Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data

Andrew O. Shelton, E.J. Dick, Donald E. Pearson, Stephen Ralston, and Marc Mangel

Abstract: Accurate landing statistics are among the most important data for the management of sustainable fisheries. For many fisheries, however, estimating species-specific landings and the associated uncertainty can be difficult, especially in the case of complex multispecies fisheries. Here we develop general and flexible methods for estimating species-specific landings, motivated by the mixed-species California groundfish fishery. We describe Bayesian generalized linear and hierarchical models for estimating species compositions from port sampling data and illustrate the application of each to several examples from California fisheries. Our hierarchical modeling approach provides a coherent statistical framework that can provide estimates of landings and uncertainty in the face of sparse and missing sampling data that compliment existing procedures for estimating landings. Furthermore, our methods provide ways to compare alternative model formulations and to maintain estimates of uncertainty when landings are aggregated across temporal or spatial scales. Our model structure is applicable to fisheries worldwide.

Résumé : Les statistiques précises sur les débarquements sont parmi les données les plus importantes pour la gestion des pêches durables. Pour plusieurs pêches, cependant, l'estimation des débarquements en fonction des espèces et l'incertitude qui leur est associée peut s'avérer difficile, particulièrement dans le cas de pêches complexes et multispécifiques. Nous mettons au point ici des méthodes générales et flexibles pour estimer les débarquements en fonction des espèces, à cause de notre intérêt pour la pêche multispécifique des poissons de fond de la Californie. Nous décrivons des modèles bayésiens généralisés linéaires et hiérarchiques pour estimer la composition en espèces à partir de données d'échantillonnage dans les ports; nous illustrons l'utilisation de chacun des modèles avec plusieurs exemples provenant des pêches de Californie. Notre méthodologie de modélisation hiérarchique fournit un cadre statistiquement cohérent qui peut produire des estimations des débarquements ainsi que de l'incertitude dans les cas de données d'échantillonnage rares ou manquantes; elle sert donc de complément aux procédures actuelles d'estimation des débarquements. De plus, nos méthodes fournissent des façons de comparer les différentes formulations des modèles et de maintenir les estimations de l'incertitude lorsque les débarquements sont répartis de manière contagieuse aux échelles temporelles ou spatiales. Notre structure de modélisation peut s'appliquer aux pêches à l'échelle mondiale.

[Traduit par la Rédaction]

Introduction

Historical landing records and catch–effort statistics derived from landing records form the backbone of most fisheries stock assessments. Therefore, the accuracy and reliability of landings statistics underlie sustainable management of fisheries for exploited species. Furthermore, landing records are some of the longest and most widely available

A.O. Shelton. Center for Stock Assessment Research, University of California, Santa Cruz, Mail Stop SOE-2, Santa Cruz, CA 95064, USA.
E.J. Dick and S. Ralston. Center for Stock Assessment Research, University of California, Santa Cruz, Mail Stop SOE-2, Santa Cruz, CA 95064, USA; Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 110 Shaffer Road, Santa Cruz, CA 95060, USA.
D.E. Pearson. Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 110 Shaffer Road, Santa Cruz, CA 95060, USA.
M. Mangel. Center for Stock Assessment Research, University of California, Santa Cruz, Mail Stop SOE-2, Santa Cruz, CA 95064, USA; Department of Applied Mathematics and Statistics, Jack Baskin School of Engineering, University of California, Santa Cruz, Mail Stop SOE-2, Santa Cruz, CA 95064, USA; Department of Biology, University of Bergen, Bergen, NO-5007, Norway.

Corresponding author: Andrew O. Shelton (e-mail: ashelton@soe.ucsc.edu).

quantitative records of marine species and are frequently used in models of oceanic ecosystems with the aim of understanding the ecosystem level consequences of fishing (e.g., Pauly et al. 2000). However, landings data frequently are riddled with unforeseen biases and limitations that can restrict their usefulness in both fisheries and ecological analyses (Cotter and Pilling 2007; de Mutsert et al. 2008). For example, landings estimates may be biased by the process in which they are sampled (e.g., nonrandom sampling) or by applying inappropriate statistical models to the data.

In many stock assessments and ecosystem models, analysts assume that landings are known with little or no uncertainty. Although many stock assessment programs have approaches that can be used to assess the effect of uncertainty in landings for the stock assessment (e.g., Stock Synthesis; Methot 2000, 2007), few regulatory agencies estimate, report, or incorporate landings uncertainty into stock assessments. However, uncertainty is present in estimated landings and needs to be documented and considered in fisheries management.

Landings in the complex, multispecies fisheries common in the world's oceans present a range of challenges. In some fisheries, reported landings represent the aggregations of multiple species, and landings must be sampled to disentangle the species composition. In others, multiple boat types and (or) fishing methods target a single fish stock, but landings for each fishing method may differ. Furthermore, landings may vary in space, time, or with other covariates. That landings can vary with many covariates often necessitates a stratified sampling regime to appropriately estimate landings (Cotter and Pilling 2007). However, stratified sampling programs can require an enormous amount of time and funding to ensure that adequate samples are obtained. This general problem gives rise to two closely related topics in sampling fisheries landings: (*i*) how to optimally allocate available resources to ensure adequate sampling that produces precise, unbiased estimates (e.g., Sen 1986; Crone 1995; Miller et al. 2007) and (*ii*) how to use the data that were actually collected to make point and uncertainty estimates of landings. A further difficulty for incorporating uncertainty in landings is that they are often aggregated across temporal and spatial scales (e.g., across multiple ports or seasons), and it is often unclear how to appropriately propagate uncertainty with this process of aggregation.

We present a general and flexible method that addresses a common challenge in the estimation of landings: sparse sampling of a stratified sampling regime. We develop techniques for estimating species-specific landings from aggregated multispecies landings with stratified sampling data. We use hierarchical Bayesian methods for generating point estimates and uncertainty bounds of species composition and translate these into species-specific landing estimates. Our methods can be employed when sampling data are available for all sampling strata and for predicting landings for strata with no sampling data. Our study is motivated by the mixed-species California groundfish fishery, but our methods are very general.

In this paper, we (*i*) describe statistical challenges posed by multispecies landings data in general and the California groundfish fishery in particular, (*ii*) outline Bayesian statistical models for mixed-species landings data, and (*iii*) apply our approaches to examples from California groundfish landings to illustrate model implementation and the power of our approach.

An example: California groundfish fisheries and landings sampling

A major component of California's groundfish fishery targets the state's diverse assemblage of rockfishes (genus *Sebastes*). Fifty-six rockfish species have been observed in the commercial fishery landings since 1969 (Pearson et al. 2008). Some non-rockfish species are caught in the rockfish fishery (e.g., sablefish, *Anoplopoma fimbria*) but generally contribute relatively little to total landings. Combined rockfish landings peaked in the early 1980s at greater than 25 000 tonnes (t) annually before declining rapidly through the 1990s and 2000s (Pearson et al. 2008). Currently, California rockfish landings from the directed fishery amount to ~1800 t per annum.

Many landings of rockfish in California are not sorted by species when landed. Although regulation requires some species to be sorted and landed into specified market categories, most species are sorted at sea or at the dock into mixed species "market categories" and the pounds landed in each market category are reported to the California Department of Fish and Game (CDFG) on landing receipts ("fish tickets"). The process of sorting species into market categories is a complex process determined by the fishermen's sorting of the catch. The species composition of a given market category is therefore driven by fish size, price, and purchaser preference among other considerations. Such aggregated landings present an obstacle for documenting historical fisheries landings and species abundances. A persistent challenge has been determining how to connect the data from market category landings to landings for individual species. Since 1978, the California Cooperative Groundfish Survey (CCGS) has sampled rockfish landings to identify the rockfish species that comprise each market category and estimated the pounds for each species landed within each market category (Sen 1984, 1986; see also Crone 1995). The California sampling protocol is a "fleet targeted" program, not a "stock targeted" program (sensu Cotter and Pilling 2007); landings for individual species are estimated from a mixed-species fishery. A single species may occur in multiple market categories. An important aspect of sampling is that it focuses on market categories rather than boat trips. Samples from the same market category from multiple boats are used to estimate the species composition in a market category. Individual landings of a market category are treated as independent sampling units. Because all sampling and landing data are reported in pounds, we report all mass in pounds (1 pound (lb) = 0.4536 kg).

Originally, protocols for sampling rockfish in California were derived by Sen (1984, 1986) as design-based sampling. Detailed descriptions of the sampling program and current process for expanding species composition estimates to species-specific landings can be found elsewhere (Sen 1984, 1986; Pearson and Erwin 1997). We briefly summarize the sampling process here. Sampling follows a two-stage stratified sampling plan (Sen 1986). Vessels are selected arbitrarily by the port sampler and two 50 or 25 lb clusters of fish are selected arbitrarily from bin(s) containing a single market category. The mass sampled depended on the size of individual fish in the landings, with landings consisting of small fish having 25 lbs sampled and those consisting of large fish hav-

ing 50 lbs sampled. The mass of each species in the clusters are recorded. The mass sampled relative to the total landing mass is highly variable. In some cases, the entire catch is sampled (e.g., some hook-and-line landings), whereas in others (e.g., trawl landings), the proportion of landed mass sample is very small. However, as the target population of port sampling is the total landed mass in each market category, the proportion of landed mass that is sampled is very small.

Multiple market categories are often landed by a single boat, and samples from several different market categories may be collected from a given landing. The species present in a market category are known to vary with location (e.g., the port; landings are grouped into 10 major port complexes in California that span ~800 miles (1 mile = ~1.61 km) of coastline), time (year and the four quarters within the year), fishing technique (trawl, hook and line, gillnet, fish pot, or other minor categories), and if the fish are landed alive or dead. Each combination of market category, location, time, fishing technique, and alive–dead comprise a stratum, and the many categories result in a very large number of potential strata. Annually, ~1500 strata have landings reported.

Multiple samples of a market category within a stratum are combined to generate a point estimate of the species composition for each stratum (see Sen 1984, 1986). This species composition is multiplied by the pounds reported by fish tickets of landings in a stratum to generate the poundage landed for each species (Sen 1986; Crone 1995). Although every effort is made to sample all landed strata, the number of landed strata is large and port sampling personnel are few, so some strata go unsampled. The difficulty in sampling strata is exaggerated by California regulations. California does not have mandatory sampling of landings and so fisherman can lawfully refuse to allow sampling of catches. In recent years, only ~25% of strata with documented landings have at least one sample recorded (CALCOM (California Cooperative Goundfish Survey) 2010). However this figure overstates the sparsity of sampling landings data, as many strata with landings have only a very small poundage landed and as such are fairly inconsequential contributors to the total landing mass (Pearson and Almany 1995; CALCOM (California Cooperative Goundfish Survey) 2010). The fraction of samples within a stratum is typically strongly correlated with the mass of landings in that stratum (Pearson and Almany 1995; CALCOM (California Cooperative Goundfish Survey) 2010).

