A general procedure for estimating the composition of fish school clusters using standard acoustic survey data

T. R. Hammond and G. L. Swartzman

(AP)

An algorithm to identify classes of fish in acoustic backscatter images would improve the accuracy of acoustic biomass estimates over manually scrutinized images. A generalized Bayesian procedure for such identification called BASCET is presented, and two implementation strategies for the procedure are compared using simulated acoustic survey data. The procedure has several unusual characteristics: it evaluates schools not individually but in clusters; it makes use of human experience at cluster identification; it presents measures of uncertainty in all estimation results; and it constructs the training set required for supervised learning automatically using spatial and temporal assumptions. The simulation study comparison suggests that making use of temporal and spatial structure in the acoustic data leads to improved estimation performance. On the simulated data, the BASCET algorithm correctly identified the dominant fish class in 15 of 16 cases. However, the simulation model generates acoustic survey data based on the same assumptions used in BASCET, assumptions that may differ from a real acoustic survey. The study also assumed that the human experience incorporated in the Bayesian prior distributions was not misleading. Performance of BASCET on real acoustic data is presented in a companion paper.

T. R. Hammond, School of Fisheries, University of Washington, Seattle, WA, USA. G. L. Swartzman, Applied Physics Laboratory, University of Washington. Correspondence to: T. R. Hammond, CEFAS, Pakefield Road, Lowestoft, Suffolk NR33 0HT, England, UK; tel: +44-1502-524213; fax: +44-1502-524511; e-mail: t.hammond@cefas.co.uk

Introduction

When acoustic surveys are used to estimate the biomass of a particular class of fish, it is valuable if the acoustic equipment can distinguish that class from others. For such discrimination to be possible, the relevant classes, whether identified by size or species, must differ in any of three ways: in acoustic characteristics of their echoes, in the shape of their schools, or in the position of their schools relative to environmental conditions. It has already been shown that such differences can be exploited to achieve successful classification (Rose and Leggett, 1988; Richards et al., 1991; Weill et al., 1993; Simmonds et al., 1996). However, a general procedure for doing so using the data from a standard acoustic survey has yet to be presented.

Acoustic data

In an acoustic survey, data are often recorded in backscatter images. The individual pixels in these images represent the mean volume backscatter cross-section (S_v) in regular depth intervals beneath the path of the transducer. Schools of fish can be extracted from the images using standard image-processing techniques (Haralick et al., 1987; Haralick and Shapiro, 1992). Extraction algorithms based on such techniques have been implemented by several scientists (Nero and Magnuson, 1989; Nero et al., 1990; Baussant et al., 1993; Reid and Simmonds, 1993; Weill et al., 1993). As implemented by Swartzman et al. (1994b), this extraction process records (among other things) mean S_v, S_v range, area, location, depth, and several environmental measurements for every school. Except for ocean depth, which is available from the acoustic images, the

environmental measurements are obtained by inter-polation between CTD (conductivity, temperature, depth) casts (Swartzman *et al.*, 1994a). Here, a school will refer to a record of such measurements that has been extracted from a backscatter image.

Schools can be numerous, so it is often better to deal with school clusters. These clusters consist of collections of schools that are all near each another. In more precise terms, two schools are placed in the same cluster whenever they are within threshold horizontal range (d_h) and within threshold vertical range (d_v) of each other. A cluster retains the average of the school measurements for the schools it contains.

The data from a standard acoustic survey consist of the following: location, time, and depth of all trawl samples and school clusters, haul catches, and the mean of each acoustic feature in each school cluster. Although variances can be recorded for acoustic fea-tures, these are not used here. Environmental feature vectors are also available for every cluster and haul. This paper is devoted to the question of how to design a general procedure for using these data to estimate the composition of all clusters.

Standard classification algorithms

Most standard classification algorithms require a train-ing set (James, 1985). In an acoustic survey context, this is a set of school clusters for which the true composition is known. It is used to teach the algorithm how to differentiate between the classes. The quadratic discrimi-nant analysis method in the pioneering work by Rose and Leggett (1988) is an example of a standard classifi-cation algorithm built with a training set. Other examples include the k-nearest-neighbour algorithm in Richards *et al.* (1991), the linear discriminant analyses in Weill *et al.* (1993) and Barange (1994), and the neural networks of Simmonds *et al.* (1996) and Haralabous and Georgakarakos (1996).

Once a standard classification algorithm has been taught to discriminate between classes using the training set, its parameters are not modified further. The school clusters are presented to it individually, and they are identified as if they were independent objects. The fact that the classification parameters are fitted only to the training set means that only a small portion of the available data is actually used. All the information contained in clusters that are not in the training set is discarded. The loss is considerable because the distribu-tion of acoustic and environmental attributes is quite informative.

The first scientist to discriminate between object types using acoustic information outside a training set was François Gerlotto. The hierarchical clustering algorithm used in his paper (Gerlotto, 1993) is notable for two

reasons: it does not require a training set, and it uses all acoustic data. The paper shows that areas of the ocean can be grouped in such a way that the acoustic measure-ments within each group are similar. He called these groups ''acoustic populations'', although their species composition remains unknown. Although this level of resolution is not particularly useful for stock assessment, the paper does serve to show the value of using the entire acoustic data set. After all, if objects can be sorted into natural groupings, then identification is just a matter of labeling these groups correctly.

Most classification algorithms can be used either to estimate the composition of an object or to classify it. Here it is considered appropriate to estimate the com-position of schools and school clusters rather than to classify them, because schools are not guaranteed to be composed purely of a single class.

The troublesome trawls

It has proved difficult to design a general procedure for estimating the composition of school clusters because it is hard to come up with the training set required by most standard classification algorithms. In a standard acoustic survey, the only information about the true composition of school clusters comes from trawl samples (Simmonds *et al.* 1992). Trawl samples are not a reliable means of determining the species or size composition of school clusters. There are three reasons for this: gear selectivity, spatial and tem-poral separation between the trawl and the cluster, and uncertainty about which trawls are informative about a particular cluster and which provide little information about it.

Trying to construct the training set for a classification algorithm using trawl data is awkward, but automating the process can relieve some of the burden on surveyors. In this paper, an algorithm for constructing a training set automatically using four spatial and temporal parameters (called R_{trawl}^S, R_{trawl}^T, $R_{cluster}^S$, and $R_{cluster}^T$) is suggested. It is hoped that reducing the process of training set construction to the choice of these four parameters will facilitate communication, as well as reduce the scope for between-surveyor variability. It is also hoped that the focus on spatial and temporal structure that this algorithm requires of the user will lead to improved estimation performance. Such improvement would come from what is here called ''context sensitive anticipation''.

The identities of schools located in close proximity to one another in the ocean are far from independent. This fact is ignored in standard classification algorithms, which assume independence. This paper will demon-strate how to construct a composition estimation tool that can anticipate the composition of a school cluster

based on information from other clusters and trawl samples nearby.

The human brain routinely anticipates the properties of observations based on their context. If a man observes that a number of school clusters in a small area all appear to be adult pollock, then when he comes upon another cluster in the same area, he would surely be inclined to think the cluster was also composed of adult pollock. Computer algorithms can be designed to mimic such anticipation. When they are so constructed, classification performance should show improvement over standard methods. This prediction is tested here using simulated data.

What is a general procedure for school cluster identification?

A general procedure for school cluster identification uses standard acoustic survey data to produce composition estimates for all school clusters in the data set. These composition estimates are vectors of length K, where K is the number of fish species or size classes that the user wishes to distinguish. The classes are assumed to be specified in advance by the user. Practical considerations suggest that they should be few in number ($K < 5$), especially if the user hopes to achieve reasonably precise estimates of composition. The entries of a composition vector are constrained to sum to 1 and represent the numerical proportion of the respective fish classes in a school cluster.

A general procedure should be a statistical algorithm rather than a sequence of *ad hoc* steps. The algorithm can and should be able to improve its estimates using auxiliary information, when such is available, but it must be prepared to make do with standard acoustic data alone. A general procedure should also be able to provide some reasonable measure of precision for the composition estimates it makes.

In this paper, two general procedures for estimating the composition of all the school clusters identified in an acoustic survey are described and compared. The procedures are designed to distinguish a small number of classes, although each class can contain several species or size groups. Both algorithms are Bayesian (Gelman *et al.*, 1995), requiring prior distributions that represent *a priori* beliefs about the values of key parameters. These priors may be elicited from an acoustic expert or obtained from experience with other surveys. Recommendations for obtaining these priors are presented in a companion paper (Hammond *et al.*, 2001). The procedures are intended for situations where superior estimation performance is crucial but speed is not. The methods usually take several days to calculate results for a real survey on a 300 MHz Pentium PC.

Methods

A simulation experiment

A primary objective of this paper is to conduct a simulation experiment comparing the performance of estimation tools. The model used to simulate the acoustic data for this experiment is presented fully in Hammond (2000), although we provide an overview below. The three classes used in both simulation and estimation were adult pollock, juvenile pollock, and jellyfish.

A test survey region and a test survey track were constructed for the simulation experiment. The region and its accompanying survey track are shown in Figure 1. The ocean depth in the test region is shown as a light-shaded grey surface, and contours of sea surface temperature are indicated on the top face of the plotting cube.

The parameters of the simulation model were selected at random from prior distributions elicited from an acoustic expert (Hammond *et al.*, 2001). Estimation tools were compared in terms of their ability to recover the true composition of the simulated school clusters. For this task, each model was allowed to use both the prior information and the simulated acoustic survey data. Both models were designed to use the same classification attributes.

The estimation tools are called "BASCET" and "the independence scheme". BASCET is an acronym for Bayesian spatial composition estimation tool for clusters of acoustically identified schools. Both algorithms are general approaches to estimating cluster composition in the sense that they use all the data and neither requires a user-specified training set. Parameters are estimated from the results of an entire survey. We now provide an overview of the simulation model, followed by an overview of the BASCET approach. Full descriptions of BASCET, including all the equations, are given in Appendix 2.

The simulation model

The model was designed to represent two aspects of fish behaviour. First, it creates patchy distributions of fish schools; second, it allocates school compositions so that local average species compositions can differ from the overall average. In other words, fish species are not mixed homogeneously in the survey area.

