INTRODUCTION

Ecological processes in marine systems are characterized by high variability and occur on large spatial and temporal scales (Stommel 1963). These large scales are difficult to explore experimentally, contributing to the wide use of statistical methods to describe marine ecological processes. Stock assessment methods have consequently been developed that integrate multiple data streams to estimate marine species productivity, stock status, relative cohort strength, compensatory production, and many other ecological processes (Quinn & Deriso 1999).

Larval and juvenile survival is one example of an ecological process occurring at large spatial scales. It is ecologically important because it drives large-scale ecological patterns.
annual fluctuations in cohort strength for many marine species (Hjort 1926, Cushing 1990). The relative abundance of cohorts can be estimated relatively easily for anadromous species such as salmon, where individuals can be counted as they return to freshwater environments to spawn (Mueter et al. 2002, Lawson et al. 2004). By contrast, cohort strength in marine groundfish is difficult to estimate, and generally requires age-composition data (Francis 2011). For these groundfish species, cohort strength is typically estimated using assessment methods such as virtual population analysis (Myers 2001) or integrated analysis (Maunder & Punt 2013).

Relative cohort strength is generally believed to be determined during early larval or juvenile survival, as well as the distances over which marine survival is correlated (Mueter et al. 2002). However, the analysis of among-species correlations in recruitment for groundfishes is comparatively sparse. The few studies that exist on correlations in groundfish recruitment typically use recruitment estimates from stock assessment methods as ‘data’ in a secondary regression model (Wespestad et al. 2000, Field & Ralston 2005, Mueter et al. 2006). This approach is questionable because assessment methods will induce autocorrelation between consecutive recruitment estimates. This autocorrelation may underlie apparent correlations with environmental time-series data and will drastically decrease the effective degrees of freedom available for identifying any environmental relationship. Analysis of outputs from stock assessments is most appropriate when the assessment is based on virtual population analysis (VPA; Gulland 1965) because VPA does not specify a parametric model for the stock–recruit relationship and provides recruitment estimates that are statistically independent among years (Myers 2001). However, VPA models cannot incorporate the data types or account for the types of sampling and process errors that are typically used by modern stock assessment methods, and illustrate errors when selectivity is mis-specified or catch-at-age data have measurement errors (Radomski et al. 2005).

Besides environmental effects, cohort strength is also affected by spawning biomass and fecundity (Morgan et al. 2011). Therefore, an estimate of environmental effects on recruitment must account for the influence of spawning biomass. Because groundfish cohort strength is typically estimated using assessment methods, the most promising avenue for ecological research regarding shared patterns in recruitment is developing stock assessment methods that incorporate data from multiple species simultaneously, hereafter referred to as recruitment-linked multispecies assessment methods. Such methods should use observational-level data from several species simultaneously, estimate imprecision accurately to account for processes ranging the observation-level data through to the level of ecological inference, and control for the confounding effect of the stock–recruit relationship on environmental patterns in cohort strength. Recruitment-linked multispecies assessment methods should also use modeling assumptions that are appropriate for each species individually, e.g. use the age- or length-based selectivity that is deemed appropriate for each species and fishery, rather than applying a ‘cookie-cutter’ model with a single, pre-specified selectivity pattern for each species.

Here we demonstrate a generic method for linking existing stock assessments and data inputs for several species simultaneously. This method jointly optimizes an overall likelihood while simultaneously estimating an unobserved but shared time series representing relative cohort strength. We demonstrate the method by estimating similarities in recruitment for several species of groundfish off the US West Coast while modifying peer-reviewed stock assessments conducted using Stock Synthesis (Methot 1990, 2009, Methot & Wetzel 2013). We chose Stock Synthesis because this or very similar methods are used for stock assessments in much of the United States, Canada, and Australia. We also note that the algorithm for linking individual stock assessments itself could be generalized to many other ecological processes and modeling platforms.

