Abstract—A generalized Bayesian population dynamics model was developed for analysis of historical mark-recapture studies. The Bayesian approach builds upon existing maximum likelihood methods and is useful when substantial uncertainties exist in the data or little information is available about auxiliary parameters such as tag loss and reporting rates. Movement rates are obtained through Markov-chain Monte-Carlo (MCMC) simulation, which are suitable for use as input in subsequent stock assessment analysis. The mark-recapture model was applied to English sole (*Parophrys vetulus*) off the west coast of the United States and Canada and migration rates were estimated to be 2% per month to the north and 4% per month to the south. These posterior parameter distributions and the Bayesian framework for comparing hypotheses can guide fishery scientists in structuring the spatial and temporal complexity of future analyses of this kind. This approach could be easily generalized for application to other species and more data-rich fishery analyses.

Defining plausible migration rates based on historical tagging data: a Bayesian mark-recapture model applied to English sole (*Parophrys vetulus*)

Ian J. Stewart

Email address: Ian.Stewart@noaa.gov

National Oceanic & Atmospheric Administration
National Marine Fisheries Service
Northwest Fisheries Science Center
2725 Montlake Boulevard East (F/NWC4)
Seattle, Washington 98112-2097

Mark-recapture data are used to estimate growth parameters, mortality rates, and population size (Hilborn and Walters, 1992; Quinn and Deriso, 1999). However, researchers are often interested in home range, site-fidelity, or migration rates which are important quantities for management and the design or evaluation of marine protected areas. The standard approach with mark-recapture data is to use an integrated model linking the underlying dynamics of the tagged population with an observation model describing the predicted recoveries and a likelihood function relating observations with model predictions (Hilborn, 1990). This integrated method has been applied to many fisheries, ranging from those for sablefish (*Anoplopoma fimbria*; Heifetz and Fujioka, 1991) to those for yellowfin tuna (*Thunnus albacares*; Hampton and Fournier, 2001).

Requirements of the integrated method include extensive tag recovery, data on fishing effort, auxiliary information on tag loss, as well as reporting rates in order to adequately estimate movement rates and other parameters (Punt et al., 2000). Analysis often benefits from fixing some model parameters at reasonable values based on external analysis or expert opinion. However, the values selected for these parameters can represent a substantial source of uncertainty in the estimates of movement rates because these parameters are poorly known for many historical tagging projects.

Bayesian methods start with prior distributions for the parameters of interest (information available before the analysis), and integrate over the joint posterior distribution of all model parameters, capturing parameter uncertainty as well as the correlation structure among these parameters. The Bayesian approach provides a logical alternative to likelihood methods when the researcher is faced with substantial uncertainty in the data and input parameters and has a desire for a probabilistic interpretation of the results (Punt and Hilborn, 1997).

One way in which uncertainty and auxiliary information about migration rates could be included in stock assessments is through the use of informative priors based on Bayesian analysis of mark-recapture data. Priors specifically applicable to west coast groundfish stock assessments have been derived for survey catchability (Millar and Methot, 2002), the steepness of the stock-recruit function (Dorn, 2002), the relationship between catch per unit of effort (CPUE) and abundance (Harley et al., 2001), and for other studies currently underway. Researchers in other regions have aggregated historical tagging information for commercially important species, such as north Atlantic cod (*Gadus morhua*; Robichaud and Rose, 2004). However, in the northeast Pacific there are many groundfish tagging studies that have never been analyzed simultaneously or used in stock assessments.

Stock assessments of west coast flatfish (and other groundfish species) have been based on one of two assumptions about latitudinal movement: no movement at all (multiple isolated stocks) or complete mixing (single stock models). In a Bayesian context, these two opposite assumptions fix the magnitude of movement rates before analysis and therefore can be considered highly informative priors. The goal of this study was to provide a generalized method with which to develop informative priors on movement rates based on quantitative analysis of historical tagging data, thereby adding a third choice of prior for use in stock assessments.

Materials and methods

Model development

A model very similar to those used in other mark-recapture studies (e.g., Hilborn, 1990; Hampton and Fournier, 2001) was developed to predict the number of tags returned from multiple tag-release events (tags released in one spatial area over a short period of time; hereafter referred to as a data set) by projecting each population forward in time. Predicted returns are tracked by month and for each spatial area. The predicted number of tags present ($\hat{N}$) in each data set (d), month (t), and area (i) are given by

$$\hat{N}_{d,t,i} = \sum_j \left(\hat{N}_{d,t-1,j} P_{t,j \to i} \right) e^{-\frac{(F_d + M + \Omega)}{12}}, \qquad (1)$$

where F_d = the average fishing mortality rate for each data set;
M = the average natural mortality rate; and
Ω = the instantaneous rate of tag loss (ongoing "attrition" due to fouling or mechanical failure; often referred to as "type-2 tag loss" in traditional terms; Beverton and Holt, 1957).

The subscript j denotes all possible source areas, and the proportion of individuals moving from each area to another in any month is given by the matrix P_t. The P_t matrix (area × area for month t) includes nonzero values in only the first off-diagonals, and a variable number of parameters within each diagonal depending on the movement hypothesis to be explored. Instantaneous rates are divided by 12 because they are applied on a monthly basis. Predicted numbers in the first month are the reported tag releases (alternately, type-1 tag loss, those tags that are shed immediately after tagging, could be included by multiplying the initial releases by 1 – type-1 tag-loss rate). The predicted recoveries ($\hat{R}$) by data set, month, and area are then

$$\hat{R}_{d,t,i} = \phi \hat{N}_{d,t,i} \left(\frac{F_d}{F_d + M + \Omega} \right) \left(1 - e^{-\frac{(F_d + M + \Omega)}{12}} \right), \qquad (2)$$

where ϕ = the reporting rate of captured tags during the time period over which tag recoveries occurred. Given these dynamics, the tagged population and predicted recoveries for all data sets available may be projected forward simultaneously.

The major departure from previous models is the use of a single average fishing mortality rate for each data set (an approach that reflects a lack of direct effort or of fishing mortality information). If only a single data set is analyzed in this manner, it is clear that any heterogeneity in fishing mortality over time or space could result in substantially biased estimates of movement rates. However, if multiple tagging events are analyzed simultaneously, and there is no consistent relationship between location of tag releases and areas of increased fishing mortality, this potential source of bias may be reduced. Were information on the spatial and temporal variability in fishing mortality available, it would be simple (and recommended) to extend the notation further to either input mortality rates directly into the analysis or to estimate them from relative effort.

Variability in observed recoveries is caused by many factors, including schooling behavior, heterogeneous distribution of fishing effort, tag loss or tag reporting over time and space, and by the stochastic nature of very low recovery rates. Because of the many potential sources of extra-model error, a likelihood function that allows for substantial variation among observations is desired. The negative binomial likelihood is the logical choice for this type of tagging data (e.g., Cormack and Skalski, 1992; Hampton and Fournier, 2001). If each tag-release group is assumed to be independent, the full likelihood (L) of the observed recoveries (R) is given by

$$L\left(R_{d,t,i} \middle| \hat{R}_{d,t,i} k\right) = \qquad (3)$$
$$\prod_d \prod_t \prod_i \frac{(k + R_{d,t,i} - 1)!}{(R_{d,t,i})!(k-1)!} \left(1 + \frac{k}{\hat{R}_{d,t,i}} \right)^{-R_{d,t,i}} \left(\frac{\hat{R}_{d,t,i}}{k} + 1 \right)^{-k},$$

where $\hat{R}$ = the predicted recoveries in a data set, month, and area; and
k = the overdispersion (variance) parameter.

The negative binomial asymptotically approaches the Poisson distribution as the value of the overdispersion parameter moves to infinity (Bishop et al., 1988).

