Uniform}(0.0, 0.5)$ and $\sigma_i^s \sim \text{Uniform}(0.1, 0.8)$. These values were resampled for each of 1000 simulations.

Computation

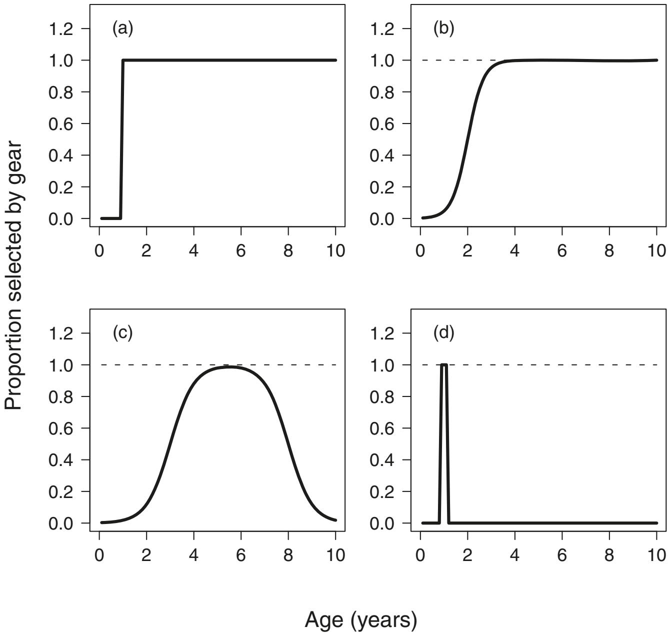
For each simulation scenario, I fit four independent Markov chains of length 60 000 to each simulated data set. Each chain was started at random starting values, and the first 10 000 iterations were treated as a burn-in phase and discarded. The remaining 50 000 iterations from each chain were combined to yield a sample of the posterior, with one in every 10 values recorded to save disk space. The result was 20 000 posterior samples for each simulated time series. Several sets of Gelman–Rubin statistics (Gelman et al. 2004) were examined to confirm that Markov chains were indeed converging to the posterior distribution. Chains appeared to converge after several thousand iterations for the most highly parameterized simulations, suggesting that chain lengths were relatively conservative.

Several measures of posterior performance were recorded, including bias of the posterior mean and 95% credible interval coverage for μ_t and σ_i^p. Because the posterior is written in terms of ν_t, inference about μ_t was made by calculating

Table 1. Formulation of population dynamics model used to simulate index data in simulation scenario 4.

Quantity	Symbol	Description or definition
Weight at age	w_a	$w_a = \frac{5}{1+\exp[-(a-3)]}$; note that the maximum weight is 5.0 and the age at 50% of maximum weight is 3.0
Maturity	m_a	$m_a = \frac{1}{1+\exp[-(a-3)]}$; note that the age at 50% maturity is 3.0
Fishery selectivity	s_a	Three models are considered: (A) all selected; $s_{ia} = 1$ (B) logistic; $s_{ia} = 1/[1 + \exp(-\eta(a - \alpha_1))]$ (C) dome-shaped (double logistic); $s_{ia} = \left(\frac{1}{1+\exp[-\eta(a-\alpha_1)]}\right)\left(1 - \frac{1}{1+\exp[-\eta(a-[\alpha_1+\alpha_2])]}\right)$ (D) recruitment index; $s_{i1} = 1$, $s_{ia} = 0$ for $a > 1$
Fishing mortality rate	$F_{a,y}$	$F_{a,y} = s_{a,y} F_y$, where F_y is the fully selected fishing mortality rate
Total mortality rate	$Z_{a,y}$	$Z_{a,y} = M + F_{a,y}$
Abundance at age	$N_{a,y}$	$N_{1,y+1} = \frac{0.8\,R_0\,h\,S_y}{0.2\,\Phi_0\,R_0\,(1-h)+(h-0.2)\,S_y}\exp(\varepsilon_y^R)$ $N_{a+1,y+1} = N_{a,y}\exp(-Z_{a,y})\;\forall\, a \in (1,\ldots,A-1)$ $N_{A,y} = N_{A-1,y-1}\,\frac{\exp(-Z_{A-1,y-1})}{1-\exp(-Z_{A,y-1})}$, where $\varepsilon_y^R \sim \text{Normal}(0, \sigma_R)$ and Φ_0 gives unfished spawning stock biomass per recruit. The latter quantity is a function of growth and mortality schedules. For purposes of this simulation study, I set $h = 0.75$ and $R_0 = 10\,000$
Mature biomass	S_y	$S_y = \sum_a N_{a,y}\,w_a\,m_a$; also referred to as spawning stock biomass
Population biomass	B_y	$B_y = \sum_a N_{a,y}\,w_a$

Fig. 4. Functional forms for selectivity used in simulation scenario 4. Recruitment to the fishery was assumed to occur at age 1.



posterior moments for transformed samples (i.e., $\mu_t^{(j)} = \exp(\nu_t^{(j)})$ for MCMC iteration j. For simulation scenarios 1–3, I also calculated bias and coverage with respect to the observed values of a single index (this was done for the longest running index in simulation scenario 2 and a randomly selected index for scenarios 1 and 3). For simulation scenarios 1, 3, and 4 (as well as the post hoc scenario), I calculated performance relative to an arithmetic mean index, which was calculated by (*i*) standardizing each index to its mean, and (*ii*) calculating the average index value in year t. Comparisons of the performance of the combined index with that of a single, unbiased index and with the arithmetic mean index provide one measure of how much one "gains" by employing the hierarchical modeling approach. For simulation scenario 4, I recorded performance statistics using all indices, and also for the case in which recruitment indices were censored prior to analysis.

