BULLETIN OF MARINE SCIENCE, 79(3): 545–559, 2006

DEVELOPING INDICES OF ABUNDANCE USING HABITAT DATA IN A STATISTICAL FRAMEWORK

Mark N. Maunder, Michael G. Hinton,
Keith A. Bigelow, and Adam D. Langley

ABSTRACT

We describe the statistical application of the habitat-based standardization (statHBS) of catch-per-unit-effort (CPUE) data to derive indices of relative abundance. The framework is flexible, including multiple component models to accommodate factors such as habitat, sampling, and animal behavior. It allows the use of prior information or the completely independent estimation of model components (e.g., habitat preference). The integration with a general linear model framework allows convenient comparison with traditional methods used to standardize CPUE data. The statistical framework allows model selection and estimation of uncertainty. The statHBS model is applied to bigeye tuna in the western and central Pacific Ocean. We describe several additional improvements to the methodology.

Knowledge of trends in population abundance is important for the development of appropriate management actions. Methods used to estimate trends in abundance range from population dynamics models and observational studies to surveys and experiments. Unfortunately, there are many factors that can influence our perception of results obtained from these methods, and these become more problematic as our control over the data collection process decreases. For example, opportunistic observational studies can be influenced by factors such as the time of year in which observations were recorded, which may change from year to year, resulting in bias in the estimated trends. Attempts are made to minimize bias by designing the data collection appropriately or by collecting additional variables than can be used to adjust for these factors. With such ancillary information, it is common to employ a general linear model (GLM) to model the dependent variable, which is assumed to be related to abundance (e.g., catch-per-unit-of-effort, CPUE) and to include the ancillary data as explanatory variables to remove variation that is not related to abundance (Maunder and Punt, 2004). In GLMs, time is usually modeled as a categorical variable, and is used to represent the temporal changes in relative abundance.

One of the principal factors influencing observations of animal abundance is habitat. Habitat is generally defined as an abiotic factor such as temperature or a physical structure that may be abiotic (e.g., bottom type) or biotic (presence of sea grass), but could be extended to any relevant factor such as presence of other species. Habitat influences observations, either because the density of individuals varies among habitat types, or because the ability to observe the individuals varies among habitat types. Therefore, analyses of information on abundance should be adjusted for habitat type. For example, habitat type could be included as one of the explanatory variables in a GLM, but this requires that habitat type be recorded for each observation.

GLMs are just one group of a multitude of methods that have been used to derive indices of abundance. They have some desirable properties, in that they can include habitat data and in that they are carried out in a statistical framework. The statistical framework allows estimation of parameters and a description of uncertainty. However, GLMs exhibit weakness in their limited ability to include scientific under-

Bulletin of Marine Science

standing about the system, particularly nonlinearities, and in some situations, they are unable to adequately model observations that sum information across multiple habitats.

Hinton and Nakano (1996) derived a general framework, commonly called habitat-based standardization (HBS) that overcomes the weaknesses of the GLM models described above. The HBS method matches the sample effort data, in this case fishing effort, with distributions of the habitat and the habitat preference of the species. It does not require that the exact location of capture within the habitat be known, only that the total catch resulting from the effort (e.g., set) and a measure of the habitat for each unit of effort (e.g., hook) be known, which provides a significant advantage when designing sampling schemes for cryptic or hard-to-observe captures. The HBS method is a process model that incorporates the sampling process and scientific understanding of the system. Component models can be developed for each of the processes (e.g., habitat, sampling, and animal behavior).

Hinton and Nakano (1996) illustrated their method with a simple deterministic application to Japanese longline catch and effort data for blue marlin (*Makaira nigricans* Lacépède, 1802) in the Pacific Ocean. The sampling effort component model was generated using the depth of hooks between the floats of a longline from a catenary curve function of the number of hooks deployed between the floats (Fig. 1). This was done because the Japanese longline fleet has increased the number of hooks between floats over time to increase the depth of the hooks so as to target bigeye tuna (*Thunnus obesus*, Lowe, 1839), a species that occurs at deeper depths (Nakano and Bayliff, 1992). The depths of the hooks estimated from the catenary curve model were then matched with habitat distribution from a component model based on the temperature difference from the mixed layer and the time at temperature for blue marlin relative to the mixed layer derived from acoustic telemetry data (Holland et al., 1990). The effort (number of hooks) was then converted into effective effort by weighting each hook by the appropriate habitat preference and summing over all hooks on the entire longline.

Much controversy has surrounded the applications of the Hinton and Nakano (1996) HBS method (Goodyear et al., 2003; Ward and Myers, 2005; Prince and Goodyear, 2006). Much of this is related to the details of the specific illustration of HBS using blue marlin data as presented by Hinton and Nakano (1996), which they stated was intended "To illustrate the method ..." (p. 176). However, the HBS method is a general framework, and many of the criticisms are unfounded or have been addressed in later works (e.g., Bigelow et al., 2002). For example, a criticism that the method does not consider when individuals are feeding is not a deficiency of the method, but a deficiency in the component process model and data used to determine the habitat, and if "differences are found, then ... [HBS] may be structured to account for [them]" (Hinton and Nakano, 1996, p. 178). In limited simulation studies, Hinton (1996) illustrated that the method could perform well in the presence of environmental variability.