**METHODS**

**Joint optimization of single-species assessment models**

We first present a generic algorithm for jointly optimizing several single-species assessments simultaneously given a set of shared parameters (we present its application to recruitment-linked multispecies assessment models in the following section). Suppose there are $n_{\text{species}}$ species and each species $i$ has a vector of single-species parameters $\mathbf{\hat{\beta}}_i$ and data $\mathbf{x}_i$, where $\mathbf{\hat{\beta}}$ and $\mathbf{X}$ refer to the matrices of parameters and data for all species. These species also share a vector of parameters $\mathbf{\hat{\theta}}$ which represent a generic and shared effect, e.g. an environmental time series, spatial off-
set, or other influence on the population dynamics of all species (in this study, it will represent the component of annual cohort strength that is shared among species). We seek the set of parameters \( \hat{B} \) and \( \hat{\theta} \) that maximize the joint log-likelihood:

\[
\ln L_{\text{joint}}(\hat{B}, \hat{\theta} \mid X) = \sum_{i=1}^{n_{\text{species}}} \ln L(\hat{B}_i, \hat{\theta} \mid x_i)
\]  

(1)

This joint log-likelihood \( \ln L_{\text{joint}} \) can be maximized with minimal modification to existing model software using the following iterative optimization algorithm:

1. Identify starting values for the shared parameters \( \hat{\theta} \) and the single-species parameters \( \hat{B} \).

2. Optimize \( \hat{B} \) given the current values for \( \hat{\theta} \). Each species is statistically independent conditional on a fixed value of \( \hat{\theta} \), so this can be done sequentially for each individual species:

\[
\hat{B}_i = \max_{\theta_i} \left[ \ln L(\hat{B}_i \mid \hat{\theta}, x_i) \right]
\]  

(2)

Optimization for each individual species \( i \) can be conducted using existing stock assessment software, data inputs, and model configurations, which represents a peer-reviewed approximation to population dynamics for species \( i \).

3. Improve \( \hat{\theta} \) given the current values for \( \hat{B} \). This is done sequentially for each individual element \( \theta_j \) of \( \hat{\theta} \) by evaluating the joint log-likelihood (Eq. 1) given different proposal values for \( \theta_j \) while leaving \( \theta_{-j} \) (all elements of \( \hat{\theta} \) except \( \hat{\theta}_j \)) fixed, and selecting the best of the proposed values:

\[
\hat{\theta}_j = \max_{\theta_j} \left[ \sum_{i=1}^{n_{\text{species}}} \ln L(\hat{\theta}_j \mid \hat{\theta}_{-j}, \hat{B}_i, x_i) \right]
\]  

(3)

4. Repeat steps 2 and 3 until a full iteration does not result in an improvement in \( \ln L_{\text{joint}} \) above a chosen threshold.

5. For final tuning of the parameter estimates, simultaneously optimize the joint likelihood of all model parameters \( \hat{B} \) and \( \hat{\theta} \) using a conventional optimization procedure:

\[
<\hat{\theta}, \hat{B}> = \max_{\theta, B} \left[ \sum_{i=1}^{n_{\text{species}}} L(\hat{\theta}, \hat{B}_i \mid x_i) \right]
\]  

(4)

**Stock Synthesis treatment of relative cohort strength**

We now illustrate how this generic approach can be applied to recruitment. The shared parameter vector \( \hat{\theta} \) in the generic algorithm is replaced by \( \hat{s} \), an index of cohort strength that is shared among species and accounts for among-species correlations in recruitment. Similarly, the matrix of single-species parameters \( \hat{B} \) is replaced by many parameters that are specific to Stock Synthesis, as subsequently defined. We present components of the Stock Synthesis model as necessary, but see Methot & Wetzel (2013) and its supplementary materials for a complete description of the Stock Synthesis model.

Stock assessment models frequently approximate the median expected abundance of age-0 individuals (termed recruits) as a function of spawning biomass, using a common re-parameterization of the Beverton-Holt stock–recruit model (Methot & Wetzel 2013):

\[
\hat{R}_{i,t} = \frac{4h_i R_{i,0} S_{B,i,t}}{S_{B,i,t}(1-h_i) + S_{B,i}(5h_i - 1)}
\]  

(5)

where \( \hat{R}_{i,t} \) is the median expected recruitment given spawning biomass \( S_{B,i} \) for species \( i \) in year \( t \), \( R_{i,0} \) is the estimated unfished recruitment for species \( i \), \( S_{B,i,0} \) is unfished biomass for species \( i \) (calculated from \( R_{i,0} \), growth, and mortality parameters), and \( h_i \) is the steepness parameter (a measure of the magnitude of recruitment compensation) for species \( i \). The number of recruits in a given year is additionally affected by unobserved environmental, biological, and other factors that cause recruitment to be greater or less than otherwise expected on the basis of spawning biomass:

\[
\ln(R_{i,t}) = \ln(\hat{R}_{i,t}) + \hat{r}_{i,t}
\]  

(6)

where \( \hat{r}_{i,t} \) is an estimated deviation from median recruitment given spawning biomass \( S_{B,i,t} \).

