Methods for spatial survey modelling - program of work for 2018

PREPARED BY: IPHC SECRETARIAT (R. WEBSTER; 20 MAY 2018)

PURPOSE

To present results on spatio-temporal survey modelling undertaken to date in 2018, and describe plans for the remainder of the year.

BACKGROUND/INTRODUCTION

In 2016, IPHC Secretariat staff began using a space-time modelling approach to estimate indices of density and biomass for use in stock assessment modelling and estimation of stock distribution. Survey station weight and number per unit effort (WPUE and NPUE) indices are used as input data, following a standardisation for competition for baits among Pacific halibut and other species. Prior to the introduction of the space-time modelling, this standardisation was calculated from data aggregated across all IPHC setline survey stations within each Regulatory Area, rather than at the level of survey station. The adjustments are calculated from the proportion of baits returned on each setline survey station, and the proportions themselves are calculated from data obtained from the first 20 hooks on each skate rather than the full 100 hooks otherwise used for data on Pacific halibut. An exception is Regulatory Area 2B, where returned baits have been counted on 100% of hooks on the setline survey since 2003 (with the exception of 2013). An analysis of the Regulatory Area 2B data (Webster and Leaman 2014) showed that, with data aggregated across the Regulatory Area, 20-hook-count estimates of the proportion of baits returned were unbiased and precise relative to the 100% hook counts. In this report, we revisit the Regulatory Area 2B bait return data to assess whether the use of 20-hook counts in other Regulatory Areas is likely to affect the density index estimates from the space-time model.

The use of environmental covariates to improve the space-time model has been discussed at previous meetings of the Scientific Review Board. The IPHC uses water column profilers on each setline survey station to record several environmental variables, including dissolved oxygen, temperature and salinity. Due to a hypoxic event off the Washington coast, the effect of dissolved oxygen on Pacific halibut density in Regulatory Area 2A in 2017 has been subject to much discussion, and this variable is therefore of particular interest to us when examining how environmental covariates can be used to improve the density index estimates from the space-time model. We summarize results of exploratory modelling of relationships between O32 WPUE and bottom dissolved oxygen and temperature for Regulatory Area 2A, and describe a proposed framework for including these variables in models for the full 1993-2017 data. For the latter, we note that wide-scale use of the water column profilers began in 2009, and that since 2009, profiler data are missing at many fished setline survey stations and unavailable when expansion stations are not fished. The method used should avoid omitting Pacific halibut data even in the absence of corresponding environmental data.

EFFECT OF USING 20-HOOK COUNTS ON WPUE

Data from Regulatory Area 2B were used to compare the output from space-time models using hook competition standardisation based on data from 20 hooks/skate, and from the output using adjustments based on 100% of hooks. Figure 1 compares the estimated WPUE time series from the space-time model using 20 hooks versus 100%. The two time series are very close, with differences well within 95% credible intervals.

