# The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models.

Maunder, M.N.[1,2], Thorson, J.T.[3], Xu, H.,[1], Oliveros-Ramos, R.[1], Hoyle, S.D.[4], Tremblay-Boyer, L.[5], Lee, H.H.,[6], Kai, M.[7], Chang, S.K.[8], Kitakado, T.[9], Albertsen, C.M.[10], Minte-Vera, C.V.[1], Lennert-Cody, C.E. [1], Aires-da-Silva, A.M. [1], and Piner, K.R.[6]

1: Inter-American Tropical Tuna Commission, 8901 La Jolla, Shores Dr., La Jolla, CA, 92037-1508, USA
2: Center for the Advancement of Population Assessment, Methodology, La Jolla, CA, USA
3: Habitat and Ecosystem Process Research program, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA
4: National Institute of Water and Atmospheric Research Ltd (NIWA), 217 Akersten Street, Port Nelson, New Zealand 7011
5: Dragonfly Data Science, 158 Victoria St, Wellington, New Zealand 6011
6: Southwest Fisheries Science Center, National Marine Fisheries Service (NMFS), NOAA, La Jolla, CA, U.S.A.
7: National Research Institute of Far Seas, Fisheries (NRIFSF), Japan Fisheries, Research and Education Agency, Shimizu, Shizuoka, Japan
8: Institute of Marine Affairs, National Sun Yat-sen University, Kaohsiung, Taiwan
9: Tokyo University of Marine Science and Technology, 5-7, Konan 4, Minato-ku, Tokyo 108-8477 Japan
10: National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet 201, DK-2800 Kgs. Lyngby, Denmark

**Abstract**

We describe and illustrate a spatio-temporal modelling approach for analyzing age- or size-specific catch-per-unit-effort (CPUE) data to develop indices of relative abundance and associated composition data. The approach is based on three concepts: 1) composition data that are used to determine the component of the population represented by the index should be weighted by CPUE (abundance) while the composition data used to represent the fish removed from the stock should be weighted by catch; 2) due to spatial non-randomness in fishing effort and fish distribution, the index, index composition, and catch composition, should be calculated at a fine spatial scale (e.g., 1°x1°) and summed using area weighting; and 3) fine-scale spatial stratification will likely result in under-sampled and unsampled cells and some form of smoothing method needs to be applied to inform these cells. We illustrate the concepts by applying them to yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean.

**Key Words:** catch-per-unit-effort, CPUE, spatio-temporal model, index of abundance, catch-at-age, length composition

## 1. Introduction

Fisheries stock assessment is the gold standard for providing management advice. Age- or size-structured population dynamics models are fit to multiple data sets to estimate model parameters and associated derived management quantities. The main data types, other than catch, are indices of relative abundance and composition data representing the proportions of the sampled population within different age, length, sex, and/or weight categories. The indices provide information on trends in abundance. The composition data provide information on the component of the population represented by the index, and the size or age of the fish removed by the fishery. They both provide information on absolute abundance (Maunder and Piner, 2015). Therefore, it is essential that the indices of relative abundance and composition data are analyzed appropriately to ensure they are as precise and accurate as possible.

Preferably, indices of abundance are based on well-designed surveys, are proportional to abundance, and are precise. Unfortunately, surveys are not possible for many stocks due to logistical and funding limitations. Therefore, many stock assessments, such as those conducted for tunas worldwide, rely on indices of relative abundance based on fishery catch-per-unit-of-effort (CPUE) data. These fishery-dependent indices are influenced by several factors that may invalidate the assumption that the index is proportional to abundance (Harley et al., 2001; Maunder et al., 2006a; Thorson et al., 2017c). Of particular concern is that fishing effort is not randomly or systematically distributed over the whole stock area, and is rather likely to be concentrated where fish are abundant. Therefore, expanding indices from sampled to under-sampled or unsampled areas may lead to positive bias. Similar issues also apply to the composition data, but are typically not addressed.

There is a large body of literature describing alternative CPUE "standardization" approaches to minimize the influence of factors other than abundance on the final index (Maunder and Punt, 2004). Typically, the CPUE is standardized using a Generalized Linear Model (GLM), or a similar method, to account for factors that impact CPUE (e.g., vessel and gear characteristics, season, location, and environmental conditions). There are also numerous examples of more sophisticated approaches to deal with specific issues or fine-tune a component of the standardization method. For example, Hinton and Nakano (1996) used a mechanistic model to match the three-dimensional spatial distribution (latitude, longitude, and depth) of fishing effort, with environmental conditions and fish habitat preference to standardize CPUE for blue marlin. Many authors have focused on the fine scale spatial distribution of CPUE (e.g., Walters, 2003; Carruthers et al., 2011; Thorson et al., 2017c), while others have used broader scale spatial strata to standardize CPUE (e.g., Punt, et al., 2001; Gruss et al., 2019).

Historical approaches to deal with spatial variation in CPUE commonly used a simple GLM that included location as a factor without an additional term for time-space interaction. This approach implicitly assumes that the estimated year effect (i.e., the temporal trend), which is assumed to be a proxy of relative abundance, is the same in each spatial stratum, and that only the average CPUE differs among strata. The assumptions underlying this model can lead to bias in the estimated index of relative abundance in several situations, including when the spatial distribution of the stock changes over time (e.g., Punt et al., 2000b). Such a situation can typically be identified when the interaction term between spatial stratum and year is statistically significant, or the time series of year effects from different strata show different trends.

In general, statistically significant interaction terms between year and another categorical variable that result in meaningful differences in standardized trends are problematic. Calculating the index requires choosing a level for the variable interacting with year, and thus the estimated trend will depend on the chosen level (Maunder and Punt, 2004). If the interaction is treated as a random effect, calculating the index requires specifying the average value for the variable that is treated as random. When the variable interacting with year is a spatial factor, a more appropriate approach may be to use "area-weighting", where the index is calculated as the weighted sum of model predictions over the levels of the spatial factor. The weights are equal to the spatial area associated with each factor. However, this approach assumes that the sampling represents all locations in a spatial stratum, including poorly-sampled locations, which is unlikely to be even approximately true for large spatial strata. Spatio-temporal modeling methods, which can use information on CPUE from neighboring locations to improve estimation of the spatial effects throughout the area occupied by the stock, can be based on finer spatial strata and should lead to improved indices compared to those derived from simple area-weighting. In cases of large differences in the year effect among spatial strata, the stock may be modelled as multiple independent or interacting populations, and each population assessed based on its respective stratum's index of relative abundance (Punt, 2019).

CPUE-based indices of relative abundance used in stock assessment models are representative of the population component caught in the fishery, rather than the entire population. Typically, this issue is addressed by using an age- or size-based selectivity curve that is estimated by fitting to fishery composition data. In most cases, the selectivity curve is used to characterize both the catch and the index of abundance. Naively, this makes sense, since both catch and the index of abundance are derived from the same fishery (e.g. gear). However, selectivity in the stock assessment model does not simply represent contact selectivity (e.g., a fish being trapped in a gillnet as it tries to pass through), but also availability, which can be a consequence of the spatial structure of the fleet relative to the stock, and is likely to change over time (Sampson, 2014; Waterhouse et al., 2014). The index is used in the assessment model to represent changes in abundance while the catch represents mortality due to fishery removals. Catch is not necessarily distributed spatially in proportion to abundance, and the "selectivity" in the stock assessment will differ between the index and the catch when the composition data differ systematically among spatial strata, and catch distribution among strata changes through time. In general, the index selectivity should represent the total vulnerable (i.e. filtered through the gear selectivity) abundance across the domain of the index, while the catch selectivity will represent the vulnerable abundance available to the fishery adjusted as the fishery spatial distribution changes over time. This concept is advantageous because 'index selectivity' will be unaffected by fleet movements so may be relatively stationary over time (assuming no major changes in the gear that are not accounted for), while 'removals selectivity' does not need to be stationary because, in the ideal case when catches are well characterized, harvest by age-and-year can just be removed exactly (i.e., it does not need to be assigned a likelihood component). In practice, approximations are typically used by modelling temporal and/or spatial variability in removals selectivity and fitting to the composition data.

The composition data should be calculated differently for the index of abundance and the catch. These data should be calculated by spatial stratum and summed as in the construction of a CPUE- based index of relative abundance. However, the composition data for the index should be weighted by the product of the CPUE by stratum and the areas of the spatial strata, while the catch composition data should simply be weighted by the catch for the spatial strata.

Here we discuss the use of spatial-temporal models to deal with changes in the spatial distribution of the fishery and the stock, and to standardize size composition data. We then discuss how this applies to yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean

(EPO) to further illustrate the approach. We apply a spatiotemporal delta model (Thorson and Barnett 2017) to standardize the catch-per-day-fished and length-composition data from the purse seine fishery on yellowfin tuna associated with dolphins, and evaluate its influence on the stock assessment compared to conventional CPUE indices.

## 2. Eastern Pacific Ocean yellowfin tuna application

We illustrate the impact of area weighting of CPUE and length-composition data using data for yellowfin tuna caught purse-seine sets associated with dolphins in in the eastern Pacific Ocean during 1975 - 2016. The CPUE-based indices of abundance and length-composition data are then used in a stock assessment model to determine their impact on the assessment results. The data are divided into three fisheries based on spatial strata (Fig. 1) to account for possible differences in selectivity and catchability. We only used CPUE and associated composition data for large (class-6) purse-seine vessels that made at least 75% of their sets on tunas associated with dolphins (Fig. 2). Length-composition data are grouped into 10cm bins from 20 to 200 cm.

An integrated age-structured stock assessment model fit to CPUE-based indices of abundance and length-composition data developed in Stock Synthesis V3.23b (Methot and Wetzel, 2013) is used to assess yellowfin tuna in the EPO (Minte-Vera et al., 2019). The full specification of the assessment can be found in Aires-da-Silva and Maunder (2012) and Table 1. The model operates on a quarterly time step, so the index of abundance and length composition data are calculated by quarter. Natural mortality is age and sex-specific, with higher natural mortality for females than males starting from 30-month-old and higher for juveniles. Growth follows a Richards' curve. Separate dome-shaped selectivity curves are estimated for the majority of fisheries. Selectivity for the southern longline fishery is assumed to be asymptotic. Maximum likelihood techniques are used to estimate the population scaling parameter (virgin recruitment, $R_0$), lognormal recruitment deviates for each quarter ($sd = 0.6$), parameters to construct the initial numbers at age in 1975, and selectivity parameters. Recruitment is assumed to be independent of stock size. Fisheries are defined as combinations of gear used (longline or purse seine), set type (for purse seiners) and area of operation. The purse seine sets are of three types: sets associated with dolphins, sets associated with floating objects, and sets on free-swimming schools. The indices of abundance are fit assuming a lognormal likelihood function and the length-composition data are fit assuming a multinomial likelihood function. The current analysis differs from Minte-Vera et al. (2019) by estimating a change in catchability and selectivity for the southern longline fisheries and their related indices of abundance since 2010, as well as by not using the indices of abundance from the purse seine fisheries for free-swimming schools. These changes are considered improvements to the stock assessment and also allow the data from the purse seine fisheries on yellowfin associated with dolphins to have more impact on the assessment results.

The specific analyses conducted are described in detail below under the corresponding sections.

## 3. Dealing with spatial data

### *3.1 Spatial weighting*

Addressing changes in the spatial distribution of the fishery and/or stock when developing indices of relative abundance should be an important component of CPUE standardization. Area-weighting can be applied to avoid bias due to temporal variation in the spatial distribution of the statistical weights if the spatial distribution of the fishery has changed substantially (Punsley, 1987; Campbell, 2004). Area-weighting for nominal CPUE is calculated as follows

$$I_t = \sum_s \frac{c_{s,t}}{f_{s,t}} \frac{A_s}{\sum_k A_k} \quad (1)$$

where $I_t$ is the index for time $t$, $c_{s,t}$ is the catch in spatial stratum (or station) $s$ during time $t$, $f_{s,t}$ is the effort in spatial stratum $s$ during time $t$, and $A_s$ is the area of spatial stratum $s$. In many cases the area might be assumed equal and left out of equation 1.

Area weighting contrasts with commonly used approaches to standardize CPUE data such as naive use of GLMs, which when applied to the entire fishery region, can be considered data-weighted in the sense that each data point is implicitly given equal weight in the log-likelihood of the standardization model, independent of the spatial stratum to which it belongs. Data-weighting for nominal CPUE is calculated as follows

$$I_t = \frac{1}{n} \sum_{i \in t} \frac{c_i}{f_i} \quad (2)$$

where $n$ is the number of observations, $c_i$ is the catch for observation $i$, and $f_i$ is the effort for sample $i$. Area weighting and data weighting both differ from the simple ratio of total catch to total effort, which is weighted by effort. Effort-weighting is calculated as follows

$$I_t = \frac{\sum_{i \in t} c_i}{\sum_{i \in t} f_i} \quad (3)$$

Spatial strata that have more data will be given more weight in the analysis as a result of using data-weighting. By contrast, area-weighting will adjust the total statistical weights within each time-area stratum to be proportional to the area, by adjusting the relative weights of individual data points. This can cause large strata with small sample sizes to overwhelm small strata with better coverage, and to have similar influence as large strata with large sample sizes. Abundance estimates from large strata with small sample sizes may have high variance and thus increase the uncertainty of estimates of total abundance. Nevertheless, this increased uncertainty might be representative of true knowledge about population density when data are not available for large segments of the population's range (e.g., Walters, 2003).

When substantial changes in the spatial distribution of the stock have occurred, the abundance trends in each stratum should ideally be calculated and combined in some manner to obtain a representative time series of overall abundance estimates. In this instance, the handling of strata with missing data becomes a key issue (Walters 2003; Punt et al., 2000a; Carruthers et al 2011; McKechnie et al., 2013). We address this issue here using a spatio-temporal model to impute density even for strata with missing data, and to propagate the increase in variance with the resulting predictions into the overall index of abundance.

3.1.1 YFT application

Spatial weighting is investigated using four approaches for the purse seine fishery on EPO yellowfin associated with dolphins:

1) Nominal index (effort weighted): The index is calculated as the total catch divided by the total effort in each year-quarter.
2) GLM-1 (data weighted): The index is calculated using a delta-lognormal GLM with year-quarter as a categorical variable.
3) GLM-2 3 spatial strata: The same as 'GLM-1', but with spatial stratum as a categorical variable. The strata are the three dolphin-associated fisheries included in the current assessment model (Fig. 1). These areas have been used in the assessment since its inception and are based on spatial differences in length-composition data. More spatial

strata might be used in a more rigorous GLM analysis, but we use the three previously-defined spatial strata for consistency with previous analyses.

4) GLM-3 3 spatial strata year-quarter and stratum interaction: The same as 'GLM-2 3 spatial strata', but with a year-quarter and stratum interaction term (implemented as separate GLMs for each stratum). The index is calculated by summing the year-quarter terms for each separate GLM weighted by the area of each stratum for each year-quarter.