A common problem with historical data is that only summarized reports are available for analysis. Where tag recoveries have been aggregated across time or space, the model predictions and the original observations are no longer on an equivalent scale. This problem is easily dealt with by aggregating the predicted recoveries to match the observed recoveries, while still maintaining the same predictive model structure. However, this method creates different types of comparisons within the likelihood (monthly recoveries compared to

monthly recoveries vs. annual recoveries compared to annual recoveries). The method and notation above is therefore extended to estimate a separate overdispersion parameter (k) for each type of data aggregation included in the analysis. The likelihood equation (Eq. 3) becomes

$$L\left(R_{c,d,l}\big|\hat{R}_{c,d,l},k_c\right)=$$

$$\prod_c\prod_d\prod_l\frac{(k_c+R_{c,d,l}-1)!}{(R_{c,d,l})!(k_c-1)!}\left(1+\frac{k_c}{\hat{R}_{c,d,l}}\right)^{-R_{c,d,l}}\left(\frac{\hat{R}_{c,d,l}}{k_c}+1\right)^{-k_c}, \quad (4)$$

where the likelihood component notation remains the same, but the subscripts are revised to include the following: data aggregation type (c), data set (d), and time-space combination (l, identical within each c; e.g., month-area or year-area depending on the level of data aggregation).

Data sources

English sole (*Parophrys vetulus*) was selected for analysis because of the large amount of tagging data available and the commercial importance of the species. English sole are widely distributed from southern California to Alaska (Hart, 1973) and are frequently captured by the bottom-trawl fisheries of both the United States and Canada. Many English sole tagging programs have been conducted by both U.S. and Canadian scientists since the 1930s. These have included

Figure 1

Map of areas based on historical Pacific Marine Fisheries Commission (PFMC) boundaries used in this analysis. The largest PMFC areas, 1A and 2B, were subdivided to make the latitudinal ranges more consistent across all areas.

releases off the coast of British Columbia (Ketchen, 1956; Forrester, 1969), Washington (Pattie, 1969), Oregon (Harry, 1956), and California (Jow, 1969). Additional tagging within Puget Sound (Menasveta, 1958; Day, 1976) and the Strait of Georgia have focused specifically on population dynamics within these waters. Most tagging data reported between 1946 and 1979 were available only through unpublished reports from the Research Board of Canada, Oregon Department of Fish and Wildlife, and the Washington Department of Fisheries. In total, 44 tagging events (defined as tags released in one area during a one-month period) resulted in the release of 57,839 tags of which 9988 (17.3 %) were recovered. The primary objective of most of these studies was to determine the amount of migration (as a percentage of the total population) that occurred among areas with the highest abundance of English sole (e.g., Harry, 1956).

Tag recoveries have been most frequently reported for the historical Pacific Marine Fisheries Commission (PMFC) areas; these areas were therefore the logical (and only) spatial context in which to analyze the results. To reduce the latitudinal range of the largest PMFC areas, PMFC area 1A was divided into two subareas (north and south of Point Conception) and PMFC area 1B was divided into three subareas (north of Point Arena, between Point Arena and Pigeon Point, and south of Pigeon Point). This division of areas resulted in 17 PMFC-like areas (area 5E is shown for reference only; for simplicity, all data from 5E were treated as if they were collected from 5C or 5D (i.e., 5E data were integrated into similar 5C and 5D areas to make the areas linear) with an average latitudinal span of 138 km (range=83–204 km, Fig. 1). Reporting of tag releases and recoveries was sufficiently detailed to allow an analysis that included these additional boundaries and that did not exclude any studies.

Only studies reporting the area, month, and year for each tagging event were included in this analysis. Additionally, tag recoveries must have been reported at one of four levels of resolution: type-1 resolution, where data were available for year, month, and area for each individual tag recovery; type-2 resolution, where data were available for month and area only; type-3 resolution, where data were available for area only; and type-4 resolution, where data were only available to indicate recovery inside or outside the area of tagging. After this screening (removing 16,375 releases), there were 25 English sole tagging events from the open coast remaining in the analysis, including 17,056 releases and 3464 recoveries; these projects, summarized in Jow (1969), Forrester (1969), and Pattie (1969) ranged from southern California to northern British Columbia and from 1936 to 1965 (Table 1).

There were many differences among individual studies that had to be reconciled or ac-
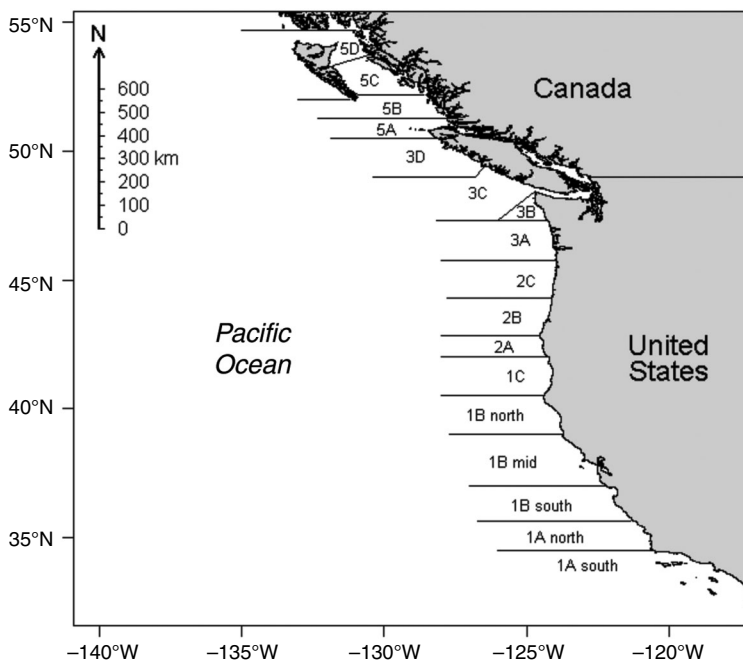
Table 1

Summary of English sole (*Parophrys vetulus*) tagging events off the coast of the United States and Canada modeled in this article. Years listed include the year of initial tagging through the final tag return. "Number released" represent total numbers tagged in a one Pacific Marine Fisheries Commission (PMFC) area (Fig. 1) in a single month. "Number recovered" and "Percent recovered" represent only those tags with complete recovery information; tags recovered within one month of tagging were excluded. Data quality categories were the following: 1, where data were available for year, month, and PFMC area for each individual tag recovery, 2, where data were available for only month and area, 3, where data were available for area only, and 4, where data were only available to indicate recovery inside or outside the area of tagging. NA = not available.

Years	PMFC area	Number released	Number recovered	Percent recovered	Number excluded	Data quality	Reference
1936–37	1B-south	16	1	6.25	0	1	Jow (1969)
1936–37	1B-south	45	9	20.00	NA	3	Jow (1969)
1938–40	1C	400	25	6.33	5	1	Jow (1969)
1939	1B-south	38	1	2.63	0	1	Jow (1969)
1939–41	1B-mid.	30	4	13.33	0	1	Jow (1969)
1940–42	1B-mid.	746	21	2.82	2	1	Jow (1969)
1940–43	1C	852	61	7.21	6	2	Jow (1969)
1940–46	1B-north	1103	27	3.35	0	1	Jow (1969)
1945–47	3C	24	1	4.35	1	1	Forrester (1969)
1945–51	5D/5E	1505	316	21.21	15	4	Forrester (1969)
1945–51	3C	132	29	21.97	NA	3	Forrester (1969)
1949–51	1B-mid.	926	32	3.47	3	1	Jow (1969)
1949–51	1C	415	69	17.16	13	2	Jow (1969)
1950–51	1C	9	1	11.11	0	1	Jow (1969)
1950–52	1B-south	19	1	5.26	0	1	Jow (1969)
1950–53	1B-mid.	200	5	2.50	0	1	Jow (1969)
1950–56	5D/5E	3039	1405	46.55	21	4	Forrester (1969)
1952–58	5D/5E	2235	737	33.21	16	4	Forrester (1969)
1956–61	3B	871	76	8.82	9	2	Pattie (1969)
1958–59	1C	6	3	50.00	0	3	Jow (1969)
1958–61	1B-mid.	97	6	6.19	0	1	Jow (1969)
1958–63	1C	4130	623	15.84	196	2	Jow (1969)
1959	1B-mid.	103	2	1.94	0	1	Jow (1969)
1963–64	1A-north	2	1	50.00	0	1	Jow (1969)
1963–65	1A-nouth	113	8	7.08	0	3	Jow (1969)
	Total	17,056	3464	20.37	287		

