

INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

EIGHTH MEETING

La Jolla, California (USA)

8-12 May 2017

DOCUMENT SAC-08-05d

THE NEED FOR SPATIAL-TEMPORAL MODELING OF CATCH-PER-UNIT-EFFORT DATA WHEN USED TO DERIVED INDICES OF RELATIVE ABUNDANCE TO INCLUDE IN STOCK ASSESSMENT MODELS

Maunder, M.N., Thorson, J.T., Lee, H.H., Kai, M., Chang, S.K., Kitakado, T., Albertsen, C.M., Lennert-Cody, C.E., Aires-da-Silva, A.M., Piner, K.R.

CONTENTS

Abstract..... 1

1. Introduction 2

2. The issues..... 3

2.1. Spatial weighting..... 3

2.2. Composition data..... 4

3. Pacific bluefin tuna 7

3.1. Fishing effort distribution 8

3.2. Abundance (CPUE) distribution 8

3.3. Targeting..... 8

3.4. Correcting for targeting 9

3.5. Size composition 9

4. Suggested PBF modelling..... 9

5. Other methods..... 11

6. Other applications 12

References 13

ABSTRACT

Standardizing for spatio-temporal distribution of the fishing effort has been a main issue in catch-per-unit-effort (CPUE) analysis. Of particular concern is the change in spatial distribution over time due to movement of the stock, recruitment dynamics, or local depletion. A simple Generalized Linear Model (GLM) including area as a factor and no interaction terms assumes that the year effect (or relative year to year variability in catch rates), which is assumed to represent relative abundance, is the same in each area and only the average catch rates differ among areas. The assumptions underlying this model can cause bias in the estimated index of relative abundance if the stock or fishery spatial distributions change over time. In addition, a somewhat overlooked component of using indices of relative abundance in stock assessment models is the component of the population that is represented by the index with respect to age or size. Typically, this is modelled using a selectivity curve that is estimated by fitting to composition data. The selectivity curve represents 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 gear. However,

“selectivity” in the stock assessment model does not simply represent contact selectivity, but also represents availability, which is a consequence of the spatial structure of the fleet relative to the stock. Therefore, due to the index representing abundance in each area, the fishery catch representing catch in each area, and catch not necessarily being spatially distributed proportional to abundance, when the composition differs among areas then the “selectivity” in the stock assessment differs between the index and the catch. The index and the composition data should both be derived using the same spatio-temporal model. We outline the use of spatio-temporal models for standardizing CPUE and composition data, and discuss the issues using North Pacific bluefin tuna as an example.

1. INTRODUCTION

Indices of relative abundance are an integral part of most stock assessments. Preferably, they would be based on sampling data from well-designed surveys and analyzed such that they are proportional to abundance while maximizing precision. Unfortunately, surveys are not possible to design or implement for all stocks due to logistical and funding issues. Therefore, many stock assessments, like 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 a number of factors that may invalidate the assumption that they are proportional to abundance (Harley *et al.* 2001; Maunder *et al.* 2006a, Thorson *et al.* In press).

There is a huge body of literature describing alternative approaches to “standardize” CPUE and minimize the factors other than abundance that influence CPUE (Maunder and Punt 2004). Typically, the CPUE is standardized using a Generalized Linear Model (GLM) or similar method for factors such as season, location, vessel and gear characteristics, and the environment. There are also numerous examples of more sophisticated approaches that deal with particular issues or increase the complexity of a component of the modelling. 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. Their approach was extended to a statistical framework by Maunder *et al.* (2006b). Many authors have also focused on the fine scale spatial distribution of the CPUE (Walters 2003, Carruthers 2011, Thorson *et al.* In press), while others used more broad scale spatial strata to standardize CPUE.

Standardizing for spatial distribution of the fishing effort has been a main issue in CPUE analysis. Of particular concern is the change in spatial distribution over time due to movement of the stock, recruitment dynamics, or local depletion. A simple GLM including area as a factor and no interaction terms assumes that the year effect (or relative year to year variability in catch rates), which is assumed to represent relative abundance, is the same in each area and just the average catch rates differ among areas. The assumptions underlying this model can cause bias in the estimated index of relative abundance for a number of reasons including if the stock spatial distribution changes over time. This effect can typically be identified when an interaction term between area and year is statistically significant and the time series of year effects from different areas show visual differences in trends. In general, statistically significant interaction terms between year and another variable that result in meaningful differences are problematic because calculating the index requires specifying a value for the variable interacting with year. If the interaction is treated as a random effect, calculating the index similarly 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 while summing across the spatial factor with weighting equal to the area associated with each level of the spatial factor. In extreme cases of differences in the year effect among areas, the stock may be modelled as multiple independent or interacting populations and each population fit to its respective areas index of relative abundance.

A somewhat overlooked component of using indices of relative abundance in stock assessment models is the component of the population that is represented by the index with respect to age or size. Typically, this is modelled using a selectivity curve that is estimated by fitting to composition data. The selectivity curve represents 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 gear. However, selectivity in the stock assessment model does not simply represent contact selectivity, but also represents 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). Therefore, due to the index representing abundance in each area, the fishery catch representing catch in each area, and the spatial distribution of the catch differs from the abundance, when the composition differs among areas, the selectivity in the stock assessment differs between the index and the catch. In general, the index selectivity will represent the contact selectivity while the catch selectivity will represent both contact selectivity and availability and will change over time as the fishery or stock distribution changes over time.