Currently, the CCGS does not directly quantify the uncertainty associated with species landings. This is at least in part a result of missing observations; it is unclear how to estimate uncertainty if no samples were observed from a stratum. At present, the CCGS uses a system of "borrowing" information to provide point estimates of species compositions for unobserved strata. Strata with species composition data are used to fill in species compositions for unobserved strata (Jermyn and Robb 1981). For example, if quarter 2 for a given market category, port, and fishing gear type had sampling data and quarter 3 with the same categorical variables was unobserved, the estimated species composition for quarter 2 is used for quarter 3. Such borrowing procedures beg the questions: which observed strata estimates should be borrowed for each unobserved stratum, and how certain are we about the

species composition of unobserved strata relative to strata with sampling data? The current system of borrowing is based on expert opinion and a poorly justified method for producing estimates of species landings (Pearson and Erwin 1997). For the California rockfish fisheries, we are most interested in developing better methods that can provide point and uncertainty estimates of species composition in both sampled and unsampled strata.

The general philosophy of this paper is to combine aspects of design-based and model-based sampling approaches to landings estimates. The behavior of design-based estimates landings are derived from normal sampling theory that relies on the central limit theorem; under repeated sampling, the means of independent samples converge to a normal distribution with well known statistical properties that are discussed in most basic texts on sampling theory (e.g., Cochran 1977). Design-based sampling protocols perform excellently when the strata are defined appropriately, collected samples are independent, and enough samples are collected in each strata. The design-based approach to sampling can face some difficulties in practice, however. First, if a stratum does not receive enough samples (e.g., less than the minimum required number of samples are collected), estimates derived from analytic expressions for the mean and variance may be biased or, in the case that ≤1 samples were collected, no estimates of variance are possible.

Second, in classical sampling theory, strata are assumed to be independent; sampling information taken from a given stratum has no influence on other strata. This approach makes a great deal of sense if the strata delineations are obvious. In practice, though, the best way to impose sampling strata may be unclear, though optimal methods for allocating sampling effort is an important area of statistical research (e.g., Miller et al. 2007). In the California groundfish fishery, for example, assigning the different fishing gear types to different categories is viewed as easily justified (e.g., trawl and hook-and-line gear do target and catch different species). However, the temporal division of the year into four quarters is not the only possible way to divide the calender. Compelling arguments can be made for dividing the year into single-month blocks, quarters, or half-year blocks, among other scenarios. The broader point is that for a given sampling scheme, it is reasonable to consider some groups of strata more similar to each other, and treating each stratum as independent is not always best. In the case that few or no samples are available, we would like to estimate the relatedness among strata and use this relatedness to make statistically justifiable estimates of landings in undersampled strata.

From a practical standpoint, the current design-based sampling design has been implemented for over 30 years, and we need to make use of the available data and statistical methods to provide the best possible point estimates and uncertainty bound of historical landings for rockfish in California. An ongoing avenue of research is how to modify and simplify the sampling protocol to reduce the number of strata and thereby improve the accuracy and precision of landings estimates in the future. However, improving the sampling regime moving forward does not provide guidance for how to make best use of available data from the past 30 years.

We develop an alternative method for estimating the species composition of landings and their uncertainty under

sparse sampling. We accomplish this by recognizing that species compositions are related across some strata and model these relationships to predict landings in strata for which direct samples are unavailable. We first build a generalized linear model for the case in which sampling data are available for each stratum and then extend our model to create a hierarchical model that can provide predictions for unsampled strata.

Methods and results

Generalized linear model

To introduce the modeling framework, we first develop a basic generalized linear model that calculates the species composition for a single stratum. Fundamentally, we are interested in the total landings for each species within a given stratum. By sampling some of the landings from individual boats, we can estimate the species composition of the landings and then combine species composition estimates with independent estimates of total landed mass to estimate the landings of each species. The basic model underlying our statistical approach is that fishing boats land pounds, X_j, for $j = 1,\ldots, J$ total species within a given stratum. Some portion of the landing is sampled and Y_j discrete pounds of species j are observed; $Y_j \geq 0$. Although pounds landed is a continuous quantity, landings are reported in discrete pounds. We use landings mass in pounds as our sampling unit and use discrete distributions to describe landings; landings are assumed to occur in discrete units of mass. From these sampling data, we estimate the proportion of landings that arise from each species, π_j, and use π_j to estimate X_j in a stratum. Throughout, capital letters indicate random variables, lowercase letters indicate parameters or realizations of the random variables, and bold symbols indicate vectors or matrices.

For simplicity, we start by estimating the species composition with no categorical covariates; we estimate the species composition in single market category in a single port for a given fishing gear type. Initially, we also ignore within-year variation and consider the case in which only a single landing was sampled during the year. We model this situation as

$$(1) \qquad Y_j \sim \text{Poisson}(e^{\theta_j})$$
$$\theta_j = \lambda + \beta_j$$

where λ is a scaling parameter shared by all species, and the β_js are species-specific parameters (note that λ and β_js are numbers on the real line, not proportions). As explained above, we assume that the mass of fish in pounds is the unit of interest. The only unfamiliar formulation in eq. 1 is the e^{θ_j} term. Writing the model in this way ensures that the Poisson parameter is positive for all values of θ_j and allows great flexibility in modeling λ and β_j. This formulation is known as the multinomial–Poisson transformation (Baker 1994) because the pounds sampled for each species are treated as Poisson random variables linked by the shared parameter λ so that inference for species composition proportions is equivalent to using a single multinomial distribution.

We observe y_j pounds of species j and $n = \sum_{j=1}^{J} y_j$ total pounds sampled. The likelihood of the vector of species-specific parameters, θ, for observed data vector y is proportional to

$$(2) \qquad p(y|\theta) \propto \prod_{j=1}^{J}(e^{\theta_j})^{y_j}\exp(-e^{\theta_j}) =$$
$$\exp\left(n\lambda + \sum_{j=1}^{J}\beta_j y_j - e^{\lambda}\sum_{j=1}^{J}e^{\beta_j}\right)$$

Throughout, we use "exp" and "e" interchangeably to indicate the exponential function. To ensure that λ is identifiable, we set $\beta_1 = 0$. Baker (1994) demonstrated that maximizing eq. 2 over λ and the vector of β_js, β, yields identical maximum likelihood estimates and asymptotic variances as using a multinomial likelihood. Therefore, λ serves to scale the likelihood such that the estimated proportion of species j in a given stratum is

$$(3) \qquad \widehat{\pi}_j = \frac{e^{\theta_j}}{\displaystyle\sum_{j=1}^{J}e^{\theta_j}}$$

and $\sum_{j=1}^{J} \widehat{\pi}_j = 1$. The estimated species composition in a given strata is the same as if it were estimated from a multinomial model and has the same properties: a larger proportion estimated for one species necessarily produces a smaller proportion for the other species. The advantage of the formulation in eq. 1 is that it provides a linear structure in which to formulate the multinomial model, as well as providing computational advantages that we describe below.

For Q independent samples of size n_q (for California rockfish, independent samples correspond to an individual boat's landing of a market category within a stratum; n_q can vary among samples), the likelihood becomes

$$(4) \qquad p(y|\lambda, \beta) \propto \prod_{q=1}^{Q}\exp\left(n_q\lambda + \sum_{j=1}^{J}\beta_j y_{jq} - e^{\lambda}\sum_{j=1}^{J}e^{\beta_j}\right) =$$
$$\exp\left(\lambda\sum_{q=1}^{Q}n_q + \sum_{j=1}^{J}\beta_j\sum_{q=1}^{Q}y_{jq} - Qe^{\lambda}\sum_{j=1}^{J}e^{\beta_j}\right)$$

where y_{jq} is the observed mass of species j in the qth sampled landing. Because $\sum_{q=1}^{Q} n_q$ is the total pounds sampled in a market category and $\sum_{q=1}^{Q} y_{jq}$ is the number of pounds observed for species j in a market category, the likelihood depends only on the sum of pounds for each species and the total number of pounds. We implicitly assume in this model formulation that each sampled landing is an independent sample of a single, multinomial distributed species composition. If the pounds sampled from each landing is approximately constant, as in the California rockfish fishery, each sampled landing will contribute roughly equal information to the estimated species composition.

We develop a Bayesian model for eq. 1 not for the express purpose of including prior information, but rather because this framework facilitates extension to the hierarchical models in later sections. Defining $p(\lambda, \beta|y)$ as the posterior distribution for the parameters, $p(\lambda)$ and $p(\beta)$ as independent prior distributions for λ and β, and $p(y|\lambda, \beta)$ as the likelihood,

$$(5) \qquad p(\lambda, \beta|y) \propto p(\lambda)p(\beta)p(y|\lambda, \beta)$$

We use normal priors for λ and each element of β; $p(\lambda) \sim N(\mu_\lambda, \sigma_\lambda)$ and $p(\beta_j) \sim N(\mu_{\beta_j}, \sigma_{\beta_j})$. Using large values of σ_λ