The key to the simulation model is the use of hotspots, which represent large-scale but temporary oceanic features. Examples of hotspots include plankton blooms, fronts, and upwelling areas. These features are assumed to be attractive to a particular composition of species. Hotspots are taken to have a circular shape. Although the shape assumption may not always be suitable, one can represent complex oceanic structures by using several hotspots.
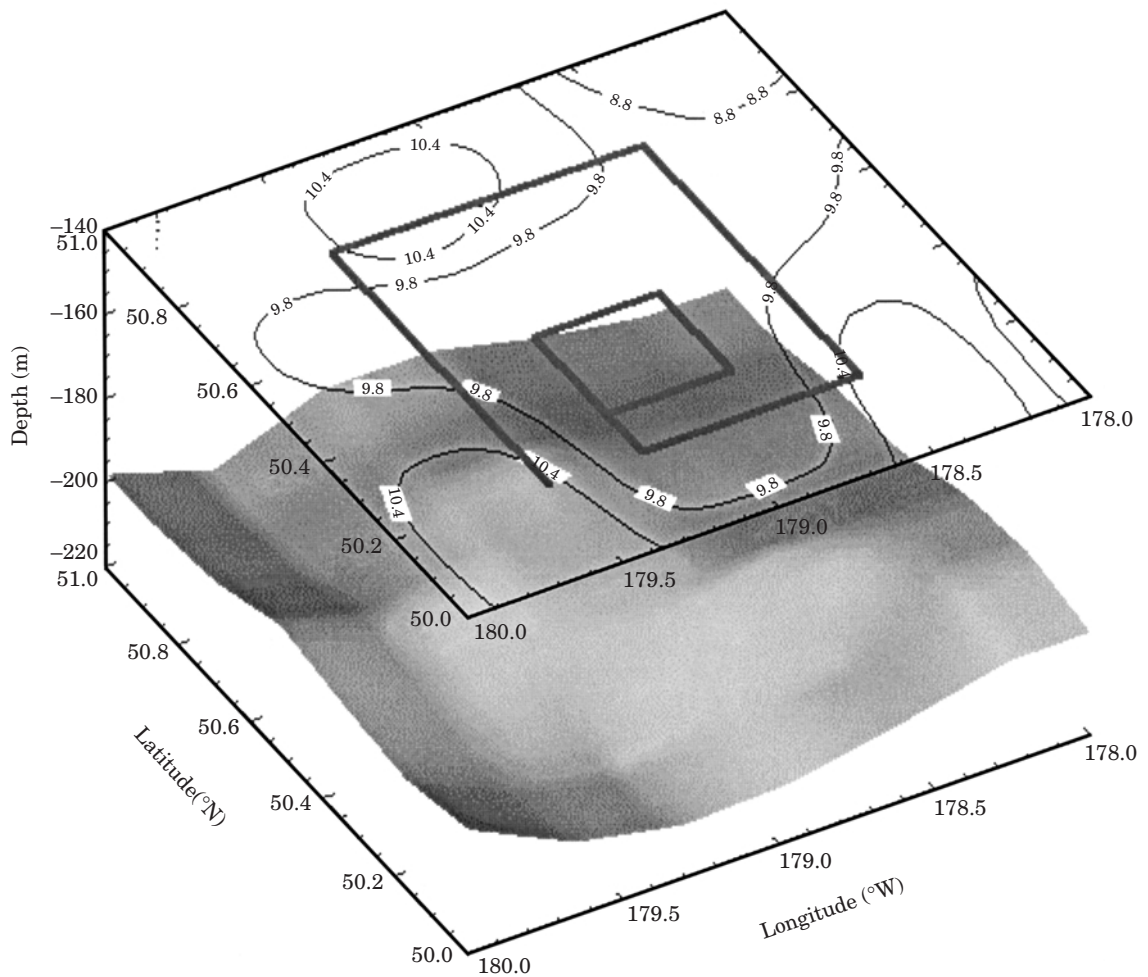
Figure 1. The survey region used for evaluation is the rectangle shown. A shaded surface plot shows the bottom depth in the survey region, while contour lines of sea surface temperature in °C are depicted on the top face of the cube. The survey track is shown by a thick solid line.

In the simulation model, fish schools are placed around hotspots using a parent-child random point process. This placement is implemented in two phases. In the first phase, hotspots are allocated randomly in the survey region, their number controlled by the hotspot intensity parameter λ_H. Each hotspot is marked with a spatial range, a temporal range, and a composition for which it is favourable. All these markings are drawn randomly from their respective distributions. In the second phase, schools are allocated so they lie on the curtain of water beneath the survey track and within spatial and temporal range of the hotspots. The number of schools in a hotspot is controlled by the school intensity parameter λ_S.

Schools are assigned a composition that depends on the composition of their hotspot "parent" (i.e. on the hotspot that contains them), on the school depth, and on the environmental features at the school's location. Pure schools of each class are assumed to have different acoustic feature distributions. The distribution of acoustic features in an impure school is modelled using a mixture of the pure school feature distributions. The mixture proportions are given by the school composition. Acoustic features of a school are generated randomly from the school's acoustic feature distribution. Schools that are within a threshold horizontal range (d_h) and within a threshold vertical range (d_v) of each other are merged into school clusters.

The catch from trawl samples is also simulated using a point process approach. Three assumptions underlie the process: that trawl time is a constant, that trawl samples are only done within hotspots, and that the depth of the trawl samples has the same distribution as the local depth distribution of the fish. The idea is to allocate
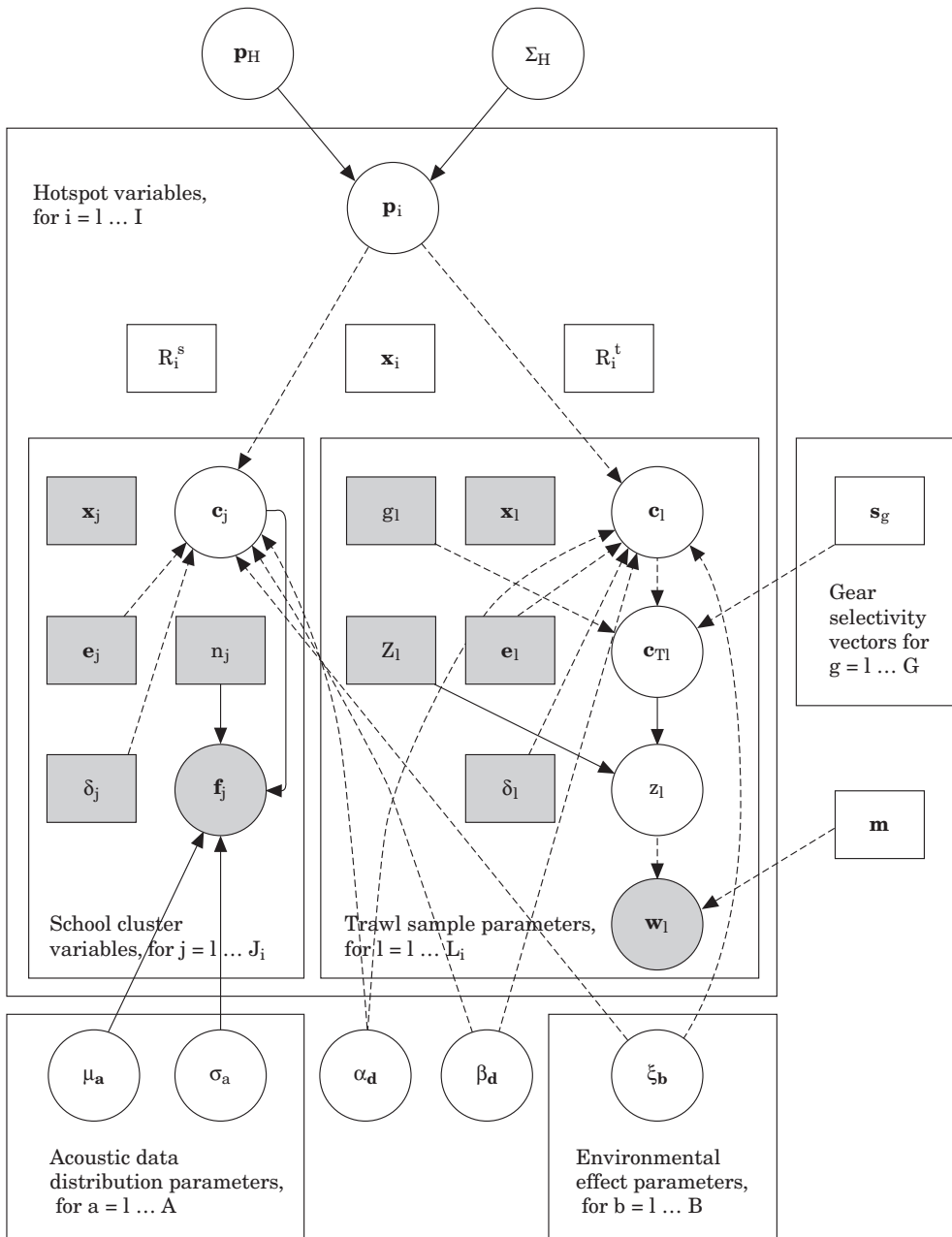
Figure 2. A graphic representation of the conditional independence properties of the BASCET model. Shaded nodes indicate variables found in the acoustic data set. Broken arrows indicate a deterministic dependency between variables, and solid arrows indicate a stochastic one. Thus, a variable with three incoming broken arrows would be a function of three arguments. Rectangular nodes indicate constants (that were not treated as uncertain).

schools within the area swept by the gear using the same process that allocates schools beneath the survey track. Schools in the path of the gear are retained with a certain average probability, and their composition is adjusted for gear selectivity. Computing the catch data is a matter of converting the numerical composition of the schools retained by the gear to a weight composition.

An overview of BASCET

The hotspot idea is also central to the estimation approach, because BASCET uses hotspots to help estimate the composition of school clusters. These hotspots represent regions of space and time within which the expected composition of school clusters can be different

Table 1. Simulation parameters and the prior distributions from which they were sampled.