Here we discuss in general the use of spatial-temporal models to deal with spatial changes in the fishery and stock, and to standardize composition data. Then we discuss how this applies to Pacific Bluefin tuna to further illustrate the approach.

2. THE ISSUES

2.1. Spatial weighting

Dealing with spatial changes in the fishery and stock when developing indices of relative abundance is an important component of CPUE standardization. Ideally the abundance in each area should be calculated and summed to get the overall abundance. This can be thought of as a form of area weighting of the data. This contrasts with commonly used approaches to standardize CPUE data like naive use of GLMs, which can be considered data weighted in the sense that each data point is given equal weight independent of what area it comes from. Areas that have more data will be given more weight in the analysis. A consequence of area weighting is that large areas with small sample sizes might overwhelm small areas and will have similar influence as large areas with large sample sizes. Abundance estimates from these large areas with small sample sizes may have high variance and may overwhelm the estimates of total abundance with uncertainty. However, this increased uncertainty will often reflect uncertainty about population density in those areas, which is appropriate when data are not available for large segments of the population's habitat (*e.g.*, Walters 2003).

A method is needed to improve the estimates of areas with low sample size and also impute abundance for areas with no samples, particularly if the spatial stratification is based on a fine scale. Contemporary spatio-temporal models are ideal for this purpose and recent developments in statistical methodology, computational algorithms, and software packages (*e.g.* Kristensen *et al.* 2016) make them practical. These models are based on the assumption that catch rates in nearby areas should be similar. Spatio-temporal models can be configured to also share information among similar time periods (*e.g.*, using autocorrelation, Thorson *et al.* 2016). Spatio-temporal models generally estimate the degree of information-sharing between nearby locations based on the estimated "correlation function", and this allows the model to incorporate either strong or weak smoothing for data that are close in space and time.

Generalized linear mixed models with spatio-temporal effects have become the state of the art in modelling spatial data (Lewy and Kristensen 2009; Kristensen *et al.* 2014; Nielsen *et al.* 2014; Thorson *et al.* 2015). Kai *et al.* (2017a) present a spatio-temporal model for shark CPUE and much of the following comes from their description. Spatio-temporal models share information through the estimated correlation function of the spatio-temporal random effects to predict density at unsampled locations and

times, and improve estimates at location and times with low sample size. Space and time are modeled both as main effects as well as an interaction term between space and time. The spatial components (main effect and spatio-temporal interaction terms) are modeled as random effects, which allows the sharing of information, and 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.*, 2015b). The spatial-temporal interaction term is modeled by combining the GRF for the spatial component with first-order autoregressive process for temporal component at each location. Seasonal spatial effects are also often modelled (*e.g.* Kai *et al.* 2017a).

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

$$d(s, t) = \exp\left(d_0(t) + \gamma(s) + \theta(s, t) + \sum_{j=1}^{n_j} \beta_j x_j(s, t)\right), \quad (1)$$

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 β_j represents the impact of covariate j with value $x_j(s, t)$ on density at station s and time t .

Spatial variation $\gamma(s)$ is modeled using a GRF and reduces to a multivariate normal distribution when evaluated at a finite set of stations (Thorson *et al.*, 2015c). The Matérn correlation function is used for computational efficiency (Diggle and Ribeiro, 2007; Roa-Ureta and Niklitschek, 2007; Lindgren *et al.*, 2011). The spatial-temporal interaction term, $\theta(s, t)$, is modeled by combining the GRF for the spatial component with a first-order autoregressive model for the temporal effect at each site.

Expected catch c_i^* is the product of fish density, as represented by the spatio-temporal model, and fishing effort f_i , $c_i^* = d(s_i, t_i)f_i$, and is fit to the observed catch c_i for the i -th observation, in station s_i and time t_i , using a likelihood function (*e.g.* log-normal, negative-binomial, or a zero-inflated model). Covariates $\hat{x}_{k,i}$ for each data point i and covariate k can be added to model catchability (*e.g.* gear effects) $c_i^* = d(s_i, t_i)f_i \sum_{k=1}^{n_k} \hat{\beta}_k \hat{x}_{k,i}$. The parameters are estimated by maximizing the likelihood function while integrating across the random effects, which represent the spatial and spatio-temporal variations, using Template Model Builder (TMB). TMB is an R package (R Core Team, 2013) that efficiently fits latent variable models to data (<http://folk.uib.no/hsk021/tmbdoc/index.html>; Kristensen *et al.* 2016; Thorson *et al.* 2015b, c), through the use of the Laplace approximation for integration and automatic differentiation for calculating derivatives. The estimated parameters include those representing the temporal main effect (*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.

The index of relative abundance is calculated by summing up the predicted density for each location in a time period. Care needs to be taken to identify which factors effect catchability and should not be used to estimate density, but are still used to calculate the expected catch used in the likelihood function, and those that effect density and are used to calculate the quantities that are summed to generate the index of relative abundance. The bias-correction algorithm to account for retransformation bias when predicting and visualizing total abundance and size composition (Thorson and Kristensen 2016) should be used when appropriate.