Recruitment deviations are penalized in the likelihood to approximate their treatment as normally distributed random effects:

\[
L(\hat{r}_{i,t} \mid \theta_i, x_i) = \text{Normal}(\hat{r}_{i,t} \mid -d_{i,t} \frac{\sigma_i^2}{2}, \sigma_i^2)
\]  

(7)

where \( L(\hat{r}_{i,t} \mid \theta_i, x_i) \) is the likelihood penalty for recruitment deviation \( \hat{r}_{i,t} \), \( \text{Normal}(\hat{r}_{i,t} \mid -d_{i,t} \frac{\sigma_i^2}{2}, \sigma_i^2) \) is the probability density for \( \hat{r}_{i,t} \) given a bias-corrected normal distribution with mean \(-d_{i,t} \frac{\sigma_i^2}{2}\) and variance \(\sigma_i^2\), and \(d_{i,t}\) is fixed at the value from the stock assessment for species \( i \). The term \(-d_{i,t} \frac{\sigma_i^2}{2}\) is a bias-correction factor that ensures that the lognormal term \(\exp[\hat{r}_{i,t} - (d_{i,t} \frac{\sigma_i^2}{2})]\) has an expected mean of one for each year \( t \) (Methot & Taylor 2011), and \(\sigma_i^2\) represents the variance of recruitment deviations. \(d_{i,t}\) depends upon how informative the data for species \( i \) are about the recruitment deviation for year \( t \), and is fixed at the
value identified in each single-species assessment model.

The shared index $\hat{s}$ accounts for correlated recruitment deviations among species, and is composed of values $\hat{s}_i$ for all years $t$ within a chosen range. $\hat{s}$ corresponds to the shared vector $\hat{\theta}$ from the generic algorithm, and we use subscript $t$ because $\hat{s}$ is time-indexed in this application. The vector $\hat{s}$ is included in each single-species model as a penalty on single-species recruitment deviations, and this is the only way in which shared parameters $\hat{s}$ affect each single-species model (Schirripa et al. 2009):

$$L(\hat{s}_i | \beta_i, x_i) = \text{Normal} \left[ \hat{s}_i | \hat{q}_{i,1} + (1 + \hat{q}_{i,2})\hat{r}_{i,t}, (\bar{\sigma} + \hat{\tau})^2 \right]$$

where $L(\hat{s}_i | \beta_i, x_i)$ is the likelihood term that penalizes recruitment deviations towards the shared index of recruitment deviations, Normal$[\hat{s}_i | \hat{q}_{i,1} + (1 + \hat{q}_{i,2})\hat{r}_{i,t}, (\bar{\sigma} + \hat{\tau})^2]$ is the probability density for $\hat{s}_i$ from a normal distribution with mean $\hat{q}_{i,1} + (1 + \hat{q}_{i,2})\hat{r}_{i,t}$, and variance $(\bar{\sigma} + \hat{\tau})^2$. $\hat{q}_{i,1}$ and $\hat{q}_{i,2}$ are estimated parameters that relate the recruitment deviations $\hat{r}_{i,t}$ for each species $i$ to the shared index of recruitment deviations $\hat{s}$. $\bar{\sigma}$ is the average of the standard errors of the elements of the vector $\hat{s}$ and $\hat{\tau}$ is a parameter representing the additional variability in recruitment deviations around $\hat{s}$ for species $i$. Eq. (8) effectively penalizes recruitment deviations towards a time series $\hat{s}$ that is common to all species. The penalty is in addition to Eq. (7), which penalizes the recruitment deviations towards the estimated stock–recruit relationship for each species.

The parameter $\hat{q}_{i,1}$ accounts for a difference between the mean of the recruitment deviations for a species over the chosen range of years and the mean of $\hat{s}$ over those same years, while $\hat{q}_{i,2}$ changes the scale of recruitment deviations when compared with $\hat{s}$. In this application, we chose to fix $\hat{q}_{i,2}$ at $(1/\sigma_i - 1)$ while retaining $\sigma_i$ at its original value for each stock assessment. Given that recruitment deviations $\hat{r}_{i,t}$ have standard deviation $\sigma_i$, this causes $(1 + \hat{q}_{i,2})\hat{r}_{i,t}$ (from Eq. 8) to have a standard deviation of 1 for all species, ensuring that recruitment deviations have a common scale for all species. We then freely estimate species-specific values for $\hat{q}_{i,1}$ and $\hat{\tau}$. We also iteratively tune $\bar{\sigma}$ to approximately match the estimated standard error for $\hat{s}_i$ averaged over all years, i.e.