Name	Description	Prior distribution	Simulation value
λ_H	Hotspot intensity	Gamma (37.24, 0.1719)	8.7507
μ_s	Expected hotspot log spatial range	$N(2.708, 0.3^2)$	3.2024
μ_t	Expected hotspot log temporal range	$N(1.792, 0.3^2)$	1.5512
σ_s^2	Variance in hotspot log spatial range	IG(18.0, 0.1634)	0.5461^2
σ_t^2	Variance in hotspot log temporal range	IG(18.0, 0.1634)	0.4704^2
$\mathbf{p}_H$	Expected hotspot composition	ALN $(\boldsymbol{\mu}_H, \boldsymbol{\Phi}_H)$ $$\boldsymbol{\mu}_H = \begin{bmatrix} 0.3918 \\ -1.2966 \end{bmatrix}$$ $$\boldsymbol{\Phi}_H = \begin{bmatrix} 0.9896 & 0.8987 \\ 0.8987 & 1.0768 \end{bmatrix}$$	$$\begin{bmatrix} 0.4145 \\ 0.1022 \\ 0.4833 \end{bmatrix}$$
Σ_H	Variance in hotspot composition	$\pi(\Sigma_H^{-1}) = $ Wishart $(15, \boldsymbol{\Omega}_H)$ $$\boldsymbol{\Omega}_H = \begin{bmatrix} 0.1187 & -0.0647 \\ -0.0647 & 0.0724 \end{bmatrix}$$	$$\begin{bmatrix} 0.8088 & 0.7967 \\ 0.7967 & 2.2295 \end{bmatrix}$$
λ_S	School intensity	Gamma(11.25, 20)	220.113
ξ_{temp}	Effect of temperature increase on expected cluster composition	ALN $(\text{alr}(\overline{\omega}_{temp}), \boldsymbol{\Psi}_{temp})$ $$\bar{\omega}_{temp} = \begin{bmatrix} 0.1867 \\ 0.4075 \\ 0.4058 \end{bmatrix}$$ $$\boldsymbol{\Psi}_{temp} = \begin{bmatrix} 0.2442 & 0.1075 \\ 0.1075 & 0.1498 \end{bmatrix}$$	$$\begin{bmatrix} 0.2055 \\ 0.5075 \\ 0.2869 \end{bmatrix}$$
ξ_{depth}	Effect of ocean depth increase on expected cluster composition	ALN $(\text{alr}(\omega_d), \boldsymbol{\Psi}_d)$ $$\bar{\omega}_d = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$ $$\boldsymbol{\Psi}_d = \begin{bmatrix} 0.0002 & 0.0001 \\ 0.0001 & 0.0002 \end{bmatrix}$$	$$\begin{bmatrix} 0.3343 \\ 0.3301 \\ 0.3356 \end{bmatrix}$$

from the overall average class composition in the entire survey region. In other words, the use of hotspots allows the algorithm to avoid assuming that the species classes are mixed homogeneously in the survey region.

For the sake of convenience, the estimation model assumes that hotspots have circular shape and finite time duration. All have a composition vector that represents the local average class composition (this vector is often called the hotspot composition). In BASCET, hotspots are always created in such a manner that they never overlap (in either space or time). The assumption that hotspots do not overlap is made so that no school cluster will lie in more than one hotspot.

In BASCET, hotspots provide the mechanism by which the catch from trawl samples is related to the composition of school clusters. The model assumes that the catch from a trawl sample is informative about the composition of the hotspot that contains the haul. The catch composition is also relevant to the composition of any school clusters contained in that same hotspot. Thus, a configuration of hotspots in the survey region

determines a set of school clusters for which there is direct composition evidence, i.e. a training set. In other words, a method for generating a configuration of hotspots in the survey region is also a method for constructing a training set. As BASCET incorporates an algorithm for using simple spatial and temporal hypotheses to generate hotspot configurations, it also provides a method for constructing training sets automatically. The hotspotgen algorithm used for constructing hotspot configurations is described in Appendix 2. There is, however, no need to accept the configuration suggested by hotspotgen: an operator can certainly edit the hotspots or even specify them all manually.

Another advantage of using hotspots derives from their ability to help anticipate the composition of school clusters. Suppose, for example, that it is necessary to predict the composition of a school cluster about which nothing is known. The prediction would be that it has the overall class composition that the survey region is known to have. If, however, the composition of the hotspot containing the cluster of interest (i.e. the local

Table 2. Remaining simulation parameters and the prior distributions from which they were sampled.

Name	Description	Prior distribution	Simulation value
$\mu_{a,k}$ for $a=1\ldots A$ $k=1\ldots K$	Acoustic feature means for each class and feature	$\pi(\mu_{a,k})=N(\theta_{a,k},\tau_{a,k}^2)$ $$[\theta_{a,k}]=\begin{bmatrix} -43 & -45 & -55 \\ 25 & 30 & 15 \\ 1500 & 2500 & 1000 \end{bmatrix}$$ $$[\tau_{a,k}]=\begin{bmatrix} 5 & 5 & 2.5 \\ 5 & 10 & 4 \\ 300 & 350 & 100 \end{bmatrix}$$	$$[\mu_{a,k}]=\begin{bmatrix} -42.65 & -36.76 & -53.53 \\ 2.822 & 5.416 & 1054.9 \\ 1455.2 & 1731.4 & 1054.9 \end{bmatrix}$$
σ_a^2 for $a=1\ldots A$	Acoustic feature variances	$\pi(\sigma_a^2)=IG(\alpha_a,\beta_a)$ $$[\alpha_a]=\begin{bmatrix} 10 \\ 2.98 \\ 8.25 \end{bmatrix}$$ $$[\beta_a]=\begin{bmatrix} 0.0056 \\ 0.0020 \\ 5.52\times 10^{-7} \end{bmatrix}$$	$$[\sigma_a^2]=\begin{bmatrix} 17.33 \\ 1.77 \\ 201440 \end{bmatrix}$$
$\alpha_{d,k}$ for $\beta_{d,k}$ $k=1\ldots K$	Depth preference parameters by class	$\pi(\alpha_{d,1})=\text{lognormal}(0.45,\ 0.28)$ $\pi(\alpha_{d,2})=\text{lognormal}(-1.03,\ 0.64)$ $\pi(\alpha_{d,3})=\text{lognormal}(-0.25,\ 0.62)$ $\pi(\beta_{d,1})=\text{lognormal}(-1.03,\ 0.59)$ $\pi(\beta_{d,2})=\text{lognormal}(0.70,\ 0.33)$ $\pi(\beta_{d,3})=\text{lognormal}(0.98,\ 0.36)$	$$\alpha_d=[\alpha_{a,k}]=\begin{bmatrix} 1.38 \\ 0.30 \\ 0.72 \end{bmatrix}$$ $$\beta_d=[\beta_{d,k}]=\begin{bmatrix} 0.24 \\ 1.04 \\ 2.00 \end{bmatrix}$$
$\mathbf{s}_g$ for $g=1\ldots G$	Gear selectivity composition vectors	$\mathbf{s}_g=\begin{bmatrix} 0.03 \\ 0.02 \\ 0.95 \end{bmatrix}$ for all g	$\begin{bmatrix} 0.03 \\ 0.02 \\ 0.95 \end{bmatrix}$
$\mathbf{m}$	Average weights by class	$\mathbf{m}=\begin{bmatrix} 100 \\ 50 \\ 70 \end{bmatrix}$	$\begin{bmatrix} 100 \\ 50 \\ 70 \end{bmatrix}$

composition) is known, then it would be better to predict that the cluster composition is the same as that of the hotspot. Therefore, in BASCET, the context of a cluster (i.e. the hotspot it is contained in and the clusters or hauls in that hotspot) provides information about the composition of that cluster. Those familiar with the Kriging approach may think of the use of hotspots as not unlike using a step-function variogram model (assuming that the covariance between random composition vectors could be defined).

BASCET begins an analysis of a set of acoustic data by generating a configuration of hotspots compatible with the data, often using hotspotgen. A configuration of hotspots is said to be compatible with the data under two conditions: first, for any school cluster C, there is exactly one hotspot H such that C is within temporal and spatial range of H; second, for any trawl sample T, there is exactly one hotspot H such that the start of haul

T is within temporal and spatial range of H. The hotspotgen algorithm takes arguments that specify the spatial and temporal ranges of the generated hotspots. Of course, an experienced operator can modify the hotspots to make them correspond to known oceanic features (e.g. shelf breaks or ocean fronts). One means of introducing expert supervision of the process is to have the expert specify a few hotspots near important features and then have hotspotgen complete the task, ensuring compatibility with the data.

Once a hotspot configuration is determined and priors have been chosen for model parameters, BASCET uses Markov Chain Monte Carlo (MCMC) to generate a sample of parameter values from the joint posterior distribution. An introduction to MCMC methods is given in Appendix 1. All estimates in this paper are derived from the posterior samples generated by MCMC. Appendix 3 of the companion paper

Table 3. Performance results for the BASCET and independence scheme estimation methods using the simulation parameters from Tables 1 and 2. The first eight clusters were simulated from one hotspot and the others were simulated from another.

	Performance results								
	True composition			Independence scheme estimate			BASCET estimate		
Cluster	Adult pollock	Juvenile pollock	Jellyfish	Adult pollock	Juvenile pollock	Jellyfish	Adult pollock	Juvenile pollock	Jellyfish
1	0.036	0.055	0.908	0.204	0.303	0.494	0.149	0.185	0.666
2	0.681	0.009	0.311	0.981	0.001	0.018	0.972	0.002	0.025
3	0.997	0.000	0.002	1.000	0.000	0.000	1.000	0.000	0.000
4	0.033	0.390	0.577	0.096	0.406	0.498	0.084	0.348	0.568
5	0.505	0.013	0.482	0.914	0.004	0.082	0.921	0.008	0.071
6	0.000	0.819	0.181	0.000	1.000	0.000	0.000	1.000	0.000
7	0.875	0.004	0.121	0.974	0.001	0.025	0.919	0.002	0.080
8	0.893	0.004	0.103	0.994	0.000	0.006	0.977	0.000	0.023
9	0.000	1.000	0.000	0.000	0.999	0.000	0.000	0.999	0.000
10	0.906	0.089	0.005	0.984	0.004	0.012	0.968	0.005	0.027
11	0.043	0.941	0.016	0.015	0.907	0.077	0.018	0.928	0.053
12	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
13	0.067	0.906	0.026	0.343	0.430	0.227	0.338	0.401	0.261
14	0.902	0.093	0.005	0.995	0.000	0.004	0.992	0.001	0.007
15	0.016	0.972	0.012	0.024	0.885	0.091	0.030	0.894	0.075
16	0.321	0.648	0.031	0.783	0.062	0.155	0.746	0.083	0.171

(Hammond *et al.*, 2001) provides a glossary of BASCET jargon that may be useful here. The parameters of the BASCET model will shortly be introduced using a graphic representation of the estimation model. To assist the reader in interpreting this representation, a short introduction to graphic modelling is presented next.