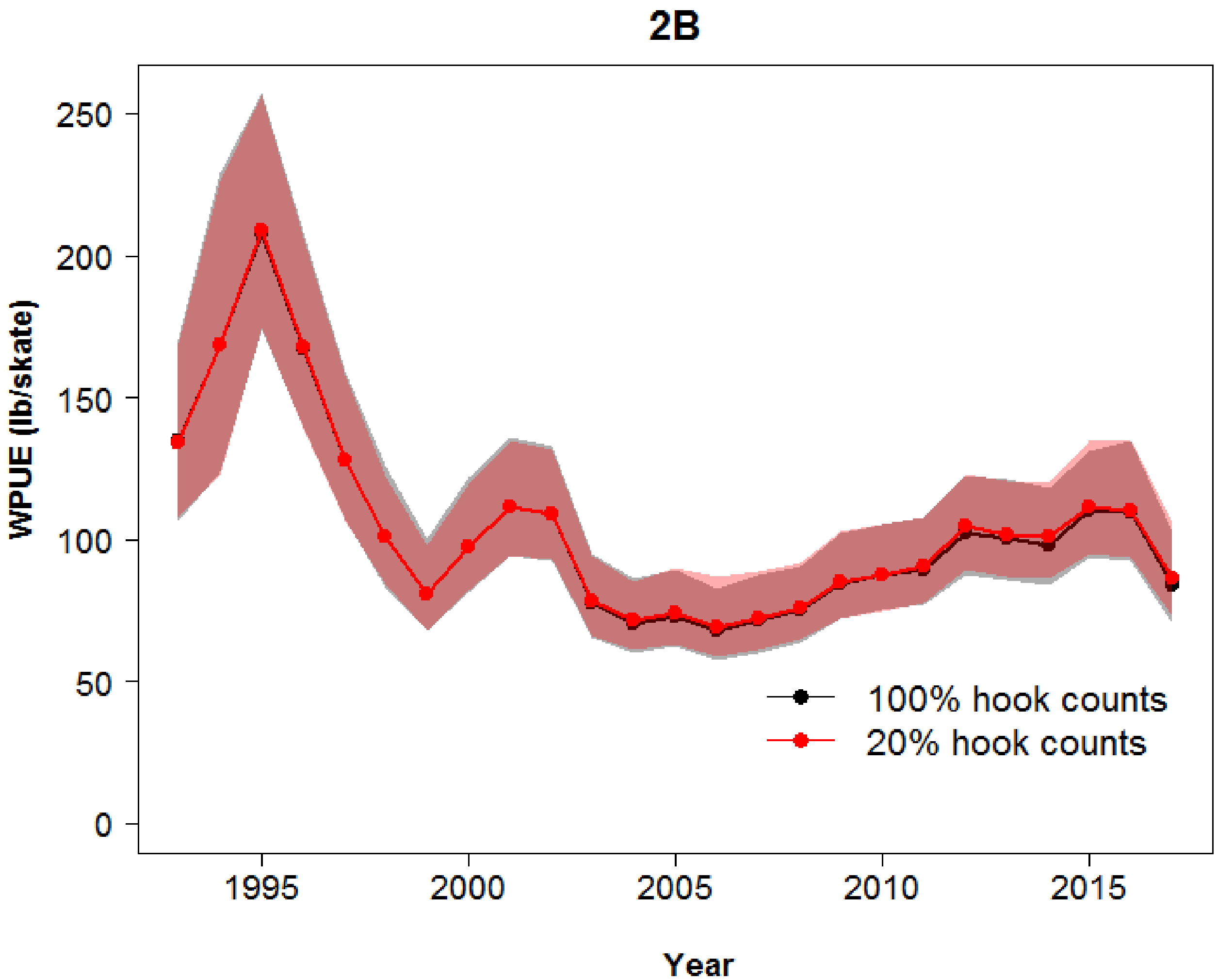

Figure 1. Space-time model posterior means and 95% credible intervals of O32 WPUE in Regulatory Area 2B with hook competition standardisations based on 100% of hooks and 20 hooks/skate.

From 2003, when 100% hook sampling was implemented, the 20-hook-count series is consistently slightly above the 100% series (Figure 1), which at face value may imply a very small positive bias in using data from just 20 hooks/skate. However, some slight difference is almost guaranteed because of the way the hook competition standardisation is calculated to accommodate the case of zero baits returned on a station. Recall that in computing the standardisation at each station, the quantity *Z* is required:

$$Z = \log\left(\frac{h}{b}\right)$$

where *h* is the number of hooks on a set and *b* is the number of baits returned. To avoid division by zero when no baits are returned, we add a small quantity δ to both *h* and *b*:

$$Z = \log\left(\frac{h+\delta}{b+\delta}\right)$$

where $\delta = h/100$. This choice means δ is proportional to the number of hooks set, ensuring that if no baits are returned, *Z* will be the same for sets of different lengths (e.g., 5 skates vs 6 skates). The adjustment factor for the standardisation is given by

$$f_H = \frac{Z}{1-e^{-Z}}$$

When 20 hooks are used instead of 100%, there is a far greater chance that any given station has zero or close to zero baits returned, likely leading to a larger adjustment factor for that station than if 100% sampling had been used. This greater chance of a larger adjustment factor is the cause of the very slightly higher WPUE values for 20 hooks in Figure 1. No value of δ is going to lead to 20 hooks and 100% giving the same adjustments on average all the time, and our experiments with alternative values (e.g., $\delta = 1$ or something intermediate to this and the current choice) led to time series that differed by a greater amount.