$$\bar{\sigma} \approx 1 / n_{\text{years}} \sum_{t=1}^{n_{\text{years}}} \text{SE}(\hat{s}_t)$$

as explained in detail in Appendix 1 $\bar{\sigma}$ represents an upper bound on the shrinkage of recruitment deviations towards $\hat{s}$, where the degree of shrinkage is determined by the sum of $\bar{\sigma}$ (the inherent uncertainty about the value of $\hat{s}$) and $\bar{\tau}$ (additional variability due to species-specific differences in the link between $\hat{s}$ and recruitment deviations). The recruitment deviations for each species are therefore never penalized towards $\hat{s}$ more strongly than the inherent estimation error associated with $\hat{s}$.

We introduce an additional step in the generic model fitting algorithm that is relevant to the application. Unobserved environmental variables (e.g. $\hat{s}$) are often modeled as having either a positive or negative effect on processes such as recruitment (Zuur et al. 2003). We therefore modify step #2 (Eq. 2) of the optimization algorithm to maximize the log-likelihood of single-species fixed effects $\beta_i$ when including either $\hat{s}$ or $-1 \cdot \hat{s}$ as the shared index of cohort strength for that species (Eq. 8), i.e. including either a positive or negative loading of $\hat{s}$ on recruitment deviations for a given species. The positive loading ($\hat{s}$) or negative loading ($-1 \cdot \hat{s}$) for species $i$ is then selected depending upon which leads to a greater log-likelihood. We subsequently use this positive or negative loading when optimizing $\hat{s}$. This modification allows the estimated time series $\hat{s}$ to have either a positive or negative relationship with recruitment deviations for each individual species included in the analysis.

Exploratory analysis using simulated data confirmed that the algorithm could appropriately identify positive and negative loadings when both were present. During model optimization, we monitored the standard deviation of estimated recruitment deviations $\text{SD}(\hat{r}_{i,t})$ and the maximum bias correction factor $d_{\text{bias}} = 1 - \text{SE}(\hat{r}_{i,t})^2/\sigma_i^2$ (Methot & Taylor 2011) to identify whether $\sigma_i^2$ or the bias correction for any species needed to be modified from the value used in the original stock assessment. We also confirmed that the algorithm results in the same final estimates given multiple different starting values.

**Case study data and application**

We compiled a database of 10 stock assessments conducted using Stock Synthesis version 3.20 or greater (Tables 1 & 2). Species were selected based on availability of an existing assessment model, rather than life history characteristics. These 10 species include 7 rockfishes, 2 flatfishes, and 1 sabrefish. Exploratory analysis using simulated data confirmed that the recruitment-linked multispecies assessment model is able to downweight species with conflicting recruitment time series by increasing $\sigma_i$ for these species, which decreases their influence on $\hat{s}$. 

...
The 10 assessments vary in terms of the time span for which recruitment deviations are estimated and the variability in recruitment deviations ($\sigma_i = 0.35–1.1$). Each assessment was used as peer-reviewed during the stock assessment process, except for the following changes. First, a parameter governing compensation in the stock-recruitment relationship (‘steepness’) was changed from fixed to freely-estimated in all assessments where it was previously fixed (i.e. sablefish, canary rockfish, Dover sole, Pacific ocean perch, and yelloweye rockfish). This was done so that uncertainty regarding the magnitude of recruitment compensation is propagated through to estimates of recruitment deviations. Second, all assessments used the standard Beverton-Holt model for recruitment, except for canary rockfish which used the flat-topped Beverton-Holt model. We therefore changed the recruitment model for canary rockfish to the standard Beverton-Holt to ensure consistent interpretation for recruitment deviations. Third, recruitment deviations were ‘turned on’ in the yelloweye and blackgill rockfish assessments. $\sigma_i$ was fixed at 0.45 and 0.35 for yelloweye and blackgill, respectively, after tuning to the variability in estimated recruitment deviations, and bias correction was also updated for each species using methods recommended by Methot & Taylor (2011). Finally, all assessment models were modified to incorporate the shared recruitment index $\hat{s}$.