2.2. Composition data

Catch composition data (*e.g.* length composition) are a key component of using CPUE based indices of abundance in stock assessments because they provide information on the portion of the population

represented by the index with respect to age or size. However, the composition data is typically used in its raw form by simply summing up all the samples, each sample possibly weighted by the corresponding catch. Weighting by the catch makes sense because the composition data is also used to inform the age or size distribution of the catch removed from the fishery. Although, there is the issue of large catches represented by small sample sizes inflating the overall uncertainty. Simply summing up the composition data implicitly assumes that each sample is a random sample of the population (*e.g.* the fish are randomly distributed throughout the whole area). However, this is unlikely because many species show heterogeneity in age or size spatially. Composition data from surveys are generally raised to the catch (or catch rate) in the survey and to the strata (station) size. This is appropriate because surveys are generally designed to estimate abundance and the composition data do not need to be used to estimate the age or size structure of the catch being removed.

The process of raising compositional data can be interpreted as one step towards a model-based framework for “standardizing” compositional data (Thorson 2014). Standardizing compositional data has several potential benefits including:

1. Raising compositional data to mean catch rates in areas with low sample sizes, based on model-estimates of population abundance in those areas;
2. Accounting for confounding factors (*e.g.*, vessel type or season) when interpreting compositional samples to estimate proportions for each category (*e.g.* length or age);
3. Estimating effective sample sizes based on the variance in compositional sampling data.

In particular, the spatial-temporal modelling approach described above could be modified to include composition information (see also Kristensen *et al.* 2014; Nielsen *et al.* 2014; Thorson *et al.* In press). For example, the model could be applied to each age or size group separately and independent indices of relative abundance could be calculated for each group and used in the stock assessment model. Using independent indices of abundance for each age class has been common in traditional use of virtual population analysis, and is commonly used in state-space age-structured models (the base-model considered in Nielsen and Berg 2014). However, this approach ignores the correlation among age classes within the stock assessment model. In addition, it ignores information that can be gained by assuming that similar ages or sizes should have similar catch rates. In many cases there may not be enough information for an age class and particularly a length class, and combining ages or lengths might be required, which reduces the detail on specific age or sizes to an extent that may reduce information about important population or fishery processes (*e.g.* recruitment or selectivity).

The spatial temporal model that includes the three dimensions (time, latitude, and longitude) can be modified to include a fourth dimension of either age or length (Lewy and Kristensen 2009, Kristensen *et al.* 2009; Nielsen *et al.* 2014; Kai *et al.* 2017b; Thorson *et al.* In press). When using a spatio-temporal model to standardize compositional data to be used in a subsequent assessment, we recommend against including an explicit growth and survival model (*e.g.*, as used in Kristensen *et al.* 2014) because resulting estimates of proportion-at-age or length are used in an assessment model that itself smooths across data using population-dynamics assumptions. One complication with our recommended approach is that it greatly increases the computational demands of the model. Because the composition data is usually not collected for each catch event, the catch and composition data may need to be fit using separate likelihood functions, but simultaneously in the same model.

As one concrete example, Kai *et al.* (2017b) developed a spatio-temporal model that also included the length of the fish caught and much of the following comes from their description. The spatio-temporal model incorporating length data estimates the density

$$d(s, t, l) = \exp\left(d_0(t) + \gamma(s) + \tau(l) + \theta(s, t, l) + \sum_{j=1}^{n_j} \beta_j x_j(s, t, l)\right)$$

where $\tau(l)$ represents the impact of length on expected catch rates, $\theta(s, t, l)$ represents an interaction term of location, time and length, and each covariate can be a function of length, expressed as $x_j(s, t, l)$. The marginal (common to all spatial stations and times) length effect, $\tau(l)$, is modeled using a first-order autoregressive process (AR1) producing a semi-parametric representation of the expected density at different length bins (Thorson *et al.* 2014). The spatio-temporal and length variation, $\theta(s, t, l)$, is modeled by combining the GRF for spatial variation with first-order autoregressive process (AR1) for temporal and for length variation.

Expected catch λ_i is the product of density and fishing effort f_i , $\lambda_i = d(s_i, t_i, q_i, l_i) f_i$, where density is a function of length, and is fitted to the observed catch c_i for the i -th observation, which is in station s_i , year t_i , and length l_i . This likelihood does not separate the catch data from the composition data (see below).

The above approach estimates a multivariate index of relative abundance and it is preferable to fit the index in the stock assessment model using a multivariate likelihood function. However, if the assessment software does not have this capability, the index can either be broken into separate indices for each age or into a total index and an estimate of proportion-at-age or proportion-at-length that is then treated as “compositional data”. The variance in estimates of proportion-at-age or –length could be used to calculate an input sample size (Thorson 2014), and this input sample size could then be down-weighted to represent the impact of model mis-specification (Francis 2017; Thorson *et al.* 2017). The advantage of this spatial method of standardizing catch rates and developing composition data is that it removes issues of spatial variability in fisheries or the size or age composition of the stock on the assumption of time-invariant catchability and selectivity. However this approach still assumes that the gear (contact) component of selectivity is temporally invariant. This approach to deal with spatial patterns in fisheries and fish populations may be more appropriate than other proposed methods (*e.g.* Stewart and Martell 2014).