counted for to perform a simultaneous analysis. Where possible, recoveries in the same month as the tagging program were excluded (and subtracted from tag releases, n=287) to avoid potentially skewed estimates of movement and fishing mortality rates before any movement could reasonably have occurred. Exclusion of immediate recoveries is common practice when analyzing tag-recovery data (McGarvey and Feenstra, 2002; Robichaud and Rose, 2004). This decision also obviated having to scale the fishing mortality rate during the first month by the number of days after tagging, which is problematic when tagging has occurred over a number of days or is reported only by the month in which it took place. Many different methods of tagging were used in these studies, but all tags were assumed to have the same rate of tag loss. Comparison of tag loss rates for disk and spaghetti tags, although limited, has not identified substantial differences (Meehan and

Milburn[1]), although variation certainly exists given the advances in tagging methods over the four decades spanned by these studies. The reporting rates for tags captured by different fisherman were another source of variability; all recoveries were assumed to have been reported at the same rate in all time periods because there was no information with which to address this issue. Researchers conducting all studies primarily tagged adult fish of both sexes, and recoveries were obtained with commercial or similar fishing gear; however little detail regarding the age or length structure of the fish tagged was available. No accounting was made for

[1] Meehan, J. M., and G. S. Milburn. 1965. Comparison of returns from dart and Peterson disc tags on Dover sole. Fish Commission of Oregon Research Briefs. 13:127. Oregon Department of Fish and Wildlife, 3406 Cherry Avenue NE, Salem, OR 97303.

Table 2

General structure of the monthly movement parameter matrices (P) for all models explored in this analysis; p_n = proportion of population moving north, p_s = proportion of population moving south. Movement rates assumed to be zero in all models are indicated in the off-diagonals. The table is compressed over the central 11 areas (denoted by "…"), where the structure of the matrix did not vary from those cells shown.

	To						
From	1A south	1A north	1B south	…	5B	5C	5D
1A south	$1-p_n$	p_n	0	…	0	0	0
1A north	p_s	$1-p_n-p_s$	p_n	…	0	0	0
1B south	0	p_s	$1-p_n-p_s$	…	0	0	0
…	…	…	…	…	…	…	…
5B	0	0	0	…	$1-p_n-p_s$	p_n	0
5C	0	0	0	…	p_s	$1-p_n-p_s$	p_n
5D	0	0	0	…	0	p_s	$1-p_s$

fish size or age in this analysis. Relative fishing effort at the spatial scale of PMFC areas was not available for these years.

Initial examination of the data (for a qualitative assessment) was performed for tagging projects conducted in Puget Sound and the Strait of Georgia (24,408 tags released and 5756 recovered). Many of these experiments were conducted with varied goals (other than that of estimating movement rates) and an attempt to recover tags from distant areas was not undertaken. Therefore, these data were excluded from the mark-recapture model but were used to structure the spatial extent of the analysis.

Movement hypotheses

Many researchers have noted seasonal changes in catch rates in specific areas, and the temporally transient appearance of aggregations of flatfish. For English sole, these aggregations seem to be associated with the winter spawning season (Alverson, 1960). Spawning of English sole occurs from early fall through late spring, and most growth occurs during the rest of the year. December and April appear to be the first and last months of strong spawning activity across all latitudes for English sole (Castillo, 1995); migration associated with movement to and from the spawning grounds could therefore reasonably be expected to take place in the fall and spring just before and following this spawning activity.

For all modeled tag recoveries, movement was restricted to adjacent areas (north or south), and the same rates of movement were applied to all areas along the coast (Puget Sound and the Strait of Georgia were excluded). This simplification restricted the P matrices (area×area for each month) to nonzero values in only the first off-diagonals, and repeated the same parameters within each diagonal (Table 2). The time-increments considered reflect the trade-off between biologically realistic hypotheses and the likely constraints on complexity in future stock assessments. English sole move-

ment over large distances has been observed to occur at a rate of three to eight kilometers per day (Forrester, 1969). With an average latitudinal span of 138 kilometers per area, it could therefore take 17 to 46 days for at least some English sole to cross a single area; this rate of movement indicated that a one-month timestep would be appropriate to accommodate interarea migration. In the simplest hypotheses, movement was considered to take place only at one time per year, and all P matrices contained only zeros except for the month in which movement occurred but allowed northerly and southerly movement to differ. These models included movement occurring in January (the standard break for assessment years), October, November, May, and June. An additional four hypotheses were considered that included movement in the fall and spring and that resulted in two different P matrices—one movement applied in a spring month and one in the fall with two unique parameters (a north and south movement) in each. Five hypotheses included movement during more than two months of the calendar year (Table 3).

Bayesian implementation

Prior distributions are required for each of the model parameters (Table 4). Priors were selected to be non-informative, allowing the likelihood function to dominate the posterior probability distribution. However, the choice of appropriate noninformative priors is difficult and case specific, requiring estimation on the appropriate scale for each parameter, often either diffuse (near uniform) or uniform over log-space (Gelman et al., 1995). In some cases, additional information was available with which to constrain the priors. All movement parameters were bounded between zero and one, and had uniform prior density. The maximum possible reporting rate was estimated by calculating one minus the proportion of recovered tags with incomplete information (no area, date, or both, and were not included in the analysis); this value was 0.92 per tag lifetime. If tag loss rate was

Table 3

Specific structure of the movement parameter matrices (*P*) for the 14 models explored in this analysis. One-season models (1–5) allowed movement to occur only once per year in the month shown, two-season models (6–9) allowed movement in two months per year, and monthly movement models (10–14) allowed movement in >2 months per year, as specified.

Model	Movement parameter matrices
One-season: two movement parameters (north and south in the month indicated)	
1	October
2	November
3	January
4	May
5	June
Two-seasons: four movement parameters (north and south in the months indicated)	
6	October, May
7	October, June
8	November, May
9	November, June
Monthly movement	
10	One parameter, all months, north=south
11	Two parameters, all months, north and south
12	One parameter, movement from May through November only
13	Two parameters, movement from May through November only, north and south
14	Four parameters, movement from May through October, north and south, and movement in November, north and south

Table 4

Prior distributions applied to parameters in all models. NA = not applicable.