Fig. 1. Estimated species composition for California groundfish landings. Boxplots show median, interquartile, and 95% credible intervals of 1000 samples from the posterior distribution of species compostitions. Maximum likelihood estimates (MLE) for each species are shown ($\Diamond$). Species names are shortened to four letter abbreviations; full species names can be found in the Supplementary materials.[1] (*a*) Species compositions from 1990 trawl landings in market category 956 at a port with 149 pounds and four landings sampled; (*b*) species compositions from longline landings from 1995 in market category 250 at a port with 2880 pounds and 49 landings sampled.
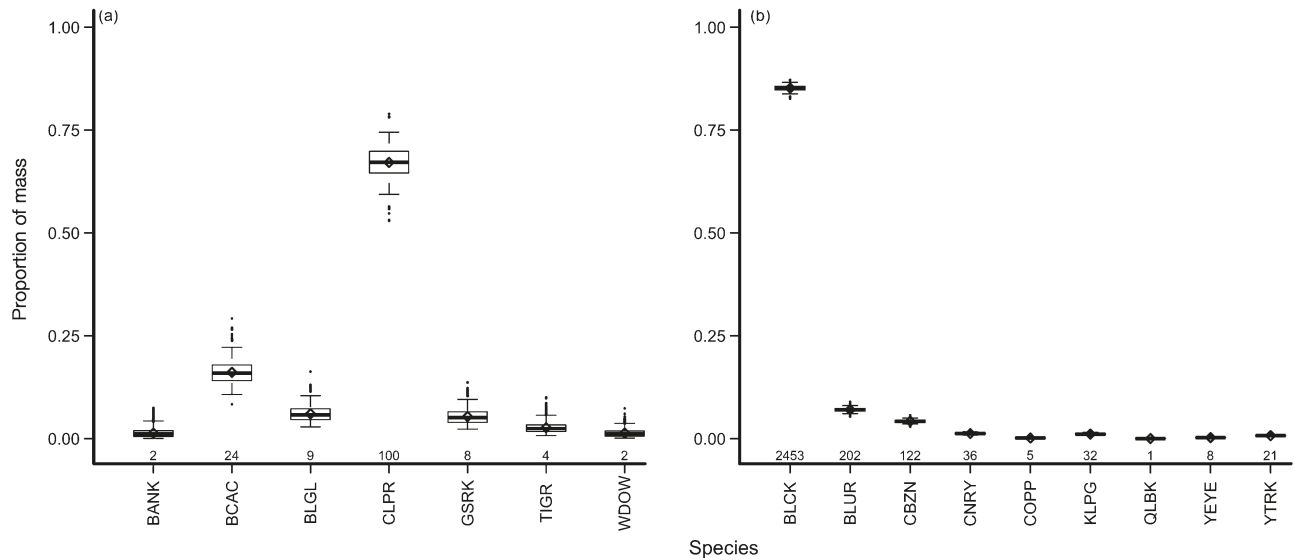
and σ_{β_j} (e.g., $\sigma_\lambda = 1000$) provides uninformative priors for the parameters (see Supplementary materials).[1] To sample the posterior distribution using Markov chain Monte Carlo (MCMC) techniques, we write down the conditional posterior distributions for λ and all β_j. Retaining only terms that include λ, μ_λ, and σ_λ, the conditional posterior distribution for λ is

$$(6) \quad p(\lambda|\mathbf{y}, \boldsymbol{\beta}) \propto \frac{1}{\sigma_\lambda} \exp\left(-\frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2\right) \times \exp\left(\lambda \sum_{q=1}^{Q} n_q - Q e^\lambda \sum_{j=1}^{J} e^{\beta_j}\right)$$

Similarly, the conditional posterior for each β_j with relevant terms retained is

$$(7) \quad p(\beta_j|\mathbf{y}, \lambda) \propto \frac{1}{\sigma_{\beta_j}} \exp\left(-\frac{1}{2\sigma_{\beta_j}^2}(\beta_j - \mu_{\beta_j})^2\right) \times \exp\left(\beta_j \sum_{q=1}^{Q} y_{jq} - Q e^\lambda e^{\beta_j}\right)$$

Note that the conditional posterior for λ includes only total pounds sampled in a market category, $\sum_{q=1}^{Q} n_q$, and the number of independent landings sampled, Q, whereas the conditional posterior for β_j includes only the observed count for each species and the number of trips sampled. In the Supplementary materials,[1] we describe and provide R code for sampling from these conditional posterior distributions using a Metropolis–Hastings algorithm (Gelman et al. 2004). Posterior samples, after MCMC burn-in and appropriate thinning of the MCMC chain, provide both point estimates and uncertainty bounds for species composition in a given market cate-

gory. Specifically, each retained iteration of the MCMC chain is a draw from the joint posterior $p(\lambda, \boldsymbol{\beta}|\mathbf{y})$, which can be converted into species compositions, $\hat{\pi}_j$, using eq. 3. A large number of draws from the posterior will produce a distribution of species compositions. With many posterior samples and diffuse prior distributions for the parameters, the mean of these posterior draws converge to the multinomial maximum likelihood estimate and provide a description of the uncertainty associated with species composition. As will be seen in examples of the application of our models in subsequent sections, the Bayesian approach provides additional advantages, most notably the ability to generate predictive distributions. We present results that illustrate the correspondence between the maximum likelihood estimates and eq. 4 using two examples of landings of California rockfish (Fig. 1). Posterior estimates correspond directly to the maximum likelihood estimates, and as expected, uncertainty estimates for species composition decline as the total pounds sampled increases.

Extending this basic multinomial–Poisson to more than one categorical variable is straightforward. For example, a model to estimate the proportion of each species in a single market category, quarter, and port in California, where Y_{ijk} are discrete counts of sample pounds for species j in port k in quarter i, is

$$(8) \quad Y_{ijk} \sim \text{Poisson}(e^{\theta_{ijk}})$$
$$\theta_{ijk} = \lambda + \beta_j + \gamma_{jk} + \eta_{ijk}$$

where γ_{jk} is the port-specific coefficient for each species, and η_{ijk} is the quarter-specific coefficient for each port in each species. To ensure parameter identifiability, for the first port observed for each species, $\gamma_{j1} = 0$, and for the first quarter in each port, $\eta_{1j1} = 0$. This formulation assumes that coeffi-

---

[1]Supplementary materials are available with the article through the journal Web site at http://researchpress.com/doi/suppl/10.1139/f2011-152.

cients for ports and quarters are nested within each species and not independent of the species effects. In preliminary attempts to model species compositions, we attempted to use a non-nested parameter formulation that included port, species, and quarter effects, as well as various interactions terms. We found that the nested parameterization provided better descriptions of the California landings data. In particular, the nested formulation avoided the possibility of a nonzero probability of a species if it was never observed in the sampling data from a given port and quarter. In other applications, however, a non-nested parameterization might be advantageous. If we wished to ensure positive posterior proportions for some species even though they were not observed in the samples, we could include those species in the prior distributions of β, γ, or η. Incorporating prior information should be particularly useful when managers think that rare species are likely to be unaccounted for or there are particular species of management concern. Thus, the potential for including prior information is another advantage of the Bayesian framework. The likelihood, joint posterior distribution, and conditional posterior distribution for eq. 8 are very similar to those derived for eq. 1. For completeness, we present the conditional posterior distributions for this model in the Supplementary materials.[1]

Hierarchical model

Although the previous section provides a method to estimate the posterior distribution of species compositions when samples are available for all strata, it provides no guidance when a stratum is unsampled. In this section, we develop a hierarchical extension of our model that shares information among strata and can predict the species composition in the face of unsampled strata. Currently, the CCGS divides the year into three-month quarters corresponding to calender months January to March, April to June, July to September, and October to December (quarters 1, 2, 3, and 4, respectively; hereafter Q1, Q2, Q3, and Q4). The year is subdivided because the behavior of fishing boats changes somewhat with the season — e.g., fishing boats tend to not travel as far off shore during the winter — and fish perform seasonal migrations; the areas fished and the species composition of the landings are expected to change accordingly. However, there is good evidence that the species composition among quarters can be similar (Pearson and Erwin 1997) and we represent this relation among quarters hierarchically.

For simplicity, we begin by writing a model for species composition in a market category in a single year, port, and gear type. We include the quarter of the year as a hierarchical effect. We divide our sampling data into four quarters, so the data, Y_{ij}, are discrete pounds for species j in quarter i,

$$(9) \quad Y_{ij} \sim \text{Poisson}(e^{\theta_{ij}})$$
$$\theta_{ij} = \lambda + \beta_j + \eta_{ij}$$
$$\eta_{ij} \sim \text{N}(0, \nu_j)$$

Then η_{ij} characterizes among-quarter variation in the abundance of species j and is drawn from a normal distribution with mean 0 and species-specific standard deviation ν_j. Thus, we treat observed species counts from each quarter as exchangeable samples from a single, annual multinomial species-composition distribution with quarter-specific coeffi-

cients being drawn from a normal distribution for each species. Generally, the addition of hierarchical effects allows for overdispersion relative to the Poisson distribution. A common way to model overdistribution in count data is to use a negative-binomial distribution instead of a Poisson distribution. One way in which the negative binomial distribution arises is by assuming that the rate parameter of the Poisson distribution follows a gamma distribution (given certain constraints; Venables and Ripley 2002). Equation 9 describes a Poisson distribution in which the rate parameter follows a lognormal distribution. The properties of gamma and lognormal distributions are very similar, so eq. 9 produces a distribution qualitatively similar to a negative-binomial distribution.

An open question in Bayesian statistics is how to construct prior distributions for variance parameters in hierarchical models (Gelman 2006). After considering several prior distributions, we used gamma-distributed priors for all variance parameters, thus providing weakly informative prior information (Gelman 2006; see also Supplementary materials).[1] Hierarchical models are easier to sample by treating the θ_{ij} as unobserved, latent variables and sampling them as part of the MCMC algorithm. The full posterior for eq. 9 includes the vector of observed pounds, y, independent prior distributions of the parameters $p(\lambda)$, $p(\beta)$, and $p(\nu)$, the probability of latent variables given the parameters, $p(\theta | \lambda, \beta, \nu)$, and the likelihood $p(y | \theta)$.

$$(10) \quad p(\theta, \lambda, \beta, \nu | y) \propto p(\theta, \lambda, \beta, \nu)p(y | \theta, \lambda, \beta, \nu)$$
$$= p(\lambda)p(\beta)p(\nu)p(\theta | \lambda, \beta, \nu)p(y | \theta)$$

The complete conditional distribution for each parameter and procedures for sampling this model with standard Metropolis–Hastings MCMC approaches are available in the Supplementary materials,[1] which also contain example R code for estimating model parameters. Similar to the generalized linear model, we can extend this model to multiple ports. For example, a hierarchical version of eq. 8 is

$$(11) \quad Y_{ijk} \sim \text{Poisson}(e^{\theta_{ijk}})$$
$$\theta_{ijk} = \lambda + \beta_j + \gamma_{jk} + \eta_{ijk}$$
$$\eta_{ijk} \sim \text{N}(0, \nu_{jk})$$

As before, γ_{jk} is the port-specific coefficient for each species with K total ports. The hierarchical effect, η_{ijk}, assumes that quarter coefficients for each species in each port are normally distributed with species- and port-specific standard deviation, ν_{jk}.

The benefit of the hierarchical model over the generalized linear model is that it shares information across strata. It provides a way to predict species compositions for strata with no sampling data via the posterior predictive distribution (PPD). The PPD describes the predicted species composition for an unobserved stratum and arises directly from the hierarchical formulation. Specifically, for eq. 9, the PPD describes the rate parameter for species j in unobserved quarter z: $\theta_{zj} \sim \text{N}(\lambda + \beta_j, \nu_j)$. In practice, we estimate the PPD by drawing a single θ_{zj} for each MCMC sample of the posterior distributions of $\{\lambda, \beta_j, \nu_j\}$. This process produces a distribution of θ_{zj} for each species that is converted to species composition proportions (π_{zj}) using eq. 3. Thus, the

PPD provides a formal way of combining sampling information to generate species composition prediction for unobserved strata. We demonstrate the application and properties of the PPD below.