Length or age compositions describing the population abundance (index) are unlikely to be the same as those describing fishery catches. The above methods define an approach to determine composition data for the index of relative abundance, which is complicated for fishery-dependent CPUE because fishery compositional data are used both to estimate population proportions in each category, and to estimate the selectivity governing fishery removals. For these reasons, it is easiest to apply the described methods using data from a survey. Composition data representing fishery removals should be raised to the total catch by weighting the spatial explicit composition data by the catch. However, this raises two problems. The first is the appropriate weight to give to the composition data. This is a standard problem in contemporary fisheries stock assessment and will not be addressed here (see Francis 2017; Maunder *et al.* 2017; Punt 2017). 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 is a violation of standard statistical practices. However, given the arbitrariness of weighting data 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.

The method used to calculate the catch at size within the spatio-temporal analysis is summing up the product of the predicted catch (or observed catch, if it is assumed to be known with little error) and the predicted composition for each station 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 because some data is discarded to avoid bias when estimating the relative index of abundance, then the calculations need to be conducted using the total effort or the total observed catch. 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. Stewart and Monnahan 2017).

Since the composition associated with the abundance index represents the population, only factors representing the density effects on the composition data should be used in calculating the length structure of the index, and catchability effects should be ignored. However, when calculating the fishery catch length structure both the density and catchability effects should be included.

3. PACIFIC BLUEFIN TUNA

Oshima *et al.* (2012) described how the Japanese longline fishing effort distribution changed over time relative to the spatial distribution of Pacific bluefin tuna (PBF) and suggested that a large portion of the fishery changed from targeting PBF to targeting yellowfin tuna (YFT). Japanese longliners not only target PBF but also YFT and albacore (ALB) in the same season and area. In addition, capture of PBF is rare for many longliners (1.2% of the catch). Oshima *et al.* (2012) suggested that several factors may influence the fishing behavior of longliners: abundance of target species, fish price, fuel price, and distance between fishing grounds and landing ports. For example, a global increase in fuel price starting in 2006 may have changed fishing behavior to limit fuel consumption. Fishing behavior may have also adjusted to account for the tradeoff between the increasing and higher prices of PBF compared to stable lower prices of YFT and the declining abundance levels of PBF to target the more profitable species.

To explore the change in spatial distribution of the Japanese longline fishery, Oshima *et al.* (2012) defined three areas based on spatial distribution of PBF CPUE and longline sets (their Fig. 3). Area 1 is a small area in the southernmost part of the fishery covering a major spawning ground with high PBF CPUE. This area also has high YFT CPUE, but low ALB CPUE. Area 2 is a large area including most of the central component of the fishery and covers a wide range of PBF CPUE. Within this area, the PBF CPUE is highest to the south and further from the coast of Japan, while YFT CPUE is also highest to the south, but closer to the coast of Japan. ALB CPUE is highest to the north. Area 3 is a small area on the northern part of the fishery and has high ALB CPUE, but low PBF and YFT CPUE. In the following, we generally ignore Area 3 as it does not help interpret the fishery with respect to PBF. Oshima *et al.* (2012) suggest that given the differences in CPUE among the species, that the longliners select their fishing ground in accordance with their target, which may be for multiple species.

Similar situations were observed in Taiwanese PBF fishery. PBF were caught mainly by the longline fishery as a seasonal target species during May to June (or to early July) in the eastern waters off Taiwan. At other times of the year, these longliners target mainly yellowfin tuna, as well as bigeye, swordfish and sharks, around Taiwan or the eastern Indian Ocean. During the PBF fishing season, they register to be PBF target vessels while at the same time fish for yellowfin since PBF catch is low with an annual average of 3–15 PBF per vessel (since 2001). The number of vessels has fluctuated substantially depending mainly on fuel cost, market value and catch rate.

Catch of PBF is distributed mainly between 20–26N, 121–126E off Taiwan with two clusters (fishing areas) noted that can be separated by 24.3N (Chang and Liu 2016). PBF catch in the northern area has a wider size range (170–260 cm FL) than the southern area (190–260 cm FL) and is about 25 kg smaller on average. The southern area was the traditional fishing ground with 90%–100% PBF catches coming from this area before 2009. However, the catch from the northern area increased subsequently and almost half of the catch was from the northern area in 2015, indicating a clear spatial change of fishing effort. Even within a fishing area, spatial changes also occurred due to political considerations such as a change in the relationship between Taiwan and the Philippines.

3.1. Fishing effort distribution

Japanese longline fishing effort was highest in area 2 throughout 1994-2011 (see Figure 4 in Oshima *et al.* (2012)). However, it declined after 2006 in Area 2, while increasing in Area 1. PBF catch was dominant in Area 2 until 2005, and then it sharply decreased. Despite having lower effort, catch of PBF in Area 1 has exceeded that in Area 2 since 2007. Consequently, Oshima *et al.* (2012) suggested that the main fishing ground for PBF shifted from Area 2 to Area 1 in 2007 and thereafter. Area 3 has been minor in both number of sets and PBF catch.