Several tests have been used to determine the appropriateness of the HBS-derived indices of abundance. Comparison of total likelihoods from stock assessment models, including indices of abundance based on nominal effort and those based on HBS (e.g., Hinton, 2001; Hinton and Bayliff, 2002), have been used to determine if the HBS effort was more consistent with the assumed population dynamics and other data (e.g., total catch) compared to nominal effort (Hinton and Maunder, 2004a).

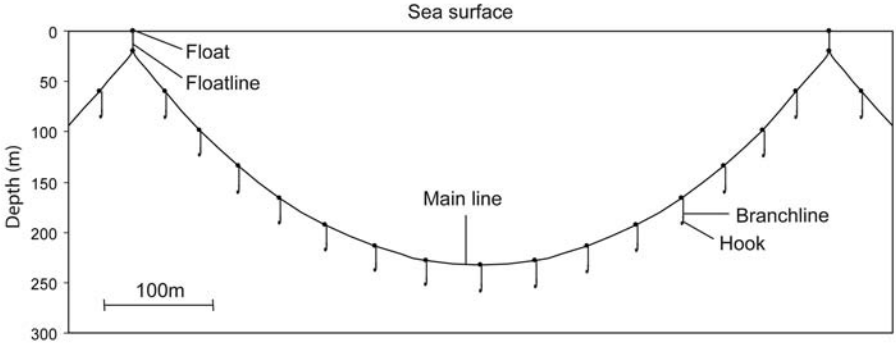

Figure 1. Vertical distribution of hooks based on catenary geometry from a typical longline set in the Japanese longline fishery that deployed 15 hooks between floats.

In another indirect test, population dynamics model (MULTIFAN-CL) results with nominal effort were unreasonable, but results obtained from effort obtained from HBS allowed model fits with reasonable estimates of ancillary parameters, such a growth and mortality (Kleiber et al., 2003). The controversies continued over the appropriateness of the results obtained from applications of the HBS method, which inspired further testing of estimation and modeling using the method.

A more traditional testing approach using several testing criteria (Akaike Information Criterion [AIC], Bayesian Information Criterion [BIC], Bayes factors) was applied by Maunder et al. (2002). This method compared observed and predicted catch using nominal and HBS effort for yellowfin (*Thunnus albacares* Bonnaterre, 1788) and bigeye tuna in the Pacific Ocean. They found that the HBS effort was substantially better than nominal effort. However, these and other results indicated that certain habitat preference models developed from archival tag information may not always be the best or appropriate models to use in a given habitat-based standardization. More discussion about the inappropriateness of the archival tag data is presented later.

The method of testing the HBS method used by Maunder et al. (2002) led to a logical process of estimating the parameters of the HBS method in a statistical framework (statHBS). In statHBS, the estimation of the habitat preference parameters (or parameters for other components of the model) can be allowed to improve the fit of the model to the observed catch data. This allows for more direct model testing and calculation of confidence intervals than would be possible in a non-statistical framework. There have since been a number of applications of statHBS (e.g., Bigelow et al., 2003; Hinton and Maunder, 2004b; Langley et al., 2005). However, the statHBS model has yet to be described in the primary literature. Applications of HBS and statHBS are not limited to situations involving longline catch and effort data. The method is more general and can be applied in many other situations.

First we describe why GLM and related approaches may not be suitable for application when observations sum information across multiple types of habitat. Next we describe the basic concepts of statHBS. Then we apply statHBS to the Japanese longline catch of bigeye tuna and effort data in the western and central Pacific and compare it to two alternative approaches, a “deterministic” application of the HBS method (detHBS), which is similar to the Hinton and Nakano (1996) illustration shown with blue marlin and nominal effort. Finally, we describe potential modifications of the statHBS.

Summing Data Across Multiple Habitats

There are many possible scenarios for which observed data can be summed across multiple habitat types. The longline catch and effort application of Hinton and Nakano (1996) is a common example in fisheries. Other marine examples include plankton tows with continuous or periodic oceanographic (e.g., temperature) recorders, and trawl surveys with continuous or periodic depth recorders. To model this type of data, an equation is needed to predict the observations. The equation must sum the predicted catches for each of the units of effort with reference to the habitat type and compare that sum to the total observed catch. For a simple example, let $C_{i,j}$ represent the predicted catch for component *j* for observation *i*, and $T_{i,j}$ represent the temperature as a continuous variable for component *j* of observation *i*. If a linear relationship is assumed between catch and temperature

$$C_{i,j} = \alpha + \beta T_{i,j}$$

$$C_i = \sum_j C_{i,j} = \sum_j [\alpha + \beta T_{i,j}] = J\alpha + \beta \sum_j T_{i,j}$$

where *J* is the total number of components. A GLM can now be applied, using C_i as the dependent variable $\sum_j T_j$ and as the explanatory variable. However, if the standard log-linear multiplicative model is used, then

$$C_i = \sum_j \alpha T^{\beta}_{i,j} \neq \alpha \left[\sum_j T_{i,j}\right]^{\beta} \text{ or equivalently,}$$

$$\ln[C_i] = \ln\left[\sum_j \exp[\alpha + \beta T_{i,j}]\right] \neq \sum_j \alpha + \beta T_{i,j}.$$

Therefore, a GLM framework can no longer be used, and a nonlinear model estimation procedure is required.