For the remainder of the paper, we analyze a series of example data scenarios from the California groundfish data using eq. 11 (hereafter the hierarchical model, HM) to illustrate its properties under a range of data situations. We emphasize the role of the PPD in providing species composition predictions for unsampled strata. Where appropriate, we compare the results from the HM with eq. 8 (hereafter the generalized linear model, GLM) and the multinomial maximum likelihood estimates (MLE). To ensure compliance with confidentiality requirements, in all examples, we omit the specific port from which sampling data were taken. All of the details for performing MCMC sampling, assessing model convergence, and sampling efficiency are described in detail in the Supplementary material.[1] All analyses were performed with R (version 2.11.1; R Development Core Team 2008).

Scenario 1: full sampling data

Initially, we consider a situation in which sampling data are available for each quarter with landings. Thus the species composition for each quarter can be directly estimated by both the GLM and HM. Under this data scenario, each posterior sample provides an estimate of the species composition for each species in each quarter (see eq. 3). We model the species composition for market category 253 for a single port (Fig. 2). We plot the MLEs and posterior estimates for each quarter derived from the GLM and the HM. There are three points of note. First, in a situation with abundant data for all quarters, both the GLM and HM produce similar estimated species proportions and they approximate MLE estimates well (Figs. 2a, 2b). Both the GLM and HM exhibit reduced uncertainty for quarters with abundant sampling data (compare estimates for Q4 with those for any other quarter).

Second, although the posterior mean of the GLM and the MLE coincide exactly, the posterior species compositions from the HM are pulled toward the posterior mean of the PPD (Fig. 2). Such shrinkage toward the posterior mean occurs because information on species composition is shared across quarters, with periods with more information (larger sample sizes) shrinking less toward the posterior mean than quarters with small sample sizes. Also, quarters with species composition very different from the PPD mean are affected more than quarters near the PPD mean. This result emphasizes the importance of regarding the quarters as exchangeable samples from a single annual species composition. The information contributed by each quarter is determined by the observed sample mass, not the mass landed; therefore if strata with small landings are disproportionately sampled, they will contribute disproportionately to the estimated species composition. This may or may not matter, depending on the application, but illustrates the importance of defining hierarchical models carefully to match biological and operational conditions. For California rockfish, the sampled masses are typically proportional to the landed mass strata; strata with large total landed mass are sampled more extensively than strata with small total landings (Pearson and Almany 1995; CALCOM (California Cooperative Groundfish Survey) 2010), and this issue is largely irrelevant. For illustration, we have plotted GLM and HM posterior species composition estimates for two species, bocaccio (BCAC; *Sebastes paucispinis*) and chilipepper (CLPR; *Sebastes goodei*) rockfish (Fig. 2d).

An important distinction between GLM and HM is the estimation of species composition for rare species. If a species is unobserved in a quarter (i.e., sampled mass = 0), the GLM and MLE estimate a species compositions of 0 for those species. However, because the HM shares information across quarters, as long as a species was observed at least once during the year, its estimated species composition will be nonzero for all quarters. For example, compare splitnose rockfish (SNOS; *Sebastes diploproa*) species composition (Fig. 2a versus Fig. 2b). Note that estimates of species composition for SNOS in Q1, Q2, and Q3 are zero in the MLE and GLM but nonzero in the HM.

In scenario 1, in which data are available for all quarters, we have no immediate use for the PPD to estimate unobserved strata. However, the PPD for species composition is centered between the quarter distributions with sampling data, and intuitively, the PPD generates species composition with increased uncertainty relative to quarters with sampling data (Fig. 2c).

Scenario 2: missing sampling data from one or two quarters

Another common situation is that sampling data are available for some but not all quarters. To illustrate this situation, we estimate both the HM or GLM from a year's worth of trawl sample data from market category 961 at a single port (Fig. 3). In this port, sample data are available for Q1, Q3, and Q4, but not for Q2. Rockfish were landed in this stratum in Q2, so an estimate of the species composition is necessary for estimating the pounds landed for each species. As in scenario 1, we can estimate the posterior species composition distributions with both the GLM and HM where data are available (Fig. 3). However, the question of how to provide an estimate of the species composition for Q2 remains. The only available solution with the GLM is to "borrow" species composition estimates from other quarters. In this case, species composition estimates would be borrowed from the nearest quarters with available sampling data (either Q1 or Q3 or by pooling across quarters). As frequently occurs, the estimated species compositions in Q1 and Q3 differ substantially (Fig. 3d), but there is no objective way to determine which species composition best approximates Q2 and should be borrowed. With the HM, however, the PPD provides a coherent estimate of the unobserved species composition for Q2 that incorporates information from all quarters with sampling data (Fig. 3c). Estimates from the PPD overlap both potential borrowed distributions, and the PPD have greater uncertainty than either borrowed GLM distributions, as is appropriate for quarters with no sampling data (Fig. 3d). Thus, the PPD provides a statistically well-justified method for estimating strata when direct sample data are unavailable. Furthermore, using a borrowed estimate of species composition will likely underestimate the variability present in the unobserved quarter. The procedure and inferences from the HM are virtually identical in situations in which two quarters have no sampling data.

238

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Fig. 2. Estimated species composition for market category 253 for trawl at a single port in California in 1981. Boxplots show median, inter-quartile, and 95% credible intervals of 1000 samples from the posterior distribution of species compostitions. "×" and horizontal bar show posterior predictive distribution (PPD) means; ◊ indicates the maximum likelihood estimates (MLE). (*a*) Species composition estimates for the hierarchical model (HM) for each quarter; (*b*) species composition estimates for the generalized linear model (GLM); (*c*) PPDs for the HM for each species; (*d*) detailed comparison of HM and GLM estimates for two species (BCAC and CLPR). Note how quarter-specific estimates are shrunk toward the cross-quarter mean species composition.
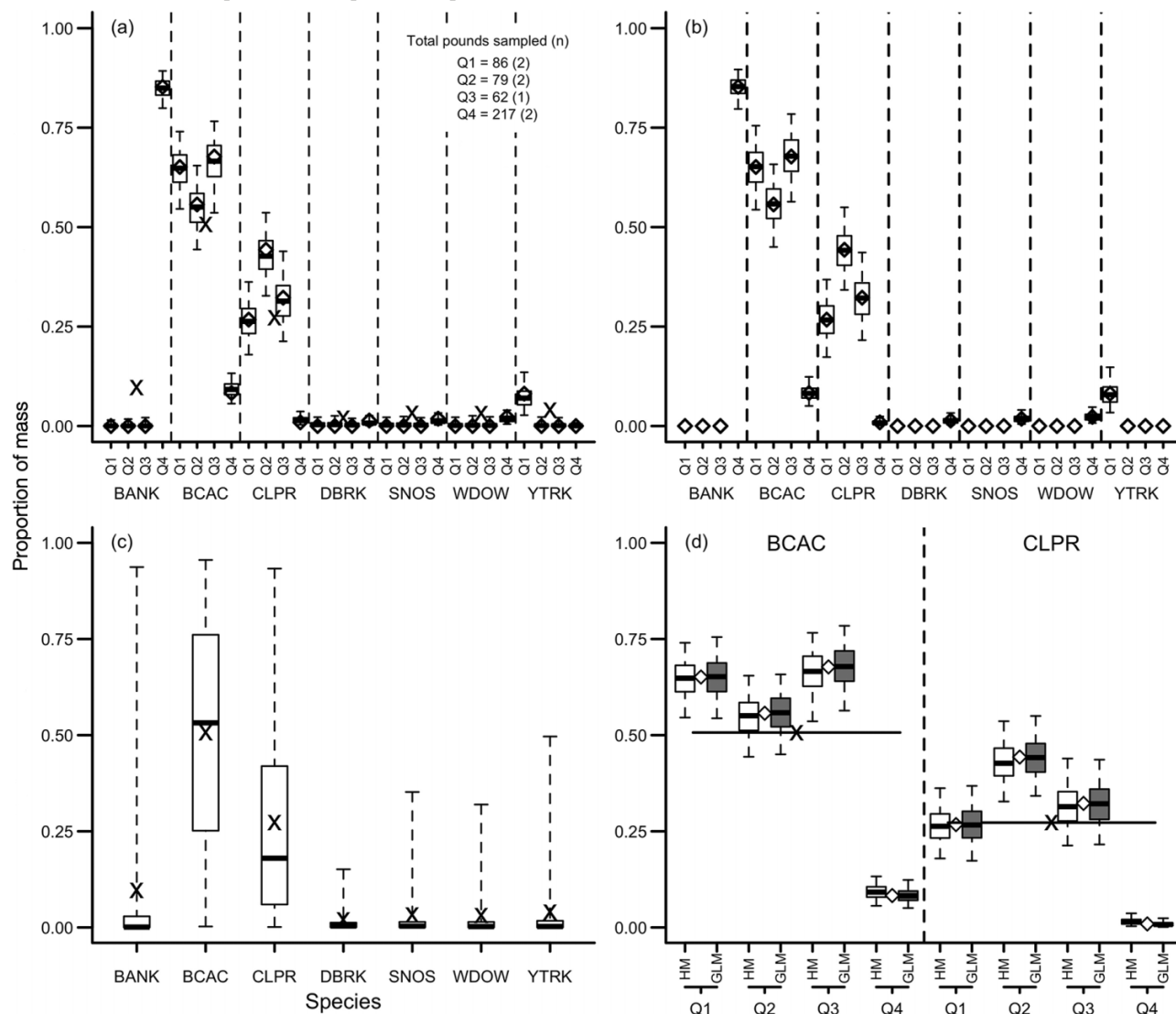

Scenario 3: data available from only one quarter

Another potential situation in the California groundfish landings is that sample data are available for a single quarter within a year (Fig. 4). The GLM from these data will provide the species composition in the single quarter with sampling data, but the HM fails to converge because the among-quarter variation parameter, v_{jk}, becomes unidentifiable. In such cases, there are easy options for estimating the species composition of unobserved quarters. First, analogously to the cur-rent CCGS procedure, species compositions could be directly borrowed across quarters from the posteriors of GLM. Sec-ond, the identifiability problem with the HM could be avoided by specifying a fixed value of v_{jk} and using this value to estimate the PPD for unobserved quarters. Alter-nately, we can provide a strong prior distribution to v_{jk}, most likely derived from a nearby port's estimate of v_{jk}. The use of

nondiffuse priors allows v_{jk} to vary slightly during the MCMC sampling procedure but produces results very similar to simply fixing v_{jk}. Both borrowing from the GLM or using the HM but fixing v_{jk} or specifying a strong prior are ad hoc solutions, but we suggest using a fixed value or strong prior on v that will produce a PPD that reflects increased uncer-tainty than that which would arise from directly borrowing from the GLM.