Graphic models

We use a directed acyclic graph (DAG) to present the estimation model. This method provides a simple description of the relationships between model parameters. It also provides a basis for rapidly determining conditional independence relationships between parameters, and allows the joint distribution of model parameters to be determined from "local" dependencies in the graph. In our DAG, the parameters of the model form the vertices of the graph, and causal relationships between parameters (often called variables, in accordance with Bayesian norms) are indicated by directed edges from cause to effect. The wording of family relationships is used in discussing DAGs, so, for example, if there is an edge from variable A to variable B, one describes A as a parent of B or B as a child of A. Such parental relationships are extrapolated to define both the set of ancestors of a given variable and the set of its descendants.

When the relationships between a set of variables (V) are represented with a DAG, a number of conditional independence relationships are implied. In general, any variable is considered conditionally independent of all non-descendants, given the values of the variable's parents. From this interpretation it follows that the DAG can be used to determine all conditional independence relationships between variables in V. In other words, one may determine whether any two variables are conditionally independent, given evidence about the state of any variables in V. Rules for determining such conditional independence relationships are presented in Jensen (1996).

The joint probability distribution of the variables in V may be factored as follows:

$$p(V) = \prod_{v \in V} p(v \mid parents(v))$$

This factorization is useful because it eliminates the need to present the joint distribution of parameters. Instead, only the conditional distribution of each parameter given its parents is required.

The BASCET DAG

Figure 2 is a graphic representation of the BASCET model. In this figure, the edges between variables are of two types: dashed grey arrows and solid black ones. The dashed grey arrows indicate a deterministic relationship between variables, whereas the solid black arrows indicate a stochastic relationship. Thus, a broken grey arrow
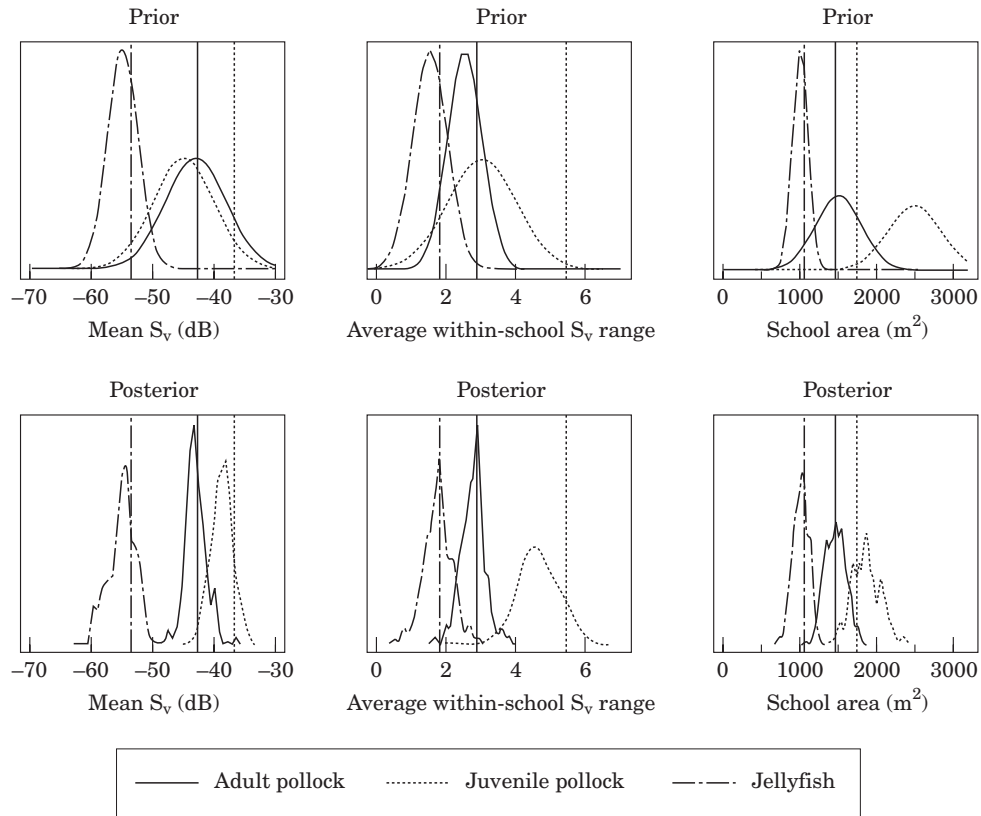
Figure 3. In the upper three panels, priors for the mean of each acoustic feature are shown for each of three species classes. Values used in the simulation exercise are shown as vertical lines for each class. The lower panels show posterior distributions computed using the BASCET estimation method.

from x to y implies that y is a function of x, and a solid black arrow indicates that the distribution of y depends on the state of x. The vertices (or nodes) of the graph are also of two types: circles and rectangles. The rectangles indicate constants, and the circles indicate general random variables. By constants, we mean quantities and vectors that are not treated as uncertain. Those variables that appear in a standard set of acoustic data are shaded grey.

Rectangular plates are used to indicate repeated structures. For example, the large plate in Figure 2 indicates that each of the I hotspots produced by the hotspotgen algorithm has an associated composition vector p_i, an associated spatial range R_i^S, an associated temporal range R_i^T, and an associated location in space-time indicated by x_i. Each hotspot i is associated with J_i school clusters and L_i hauls (note that either J_i or L_i could be 0, but not simultaneously) and variables associated with these objects are arranged on subplates. Naturally, the locations and sizes of the hotspots placed by hotspotgen determine the number of clusters and hauls associated with each hotspot.

The purpose of BASCET is to estimate all cluster composition vectors c_j. Figure 2 indicates that these compositions are deterministic functions of the respective hotspot parent composition, of environmental features (e_j) at the cluster location (x_j), and of the depth of the cluster divided by local ocean depth (δ_j). Parameter vector ξ_b controls the effect of the bth environmental attribute in vector e_j on cluster composition, whereas parameter vectors α_d and β_d control the effect of δ_j on cluster composition. The distribution of the ath acoustic feature of cluster j (in vector f_j) is determined from c_j, from the number of schools in the cluster (n_j), and from acoustic feature distribution parameters μ_a and σ_a.

The composition of trawl sample l (c_l) is determined with the same parameters used to determine school cluster compositions (c_j). This vector is then adjusted using the selectivity vector s_g that corresponds to the gear type (g_l) used in haul l, giving c_{Tl}, the final expected composition of the trawl sample. The numerical composition of the catch is denoted by z_l. As indicated in Figure 2, BASCET treats the total number of individuals caught (Z_l) as if it had been determined *a priori*.
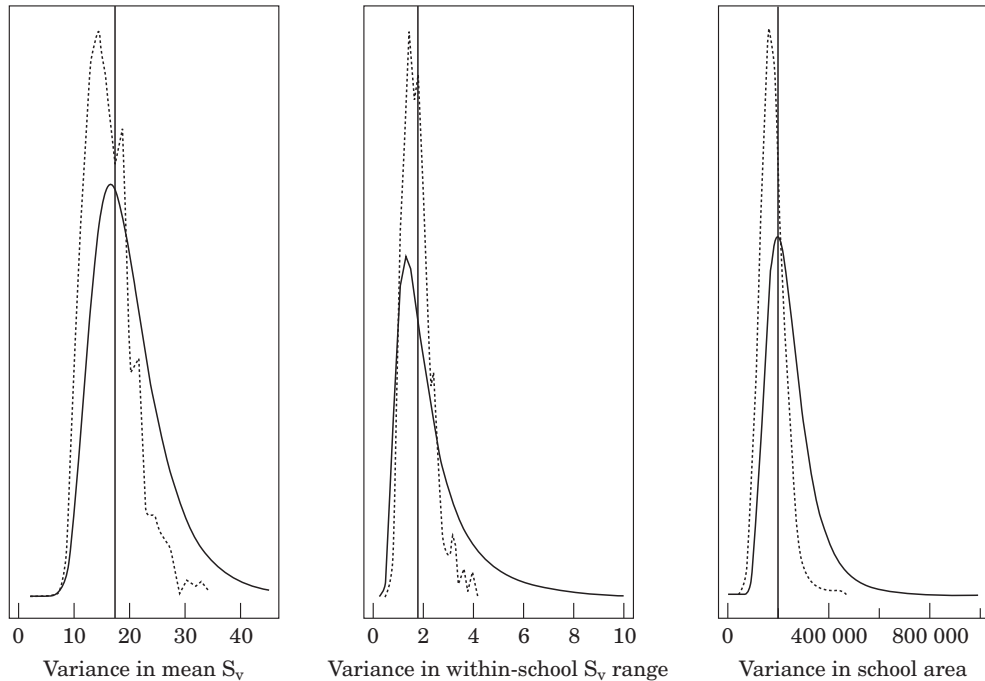
Figure 4. Prior distributions for the variance of each acoustic feature (solid lines), posterior distributions (broken lines), and true simulation values (solid vertical lines).

This departure from reality was taken in order to avoid modelling the size of the trawl catch. Finally, vector **m** is used to convert the numerical composition of the catch into the weight-by-class vector $\mathbf{w}_1$ actually observed in the acoustic data set.

All equations describing the relationships between the parameters of Figure 2 are presented in Appendix 2. That Appendix also provides the techniques used to update individual parameters in MCMC.

The independence scheme

The independence scheme provides a "control" case against which to compare performance. It is designed around the principle that no context sensitive anticipation should be attempted. In other words, the spatial and temporal proximity of clusters should have no bearing on composition estimates. In this sense, it is similar to standard classification tools.

Such is the flexibility of the hotspot model that the independence scheme is easily incorporated within it. Saying that no context sensitive anticipation should be attempted is the same as saying that each cluster should be located within its own hotspot. As this can be accomplished easily within the hotspotgen algorithm (by simply making parameters $R_{cluster}^{S}$ and $R_{cluster}^{T}$ very small; see Appendix 2), the independence scheme can be viewed as a special case of BASCET. Therefore, the same estimation algorithm can be used.