The GLM index of abundance (#2) was found to be similar to the nominal index (#1; Fig. 3a). The index for the southern stratum is different from those of the north and coastal strata, showing more extreme fluctuations (Fig. 3b). This makes the area-weighted index (#4) the most different from the nominal index out of the three GLM approaches both in terms of scale and variability (Fig. 3a). However, the differences in the indices for the whole EPO are near-trivial due to the low CPUE (abundance) of yellowfin in the southern stratum.

*3.2 Spatio-temporal modeling*

A method is needed to improve the estimates for spatial strata with low sample sizes and to impute abundance for spatial strata within the stock distribution for which no data exist for one or more time periods. This is particularly the case if the spatial stratification is based on a fine scale. Contemporary spatio-temporal models are useful for this purpose and have been made practical by recent developments in statistical methodology, computational algorithms, and software packages (e.g., Kristensen et al., 2016). These models are based on the assumption that catch rates in nearby locations should be similar, but that the degree of similarity should decrease as the distance between locations increases. This relationship is called the 'covariance function' and the rate at which correlation decreases with distance is referred to as the 'decorrelation rate'. Spatio-temporal models can also be configured to share information among periods close in time (e.g., using temporal autocorrelation, Thorson et al., 2016). Spatio-temporal models estimate the degree of information-sharing between neighboring points by estimating the shape of the covariance function from a specified family, which then allows the model to incorporate either strong or weak smoothing for predictions that are close in space and time. This can improve estimates for locations and times with low sample size, including for unsampled strata. Research is ongoing regarding additional computational improvements, e.g., for nonstationary correlation functions, which would allow decorrelation rates to differ among stock habitats.

In common with any other analysis method, several terms can be included in the model and some form of model selection used to determine the "best" model. Of particular interest are models with just spatial effects and those with spatial-temporal effects. Including only spatial effects is appropriate when the spatial distribution of the stock does not change over time. For example, when physical habitat (e.g., rocks, kelp, sea grass) determines the spatial distribution of the stock. Spatial-temporal effects are appropriate when the factors determining the distribution of the stock change over time; for example, when unmeasured oceanographic variables (e.g., temperature, chlorophyll), which change over time, determine the spatial distribution of the stock. In either case, model selection tools should be used to evaluate the alternative models.

Generalized linear mixed models with spatio-temporal effects are now seeing broad usage in analysis of spatial data (e.g., Lewy and Kristensen, 2009; Kristensen et al., 2014; Nielsen et al., 2014; Thorson et al., 2015a). Kai et al. (2017a) presented a spatio-temporal model for shark CPUE and much of the following comes from their description (other examples are provided in Table 2). Space and time are modeled as main effects with an additional term for the interaction between space and time. The random effects are integrated out during statistical inference. The spatial components are implemented using a Gaussian random field (GRF), which is a computationally efficient approach for implementing multi-dimensional smoothers (Thorson et al., 2015a). The spatial-temporal interaction term can be modeled in a computationally efficient manner by using a GRF for each time period so that the spatial-temporal component distribution is uncorrelated over time or using a first-order autoregressive process to include temporal correlation. Seasonal spatial effects are also often modelled (e.g. Kai et al., 2017a). Seasonal models could be developed by including a spatial term for each season and a spatio-temporal term for each season-year combination. We recommend further research regarding seasonal models but do not discuss these in detail here.

The spatio-temporal model estimates the density of individuals, $d$ ($s$, $t$), for each stratum (station) $s$ (latitude and longitude) and time $t$ as:

$$\log\big(d(s,t)\big) = d_0(t) + \gamma(s) + \theta(s,t) + \sum_{j=1}^{n_j} \beta_j x_j(s,t), \quad (4)$$

where $d_0(t)$ represents a temporal main effect, $\gamma(s)$ represents the spatial component, $\theta(s,t)$ represents the spatio-temporal interaction term, and $\beta_j$ represents the impact of covariate $j$ with value $x_j(s,t)$ on density at stratum $s$ and time $t$.

Spatial variation $\gamma(s)$ is modeled using a GRF, which reduces to a multivariate normal distribution when evaluated at a finite set of strata (Thorson et al., 2015b). The Matérn correlation function is used for computational efficiency (Diggle and Ribeiro, 2007; Roa-Ureta and Niklitschek, 2007; Lindgren et al., 2011). Computational efficiency is often improved by adapting a "predictive process" framework where spatial/spatio-temporal variation is only modeled between a small number of locations (termed "knots", which together form a "mesh") and variables are then interpolated between these knots. However, ongoing research is needed to evaluate this "predictive process" framework including: (1) whether the number of knots impacts results; and (2) how best to determine the mesh configuration. For example, the R package VAST by default distributes knots proportionally to the density of available data (Thorson, 2019a), which results in poorly sampled areas receiving fewer knots, which then impacts the resolution of the spatial imputation.

Expected catch, $c_i^*$, which is used to fit to the observed catch during the parameter estimation process using a likelihood function (e.g., log-normal, negative-binomial, or a zero-inflated model), is the product of relative fish density, as represented by the spatio-temporal model, and fishing effort $f_i$, $c_i^* = d(s_i, t_i) f_i$, for the $i$-th observation, at stratum $s_i$ and time $t_i$. Covariates $\acute{x}_{k,i}$ for each data point $i$ and covariate $k$ can be added to model catchability (e.g., gear effects) $\log(c_i^*) = \log\big(d(s_i, t_i)\big) + \log(f_i) + \sum_{k=1}^{n_k} \acute{\beta}_k \acute{x}_{k,i}$.

The parameters are estimated by maximizing the likelihood function while integrating across the random effects representing spatial and spatio-temporal variation using Template Model Builder (TMB). TMB is an R package (R Development Core Team, 2013) that efficiently fits latent variable models to data (https://www.github.com/kaskr/adcomp; Kristensen et al., 2016), through the use of the Laplace approximation for integration, and automatic differentiation for calculating derivatives. The estimated fixed effects parameters include those representing the temporal main effects (e.g., coefficients associated with a categorical variable for year), the covariance structure associated with the spatial component, the spatial-temporal interaction (the variance of the first-order autoregressive model and the covariance structure of the GRF), the coefficients associated with density covariates, and the catchability covariate coefficients. For a discussion of catchability and density covariates, see Thorson (2019a).

The index of relative abundance for a particular time period is calculated by summing the predicted densities (or the product of predicted density and area, if area differs among strata) for each location in that time period (Thorson, 2019a). Care needs to be taken to identify which factors affect catchability ($\acute{x}$) and should not be used to estimate density, but are used to calculate the expected catch used in the likelihood function. Covariates that effect density ($x$) are used to calculate the quantities that are summed to generate the index of relative abundance. When random effects are used to model abundance on the log-scale, TMB will report the median instead of the mean of the resulting distribution on the natural scale. Therefore, a bias-correction algorithm to account for retransformation bias when predicting and visualizing total abundance and size composition (Thorson and Kristensen, 2016) should be used where appropriate (Thorson, 2019b).

3.2.1 YFT application

A VAST implementation of a spatio-temporal delta model is used to standardize the CPUE data from the dolphin-associated yellowfin fishery (Xu et al., 2019a). The resulting index of relative abundance is compared with those obtained from the raw data and the GLM standardizations.

The spatio-temporal model is a delta-lognormal Generalized Linear Mixed Model (GLMM) with a time-invariant spatial variation component and a time-varying spatio-temporal component The temporal main effect is modelled as a separate intercept for both components of the delta-model for each season-year combination (season-year is a categorical variable). The spatio-temporal component is independent across years (i.e., the optional autoregressive process with a one time-step lag (AR1) is not included) due to computational limitations. No other covariates are included in the model. The whole EPO is modelled simultaneously, and indices of abundance are extracted for each spatial stratum (Fig. 1). The data used in the model are aggregated at the year-quarter level and by 1° by 1° stratum, with two hundred knots distributed over the domain proportionally to effort in days fished. Two sets of indices are developed from the same model:

1) an index for each of the three fisheries used in the assessment (Fig. 1); and
2) an overall index for the whole EPO.

The spatio-temporal model produces an index for the whole EPO similar to that produced using the raw data (Fig. 3c). The index from the spatio-temporal model shows larger fluctuations and is higher than the GLM index for later years. The three fishery-specific indices of abundance from the spatio-temporal model are similar to each other (Fig. 3d). The spatio-temporal model-based index for the southern stratum fluctuates less than the GLM-based index (compare Fig. 3).

There are substantial differences in the spatial distribution of effort (Fig. 2b) and the predicted CPUE (Fig. 2a), thus the implied yellowfin densities, among years. The 4th quarter of 1998 was an El Nino and had higher CPUE in the coastal areas (Fig. 2ai) compared to the 4th quarter of 2003 (Fig. 2aiii), which was neutral. The 4th quarter of 1998 was a La Nina and had more restricted effort distribution but predicted CPUE was high towards the west and to the south (Fig. 2aii) where there was no effort (Fig2bii).

*3.3 Composition data*

Analysis of composition data (e.g., age, length, or weight composition) is a key component of developing CPUE-based indices of abundance in stock assessments. The composition data provide information on the portion of the population represented by the index with respect to age or size.

3.3.1 Simple assembly

Composition data are typically used in their raw form by summing all the samples in size or age bins, each sample possibly weighted by the corresponding catch. Simply summing the composition data implicitly assumes that each sample represents a random draw from the population (e.g., fish sizes are randomly distributed throughout the whole area). In practice, this is unlikely to be the case because many species exhibit spatial heterogeneity in age or size. Reweighting samples by the associated catch or expanding sampled data to the total catch by stratum (e.g., gear, month, 1° by 1° square) made sense in the context of earlier assessment methods such as virtual population analysis, where the composition data are directly used to inform the age or size distribution of the total catch removed by the fishery. However, catch-at-age methods, which predict catch composition based on selectivity and population structure, require different methods for preparation of composition data, depending on whether it is meant to represent the catch, or to represent the index of abundance.

### 3.3.2 Spatio-temporal modelling of composition data

The traditional process of reweighting composition data can be interpreted as one step towards a model-based framework for "standardizing" composition data (Thorson, 2014; Thorson and Haltuch, 2018). Standardizing composition data using model-based methods has several potential benefits including:

1) accounting for confounding factors (e.g., vessel type, gear configuration or season) when using composition samples to estimate proportions for each category (e.g., size, area, or age);
2) using auxiliary information to improve predictions of age/size/sex composition in spatial strata with low samples sizes, e.g., by basing predictions upon estimated environmental relationships, persistent spatial or temporal patterns, or alternative sources of information such as tags or fishery CPUE; and
3) calculating the multinomial sample size for estimated composition data based on the variance in composition sampling data, where this sample size is then used as a starting point (or ceiling) for the weight that these data should receive in an assessment model (Thorson and Haltuch, 2018) - although there are approaches to calculate the sample size for raw data (e.g., bootstrapping), they are seldom used.

The spatio-temporal modelling approach described in previous sections can be modified to include composition information (Kristensen et al., 2014; Nielsen et al., 2014; Thorson et al., 2019c). For example, an independent model could be applied to each age- or size- group separately to create indices of relative abundance for each group for use in the stock assessment model. Using independent indices of abundance for each age class was common practice when tuning virtual population analysis, and is commonly used in state-space age-structured models (e.g., the base-model in Nielsen and Berg, 2014). However, estimating a separate index of abundance for each age/size category ignores the correlation among age-classes specified within the stock assessment model. In addition, it does not take full advantage of the data because it ignores the fact that similar ages or sizes likely have similar catch rates. In many cases, there may not be sufficient data by size or age bin, especially if modelling length or weight, such that if age/size groups are assumed independent, age or size bins may have to be combined, which could dampen signals about important population or fishery processes (e.g., recruitment or selectivity).

The three dimensions of the spatio-temporal model (time, latitude, and longitude) need to be modified to include a fourth dimension of either age or size to incorporate composition information in a spatio-temporal model (Lewy and Kristensen, 2009; Kristensen et al., 2009; Nielsen et al., 2014; Kai et al., 2017b; Thorson et al., 2017a). One complication with our recommended approach is that its implementation is computationally intensive. Also, because composition data are usually not collected for all catch events, the catch and composition data may need to be fit using separate likelihood functions in the size-composition standardization, but simultaneously in the same spatio-temporal standardization model.

Of note, a growth and survival model could also be used to inform the standardization of size composition data (e.g., Kristensen et al., 2014). However, this approach is not considered here as the aim of the composition data standardization is to create data for use in an age-structured stock assessment, where the resulting estimates of proportion-at-size are subsequently fitted based on an assumed or estimated growth function. Accounting for growth in the standardization would result in the growth information being used twice, and compromise variance estimates in the assessment model.

One concrete example of our proposed approach was implemented by Kai et al. (2017b) who developed a spatio-temporal model that also included the size of the fish caught (see Table 3 for other examples). The following comes from their description.

The spatio-temporal model incorporating size data estimates the density:

$$\log\big(d(s,t,l)\big) = d_0(t) + \gamma(s) + \tau(l) + \theta(s,t,l) + \sum_{j=1}^{n_j} \beta_j x_j(s,t,l) \tag{5}$$

where $\tau(l)$ represents the impact of size (length) on expected catch rates, $\theta(s,t,l)$ represents an interaction term of stratum, time and size, and each covariate $j$ can be a function of size, expressed as $x_j(s,t,l)$. The marginal (common to all strata and times) size effect, $\tau(l)$, is modeled using a first-order autoregressive process (AR1) leading to a semi-parametric representation of the expected density at each size bin (Thorson et al., 2014). Covariates could also be included to model catchability as described above. Expected catch $c_i^*$ is the product of density and fishing effort $f_i$, $c_i^* = d(s_i,t_i,l_i)f_i$, where density is a function of size, and is fitted to the observed catch $c_i$ for the $i$-th observation, which is at stratum $s_i$, year $t_i$, and size $l_i$. The spatio-temporal-at-size variation, $\theta(s,t,l)$, is modeled by combining the GRF for spatial variation with a first-order autoregressive processes (AR1) for temporal and for size variation. However, more complicated models for covariation among sizes could be explored. For example, different cohorts often partition habitats spatially such that different sizes/ages may be negatively correlated, and negative correlations are not approximated well using the AR1 process used in Kai et al. (2017b). In these cases, researchers could instead explore a factor model for size/age covariance (e.g., Thorson et al., 2017a), where a full-rank or rank-reduced covariance is estimated.