Our conclusion is that the very small difference between the series based on data from 20-hook and 100% hook sampling is not meaningful for scientific and management purposes, and is not sufficient to justify the additional expense of sampling 100% of hooks in all Regulatory Areas.

ENVIRONMENTAL COVARIATES IN SPACE-TIME MODELS

In this section we outline a proposed approach to including environmental data from the setline survey's water column profilers as covariates in space-time models. The focus here is on Regulatory Area 2A, due in part to the apparent large effect of dissolved oxygen on this area's WPUE motivating the desire to explore such models, and that for Regulatory 2A, space-time models can be run relatively quickly, allowing us to fit and compare several models more easily.

We began by fitting exploratory models using observed data only. That is, if a setline survey station has missing covariate information, it was excluded from the modelling. Thus, we are only using setline survey WPUE from 2009 onwards, and exclude a number of stations for which technical or data quality problems meant no reliable covariate data were available. The advantage of starting with the observed data is that models can be fit very quickly to this smaller data set, while the results can demonstrate if it is worthwhile trying to fit models with all the data from 1993-2017. Exploratory models included covariates for bottom temperature, which previous models fitted to Bering Sea data showed was an important predictor in that region, and dissolved oxygen. All models also included the depth and latitude covariates used in the modelling of the full 1993-2017 data, and considered base-model covariates for Regulatory Area 2A. Model fit was compared using the Deviance Information Criterion (*DIC*), with lower values indicating better fitting models.

Table 1 presents the *DIC* for the models exploratory models we fitted to the Regulatory Area 2A data. The space-time models include two components, one for modelling the probability of zero WPUE, and a gamma model for non-zero WPUE observations. Each component can include covariates, and unless indicated by “z” or “nz”, covariates were included in both.

Table 1 Deviance Information Criterion values for models fitted to observed setline survey data in Regulatory Area 2A (2009-2017).

Model	***DIC***	**Notes**
Base model (depth + latitude)	8605.9	
Base + temp	8603.7	
Base + $temp^2$	(Did not converge)	Correlation with depth?
Base + $temp^2$(z) + temp(nz)	8605.5	
Base + $temp^2$(nz)	8605.7	
Base + O_2	8593.7	
Base + $I(O_2 < 0.9)$(z) + O_2(nz)	8590.2	

Note that I( ) is the indicator function, taking values of 1 is the statement is true, and zero otherwise. In our model, this was used to create a binary variable with value 1 if dissolved oxygen was less than 0.9 (the level previous work and last year’s data showed led to zero or almost zero halibut) and value zero otherwise. In that model, non-zero WPUE was modelled as linearly dependent (on the log scale) on dissolved oxygen.

The results in Table 1 show little evidence for the importance of bottom temperature as a covariate in WPUE, although we note the linear temperature model had a slightly lower *DIC* than the baseline. Inclusion of dissolved oxygen led to a much larger decrease in *DIC*, with the model with the binary variable for the probability of zero halibut having the lowest *DIC* of those we considered. In summary, dissolved oxygen is a promising variable for inclusion in models which use the full 1993-2017 WPUE data set.

Next we must find an approach which allows for inclusion of environmental covariates without excluding any of the Pacific halibut catch data. The first step is to impute values for stations with missing data from 2009-2017. To do this efficiently, we chose to fit simple exponential spatial models to the dissolved oxygen data for each year separately, and then predicted dissolved oxygen at stations with missing values (including stations unsurveyed in a given year). Dissolved oxygen has a relatively smooth spatial distribution, and predictions within the spatial range of observed data are likely quite good. For consistency, we also replaced observed values with predicted values, although these were almost always very similar.