As to the fishing grounds off Taiwan, fishing effort in the northern area was lower than 10% of the total effort before 2007 and increased thereafter to about 35% in 2015. Meanwhile, a similar increasing trend was observed in the number of vessels fishing in the northern area, suggesting that more vessels are likely to move their fishing ground from the traditional southern area to the northern area.

3.2. Abundance (CPUE) distribution

CPUE can be used as an indicator of abundance assuming that fishing behavior such as targeting does not change over space or time. Oshima *et al.* (2012) showed that Japanese longline PBF CPUE in Area 1 increased drastically in 2003, whereas CPUE in Area 2 displayed a decreasing trend from 1994 to 2011. The increase in CPUE in Area 1, which is a spawning ground and is dominated by large fish, is supported by the large cohort entering and moving through the fishery as shown in Maunder *et al.* (2014). The decline in CPUE in area 2 may be partly due to this cohort moving out of the area, but may also be due to a change in targeting (see below).

Standardized CPUE for the Taiwanese PBF fishery indicated that the CPUE declined since the beginning of the series in 2001 for both fishing areas. Later, for the southern area, the series reached the lowest level in 2012–13 and started to slightly recovered; while, for the northern area, after reached the lowest level in 2012, the series showed a strong increasing trend to a high level in 2014 and 2015 (Chang and Liu 2016). The trend in the northern area might not be representative of abundance because the effect of spatial change due to a substantial increase of fishing effort in a small area that might have not been properly addressed in the analysis.

3.3. Targeting

Targeting can be explored by looking at the spatial distribution of effort with respect to the typical areas of high CPUE for the different species. For example, within Area 2 the effort of the Japanese longline in the higher PBF CPUE areas decreased after 2005 (Oshima *et al.* 2012). Oshima *et al.* (2012) showed that the proportion of sets made in the high PBF CPUE locations declined from 64% in 2005 to 22% in 2011. In contrast, the effort in the highest YFT CPUE locations increased in 1999 and further increased after 2005. Additionally, the number of sets in the lowest YFT CPUE locations decreased after 2001. The mean number of hooks per basket (HPB), an indicator of targeting, with less HPB used to target YFT, decreased since 2002 in the highest YFT CPUE locations. Despite YFT CPUE being higher in Area 1, YFT CPUE generally did not change in this area until 2009, although it spiked in 2002. Oshima *et al.* (2012) suggested based on this information that PBF targeting moved to area 1 while effort in area 2 switched to targeting YFT. Targeting of YFT in area 2 included both reducing HPB and changing the location of fishing. They also argued that the PBF CPUE decline after 2005 might be exaggerated by the shift of target species.

For Taiwanese CPUE, the target effect was not investigated due to lack of supporting information. The CPUE standardization was performed on registered PBF targeting vessels that are supposed to target on PBF during the fishing season. Vessels that caught less than 5 fish in the year were excluded from the analysis. Therefore, Chang and Liu (2016) considered that the target effect may not have a significant impact on their standardized CPUE. However, this may need further testing using data since 2010 that

have higher logbook coverage.

3.4. Correcting for targeting

Oshima *et al.* (2012) attempted to account for the change in targeting by conducting CPUE standardization with new area definitions, two HPB categories, and including YFT and ALB catch rates as explanatory variables. Unfortunately, they concluded that their new analysis did not completely correct for the decrease of CPUE in recent years and that their procedure of CPUE standardization for PBF could not remove the effect of changes in the fishery. Other authors (*e.g.*, Thorson *et al.* In press) have accounted for targeting by including multispecies data, and increasing (or decreasing) the statistical leverage of observations that have catches associated with the target species (or associated with targeting a different species).

3.5. Size composition

Maunder *et al.* (2014) showed how the size composition of the Japanese and Taiwanese longline catch changed over time as they followed a few very strong cohorts as they grew in size. They suggested that these strong cohorts moved spatially through the Japanese longline fishery and into the Taiwanese longline fishery in the south. They also suggested that the Japanese longline fishery followed these cohorts spatially and therefore the strong cohorts had an influence on the spatial distribution of the fishery. Hence, any use of the composition data used in the assessment for representing the component of the computation represented by the CPUE based index of relative abundance should account for the spatial structure of the population and the fishing effort.

Unfortunately, the Japanese longline length composition data, which is collected by port sampling, is not linked to a location. Attempts to link the length composition data to the logbooks, which contain location information, by date and port of landing are problematic. Training vessels may be an alternative source of length composition data, but their catches of Pacific bluefin tuna may be too small to be useful. Development of more creative methods might be needed to allow for modelling of the size composition of the data when creating the index of abundance.

Length composition data with location of capture is available for the Taiwanese longline fishery starting in 2010.

Other tuna stocks (*e.g.* albacore tuna in the north Pacific Ocean) show spatial differences by gender. Therefore, obtaining gender information in addition to length composition data should be considered.