For the simulation study presented in this paper, the arguments to the hotspotgen algorithm in BASCET were set as follows: $R_{trawl}^{S}=5$ km, $R_{trawl}^{T}=3$ h, $R_{cluster}^{S}=5$ km and $R_{cluster}^{T}=3$ h. For the independence scheme, trawl-related arguments were kept the same, but the values used for the cluster-related arguments were $R_{cluster}^{S}=0.01$ km, and $R_{cluster}^{T}=0.01$ h.

Results

The parameters of the simulation model were drawn at random from prior distributions elicited from an acoustic expert (Professor Gordon Swartzman) using techniques described in the section of Hammond *et al.* (2001) entitled "Configuring BASCET". The prior distributions are shown in Tables 1 and 2 along with a parameter set drawn randomly from them. A full explanation of the role of each parameter is given in Appendix 2. Each set of parameters was used to generate acoustic data using the simulation model of Hammond (2000).

The performance of the two estimation approaches was compared on the data generated from the set of simulation model parameters in Tables 1 and 2. Performance was evaluated using an estimation loss function that measures how close each estimated composition vector is to the true value. If, for each school cluster j, the estimated composition is $\mathbf{e}_j$, the number of
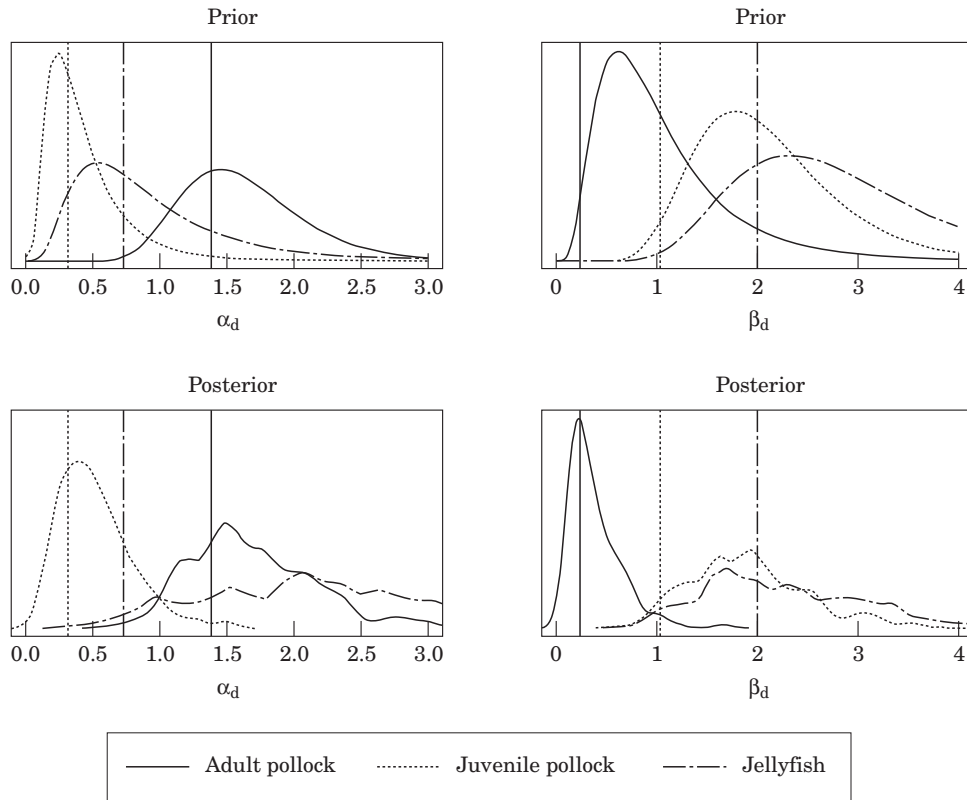
Figure 5. Prior and posterior results for depth distribution parameters. True values are shown as vertical lines. The parameters determine the Beta distribution that describes the depth preferences of each class. It is helpful to bear in mind that the mean of a Beta distribution is $\alpha_d/(\alpha_d + \beta_d)$.

schools is n_j, and the true composition is $\mathbf{c}_j$, then the estimation loss is given by:

$$\text{loss} = \frac{\sum\limits_{j=1}^{j} n_j(\text{alr}(\mathbf{c}_j) - \text{alr}(\mathbf{e}_j))'(\text{alr}(\mathbf{c}_j) - \text{alr}(\mathbf{e}_j))}{\sum\limits_{j=1}^{J} n_j}$$

wherein the additive log ratio (alr) transformation is defined in Appendix 2. We used the alr transformation in measuring loss because the elements of untransformed composition vectors must sum to 1.

Point estimates of cluster composition were obtained from each approach using 300 000 iterations of MCMC (by averaging cluster composition estimates over the MCMC iterations, see Appendix 1). In this simulation experiment, BASCET was superior to the independence scheme at the end of 300 000 iterations (loss=4.65 vs. 6.52).

Table 3 shows the true school cluster compositions generated by the simulation model next to estimates of those from BASCET after 300 000 iterations of MCMC. It also gives the results for the independence scheme.

The rate of successful identification of the dominant class was 15/16 in both estimation schemes.

Estimates of classification parameters can reveal flaws in the estimation method. To this end, BASCET estimates of key classification parameters were compared with simulated values. Figure 3 displays such a comparison for the mean (μ_a) of each acoustic feature. Vertical lines in this figure show the true values for each class and acoustic feature. As one would hope, the posterior distributions are clustered tightly around the true values. Figure 4 makes a similar comparison for the variances in acoustic features (σ_a^2). In Figure 4, the posterior distributions are also clustered tightly around the true values.

Figure 5 compares the prior, posterior, and true values for depth distribution parameters $\alpha_{d,k}$ and $\beta_{d,k}$, for classes $k=1 \ldots 3$. The posterior distributions for the three $\beta_{d,k}$ parameters are sensible, but the alpha parameter associated with jellyfish has a posterior that seems surprisingly broad. Still, because there were only two clusters dominated by jellyfish (Table 3), and because the true value was plausible as a random sample from this posterior, we saw no reason for concern.

The upper part of Figure 6 shows contours of the prior distribution for the overall class composition in the

Prior for expected hotspot composition

Adult pollock

Juvenile pollock   Jellyfish

Temperature effect prior

Adult pollock

Juvenile pollock   Jellyfish

Posterior for expected hotspot composition

Adult pollock

Juvenile pollock   Jellyfish

Temperature effect posterior
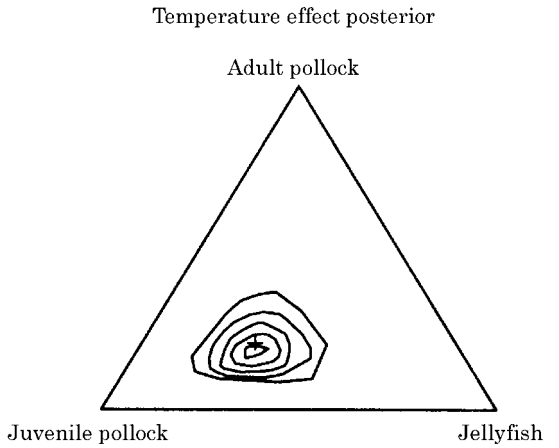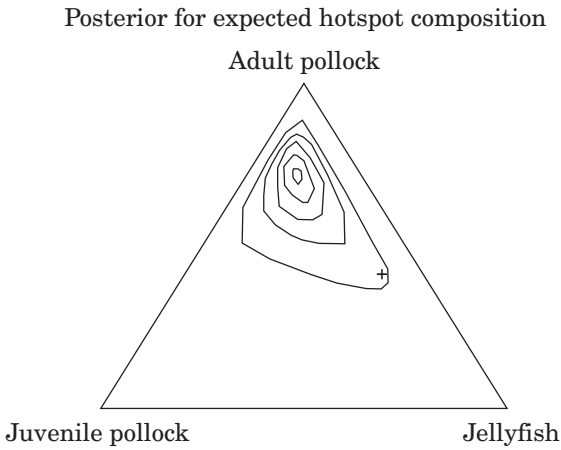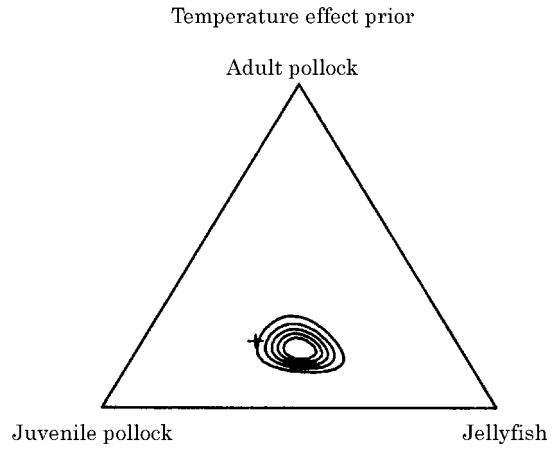
Adult pollock

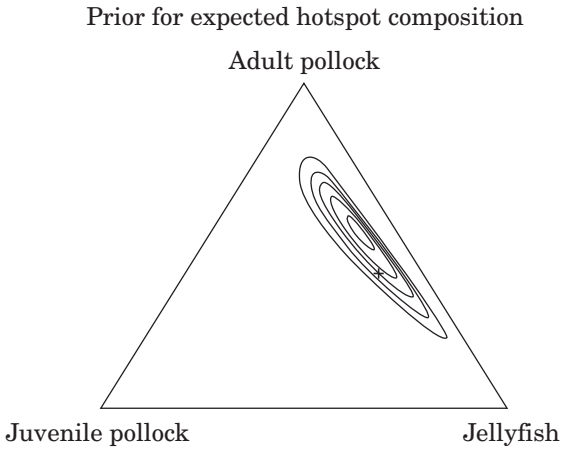Juvenile pollock   Jellyfish

Figure 6. Contour lines of the prior distribution for expected hotspot composition. The true value is shown with a cross. The vertices of the triangle are pure compositions, and points near the centre are even mixtures of the classes. In the lower triangle plot, the posterior distribution is indicated by 95%, 75%, 50%, 25%, and 5% credible regions.