To illustrate this problem, we apply the GLM and HM models using fixed values for v_{jk} ($v_{jk} = 0.1$ to 2) to landings data for trawl gear in market category 956 from a single port in 1988. This range of v_{jk} approximates values estimated from strata in which all quarters are observed. Sampling data are only available for Q4 in this stratum, and thus we need a way to predict species composition for quarters 1, 2, and 3. Fitting the GLM and HM with a $v_{jk} = 100$ yields virtually

Fig. 3. Estimated species composition for market category 961 for trawl at a single port in California in 2005. Note that quarter 2 has no sampling data. Boxplots show median, interquartile, and 95% credible intervals of 1000 iid samples from the posterior distribution of species compostitions. (*a*) Species composition estimates for the hierarchical model (HM) for each quarter with sampling data; (*b*) species composition estimates from the generalized linear model (GLM) for each quarter with sampling data; (*c*) posterior predictive distributions (PPD) for the HM for each species; (*d*) comparing alternate estimates of the species composition for the quarter with missing data (Q2) for three species. The estimated species composition of the PPD is derived from the HM, whereas Q1 and Q3 are borrowed from the GLM.
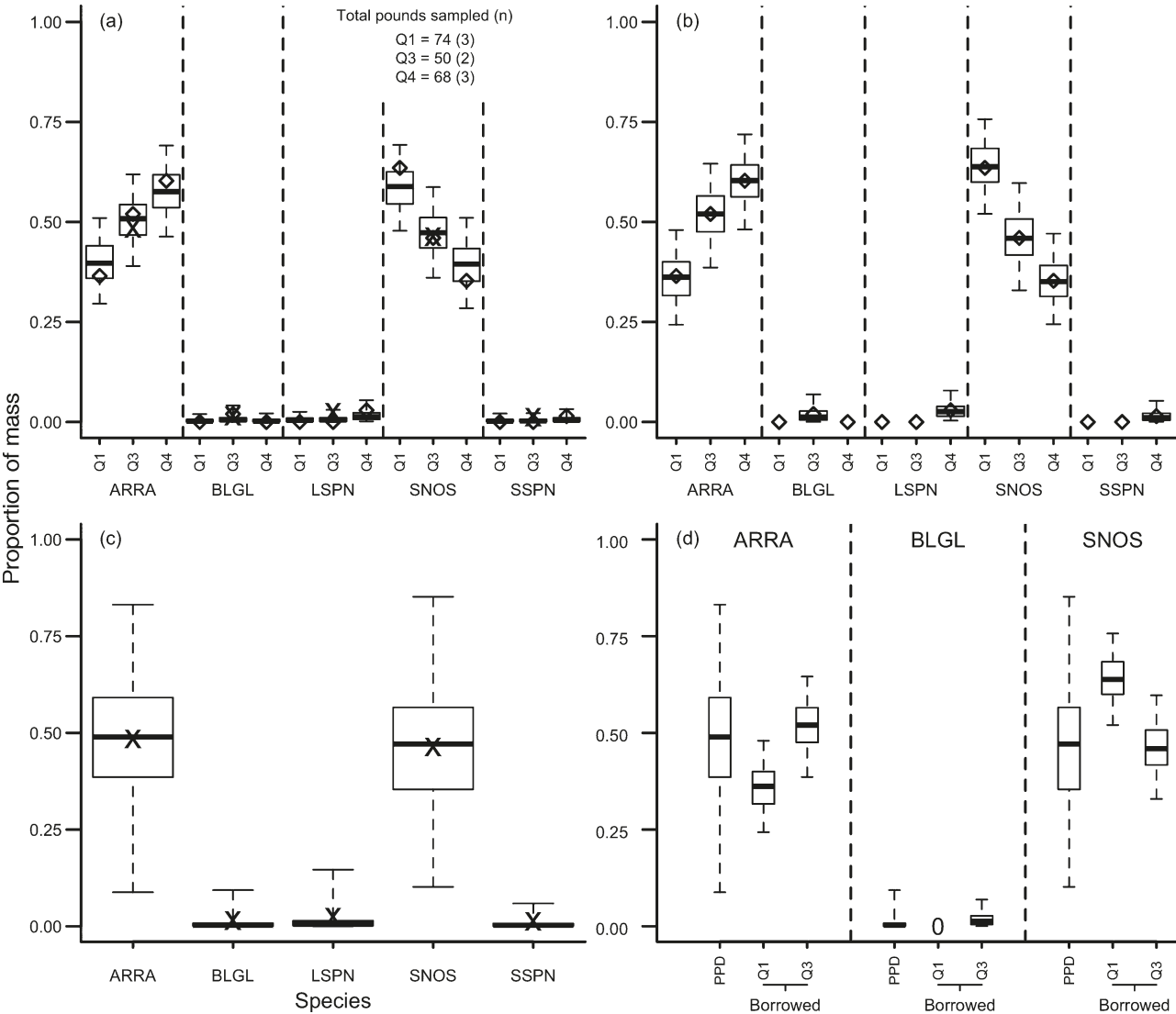
identical species composition estimates for the quarter with sampling data (Fig. 4*a*). Indeed, prespecifying any large value of ν_{jk} has little effect on the estimated species composition for the strata with sampling data (data not shown). Choosing a value for ν_{jk} has significant consequences for the PPD, however. For example, gradually increasing the value of ν_{jk} (from 0.1 to 2.0) for a single species composition estimate (Chilipepper rockfish; Fig. 4*b*) greatly expands the uncertainty of the PPD.

The two possibilities discussed above are not the only solutions to sparse data. One additional option would be to rewrite eq. 11 so that there are always sufficient samples to estimate a PPD for unobserved strata. For example, instead of treating quarters within a port hierarchically, we could combine California landing ports into geographical regions (e.g., northern, central, and southern California) and con-
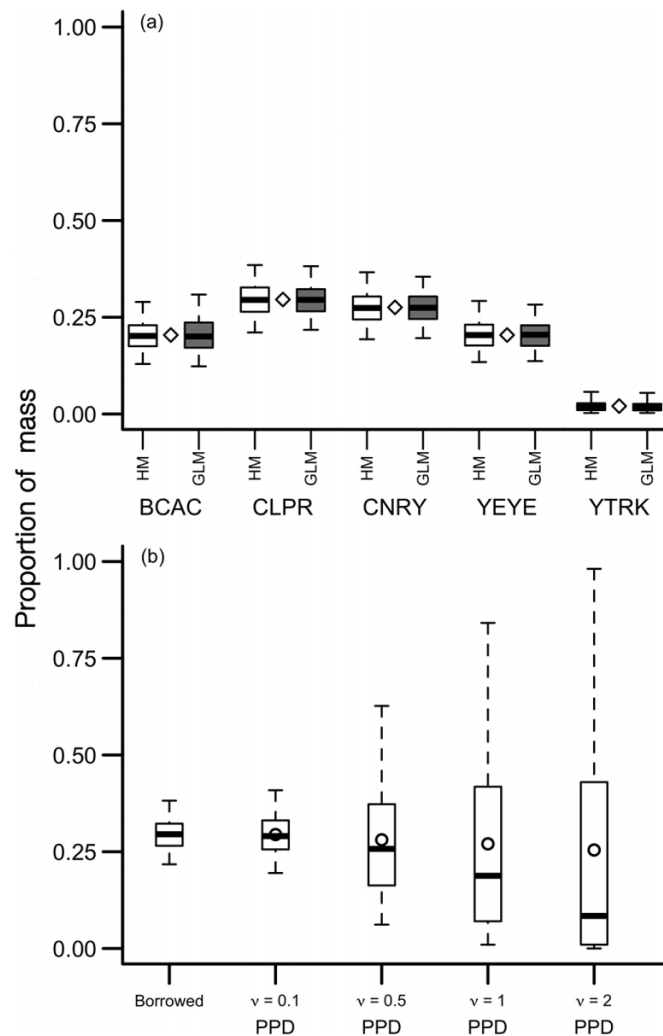
struct a hierarchical model using multiple parameters to share information within geographical regions and quarters. Care is needed when constructing such models to ensure that data represent exchangeable samples from a single species composition distribution. Indeed, respecifying the model is the only option when no samples are available for a market category in an entire year. We defer comparing such alternative landings models and how to choose a "best" model from available candidate models to future analyses.

Comparing GLM and HM

An important component of devising and applying new statistical models is comparing the relative effectiveness of various models at describing available data. In this section, we illustrate a method for comparing the GLM and HM for California landings. An exhaustive comparison of the GLM

Fig. 4. Estimated species composition for trawl landings in market category 956 in 1988 at a single port in California. Only a single quarter of sampling data was available for this port and market category. Boxplots show median, interquartile, and 95% credible intervals of 2000 samples from the posterior distribution of species compositions. Maximum likelihood estimates (MLE; ◇) and posterior predictive distribution (PPD; ○) means are shown. (*a*) Comparing species composition estimates from the hierarchical model (HM; open) and generalized linear model (GLM; shaded) for the single quarter with sampling data; to fit the HM, we fixed $\nu_{jk} = 100$; (*b*) alternative species composition estimates for CLPR (chilipepper rockfish) for the unobserved quarters (PPD) for the HM for each species under different fixed values of among-quarter variation, ν). "Borrowed" indicates the species composition estimate borrowed from the GLM.

ters in a year were available. Instead of using all four quarters of data, though, we applied the HM and GLM to three quarters of data and then predicted the species composition in the left out quarter. We then compare each model's predictive accuracy for the excluded data. The following is therefore a type of posterior predictive model checking.

We let $\boldsymbol{y}_{-i}$ denote the data vector for all species excluding quarter i and $\boldsymbol{y}_i^*$ be the vector of excluded data. Then $p(\boldsymbol{y}_i^*|\boldsymbol{y}_{-i}, M)$ is the probability of observing the excluded data $\boldsymbol{y}_i^*$ given model M and $\boldsymbol{y}_{-i}$. Because the excluded data only interact with the model via the predicted proportions of mass for each species, $\hat{\pi}_j$ (eq. 3), we rewrite the probability of interest using the vector of predictions for model M, $\boldsymbol{\pi}_M$, yielding $p(\boldsymbol{y}_i^*|\boldsymbol{\pi}_M)$. An appropriate method to score the accuracy of our prediction is to calculate the logarithm of the probability of observing $\boldsymbol{y}_i^*$ given the prediction (Gneiting and Raftery 2007; Czado et al. 2009). This is called the log score (LS) for model M given excluded quarter i:

$$(12) \qquad \mathrm{LS}_{i,M} = -\log(p(\boldsymbol{y}_i^*|\boldsymbol{\pi}_M))$$

That is, LS is the negative logarithm of the height of the probability distribution at the observed species composition given the parameters. We assume that landings are drawn from a multinomial distribution, and because $\boldsymbol{y}_i^*$ is a vector of observed counts for predicted proportions, $\boldsymbol{\pi}_M$,

$$(13) \qquad p(\boldsymbol{y}_i^*|\boldsymbol{\pi}_M) = \mathrm{Multinomial}(\boldsymbol{y}_i^*|n_i^*; \boldsymbol{\pi}_M)$$

and n_i^* is the total pounds sampled in the excluded quarter. A smaller LS indicates a better match between the model and the species composition from the excluded quarter.