An alternative is to change the habitat variable to a categorical variable

$$C_i = \sum_j \alpha + H_{i,j \in h}$$

$$h = \{T \in (0,5), T \in (5,10), T \in (10,15), T \in (15,20)\}$$

$$C_i = J\alpha + E_1 H_1 + E_2 H_2 + \ldots$$

where $E_h = \sum_j \begin{cases} j \notin h & 0 \\ j \notin h & 1 \end{cases}.$

A GLM can then be applied by summing the amount of effort in each habitat and using each of these habitat groups as explanatory variables. Unfortunately, most desirable modifications of the model make the GLM framework impossible to use. For example, including a multiplicative year effect I_y (the desired quantity for an index of abundance) makes the model a combination of additive and multiplicative models,

$$C_i = (E_1 H_1 + E_2 H_2 + \ldots) I_y.$$

If a nonlinear estimation framework is used, there is no restriction on the model structure, and this allows for the scientific understanding of the system to be included in the model. Each process can then be modeled as a component model that includes the understanding of that process. The component models can be combined into a single model used for parameter estimation in the statHBS framework.

The Statistical Habitat-based Standardization (statHBS) Model

We describe the statHBS model using a fisheries longline CPUE example with the main goal to estimate the year effect, but CPUE can be replaced with other quantities and year can be replaced with another measure of time or some other dimension. Let i index the longline set and j index the hook in that set.

Basic Model.—The CPUE is assumed to be proportional to abundance, N_t, at time t

$$CPUE = \frac{C}{E} = qN_t,$$

where C is catch, E is the effort, and q is the constant of proportionality (catchability in this example). This equation can be rearranged to predict the catch.

$$C = qEN_t$$

A basic premise of this application of statHBS is that catchability is proportional to habitat type,

$$q \propto H$$

For this illustration, let the habitat be discrete categories. Let q_{base} be the overall catchability and define H_h as the difference from the overall catchability for habitat h.

$$q_{i,j} = q_{base} H_{i,j}$$

where $H_{i,j}$ represents the habitat effect for hook j in set i, and $H_{i,j} = H_h$ if the habitat type for observation i,j is h.

The main objective is to estimate the year effect, I_t, which represents the relative abundance and is proportional to N_t. Therefore, the predicted catch is

$$C_{i,j} = q_{base} H_{i,j} I_t E_{i,j}$$

where t represents the time for observation i,j and $E_{i,j}$ is the effort for observation i,j. $E_{i,j} = 1$ in this example.

To accumulate the catch for the whole longline, the predicted catch is summed over the habitat types,

$$C_i = q_{base} I_t \sum_h \left(H_h \sum_{i,j \in h} E_{i,j} \right)$$

The parameters of the model are q_{base} and the vectors H and I. However, q_{base} is confounded with I and H, so the initial I is set to one and the H's are constrained to sum to 1.

The parameters of the model are estimated by fitting to the observed catch. This is accomplished by minimizing the negative log-likelihood. For this example, we use the lognormal likelihood function

$$-\ln L\left(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{C}}\right) = \sum_i \ln[\sigma] + \left[\frac{\left(\ln\left[\widetilde{C}_i + \delta\right] - \ln[C_i + \delta]\right)^2}{2\sigma^2}\right]$$

where $\widetilde{C}_i$ is the observed catch, and δ is a small constant (e.g., 1.0) added to avoid computational problems when the observed or predicted catch is zero. The standard deviation, σ, of the likelihood function is an estimated parameter in the model.

Prior on Habitat.—The original illustration of Hinton and Nakano (1996) was deterministic, in that all the HBS model parameters were estimated externally using auxiliary data. In the HBS framework, it was assumed that all the components of the model, including the habitat preference, were known. This approach can be approximated by fixing the H's in the above description at set values. This model will be referred to as detHBS. Hinton and Nakano (1996) recognized that there is uncertainty in the habitat preference estimates, and they sampled the habitat preference from a multinomial distribution to obtain confidence bounds for abundance trends. This simulation method logically leads to the use of priors in a Bayesian context for the habitat preference or other model parameters.

For example, the prior for the preference of each habitat could be assumed log-normal with a given mean and standard deviation:

$$-\ln P(H_h) = \sum_h \frac{(\ln[H_h] - \ln[\mu_h])^2}{2\sigma_h^2}.$$

The negative of the logarithm of the prior is added to the negative log-likelihood.

GLM Component.—There are many factors other than the habitat that may affect catchability or relative abundance of the individuals available to the gear. GLMs allow for the inclusion of multiple variables, either categorical or continuous. The GLM framework can be integrated with the statHBS approach. In this case, the GLM explanatory variables are related to the longline set.