**Model fitting and exploration**

We fitted the model using data for all stock assessments that originally estimated recruitment deviations (i.e. all except yelloweye and blackgill rockfishes) and estimated the common factor $\hat{s}$ for all years where all of these assessments estimated a recruitment deviation, i.e. the period 1978–2007. Model results and performance were explored in several ways. First, we present the estimate of $\hat{s}$ with associated standard errors. Standard errors were estimated using the inverse of a finite-difference
approximation to the Hessian of the log-likelihood at the maximum likelihood estimate. Due to the extremely large number of parameters, the Hessian was calculated for only the shared environmental factor \( \hat{s} \) and all recruitment deviations for the selected years, and it is this Hessian that was used to tune the value of \( \tau \) used in Eq. (7).

We also explored whether, and to what degree, recruitment deviations are ‘shrunk’ towards \( \hat{s} \) for the species that are used to estimate \( \hat{s} \). This shrinkage shows whether relatively uninformative species will borrow information about recruitment deviations from informative species (Punt et al. 2011). We demonstrate this effect by presenting estimates and standard errors for recruitment deviations with and without \( \hat{s} \). We then present estimates of recruitment deviations with and without \( \hat{s} \) for the species that were not used to estimate \( \hat{s} \) (yelloweye and blackgill rockfishes). These species were chosen for this test application because the previous stock assessments concluded that, while estimable, recruitment deviations were not parsimonious (Stewart et al. 2009). This example illustrates the effect of the shared index of cohort strength on data-poor species in the same region (Thorson 2011).

RESULTS

Single-species assessment model estimates of recruitment deviations are highly variable. Some species (e.g. canary rockfish) have relatively little variability in recruitment deviations, while others (e.g. sablefish) have relatively greater variability, and this is reflected in the relative magnitude of their variability in recruitment (Table 1: canary \( \sigma_i = 0.5 \), sablefish \( \sigma_i = 1.1 \)). Several species show a recruitment pulse around the years 1999–2000 (Fig. 1A). This spike is reflected in the first principal component of recruitment deviations, which explains 39% of variation in recruitment deviations, while second and third components explain an additional 26% and 17%, respectively (Fig. 1B). This principal component analysis is presented to demonstrate that a significant portion of total recruitment variability could be explained by a single recruitment factor. Tuning of \( \tau \) (Eq. 7) yields \( \tau = 0.2 \), while the average estimated standard error for \( \hat{s} \) is 0.207 (range of SE[\( \hat{s} \)] for all t: 0.168 to 0.255). Monitoring SD(\( \hat{r}_i \)) and \( 1 - \text{SE}(\hat{r}_i)^2 / \sigma_i^2 \) during the estimation process provides no evidence that bias correction for any species is misapplied (Methot & Taylor 2011). A positive loading for \( \hat{s} \) was selected by the model estimation algorithm for all 8 species, implying that all species were positively associated with the shared index of cohort strength. The shared index of cohort strength (Fig. 2) showed peaks around 1990–1991 and 1999–2000, followed by an extended period of low recruitment during the 2000s, and does not exactly mimic any of the 3 first principal components (Fig. 1B). The period of negative recruitment deviations 2002–2007 is more prolonged than any other seen in the time series, and is estimated to occur after accounting for the stock–recruit relationship for each species (i.e. the effect of changes in spawning biomass on recruitment).
Estimates of $\hat{\tau}_i$ (representing the variability in recruitment deviations around $\hat{s}$ for each species) show how influential $\hat{s}$ is on recruitment deviations for species $i$, where a species with a low estimate of $\hat{\tau}_i$ indicates that the shared index will be highly influential. Estimates of $\hat{\tau}_i$ (Table 3) show that Dover sole and Pacific ocean perch are highly influenced ($\hat{\tau} + \hat{\tau}_i = 0.2$) and darkblotched rockfish and bocaccio are influenced minimally ($\hat{\tau} + \hat{\tau}_i \approx 1$), with other species being intermediate. Recruitment deviations and standard errors for each species (Fig. 3) show the strong effect of $\hat{s}$ on recruitment deviations, e.g. for Dover sole, which has recruitment deviation estimates when incorporating $\hat{s}$ that show a compromise between $\hat{s}$ and originally estimated recruitment deviation time series. Specifically, the original time series shows recruitment pulses (i.e. $\hat{r}_i, t > 0.3$) in 1988, 1991–1992, 1997, and 2000, while the time series when incorporating $\hat{s}$ shows recruitment pulses in only 1991 and 2000, with greater stability during the previously-estimated 1988 and 1997 pulses. The estimated standard errors for Dover sole are also much lower when recruitment is informed by $\hat{s}$. Recruitment deviations are also visibly influenced for Pacific ocean perch and widow rockfish, although the recruitment deviations for these species are noticeably different from $\hat{s}$ for at least a few years. Including the shared index of cohort strength decreases the width of 95% confidence intervals for recent years (2006–2007) even for species that are not strongly influenced in earlier years, e.g. sablefish and canary rockfish (Fig. 4).