Analysis of simulation output

For each scenario, I used a variety of graphical and statistical approaches to examine percent relative bias, mean squared error (MSE), and 95% credible interval coverage (95% COV) resulting from different simulation scenarios. Examination of bias was difficult to do in a statistical sense because each estimated index was scaled to its mean before computation of performance statistics. As such, mean bias is zero by definition; thus, bias was primarily examined visually. For MSE and 95% COV, I used GLMs to investigate factors influencing performance. In particular, I used stepwise Akaike's information criterion (AIC) with both forward and backward selection (Venables and Ripley 2002) to evaluate potential predictor variables. Potential predictor variables used in each analysis are identified below.

For simulation scenario 1, MSE analysis was done with $\text{MSE}^{0.25}$ as the dependent variable and T, I, σ^p, σ^s, and I_T as predictor variables. Here, I_T is an indicator variable that equals 1.0 when the estimate is from the final year of the study and 0.0 otherwise. The inclusion of I_T was suggested by visual inspection of simulation results. The 0.25 exponent on MSE was necessary for residuals to be approximately normally distributed (as confirmed by inspection of quantile–quantile plots). The most complicated model in stepwise AIC selection included all main effects. I did not consider higher order terms because (*i*) they were more difficult to interpret, and (*ii*) estimates of MSE in individual time series were not truly independent, and thus AIC could be expected to erroneously choose models that were too complicated (i.e., assumed sample sizes were overinflated). I examined 95% COV in an analogous fashion, this time assuming that COV was a Bernoulli response variable. Predicted 95% COV that was higher than 95% indicated that credible intervals were too conservative (that is, include true parameter values >95% of the time), whereas predictions less than 95% indicate that estimated precision is overstated. I also examined bias and COV of the estimator of process standard deviation, σ^p. Again, GLMs were used to examine factors influencing bias and coverage. In this case, the highest dimensional model included additive effects of T, I, σ^p, and σ^s, as well as an interaction between σ^p and σ^s. The importance of the interaction was apparent when examining simulation data visually.

For simulation scenario 2, the length and placement of indices varied considerably; in this case, overall bias, MSE, and 95% COV were examined to make sure that the posterior mean still had reasonable performance. For simulation scenario 3 (spatial dynamics), effort was made to relate the three performance measures to σ_{spatial}, which controlled autocorrelation in spatial mixing proportions. This was done via GLMs in addition to plotting and comparison with individual indices. This same approach was used for simulation scenario 4 (selectivity differences), although in this case, possible predictors were the expected population trend (three levels: increasing, constant, decreasing) and recruitment variability (σ_R: low or high).

Spanish mackerel example

Traditionally, Spanish mackerel off the Atlantic coast have been split into Gulf and Atlantic stocks for management purposes. Here, I focus on the Atlantic stock, which overwinters off the eastern coast of Florida and migrates north in the summer. The bulk of the commercial fishery occurs off the Florida coast in winter, but landings occur all the way to southern New England in the summer (landings this far north are rare).

I gathered Spanish mackerel indices of abundance from a recent stock assessment of Spanish mackerel (SEDAR 2008*a*) for hierarchical analysis. The fishery-dependent indices had all been adjusted to account for an assumed linear increase in catchability thought to result from technology creep prior to use in the assessment, an adjustment which I retained here. A total of nine indices were available, representing a number of fishing gears, data sources, time periods, and geographical areas (Table 2). I performed analysis on all seven fishery-dependent CPUE indices, electing not to use two fishery-independent index time series because these only selected for small fish (young-of-year and some one-year-olds). In contrast, estimated selectivities from a statistical catch-at-age model (SEDAR 2008*b*) were similar for fishery-dependent gear types, at least for high abundance age classes. However, there was poor correlation between indices, with individual Pearson's correlation coefficients ranging from –0.64 to 0.44. Assessment panelists were unable to justify why one index should be favored over the rest. Analysts were also concerned about numerical stability when using multiple, conflicting indices in assessment models, particularly for auxiliary analyses such as surplus production and stochastic stock reduction models. The collection of indices thus seemed ideally suited for hierarchical analysis. For model fitting, I employed the same MCMC configuration and set of prior distributions used in the simulation study with the exception that the chains were not thinned (apart from saving disk space, there is little reason to do so; Spiegelhalter et al. 2003).

Results

Simulation performance

Scenario 1: all assumptions met

When indices were continuous and data were simulated according to the same error distribution assumed during estimation, the combined index was unbiased, with a similar MSE as the arithmetic mean of indices (when all indices have the same level of process and sampling error; Fig. 5). Inspection of bias in relation to simulation inputs suggested that the index was approximately unbiased in all cases. The spread of relative bias over individual data points appeared Gaussian and exhibited less spread than when inference about relative abundance was based on one index alone. Computations of MSE agreed, with MSE significantly lower for the posterior mean from the hierarchical analysis (0.011; SE < 0.001) and the arithmetic mean index (0.011; SE < 0.001) than for the single index (0.071; SE = 0.005). Stepwise model selection procedures using AIC (Venables and Ripley 2002) suggested that only the number of indices and the levels of process and sampling errors were important predictors of MSE. Point estimates from the highest ranked linear model suggested that MSE decreased with the number of indices available for analysis and increased as the assumed level of process and sampling error increased.