Figure 7. Contour lines of the prior distribution for the composition that represents the effect of a temperature increase of one degree. The true value is shown with a cross. The vertices of the triangle are pure compositions, and points near the centre are even mixtures of the classes. In the lower triangle plot, the posterior distribution is indicated by 95%, 75%, 50%, 25%, and 5% credible regions.

survey region ($\mathbf{p}_H$) using a triangle plot. The true overall class composition is shown with a cross. The vertices of the triangle indicate pure compositions, whereas the point at the centre of the triangle indicates an even mixture of classes. The posterior estimate for $\mathbf{p}_H$ from the BASCET method is shown in the lower part of Figure 6, using 95%, 75%, 50%, 25%, and 5% credible regions. The true expected hotspot composition ($\mathbf{p}_H$) is not recovered very accurately, but it is within the 95% credible region.

The upper part of Figure 7 shows contours of the prior distribution for ξ_{temp}, the composition vector that describes the effect of temperature on cluster composition. The value used in the simulation experiment is

shown with a cross. In the lower part of Figure 7, the posterior distribution for ξ_{temp} from the BASCET model is described with 95%, 75%, 50%, 25%, and 5% credible regions. It is encouraging to see the posterior shifting over onto the true value.

Discussion

The performance results above are consistent with our claims about the value of context-sensitive anticipation. As predicted, trying to anticipate the composition of school clusters appears to improve estimation results.

This improvement occurs despite the fact that the hotspots used in the anticipation process are not observed at all. It must be said, however, that the study was performed with just a single set of simulation parameters, sampled randomly from prior distributions. It is conceivable, therefore, that the improved performance is particular to the simulation parameters used. Moreover, the similarities between the simulation model and the estimation tool (e.g. both use hotspots) might possibly have made context-sensitive anticipation perform better than it would on real data, even though the hotspots used in simulation were not available to BASCET. These limitations are, of course, common to many a simulation study.

BASCET is general enough to improve with developments in acoustic technology. As technology advances, new features will be discovered that will help discriminate between the classes (Scalabrin *et al.*, 1996). For example, recent developments in wide-band technology promise improved discrimination (Zakharia *et al.*, 1996). The estimation tools easily allow for the addition of new acoustic features. The incorporation of additional environmental variables is also straightforward.

Another valuable feature of the BASCET approach is its transferability. If the estimation tool has been applied to an area, and a new acoustic survey is carried out nearby, then the prior distributions allow the user to incorporate some or all of the old results in the new survey. Old parameter estimates can be incorporated to the degree that they are deemed relevant. In a general estimation approach this feature is essential, and using a Bayesian approach is probably the simplest way to achieve it.

Despite encouraging performance results, several problems with BASCET are apparent. It is not very satisfying that the algorithm asks the user to specify hotspotgen arguments and selectivity parameters. It would be better if such selection was guided by the data. These problems will have to be addressed in future research. There were also some indications of bias in the recovery of secondary model parameters, especially the average hotspot composition (although the true values were not outside 95% credible regions). The focus of this paper is firmly fixed, however, on making good cluster-composition estimates. By this standard, the results are certainly encouraging. On simulated data, the central prediction about the value of context-sensitive anticipation appears to be vindicated.

Acknowledgements

References

Aitchison, J. 1986. The statistical analysis of compositional data. Chapman & Hall, New York.

Aitchison, J., and Shen, S. M. 1980. Logistic normal distributions: some properties and uses. Biometrika, 67: 261–272.

Barange, M. 1994. Acoustic identification, classification and structure of biological patchiness on the edge of the Agulhas Bank and its relation to frontal features. South African Journal of Marine Science, 14: 333–347.

Baussant, T., Ibanez, F., and Étienne, M. 1993. Numeric analysis of planktonic spatial patterns revealed by echograms. Aquatic Living Resources, 6: 175–184.

Billheimer, D. D. 1995. Statistical analysis of biological monitoring data: state-space models for species compositions. PhD thesis, Statistics Department, University of Washington, Seattle.

Gelfand, A. E., and Smith, A. F. M. 1990. Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association, 85: 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 1995. Bayesian data analysis. Chapman & Hall, London.

Gerlotto, F. 1993. Identification and spatial stratification of tropical fish concentrations using acoustic populations. Aquatic Living Resources, 6: 243–254.

Hammond, T. R. 2000. Classification of fish schools from acoustic survey data. PhD thesis, School of Fisheries, University of Washington, Seattle.

Hammond, T. R., Swartzman, G. L., and Richardson, T. S. 2001. Bayesian estimation of fish school cluster composition applied to a Bering Sea acoustic survey. ICES Journal of Marine Science, doi:10.1006/jmsc.2001.1093.

Haralick, R., and Shapiro, L. 1992. Computer and robot vision. Addison Wesley, Reading, MA.

Haralick, R., Sternberg, J., and Zhang, C. 1987. Image analysis using mathematical morphology. IEEE Transactions, PAMI-9, 4: 532–550.

Haralabous, J., and Georgakarakos, S. 1996. Artificial neural networks as a tool for species identification of fish schools. ICES Journal of Marine Science, 53: 173–180.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57: 97–109.

James, M. 1985. Classification algorithms. John Wiley, New York.

Jensen, F. V. 1996. An introduction to Bayesian networks. UCL Press, London.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. 1953. Equations of state calculations by fast computing machines. Journal of Chemistry and Physics, 21: 1087–1091.

Nero, R. W., and Magnuson, J. J. 1989. Characterization of patches along transects using high-resolution 70 kHz integrated acoustic data. Canadian Journal of Fisheries and Aquatic Sciences, 46: 2056–2064.

Nero, R. W., Magnuson, J. J., Brandt, S. B., Stanton, T. K., and Jech, J. M. 1990. Finescale biological patchiness of 70 kHz acoustic scattering at the edge of the Gulf Stream. EchoFront 85. Deep-Sea Research, 37: 999–1016.

Reid, D. G., and Simmonds, E. J. 1993. Image analysis techniques for the study of fish school structure from acoustic survey data. Canadian Journal of Fisheries and Aquatic Sciences, 50: 886–893.

Richards, L. J., Kieser, R., Mulligan, T. J., and Candy, J. R. 1991. Classification of fish assemblages based on two echo integration surveys. Canadian Journal of Fisheries and Aquatic Sciences, 48: 1264–1272.

Rose, G. A., and Leggett, W. C. 1988. Hydroacoustic signal classification of fish schools by species. Canadian Journal of Fisheries and Aquatic Sciences, 45: 597–604.

Scalabrin, C., Diner, N., Weill, A., Hillion, A., and Mouchot, M.C. 1996. Narrowband acoustic identification of mono-specific fish shoals. ICES Journal of Marine Science, 53: 181–188.

Simmonds, E. J., Williamson, N., Gerlotto, F., and Aglen, A. 1992. Acoustic survey design and analysis procedure: a comprehensive review of current practice. ICES Cooperative Research Report 187.

Simmonds, E. J., Armstrong, F., and Copland, P. J. 1996. Species identification using wideband backscatter with neural network and discriminant analysis. ICES Journal of Marine Science, 53: 189–195.

Swartzman, G. L., Stuetzle, W., Kulman, K., and Powojowski, M. 1994a. Relating the distribution of pollock schools in the Bering Sea to environmental factors. ICES Journal of Marine Science, 51: 481–492.

Swartzman, G. L., Stuetzle, W., Kulman, K., and Wen, N. 1994b. Modeling the distribution of fish schools in the Bering Sea: morphological school identification. Natural Resource Modeling, 8: 177–194.

Tanner, M. A. 1996. Tools for statistical inference: methods for exploration of posterior distributions and likelihood functions. Springer, New York.

Weill, A., Scalabrin, C., and Diner, N. 1993. MOVIESB: an acoustic detection description software. Application to shoal species classification. Aquatic Living Resources, 6: 255–267.

Zakharia, M. E., Magand, F., Hetroit, F., and Diner, N. 1996. Wideband sounder for fish species identification at sea. ICES Journal of Marine Science, 53: 203–208.

Appendix 1: Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methodology is useful for drawing inference about a statistical distribution (F) whose density (f) can be identified up to a constant of proportionality. As this scenario arises constantly in Bayesian estimation problems, MCMC has seen its greatest use in that field. However, the technique is not necessarily limited to Bayesian problems. For a given distribution F, MCMC is able to accomplish two valuable tasks: it can draw a sample from it, or it can compute the mean and variance. Therefore, in Bayesian applications, MCMC is used either to draw a sample from the posterior or to compute point estimates and variances for parameters of interest. Both results are approximations, as one would expect. A particular strength of MCMC is its ability to handle problems with missing data or latent variables (Tanner, 1996). As the

hotspots introduced in this paper are latent variables, MCMC should be suitable for the BASCET model.

Within the MCMC field, there are a number of algorithms from which to choose. Three of these are the Gibbs sampler (Tanner, 1996, p. 137), the Metropolis algorithm (Metropolis *et al.*, 1953; Tanner, 1996, p. 174) and the Hastings algorithm (Hastings, 1970). The Gibbs sampler is a special case of the Metropolis algorithm (Tanner, 1996, p. 181) and both are special cases of the Hastings algorithm. All three methods proceed from an arbitrarily chosen starting vector ($\mathbf{x}^1$). All are designed to create a Markov Chain of parameter vectors ($\mathbf{x}^1$, $\mathbf{x}^2$, $\mathbf{x}^3$, . . ., $\mathbf{x}^n$) that has F as its equilibrium distribution. In other words, as n tends to infinity, $\mathbf{x}^n$ converges in distribution to a random variable $\mathbf{x}$ with distribution F, and furthermore

$$\frac{1}{n} \sum_{i=1}^{n} \phi(\mathbf{x}^i) \xrightarrow{a,s} E_F[\phi(\mathbf{x})]$$

for any integrable function ϕ.

In practice, the chain is used in one of two ways: either one treats every mth vector beyond a certain index b in the chain as a sample from F, or one computes

$$\frac{1}{(n-b)} \sum_{i=b+1}^{n} \mathbf{x}^i$$

as an estimator for the expected value of F. Index b, often called the burn-in period, is needed in order to give the algorithm a chance to ''forget'' the fact that it started from an arbitrary initial vector. Parameter m is used to reduce the dependence between samples in the chain and to reduce disk storage space.