The standard log-transformed GLM model used in fisheries is

$$\ln[\boldsymbol{CPUE}_i] = \boldsymbol{\beta X}_i + \boldsymbol{\varepsilon}_i$$

Using this parallel, the expected logarithm of catchability is a linear combination of explanatory variables

$$q_i = \exp(\boldsymbol{\beta X}_i + \boldsymbol{\varepsilon}_i)$$

where $\boldsymbol{\beta}$ is a vector of parameters, $\boldsymbol{X}_i$ is a matrix of data, and $\boldsymbol{\varepsilon}_i$ is the error term for observation i [usually iid N(0, σ^2)]. Consider a model with an intercept, a continuous variable, and a categorical variable with three levels. The model has four parameters, the intercept (α), a coefficient for the continuous variable (β), and two parameters for the categorical variable (p_2 and p_3; one of the levels is confounded with the intercept). The parameters of the categorical variable are offsets from the base category that is represented by the intercept. For example, the first row in the matrix below represents the base category, the second row the second category, and the third row the third category. α represents q_{base} in the previous description.

$$\boldsymbol{\beta}^T = \{\alpha \ \beta \ p_1 \ p_2\}$$

$$\boldsymbol{X} = \begin{Bmatrix} 1 & 0.3 & 0 & 0 \\ 1 & -1.7 & 1 & 0 \\ 1 & 1.2 & 0 & 1 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{Bmatrix}$$

Traditionally, the year effect (I) is included as a categorical variable in a GLM analysis, and can be included in the GLM component of the statHBS model, rather than using the explicit method described above.

When the intercept is replaced with q_{base} and I is used as a separate categorical variable, the statHBS model is:

$$\ln[C_i] = \ln\left[q_{base} I_t \sum_h \left(H_h \sum_{i,j \in h} E_{i,j}\right)\right] + \boldsymbol{\beta X}_i$$

NONLINEAR ESTIMATION.—To complete the estimation process, the habitat must be determined for each component of effort (e.g., for each hook). Habitat data (e.g., ambient temperature) can be obtained or estimated from an external source (e.g., an Ocean General Circulation Model, OGCM), which in some cases may require interpolation to ensure values are available for each piece of effort. In general terms, this requires computer code that determines the habitat for each hook and sums over the number of hooks in each habitat for each longline set. The parameters can be estimated by minimizing the negative log-likelihood adjusted for the prior of the habitat preference (i.e., the negative log of the posterior),

$$-\ln L\left(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{C}}\right) - \ln P(H_h)$$

then finding the posterior mode. This can also be viewed as a penalized likelihood, or the habitat preference prior could be viewed as a prior likelihood (Pawitan, 2003). Alternatively, the prior could be replaced with an approximation to the likelihood from the analysis that estimated the habitat preference prior, or the data could be integrated directly into the statHBS analysis. Or, the prior could be eliminated from the analysis.

MODEL TESTING.—Standard model selection methods (e.g., likelihood ratio test, AIC) can be used to select different components of the statHBS model. For example, model selection can be used to determine which habitat variables to include or which explanatory variables to include in the GLM component. Most components of the model are nested and therefore likelihood ratio tests can be used.

APPLICATION

We apply the statHBS model to the main area of longline catch for bigeye tuna in the western and central component (15°N–10°S, 140°E–150°W) of the Pacific Ocean population (Fig. 2). Oceanographically, this spatial area represents three tropical bio-ecological provinces (western Pacific warm pool, North Pacific Equatorial Countercurrent and Pacific Equatorial Divergence) based on primary productivity regimes (Longhurst et al., 1995). We fit the statHBS model to bigeye catch from the Japanese distant-water longline fishery grouped by time (month), area (5° latitude × 5° longitude), and gear configuration [hooks between floats (HBF); range 5–23]. The area contained 52,741 strata (1% zero catch observations) from 1975 to 2004. The vertical distribution of hooks within each configuration was based on longline characteristics and catenary geometry (Bigelow et al., 2002) and considered deterministic in the statHBS model. Habitat data (ambient temperature) were obtained from 0–600 m at 40 m intervals from an OGCM (Simple Ocean Data Assimilation (SODA), Carton et al., 2000a,b). The original OGCM spatio-temporal resolution was monthly, 1° latitude × 1.5° longitude, and temperature data were spatially aggregated to the scale of the fishery data.

We compare statHBS indices of relative abundance to those estimated without standardization and using detHBS. The habitat parameters are water temperature grouped into 2 °C bins. No GLM variables were included in the analysis. The AIC was used to determine which model best fit the data. The application shown here is used only to illustrate the statHBS model and a different spatial stratification may be more appropriate for assessment and management of this stock.

The results show that the statHBS model (AIC = 18,408) fit the data better than the nominal (AIC = 19,618) and detHBS (AIC = 31,130) models. The detHBS model performed worse than just using nominal effort. The detHBS and statHBS models

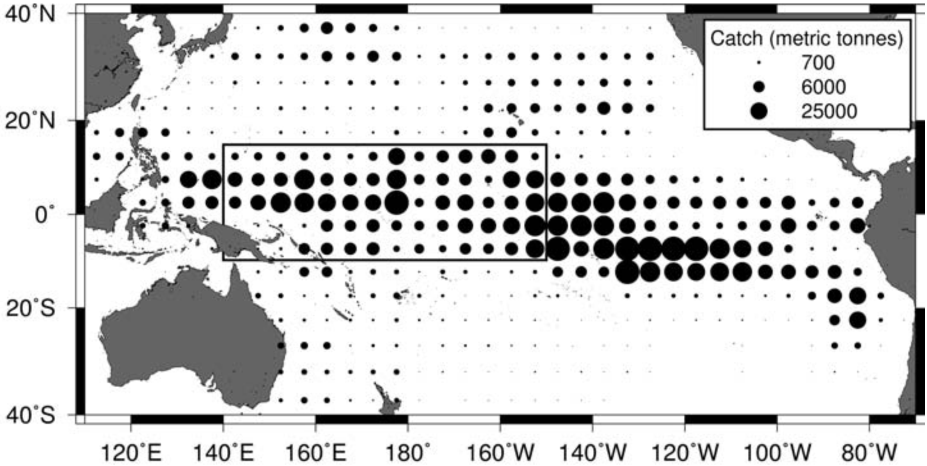

Figure 2. Distribution of cumulative bigeye tuna longline catch from 1990–2004 by 5° latitude and 5° longitude. Maximum circle size represents a catch of 50,000 tons. The box indicates the spatial stratification for applications of statistical models to bigeye tuna in the western and central component of the Pacific Ocean population.