In practice, the MCMC approaches for both the GLM and HM provide W samples of the predictive distribution — the distribution taken from an adjacent quarter in the case of the GLM and the PPD from the HM — not a single prediction as indicated by eq. 12. We summarize the predictive accuracy using LSs for each model by calculating the LS of the mean of the predictive distribution, $\mathrm{LS}_{i,M}(\overline{\pi})$, which provides a description of the predictive accuracy of the mean of the distribution across all species simultaneously. A second way to examine predictive accuracy is to calculate the average LS across the W samples from the predictive distribution, $\overline{\mathrm{LS}_{i,M}(\boldsymbol{\pi})} = \frac{1}{W}\sum_{w=1}^{W} -\log(p(\boldsymbol{y}_i^*|\boldsymbol{\pi}_{M,w}))$. This approach provides an metric for judging predictive accuracy that integrates across the uncertainty in the predictive distribution. In addition to these formal summaries of model performance, we also produced an intuitive description of predictive accuracy by calculating the 95% credible interval (CI) for the predictive distributions. We then asked what fraction of the observed species compositions in the excluded quarter fell within the CI.

We implement all metrics to compare the predictive accuracy of the GLM and HM for three examples. For each of the three examples, we estimated parameters for a model using a single market category in a single port where sampling data were available for all four quarters. For the GLM, we estimated a single model and borrowed the appropriate distribution from an adjacent quarter to predict the observed sampling data for each quarter and calculate the LSs. We apply the procedure of borrowing among quarters currently employed by NOAA to provide species composition

and HM is beyond the scope of this study and would involve a comparison of multiple structures for the HM, not just the single HM structure considered here. The method presented below is a general way to compare model structures. Since the emphasis is to provide estimates of species compositions for strata that are unobserved, a reasonable metric of model performance would be one that makes good predictions about the species composition of unobserved strata. Therefore we use cross-validation; specifically, we estimated the GLM and HM to market categories where samples from all four quar-

Fig. 5. An example of a cross-validation comparison of predictive distributions for market category 956 for trawl in 1985. Predicted distributions from the hierarchical model (HM; open) and the generalized linear model (GLM; shaded) when data from (*a*) quarter 2 or (*b*) quarter 3 are excluded. Boxplots show mean ($\Diamond$), median, interquartile, and 95% credible intervals of 1000 samples from the predictive distribution of species compositions. Observed data excluded from model fitting are shown ($\bullet$).
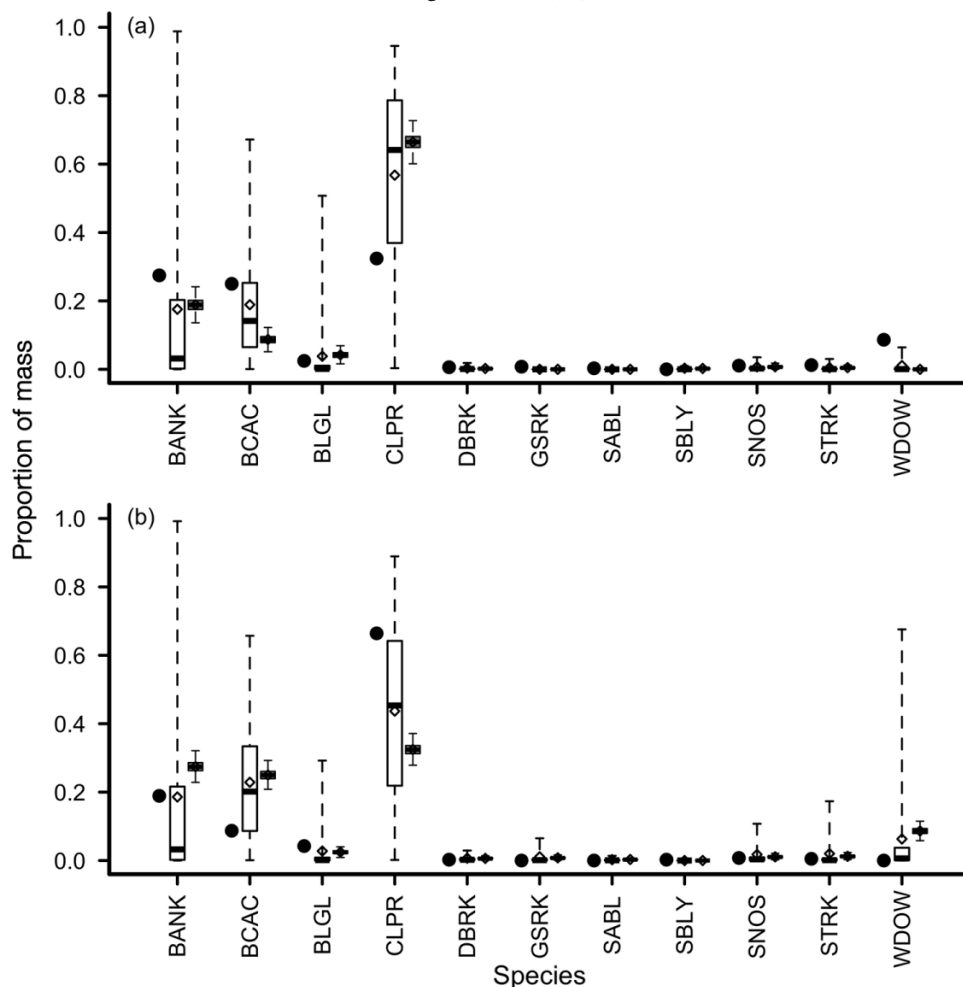
distributions for an unsampled stratum. Specifically we use the posterior distribution of π from Q2 to predict Q1, Q3 to predict Q2, Q2 to predict Q3, and Q3 to predict Q4 (Fig. 5).

For the HM, we used four independent MCMC samplers for each example, one each with a single quarter excluded from the data. For example, if Q1 was excluded, we used data from Q2, Q3, and Q4 to estimate the HM and produced a PPD and compared the predictions with the observed data from Q1 with LSs and CIs.

When we applied the HM and GLM to data from market category 956 at a single port in 1985, the GLM provided a much narrower predicted distribution of proportions than the HM (Fig. 5; for an additional example from different strata, see Fig. 3*d*). For the two quarters plotted for illustration, for most species, the 95% CI for the predictive distribution of the HM includes the observed data point, whereas the similar CI for the GLM rarely does (Table 1). However, the spread of the predictive distribution also means that much of the predictive distribution is far away from observed data. Thus, $\mathrm{LS}_{i,M}(\overline{\pi})$ identifies the HM as providing the best prediction of the mean in Q2 and Q3 (i.e., $\mathrm{LS}_{HM} \leq \mathrm{LS}_{GLM}$), but GLM as the better predictor in Q1 and Q4 (Table 1). $\overline{\mathrm{LS}_{i,M}(\pi)}$ in-

dicated that the HM only outperformed the GLM in Q2. However, the CI from the HM always outperforms the GLM.

In the second example, we compare the predictive success for market category 253 (Fig. 2). In this case, the GLM proves a better predictor for all LS metrics, indicating that borrowing from the GLM predicted the observed mean of the excluded quarter better than the HM. This is largely a result of the strange sample data for the species BANK (*Sebastes rufus*; Fig. 2), which is unobserved in Q1, Q2, and Q3 but comprised ~80% of sampled pounds in Q4. The HM responds to such variation among quarters by making the predictive distribution highly uncertain, whereas borrowed distributions in the GLM are relatively precise. Note the large LS in Q4, which indicates that both models are very poor at estimating the species composition (Table 1). Again, comparing CIs suggests that intervals from the HM tend to include the excluded data point, whereas the GLM does not.

In the third example, we analyze landings in market category 250 for a single port in 1995 (Table 1). In this case, the HM outperforms the GLM for all metrics of predictive success. The GLM makes precise but inaccurate predictions for the excluded data.

242

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Table 1. Comparison of GLM and HM predictive accuracy from cross-validation with log scores (LSs) and credible intervals (CIs).

Market category	Quarter excluded	$LS_M(\bar{\pi})$		$\overline{LS_{i,M}(\pi)}$		95% CI includes estimate from excluded quarter	
		GLM	HM	GLM	HM	GLM	HM
956	1	34.9	38.4	35.3	265.5	0 / 4	4 / 4
	2	1389.5	273.7	1398.3	1216.0	4 / 10	7 / 10
	3	161.2	110.0	162.8	546.2	2 / 8	7 / 8
	4	190.8	285.0	193.0	450.7	2 / 4	4 / 4
253	1	144.9	165.3	145.4	224.4	1 / 3	2 / 3
	2	4.9	32.7	5.7	104.3	0 / 2	2 / 2
	3	4.1	26.9	4.4	80.8	0 / 2	2 / 2
	4	4422.2	4424.2	4422.5	4426.5	0 / 6	2 / 6
250	1	2232.3	2198.3	2232.3	2208.5	0 / 5	2 / 5
	2	2195.1	613.7	2195.1	896.5	0 / 4	2 / 4
	3	1379.6	932.9	1497.6	1412.8	0 / 3	1 / 3
	4	2777.6	1248.3	2777.6	1886.2	0 / 5	3 / 5

Note: Lower LSs indicate a better match between the mean prediction and the excluded species composition. CI results are reported as the number of observed species compositions falling within the predicted interval / total number of species sampled.

The predictive accuracy of our cross-validation metrics were highly variable and depended on which quarter was excluded; LSs in all three examples vary over several orders of magnitude among quarters (Table 1). Generally, the borrowed estimates from the GLM provide a narrow predictive interval that does not include the data from the excluded quarter, whereas the HM provides a broad predictive interval that does include the excluded quarter.