Analysis of 95% COV suggested that uncertainty about μ was overestimated using the hierarchical approach, with true coverage around 99.9% (SE = 0.1%). In contrast, coverage for the arithmetic mean index was underestimated at 91.8% (SE = 0.9%). The highest ranked AIC model provided no evidence that 95% COV was a function of simulation inputs.

Overall, bias in the posterior mean for process error was positive (0.039; SE = 0.002), which may reflect the adoption of a flat prior for σ^p. When investigating bias using a GLM, diagnostics indicated a substantial lack of fit when bias was related to predictor variables. This problem was not easily remedied by transformations of the predictor variable, so an overdispersion parameter ($\widehat{c} = \text{Deviance/df}$) was estimated to adjust model selection criterion (i.e., model selection was based on quasi-likelihood formulation of Akaike's information criterion (QAIC); Burnham and Anderson 2002). Using this approach, $\widehat{c} = 2.33$, and the highest ranked QAIC model included effects for the number of years, σ^p, and σ^s, along with an interaction between σ^p and σ^s. Bias under this model was positive when the CV of true process error was less than about 0.2 and negative otherwise. In general, the magnitude of bias increased as the CV of sampling error increased (Fig. 6). Overall, 95% COV for σ^p was greater than nominal at 96.8% (SE = 0.6%). However, there was considerable variation in coverage depending on simulation inputs. For instance, the highest ranked AIC model fit to the simulation data included all simulation inputs as predictor variables and suggested that coverage was higher than nominal when errors were high (e.g., 0.998 when $\sigma^p = \sigma^s = 0.3$) and lower than nominal when errors were low (e.g., 0.860 when

Table 2. Spanish mackerel indices used in hierarchical analysis (sampling error CV in parentheses).

Year	FL.GN1	FL.GN2	FL.CN	FL.HL	MRFSS	LB.GN	LB.HL
1985	0.46 (0.07)			0.69 (0.08)			
1986	0.59 (0.07)			0.94 (0.08)			
1987	0.83 (0.07)			1.03 (0.08)	0.84 (0.07)		
1988	0.64 (0.07)			1.21 (0.08)	1.17 (0.06)		
1989	0.93 (0.07)			1.16 (0.07)	1.00 (0.08)		
1990	0.79 (0.06)			1.12 (0.07)	0.90 (0.06)		
1991	0.65 (0.06)			0.87 (0.06)	0.77 (0.05)		
1992	0.63 (0.07)			0.85 (0.07)	0.83 (0.04)		
1993	2.1 (0.19)			0.87 (0.08)	0.66 (0.06)		
1994	2.4 (0.12)			0.68 (0.07)	1.01 (0.06)		
1995				0.69 (0.07)	0.81 (0.08)		
1996		1.25 (0.17)		0.63 (0.08)	1.06 (0.07)		
1997		0.77 (0.34)		0.67 (0.07)	1.09 (0.06)		
1998		1.05 (0.17)		0.95 (0.06)	0.93 (0.07)	0.79 (0.13)	0.87 (0.14)
1999		1.05 (0.17)	0.77 (0.15)	0.82 (0.07)	1.16 (0.06)	0.72 (0.13)	1.12 (0.15)
2000		1.09 (0.14)	0.77 (0.13)	0.92 (0.06)	1.07 (0.06)	0.92 (0.12)	0.88 (0.16)
2001		0.88 (0.14)	0.83 (0.13)	1.40 (0.07)	1.13 (0.07)	1.23 (0.13)	0.97 (0.15)
2002		0.85 (0.15)	0.95 (0.12)	0.85 (0.06)	1.36 (0.07)	1.34 (0.12)	1.19 (0.18)
2003		0.94 (0.15)	1.39 (0.12)	1.22 (0.07)	1.19 (0.08)	1.20 (0.12)	0.93 (0.19)
2004		0.62 (0.15)	1.48 (0.13)	1.52 (0.06)	0.99 (0.07)	1.00 (0.12)	1.00 (0.21)
2005		1.11 (0.14)	1.17 (0.12)	1.22 (0.07)	1.29 (0.07)	0.85 (0.12)	0.86 (0.14)
2006		1.17 (0.14)	0.86 (0.13)	1.46 (0.06)	0.82 (0.07)	0.90 (0.11)	1.16 (0.19)
2007		1.21 (0.13)	0.78 (0.13)	1.22 (0.06)	0.88 (0.06)	1.06 (0.11)	0.80 (0.15)

Note: The indices include four standardized trip ticket indices for the state of Florida: gillnet prior to net ban in state waters (FL.GN1); gillnet after state net ban (FL.GN2); castnet (FL.CN); and hook-and-line (FL.HL). Also available were indices based on recreational survey data (MRFSS) and from logbooks in states north of Florida, which included standardized CPUE from gillnet (LB.GN) and hook-and-line (LB.HL) fisheries. For details on index construction, see SEDAR (2008*a*).

$\sigma^p = \sigma^s = 0.1$. Coverage also increased with the number of indices, whereas the number of years had a negligible effect.

Scenario 2: all assumptions met, heterogeneous indices

When there was a small number of heterogeneous indices, the hierarchical approach resulted in a lower MSE (0.028) than a single index that spanned the entire period (MSE = 0.075). There was no obvious way to compute an arithmetic mean index for comparison, as indices were available for different lengths of time and did not all overlap at a common time point. There was no visible bias pattern (the estimators were unbiased throughout simulated time series), but COV was again substantially greater than the 95% nominal level, this time at 99.7% (SE = 0.2%), suggesting that credible intervals were overly conservative.