### 3.3.3 Yellowfin application

A VAST implementation of a spatiotemporal delta model is used to standardize the length-composition data from the dolphin-associated yellowfin fishery. The model is an extension of that used for the CPUE described in the previous sections. The lengths are grouped into 10 cm bins for computational efficiency, and preliminary stock assessment results (not shown here) show that moving from 2cm to 10cm length bins had a negligible impact on results. Interactions in the model include 1) length bin and time, 2) length bin and space, and 3) length bin, and time and space . We specify that spatio-temporal variation and intercepts are independent for each combination of year, quarter, and length bin (i.e., not using the AR1 components) to minimize estimation covariance among bins (because the resulting estimated covariance matrix is not typically provided to the stock-assessment model). The resulting length-compositions are compared with those obtained from the raw data simply by weighting each sample by the number of sampling events (the number of wells, which are the storage compartments for the fish on board the vessel). We include a separate intercept for each component of the delta-model for each season-year combination (i.e., season-year is a categorical variable). No covariates are included in the model. The whole EPO is modelled simultaneously, and composition data are extracted for each spatial stratum. The data used in the model are aggregated at a year, quarter and 5° x 5° level to match the resolution of the length-composition data. Twenty knots were used for the spatio-temporal model for size, as the increased model complexity prevented using the higher mesh resolution of the CPUE model. Also, the dataset covers 70 unique 5° x 5° cells during 1975-2016, and on average less than 10 unique 5° x 5° cells in each quarter, so 20 knots balances estimation accuracy and computation efficiency. With this lower number of knots, the model still took more than a day to provide results in a 6-CPU parallel R environment.

Two sets of length-composition data are developed from the same model and each spatial stratum (the sizes of spatial stratum are assumed to be equal) is weighted by CPUE and by catch (see the next section for details and rationale for using catch):

1) composition for each of the three fisheries used in the assessment (Fig. 1); and
2) composition for the whole EPO.

The size-composition data estimated by the spatio-temporal model using either catch weighting or CPUE weighting are similar to the nominal size compositions for all three fisheries (Figs 4-6) and the EPO as a whole (Fig. 7). However, there are some differences in the size of the fish in the composition data, particularly between the nominal length-composition and the two spatio-temporal model-based composition estimates. In general, the two types of spatio-temporal model-based length compositions are more similar to each other than they are to the nominal compositions, but there are also instances where the catch weighted and the CPUE weighted compositions are different. Overall, the difference between the three length compositions is larger in the early period and in the southern fishery due to small sample size.

The spatial distribution of the length frequency samples was much more restricted than the data used for the CPUE analysis (Fig. 2d). Therefore, since no temporal correlation was used in the analysis, the VAST model substantially augments the spatial distribution for a particular year-quarter using the spatial main effect given that the data is very limited for each year-quarter. This can be seen in the similarity in the spatial distribution of mean length among years (Fig. 2c).

*3.4 Use in the stock assessment model*

The spatio-temporal modeling approach described above estimates a multivariate index of relative abundance for size composition, such that it is preferable to fit the index in the stock assessment model using a multivariate likelihood function that takes correlation among sizes and time into consideration. However, if the assessment software does not have this capability, the index can either be: 1) broken into separate indices for each age (or size composition group), or 2) broken into a total abundance index with a separate estimate of proportion-at-age or proportion-at-size that is then treated as "composition data" within the stock assessment model. The variance in estimates of proportion-at-age or -size could be used to calculate an input sample size for likelihood function used to fit the composition data in the stock assessment (Thorson, 2014), and this input sample sizes could then be down-weighted to represent the impact of model mis-specification (Francis, 2017; Thorson et al., 2017b; Xu et al., 2020).

Size- or age-compositions representing the component of the population associated with the abundance index are unlikely to be the same as those describing fishery catches. The above methods define an approach to estimate the composition for the index of relative abundance, which is complicated for fishery-dependent CPUE because fishery composition data are used both to estimate population proportions in each category, and to estimate the selectivity governing fishery removals. Composition data representing fishery removals should be raised to the total catch by weighting the spatial-explicit composition data by the respective catch for each location. However, this raises two problems. The first is the appropriate weight to give to the composition data likelihood function in the stock assessment. This is a standard problem in contemporary fisheries stock assessment (Francis, 2017; Maunder et al., 2017; Punt, 2017), and will not be addressed here. The second is that the composition data will generally be used twice due to limitations of standard stock assessment approaches, once for the index of relative abundance and once for the catch. Double use of data under the typical assumption of independent likelihood functions is a violation of standard statistical practices. However, given the arbitrariness of data weighting and the common approach of internally estimating the weighting of composition data, the double use of the data is probably less of an issue than using biased composition data for indices of relative abundance. A simple *ad hoc* approach to downweighting the data (e.g., Tremblay-Boyer et al., 2018) might be all that is needed.

The method used to calculate the catch-at-size within the spatio-temporal analysis is to sum predicted catch in number at size (or observed catch-in-number, if it is assumed to be known with little error) for each stratum to give the overall catch-at-size to use in the stock assessment model. If the data used in the spatio-temporal model is not the total catch, e.g., if some data were discarded at the grooming stage to avoid bias when estimating the index of relative abundance, then the calculations need to be adjusted to use the total effort or the total observed catch by stratum. The stock assessment model could then remove the catch-at-size directly as estimated from the spatio-temporal model paralleling a VPA or with flexible time varying selectivity if used in a contemporary statistical stock assessment model (e.g., Nielsen and Berg, 2014; Stewart and Monnahan, 2017). For example, a flexible semi-parametric time-varying selectivity has been implemented in Stock Synthesis (Xu et al., 2019b). Alternatively, consistent with the underlying assumptions of the spatio-temporal model, the temporal variation may be due to spatial changes in the fleet distribution and the method of Hoyle and Davies (2009) using many spatially defined fisheries with time-invariant selectivity might be appropriate. Seasonal selectivities may be needed if this approach is taken since many stocks experience seasonal movement. Other shifts in distribution due to environmental factors (e.g., El Niño) may also have to be addressed.

Since the composition associated with the abundance index represents the population, only factors representing density effects on the composition data should be used in calculating the size structure of the index in the spatio-temporal model, and catchability effects should be ignored when predicting compositions. In contrast, catchability effects should be included when predicting the fishery catch size structure in the spatio-temporal model as these will impact the expected composition of the catch from each fishery.

#### 3.4.1 Yellowfin application

The stock assessment was applied to the datasets presented in the previous sections. The length-composition data were also calculated for the catch so that the abundance length-composition and the catch length-compositions could be used simultaneously. The data were included in the assessment as either: a) length compositions and an overall index of abundance or b) indices for each 10cm length bin. The index for the southern fishery was not included in the stock assessment due to low sample size, but the southern area was included in the calculation for indices based on the whole EPO. Length composition sample size is equal to the number of wells samples except where noted otherwise.

The stock assessment model runs compared seven ways of treating length-composition and length-based catches (labeled 1-5, 6A and 6B below), and three alternative ways of weighting the length composition data (labeled DW1, DW2, and DW3):

1) Nominal CPUE and length composition data for each of the three fisheries.
2) VAST-standardized CPUE and length-composition data for each of the three fisheries with
    a. CPUE-weighted length-composition; and
    b. Catch-weighted length-composition.
3) VAST-standardized CPUE and CPUE-weighted length composition data for each of the three fisheries treated as "surveys" (i.e., fisheries with no associated catch but an index of abundance and composition data), and catch-weighted composition data for the fisheries. The fisheries and surveys have different estimated selectivities.
4) VAST-standardized CPUE and CPUE-weighted length composition data for the whole EPO, and catch-weighted composition data for the three fisheries. The fisheries and surveys have different estimated selectivities
5) The same as 4), except that the fisheries have time-varying selectivities. A double-normal selectivity function was used for the fishery and the parameters of the function have a random walk process in time with a 20% coefficient of variation (CV).
6) A: VAST-standardized length-bin-specific CPUE indices (time-invariant estimated CV) for the whole EPO and catch-weighted composition data for the three fisheries. The fisheries and surveys have different selectivities. The fisheries have estimated selectivities. The selectivities for each survey that represent a length bin-specific CPUE index are fixed at 1 for lengths in that length bin and zero for others lengths. The lognormal likelihood function used to fit the index data is not defined for zero observations so the first three and last two 10cm length indices were not used in the analysis because they were predominantly zeros. Initially, 1 was added to all the index values to avoid computational errors due to zeros for the other age bins. However, the results were still impacted by the zero observations, which often occurred between bins with observations substantially greater than zero, and therefore the zero observations were removed from the calculations. This issue needs further investigation if the approach is used in applications.

    B: The same as 6A, but with time-varying CVs in the likelihood function used to fit the abundance index based on those estimated from the spatio-temporal model with an estimated additive CV.

A. DW1). The same as scenario 2b.
B. DW2). DW1, but with the length-composition samples re-weighted using the Francis method. The Francis method calculates the sample sizes that lead to confidence intervals on the observed mean size that are consistent with the hypothetical fit to mean size and therefore takes correlated residuals into consideration.
C. DW3). DW2, but with the initial input sample size based on the VAST-estimated sample size then re-weighted using the Francis method.

The estimates of the spawning biomass ratio (SBR: the spawning biomass divided by the average spawning biomass in the absence of fishing) were similar for all model runs (Figs 8 and 9). The largest differences occurred for the runs that used length bin indices of abundance rather than composition data. However, these models were problematic due to issues related to dealing with zero observations. There are differences in the management quantities among the runs that could result in different management actions (Table 4). Using catch-weighted length-composition data had the largest influence on management quantities.

The sample sizes for the composition data estimated from the spatio-temporal model differ somewhat from the number of wells sampled (Fig. 9). However, the data weighting had little influence on the results (Fig. 10).

**4. Discussion**

Spatio-temporal modeling of CPUE and associated composition data have some clear advantages (Thorson, 2019a) but are still rarely used as inputs to stock assessments. We have outlined here several approaches to using the fishery-dependent indices of abundance and composition data within stock assessment models. The approach is based on three concepts: 1) composition data that are used to determine the component of the population represented by the index should be weighted by CPUE (abundance) while the composition data used to represent the fish removed from the stock should be weighted by catch; 2) due to spatial non-randomness in fishing effort and fish distribution, the index, index composition, and catch composition, should be calculated at a fine spatial scale (e.g., 1°x1°) and summed; and 3) fine-scale spatial stratification will likely result in under-sampled and unsampled cells and some form of smoothing method needs to be applied to inform these cells.

Our application to yellowfin tuna showed some sensitivity to the assumptions about how the indices of abundance and length-compositions are created, but these sensitivities are probably small compared to other uncertainties in the stock assessment model (e.g., about natural mortality, asymptotic length, selectivity, data weighting, or the stock-recruitment relationship). The results will be more sensitive to the approaches we have discussed when there is clear spatial structure in the abundance and size of fish while fishing effort is nonrandom. The sensitivity will also depend on what other data are used in the assessment and how informative they are. The lack of sensitivity of the results of the yellowfin assessment to the use of alternative indices could be partly due to the inclusion of composition data from other fisheries (longline and purse seine sets on floating objects and unassociated schools) in the assessment.

The benefit of spatio-temporal modelling was illustrated by the application of area weighting to GLM-based indices of abundance, which led to higher weight to the poorly sampled southern stratum, creating an index of abundance for the whole EPO that is more variable. In contrast, the spatio-temporal model allowed the sharing of information among areas and removed much of the variability in the final index of abundance. This result suggests that spatio-temporal models can be useful tools to derive indices of abundance and composition data when sampling intensity varies across the spatial domain.

One advantage of using spatio-temporal analyses is that they account for variability in sampling over space and time, which otherwise violates the assumptions underlying the use of time-invariant catchability and selectivity. However, these analysis methods still assume that the gear component of selectivity is temporally invariant. The inclusion of relevant catchability covariates may partially address this issue. Applications of spatio-temporal models thus appear a more effective approach to handling time-varying selectivity than alternatives (e.g., Stewart and Martell, 2014) because they may retain more of the information on abundance.

Computational demands are probably the largest roadblock to conducting the desired analyses. Compromises, such as broader size bins and simplified correlation structures, have to be made until the computational demands can be solved. Other issues such as the appropriate multivariate likelihood functions for the index, index composition, and catch composition data are academically interesting, but probably of a lesser priority.

*4.1 Issues*

Several issues arise when conducting spatio-temporal analyses, which will need to be addressed in the future. Most of these come down to a tradeoff between the computational demands of the approach and the desire to implement more accurate modelling of the system. Here we focus on several issues. The first is the extent to which the correlation structure of the model approximates what might be expected given the underlying habitat complexity (i.e., how habitat changes in space). For example, depending on the habitat, CPUE may change more with latitude than with longitude, or vice versa (i.e., geometric anisotropy). In such cases, the parameters of the correlation function should be estimated separately for latitude and longitude, and this should generally be the default assumption.

Temporal and length correlation (e.g., AR1) were not used in the yellowfin application. This choice was made to (1) minimize the estimation covariance between estimated proportion-at-lengths for adjacent years (this estimation covariance would likely increase when specifying a temporal correlation in the spatio-temporal model), and (2) avoid the increased computational demands when modelling temporal autocorrelation. Length bin – time, length bin – space, and length bin – time – space interactions are included in the model, but the spatial effects were assumed independent across time and length. We specified spatio-temporal variation and intercepts as independent for each combination of year, quarter, and length bin; this specification minimizes any sharing of information over time, and may be appropriate given that the stock assessment model's likelihoods used in the yellowfin application assume that length composition data are independent for every year-quarter combination. The correlation structure is also complicated due to strong cohorts growing or aging over time. The amount of correlation might be related to the biology of the species. For example, less temporal correlation might be expected for short-lived species that exhibit large recruitment variation. More research needs to be conducted to determine how much sharing of information should be carried out within the spatio-temporal model versus within the stock assessment itself.

There is a tradeoff between the level of data aggregation and computational demands. The least computationally demanding approach is to define a data point as the data aggregated to the level of the factors included in the spatio-temporal model. For example, a data point could be a year by 1x1 degree square by 1 cm length interval in a simple spatio-temporal model. The variance within that stratum could then be associated with the data point and used in the likelihood function. However, the data would likely have to be disaggregated if it was later decided that other factors should be added to the analysis. For example, the data would also have to be disaggregated by vessel if vessel was included as a factor. Conversely, each fishing operation could be considered as a data point to allow full flexibility for covariates and variance estimates. A related problem is that composition data may not be available for every set of the gear in some, if not most, applications. In this case, the catch and composition data could be aggregated separately by stratum and separate likelihoods used for each component, with the catch likelihood weighted by the effort and the composition data (perhaps using a multinomial distribution-based likelihood) weighted by the sample size. Scaling factors may need to be estimated for the variance components of each of the likelihoods. In the yellowfin application we use the full data set to calculate the index of abundance and a data set limited to locations that had composition data to calculate the length compositions, which may lead to some inconsistencies.

Inclusion of age or size data extends the standard spatio-temporal model from 3 to 4 dimensions, greatly increasing the computational demands of the analysis. Analysts may need to decrease the resolution of size structure included in the model. For example, Kai et al. (2017b) only used 13 size-classes in their analysis of data for shortfin mako shark ( ) and we aggregated the data to 10 cm bins for the yellowfin applications. Most stock assessment models use composition data with the intention to mainly provide information on selectivity and recruitment, and therefore may need a finer resolution. For our yellowfin tuna example, the 2cm length bins used in the current stock assessment (Minte-Vera et al., 2019) led to a VAST model that was too computationally demanding to apply, and 10cm bins had to be used instead. Fortunately, we found that the results from the stock assessment using the nominal length composition data were essentially the same between 2cm and 10cm bins for all fleets. Additional research is needed to investigate efficient ways to implement the model to ensure that the desired resolution is practical.