There are many proposed methods for predictive model comparison (Gelfand and Ghosh 1998; Gneiting and Raftery 2007; Czado et al. 2009). The approach used here — leave-one-out cross-validation with log scoring and CIs — provides useful metrics of predictive accuracy. However, it is computationally intensive for the HM because a separate MCMC chain must be run for each data point that is excluded. As such, application of leave-one-out approaches on HMs for all data points in the California groundfish database is not practical. However, it can be used to spot-check the relative performance of alternate models for select, representative scenarios. Our application of log scoring to compare the performance of the GLM and HM for three market categories shows that the best model is not always clear. Generally, using the GLM tends to produce precise but inaccurate predictions, whereas the HM produces broader prediction intervals that tend to include the omitted data points. We conclude that in the presence of sparse or missing data, the HM produces a more reasonable description of uncertainty in landings composition and should be favored over a GLM approach, which relies on an ad hoc system of borrowing information across quarters. The use of the HM also matches intuition about how uncertainty in species composition should change with sampling effort; using the HM, strata with no sampled landings have greater uncertainty than strata with sampled landings. Using the GLM and borrowing, sampled and unsampled strata have equivalent uncertainty. As a result, in the next section, we focus on using the HM but highlight differences between GLM and HM landings estimates.

Estimating species-specific landings

In addition to providing robust methods for estimating spe-
cies composition from landings sampled, our approach provides a direct method for translating species compositions into species-specific landing masses and maintaining estimates of uncertainty. The HM produces species composition estimates for landings for a given gear group and market category for each quarter and port. Frequently, however, landings estimates are of interest only when they have been aggregated across temporal or spatial scales. For example, fisheries managers may be interested in the annual landings by species at a port or the species-specific landings by a certain gear group. The output from either the GLM or HM in concert with estimates of the total mass landed in each stratum can be used to aggregate landings across scales while maintaining uncertainty estimates.

We define π_{ijk} as the vector of posterior species composition estimates for species j in quarter i at port k, so that each element of π_{ijk} is a proportion between 0 and 1 (eq. 3). For a given i and k, the species composition for all species can be combined into a matrix $\mathbf{A}_{ik}$ in which the columns are π_{ijk} and each row is a draw from the joint posterior distribution for the species composition. Each row sums to 1. Each matrix $\mathbf{A}_{ik}$ is derived either from latent variables θ_{ijk} if sampling data were available in port k and quarter i or from the PPD if direct samples were unavailable. We use $\mathbf{A}_{ik}$ in concert with the total pounds landed within a market category in each quarter and port, x_{ik}, to estimate the landings of each species. The total pounds landed within each stratum is calculated from fish tickets. If the landings mass estimated from fish tickets is uncertain, we can use an observed landings vector x_{ik}, which contains independent draws from the uncertainty distribution of total landings mass. We combine $\mathbf{A}_{ik}$ and x_{ik} by summation to estimate the total landings for each species. For example, the estimates for the landings of species j in port k across all quarters, n_{jk}, is

$$(14) \qquad n_{jk} = \sum_{i=1}^{I} \pi_{ijk} x_{ik}$$

The elements of n_{jk} are thus posterior estimates of the landing mass for species j in port k. Similarly, we could obtain an

Fig. 6. Estimated species-specific landed mass for three ports in California for trawl and market category 956 in 1985. Each panel shows 2000 posterior estimates of species landings by quarter for (*a*) port A, (*b*) port B, and (*c*) port C. The total pounds sampled and number of independent samples in each port and quarter are noted at the top of each panel. Mean (○), interquartile (solid line), and 95% CI (broken line) are shown. Zero pounds were sampled for quarters 2 and 3 for port A and for quarter 3 for port B.

244

Can. J. Fish. Aquat. Sci. Vol. 69, 2012
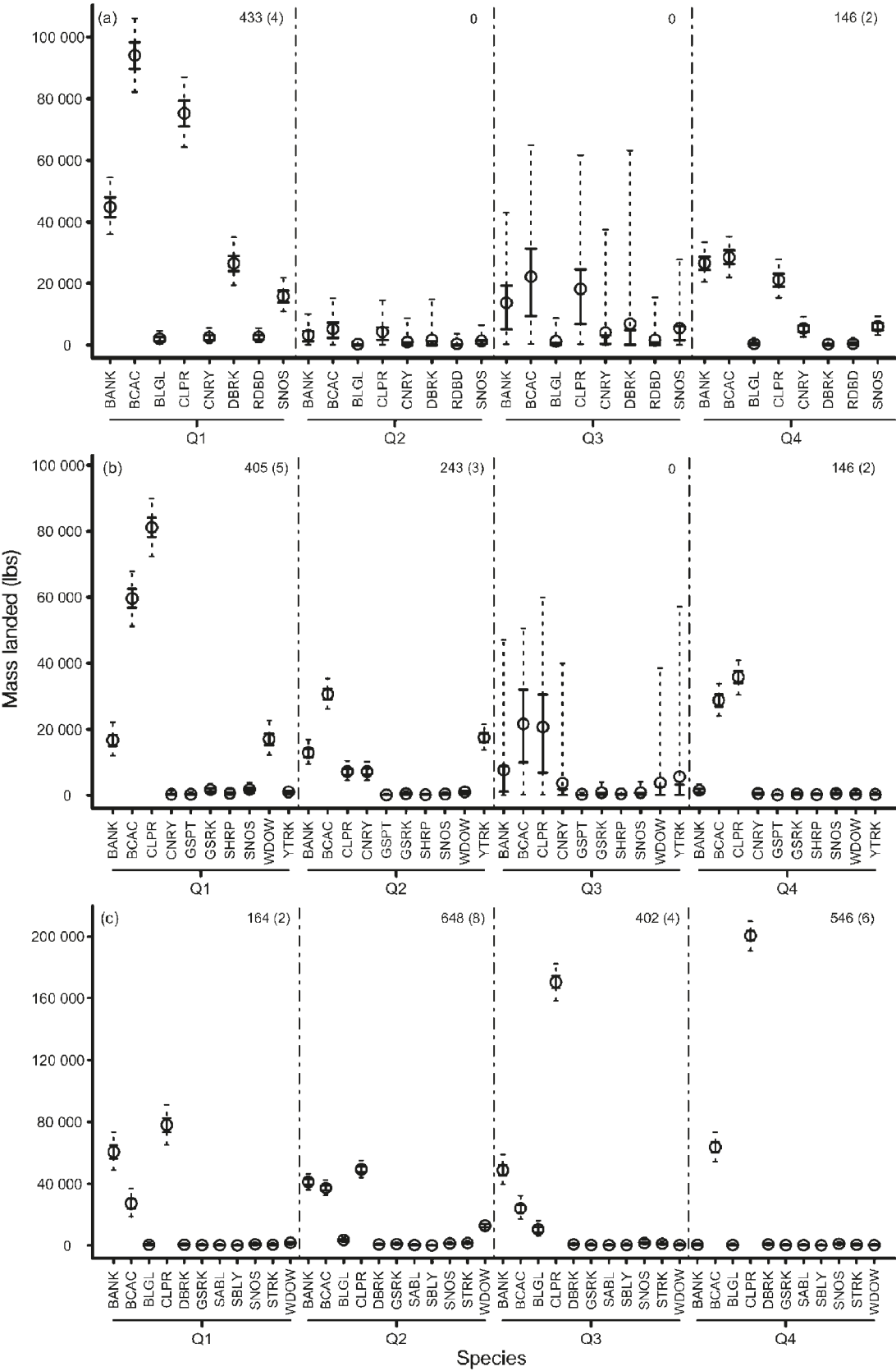
Fig. 7. Posterior estimates of species-specific landed mass for three ports in California for trawl and market category 956 in 1985: (*a*) species-specific landings estimates for three ports illustrated in Fig. 6; (*b*) total species-specific landings estimates and estimated coefficients of variation (CV) for the combined landings at the three ports; and (*c*) trends in CV with mass landed for the 16 species observed at the three sampled ports from the hierarchical model (HM; ◆) and generalized linear model (GLM; ○). Species estimates of CV from the GLM and HM are connected by broken lines.
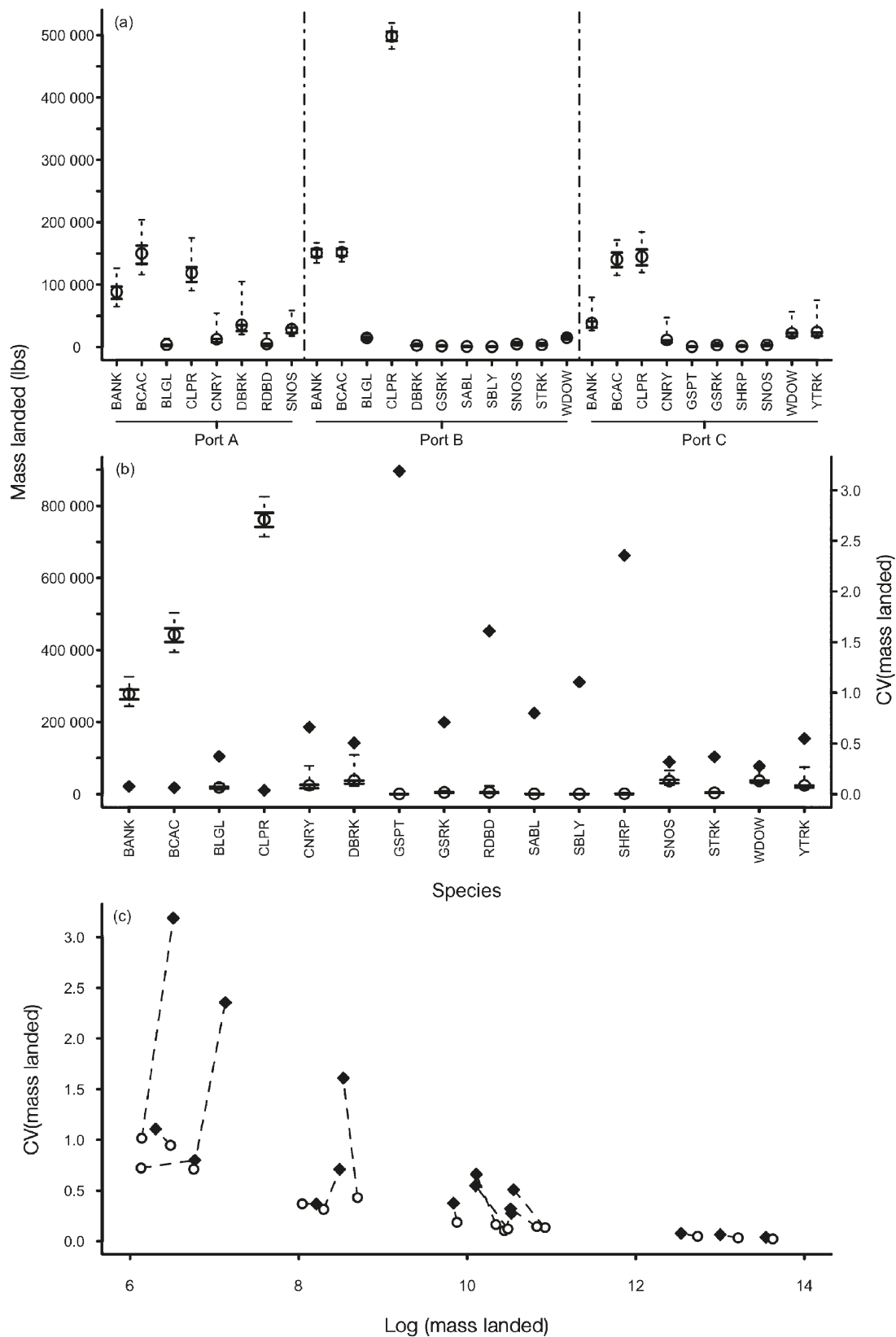
estimate for the coastwide landings for all ports by summing across ports. If the total pounds landed in each stratum are known without error, we can replace the vector x_{ik} in eq. 14 with a scalar, x_{ik}.