The next step is to construct a model for the relationship between WPUE (both zero and non-zero components) and dissolved oxygen that does not require having environmental data prior to 2009. For this, we need to define the dummy variable D as $D = 0$ if year < 2009 and $D = 1$ if year ≥ 2009. Then the covariate model including dissolved oxygen is:

$$\text{(intercept + other covariates)} + \beta_0 D_i + \beta_1 X_i D_i$$

where β_0 is the intercept difference for stations with dissolved oxygen values, β_1 is the slope of the linear relationship, and X_i is the value of dissolved oxygen for the ith station in a given year. Thus in the absence of dissolved oxygen values, the model defaults to the base model, which (as above) includes depth and latitude as covariates. Non-linear relationships, and other more general models can be also be constructed using the same dummy variable.

Such covariate models are currently being fitted to the full 1993-2017 data, and it is expected that some results will be available for presentation at SRB012. Based on the final results and SRB input, a decision will be made as to whether to include dissolved oxygen as a covariate in space-time models for Regulatory Area 2A in 2018.

One point of discussion that we would appreciate SRB input on is how to account for variables such as dissolved oxygen in estimating the WPUE time series from the space-time model. Two approaches have been discussed among staff and stakeholders. The first is that we simply estimate WPUE at the values of dissolved oxygen used in the original modelling, so that the purpose of adding this variable is to add information that leads to improved estimation of WPUE. The second option is to predict WPUE at a fixed value of dissolved oxygen for all years and stations, leading to a WPUE index that has been adjusted for variation in dissolved oxygen. The argument for the latter approach is that years of very low dissolved oxygen are anomalous and that by “adjusting away” the effect of this variable, the resulting WPUE series will provide a better index when estimating stock distribution to inform management decisions for the following the year. A counter-argument is that the WPUE index should reflect the underlying Pacific halibut density in each year, and not some hypothetical density under environmental conditions which were not in fact observed.

SPACE-TIME MODELLING CHANGES AND PLANS FOR 2018

The following is a list of other potential changes to the space-time modelling in 2018, and other work related to these models:

- Noting that three Pacific halibut were found near the outside of Kotzebue Sound in a 1998 trawl survey, we will examine adding geographic area in the southern Chukchi Sea to Area 4CDE space-time models. We note that no Pacific halibut were found during the 2012 National Marine Fisheries Survey of the Chukchi Sea.
- Summarise measures of geographic area currently used by other fisheries organisations when estimating stock distribution among management areas. The IPHC

uses the geographic area of the ocean surface in each Regulatory Area to create coastwide indices of density and estimate stock distribution, but Commissioners have directed staff to consider alternative metrics that account for the slope or rugosity of the ocean bottom. Examining current practice elsewhere is the first step.

- Use a version of the space-time models to study the spatial distribution of Pacific halibut larvae, with particular focus on differences between warm and cold periods, and between the Bering Sea and the Gulf of Alaska. This work is in progress at present in collaboration with Lauri Sadorus (IPHC staff) and scientists from NMFS.
- In 2018, the IPHC revised the criteria for effectiveness of setline survey sets given evidence of whale depredation. Once databases have been updated to reflect this change, we intend to examine its effect on space-time model estimates of density indices.
- Report on the results of the setline survey expansions in Regulatory Area 2B and 2C in 2018, and incorporate all new data into the 2018 space-time modelling for these areas.

Reference

Webster, R. A. and Leaman, B. M. 2014. Setline survey hook counts: are the first 20 hooks sufficient? Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2013: 421-432.

Recommendation/s

The IPHC secretariat requests that the SRB:

NOTE this document summarising spatial survey modelling updates for 2018.

NOTE any discussion occurring during SRB012, and **RECOMMEND** any conceptual or technical improvements for survey modelling.