Scenario 3: spatial dynamics

A visual inspection of bias from simulations employing variation in mixing proportions suggested that there was potential for positive bias in the hierarchical index estimator at the beginning of the time series, particularly when the proportion of fish using each area could vary considerably over time (Fig. 7). Over all simulations, there was a spurious 6.1% decline in relative abundance. However, post hoc experimentation with the study design suggested that this bias was at least partly an artifact of the random walk process occurring in multinomial logit space, as it was relatively common for indices with the least variable mixing proportions to also be those with mixing proportions that were declining (e.g., Fig. 3). It appears that the hierarchical estimator may have been keying in on the least variable (and also declining) indices.

Perhaps reflecting this bias, MSE was lower for the arithmetic mean index (0.009) than the hierarchical index estimator (0.012). Both of these options were substantially better than basing inference on a randomly selected index (MSE = 0.083). In general, mean squared error increased substantially for simulations employing higher levels of variation in spatial mixing proportions. Overall COV was 99.9% for a nominal 95% interval, suggesting that the estimation procedure (or selected set of prior distributions) is overly conservative. Model selection provided no evidence that coverage was related to σ_{spatial}. In comparison, asymptotic coverage for the arithmetic mean index was 93.2%, thus much closer to nominal.

Scenario 4: selectivity differences

Randomly selected plots of individual indices and the hierarchical index suggested that estimates of relative abundance were fairly insensitive to selectivity differences among indices, although recruitment indices could be quite variable (e.g., Fig. 8). A comparison of MSE for (*i*) analyses including recruitment indices (MSE = 0.013) and (*ii*) those censoring recruitment indices prior to analysis (MSE = 0.010) indicated that there was a slight benefit to censoring recruitment indices. I thus only report results for case *ii*.

Bias of the posterior mean for μ appeared negligible. The highest ranked AIC model for mean squared error did not include any simulation inputs, suggesting that neither population trend nor recruitment variability seriously impacted overall model performance. Mean squared error was similar

Fig. 5. Kernel density estimates of observed percent relative bias of index estimators for different simulation scenarios: (*a*) simulation scenario 1 (all assumptions met); (*b*) post hoc simulations (noisy, heterogeneous indices); (*c*) simulation scenario 3 (spatial dynamics); and (*d*) simulation scenario 4 (selectivity differences). The continuous line shows the distribution resulting from hierarchical analysis, the dotted line shows results for the arithmetic mean index, and the broken line shows the distribution of bias when a single time series is chosen as an index of abundance. Because time series are all scaled to their mean, the mean percent relative bias is 0 by definition.

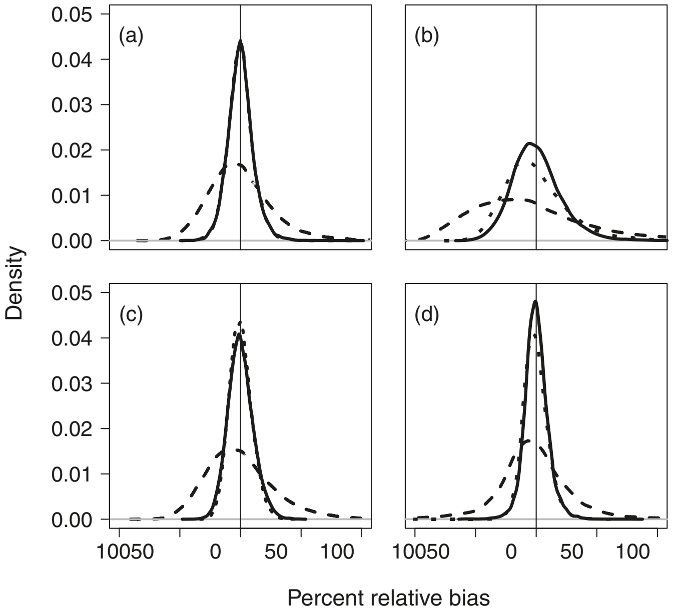
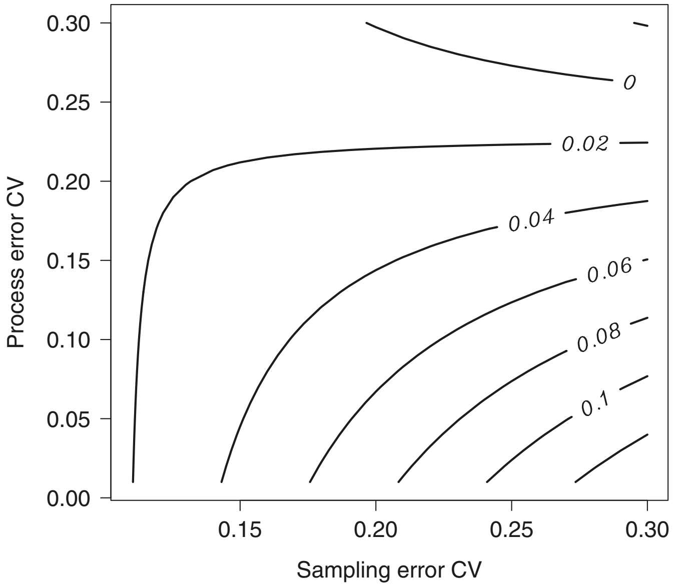
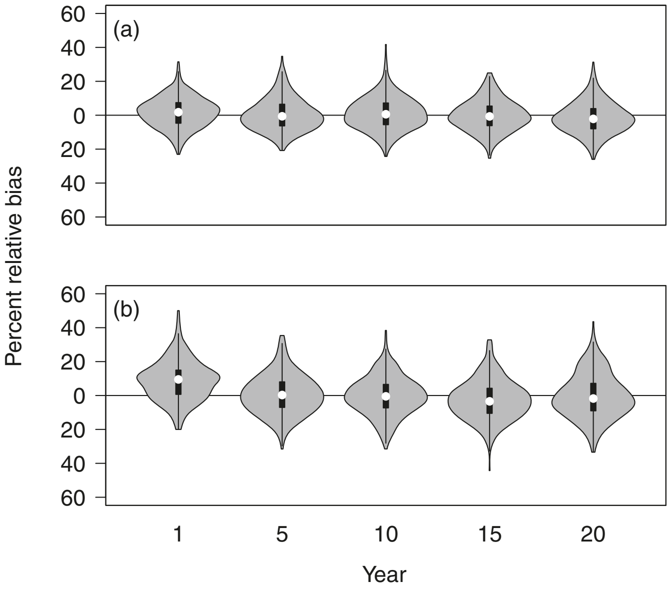