Finally, the application to yelloweye and blackgill rockfishes shows the effect of including the shared index of cohort strength on species where recruitment deviations were not previously considered to be parsimonious (Fig. 5). In this case, $\hat{s}$ has an effect on the estimated value for recruitment deviations, and a particularly strong effect for recent years (2000–2007). Including $\hat{s}$ additionally decreases the standard error for estimated recruitment deviations for 2007 by approximately 42% for blackgill (from 0.34 to 0.20) and by 40% for yelloweye (from 0.53 to 0.32). Such a decrease in standard error shows that expectations of recruitment for data-poor species in the same region can be significantly improved using multispecies recruitment indices.

**DISCUSSION**

Environment–recruitment relationships have previously been explored by analyzing stock assessment estimates of recruitment as ‘data’ in a secondary analysis (Wespestad et al. 2000, Mueter et al. 2006). Analyses have also tested presumed environmental relationships by testing an index directly within single-species assessments (Schirripa 2007, Schirripa et al. 2009). However, previous simulation work (Megré et al. 2005, Haltuch & Punt 2011) suggests that estimates of environmental forcing using single-species data, whether estimated internally in assessment models or using stock assessment model output, may be subject to substantial Type 1 (i.e. detecting a relationship when none is present) and Type 2 (failing to detect a present relationship) error.

**Table 3.** Estimates for parameters relating the shared environmental index $\hat{s}$ to recruitment deviations in each single-species assessment model ($q_{i,1}$: difference between average values for $\hat{s}$ and recruitment deviations; $q_{i,2}$: parameter scaling $\hat{s}$ to recruitment deviations; $\hat{\tau}_i$: variability of recruitment deviations around $\hat{s}$ in excess of $\hat{\tau} = 0.2$; $\hat{\tau} + \hat{\tau}_i$: variability of recruitment deviations around $\hat{s}$). See Table 1 for full species names

<table>
<thead>
<tr>
<th>Fish species</th>
<th>$q_{i,1}$</th>
<th>$q_{i,2}$</th>
<th>$\hat{\tau}_i$</th>
<th>$\hat{\tau} + \hat{\tau}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bocaccio</td>
<td>0.191</td>
<td>0.000</td>
<td>0.808</td>
<td>1.008</td>
</tr>
<tr>
<td>Sablefish</td>
<td>0.001</td>
<td>-0.091</td>
<td>0.348</td>
<td>0.548</td>
</tr>
<tr>
<td>Darkblotched</td>
<td>-0.055</td>
<td>0.250</td>
<td>0.749</td>
<td>0.949</td>
</tr>
<tr>
<td>Widow</td>
<td>-0.152</td>
<td>0.538</td>
<td>0.368</td>
<td>0.568</td>
</tr>
<tr>
<td>Petrale sole</td>
<td>-0.113</td>
<td>1.500</td>
<td>0.459</td>
<td>0.659</td>
</tr>
<tr>
<td>Canary rockfish</td>
<td>0.048</td>
<td>1.000</td>
<td>0.309</td>
<td>0.509</td>
</tr>
<tr>
<td>Pacific ocean perch</td>
<td>-0.108</td>
<td>0.429</td>
<td>0.000</td>
<td>0.200</td>
</tr>
<tr>
<td>Dover sole</td>
<td>0.051</td>
<td>1.857</td>
<td>0.000</td>
<td>0.200</td>
</tr>
</tbody>
</table>
These Type 1 and 2 errors occur because recruitment deviations incorporate statistical ‘noise’ from many different processes, and hence environmental time series likely explain only a portion of total variance. Recruitment estimates typically have high autocorrelation due to factors such as ageing imprecision and imprecise resolution of cohorts in length data, and hence the effective degrees of freedom when analyzing a single species can be considerably lower than the number of years being analyzed. However, a multispecies analysis will have substantially higher effective degrees of freedom, due to the independence of recruitment deviation estimates among species. We therefore hypothesize that the use of data from multiple species may improve Type 1 and 2 error rates when screening for environmental relationships within a given fishery management region. This is especially likely given information regarding the degree of autocorrelation for each species, as is available when analyzing observation-level data.