4. SUGGESTED PBF MODELLING

Given that the longline CPUE is the main source of information about adult relative abundance in the PBF stock assessment, further analysis of the data to remove the effects of targeting and changes in spatial distribution is required to ensure that it provides a reliable index of relative abundance. Here we make several suggestions on possible approaches to improve the analysis of the longline data.

First, the changes in the spatial distribution should be accounted for using a spatial-temporal model as outlined above. Given the sharp changes in CPUE latitudinally compared to the changes longitudinally, the parameters of the correlation function should be estimated separately for latitude and longitude (*e.g.*, termed geometric anisotropy). The PBF fishing season is short from April to June so it may not be important to include season in the model. However, Oshima *et al.* (2012) did find that periods of 10 days intervals were significant in their standardization model. We suggest that initial models should not include season, but the residuals of the model fit investigated to see if some finer temporal scale is required. This implies that the time step of the model is a year and this will greatly reduce the computational demands of the model. A zero-inflated model will likely be needed due to the rarity of PBF in the catch, depending

on the choice of data aggregation. For example, Oshima *et al.* (2012) used a delta-lognormal model following Ichinokawa and Takeuchi (2012). Sea surface temperature should be considered as a predictor of fish density because most species are highly influenced by temperature. Other covariates should also be considered. HPB should be used as the primary measure for targeting of PBF (other than location) as a catchability component of the model. However, adding HPB in the model may require different spatial variance and correlation parameters for each HPB category and it is not clear if this would be related to density (by depth) or catchability. Therefore, initial analyses should be based on sets with HPBs that are considered to target PBF and not use HPB as an explanatory variable. For example, Oshima *et al.* (2012) used ≥ 16 HPB to define deep sets that presumably target PBF. Oshima *et al.* (2012) also included YFT and ALB catch rates in their analysis to account for targeting. However, this can bias the analysis and we recommend initially trying to deal with targeting using spatial and HPB factors. Other methods to determining targeting should be investigated. One by one degree stratification of the data should be an adequate resolution for the analysis and will limit the computational demands, but further evaluation of the spatial resolution may be needed after initial analyses.

Due to the spatial movement of strong cohorts through the fishery as they age and grow, and the apparent targeting of these strong cohorts, inclusion of length in the analysis is important. All possible methods should be evaluated to determine the location of the length frequency samples. The composition data is a sub-sample of the total catch and therefore separate likelihoods should be used for the two components so that the sample size of the composition data is taken into consideration. For example, this would mean that stations with low sample size for composition data, but with a large amount of effort, will have little weight on the size component of the predicted distribution, but substantial weight on the predicted overall CPUE. This also implies that the overall CPUE component of the likelihood function should be weighted by fishing effort (*e.g.* number of sets or hooks). The choice is a tradeoff among computational demands, more precise modelling of the system, factors used in the standardization, and desired resolution of the estimated catch size composition. The issue is partially controlled by the definition of a data point in the analysis. The least computationally demanding approach is to define a data point as the data aggregated to the level of the factors included in the model. For example, in a simple spatial-temporal model, the data point could be the catch aggregated by year, 1x1 degree square, and 1 cm length interval, and the variance component of the likelihood function appropriately specified. The data would likely have to be disaggregated if other factors were included in the model. For example, if vessel was considered as a factor in the analysis, then the data would also have to be disaggregated by vessel. At the other extreme, each longline set could be considered as a data point. To simplify the model, a single likelihood could be used to fit the catch by strata (time and 1x1) and length bin and each strata-length bin given the same weight (variance). Alternatively, the catch and composition data could be aggregated separately by strata 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. The catch-at-size data for use in the stock assessment model should be calculated simultaneously as described above.

The inclusion of length composition data into the analysis is complicated by the issues involved with assigning a location to the data. Creative methods may need to be developed to model the length component of the spatial model. For example, if the length composition data can be assigned to a trip, then the observed composition data can be fit to the predicted length composition data for the whole trip based on the locations of all the sets in the trip and the total catch for each set. Inclusion of gender in the model should also be considered if data is available.

Consideration should be given to analyzing the Japanese and Taiwanese data simultaneously in the same

model, thereby allowing for an “integrated” index of abundance to be developed. The concept behind this recommendation is that in areas where the fleets overlap, the length specific density is the same for both fleets, but the catch composition may differ between the two fleets in a given area due to differences in the contact selectivity (*e.g.* the gear characteristics such as depth of fishing). The Taiwanese data may increase the spatial coverage of the data set, particularly to areas where the largest fish are caught.

Inclusion of length extends the standard spatial temporal model from 3 dimensions to 4 dimensions greatly increasing the computational demands of the analysis. Kai et al (2017b) used 13 size-classes in their analysis, but aggregated output to represent only three categories (juveniles, immature, and subadults and adults for shortfin mako). Stock assessment models use composition data to mainly provide information on selectivity and recruitment and therefore need a much finer resolution. The current PBF stock assessment uses 2 to 6 cm width length composition bins and the CPUE analysis should be based on the same resolution. Therefore, additional research is needed to investigate efficient ways to implement the model to ensure that the desired resolution is practical.