When data from multiple fleets are available, consideration should be given to analyzing those data simultaneously in the same model, thereby allowing for an "integrated" index of abundance to be developed. Although the catch size-composition may differ between the fleets in a given spatial stratum due to differences in their selectivity (e.g., the gear characteristics such as depth of fishing), in spatial strata where the fleets overlap, the underlying size-specific density encountered by the fleets will be the same. This concept may be utilized to better model age- or size-composition spatially, but will require estimation of gear specific selectivity. Several studies have developed indices of abundance using spatio-temporal analysis of multiple surveys (e.g., Dolder et al., 2018; Grüss et al., 2018; Runnebaum et al., 2017), but to our knowledge only Ono et al. (2018) has done so while generating age- or size-composition estimates.

The final, and possibly the most important issue, is preferential sampling due to the use of fishery-dependent CPUE data. In general, fishers target areas where the fish are abundant, and therefore the available data will have higher CPUE than in under-sampled and unsampled areas. It should also be noted that fishers may fish in areas of lower abundance due to convenience, safety, or profitability. The preferential sampling issue is somewhat reduced using spatial weighting compared to data or effort weighting. However, there is a tradeoff between accuracy and variance when applying spatial-temporal models since areas with abundant data (preferentially sampled areas) will inform under-sampled and unsampled areas. This is particularly concerning when information for large contiguous spatial areas has to be imputed for some years. Including covariates that relate to abundance may further reduce the impacts of preferential sampling. However, separating the effect of covariates on abundance from catchability may be problematic.

### *4.2 Other methods*

The spatio-temporal modelling approach we described is based on Gaussian random fields and follows the work of Lewy and Kristensen (2009), Kristensen et al. (2014), Nielsen et al. (2014), Thorson et al. (2015a), Kai et al. (2017a,b), and Thorson and Haltuch (2018). However, there are several other approaches that have been used for spatio-temporal modelling. For example, the 2-D spatial surfaces of generalized additive models (GAMs; e.g., Wood, 2006), commonly fitted with tensor product smooth terms, can be extended to allow for changing spatial structure through time by specifying a 3-D surface (e.g., Rooper et al., 2016). Use of separate smoother types for space and time allows for different amounts of smoothing in space and in time (Augustin et al., 2013). Other extensions of classical generalized additive models permit the modelling of spatio-temporal structure where boundaries exist in space. This is done with the use of soap film smoothers (Wood et al., 2008) that do not require spatial structure to be connected across boundaries, such as stock boundaries (Augustin et al., 2013), or across physical barriers such as peninsulas. Also, spatio-temporal GAMs allow for other covariates in the model, which could permit the simultaneous modelling of size, in addition to the spatio-temporal effects. GAMs can be extended to generalized additive mixed-effects models (GAMMs) to account for factors, such as vessel effects, that are better parameterized as random effects. This can be done in the context of GAMs by treating the random effects as penalized fixed effects (Wood, 2006; Augustin et al., 2013). There are also simpler approaches for improving the fit to size data when there is spatial size variation and time-varying sampling. These include stratifying and weighting the size data in proportion to the long-term spatial distribution of either relative abundance (Hoyle and Langley, 2011) or catch (Hoyle et al., 2012).

Future research could also explore the range of machine-learning regression techniques to model spatio-temporal variation (in the following, we refer to tree-based methods for regression, but note that recursive neural networks have shown promise in time-series learning environments). Tree-based methods (e.g., Classification and Regression Trees (CART) and random forests; Breiman et al., 1984; Breiman, 2001) can be used to explore interactions of space (using multiple, possibly correlated, measures, e.g., latitude, longitude, and/or distance from port), time (year, season), and size (length, weight, age), while also including other explanatory variables. They can easily be adapted to multivariate response variables and loss functions other than squared error loss (e.g., length-frequency distributions or CPUE trends; Lennert-Cody et al., 2010, 2013). There are no structural constraints on the space and time scales of the spatio-temporal structure captured using these types of algorithms, except in a limited manner by way of the spatial and temporal resolution of the predictors. Therefore, by their very nature, tree-based algorithms have the flexibility to implicitly capture complex spatio-temporal structure over a range of scales, and this can be helpful in exploratory analyses in the case of CART, or for challenging predictive problems in the case of random forest methods, when the underlying process behind the data are unknown.  Ongoing research suggests that machine learning techniques can decrease predictive errors for local predictions of resource density, but also result in increased bias in some instances (Stock et al., 2019).

### *4.3 Potential applications*

There are numerous current and possible applications of spatio-temporal models that are relevant to the management of tuna and related species in the eastern Pacific Ocean (EPO) or other species in other oceans. For example, Kai et al. (2017a,b) applied spatio-temporal models to blue and mako sharks, and Thorson et al. (2017a) explored the relative explanatory power of local and regional temperature, size-structure, and otherwise unexplained processes in explaining the shifting distribution for Alaska pollock (*Gadus chalcogrammus*) in the Bering Sea. Similarly, spatio-temporal models could be used to generate short-term forecasts of distribution that may in some cases improve upon a default "persistence" forecast (Thorson, 2019c), and these could be useful for fishery stakeholders, e.g. for planning spatial management such as move-on rules (Eveson et al., 2015). In all cases, simulation experiments would be useful to determine the effectiveness of the various approaches and their specific applications.

Table 2 lists example applications using spatio-temporal models to standardize CPUE data and Table 3 lists example applications that include composition data. The applications use a variety of methods and make various assumptions. Some use formal statistical tests to determine what assumptions are used in the final model, while others are *ad hoc* or are made due to computational constraints. For example, some models use only a spatial-temporal component, while others used both a spatial component and a spatial-temporal component. The spatial component may be a GRF or a function of latitude and longitude. Some applications have temporal correlation, but many assume temporal independence, which ignores some of the advantages of using spatio-temporal models. Similarly, many of the applications for composition data ignore correlation among size-classes. This indicates that there is still much research needed to determine the appropriate assumptions to make (likely to differ among applications) and what strategy to adopt for improved performance.

Management boundaries are often politically motivated and do not represent the biology of the species. This causes issues when there is spatio-temporal variation in the population distribution in relation to the management boundary (i.e., environmental conditions cause some of the stock to move across the management boundary in some years) and application of spatio-temporal models may better inform management decisions. For example, purse seine data have been used to develop indices of relative abundance for silky sharks (*Carcharhinus falciformis*) in the eastern Pacific Ocean. However, these indices, particularly for juvenile sharks, appear to be biased due to movement of individuals in and out of the EPO, or in and out of the area fished by purse seiners, possibly due to changing environmental conditions (Lennert-Cody et al., 2019). Therefore, integrating additional data sets (e.g., purse seine data from the western and central Pacific Ocean and longline data for the whole Pacific Ocean) into a spatio-temporal analysis to extend the northern and western range of the data may help determine the influence of movement and improve the indices of abundance, since they provide a way to account for data beyond the spatial domain of interest (e.g., the EPO), but allow extraction of indices specific to the area of choice.

Temporal change in gear characteristics is a common issue in CPUE standardizations and, if not addressed, can bias indices of abundance. The issue is typically addressed by including gear characteristics as covariates in the analysis. However, gear characteristics may also vary spatially, and spatial changes in the effort may confound temporal changes in gear characteristics. Therefore, spatial patterns in gear characteristics need to be addressed. The Japanese longline fleet that targets tuna in tropical waters has shown changes in spatial distribution and in gear characteristics over time in several oceans including the EPO (e.g., Lennert-Cody et al., 2012; Hoyle and Okamoto, 2011; Hoyle et al., 2017). The catch and effort data from this fleet are used to compute the main indices of relative abundance for the EPO stock assessments of yellowfin and bigeye tuna. Spatio-temporal models that consider gear factors as catchability covariates should be used to develop the indices of abundance. For example, the spatio-temporal component of the model might depend on the hooks between floats, which influences the depth of the hooks, and this could combine the statistical habitat based standardization approaches (e.g., Maunder et al., 2006b) with the newer spatio-temporal methods.

Spatial variation in sex as well as size is common for many species and should be taken into account when developing indices of abundance and composition data. For example, albacore tuna (*Thunnus alalunga*) shows differences in distribution by size and sex (Chen et al., 2010; Ichinokawa et al., 2008). Therefore, the analysis of longline catch and effort data, which is used as the main index of abundance, could be conducted using spatio-temporal models in which the spatio-temporal component is a function of both length and sex.

Cohort targeting, particularly when combined with ontogenetic movement, can lead to serious issues with indices of abundance based on CPUE data that may be resolvable with spatio-temporal models. Pacific bluefin (*Thunnus orientalis*) have shown substantial changes in both the spatial distribution of the stock and the fleet, which has been further complicated by changes in targeting (Oshima et al., 2012). Standardization models for Taiwanese longline bluefin CPUE showed significant year and area interactions, and indices of relative abundance showed different trends among spatial strata (Chang et al., 2017). There appears to be spatial targeting of large cohorts of mature bluefin tuna as they move through the Japanese longline fishery into the Taiwanese longline fishery (Maunder et al., 2014). This violates the constant selectivity assumption that is typically needed for an index of relative abundance to be informative. Joint spatial-temporal modelling of CPUE and composition data may facilitate the use of constant selectivity for a longline based index of relative abundance for Pacific bluefin tuna. In addition, joint modelling of the Japanese and Taiwanese longline data, which may have different selectivities, could improve the spatial coverage.

Spatio-temporal models may provide an effective way to complement survey data with fishery- dependent data. Assessments for eastern Pacific Ocean dolphin stocks (e.g., Hoyle and Maunder, 2004) have been conducted using estimates of absolute abundance based on ship-based line transect surveys (Gerrodette et al., 2008). However, these surveys are expensive and limited in temporal scope and frequency. Therefore, it might be useful to combine the information collected by fisheries observers onboard the commercial tuna vessels, which use dolphins to locate and catch yellowfin tuna, with the survey data to obtain estimates of abundance. Although it is unclear how the biases that have previously been identified in the fisheries observer data (Lennert-Cody et al., 2001, 2016) would be addressed, spatio-temporal models would be required to combine the two data sets because the commercial tuna fishery has a different spatial and temporal distribution of effort than do the fishery-independent surveys.

### *4.4 Recommendations*

Applications of spatio-temporal models should follow the recommendations outlined in Thorson (2019a). For example, always include a spatio-temporal interaction term, carefully define the spatial domain for extrapolation, distinguish catchability and density covariates, and process results to expand estimates while accounting for area (in index standardization) and/or catch (when standardizing catch proportions as explored here).

We also provide some recommendations on best practices for stock assessment, but more research needs to be conducted before they become the standard practice. CPUE and compositional data should be analyzed in a single spatio-temporal analysis, and the resulting index of abundance for the whole stock (or sub-stock if interacting sub-stocks are modelled in the stock assessment) and associated CPUE-weighted composition data should be used in the assessment to represent abundance. The area-specific catch-weighted composition data should be used to represent fisheries with time-varying selectivity (or fine spatial scale defined fisheries with time invariant selectivity), if necessary, or it should be assumed that fishery catches are known exactly and the product of catch-expanded composition data and total fishery catch removed from the population (similar to how removals are treated in a VPA). The effective composition sample size from the spatio-temporal model should be used in the stock assessment to represent relative weight and an appropriate form of reweighting used to adjust the overall weighting for the composition data. The year-specific variance for the index of abundance from the spatio-temporal model should be used in the stock assessment likelihood function, and an additional variance should be estimated within the stock assessment model to represent unmodelled variation in catchability and other model error.

**Acknowedgements**

Andre Punt provided comments on the manuscript and editorial assistance.

## References


Aires-da-Silva, A., Maunder, M.N., 2012. Status of yellowfin tuna in the eastern Pacific Ocean in 2010 and outlook for the future. IATTC Stock Ass. Rep. 12. http://www.iattc.org/PDFFiles2/StockAssessmentReports/SAR-12-YFTENG.pdf.

Augustin, N.H., Trenkel, V.M., Wood, S.N., Lorance, P., 2013. Space-time modelling of blue ling for fisheries management. Environmetrics 24, 109-119.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman and Hall. 358pp.

Breiman, L., 2001. Random forests. Mach. Learn. J. 45, 5-32.

Cao, J., Thorson, J.T., Richards, R.N., Chen, Y., 2017. Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine Can. J. Fish. Aquat. Sci. 74: 1781–1793

Campbell, R.A., 2004. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70, 209-227.

Carruthers, T.R., Ahrens, R.N., McAllister, M.K., Walters, C.J., 2011. Integrating imputation and standardization of catch rate data in the calculation of relative abundance indices. Fish. Res. 109, 157–167.

Cavieres,J., Nicolis, O., 2018. Using a spatio-temporal Bayesian approach to estimate the relative abundance index of yellow squat lobster (*Cervimunida johni*) off Chile. Fish. Res. 208, 97–104

Chen K.S., Crone P.R., Hsu C.C., 2010. Reproductive biology of albacore tuna (Thunnus alalunga) in the western North Pacific Ocean. J Fish Biol. 17, 119–136.

Chang S-K., Liu H-I, Fukuda, H., Maunder, M.N., 2017. Data reconstruction can improve abundance index estimation: An example using Taiwanese longline data for Pacific bluefin tuna. PLoS ONE 12(10): e0185784.

Diggle, P. J., Ribeiro, P., 2007. Model-based geostatics New York: Springer. 228pp.

Dolder, P.J., Thorson, J.T., Minto, C., 2018. Spatial separation of catches in highly mixed fisheries. Sci. Rep. 8, 13886.

Eveson, J.P., Hobday, A.J., Hartog, J.R., Spillman, C.M., Rough, K.M., 2015. Seasonal forecasting of tuna habitat in the Great Australian Bight. Fish. Res. 170, 39–49.

Francis, R.I.C.C., 2017. Revisiting data weighting in fisheries stock assessment models. Fish. Res. 192, 5-15.

Gerrodette, T., Watters, G., Perryman, W., Ballance, L., 2008. Estimates of 2006 dolphin abundance in the eastern tropical Pacific, with revised estimates for 1986-2003. NOAA-TM-NMFS-SWFSC-422. https://repository.library.noaa.gov/view/noaa/3639/noaa_3639_DS1.pdf

Grüss, A., Thorson, J.T., 2019. Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. ICES J. Mar. Sci. 76, 1748–1761.

Grüss, A., Drexler, M.D., Ainsworth, C.H., Babcock, E.A., Tarnecki, J.H., Love, M.S., 2018. Producing Distribution Maps for a Spatially-Explicit Ecosystem Model Using Large Monitoring and Environmental Databases and a Combination of Interpolation and Extrapolation. Front. Mar. Sci. 5, 16.

Grüss, A., Walter III, J.F., Babcock, E.A., Forrestal, F.C., Thorson, J.T., Lauretta, M.V., Schirripa, M.J., 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fish. Res. 213, 75–9.