Fig. 6. Predicted bias of σ_P as a function of the true CVs used to generate data in simulation scenario 1. A linear model was used to relate bias to simulation input variables.



for the hierarchical approach when compared with the arithmetic mean index, suggesting similar performance (Fig. 4). Overall COV was 0.997 for the hierarchical approach, compared with 0.848 for the arithmetic mean index.

Fig. 7. The distribution of percent relative bias of the combined index as a function of the year of simulation for simulation scenario 3 (spatial dynamics): (*a*) bias from simulations with smaller annual variation in mixing proportions ($\sigma_{\text{spatial}} < 0.1$); (*b*) bias from simulations with higher variation ($\sigma_{\text{spatial}} > 0.2$). Each "violin" plot combines a kernel density estimate (essentially a smooth histogram) with a traditional box-and-whisker plot. Median values are represented by white circles, thick bars represent interquartile range, and thin vertical lines represent the distance between 2.5% and 97.5% quantiles.



Post hoc scenario

In this scenario, the hierarchical analysis produced more precise estimates of population trends than the arithmetic mean index (Fig. 5). Overall MSE for the hierarchical approach was 0.044 (SE = 0.003), compared with 0.076 (SE = 0.006) for the arithmetic mean index. Credible interval coverage was 98.9% (SE = 0.3%) for the hierarchical index and 85.9% (SE = 1.1%) for the arithmetic mean index.

Spanish mackerel analysis

The hierarchical model produced posterior estimates of μ suggestive of two "peaks" in abundance from 1985 to present (Fig. 9). The first appears around 1988, shortly after the first catch quotas were imposed by emergency rule. The second occurs in 2002–2003 following a period of increase that started in the early 1990s. For reference, a moratorium on gillnets was imposed in Florida state waters starting in 1995, a move that is widely thought to have resulted in an increase in abundance. The posterior mean of relative abundance in 2007 suggests that relative abundance has decreased somewhat from its peak in the early 2000s.

In addition to estimates of the index itself, one also can examine estimates of process error variance associated with each of the indices (Fig. 10). For instance, the analysis suggests that process errors for the Florida gillnet index (FL.GN1) prior to the state gillnet ban were sampled from a distribution that was quite diffuse. This likely is due to the two large FL.GN1 index values in 1993 and 1994 that were incongruous with the other time series and suggests that the FL.GN1 index may be a poor index of abundance in its own

Fig. 8. An example simulation in which indices are generated assuming different selectivities: circles denote true biomass; the continuous line is an index with all ages sampled; the broken line is an index with a dome-shaped selectivity function; the dotted line is an age-1 recruitment index; and the shaded continuous line gives relative abundance as estimated via hierarchical analysis of 10 indices (only three of which are pictured here). The hierarchical index had a lower MSE than any of the individual indices, a feature common to most simulations.