Parameter	Name (number)	Distribution (parameters describing the prior distribution)	Prior distribution parameter values	Parameter bounds: low, high
F_d	Fishing mortality rate/year for each data set (n=25)	Lognormal (median, coefficient of variation)	F_{med}, 0.25	–5, 5
F_{med}	Median fishing mortality rate/year	Scaled beta (*shape1*, *shape2*)	1.02, 1.02	0.01, 0.80
M	Natural mortality rate/year	Fixed	0.26	NA
Ω	Tag loss rate/year	Scaled beta (*shape1*, *shape2*)	1.02, 1.02	0.01, 1.52
ϕ	Reporting rate	Scaled beta (*shape1*, *shape2*)	1.02, 1.02	0.21, 0.92
p_n, p_s	Proportion moving (***n***=1–4, depending on the hypothesis)	Uniform (bounds)	NA	0.00, 1.00
K_c	Overdispersion (***n***=4)	Gamma (shape, rate)	1.001, 0.01	0.001, 1000

zero and all tags were recaptured immediately, then the minimum possible reporting rate would be the ratio of tags returned to total releases; this value, 0.21, is therefore a logical lower bound on reporting rate. These bounds are comparable to the range of reporting rates observed for other species (Gaertner and Hallier, 2004). Long-term tag-induced mortality (included in tag loss in this model) may be twice that of natural mortality (Manzer, 1952). Tag loss was also reported to be high

because of corrosion of the tag pins for English sole tagged in a similar fashion over roughly the same period (Forrester and Ketchen, 1955). The proportion of tags lost could not have been greater than 0.78 because 0.22 of the tags were recovered; this value (0.22) is greater than that used to calculate the minimum bound for reporting rate because it includes those tags recovered without full reporting of location information. The value of 0.78 for the proportion of tags lost corresponds to a

maximum instantaneous rate of tag loss of 1.52, which was used as the upper bound for this parameter. The tag loss rate was given a lower bound of 0.01 in the absence of other information. The median rate of fishing mortality across data sets (F_{med}) was bounded to lie between 0.01 and 0.8. Reporting rate, tag loss rate, and median fishing mortality rate were all assigned a scaled-beta distributed prior with both shape parameters equal to 1.02. This prior has the desirable properties of a nearly uniform density over most of the parameter space, except immediately adjacent to the bounds, which have zero density. Data set-specific fishing rates (F_d) were assumed to be related in a common hierarchy; the values of these parameters were constrained with a lognormal prior (F_{med}, 0.25). The overdispersion parameters (one for each category of data) were given a gamma-distributed prior (shape=1.001, rate=0.01) and bounded to lie between 0.001 and 1000. This choice reflected the desire for a generally uninformative prior, but one that favored a substantially larger variance than that in a simple Poisson likelihood. Exploration of the sensitivity of the model inference to the choice of priors was conducted by changing the values, rerunning the analysis, and comparing the results. The effect of six key prior distributions were explored through sensitivity analysis by modifying the shape of these distributions: doubling the coefficient of variation of the prior on deviations from F_{med}, using a uniform prior on the log scale for the overdispersion parameter (k), reducing the upper bound on movement parameters to 0.25, extending the prior bounds on reporting rate (0.01 to 1.0), and sequentially setting the priors on tag loss and F_{med} to be uniform.

This model was programmed in AD Model Builder© (Otter Research Ltd., Sidney, B.C., Canada), which uses a Metropolis-Hastings algorithm to sample from the joint posterior distribution of all model quantities. Markov-chain Monte-Carlo (MCMC) sampling was performed for five to fifteen million iterations for each hypothesis. Each chain was thinned by taking every 1000th (or fewer) draws to achieve low autocorrelations (<0.3) within chains and by taking nearly equal actual and effective (modified to account for autocorrelation) sample sizes. Convergence was assumed to have occurred for each chain when the criteria above were met, visual inspection of trace plots and cumulative quantiles (0.05, 0.5, 0.95) indicated stationarity in all model parameters, and most parameters had a Geweke statistic (Geweke, 1992) less than 1.96 (this statistic can be interpreted as a z-score and will produce some significant values due to random chance).

Bayes factors are frequently used in Bayesian analyses to compare the weight of evidence among various model hypotheses, accounting for differences in the number of estimated parameters (Gelman et al., 1995; Burnham and Anderson, 2002). In the present analysis, harmonic mean posterior likelihood for each model was used to calculate approximate Bayes factors (Kass and Raftery, 1995). Model support (amonth those compared) is based on twice the log of the ratio of mean likelihoods (hereafter referred to as the Transformed Bayes

Factor, TBF), judged on the following scale: 0–2, not worth more than a bare mention; 2–6, positive; 6–10, strong; and >10, very strong support for one model over another (Kass and Raftery, 1995). For this application, the TBF metric appeared quite stable and robust to sampling effects arising from the posterior distributions in preliminary testing.

Results

Most of the recoveries of tagged fish across all data sets occurred in the area that initial tagging took place, indicating relatively low rates of movement over all areas and time periods. Of the 3464 tagged English sole recovered off the open coast, only 130 (3.8%) of these had moved from the area of tagging and only 55 (1.6%) had moved more than one area. Low levels of exchange were particularly pronounced for those fish released in Puget Sound and the Strait of Georgia. Of 24,408 tagged English sole released in Puget Sound and the Strait of Georgia, only 12 (0.002%) of the 5756 recoveries were captured off the open coast. Conversely, only 3 (0.001%) of 4232 tagged fish recovered from 32,431 released fish (including some that could not be included in the quantitative analysis) on the open coast were recaptured within Puget Sound or the Strait of Georgia. A single release of 282 English sole in the Strait of Juan de Fuca (Forrester, 1969) resulted in 59 recoveries, 34 of which were from the open coast, mostly off Washington, but recoveries ranged as far south as Oregon. In aggregate, these results indicate that Puget Sound and the Strait of Georgia are substantially isolated from the open coast, but that mixing of adults does occur in the Strait of Juan de Fuca (and possibly at the north end of the Strait of Georgia). Therefore, in all model hypotheses considered in this analysis, Puget Sound and the Strait of Georgia were not included as part of the coastal population.

All five single time-step models resulted in English sole movement estimates that were three to four times more southerly than northerly, although the posterior distributions for the movement parameters were not identical (Fig. 2). TBFs of 49–152 indicated strong support for the model, allowing movement only in November over all other single time-step models (Table 5) . In this model, the posterior median proportion of English sole expected to move to the north each year was 0.08, and the 90% posterior interval ranged from 0.05 to 0.12, and 0.31 (0.25–0.39) to the south. The model including movement in January was second best (TBF=49) and qualitatively similar with 0.06 (0.04–0.09) moving to the north, and 0.26 (0.19–0.34) moving to the south each year.

The two-season models also showed some consistency in parameter estimates regardless of the months in which movement was allowed to occur (Fig. 3). TBFs of 16–26 indicated strong support for the model allowing movement in November and May (Table 5). Results from this two-season model showed that the proportion of English sole moving in the spring and to the
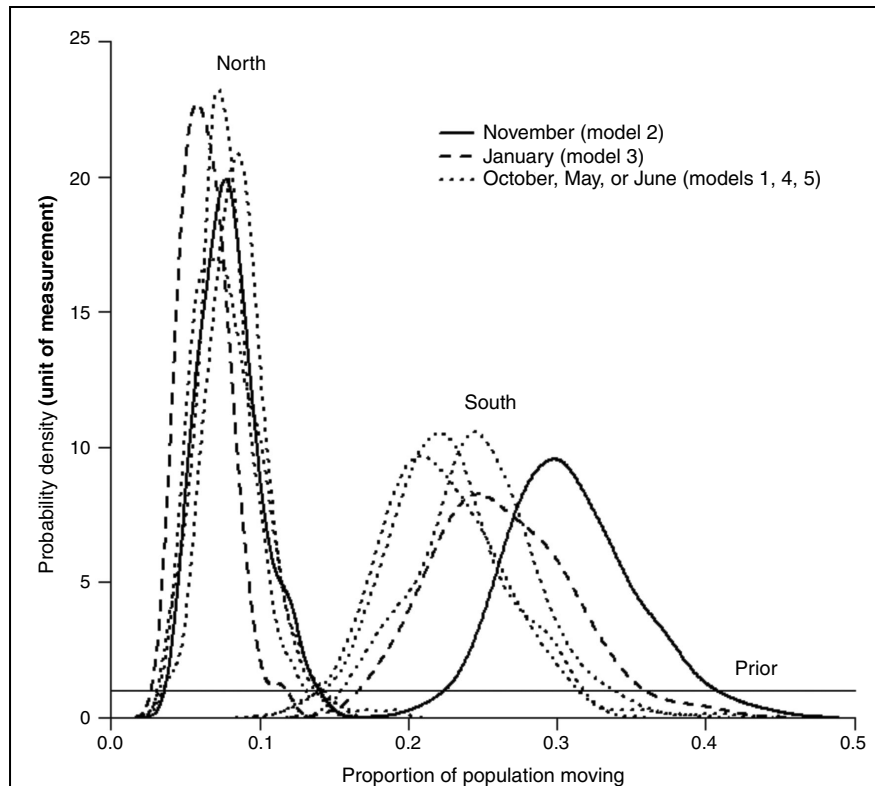
Figure 2