The Hastings algorithm moves from one parameter vector $\mathbf{x}_i$ to the next by proposing moves from a transition probability function $q(\mathbf{x}^i, \mathbf{x}')$. Proposals $\mathbf{x}'$ from q are either accepted, in which case $\mathbf{x}^{i+1} = \mathbf{x}'$, or rejected, in which case $\mathbf{x}^{i+1} = \mathbf{x}^i$. The key is to accept or reject these moves with probability p given by the equation

$$p = \min\left(1, \frac{q(\mathbf{x}', \mathbf{x}^i) f(\mathbf{x}')}{q(\mathbf{x}^i, \mathbf{x}') f(\mathbf{x}^i)}\right)$$

As one would imagine, it is necessary that the proposal distribution q be such that there always exists a sequence of moves that will take any vector in the domain of f to any other vector in that domain.

The Metropolis algorithm is a special case of the Hastings algorithm in which the proposal distribution q is symmetric. This has the advantage that the terms involving q drop out of the acceptance probability. In all other respects it is like the Hastings algorithm.

The Gibbs sampler is the most efficient of the MCMC methods because proposals are never rejected. On the

other hand, it requires analytical computations that will not always be feasible. In the Gibbs sampler, proposals are made to each parameter in the vector $\mathbf{x}$ individually. It makes use of the following observation: in many cases where the posterior distribution $\pi(\mathbf{x}|\mathbf{y}) = \pi(x_1, x_2, \ldots, x_p|\mathbf{y})$ is difficult to compute, the full conditional distributions $\pi(x_k|x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p, \mathbf{y}) = \pi(x_k|x_{-k}, \mathbf{y})$ can still be obtained analytically. When this is so, these conditional distributions can be used to construct a Markov Chain.

Given the starting point $(x_1^1, x_2^1, \ldots, x_p^1)$ at which $i=1$, the Gibbs sampler iterates the following loop:

(1) Sample x_1^{i+1} from $\pi(x_1|x_2^i, \ldots, x_p^i, \mathbf{y})$
(2) Sample x_2^{i+1} from $\pi(x_2|x_1^{i+1}, x_3^i, \ldots, x_p^i, \mathbf{y})$ if $p>2$.
(3) Sample x_3^{i+1} from $\pi(x_3|x_1^{i+1}, x_2^{i+1}, x_4^i, \ldots, x_p^i, \mathbf{y})$ if $p>3$.
(4) Sample x_4^{i+1} from $\pi(x_4|x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_5^i, \ldots, x_p^i, \mathbf{y})$ if $p>4$.
(5) . . .
(6) Sample x_p^{i+1} from $\pi(x_p|x_1^{i+1}, \ldots, x_{p-1}^{i+1}, \mathbf{y})$.
(7) Increment i.

When this is done, the vectors $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \ldots$ form the required Markov Chain.

It is possible to combine elements from several algorithms in the same chain. Such algorithms are called hybrid schemes. For example, if analytical conditional results ($\pi(x_k|x_{-k}, \mathbf{y})$) are available for a subset of parameters, one could update these parameters as prescribed by the Gibbs sampler. Each of the remaining parameters could then be assigned their own transition probability function (q) and would be updated according to the Hastings algorithm. The resulting Markov Chain could then be used for inference just as before.

MCMC sampling can result in serious errors in parameter estimates. When using MCMC, it is important to evaluate whether one is using sufficient iterations (n), and burn-in (b). In a simulation study, like the one in this paper, poor correspondence between true and estimated parameters indicates that more iterations or burn-in are required. In a real application, however, such comparison is impossible, and techniques such as those described in Appendix 2 to Hammond *et al.* (2001) are necessary.

Appendix 2: BASCET

BASCET chooses an initial configuration of hotspots using an algorithm called hotspotgen. This algorithm has four arguments: R_{trawl}^S, R_{trawl}^T, $R_{cluster}^S$, and $R_{cluster}^T$. The first two of these arguments define the spatial and temporal ranges of hotspots created around trawl samples. The second two define the same characteristics for hotspots created around school clusters. The algorithm works as follows:

(1) Start with a set Ξ composed of all the trawl samples and another set Θ of hotspots. Initially, Θ might be empty or might contain some operator-specified hotspots.
(2) If no trawl samples remain in Ξ, go to line 8. Otherwise, select a trawl sample (t) at random from Ξ without replacement.
(3) If the start of t is contained within any hotspot in Θ, then go back to line 2.
(4) Create a hotspot h at the start of t with spatial range R_{trawl}^S and temporal range R_{Trawl}^T.
(5) Add h to Θ.
(6) Merge any hotspots in Θ that overlap with each other.
(7) Go back to line 2.
(8) Construct a set Φ composed of all the school clusters.
(9) If no clusters remain in Φ, then STOP. Otherwise, select a school cluster (c) at random from Φ without replacement.
(10) If c is contained within any hotspot in Θ, then go to 9.
(11) Create a hotspot h at the location of c with spatial range $R_{cluster}^S$ and temporal range $R_{cluster}^T$.
(12) Add h to Θ.
(13) Merge any hotspots in Θ that overlap with each other.
(14) Go back to line 9.

Estimation

In the BASCET method, estimation is achieved using a hybrid MCMC scheme. In this scheme, model parameters are updated in turn, as in the description of the Gibbs sampler in Appendix 1. Whenever full conditional posterior results (see Appendix 1) can be computed analytically for a particular parameter, the parameter is updated as prescribed by the Gibbs sampler. Whenever such results are not available, the parameter is assigned a transition probability function q from which moves were proposed. These moves were accepted or rejected in accordance with the rules of the Hastings algorithm.

In the next section, the relationships between parameters of the BASCET model are defined. Certain of the parameters are assigned prior distributions, and, in many cases, analytical posterior results are available that are conditional on the values of all other parameters (often called full conditional distributions). When such analytical results are available, they will be presented along with the parameter description. When such results are not available, a proposal distribution (also called a transition probability function) will be presented instead. Thus, the description of the previous paragraph should allow the reader to determine how each parameter is updated in the MCMC scheme.

Hotspot related parameters

Following a precedent established in Billheimer (1995), hotspot composition was assigned an Additive Logistic Normal (ALN) distribution. Aitchison and Shen (1980) introduced this distribution, and its use in the analysis of compositional data is chronicled in Aitchison (1986). Here it is introduced using two vector functions: alr and ialr. The alr transformation is defined as

$$\mathrm{alr}(\mathbf{z}) = \left[\log\left(\frac{z_1}{z_K}\right), \cdots, \log\left(\frac{z_k}{z_K}\right), \cdots, \log\left(\frac{z_{K-1}}{z_K}\right)\right]$$

and its inverse is given by

$$\mathrm{ialr}(\mathbf{y}) = \frac{[\exp(y_1), \cdots, \exp(y_k), \cdots, \exp(y_{K-1}), 1]}{\sum\limits_{j=1}^{K-1}\exp(y_j) + 1}$$

The ALN distribution has a density:

$$p(\mathbf{z}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{\frac{1-K}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\left(\exp\left(-\frac{1}{2}(\mathrm{alr}(\mathbf{z}) - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathrm{alr}(\mathbf{z}) - \boldsymbol{\mu})\right)\right)\prod_{k=1}^{K}z_k^{-1}$$

where $\mathbf{z}$ is a K dimensional composition vector whose elements add up to 1.

The composition $\mathbf{p}_i$ of hotspot i was modelled as:

$$\mathbf{p}_i \sim \mathrm{ALN}(\mathrm{alr}(\mathbf{p}_H), \boldsymbol{\Sigma}_H)$$

In turn, $\mathbf{p}_H$ was assigned the prior distribution

$$\pi(\mathbf{p}_H) = \mathrm{ALN}(\boldsymbol{\mu}_H, \boldsymbol{\Phi}_H)$$

In the graphic model presented in Figure 2, it is apparent that the full conditional distribution for parameter $\mathbf{p}_H$ reduces to a distribution that depends only on $\boldsymbol{\Sigma}_H$ and on the hotspot compositions $\mathbf{p}_i$ for i=1...I. It also follows from the specification above that the full conditional distribution for $\mathbf{p}_H$ is ALN with mean

$$\boldsymbol{\mu}'_H = [I\boldsymbol{\Sigma}_H^{-1} + \boldsymbol{\Phi}_H^{-1}]^{-1}(I\boldsymbol{\Sigma}_H^{-1}\bar{\mathbf{x}} + \boldsymbol{\Phi}_H^{-1}\boldsymbol{\mu}_H)$$
$$= (\mathbf{I}_{K-1} - 1/I\boldsymbol{\Sigma}_H[1/I\boldsymbol{\Sigma}_H + \boldsymbol{\Phi}_H]^{-1})$$
$$\bar{\mathbf{x}} + 1/I\boldsymbol{\Sigma}_H[1/I\boldsymbol{\Sigma}_H + \boldsymbol{\Phi}_H]^{-1}\boldsymbol{\mu}_H$$

and variance matrix

$$\boldsymbol{\Phi}'_H = [I\boldsymbol{\Sigma}_H^{-1} + \boldsymbol{\Phi}_H^{-1}]^{-1}$$

where I is the number of hotspots, $\mathbf{I}_{K-1}$ is a K−1 dimensional identity matrix, and

$$\bar{\mathbf{x}} = \sum_{i=1}^{I}\mathrm{alr}(\mathbf{p}_i)/I.$$

This result was adapted from Gelman *et al.* (1995), p. 79).

In accordance with standard Bayesian practice (Gelman *et al.*, 1995), a Wishart prior distribution was specified for $\boldsymbol{\Sigma}_H^{-1}$:

$$\pi(\boldsymbol{\Sigma}_H^{-1}) = \mathrm{Wishart}(\rho, \boldsymbol{\Omega}_H),$$

with expected value $\rho\boldsymbol{\Omega}_H$ and degrees of freedom ρ. The Wishart distribution (Gelman *et al.*, 1995), a multivariate analog of the Gamma family, has density proportional to

$$p(\mathbf{W}|\rho,\mathbf{Q}) \propto |\mathbf{W}|^{\frac{\rho-K}{2}}|\mathbf{Q}|^{\frac{-\rho}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{Q}^{-1}\mathbf{W})\right),$$

where $\mathbf{W}$ and $\mathbf{Q}$ are K−1 dimensional, symmetric, positive-definite matrices.