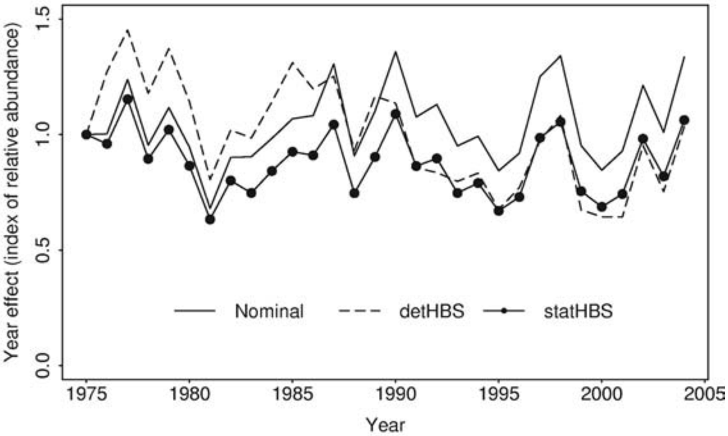

Figure 3. Indices of relative abundance for a bigeye tuna in the western and central Pacific Ocean example as estimated from nominal effort, detHBS and statHBS.

produced a greater decline in relative abundance than just using nominal effort (Fig. 3) because of the time-series change to deeper gear. Shallow gear (5–9 hooks between floats) represented 39.3% of the gear during the initial 10 yrs (1975–1984) in the study area but only 1.1% during the last 10 yrs (1995–2004). Conversely, deep gear (15–23 hooks between floats) was rarely deployed (0.6%) in the initial 10 yrs but was paramount (77.6%) during the last 10 yrs. The statHBS model estimated a different habitat preference than the one used in the detHBS model (Fig. 4), with a greater preference for warmer water. By adding a constant of 1.0 to each catch, residuals were normally distributed (Fig. 5) with a slightly longer negative tail. The addition of a constant of 1.0 was preferred over a smaller constant (0.001) as a smaller constant produced a bimodal distribution in residuals, because taking the logarithm of zero catches introduces a greater outlier. A constant of 10 produced normally distributed residuals without a negative skew; however, the estimated year effects may be biased as they were 5% less then with a constant of 1.0.

Modifications

The following modifications to the statHBS model may improve the method.

Total Habitat.—The model as specified above assumes that the total amount of habitat is constant over time. However, it is well known that the environment can have large temporal variation, and as the total amount of habitat increases or decreases, the population may expand or contract and fish densities will change. Therefore, the year effect will be confounded with the total habitat. To account for changes in the total habitat, the total relative habitat can be calculated for each time period by summing up the relative habitat preference for each area strata and the catchability multiplied by its inverse.

$$\ln[C_i] = \ln\left[\frac{q_{base} I_t \sum_h \left(H_h \sum_{i,j \in h} E_{i,j}\right)}{\sum_A H_A}\right] + \beta X_i$$

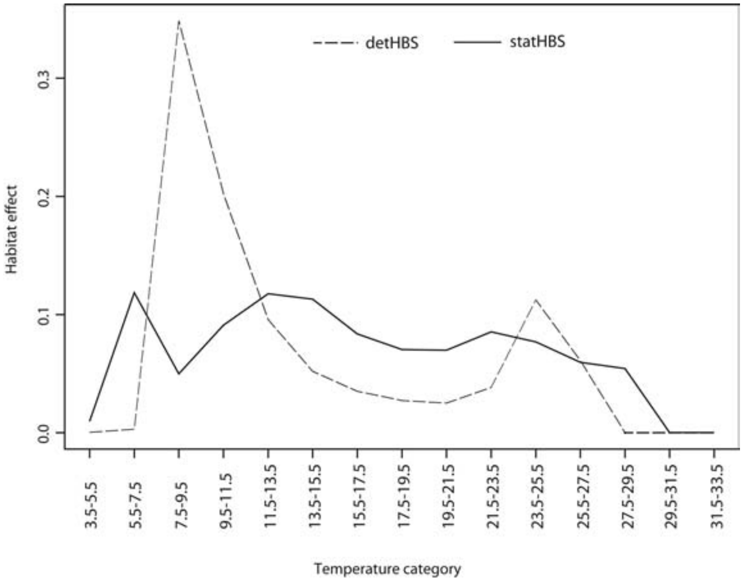

Figure 4. Relative habitat effect (habitat preference) used in the detHBS analysis and estimated by the statHBS analysis for bigeye tuna in the western and central Pacific Ocean.