Prior (horizontal line) and posterior distributions for movement parameters (proportion of the population moving) in the single-season models (1–5). Solid lines indicate movement to the north and south in November (model 2), dashed lines show movement only in January (model 3), and dotted lines show movement in October, May, or June (nodels 1, 4, 5; not separated in the legend).

north to be 0.10 (0.04–0.18) and to be 0.09 (0.04–0.l7) to the south. In the fall, parameter medians were 0.05 (0.02–0.09) to the north and 0.27 (0.17–0.38) to the south. This model was strongly supported over the best single-season model (TBF=49). The next best two-season model (TBF=16) allowed movement in October and May. Movement rates estimated from two-season models showed a similar pattern to those estimated from the single time-step models. Net movement to the south was identified in both cases, primarily in the fall, and although movement rates from two-season models were somewhat reduced, they were applied twice per year. The cumulative expected value for movement in the best two-season model was 0.11 to the north and 0.30 to the south, very close to the values from the best single-season model.

When movement was allowed in each month of the year, parameter estimates were much smaller per month (implying a similar magnitude of annual movement), but greater movement was still predicted to the south than to the north (Fig. 4). The best monthly model (no. 11) included separate proportions moving north (0.02 [0.02–0.03]), and south (0.04 [0.03–0.04]). This model received strong support over the best two-season model

(TBF=101), but only slightly more support from the data (TBF=3) than constant and equal movement all year (model10). The cumulative expected value of this movement was 0.21 to the north and 0.29 to the south, similar to the best two-season model, but with more net movement to the north. With only a single movement parameter, the median proportion moving in model 10 was 0.03 (0.02–0.04). Although more complex models including movement in only some months were explored, none were supported by the data (TBFs 16–34) over models 10 or 11 (Table 5).

Model parameters other than movement rates showed no obvious restriction by their priors, although posterior distributions included much of the marginal parameter space within the prior bounds. For all models, the overdispersion parameters (k_c) for each data type had substantial density below 1.0, indicating variability far in excess of that expected from a Poisson distribution (Fig. 4). Reporting rate was generally less than 0.75 and was highly correlated with the median fishing mortality rates. Tag loss rates were predicted to be high (often greater than 0.5), but quite uncertain. Sensitivity analysis to the shape of key prior distributions did not result in any substantial changes for the posterior

Table 5

Transformed Bayes Factors (TBFs) used for comparison of models allowing movement in only one month (one-season), in two months (two-seasons), or in >2 months of each year (monthly). Support for one model over another was based on the following TBF scale: 0–2, not worth more than a bare mention; 2–6, positive; 6–10, strong; and >10, very strong (Kass and Raftery, 1995).

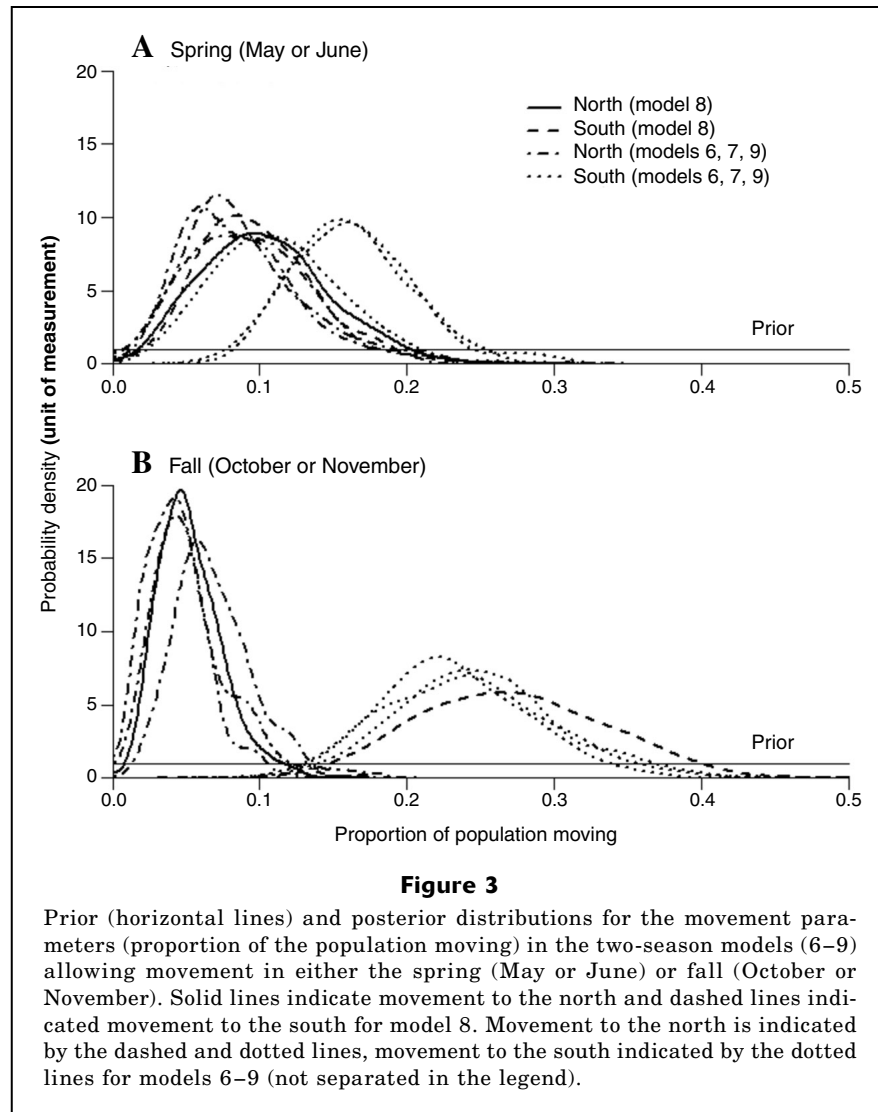| Comparison	Model	Number of estimated parameters	TBFs: 2×log (*likelihood of the best model/ likelihood of model on row*)
Among one-season models:	1	34	97
	2	34	Best one-season model
	3	34	49
	4	34	134
	5	34	152
Among two-season models:	6	36	16
	7	36	26
	8	36	Best two-season model
	9	36	20
Among monthly models:	10	33	3
	11	34	Best monthly model
	12	33	16
	13	34	14
	14	36	34
Among time-steps:			
Best one-season model	2	34	151
Best two-season model	8	36	101
Best monthly model	11	34	Best model

distributions of the movement parameters for any of the six alternate sets of parameter values considered (Fig. 5). A further check was made by rerunning the model assuming high (75%) initial tag loss; again no appreciable change in posterior distributions was observed, and median parameter values changed by less than 0.001.

The ability to quantify the plausibility of the observed data given the fitted model, Bayesian goodness-of-fit, was examined by checking the posterior predictive distribution (the probability distribution for an unobserved data point), which can indicate the degree to which the model structure, priors, and likelihoods assumed in the model are appropriate (Gelman et al., 1995). The posterior predictive distribution for expected recoveries corresponding to each of the observed recoveries was generated during MCMC sampling. The posterior predictive check compares the observed data to a distribution of predictions and summarizes the information across data types. The mean standardized residuals were calculated by dividing the raw residual (between the observed value and the xth percentile of the posterior predictive distribution) by the expected standard deviation (based on the negative binomial likelihood), and by taking the mean of these values for each number of observed recoveries. Figure 6 shows the mean standardized residual for the 95th, 50th, and 25th percentiles of the posterior distribution of expected recoveries for

each of the four types of data. There are a few observations well in excess of the range expected for standardized residuals, primarily for the type-1 data sets. In addition, some trend was observed in the residuals for the type-4 data sets with larger residuals occurring at the largest observed values. Further, although the zero-line for the 5th percentile of the predictions does lie below more than 95% of the residuals, there appears to be an excess of residuals above the zero-line for the 95th percentile of the posterior predictions. This excess of residuals indicates that model predictions generally resulted in fewer recoveries (for some time and space combinations) than were observed.