5. OTHER METHODS

The spatio-temporal modelling approach we described is based on using Gaussian random fields and follows the work of Lewy and Kristensen (2009), Kristensen *et al.* (2014), Nielsen *et al.* (2014), Thorson *et al.* (2015), and Kai *et al.* (2017a,b). 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. 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), but also physical barriers such as peninsulas. GAMs allow for other covariates in the model, which could include simultaneously modelling of length, in addition to the spatio-temporal effects. And, 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, something that can be done in the context of GAMs by treating the random effects as penalized fixed effects (Wood 2006; Augustin *et al.* 2013).

Future research could also explore, using the panoply of machine-learning regression techniques, to model the spatio-temporal variation (in the following, we refer to tree-based methods for regression, *e.g.* CART, random forest (RF), but note that recursive neural networks have shown great progress in time-series learning environments). Tree-based methods (*e.g.*, CART, random forests; Breiman *et al.* 1984; Breiman 2001) can be used to explore interactions of space (using multiple redundant measures, *e.g.*, latitude, longitude, depth, distance from port, to explore which combination has greatest explanatory power), 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). The space and time scales of the spatio-temporal structure captured by these types of tree-based algorithms are not explicitly controlled, except in a limited manner by way of the spatial and temporal resolution of the predictors. However, 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 forests when the underlying process behind the data are unknown.

We see at least three major problems with an RF-based approach for standardizing compositional data,

which may need to be addressed, all of which stem from the fact that RF was developed as a black-box method for prediction, not as a method for fitting parametric models:

1. It is difficult to characterize RF uncertainty for predictions at a point, or aggregating across space (strata, etc.). RF (regression) typically evaluates uncertainty in terms of mean squared prediction error, between observations and predictions on out-of-bag data or predictions on a test data set. As such, the typical RF measure of uncertainty does not easily lend itself to construction of covariance matrices. Currently, the uncertainty in abundance indices and composition data are frequently used to represent observation error in stock assessments. Similarly, the joint index and composition estimates from spatio-temporal models would ideally use the estimated covariance matrix to define the likelihood in the stock assessment model.
2. RF may overfit with respect to a desired parametric model. Specifically, there is no explicit control over the interaction structure that will result for each tree in the forest. This can be advantageous if the interaction structure is unknown (i.e., the underlying model is unknown), but disadvantageous if a specific parametric model is preferred. For example, spatio-temporal models can be specified to include a design matrix for gear in GMRF models such that the ratio of expected catch rates for gear 1 and 2 is identical across space and covariates. This precludes any interaction between a gear variable and other covariates, and sacrifices predictive performance in favour of interpretability. By contrast, including gear in RF will most likely result in high-level interactions, which can render interpretation of the resulting catchability-ratios (and understanding what is attributed to catchability vs. density differences) extremely difficult.
3. Stock assessment modelers (e.g., using Stock Synthesis) typically treat indices in an assessment model as if each index year is statistically independent (i.e., assuming that covariance among years is zero). If a covariance matrix could be estimated for a time series of RF predictions, we hypothesize that this covariance structure would have non-zero covariance between years (due to shared dependence upon covariates, and trees where a year-factor is not included in the predictors). Spatio-temporal models by contrast can treat year as a fixed effect and therefore be specified to achieve a near-zero covariance for index estimates among years. If an assessment model is modified to use a multivariate distribution to fit indices (so that model assumptions aren't violated when using an index generated by RF), it would still be a disadvantage to have non-zero covariance because this covariance will decrease "effective sample size" for the index (which will offset the potential gain in precision for each year in isolation when using RF or other methods that achieve increased precision via temporal shrinkage). Most importantly, we hypothesize that RF will most often shrink years towards the mean over blocks of years if year is included as a factor (or covariate). This shrinkage will result in bias in assessment estimates of status in year $T+1$ towards estimates in year T , which would in turn cause assessments to underestimate changes in biomass. Underestimating changes in biomass will generally result in increased probability of a stock being overfished. By contrast, spatio-temporal models can control the inclusion/exclusion of shrinkage among years via the temporal structure on annual intercepts -- index-based methods for setting overfishing limits will often benefit from including temporal shrinkage, but expanded estimates of composition to be used in a subsequent stock assessment presumably will not.

Despite these concerns, we recommend future research comparing the relative performance of RF vs. spatio-temporal models for compositional expansion for use in stock assessment models.

6. OTHER APPLICATIONS

We discuss the use of spatio-temporal models in respect to Bluefin tuna and modelling the length composition of the stock and how it may differ from the catch. However, there are numerous other

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.* (In press) explored the relative explanatory power of local and regional temperature, size-structure, and otherwise unexplained processes in explaining distribution shift for Alaska pollock in the Bering Sea.

Purse seine data has been used to develop indices of relative abundance for silky sharks 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 the area fished by purse seiners, possibly due to changing environmental conditions. 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.

The Japanese longline fleet targeting tuna in the EPO has shown changes in spatial distribution and gear characteristics over time. The catch and effort data from this fleet is used as the main indices of relative abundance for the stock assessments of yellowfin and bigeye tuna. Therefore, spatio-temporal models that consider gear factors should be used to develop the indices of abundance. For example, the spatio-temporal component of the model might be dependent on the hooks between floats, which influences the depth that the long lines fish.