Matching Effort with Habitat.—The statHBS requires that effort be matched with the appropriate habitat. If the habitat is recorded with the sampling then this is automatically accomplished. In the longline catch and effort applications, a component model describing hook depth is used in association with environmental data to determine the habitat associated with each hook. However, the simple model based on the catenary curve used to represent hook depth may be inappropriate, as many factors influence the depth of hooks, and differences in the characteristics of the gear and deployment methods may influence the depth of the hooks. In this and similar cases, it might be possible to parameterize the component model that associates effort with habitat and estimate or update the parameters in the statHBS model. For example, currents can reduce the depth of a hook by shoaling the longline. The amount of shoaling can be modeled based on explanatory variables (e.g., line characteristics and ocean currents).

Parameterizing the Habitat Preference.—The description of the statHBS model described above uses habitat as a categorical variable. It is possible to consider the habitat as a continuous variable, or one of multiple dimensions, and to use either a simple GLM approach or to use scientific understanding to develop and parameterize a functional form for the habitat preference. The parameters of this component model could be estimated within the statHBS framework.

Alternative Likelihood Functions.—The likelihood function we presented is a simple likelihood; however, other GLM error models and link functions could be used. The delta-lognormal methods commonly used in fisheries to deal with zero observations could also be applied with two separate statHBS models with one model used for the proportion of zero observations and a second for positive observations. If the number of individuals caught is small, then likelihoods for count data, such as the Poisson or negative binomial, may be appropriate.

Efficient Estimation.—It may be possible to use the efficient estimation equation of GLM theory for the GLM component of the statHBS model to greatly reduce

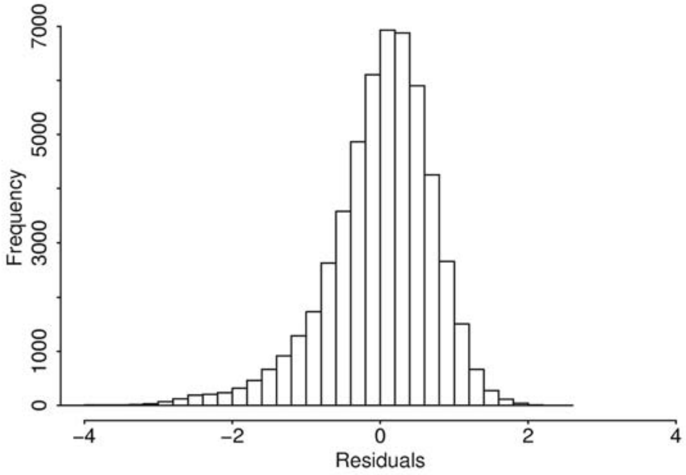

Figure 5. Residuals (log(observed catch+1)−log(estimated catch+1)) for the statHBS model applied to bigeye tuna in the western and central Pacific Ocean.

computation times. The catchability (*q*) and year effects could also be included in the GLM framework to increase efficiency. Each function evaluation in the numerical iteration estimation procedure, which is defined for a set of values of the habitat preference parameters, would use the MLEs of the parameters of the GLM component calculated using the GLM theory. This is similar to using the analytical solution for the catchability coefficient when fitting a population dynamics model to a CPUE index of relative abundance.

Setting and Retrieval.—It has been suggested by a limited number of studies that some species may be more likely to be caught on longline gear when the baited hook is moving through the water (Boggs, 1992). This occurs when the longline is being set or retrieved. For example, as a hook is retrieved it moves out of a deeper habitat and through shallower habitats. This process can be included by modeling the relative time the hook spent in each of the habitats (i.e., the depth of the habitat) as the gear was retrieved. This quantity is multiplied by a scalar to estimate the relative catchability that occurred during setting and retrieval.

Discussion

We have shown how applying a statistical framework to the habitat-based standardization method of Hinton and Nakano (1996) for determining relative indices of abundance (1) takes habitat information into consideration, (2) allows the summation of multiple units of effort associated with habitat type to predict the observation, (3) uses the scientific understanding of the system in the formulation of the model, and (4) estimates the model parameters and calculates uncertainty in a statistical framework that can also be used for model selection. The framework is very flexible and can be applied in a variety of ways. Component models can be developed for each process and the relevant parameters of these models estimated within the statHBS framework. In addition, we have shown how the model may be integrated with a GLM framework to allow convenient comparison with traditional methods used to standardize CPUE data.

The HBS was initially applied in a deterministic framework (e.g., Hinton and Deriso, 1998; Bigelow et al., 2002; Hinton and Bayliff, 2002; Kleiber et al., 2003). Although statistical tests indicated that the method provided improvement over nominal effort in some cases, the statistical approach provides advantages for modeling and testing, and it is becoming the predominant method of application. The statHBS model has been applied in assessment work on bigeye and yellowfin tuna (Bigelow et al., 2003, 2005; Langley et al., 2005) and on striped marlin (*Tetrapturus audux* Philippi, 1887) (Hinton and Maunder, 2004b) in the Pacific Ocean.