We present the shared index of cohort strength as an improved and statistically-rigorous alternative to previous methods for estimating environment–recruitment relationships. It is estimated without using environmental data, and hence represents an entirely data-driven index of West Coast productivity. Future studies can compare it with other environmental variables in the California Current, e.g. sea surface temperature and the spring transition date which affects West Coast salmon cohort strength (Ryding & Skalski 1999) and winter upwelling indices which affect groundfish individual growth rates (Black 2009, Black et al. 2010). Such a comparison could be used to generate hypotheses regarding the mechanism contributing to correlations in recruitment off the US West Coast, and would potentially add support to the estimated recruitment index if a plausible mechanism could be identified.

This recruitment-linked multispecies assessment model relies upon 2 assumptions: (1) that cohort
strength is similar within a species throughout its range, and (2) that recruitment for groundfish species is affected by a single, shared process. Previous research suggests that rockfish cohort strength for a given species is generally synchronized coastwide rather than varying regionally (Field & Ralston 2005), which supports the first assumption. Rockfishes and flatfishes exhibit weak but significant correlations in

![Fig. 4. Estimates of recruitment deviations (circles) and their 95% confidence intervals (bars) without (blue) and with (red) the shared index of cohort strength $\hat{s}$ for each fish species and year (1 = bocaccio; 2 = sablefish; 3 = darkblotched rockfish; 4 = widow rockfish; 5 = petrale sole; 6 = canary rockfish; 7 = Pacific ocean perch; 8 = Dover sole).]
cohort strength in the nearby Bering Sea but not the Gulf of Alaska, and correlations may be stronger when restricting analysis purely to rockfishes (Mueter et al. 2007), which provides mixed support for the second assumption. Model estimates indicate that the index has little explanatory power for bocaccio or darkblotched rockfishes ($\hat{\tau} \approx 1$), but that all other species are influenced at least partially by the index.

The shared index of cohort strength could also be used in future stock assessments for groundfishes off the US West Coast as an index of recruitment deviations. The impact of such an index will vary greatly from species to species, depending on how informative the data are regarding recruitment for a given species, and whether the species is estimated to have a strong or weak correlation with the shared recruitment factor. This recruitment time series could be particularly helpful within data-poor assessment methods, which often do not account for recruitment variability (Dick & MacCall 2011). Improved estimates of recent recruitment could also impact rebuilding plans for data-rich species by improving estimates of the strength of recent cohorts that may play a significant role in the rebuilding analysis forecasts.

This analysis could be easily replicated in other regions that have stock assessments using Stock Synthesis, or other integrated assessment methods that penalize estimated recruitment deviations towards a recruitment index. Other regions with several Stock Synthesis assessments include (among others) the Bering Sea, Gulf of Alaska, Eastern Pacific Ocean, and southeast Australia. The general applicability of the methodology supports the suggestion of Thorson et al. (2012) to develop a worldwide database of Stock Synthesis assessments. This database would allow analysis of observation-level fishery data, and could complement existing global databases such as the RAM Legacy Stock Assessment Database (Ricard et al. 2012) and the FAO landings database (FAO 2010).

However, we recommend that simulation modeling be used to explore the properties of the proposed estimator for correlations in recruitment deviations. Simulation could quantify Type 1 (false positive) and Type 2 (false negative) error rates when identifying significant correlations in recruitment. Simulation could also be also used to assess sensitivity to violated model assumptions, i.e. using a shared recruitment factor for a species that in actuality has radically different recruitment. It is not immediately obvious what effect incorrectly specifying $\hat{s}$ would have on the risk of overfishing, lost yield, or other common measures of fishery management performance, and this could be quantified using management strategy evaluation (Sainsbury et al. 2000).