Albacore tuna shows differences in distribution by size and gender. Therefore, the analysis of longline catch and effort data should be conducted using spatio-temporal models in which the spatio-temporal component is a function of both length and gender.

Assessments for EPO 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 try to combine information collected by onboard fisheries observers aboard the commercial tuna vessels, which use dolphins to locate and catch yellowfin tuna, to obtain estimates of abundance. Although it is unclear how the biases that have been identified in the fisheries observer data (Lennert-Cody *et al.* 2001; 2016) would be addressed, spatio-temporal models would be required in order 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.

REFERENCES

- Augustin, N.H., Trenkel, V.M., Wood, S.N. and 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. *Machine Learning Journal* 45: 5-32.
- Carruthers, T.R., Ahrens, R.N., McAllister, M.K. and Walters, C.J. (2011) Integrating imputation and standardization of catch rate data in the calculation of relative abundance indices. *Fisheries Research* 109, 157–167.
- Francis, R. I. C. C. (*in press*). Revisiting data weighting in fisheries stock assessment models. *Fisheries Research*.
- 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.
- Harley, S. J., Myers, R. A., and Dunn, A. 2001. A meta-analysis of the relationship between catch-per-

- unit-effort and abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 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., Maunder, M.N. 2004. A Bayesian integrated population dynamics model to analyze data for protected species. *Animal Biodiversity and Conservation* 27: 247-266.
- Ichinokawa, M. and Takeuchi, Y. 2012. Standardized CPUE of North Pacific Bluefin tuna caught by Japanese coastal longliners: updates until 2011. *ISC/12-1/PBFWG/8*. 1-15.
- Kai, M., Thorson, J.T., Piner, K., and Maunder, M.N. 2017a. Predicting the spatio-temporal distributions of pelagic sharks in the western and central North Pacific. *Fisheries Oceanography*.
- Kai, M., Thorson, J. T., Piner, K. R. and Maunder, M. N. 2017. Spatio-temporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Can. J. Fish. Aquat. Sci.* doi:10.1139/cjfas-2016-0327
- 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. doi:10.1139/cjfas-2013-0151
- Kristensen, K., Nielsen, A., Berg, C., Skaug, H., & Bell, B. (2016). TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70(5), 1-21. doi:http://dx.doi.org/10.18637/jss.v070.i05
- Lennert-Cody, C.E., Buckland, S.T., Marques, F.C. 2001. Trends in dolphin abundance estimated from fisheries data: a cautionary note. *Journal of Cetacean Research and Management* 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. *Fisheries Research* 102: 323-326.
- 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. *Fisheries Research* 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. *Fisheries Research* 178: 101-113.
- Lewy, P., Kristensen, K., 2009. Modelling the distribution of fish accounting for spatial correlation and overdispersion. *Can. J. Fish. Aquat. Sci.* 66, 1809–1820. doi:10.1139/F09-114
- Maunder, M.N. and Punt, A.E. (2004) Standardizing Catch and Effort Data: A Review of Recent Approaches. *Fisheries Research* 70(2-3): 141-159.
- Maunder, M.N., Sibert, J.R. Fonteneau, A., Hampton, J., Kleiber, P., and Harley, S. (2006a). Interpreting catch-per-unit-of-effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science*, 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. *Bulletin of Marine Science*, 79(3): 545–559.
- Maunder, M.N., Piner, K.R., and Aires-da-Silva, A. 2014. Stock status of Pacific Bluefin tuna and urgent need for management action. *IATTC Stock Assessment Report* 15: 47-73.
- 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. *Fisheries Research*.
- 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. doi:10.1371/journal.pone.0099151
- Sampson, D. B. 2014. Fishery selection and its relevance to stock assessment and fishery management. *Fisheries Research*, 158: 5-14.
- Stewart, I.J., and S.J.D., Martell (2014) A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment *Fisheries Research* 158:40-49.
- Stewart, I.J., and Monnahan, C.C. (2017). Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. *Fisheries Research*.
- Thorson, J.T. (2014) Standardizing compositional data for stock assessment. *ICES Journal of Marine Science: Journal du Conseil* 71, 1117–1128.
- Thorson, J.T., Ono, K. and 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. and Skaug, H. (2015b) 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. and Benante, J. (2015c) 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. and Zipkin, E.F. (2016) Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography* 25, 1144–1158.
- Thorson, J.T., Johnson, K.F., Methot, R.D. and Taylor, I.G. (2017) Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. *Fisheries Research*.
- Thorson, J.T., Fonner, R., Haltuch, M., Ono, K. and Winker, H. (In press) Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. *Canadian Journal of Fisheries and Aquatic Sciences*.
- Thorson, J.T., Ianelli, J.N., Kotwicki, S., In press. The relative influence of temperature and size structure on fish distribution shifts: a case study on walleye pollock in the Bering Sea. *Fish Fish*.
- 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. *Fisheries Research*, 158: 15-25.
- Walters, C. (2003) Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences* 60, 1433–1436.
- 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. *Journal of the Royal Statistical Society, Series B* 70: 931-955.