For the Wishart to be a proper distribution, the degrees of freedom ρ must be at least K−1. The prior becomes increasingly informative with increasing ρ, so the minimum value will often be appropriate. Suggestions for eliciting prior parameter $\boldsymbol{\Omega}_H$ from an expert are given in Hammond *et al.* (2001).

It is readily seen in Figure 2 that the full conditional distribution for $\boldsymbol{\Sigma}_H^{-1}$ reduces to a distribution that depends on $\mathbf{p}_H$ and the hotspot compositions $\mathbf{p}_i$ for i=1 ... I. The good thing about choosing a Wishart prior is that the full conditional distribution of $\boldsymbol{\Sigma}_H^{-1}$ is also Wishart-distributed:

$$\pi(\boldsymbol{\Sigma}_H^{-1}|\mathbf{p}_H,\{\mathbf{p}_i\}_{i=1}^{I}\alpha\mathrm{Wishart}(\rho+I,[\boldsymbol{\Omega}_H^{-1}+\mathbf{S}]^{-1})$$

where

$$\mathbf{S} = \sum_{i=1}^{I}(\mathrm{alr}(\mathbf{p}_i) - \mathrm{alr}(\mathbf{p}_H))(\mathrm{alr}(\mathbf{p}_i) - \mathrm{alr}(\mathbf{p}_H))'$$

Gelfand and Smith (1990) provide an algorithm for sampling from a Wishart distribution, and this Wishart posterior result is also given in that paper.

No analytical results are available for the full conditional distribution of individual hotspot compositions $\mathbf{p}_i$. Instead, (as part of MCMC inference) new proposals for these compositions are generated from an ALN distribution with expected location given by the current value. These proposals are accepted and rejected in accordance with the Hastings algorithm (Appendix 1).

Acoustic parameters

The distribution for acoustic feature mean a for a cluster with composition c was modelled by the equation

$$f_a \sim N\left(\sum_{k=1}^{K}\mu_{a,k}c_k, \sigma_a^2\right)$$

Each of the $\mu_{a,k}$ values was assigned a prior distribution as follows:

$$\pi(\mu_{a,k}) = N(\theta_{a,k}, \tau_{a,k}^2)$$

The σ_a^2 values were assigned prior distributions from the inverse Gamma family:

$$\pi(\sigma_a^2) \sim IG(\alpha_a, \beta_a)$$

On defining $m_{a,-k} = \Sigma_{i \neq k} c_i \mu_{a,i}$ then, under the model, the distribution of

$$x_{a,k} = \frac{f_a - m_{a,-k}}{c_k}$$

is Normal, with mean $\mu_{a,k}$ and variance σ_a^2/c_k^2. By indexing the school clusters by $j = 1 \ldots J$, one can refer to the $x_{a,k}$ value for cluster j by $x_{a,k,j}$, and to the variance in this by $v_{j,k}^2$. Then, one can extract all the information in the school cluster data about $\mu_{a,k}$ with two statistics:

$$M_{a,k} = \sum_{j=1}^{J} \frac{x_{a,k,j}}{v_{j,k}^2}$$

and

$$P_{a,k} = \sum_{j=1}^{J} \frac{1}{v_{j,k}^2}$$

As can be seen in Figure 2 (using rules specified in Jensen, 1996), the full conditional distribution for $\mu_{a,k}$ depends only on the number of schools in each cluster ($\{n_j\}_{j=1}^{J}$), on the observed values of feature a (i.e. $\{f_{a,j}\}_{j=1}^{J}$), on the school cluster compositions ($\{c_j\}_{j=1}^{J}$), on the other means $\{\mu_{a,m}\}_{m \neq k}$ and on σ_a^2. It has a Normal distribution

$$\pi(\mu_{a,k} | M_{a,k}, P_{a,k}) = N\left(\frac{\tau_{a,k}^2 M_{a,k} + \theta_{a,k}}{\tau_{a,k}^2 P_{a,k} + 1}, \frac{\tau_{a,k}^2}{\tau_{a,k}^2 P_{a,k} + 1} \right)$$

As noted above, school clusters store data about the mean of acoustic feature f_a for all n_j schools that are in the cluster in $\bar{f}_{a,j}$. For each cluster, let us define the following statistic:

$$SS_{a,j} = \left(\bar{f}_{a,j} - \sum_{k=1}^{K} c_k \mu_{a,k} \right)^2$$

Using this statistic one can compute the full conditional distribution of σ_a^2 (which depends only on the $\{\mu_{a,k}\}_{k=1}^{K}$ values and on the cluster compositions $\{c_j\}_{j=1}^{J}$) to be an inverse Gamma distribution with the following parameters:

$$\alpha_a' = \frac{J}{2} + \alpha_a$$

and

$$\beta_a' = \frac{2}{\sum\limits_{j=1}^{J} SS_{a,j} + 2/\beta_a}$$

Trawl sample parameters

The treatment of trawl data differs considerably from that in the simulation model of Hammond (2000). In BASCET, results are conditioned on the actual size of the catch. In addition, the weight conversion vector (**m**) and the gear selectivity composition vectors ($\mathbf{s}_g$, for $g = 1 \ldots G$) introduced in Figure 2 are assumed to be known without error. This may not be a reasonable assumption in some circumstances, but the sensitivity of the answers to the values chosen can be examined.

Each haul is assigned an expected composition (**c**) that depends on the hotspot parent composition, on the haul depth and on the environmental features at the haul (start) location. As indicated in the discussion of Figure 2, the calculation of **c** from these components is carried out exactly as if the haul were a school cluster. Equations for this computation are described in the next two sections. Once **c** has been obtained, gear selectivity can be addressed. In order to do so, **c** is adjusted using the gear selectivity vector $\mathbf{s}_g$, giving the final expected catch composition $\mathbf{c}_T$:

$$\mathbf{c}_T = ialr(alr(\mathbf{c}) + alr(\mathbf{s}_g)) = [c_1 s_{g,1}, \cdots, c_K s_{g,K}] / \sum_{k=1}^{K} c_k s_{g,k}$$

The trawl data were treated as if the total size of the catch was fixed before the haul was even carried out. The total catch vector ($\mathbf{w}_1$) is divided element-wise by the weight conversion vector **m**, giving a new vector $\mathbf{z}_1$. The distribution of a trawl sample catch is then taken to be multinomial:

$$\mathbf{z}_1 \sim \text{Multinomial } (Z_1, \mathbf{c}_{T,l})$$

for

$$Z_1 = \sum_{k=1}^{K} z_{l,k}$$

Depth preferences

The proportions of ocean depth at which each of the K species classes prefer to swim are assigned distributions from the Beta family. For example, the proportion of ocean depth preferred by species k (δ_k), is taken to have a distribution given by

$$\delta_k \sim \text{Beta}(\alpha_{d,k}, \beta_{d,k})$$

The following prior distributions were assigned to $\alpha_{d,k}$ and $\beta_{d,k}$:

$$\pi(\alpha_{d,k}) = \text{LogNormal}(\theta_{d,k}, \tau^2_{d,k})$$

and

$$\pi(\beta_{d,k}) = \text{LogNormal}(\gamma_{d,k}, \upsilon^2_{d,k})$$

where the distribution of LogNormal variable X is defined by the mean and variance of log(X).

All school clusters have a δ value defined as the cluster depth divided by the local ocean depth. A school cluster located within a hotspot with composition vector $\mathbf{p}$ is taken to have its distribution formed by a $\mathbf{p}$ mixture of the Beta distributions above. In other words, the value for a school cluster was taken to have density given by the following:

$$p(\delta) = \sum_{k=1}^{K} \frac{p_k \Gamma(\alpha_{d,k} + \beta_{d,k})}{\Gamma(\alpha_{d,k})\Gamma(\beta_{d,k})} \delta^{\alpha_{d,k}-1}(1-\delta)^{\beta_{d,k}-1}$$

The composition ($\mathbf{c}$) of a school cluster is taken to be a function of three variables: the local hotspot composition $\mathbf{p}$, the cluster's δ value, and the environmental variables at the cluster's location. First let us consider the effects of depth alone. Given that a cluster is located at depth δ times the local ocean depth, and that it is within a hotspot of composition $\mathbf{p}$, one can predict a composition for this cluster by using the depth preferences of the species classes. This is done by computing a depth-adjusted cluster composition $\mathbf{c}_d$ as follows:

$$\mathbf{c}_d = \frac{[p_1 f_1(\delta), \cdots, p_k f_k(\delta), \cdots, p_K f_K(\delta)]}{\sum\limits_{k=1}^{K} p_k f_k(\delta)}$$

where

$$f_k(\delta) = \frac{\Gamma(\alpha_{d,k} + \beta_{d,k})}{\Gamma(\alpha_{d,k})\Gamma(\beta_{d,k})} \delta^{\alpha_{d,k}-1}(1-\delta)^{\beta_{d,k}-1}$$

New values for the parameters affecting the depth preferences of each species class (i.e. $\alpha_{d,k}$ and $\beta_{d,k}$, for k=1 . . . K) were proposed by multiplying the old values by values generated from a LogNormal distribution with a mode of 1. These moves are accepted and rejected in accordance with the rules of the Hastings algorithm.

Environmental features

The vectors of environmental features (denoted by $\mathbf{e}$) associated with school clusters are not considered random: they are entirely determined by the cluster location. Their effects on cluster composition is modelled using composition effect vectors ξ_b as follows:

$$\text{alr}(\mathbf{c}) = \text{alr}(\mathbf{c}_d) + \sum_{b=1}^{B} (\mathbf{e}_b - \bar{\mathbf{e}}_b)\text{alr}(\xi_b)$$

The composition effect vectors are assigned prior distributions from the ALN family as follows:

$$\pi(\xi_b) = \text{ALN}(\mathbf{0}_{K-1}, q\mathbf{N})$$

where $\mathbf{0}_{K-1}$ is a zero vector of length $K-1$. New values (ξ'_b) for these compositions are proposed from an ALN distribution centred at the current value as follows:

$$p(\xi'_b | \xi_b) = \text{ALN}(\text{alr}(\xi_b), s\mathbf{N})$$