We illustrate the process of converting species composition estimates to species-specific landings using sampling data and total landings data from the 1985 trawl fishery at three ports using the HM (Figs. 6, 7). In these figures, we refer to the coefficient of variation (CV; CV = standard deviation / mean) as a metric of variability. Because of confidentiality requirements, we refer to these landing ports as port A, port B, and port C. In port A, sampling data were available for quarters 1 and 4 but not quarters 2 and 3; for port B, sampling data were available for all but quarter 4; and for port C, sampling data were available for all quarters (Fig. 6). The total mass landed varied by port as well as by quarter. In this example, we assume that the total pounds landed in a stratum is known without error. Note that the species present in this market category vary among ports. As expected for the PPD, quarters with no sampling data are estimated with less precision than quarters with abundant sampling data (Figs. 6a, 6b). We combine landings estimates from these ports to produce estimates of landings for annual species landings at each port (Fig. 7a) and then combine ports to estimate species-specific landings for all three ports combined (Fig. 7b). Note that the uncertainty bands are maintained across summations. In theory, we could continue to sum across multiple market categories or multiple gear groups and produce landings estimates aggregated to any scale of interest.

Because we maintain uncertainty bounds across categories, we can also estimate the precision of our species-specific landings using the CV. We plot the CV for the landings at ports A, B, and C combined for each species (Fig. 7b) and for each species as a function of its mean landed biomass (Fig. 7c). For comparison, we also plot the CV derived from the GLM for the same ports. The clearest difference between the GLM and HM is a decreased estimate of CV in the GLM. The mean estimate of landings from the GLM and HM are similar but not identical (Supplemental material, Fig. S1).[1] As discussed earlier, the HM is a better reflection of true uncertainty than the GLM (Table 1). Only the three species with the largest poundages landed have CVs of less than 0.10, whereas relatively rare species have large CV estimates (some with CV > 1.5; Fig. 7c). These methods provide a clear way to produce point estimates and uncertainty bounds for landings over any scale of interest.

Discussion

We developed a hierarchical Bayesian method for estimating species composition in a complex multispecies fishery and used it to estimate landings under scenarios with abundant and sparse sampling data. Furthermore, we provide a way to compare model performance and an intuitive way to propagate uncertainty when aggregating landings across strata. Our models draw from the well-studied class of Poisson linear mixture models (e.g., Lee et al. 2006) and are related to multinomial regression models. They present a convenient form for the analysis of count data with categorical covariates. We developed and applied a fairly basic model

to estimate the species composition of landings under sparse data conditions. However, the model structure can accommodate a great deal more complexity, including multiple hierarchical effects. For example, an alternative formulation of the model would be to treat individual boat landings hierarchically within a stratum rather than as independent random samples. Such an approach would provide a way to account for overdispersed data in species composition among landings. Generally, the ability to add hierarchical effects at any level of stratification provides a flexible method introducing overdispersion where appropriate. However, more complicated models with many hierarchical effects generally require more data and can become more difficult to interpret.

The log-linear formulation of the model enables great flexibility, and these methods should be easy to implement in a wide variety of applications with multinomial-type data. Potential applications range from similar problems such as the estimation of fisheries bycatch in at-sea observer programs or estimating species age composition from sampling data (e.g., Hulson et al. 2011) to the analysis of diet from observations of foraging animals. There are a few previous applications of the multinomial–Poisson transformation or similar approaches in the ecological literature (see Kéry and Royle 2010; Dorazio et al. 2010), but we believe that this is a novel application of this technique in fisheries. We emphasize that our approach is not recommended as a substitute for thoughtful and well-designed sampling plans; rather it is a method for recognizing the relatedness between strata that can be used to aid in quantifying uncertainty. The approach is particularly useful in the face of sparse or missing sampling data.

As with any statistical sampling program, care must be taken to ensure that the assumption of exchangeability among samples within categories is appropriate. Given the vast number of well-documented potential biases of fisheries and landing monitoring programs (e.g., Cotter and Pilling 2007; Benoît and Allard 2009), determining which samples can be combined and which cannot is a nontrivial task and currently must rely on the expert judgement of managers. In California groundfish fisheries, for example, species composition samples from a given year are never used to inform landings in other years because annual changes in harvest regulations change the species composition within strata. In general, for a constant sampling effort, a greater number of strata means fewer samples in each stratum and increased probability of unsampled strata. Thus, sampling designs must balance potential errors arising from combining sampling data from truly different categories with errors resulting from leaving some categories entirely unsampled. An ongoing area of research is understanding the consequences of different statistical models for estimating landings and uncertainty and how to identify the most appropriate statistical model for a given circumstance. If anything, the large amount of uncertainty that arises from using the PPD demonstrates the value of ensuring adequate landings sampling.

As with most Bayesian hierarchical models, computational time and computational efficiency are important concerns for implementation. More complex models have more parameters and latent variables and so computing is slowed. Additionally, MCMC simulations of complex models tend to mix poorly and thus require extensive thinning of the MCMC chain, so longer simulations are necessary to obtain the re-

246

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

quired number of independent draws from the posterior. Given the continuing increase in computing power, we view computing challenges as relatively minor.

References

Baker, S.G. 1994. The multinomial–poisson transformation. Statistician, **43**(4): 495–504. doi:10.2307/2348134.

Benoît, H.P., and Allard, J. 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? Can. J. Fish. Aquat. Sci. **66**(12): 2025–2039. doi:10.1139/F09-116.

CALCOM (California Cooperative Goundfish Survey). 2010. The CALCOM database. Managed by the California Cooperative Groundfish Survey (CCGS): California Department of Fish and Game (CDFG), Belmont, California; the Pacific States Marine Fisheries Commission (PSMFC), Belmont, California; and the National Marine Fisheries Service (NMFS), Santa Cruz, California.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. John Wiley and Sons, New York.

Cotter, A.J.R., and Pilling, G.M. 2007. Landings, logbooks and observer surveys: improving the protocols for sampling commercial fisheries. Fish Fish. **8**: 123–152.

Crone, P.R. 1995. Sampling design and statistical considerations for the commercial groundfish fishery of Oregon. Can. J. Fish. Aquat. Sci. **52**(4): 716–732. doi:10.1139/f95-072.

Czado, C., Gneiting, T., and Held, L. 2009. Predictive model assessment for count data. Biometrics, **65**(4): 1254–1261. doi:10. 1111/j.1541-0420.2009.01191.x. PMID:19432783.

de Mutsert, K., Cowan, J.H., Jr., Essington, T.E., and Hilborn, R. 2008. Reanalyses of Gulf of Mexico fisheries data: landings can be misleading in assessments of fisheries and fisheries ecosystems. Proc. Natl. Acad. Sci. U.S.A. **105**(7): 2740–2744. doi:10.1073/pnas.0704354105. PMID:18287085.

Dorazio, R.M., Kéry, M., Royle, J.A., and Plattner, M. 2010. Models for inference in dynamic metacommunity systems. Ecology, **91**(8): 2466–2475. doi:10.1890/09-1033.1. PMID:20836468.

Gelfand, A.E., and Ghosh, S.K. 1998. Model choice: a minimum posterior predictive loss approach. Biometrika, **85**(1): 1–11. doi:10.1093/biomet/85.1.1.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. **1**: 515–534. doi:10.1214/06-BA117A.

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

Gneiting, T., and Raftery, A.E. 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. **102**(477): 359–378. doi:10.1198/016214506000001437.

Hulson, P.-J.F., Hanselman, D.H., and Quinn, T.J., II 2011. Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood methods. ICES J. Mar. Sci. **68**(7): 1548–1557. doi:10. 1093/icesjms/fsr102.

Jermyn, A.S., and Robb, A.P. 1981. Review of the cod, haddock, and whiting discarded in the North Sea by Scottish fishing vessels for the period 1975–1980. ICES CM 1981/G:47.

Kéry, M., and Royle, J.A. 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. J. Anim. Ecol. **79**(2): 453–461. doi:10.1111/j.1365-2656. 2009.01632.x. PMID:19886893.

Lee, Y., Nelder, J.A., and Pawitan, Y. 2006. Generalized linear models with random effects. Chapman and Hall/CRC, Boca Raton, Florida.

Methot, R.D. 2000. Technical description of the Stock Synthesis assessment program. NOAA Tech. Memo. NMFS-NWFSC-43.

Methot, R.D. 2007. User manual for the integrated analysis program Stock Synthesis 3 (SS3). User manual for the Assessment Program Stock Synthesis 3 (SS3). Available from http://nft.nefsc.noaa.gov/.

Miller, T.J., Skalski, J.R., and Ianelli, J.N. 2007. Optimizing a stratified sampling design when faced with multiple objectives. ICES J. Mar. Sci. **64**: 97–109.

Pauly, D., Christensen, V., and Walters, C. 2000. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. ICES J. Mar. Sci. **57**(3): 697–706. doi:10.1006/jmsc.2000.0726.

Pearson, D.E., and Almany, G. 1995. The effectiveness of California's commercial rockfish port sampling program. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-218.

Pearson, D.E., and Erwin, B. 1997. Documentation of California's commercial market sampling data entry and expansion programs. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-240.

Pearson, D.E., Erwin, B., and Key, M. 2008. Reliability of California's groundfish landings estimates from 1969–2006. NOAA Tech Memo. NOAA-TM-NMBS-SWFSC-431.

R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

Sen, A.R. 1984. Sampling commercial rockfish landings in California. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-45.

Sen, A.R. 1986. Methodological problems in sampling commercial rockfish landings. Fish Bull. **84**: 409–421.

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