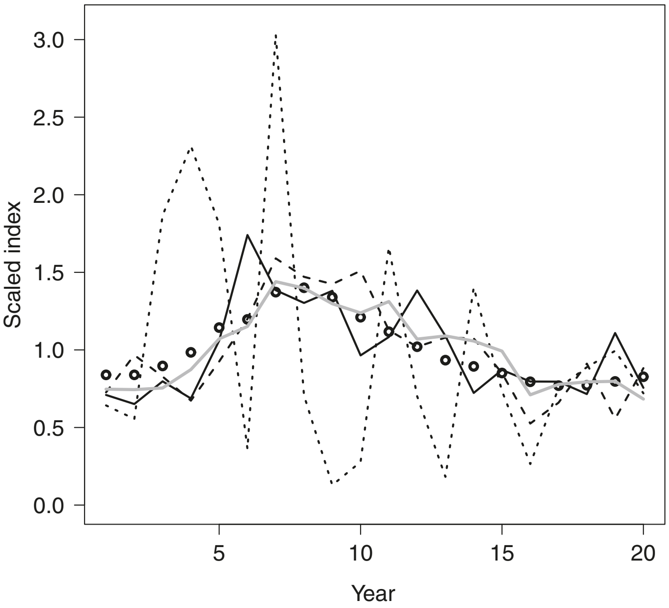
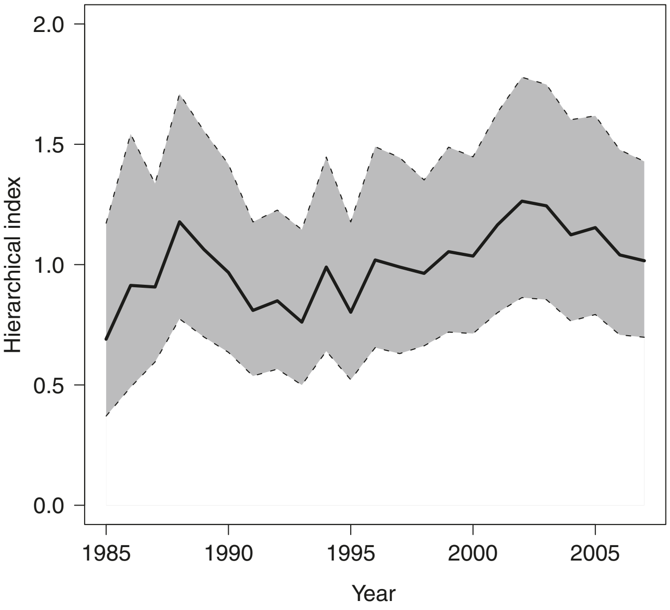
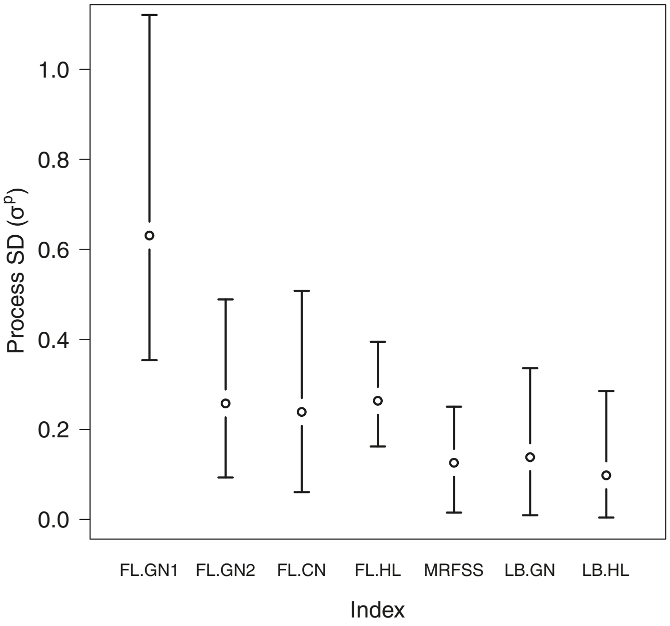


Fig. 9. Time series of Spanish mackerel relative abundance as estimated from hierarchical analysis. The continuous line gives the posterior mean, and the shaded area represents a 95% credible interval about the time series.



right. In contrast, the estimates of process error for the Marine Recreational Fisheries Statistics Survey (MRFSS) and logbook indices were quite low (<0.15), suggesting that these indices may better represent abundance (or, at least, that discrepancies between these indices were easier to reconcile).

Fig. 10. Posterior means and 95% credible intervals for the standard deviation of process error (σ^P) for the seven indices in the Spanish mackerel analysis. For index definitions, see Table 2.



Discussion

Relative abundance indices are a crucial component of many stock assessments. However, when catchability varies over time and space, different indices may provide conflicting signals. In this paper, I have shown how hierarchical analysis can be used to estimate a common population trend from multiple indices. This framework separates components of index variation into process error and sampling error. In this manner, analysts can calculate a single, “most probable” index prior to stock assessment analyses. Such an index may be of interest in its own right or may be advantageous in model fitting because it reduces the dimensionality of the likelihood and precludes numerical problems that can arise when fitting data to multiple, conflicting indices. It also has the potential to reduce the number of subjective decisions that are typically made about which indices to include in the analysis. The approach differs considerably from other approaches used to combine indices, e.g., area- or catch-weighting (cf. Quinn and Deriso 1999), which require that available habitats be correctly delineated (in the case of area-weighting) or that fishing mortality be uniformly applied across the population (in the case of catch-weighting).

In general, the simulation study suggested that hierarchical relative abundance estimates were reasonable. The hierarchical approach appeared robust to model assumption violations, including those resulting from differences in selectivities and trends in spatial mixing proportions. For the spatial scenario, there was a tendency to estimate a slight but spurious population decline (6.4% over 20 years). However, this phenomenon was likely an artifact of the way in which data were simulated. Notably, it was often the case that the random walk in multinomial logit space led to the situation in which the least variable indices all exhibited population declines.

When indices all covered the same time span, simple calculations of an arithmetic mean index (standardizing each index to its mean and then taking the average index value in each year) performed surprisingly well as a relative abundance estimator. Mean squared error for the arithmetic mean index was statistically indistinguishable from the hierarchical index in simulation scenarios 1 and 4, slightly outperformed the hierarchical index in simulation scenario 3 (spatial dynamics), but exhibited worse performance when sampling and process errors varied substantially between indices (post hoc scenario). The hierarchical approach thus appears to have the greatest utility when applied to a group of noisy, heterogeneous indices. It can also be applied in a greater range of circumstances, for example, when the time series is made up of indices that span different time periods (e.g., simulation scenario 2).

Credible intervals for the hierarchical index were frequently wider than nominal for all simulation scenarios, suggesting that the estimation procedure was overly conservative with regards to μ_t. Process error variance was frequently overestimated, which may have contributed to this phenomenon. Also, diffuse prior distributions were chosen for most parameters and reflected considerable uncertainty; more informative priors should decrease coverage. In contrast, the arithmetic mean index consistently had coverage that was below nominal.