The statHBS model has been investigated and improved by several researchers, though much of the work to date remains unpublished. Hinton and Maunder (2004b) included retrieval in their application to striped marlin with limited success. At the conclusion of the model runs they decided to not include this variable in the model selected due to questions about parameter formulation and impacts on model fit. Langley et al. (2005) illustrated that in some instances when spatial parameters are included in the GLM component of statHBS, the estimated habitat preference effect becomes constant, indicating that within the model structure, habitat and gear interactions are correlated with spatial structure. Such confounding has also been demonstrated by application of GLM models to swordfish (*Xiphias gladius* Linnaeus, 1758) in the eastern Pacific Ocean. Analyses with large spatial structure had significant factors for gear (Hinton, 2003). However, in analyses with highly detailed spatial structure, which provided closer conformity to the physical oceanography of the region, gear was not a significant factor (Y. Uozumi and K. Yokawa, National Research Institute for Far Seas Fisheries, Japan, pers. comm.). These results emphasize the need to consider the temporal and spatial scale of the data used in the model. The statHBS has been used to show that studies in which longline CPUE was standardized solely on depth (Ward and Myers, 2005) are misleading and encounter problems similar to those encountered with the detHBS model.

In their statHBS application, Langley et al. (2005) used the canonical implementation of the year effect (Francis, 1999) to develop confidence intervals independent of the base year. Langley et al. (2005) also integrated multiple spatial areas to develop regional indices of abundance that shared a common overall catchability coefficient, to be consistent with the MULTIFAN-CL assessment model (Hampton et al., 2005). Bigelow et al. (2003) used statHBS to determine if ambient temperature values or temperature relative to the mixed layer is the more important habitat variable to explain catch rates in bigeye and yellowfin tuna.

The results obtained in the bigeye tuna example illustrate a problem caused by the fact that longline sets are deployed early in the morning and completed at night. Some species (e.g., bigeye tuna and swordfish) are known to occupy different habitats in the day and night. The statHBS approach integrates over the entire longline set, and habitat preferences are sometimes bimodal due to incorporating aspects of both day and night vulnerability. A better approach would be to structure the habitat by day, by night, or by a combination of preferences for each period.

Previous studies have shown that in some cases the detHBS model is less able to predict catch than nominal effort, which is also illustrated in our bigeye tuna example in the western and central Pacific Ocean, and that the habitat preferences can change from area to area. The statHBS always improves the ability to predict catch, which has been statistically significant in all of our applications, and the results can be quite different from the detHBS model. There are several problems to consider when using archival tag data to determine habitat preference and vulnerability for use in habitat-based standardization. The most commonly cited problem is that the archival tagging data furnish information about where the individual fish is located, rather than where it is feeding or takes baited hooks, so the habitat and corresponding vertical distribution estimates may reflect different vulnerabilities to capture. However, Bach et al. (2003) found that the distribution of bigeye tuna caught in fishing experiments was similar to that estimated by electronic tags. Another major reason for some of the significant differences in performance between the detHBS and statHBS results is the failure to correctly map the spatial and temporal scales of the habitat distribution data and the habitat preference. In the examples, these were simply the oceanographic data, which are usually from OGCMs and typically represent averages over a scale of month and 1° latitude × 1.5° longitude, whereas the archival tag data reflect oceanographic conditions exactly where the fish is at any instance in time. If the fish chooses where it prefers to be, for example, following an oceanic front, then the archival tag data may not correspond to the average conditions in an OGCM stratum. In truth, the distributions of habitat and habitat preferences are complex (Hinton and Nakano, 1996), and with model misspecification inevitable in such a simplified framework, the estimation process inherent in the statistical approach will, in almost every case, result in a significantly better fit to the data than will a deterministic model. Other factors, such as undocumented technology improvements (e.g., monofilament lines and new hooks), that are also problems in other CPUE standardization methods, increase CPUE over time and it may not be possible to separate these factors from the year effects. Research is needed to determine the robustness of the statHBS approach to all the other factors mentioned above, and to improve the model.

Acknowledgments

The authors gratefully acknowledge the National Research Institute of Far Seas Fisheries for access to commercial fisheries data. W. Bayliff, H. Shono, and J. Wetherall provided comments on the manuscript. Two anonymous referees provided useful advice, particularly the suggestion to use the efficient estimation equation of GLM theory for the GLM component of the statHBS model. The views expressed herein are those of the authors and do not necessarily reflect the views of their respective organizations.

Literature Cited

Batch, P., L. Dagorn, A. Bertrand, E. Josse, and C. Misselis. 2003. Acoustic telemetry versus monitored longline fishing for studying the vertical distribution of pelagic fish: bigeye tuna (*Thunnus obesus*) in French Polynesia. Fish. Res. 60: 281–292.

Bigelow, K. A., J. Hampton, and N. Miyabe. 2002. Application of the habitat-based model to estimate effective longline effort and relative abundance of Pacific bigeye tuna (*Thunnus obesus*). Fish. Ocean. 11: 143–155.

____________, M. N. Maunder, and M. G. Hinton. 2003. Comparison of deterministic and statistical habitat-based models to estimate effective longline effort and standardized CPUE for bigeye and yellowfin tuna. Standing Committee on Tuna and Billfish 16 RG–3. 19 p.

____________, A. Langley, T. Patterson, and N. Miyabe. 2005. Relative abundance indices of the Japanese longline fishery for bigeye tuna: analysis of data from the western and central Pacific Ocean. Pages 110–129 *in* Migration and habitat preferences of bigeye tuna, *Thunnus obesus*, on the east coast of Australia — a project using archival and conventional tags to determine key uncertainties in the species stock structure, movement dynamics and CPUE trends. CSIRO Marine Research, FRDC Project no. 1999. Report 109.