Harley, S. J., Myers, R. A., Dunn, A., 2001. A meta-analysis of the relationship between catch-per-unit-effort and abundance. Can. J. Fish. Aquat. Sci. 58, 1705-1772.

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

Hoyle, S.D, Davies, N.. 2009. Stock assessment of albacore tuna in the South Pacific Ocean. WCPFC Scientific Committee 5, Port Vila, Vanuatu. WCPFC-SC5-2009/SA-WP-6. https://www.wcpfc.int/meetings/5th-regular-session-scientific-committee

Hoyle, S., Langley, A., 2011. Spatial size data stratification for length-based stock assessments. WCPFC SC7 SA IP-9, Pohnpei, Federated States of Micronesia, 9-17. https://www.wcpfc.int/meetings/7th-regular-session-scientific-committee

Hoyle, S.D., Maunder, M.N., 2004. A Bayesian integrated population dynamics model to analyze data for protected species. Animal Biodivers. Cons. 27, 247-266.

Hoyle, S.D., Okamoto, H., 2011. Analyses of Japanese longline operational catch and effort for bigeye and yellowfin tuna in the WCPO. WCPFC-SC7-2011/SP IP-01. https://www.wcpfc.int/meetings/7th-regular-session-scientific-committee

Hoyle, S., Hampton, J., Davies, N., 2012. Stock Assessment of Albacore in the south Pacific Ocean Rev 1 (29 July 2012). SA-WP-04. In: Eighth Regular Session of the Scientific Committee. https://www.wcpfc.int/meetings/8th-regular-session-scientific-committee

Hoyle, S.D., Satoh, K., Matsumoto, T., 2017. Exploring possible causes of historical discontinuities in Japanese longline CPUE. IOTC-2017-WPM08-19 Rev_1. https://www.iotc.org/sites/default/files/documents/2017/10/IOTC-2017-WPM08-19_Rev_1.pdf

Ichinokawa, M., Coan, A.L., Takeuchi, Y., 2008. Transoceanic migration rates of 544young North Pacific albacore, Thunnus alalunga, from conventional tagging data. Can. J. Fish. Aquat. Sci. 65, 1681–1691.

Kai, M., 2019. Spatio-temporal changes in catch rates of pelagic sharks caught by Japanese research and training vessels in the western and central North Pacific. Fish. Res. 216, 177–195

Kai, M., Thorson, J.T., Piner, K., Maunder, M.N., 2017a. Predicting the spatio-temporal distributions of pelagic sharks in the western and central North Pacific. Fish. Oceanogr. 26, 569-582.

Kai, M., Thorson, J. T., Piner, K. R., Maunder, M. N., 2017b. Spatio-temporal variation in size-structured populations using fishery data: an application to shortfin mako ( Can. J. Fish Aquat. Sci. 74, 1765-1780.

Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E., 2014. Estimating spatio-temporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. 71, 326–336.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., Bell, B., 2016. TMB: Automatic Differentiation and Laplace Approximation. J. Stat. Softw. 70(5), 1-21.

Lennert-Cody, C.E., Buckland, S.T., Marques, F.C., 2001. Trends in dolphin abundance estimated from fisheries data: a cautionary note. J. Cetacean Res. Manag. 3, 305-319.

Lennert-Cody, C.E., Minami, M., Tomlinson, P.K., Maunder, M.N., 2010. Exploratory analysis of spatial-temporal patterns in length-frequency data: An example of distributional regression trees. Fish. Res. 102, 323-326.

Lennert-Cody, C.E., Okamoto, H., Maunder, M.N., 2012. Analyses of Japanese longline operational-level catch and effort data for bigeye tuna in the eastern Pacific Ocean. IATTC Document SAC-04-05B. https://www.iattc.org/Meetings/Meetings2013/SAC-04/Docs/_English/SAC-04-05b_Analyses%20of%20Japanese%20longline%20catch%20and%20effort%20data%20for%20bigeye.pdf

Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A., Minami, M., 2013. Defining population spatial units: Simultaneous analysis of frequency distributions and time series. Fish. Res. 139, 85-92.

Lennert-Cody, C.E., Maunder, M.N., Fiedler, P.C., Minami, M., Gerrodette, T., Rusin, J., Minte-Vera, C.V., Scott, M., 2016. Purse-seine vessels as platforms for monitoring the population status of dolphin species in the eastern tropical Pacific Ocean. Fish. Res. 178, 101-113.

Lennert-Cody, C.E., Clarke, S.C., Aires-da-Silva, A., Maunder, M.N., Franks, P.J.S., Román, M., Miller, A.J., Minami, M., 2019. The importance of environment and life stage on interpretation of silky shark relative abundance indices for the equatorial Pacific Ocean. Fish. Oceanogr. 28, 43-53.

Lewy, P., Kristensen, K., 2009. Modelling the distribution of fish accounting for spatial correlation and overdispersion. Can. J. Fish. Aquat. Sci. 66, 1809–1820.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 423–498.

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72, 7-18.

Maunder, M.N., Punt, A.E., 2004. Standardizing Catch and Effort Data: A Review of Recent Approaches. Fish. Res. 70, 141-159.

Maunder, M.N., Sibert, J.R. Fonteneau, A., Hampton, J., Kleiber, P., Harley, S., 2006a. Interpreting catch-per-unit-of-effort data to assess the status of individual stocks and communities. ICES J. Mar. Sci. 63, 1373-1385.

Maunder, M.N., Hinton, M.G., Bigelow, K.A., Langley, A.D., 2006b. Developing indices of abundance using habitat data in a statistical framework. Bull. Mar. Sci. 79, 545–559.

Maunder, M.N., Piner, K.R., Aires-da-Silva, A., 2014. Stock status of pacific Bluefin tuna and urgent need for management action. IATTC Stock Ass. Rep. 15, 47-73. https://www.iattc.org/PDFFiles/StockAssessmentReports/_English/No-15-2014_Status%20of%20the%20tuna%20and%20billfish%20stocks%20in%202013.pdf

Maunder, M.N., Crone, P.R, Punt, A.E., Valero, J.L., Semmens B. X., 2017. Data conflict and weighting, likelihood functions and process error. Fish. Res. 192, 1-4.

McKechnie S, Hoyle S, Harley S., 2013. Longline CPUE series that account for changes in the spatial extent of fisheries. WCPFC-SC/SA-IP-05. https://www.wcpfc.int/meetings/9th-regular-session-scientific-committee

Methot, R.D., Wetzel, C.R., 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment

and fishery management. Fish. Res. 142, 86–99.
Minte-Vera, C.V., Xu, H., Maunder M.N., 2019. Status of yellowfin tuna in the eastern Pacific Ocean in 2018 and the outlook for the future. IATTC DOCUMENT SAC-10-07. https://www.iattc.org/Meetings/Meetings2019/SAC-10/Docs/_English/SAC-10-07_Yellowfin%20tuna%20assessment%20for%202018.pdf
Nielsen, A., Berg. C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158: 96-101.
Nielsen, J.R., Kristensen, K., Lewy, P., Bastardie, F., 2014. A Statistical Model for Estimation of Fish Density Including Correlation in Size, Space, Time and between Species from Research Survey Data. PLOS ONE 9, e99151.
Ono, K., Ianelli, J.N., McGilliard, C.R., Punt, A.E., 2018. Integrating data from multiple surveys and accounting for spatio-temporal correlation to index the abundance of juvenile Pacific halibut in Alaska. ICES J. Mar. Sci. 75, 572–584.
Oshima, K., Mizuno, A., Ichinokawa, M., Takeuchi, Y., Nakano, H., Uozumi, Y., 2012. Shift of fishing efforts for Pacific bluefin tuna and target shift occurred in Japanese coastal longliners in recent years. ISC/12/PBFWG-3/05. http://isc.fra.go.jp/pdf/PBF/ISC12_PBF_3/ISC12_PBFWG-3_05_Oshima.pdf
Perrettia, C.T., Thorson, J.T., 2019. Spatio-temporal dynamics of summer flounder (*Paralichthys dentatus*) on the Northeast US shelf. Fish. Res. 215, 62–68
Punsly, R., 1987. Estimation of the relative annual abundance of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean during 1970-1985. IATTC Bul. 19, 263-306.
Punt, A.E., 2017. Some insights into data weighting in integrated stock assessments. Fish. Res. 192, 52-65.
Punt, A.E., 2019. Spatial stock assessment methods: A viewpoint on current issues and assumptions. Fish. Res. 213: 132-143.
Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., Prince, J.D., 2000a. Stock assessment of school shar: Galeorhinus galeus based on a spatially-explicit population dynamics model. Mar. Freshw. Res. 51, 205–220.
Punt, A.E., Walker, T.I., Taylor, B.L., Pribac, F., 2000b. Standardization of catch and effort data in a spatially-structured shark fishery. Fish. Res. 45, 129–145.
Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., 2001. Population modeling and harvest strategy evaluation for school and gummy shark. Report of FRDC Project No. 99/102. http://frdc.com.au/Archived-Reports/FRDC%20Projects/1999-102-DLD.pdf
R Development Core Team, 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org
Roa-Ureta, R., Niklitschek, E., 2007. Biomass estimation from surveys with likelihood based geostatistics. ICES J. Mar. Sci. 64, 1723–1734.
Rooper, C.N., Sigler, M.F., Goddard, P., Malecha, P., Towler, R., Williams, K., Wilborn, R., Zimmermann, M., 2016. Validation and improvement of species distribution models for structure-forming invertebrates in the eastern Bering Sea with an independent survey. Mar. Ecol. Prog. Ser. 551, 117–130.
Runnebaum, J., Guan, L., Cao, J., O'Brien, L., Chen, Y., 2017. Habitat suitability modeling based on a spatiotemporal model: an example for cusk in the Gulf of Maine. Can. J. Fish. Aquat. Sci. 75, 1784–1797.
Sampson, D. B., 2014. Fishery selection and its relevance to stock assessment and fishery management. Fish. Res. 158, 5-14.
Stewart, I.J., Martell S.J.D., 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fish. Res. 158, 40-49.
Stewart, I.J., Monnahan, C.C., 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fish. Res. 192, 126-134.
Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E., Semmens, B.X., 2019. The utility of spatial model-based estimators of unobserved bycatch. ICES J. Mar. Sci. 76, 255–267.
Thorson, J.T., 2014. Standardizing compositional data for stock assessment. ICES J. Mar. Sci. 71, 1117–1128.
Thorson, J.T., 2019a. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210, 143–161.
Thorson, J.T., 2019b. Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics. Fish. Res. 217, 133-139.
Thorson, J.T., 2019c. Forecast skill for predicting distribution shifts: A retrospective experiment for marine fishes in the Eastern Bering Sea. Fish. Fish. 20, 159–173. doi:10.1111/faf.12330.
Thorson, J.T., Barnett, L.A.K., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74, 1311-1321.

Thorson, J.T., Haltuch, M.A., 2018. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Can. J. Fish. Aquat. Sci. 1–14.

Thorson, J.T., Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fish. Res. 175, 66–74.

Thorson, J.T., Ono, K., Munch, S.B., 2014. A Bayesian approach to identifying and compensating for model misspecification in population models. Ecology 95, 329–341.

Thorson, J. T., Shelton, A. O., Ward, E. J., Skaug, H., 2015a. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. 72, 1297−1310.

Thorson, J. T., Skaug, H., Kristensen, K., Shelton, A. O., Ward, E. J., Harms, J., Benante, J., 2015b. The importance of spatial models for estimating the strength of density dependence. Ecology 96, 1202–1212.

Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C., Zipkin, E.F., 2016. Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. Global Ecol. Biogeog. 25, 1144–1158.

Thorson, J.T., Ianelli, J.N., Kotwicki, S.,. 2017a. The relative influence of temperature and size-structure on fish distribution shifts: A case-study on walleye pollock in the Bering Sea. Fish Fish. 18, 1073–1084.

Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017b. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. 192, 84-93.

Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., Winker, H., 2017c. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Can. J. Fish. Aquat. Sci. 74, 1794–1807.

Tremblay-Boyer, L., Pilling, G., 2017. Geostatistical analyses of operational longline CPUE data. Technical report SA-WP-03 to the 13th Scientific Committee of the Western and Central Pacific Fisheries Commission. Rarotonga, Cook Islands, August 9—17th 2017. https://www.wcpfc.int/meetings/sc13

Tremblay-Boyer, L., Hampton, J., McKechnie, S., Pilling, G., 2018. Stock assessment of South Pacific albacore tuna. Technical report SA-WP-05 to the 14th Scientific Committee of the Western and Central Pacific Fisheries Commission. Busan, Republic of Korea, August 8—16th 2018. https://www.wcpfc.int/meetings/14th-regular-session-scientific-committee

Walters, C., 2003. Folly and fantasy in the analysis of spatial catch rate data. Can. J. Fish. Aquat. Sci 60, 1433–1436.

Waterhouse, L., Sampson, D. B., Maunder, M. Semmens, B. X., 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fish. Res. 158, 15-25.

Wood, S.N., 2006. Generalized Additive Models. An Introduction with R. Chapman and Hall.

Wood, S.N., Bravington, M.V., Hedley, S.L., 2008. Soap film smoothing. J. Royal Stat. Soc., Series B 70, 931-955.

Xu, H. Lennert-Cody, C.E., Maunder, M.N., Minte-Vera, C.V., 2019a. Spatiotemporal dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. Fish. Res. 213, 121–131

Xu, H., Thorson, J.T., Methot, R.D., Taylor, I.G., 2019b. A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. Can. J. Fish. Aquat. Sci. 76, 268–285. https://doi.org/10.1139/cjfas-2017-0446

Xu, H., Thorson, J.T., Methot, R.D., 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. Can. J. Fish. Aquat. Sci. 77, 247–263. https://doi.org/10.1139/cjfas-2019-0107

Zhou, S., Campbell, R.A., Hoyle, S.D., 2019. Catch per unit effort standardization using spatio-temporal models for Australia's Eastern Tuna and Billfish Fishery. ICES J. Mar. Sci. 76, 1489–1504.

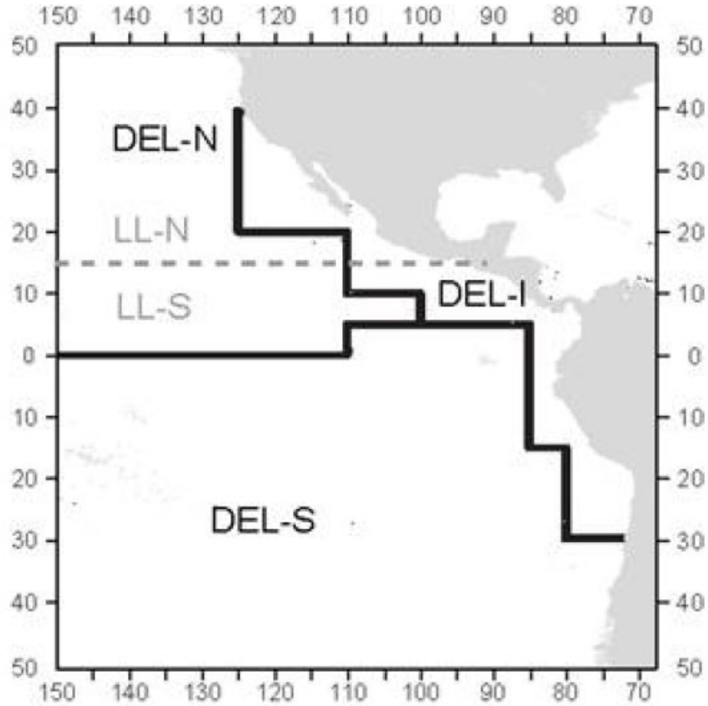


**Figure 1.** Dolphin associated purse seine fishery (DEL) and longline (LL) spatial strata definitions for the fisheries north (N), inshore (I) and south (S).