Although hierarchical analysis appears attractive, it is important to reflect on the utility of the approach when evaluating a collection of indices. For instance, Schnute and Hilborn (1993) caution that estimates of population trend from a “compromise” fit can be misleading, particularly if one admits the possibility that models are misspecified or certain data sources are invalid. For the former, model selection and (or) model averaging (Hoeting et al. 1999; Burnham and Anderson 2002; Spiegelhalter et al. 2002) may reduce the chances of basing inference on a poorly specified error distribution. However, the problem of erroneous data is a bit more insidious. For instance, consider the case in which a specific index reflects some quantity other than relative abundance (e.g., an unmeasured environmental variable). Schnute and Hilborn (1993) suggest that analysts should often consider the case in which there is some finite probability (say, p_i) that data source i is noninformative. This mixture modeling approach will tend to decrease precision but may better reflect uncertainty about parameter estimates.

To some degree, estimation of process error by the hierarchical model will tend to reduce the influence of “garbage” indices, provided that the hierarchical model can key in on indices that are providing similar (and presumably correct) signals on relative abundance. However, problems can still occur. For instance, consider the hypothetical case of several spatially referenced indices following implementation of a marine reserve in a stock that experienced overfishing. In this case, indices calculated with data outside the reserve may all suggest a decreasing stock, whereas relative abundance estimates inside the reserve may indicate stock increases. In an effort to reconcile these indices, the hierarchical model may compensate by overestimating process error for the index within the reserve, thus reducing its influence. It is thus critical that inference be based on indices collected from properly designed sampling programs (including adequate spatial balance). No amount of statistical wizardry will remedy problems with poorly collected data.

As with any modeling exercise, violations of model assumptions can invalidate results. At present, I suggest that analysts carefully prescreen indices before any sort of hierarchical analysis is attempted to make sure that sampling designs are adequate for measuring relative abundance. Even then, one should also investigate the sensitivity of model results to index choice and should make sure that process error estimates are reasonable given prior knowledge. Unfortunately, stock assessment scientists are often put in the position of selecting a “base” model run to inform management, and some sort of compromise fit is often unavoidable. To reduce problems with invalid indices, future research might consider attempting to merge the hierarchical modeling approach with the theory presented by Schnute and Hilborn (1993). For instance, each index i could be subject to some probability p_i of being uninformative. If the same index is used to standardize CPUE for multiple species, p_i might even be estimable (using meta-analysis, for instance).

The approach developed in this paper shares many features with other statistical methods such as those that employ autoregressive integrated moving-average (ARIMA) models to smooth individual CPUE indices (e.g., Pennington 1985; Stockhausen and Fogarty 2007). However, though the latter implicitly assume constant or randomly varying selectivity to maintain constant variance throughout the time series, the multiple-index approach requires a number of additional selectivity assumptions. Fundamentally, we must assume that each time series is attempting to measure a similar quantity. This issue is perhaps best looked at along a continuum. In a theoretical age- or stage-structured population that is growing, shrinking, or staying the same with a constant rate of increase (λ), there is a well defined stable age or stage distribution (Caswell 2001). In this case, the relative proportion of the population in each age class remains constant, and it does not matter if different selectivities are associated with different indices. At the other end of the spectrum is a population exhibiting large variation in annual recruitment, as is typical of many fish populations. In this case, annual changes in an index that is solely selective of new recruits would often differ dramatically from one that selects for older age classes (and, indeed, simulation results suggest that recruitment indices should not be combined with those from older ages). However, there is likely a middle ground. In particular, results from simulation scenario 4 suggest that combining indices with different selectivities often results in a reasonable “combined” index. Ostensibly, variation among indices due to different selectivities is absorbed by the process error component in the model. That is, estimated process error likely represents both variation in catchability, as well as variation between indices due to selection of different age or stage classes.

Although further investigation of the selectivity question certainly seems warranted, a few comments and suggestions may help in practice. First, I note that index-specific selectivities are entirely ignored in age-aggregated stock assessment models such as tuned surplus production models (e.g., Prager 1994). In this context, ignoring selectivities is something routinely done in fisheries science. Second, I suggest that the hierarchical approach should be applied only to indices that have similar selectivities. This was precisely the reason why fishery-independent SEAMAP indices were removed prior to the Spanish mackerel analysis. A further consideration addresses the best way to use the "combined" index within stock assessments. For instance, when multiple selectivity curves are estimated, one must decide how best to model selectivity of the combined index. One possibility, which was used in a recent Spanish mackerel assessment, was to use an *F*-weighted selectivity vector.

Further inferences about the components of process variance may be possible if the hierarchical model is embedded in a statistical catch-age model. Because selectivities are often estimated within assessment models, it should be possible to separate process variance into two components: one due to differences in the age or stage structure sampled by different indices, and one attributable to annual fluctuations in catchability. This would be somewhat similar to the "additional variance" approach suggested by several authors (e.g., Geromont and Butterworth 2001; Wade 2002) whereby an extra precision parameter was estimated for each index above and beyond the sampling error input to the model. Another interesting extension would be to consider the possibility of trends in catchability (e.g., Wilberg and Bence 2006). For instance, if catchability was assumed to be stationary for fishery-independent data, it should be possible to estimate catchability trends for fishery-dependent indices when multiple CPUE time series are available. As one reviewer noted, an autoregressive model for $q_{it}^{'}$ might also be used to stabilize estimation in this case.