Acknowledgements. J. Thorson acknowledges financial support from the National Research Council Research Associate Program with supervision by E. Holmes, E. Ward, and M. Scheuerell at the NMFS Northwest Fisheries Science Center (NWFSC), as well as support from the NWFSC Fisheries Resource and Monitoring Division. We also thank the many research scientists and assessment authors who collected, compiled and analyzed the data in the stock assessments used herein, and S. Miller for helping to compile and distribute these and other Stock Synthesis assessments. Finally, we thank J. Field, J. Hastie, M. McClure, 3 anonymous reviewers, and the associate editor J. Rice for many helpful suggestions and editorial guidance.
LITERATURE CITED

- Field J (2011) Status of bocaccio, Sebastes paucispinis, in the Conception, Monterey and Eureka INPFC areas as evaluated for 2011. National Marine Fisheries Service, Southwest Fisheries Science Center, Santa Cruz, CA
- Field JC, Pearson D (2011) Status of blackgill rockfish, Sebastes melanostomus, in the Conception and Monterey INPFC areas for 2011. National Marine Fisheries Service, Southwest Fisheries Science Center, Santa Cruz, CA
- Scheuermann MD, Zabel RW, Sandford BP (2009) Relating juvenile migration timing and survival to adulthood in two species of threatened Pacific salmon (Oncorhynchus
Appendix 1. Tuning input variance for the shared index of cohort strength

In the recruitment-linked multispecies assessment model, the shared index $s$ accounts for correlated recruitment deviations among species. The vector $\hat{s}$ is included in each single-species model as a penalty on single species recruitment deviations (Schirripa et al. 2009):

$$L(\hat{s}, \beta, x) = \text{Normal}\left[\hat{s}_i | \hat{q}_{i,1}^2 + (1 + \hat{q}_{i,2})\tau_i, (\tau + \tau_i)^2\right]$$  

(A1)

where $L(\hat{s}, \beta, x)$ is the likelihood term that penalizes recruitment deviations towards the shared index of recruitment deviations, $\text{Normal}(\hat{s}_i | \hat{q}_{i,1}^2 + (1 + \hat{q}_{i,2})\tau_i, (\tau + \tau_i)^2)$ is the probability density for $s_i$ from a normal distribution with mean $\hat{q}_{i,1}^2 + (1 + \hat{q}_{i,2})\tau_i$ and variance $(\tau + \tau_i)^2$. $\tau_i$ is an estimated deviation from median recruitment given spawning biomass for species $i$ during year $t$. $\hat{q}_{i,1}$ and $\hat{q}_{i,2}$ are estimated parameters that relate the model estimates of recruitment deviations $\hat{r}_{it}$ for each species $i$ to the shared index of recruitment deviations $\hat{s}_i$. $\tau$ is the average of the standard errors of the elements of the vector $\hat{s}$ and $\tau_i$ is a parameter representing the variability in recruitment deviations around $\hat{s}$ for species $i$. The variance term in Eq. A1 $(\tau + \tau_i)^2$ ensures that the variance of the likelihood penalty for recruitment deviations is greater than or equal to $\tau^2$, i.e. and hence that the degree of shrinkage of recruitment deviations towards $\hat{s}$ is never greater than the inherent estimation error associated with $\hat{s}$. Including $\tau$ in the likelihood penalty also ensures that the model is consistent in its weighting of data, i.e. that each element of the assessment likelihood is weighted according to the variation in its measurement errors ($\tau$) as well as a potential inflation due to process errors ($\tau_i$).

Exploratory analysis shows that an increase or decrease in the value of $\tau$ has little impact on species for which $\hat{r}_i >> 0$ (i.e. sablefish, boccaccio, darkblotched, widow, and petrale) because a change in $\tau$ for these species is exactly offset by changes in $\hat{r}_i$ to maintain a constant value of $\tau + \hat{r}_i$. By contrast, an increase in the value of $\tau$ has a large impact on the results for species for which $\hat{r}_i \approx 0$, because the model is unable to decrease $\hat{r}_i$ to offset this change, and it therefore causes a change in the likelihood penalty variance $(\tau + \hat{r}_i)^2$. For species where $\hat{r}_i \approx 0$, a decrease in $\tau$ results in tighter confidence intervals for $\hat{r}_i$ and an increase in $\tau$ results in wider confidence intervals for $\hat{r}_i$. The approach used to set $\tau$ shares several features with a mixed-effects estimator. First, $\tau$ will have a large estimated value when the $\hat{r}_i$ are poorly estimated for all species. In this case, $\hat{s}$ will have little effect on $\hat{r}_i$, which is a property of mixed-effects models (Gelman & Hill 2007). Second, a precise estimate of $\hat{s}$ will result in a low value for $\tau$, which in turn will allow $\hat{s}$ to have a large effect on $\hat{r}_i$. This is an intuitive outcome, given that a strong common trend in a measure of recruitment success should have a strong effect on recruitment deviations in this case for each species.