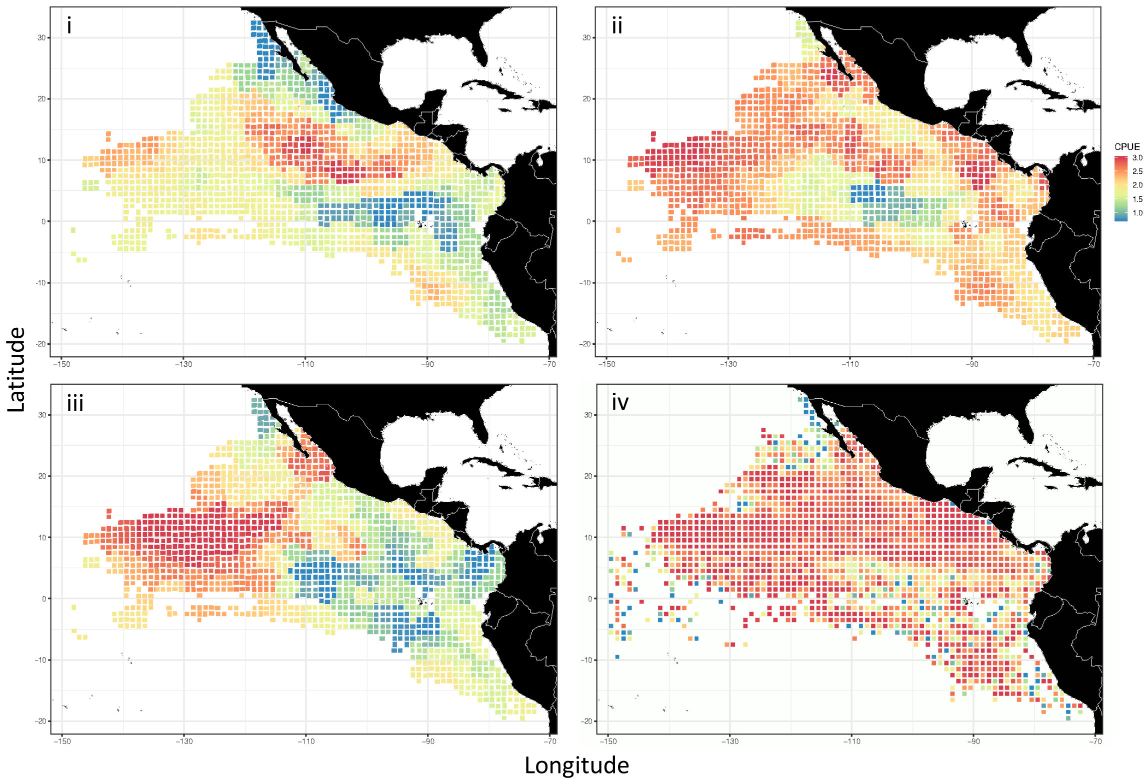


**Figure 2a**. Predicted logarithm of CPUE in tons per day fished for i) an El Nino period (fourth quarter of 1997) ii) a La Nina period (fourth quarter 1998) and iii) a neutral period (fourth quarter of 2003), and iv) logarithm of the observed average nominal CPUE over all years (1975-2016).

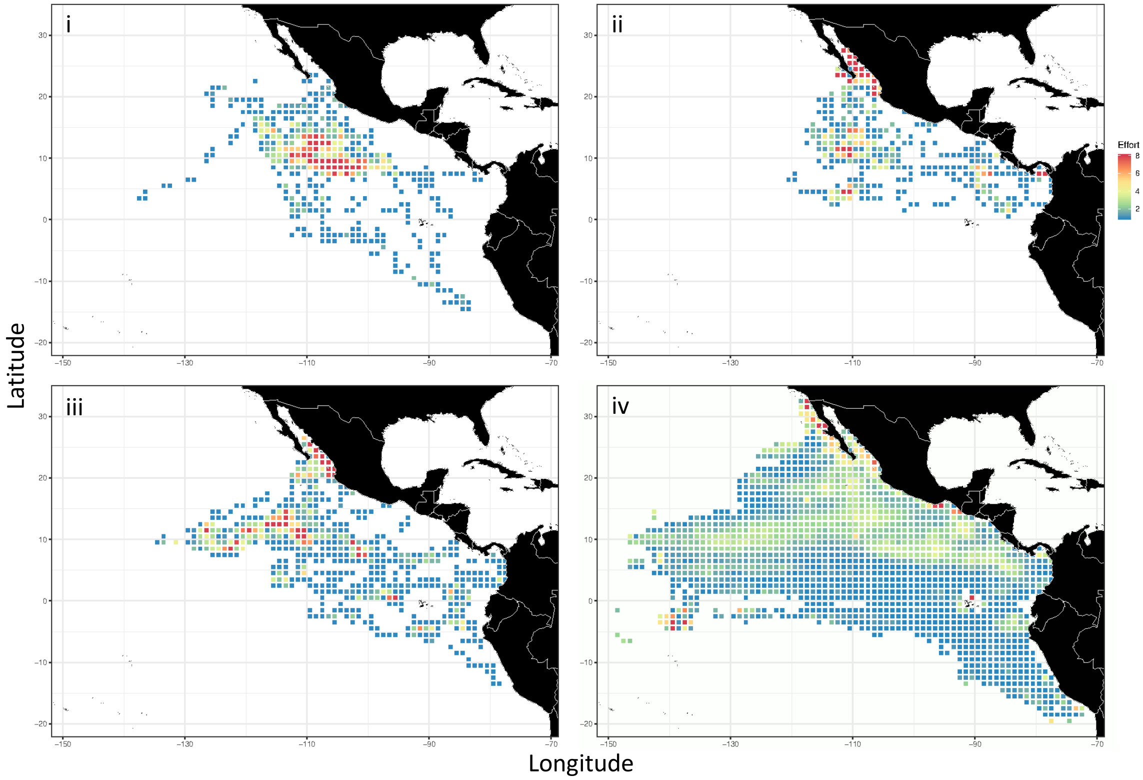


**Figure 2b**. Effort in days fished for i) an El Nino period (fourth quarter of 1997) ii) a La Nina period (fourth quarter 1998) and iii) a neutral period (fourth quarter of 2003), and iv) the average over all years (1975-2016).

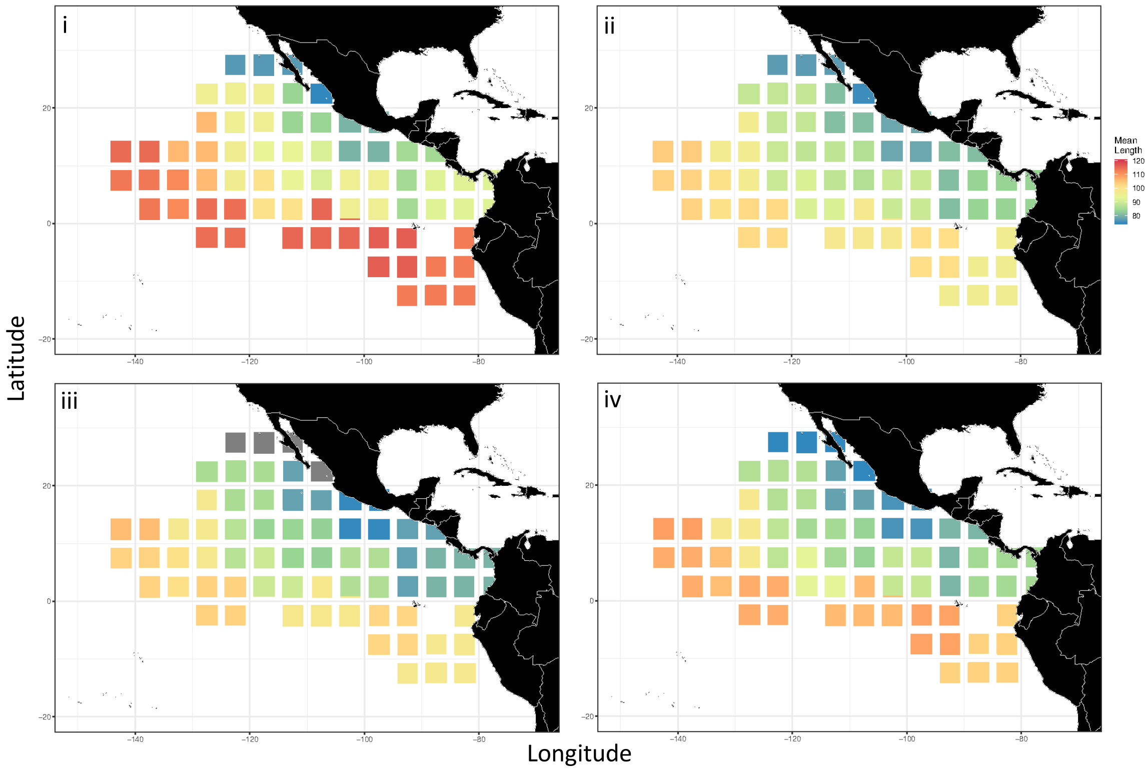


**Figure 2c**. Predicted mean length (cm) for i) an El Nino period (fourth quarter of 1997) ii) a La Nina period (fourth quarter 1998) and iii) a neutral period (fourth quarter of 2003), and iv) the observed average mean length over all years (1975-2016).

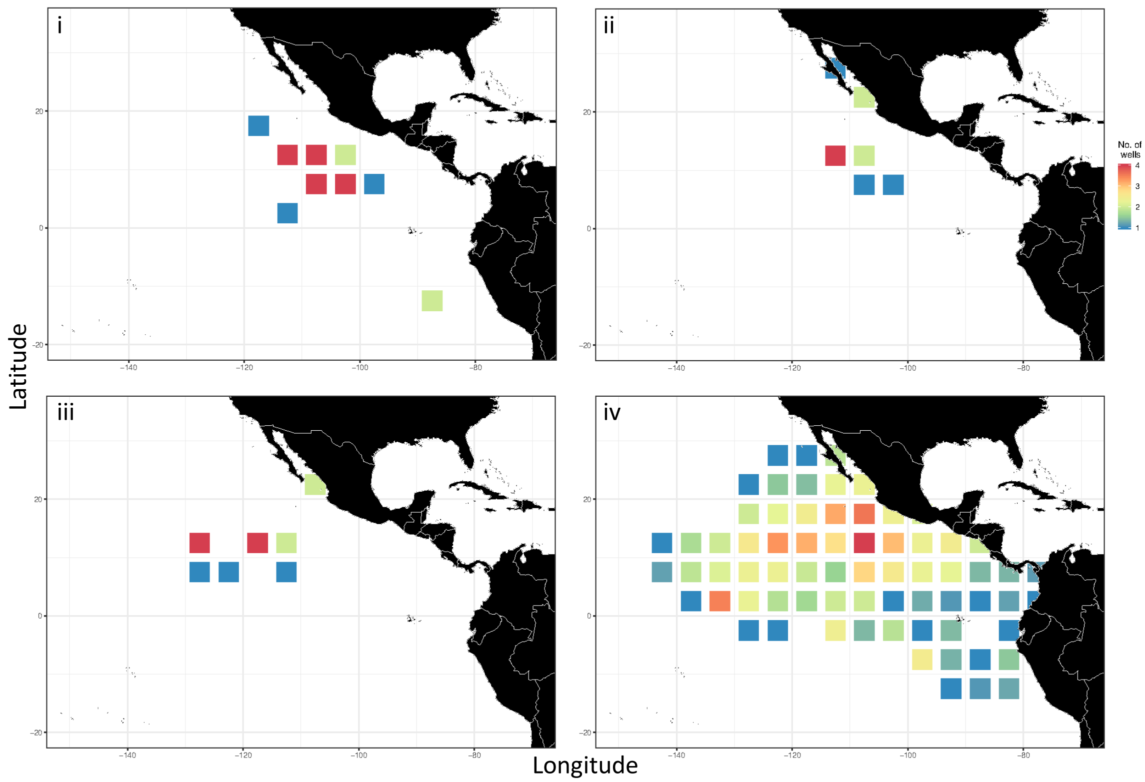


**Figure 2d**. Number of wells sampled for length composition for i) an El Nino period (fourth quarter of 1997) ii) a La Nina period (fourth quarter 1998) and iii) a neutral period (fourth quarter of 2003), and iv) the average over all years (1975-2016).