Discussion

Bayesian analyses are ideal for fisheries applications because uncertainty is explicitly and transparently incorporated into them, they allow for the use of several data sources (Hilborn and Mangel, 1997), provide easily interpretable probability inference (Wade, 2000), and yield results suitable for formal decision analysis (McAllister et al., 1994). The Bayesian framework developed here allows calculation of probability distributions for key parameters governing English sole movement rates. The results from this analysis qualitatively support what can be directly inferred from the original

Figure 3

Prior (horizontal lines) and posterior distributions for the movement parameters (proportion of the population moving) in the two-season models (6–9) allowing movement in either the spring (May or June) or fall (October or November). Solid lines indicate movement to the north and dashed lines indicated movement to the south for model 8. Movement to the north is indicated by the dashed and dotted lines, movement to the south indicated by the dotted lines for models 6–9 (not separated in the legend).
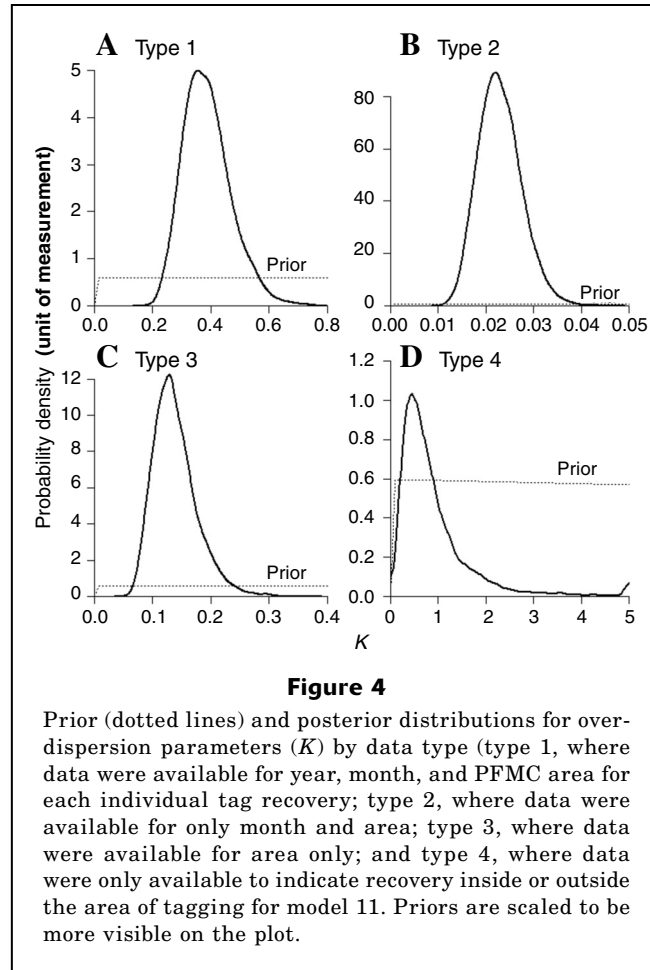
published analyses of these data sets: English sole are not highly mixed across the entire coast but are also not sedentary at the scale of hundreds of kilometers. The approach of modeling the open coast separately is supported by the observation that only rarely are tagged English sole observed to move between the open coast and Puget Sound or the Strait of Georgia. This result may be specific to the biological and ecological habits of the species but it is commonly assumed to be the result for other species as well.

Also in concordance with historical observations about the seasonal and latitudinal movements of flatfish, the current analysis supports models that include more than just one movement per year. Movement appears to be of greater magnitude in the fall, just before the spawning season, but this pattern is not supported when monthly hypotheses are explored. This lack of a consistent pattern could be due to interannual variability in spawning activity. Research shows that spawning,

in the case of English sole, is related to temperature (Kruse and Tyler, 1983; Peterman et al., 1987), as well as latitude (Castillo, 1995). Because this analysis lacked temperature as a covariate, there may not be adequate data support for specific spring and fall movement rates with the potential variation in timing of spawning activity (when compared to uniform monthly movement). If specific environmental information were available for each year of the analysis, covariates could be developed to improve assignment of the correct month for pre-spawning migration to the south and post-spawning migration to the north.

Of potential importance to stock assessment is the net movement to the south predicted by nearly all of the models. Although this pattern does not fit the data substantially better than simple diffusion (equal movement north and south) throughout the year, increased movement to the south estimated in simpler models may be worth further investigation. The effect of net

Figure 4

Prior (dotted lines) and posterior distributions for over-dispersion parameters (K) by data type (type 1, where data were available for year, month, and PFMC area for each individual tag recovery; type 2, where data were available for only month and area; type 3, where data were available for area only; and type 4, where data were only available to indicate recovery inside or outside the area of tagging for model 11. Priors are scaled to be more visible on the plot.
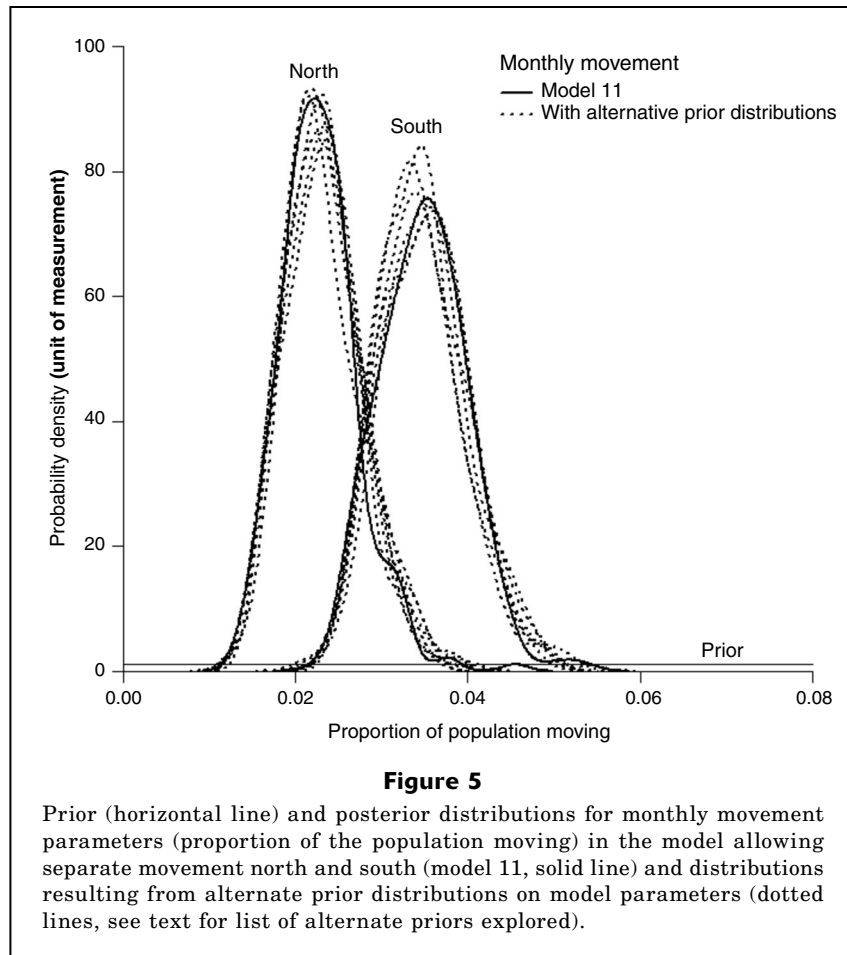
southerly movement on equilibrium harvests of adult English sole should be explored because this movement would have implications for current and future management strategies.