As a final note, in speaking about variation in catchability, I do not mean to imply that the hierarchical approach is capable of controlling for long-term trends in catchability such as those attributable to population density, trends in global sea surface temperatures, or other biotic or abiotic factors (particularly when these factors affect fishery-independent catchability). Where possible, covariates may be used to control for factors influencing catchability prior to analysis, but long-term trends in even measurable quantities may be confounded with trends in abundance (Nichols et al. 2009). This is not really a criticism of the method per se, as it is a feature of most methods that do not include independent estimates of catchability (e.g., from tagging studies).

Acknowledgments

The author is grateful to L. Brooks, M. Prager, K. Shertzer, E. Williams, and one anonymous reviewer for constructive comments and suggestions. C. Porch provided comments and modeling suggestions that greatly enhanced the utility of the modeling approach and also showed that a variant of this model could be fit with least squares when process error variance was constant across indices.

References

Besbeas, P., Freeman, S.N., Morgan, B.J.T., and Catchpole, E.A. 2002. Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. Biometrics, **58**(3): 540–547. doi:10.1111/j.0006-341X.2002.00540.x. PMID:12229988.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information–theoretic approach. 2nd ed. Springer-Verlag, New York.

Byrne, C.J., Azarovitz, T.R., and Sissenwine, M.P. 1981. Factors affecting variability of research trawl surveys. Can. Spec. Publ. Fish. Aquat. Sci. **58**: 238–273.

Caswell, H. 2001. Matrix population models. 2nd ed. Sinauer, Sunderland, Mass.

Chen, M.H., Shao, Q.M., and Ibrahim, J.G. 2000. Monte Carlo methods in Bayesian computation. Springer, New York.

Collie, J.S., and Sissenwine, M.P. 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. **40**(11): 1871–1879. doi:10.1139/f83-217.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. **3**: 515–534.

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

Geromont, H.F., and Butterworth, D.S. 2001. Possible extensions to the ADAPT VPA model applied to western North Atlantic bluefin tuna, addressing in particular the need to account for "additional variance". Col. Vol. Sci. Pap. ICCAT, **52**: 1116–1129.

Harley, S.J., Myers, R.A., and Dunn, A. 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. **58**(9): 1760–1772. doi:10.1139/cjfas-58-9-1760.

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. 1999. Bayesian model averaging: a tutorial. Stat. Sci. **14**: 382–417.

Lo, N.C.H., Jacobson, L.D., and Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. **49**(12): 2515–2526. doi:10.1139/f92-278.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. **10**(4): 325–337. doi:10.1023/A:1008929526011.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. **70**(2–3): 141–159. doi:10.1016/j.fishres.2004.08.002.

Mazzetta, C., Brooks, S., and Freeman, S.N. 2007. On smoothing trends in population index modeling. Biometrics, **63**(4): 1007–1014. PMID:17501945.

Nichols, J.D., Thomas, L., and Conn, P.B. 2009. Inferences about landbird abundance from count data: recent advances and future directions. *In* Modeling demographic processes in marked populations. *Edited by* D.L. Thomson, E.G. Cooch, and M.J. Conroy. Springer Environmental and Ecological Statistics Series, Vol. 3, New York. pp. 201–236.

Pennington, M. 1985. Estimating the relative abundance of fish from a series of trawl surveys. Biometrics, **41**(1): 197–202. doi:10.2307/2530654.

Poulard, J.C., and Trenkel, V.M. 2007. Do survey design and wind conditions influence survey indices? Can. J. Fish. Aquat. Sci. **64**(11): 1551–1562. doi:10.1139/F07-123.

Prager, M.H. 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. (Washington, D.C.), **92**: 374–389.

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

R Development Core Team. 2007. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org [accessed 10 January 2009].

Schnute, J.T., and Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessment. Can. J. Fish. Aquat. Sci. **50**(9): 1916–1923. doi:10.1139/f93-214.

SouthEast Data, Assessment, and Review. 2008*a*. SEDAR 17 Data Workshop Report [online]. Available at http://www.sefsc.noaa.gov/sedar/ [accessed 10 January 2009].

SouthEast Data, Assessment, and Review. 2008*b*. SEDAR 17 Assessment Workshop Report [online]. Available at http://www.sefsc.noaa.gov/sedar/ [accessed 10 January 2009].

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. B, **64**(4): 583–639. doi:10.1111/1467-9868.00353.

Spiegelhalter, D.J., Thomas, A., Best, N., and Lunn, D. 2003. WinBUGS user manual. Version 1.4 [online]. MRC Biostatics Unit, Cambridge, UK. Available at http://www.mrc-bsu.cam.ac.uk/bugs [accessed 10 January 2009].

Stefánsson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J. Mar. Sci. **53**(3): 577–588. doi:10.1006/jmsc.1996.0079.

Stockhausen, W.T., and Fogarty, M.J. 2007. Removing observational noise from fishery-independent time series data using ARIMA models. Fish. Bull. (Washington, D.C.), **107**: 88–101.

Sturtz, S., Ligges, U., and Gelman, A. 2005. R2WinBUGS: a package for running WinBUGS from R. J. Stat. Softw. **12**: 1–16.

Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. 4th ed. Springer, New York.

Wade, P.R. 2002. A Bayesian stock assessment of the eastern Pacific gray whale using abundance and harvest data from 1967–1996. J. Cetacean Res. Manag. **4**: 85–98.

Wilberg, M.J., and Bence, J.R. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Can. J. Fish. Aquat. Sci. **63**(10): 2275–2285. doi:10.1139/F06-111.