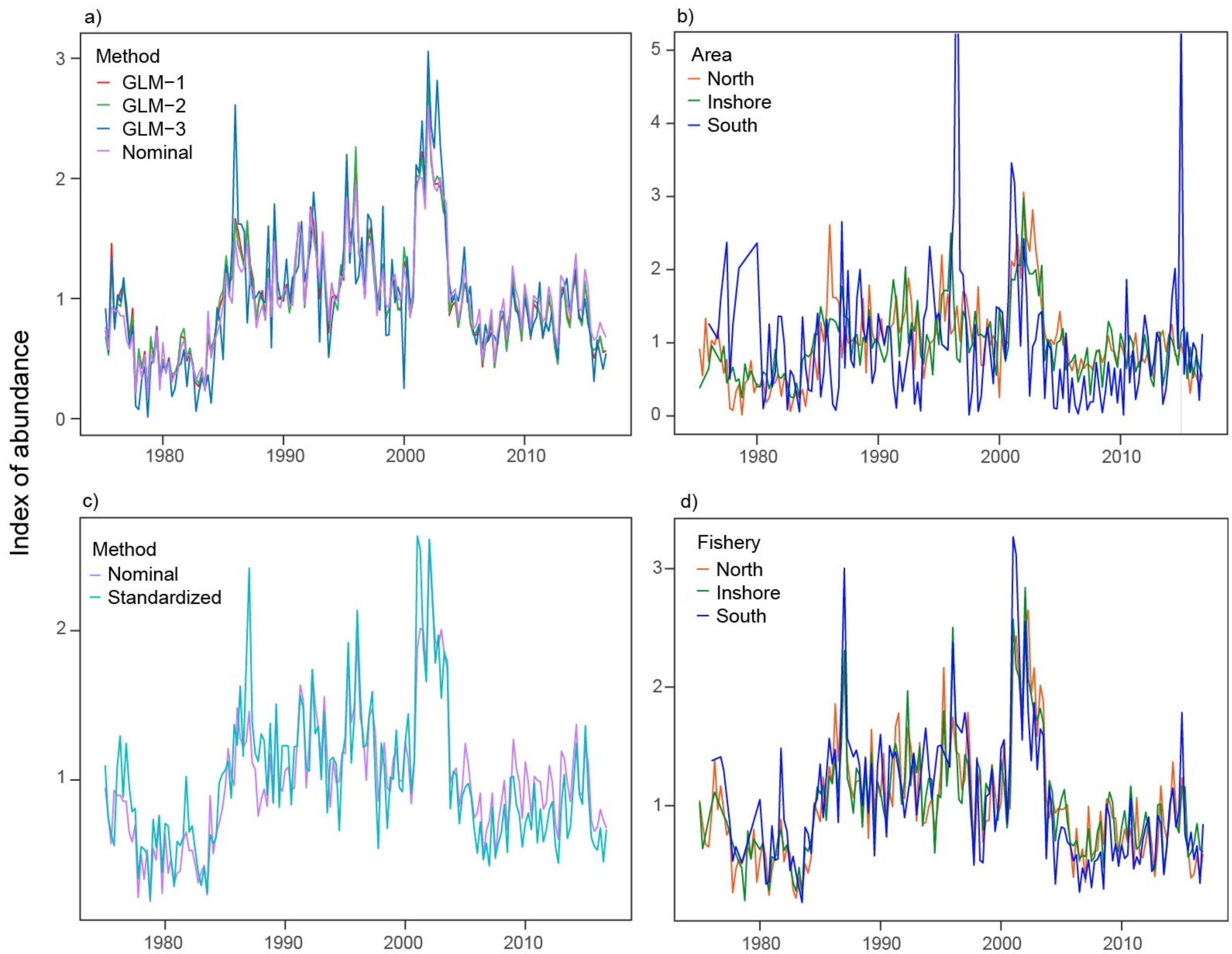


**Figure 3.** a) Indices of abundance from the GLM analyses compared to the nominal index; b) GLM-based indices of abundance for each spatial stratum; c) the index of abundance for the EPO from the spatio-temporal model compared to the nominal index; and d) the index of abundance for the three spatial strata (fisheries) from the spatio-temporal model.

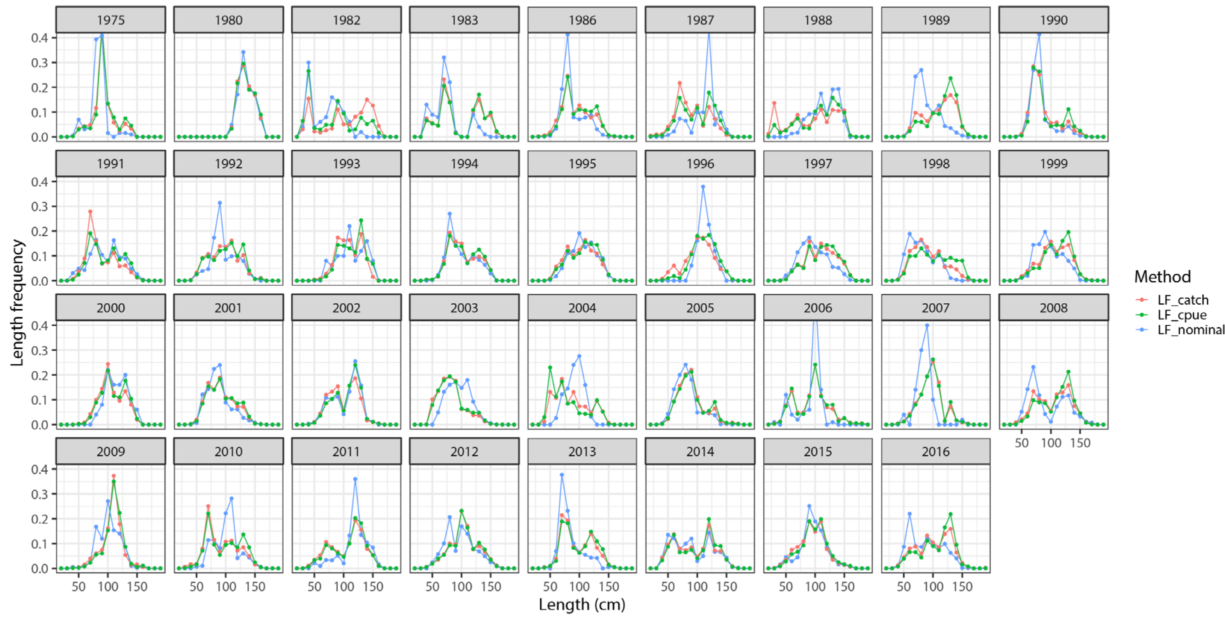


**Figure 4**. Comparison of the first quarter of each year of spatio-temporal model-based length-compositions from catch and CPUE area based on weighting with nominal length compositions for fishery 1. Catch weighting means weighting each composition data in a cell and time period by the catch. CPUE weighting means weighting each composition data in a cell and time period by the CPUE.