Given the presence of significant outliers, the degree of support among models based on the posterior probability should be interpreted with caution. There are many potential reasons for the outliers in observed recoveries present in the data. Although an attempt was made to retain a structure commensurate with what might be possible to replicate in a stock assessment, the analysis may simply not be complex enough. Many parameters, such as fishing mortality, reporting rate, and tag loss rate, are assumed to be temporally and spatially invariant because of a lack of available data. Spatial differences in the location of the initial tagging within the larger PMFC area were not accounted for and reporting rates could be variable due to a mixture of Canadian and U.S. fishing vessels with varied incentives for returning tags. During much of the time period over which tags were recovered, there was a substantial fishery for mink food; this fishery may have had very different handling and tag-recovery practices than those of the concurrent fishery targeting fish for human con-

sumption. The relative effects of violating these various assumptions could be addressed in the future through simulation testing.

Spatially and temporally local increases in fishing-induced mortality rates, resulting in additional recaptured tags, could have generated many of the positive residuals from a model that does not allow fishing mortality to vary within a single data set. Exploration of how model parameters are influenced by localized increases in fishing mortality should be explored through future simulation analysis. Another potential use of this method is to extend the hierarchal model of fishing mortality rates, allowing them to vary over space, time (or both) within a data set. Estimation of the coefficient of variation of this distribution could also be explored. The coefficient of variation of total U.S. catch over the years in which tagging projects were conducted for English sole is on the order of 0.5, indicating that interannual variation in fishing mortality rate may be an important factor absent from the analysis. Similarly, a hierarchical approach could be taken with regard to movement parameters between specific areas. Geographic regions could be defined on the basis of likely bathymetric features such as submarine canyons or rocky headlands

Figure 5

Prior (horizontal line) and posterior distributions for monthly movement parameters (proportion of the population moving) in the model allowing separate movement north and south (model 11, solid line) and distributions resulting from alternate prior distributions on model parameters (dotted lines, see text for list of alternate priors explored).
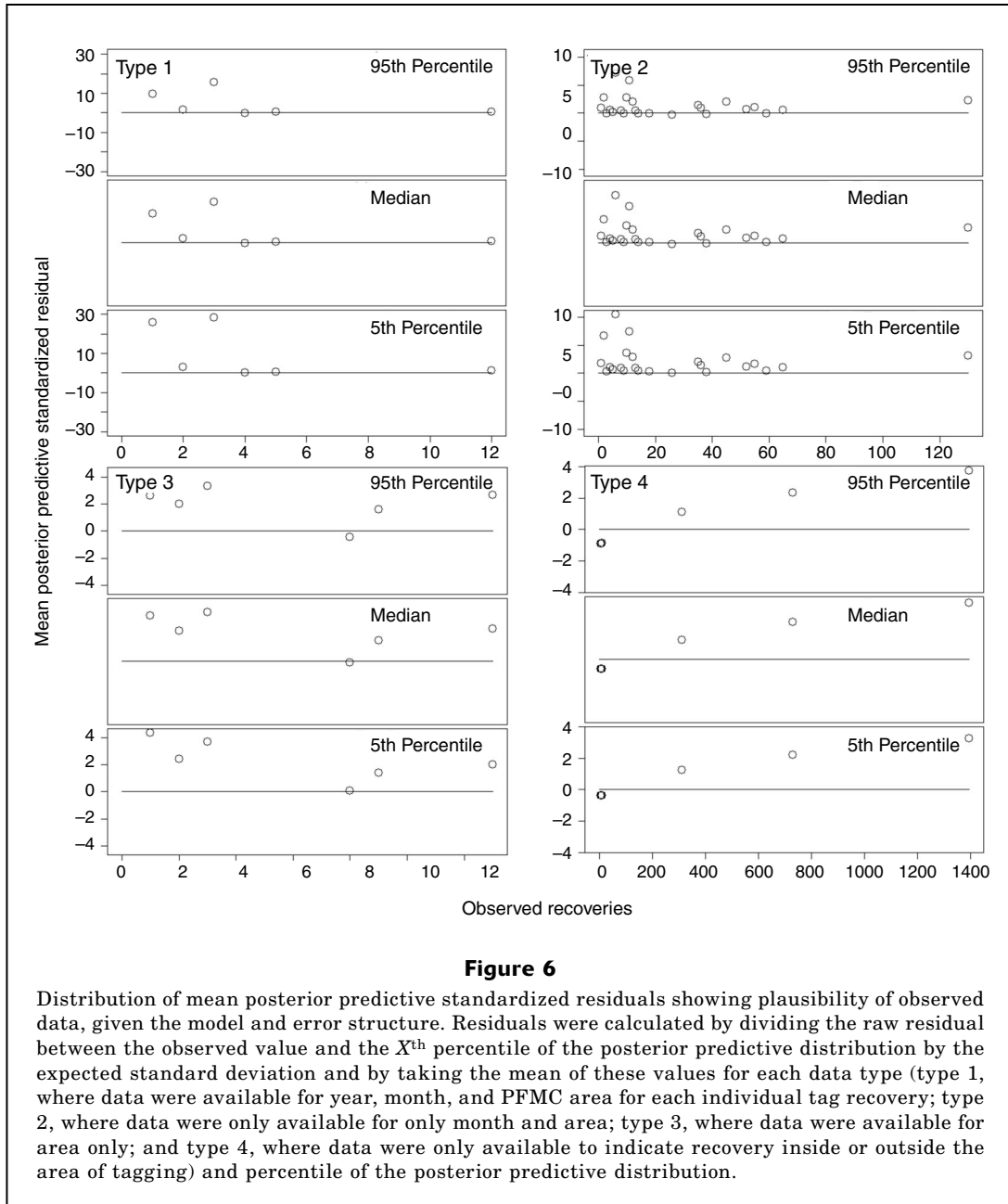
that might serve to disrupt movement along the coast. Although conceptually appealing, these extensions may cause technical problems for MCMC because of the low information content of the aggregated historical tagging data.

Movement rates are notoriously difficult to estimate (Xiao, 1996). However, the general approach to integrated tag analysis based on maximum likelihood has been found to be reliable through simulation testing (Maunder, 2001). In this application, the error structure, although intended to accommodate clustered recoveries and the inclusion of zero recoveries in many space-time combinations, may be inadequate for the observed level of variability in recoveries. Specifically, there were many cases of observed recoveries in areas where no recoveries could have been predicted under simple models, given the structure imposed by the population dynamics that were assumed. Future extension and simulation testing to evaluate other error structures, such as zero-inflated Poisson or negative binomial models, should be undertaken. Further, there may be interactive effects of the prior distributions used, despite lack of observed response to one-dimensional sensitivity testing. Some important elements of uncertainty may not have been included in the present analysis; however, uncertainty is a problem faced by most modeling applications in fisheries.

It is unlikely that future stock assessments will be structured around areas as small as PMFC areas. Conversion of the movement rates reported here for use as priors in stock assessments will require assumptions regarding the distribution of biomass within areas modeled and the correspondence of the areas modeled to PMFC areas. Research survey data may provide a fishery-independent source on which to base these assumptions. The estimation of fishing mortality in this analysis should also be considered if it is to be applied within a stock assessment framework. These issues would appear to be no more daunting than currently accepted assumptions of thorough mixing across the coast or of completely isolated stock groups.