Boggs, C. H. 1992. Depth, capture time, and hooked longevity of longline-caught pelagic fish: timing bites of fish with chips. Fish. Bull. 90: 642–658.

Carton, H., G. Chepurin, and X. Cao. 2000a. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: Methodology. J. Phys. Ocean. 30: 294–309.

__________, ________, and ______. 2000b. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part II: Results. J. Phys. Ocean. 30: 311–326.

Francis, R. I. C. C. 1999. The impact of correlations in standardized CPUE indices. NZ Fish. Ass. Res. Doc. 99/42.

Goodyear, C. P., D. Die, D. W. Kerstetter, D. B. Olson, E. Prince, and G. P. Scott. 2003. Habitat standardization of CPUE indices: research needs. Pages 613–623 *in* ICCAT collective volume of scientific papers 55.

Hampton, J., P. Kleiber, A. Langley, Y. Takeuchi, and M. Ichinokawa. 2005. Stock assessment of bigeye tuna in the western and central Pacific Ocean. Western and Central Pacific Fisheries Commission Science Committee 1 Stock Assessment Working Paper 2. 77 p.

Hinton, M. G. 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints, and environmental data, with applications to blue marlin (*Makaira nigricans*) and swordfish (*Xiphias gladius*) of the Pacific Ocean. Ph.D. Diss., Scripps Institution of Oceanography, University of California, San Diego. 139 p.

____________. 2001. Status of blue marlin in the Pacific Ocean. Inter. Amer. Trop. Tuna Comm., Stock Assess. Rep. 1: 284–316.

____________. 2003. Status of swordfish stocks in the eastern Pacific Ocean estimated using data from Japanese tuna longline fisheries. Mar. Freshw. Res. 54: 393–399.

____________ and W. H. Bayliff. 2002. Status of striped marlin in the eastern Pacific Ocean in 2001 and outlook for 2002. Inter. Amer. Trop. Tuna Comm., Stock Assess. Rep. 3: 328–364.

____________ and R. B. Deriso. 1998. Distribution and stock assessment of swordfish, *Xiphias gladius*, in the eastern Pacific Ocean from catch and effort data standardized on biological and environmental parameters. NOAA Tech. Rep. NMFS 142: 161–179.

____________ and M. N. Maunder. 2004a. Methods for standardizing CPUE and how to select among them. Pages 169–177 *in* ICCAT collective volume of scientific papers 56.

____________ and _____________. 2004b. Status of striped marlin in the eastern Pacific Ocean in 2002 and outlook for 2003–2004. Inter. Amer. Trop. Tuna Comm. Stock Assess. Rep. 4: 287–310.

____________ and H. Nakano. 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from the Japanese longline fisheries in the Pacific. Inter-Am. Trop. Tuna Comm. Bull. 21: 171–200.

Holland, K. N., R. W. Brill, and R. K. C. Chang. 1990. Horizontal and vertical movements of Pacific blue marlin captured and released using sportfishing gear. Fish. Bull. 88: 397–402.

Kleiber, P., M. G. Hinton, and Y. Uozumi. 2003. Stock assessment of blue marlin (*Makaira nigricans*) in the Pacific with Multifan-CL. Mar. Freshw. Res. 54: 349–360.

Langley, A., K. Bigelow, M. N. Maunder, and N. Miyabe. 2005. Longline CPUE indices for bigeye and yellowfin in the Pacific Ocean using GLM and statistical habitat standardisation methods. Western and Central Pacific Fisheries Commission Science Committee 1 Stock Assessment Working Paper 8. 40 p.

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill. 1995. An estimate of global primary production in the ocean from satellite radiometer data. J. Plankton Res. 17: 1245–1271.

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

_____________, M. G. Hinton, K. A. Bigelow, and S. J. Harley. 2002. Statistical comparisons of habitat standardized effort and nominal effort. SCTB 15 MWG-7 18 p.

Nakano, H. and W. B. Bayliff. 1992. A review of the Japanese longline fishery for tunas and billfishes in the eastern Pacific Ocean, 1981–1987. Inter-Am. Trop. Tuna Comm. Bull. 20: 187–237.

Pawitan, Y. 2003. In all likelihood: statistical modeling and inference using likelihood. Oxford University Press, Oxford, UK. 528 p.

Prince, E. D. and C. P. Goodyear. 2006. Hypoxia-based habitat compression of tropical pelagic fishes. Fish. Ocean. 15: 451–464.

Ward, P. and R. A. Myers. 2005. Inferring the depth distribution of catchability for pelagic fishes and correcting for variations in the depth of longline fishing gear. Can. J. Fish. Aquat. Sci. 62: 1130–1142.

ADDRESSES: (M.N.M., M.G.H.) *Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, California 92037-1508*. (K.A.B.) *Pacific Islands Fisheries Science Center, NOAA Fisheries, 2570 Dole Street, Honolulu, Hawaii 96822*. (A.D.L.) *Secretariat of the Pacific Community, BP D5, 98848 Noumea Cedex, New Caledonia*. CORRESPONDING AUTHOR: (M.N.M.) *E-mail: <mmaunder@iattc.org>*.