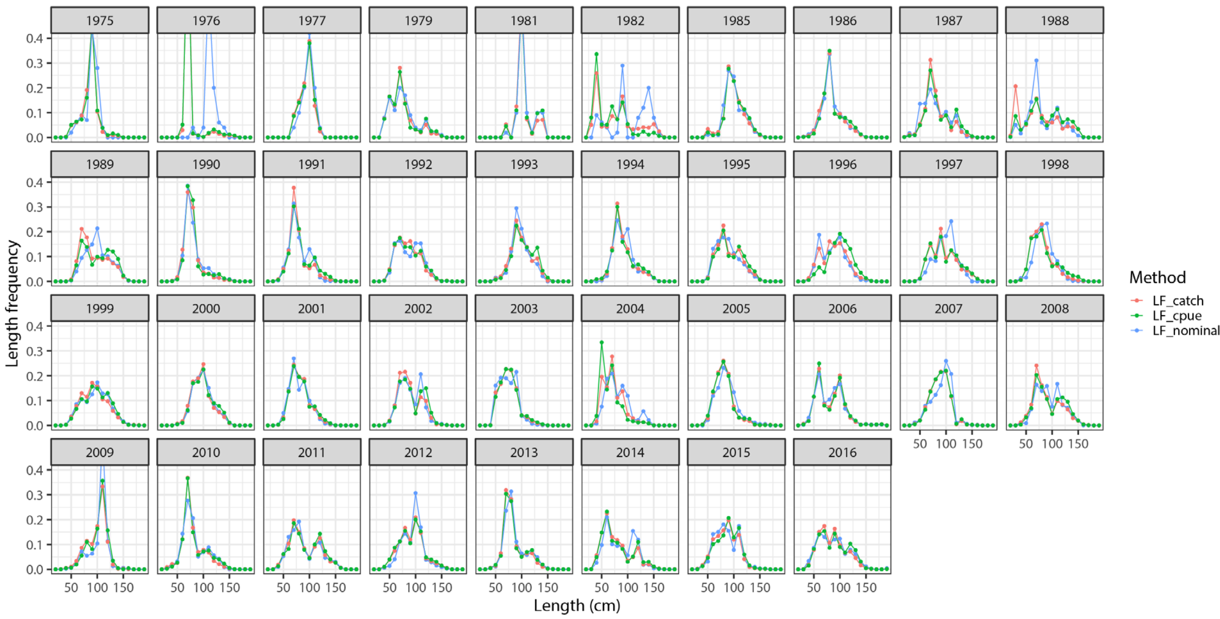


**Figure 5**. As for Fig. 3, except the results are for fishery 2.

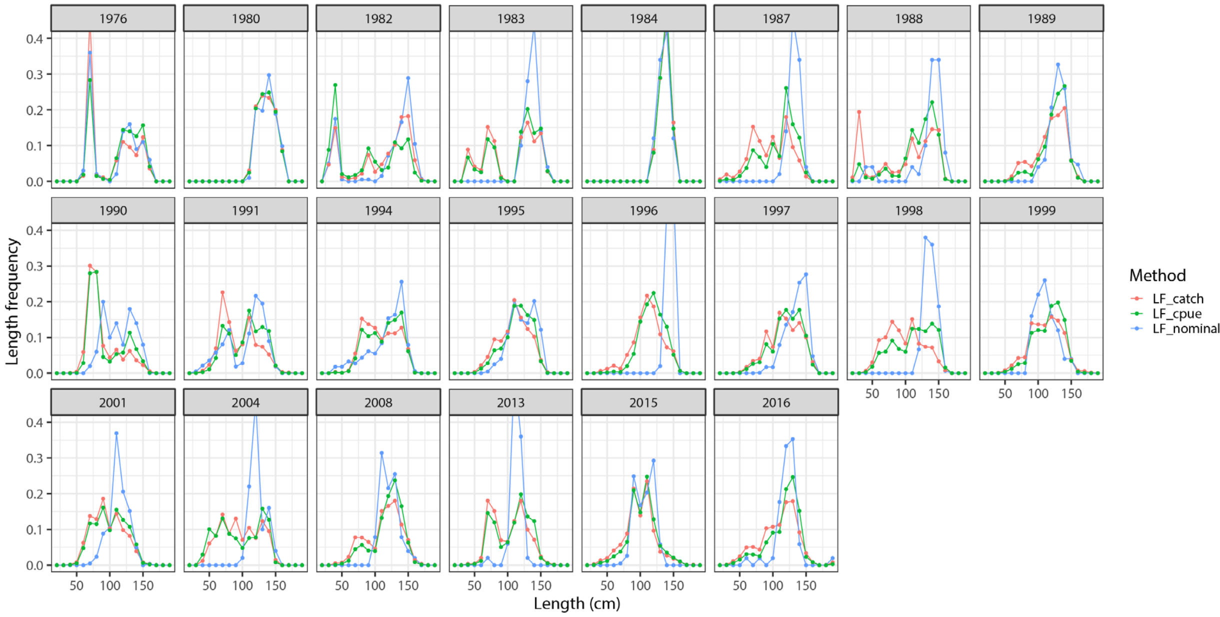


**Figure 6**. As for Fig. 3, except the results are for fishery 4.

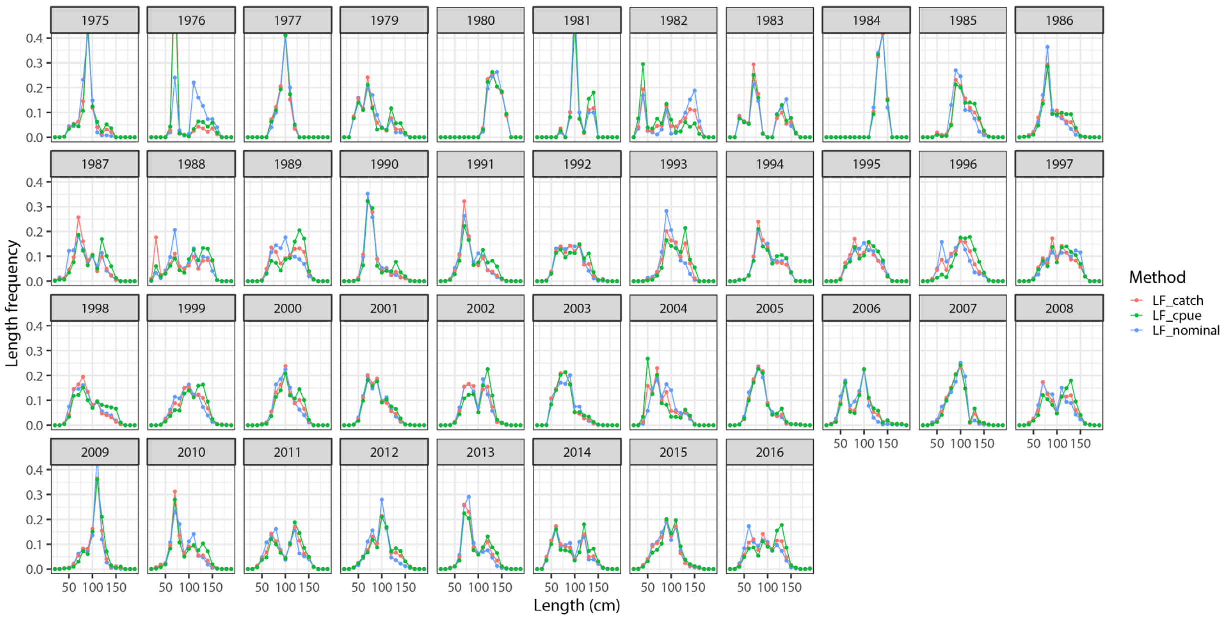


**Figure 7**. As for Fig. 3, except the results are for the entire EPO.

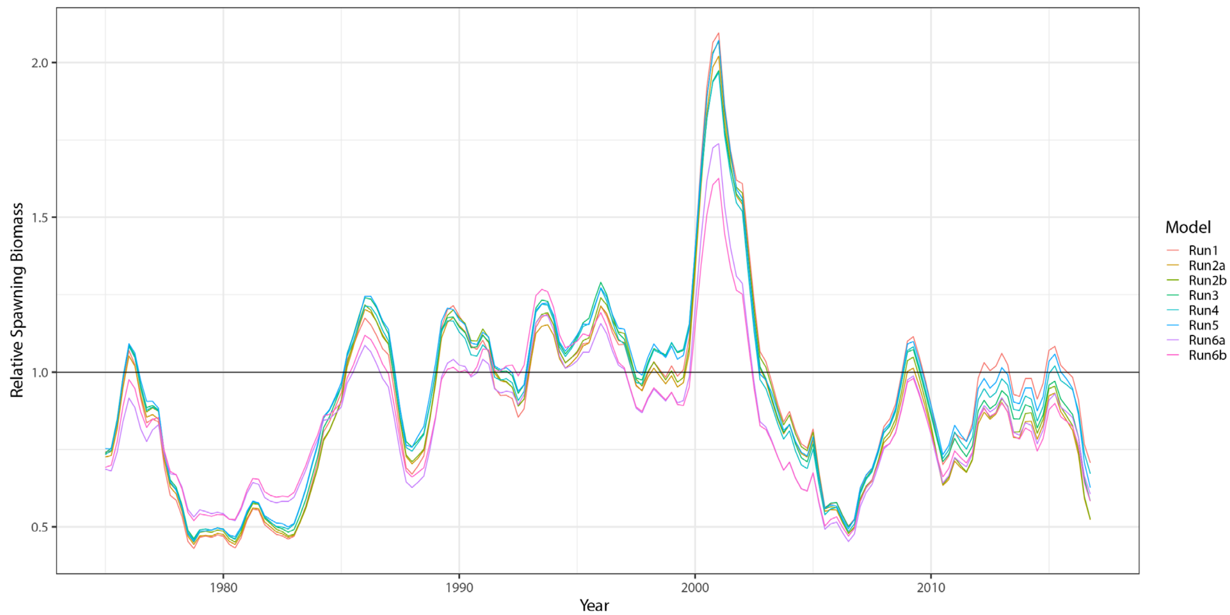


**Figure 8.** Spawning Biomass Ratio (SBR) for the models. The models are defined in section 3.4.1.

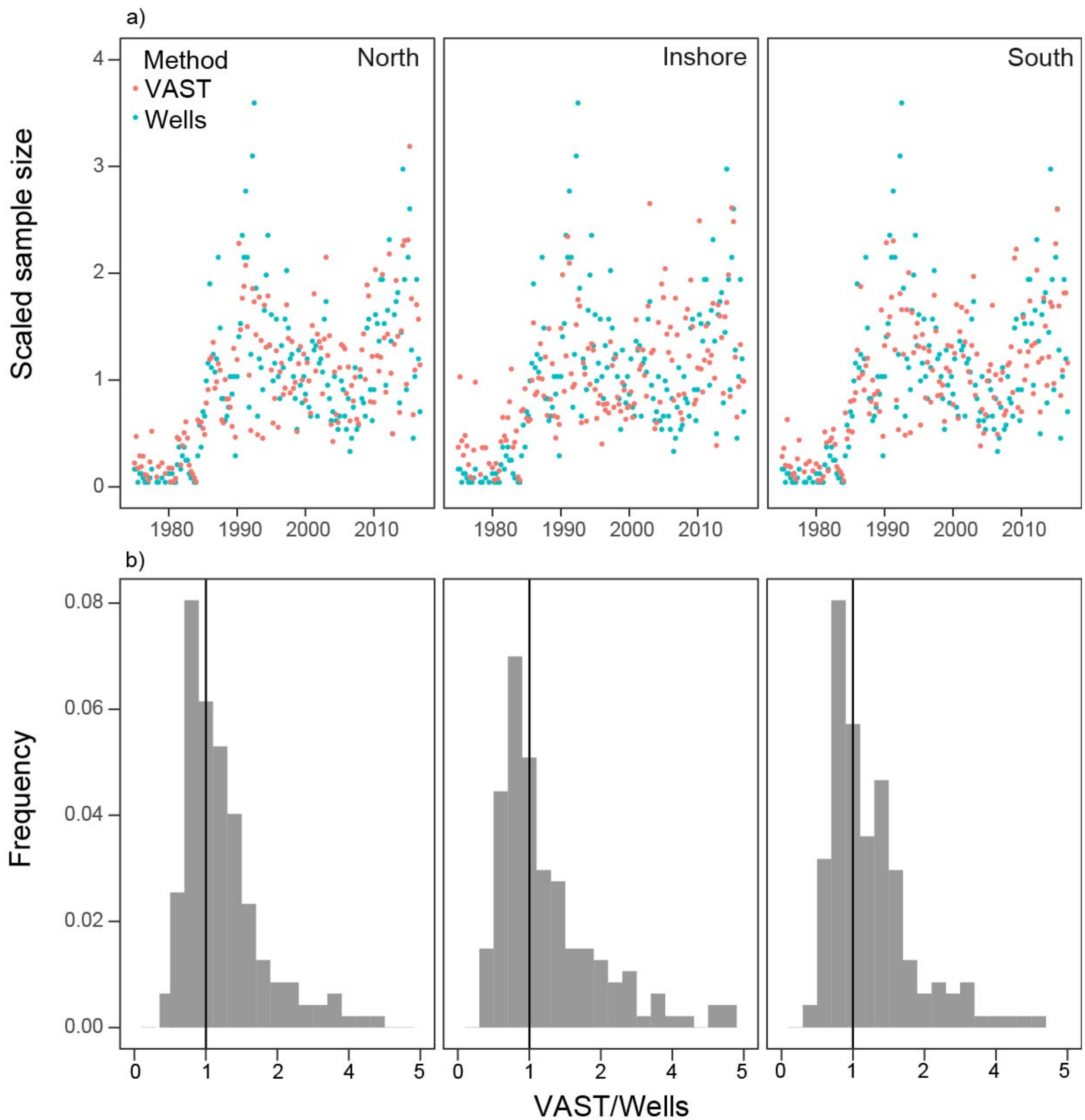


**Figure 9.** a) Relative sample size estimated from the spatio-temporal model (VAST) compared to the number of wells (Wells). b) Frequency distributions for the ratio of the sample size estimated from the spatio-temporal model to the number of wells.

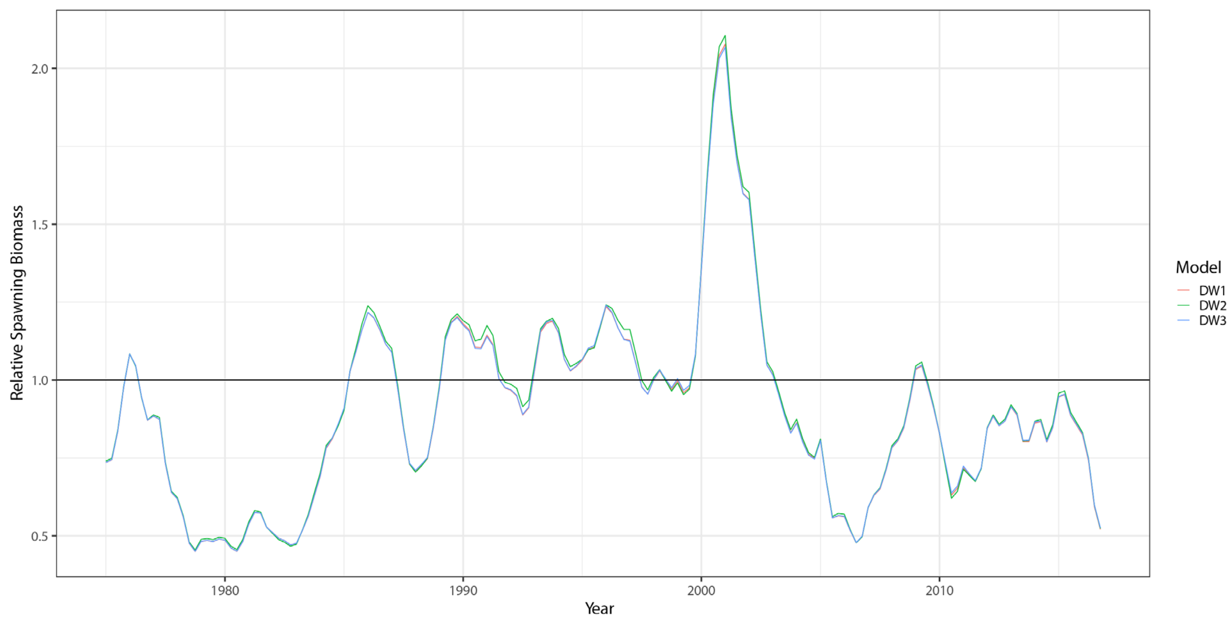


**Figure 10.** SBR for the data weigting scenarios. The data-weighting scenarios are defined in section 3.4.1.

Table 1. Relevant features of the yellowfin stock assessment application.

| Feature | Reference analysis | Changed in analyses |
|---|---|---|
| Modelling framework | Stock Synthesis | |
| Time span | 1975 to 2018 | |
| Maximum age | Plus group at 7.25 (29 quarters) | |
| Time step | Quarterly | |
| Number of fisheries | Sixteen. Defined as a combination of gear used, set type, and area of operation. | |
| Number of indices of abundance | 3 (Sothern longline, Northern dolphin associated purse seine, and inshore dolphin associated purse seine fishery). Southern longline fishery index has a change in catchability and selectivity in 2010. | Yes – the dolphin associated purse seine fishery indices are the focus of this research. Some runs combine the data into a single dolphin associated purse seine index |
| Natural mortality | Age and sex structured, fixed | |
| Growth | Richards growth curve, fixed | |
| Stock recruitment function | Constant with lognormal deviates, penalized likelihood framework | |
| Recruitment standard deviation | Fixed at 0.6 | |
| Selectivity | Dome shaped for all fisheries except the southern longline fishery, estimated for most fisheries and the index | |
| Index of abundance | Sothern longline based on a GLM analysis, dolphin associated purse sein indices based on nominal data | Yes – Different scenarios are investigated |
| Index of abundance composition data | Data weighted | Yes – area stratified and weighted by CPUE or catch |
| Fishery composition data | Data weighted | Yes – area stratified and weighted by catch |

Table 2. Example applications using spatio-temporal models to standardize survey or CPUE data. VAST includes VAST or its precursors (R packages SpatialDeltaGLMM or SpatialDFA). t = time, s = space, v = smoothness parameter.

| **Reference** | **Model type** | **Software** | **Model structure t + s + t*s** | **Spatial covariance** | **Temporal covariance (in t*s)** | **Spatial main effects** | **temporal main effects** | **Stock** | **Other** |
|---|---|---|---|---|---|---|---|---|---|
| Cao et al. (2017) | Delta-GLMM | VAST | t + s + t*s | Matern (v=1) anisotrophy | None | GMRF | Categorical | Northern shrimp (*Pandalus borealis*) in the Gulf of Maine | Survey data, compared to design based methods |
| Cavieres and Nicolis )2018) | Bayesian GLMM | INLA | t + s | Matern | NA | GMRF | Categorical/rand om walk | Yellow squat lobster (*Cervimunida johni*) off Chile | |
| Gruss and Thorson (2019) | Delta-GLMM | VAST | t + s + t*s | Matern (v=1) anisotrophy | None | GMRF | categorical | Gulf of Mexica red snapper (*Lutjanus campechanus*) | Fits to biomass, count, and presence-absence data. |
| Gruss et al. (2019) | Delta-GLMM | VAST | t + s + t*s | Matern (v=1) anisotrophy | None | GMRF | Categorical | Atlantic blue Marlin (*Makaira nigricans*) | Compared with simulated data to GLM, GLMM, GAM, with/without area*year interaction, area weighting, |
| Kai (2019 ) | Delta-GLMM | TMB | t + s + t*s | Matern (v=1) anisotrophy | AR1 | GMRF | Categorical | Blue shark and shortfin mako in the north Pacific | Uses a quarterly effect, but does not allow a random effect over quarter x year |
| Lewy and Kristensen (2009) | Log Gaussian Cox Process (LGCP) (multivariate Poisson-Lognormal distribution) | R | s | Exponential with estimated nugget | Each year analyzed independentl y | Second degree polynomials for latitude and longitude | Each year analyzed independently | Cod (*Gadus morhua*) in the North Sea and the Skagerrak | Conducted separately for age groups 1, 2, 3+ and for each year |
| Thorson and Barnett (2017) | Delta-GLMM factor analysis for species | VAST | t+t*s | Matern (v=1) anisotrophy | None | None | Categorical | US Pacific coast rockfish | Uses factor analysis to model correlations among species, a similar approach can be used for age/length |
| Thorson *et al.* (2015a) | Delta-GLMM | VAST | t + s + t*s | Matern (v=1) anisotrophy | None | GMRF | Categorical | 28 groundfish species off the U.S. West Coast | Survey data, compared to design based methods |
| Tremblay-Boyer and Pilling (2017) | Delta-GLMM | | t + s + t*s | Matern (v=1) anisotrophy | None | GMRF | Categorical | Bigeye (*Thunnus obesus*) and yellowfin tuna (*Thunnus albacares*) in the Western and Central Pacific Ocean | |
| Tremblay-Boyer | Delta-GLMM | | t + s + t*s | Matern (v=1) | None | GMRF | Categorical | Albacore tuna | Indices included in the |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| et al. (2018) | | | | anisotrophy | | | | (*Thunnus alalunga*) in the South Pacific | stock assessment for index fisheries |
| Xu et al. (2019a) | Delta-GLMM | VAST | t + s + t*s | Matern (v=1) anisotrophy | None | GMRF | Categorical | Yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean | |
| Zhou et al. (2019) | Bayesian Delta-GLMM | INLA | t + t*s<br>OR<br>t + s | Matern | AR1<br>OR<br>None | None<br>OR<br>GMRF | Categorical | Australia's Eastern Tuna and Billfish Fishery | Compares with GLM and GAM with/without area*year interactions |

Table 3. Example applications using spatio-temporal models to standardize age or size composition data. VAST includes VAST or its precursors. t = time, s = space, l = size (length) v = smoothness parameter.

| **Reference** | **Model type** | **Software** | **Model structure t + s + l + t*s + t*l + s*l + t*s*l** | **Spatial covariance** | **Temporal covariance (in t*s, s*l, or t*s*l)** | **Age or size covariance (in t*l, s*l, or t*s*l)** | **Spatial main effects** | **temporal main effects** | **Age or size main effects** | **Stock** | **Other** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kristensen et al. (2014) | Log Gaussian Cox Process (LGCP) (multivariate Poisson-Lognormal distribution) | 'lgc package | s + s*t + s*t*l | GMRF | Exponential decay | Exponential decay | GMRF | Population dynamics model | Population dynamics model | Cod (*Gadus morhua*) | Includes growth, mortality, and reproduction |
| Nielsen et al. (2014), | Log Gaussian Cox Process (LGCP) | TMB | l + s*l | GMRF | NA | Periodic, log, and logistic | None | NA | Categorical | Cod (*Gadus morhua*) and whiting (*Merlangius merlangus*) | Also looked at between species correlation |
| Kai et al. (2017b) | Delta-GLMM | TMB | t + s + l + t*s*l | Matern (v=1) anisotrophy | AR1 | AR1 | GMRF | Categorical | AR1 | Shortfin mako shark (*Isurus oxyrinchus*) in the north Pacific | |
| Perretti and Thorson (2019) | Delta-GLMM | VAST | t*l + s*l + t*s*l | Matern (v=1) anisotrophy | None | Random walk | GMRF | Random walk over t for each l | Random walk over t for each l | Summer flounder (*Paralichthys dentatus*) | Fitted to two separate data sets |
| Thorson and Haltuch (2018) | Delta-GLMM | VAST | t*l + s*l + t*s*l | Matern (v=1) anisotrophy | None | None | GMRF | Categorical (t*l) | Categorical (t*l) | Lingcod (*Ophiodon elongatus*) in the California Current | Separate spatial variances by length, but same decorrelation distance parameter |

**Table 4.** Management quantities estimated by the stock assessment model for the scenarios. MSY = maximum sustainable yield, $S_{recent}$ is the spawning biomass at the start of 2017, $S_{MSY}$ is the spawning biomass associated with MSY, $F_{multiplier}$ is the amount the current fishing mortality (averaged over 2014-2016) would have to be increased to equal the fishing mortality corresponding to MSY ($F_{MSY}/F_{recent}$).

| Quantity | Run1 | Run2a | Run2b | Run3 | Run4 | Run5 | Run6a | Run6b |
|---|---|---|---|---|---|---|---|---|
| MSY (t) | 264,903 | 264,103 | 263,313 | 266,032 | 266,467 | 272,280 | 269,189 | 270,640 |
| $S_{recent}/S_{MSY}$ | 0.87 | 0.71 | 0.65 | 0.77 | 0.86 | 0.73 | 0.87 | 0.84 |
| $F_{multiplier}$ | 1.15 | 1.08 | 1 | 1.12 | 1.15 | 1.17 | 1.19 | 1.16 |