Tagging programs for groundfish species off the U.S. west coast have generally decreased over the last 75 years, and there are no plans for large-scale tagging programs in the future. An analysis of all available tagging data may therefore provide the only quantitative source of guidance and an important avenue to allow for uncertainty in movement rates without new data with which to estimate these rates. This type of method

Figure 6

Distribution of mean posterior predictive standardized residuals showing plausibility of observed data, given the model and error structure. Residuals were calculated by dividing the raw residual between the observed value and the Xth percentile of the posterior predictive distribution by the expected standard deviation and by taking the mean of these values for each data type (type 1, where data were available for year, month, and PFMC area for each individual tag recovery; type 2, where data were only available for only month and area; type 3, where data were available for area only; and type 4, where data were only available to indicate recovery inside or outside the area of tagging) and percentile of the posterior predictive distribution.

has been demonstrated for English sole but could be extended to other species that have also been the subject of multiple tagging projects. When compared with the status quo assumptions invoked with single (unit) or multiple independent stocks, these results may provide a more realistic integration of spatial movement off the west coast of North America into assessment models. Use of these results could lead to a better representation of the uncertainty associated with estimates of biomass and in the case of stock assessments, they could lead to predictions of exploitation rates that allow a sustainable resource.

Acknowledgments

Literature cited

Alverson, D. L.
    1960. A study of annual and seasonal bathymetric catch patterns for commercially important groundfishes of the Pacific Northwest coast of North America. Pac. Mar. Fish. Comm. Bull. 4:1–99.

Beverton, R. J. H., and S. J. Holt.
    1957. On the dynamics of exploited fish populations, 533 p. Chapman and Hall, London.

Bishop, Y. M. M., S. E. Feinberg, and P. W. Holland.
    1988. Discrete multivariate analysis: theory and practice, 9th ed., 557 p. Massachusetts Inst. Technology Press, Cambridge, MA.

Burnham, K. P., and D. R. Anderson.
    2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed., 488 p. Springer-Verlag, New York, NY.

Castillo, G. C.
    1995. Latitudinal patterns in reproductive life history traits of northeast Pacific flatfish. *In* Proceedings of the international symposium on north Pacific flatfish, p. 51–72. Alaska Sea Grant College Program, Univ. Alaska, Fairbanks, AK.

Cormack, R. M., and J. R. Skalski.
    1992. Analysis of coded wire tag returns from commercial catches. Can. J. Fish. Aquat. Sci. 49:1816–1825.

Day, D. E.
    1976. Homing behavior and population stratification in central Puget Sound English sole (*Parophrys vetulus*). J. Fish. Res. Board Can. 33:278–282.

Dorn, M. W.
    2002. Advice on West Coast rockfish harvest rates from Bayesian meta-analysis of stock-recruit relationships. N. Am. J. Fish. Manage. 22:280–300.

Forrester, C. R.
    1969. Results of English sole tagging in British Columbia waters. Pac. Mar. Fish. Comm. Bull. 7:1–10.

Forrester, C. R., and K. S. Ketchen.
    1955. The resistance to salt water corrosion of various types of metal wire used in the tagging of flatfish. J. Fish. Res. Board Can. 12:134–142.

Gaertner, D., and J.-P. Hallier.
    2004. Combining Bayesian and simulation approaches to compare the efficiency of two types of tags used in tropical tuna fisheries. Aquat. Living Res. 17:175–183.

Gelman, A., J. B. Carlin, and D. B. Stern.
    1995. Bayesian data analysis, 526 p. Chapman and Hall/CRC, Boca Raton, FL.

Geweke, J.
    1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In* Bayesian statistics 4 (J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, eds.), p. 169–193. Oxford Univ. Press, Oxford.

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

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

Harry, G. Y., Jr.
    1956. Analysis and history of the Oregon otter-trawl fishery. Ph. D. diss., 329 p. Univ. Washington, Seattle, WA.

Hart, J. L.
    1973. Pacific fishes of Canada, 740 p. Fish. Res. Board Canada, St. Andrews, New Brunswick, Canada.

Heifetz, J., and J. T. Fujioka.
    1991. Movement dynamics of tagged sablefish in the northeastern Pacific. Fish. Res. 11:355–374.

Hilborn, R.
    1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. Can. J. Fish. Aquat. Sci. 47:635–643.

Hilborn, R., and M. Mangel.
    1997. The ecological detective: confronting models with data, 315 p. Princeton Univ. Press, Princeton, N.J.

Hilborn, R., and C. J. Walters.
    1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty, 570 p. Chapman and Hall, London.

Jow, T.
    1969. Results of English sole tagging off California. Pac. Mar. Fish. Comm. Bull. 7:15–33.

Kass, R. E., and A. E. Raftery.
    1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.

Ketchen, K. S.
    1956. Factors influencing the survival of the lemon sole (*Parophrys vetulus*) in Hecate Strait, British Columbia. J. Fish. Res. Board Can. 13:647–694.

Kruse, G. H., and A. V. Tyler.
    1983. Simulation of temperature and upwelling effects on the English sole (*Parophrys vetulus*) spawning season. Can. J. Fish. Aquat. Sci. 40:230–237.

Manzer, J. I.
    1952. The effects of tagging upon a Pacific coast flounder, *Parophrys vetulus*. J. Fish. Res. Board Can. 8:479–485.

Maunder, M. N.
    2001. Integrated tagging and catch-at-age analysis (ITCAAN): model development and simulation testing. *In* Spatial processes and management of marine populations (G. H. Kruse, N. Bez, A. Booth, M. W. Dorn, S. Hills, R. N. Lipcius, D. Pelletier, C. Roy, S. J. Smith, and D. Witherell, eds.), p. 123–146. Alaska Sea Grant College Program, Univ. Alaska, Fairbanks, AK.

McAllister, M. K., E. K. Pikitch, A. E. Punt, and R. Hilborn.
    1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 51:2673–2687.

McGarvey, R., and J. E. Feenstra.
    2002. Estimating rates of fish movement from tag recoveries: conditioning by recapture. Can. J. Fish. Aquat. Sci. 59:1054–1064.

Menasveta, D.
    1958. Migration and fishing mortality of English sole (*Parophrys vetulus*) in Saratoga Passage, and adjacent waters. M.S. thesis, 84 p. Univ. Washington, Seattle, WA.

Millar, R. B., and R. D. Methot.
    2002. Age-structured meta-analysis of the U.S. West Coast rockfish (Scorpaenidae) populations and hierarchical modeling of trawl survey catchabilities. Can. J. Fish. Aquat. Sci. 59:383–392.

Pattie, B. H.
    1969. Dispersal of English sole, *Parophrys vetulus* tagged off the Washington coast in 1956. Pac. Mar. Fish. Comm. Bull. 7:11–14.

Peterman, R. M., M. J. Bradford, and G. H. Kruse.
    1987.   Simulation model of English sole (*Parophrys vetu-lus*) population dynamics in Washington and Oregon coastal waters.   Can. J. Fish. Aquat. Sci. 44:1870–1878.

Punt, A. E. and R. Hilborn.
    1997.   Fisheries stock assessment and decision analy-sis: the Bayesian approach.   Rev. Fish Biol. Fish. 7:35–63.

Punt, A. E., F. Pribac, T. I. Walker, B. L. Taylor and J. D. Prince.
    2000.   Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model.   Mar. Freshw. Res. 51:205–220.

Quinn, T. J. II., and R. B. Deriso.
    1999.   Quantitative fish dynamics, 542 p.   Oxford Univ. Press, New York., NY.

Robichaud, D., and G. A. Rose.
    2004.   Migratory behaviour and range in Atlantic cod: inference from a century of tagging.   Fish Fish. 5:185–214.

Wade, P. R.
    2000.   Bayesian methods in conservation biology.   Cons. Biol. 14:1308–1316.

Xiao, Y.
    1996.   A framework for evaluating experimental de-signs for estimating rates of fish movement from tag recoveries.   Can. J. Fish. Aquat. Sci. 53:1272–1280.