SCIENTIFIC COMMITTEE
SIXTEENTH REGULAR SESSION

ELECTRONIC MEETING
11-20 August 2020

ANALYSIS OF TAGGING DATA FOR THE 2020 TROPICAL TUNA ASSESSMENTS: TAGGER AND CONDITION EFFECTS

WCPFC-SC16-2020/SA-IP-05

J. Scutt Phillips[1], T. Peatman[2], M. Vincent[1], S. Nicol[1]

[1] Pacific Community (SPC), Ocean Fisheries Programme (OFP), Noumea, New Caledonia
[2] Independent Fisheries Consultant, Wiltshire, United Kingdom

Executive summary

This information Paper describes the pre-processing of tagging data prior to their integration into the 2020 stock assessments for yellowfin and bigeye tuna.

The SC is advised to note:

The adjustment to the process for selecting data and variables for use in the statistical models to estimate the effects of tagger experience, imprecise tag placement, and fish injury prior to tagging in comparison to previous applications.

The pooling of all tagging data for estimation of models, including those from the SSAP, RTTP and PTTP.

A term for individual tagging events/schools (*tag_sch_id*) has been included previously to account for the spatiotemporal distribution of the fishing effort around the location of tagging, the behaviour of the tagged fish in relation to this effort, and the effect of the gear used to catch fish prior to tagging. In 2020 this term was replaced by "*cruise leg*" due to poor representation of most other variables when "*tag_sch_id*" was included.

Similarly, a tagger ID term has previously been used to capture variability due to the taggers experience and skill. In 2020 this was replaced by a tagger experience term, assuming that the skill of individual taggers was captured purely by the numbers of tags they had deployed in any species of tuna previous to each release. This approach further allowed better representation of variables across four tagging experience groups of beginner (less than 500 tags), intermediate (500-2500 tags), competent (2500-10,000), and advanced (with more than 10,000 tags released).

A suite of models were estimated using combinations of covariates at the tag release level, before a model selection process based on information-criteria was carried out to select the most parsimonious model, given the data. The chosen model for yellowfin tuna included *cruise leg, length, fish condition, tagging quality, school association behaviour* and *tagger experience group* as terms. Correction factors were estimated, controlling for optimum *fish condition, tagging quality* and *tagger experience group*. The median correction factor for yellowfin tuna assessment model tag release groups was 0.90.

The chosen model for bigeye tuna included only *cruise leg, length,* and *fish condition* as terms, and correction factors were calculated correcting only for *fish condition*. The median correction for bigeye tuna assessment model tag release groups was 0.99.

SC is invited to note the following recommendations:

- **Continue to apply this objective model selection framework to future tagging effects analyses prior to calculation of correction factors for future assessments**
- **Re-examine combining tags for skipjack, yellowfin and bigeye tuna in modelled datasets, with the inclusion of interaction terms between species and covariates that are likely to vary for each of the three tunas**
- **Explore potential terms with mechanisms that capture the spatiotemporal fishing effort at time of release, replacing the need for variance in recapture probability to be replicated by abstract tag event or cruise leg covariates**

1 Introduction

Conventional mark-recapture tagging data are used in stock assessments of tropical tunas using MULTIFAN-CL, and provide important information that inform estimated fishing mortality and movement parameters (Hoyle et al. 2013, Berger et al. 2014, Vincent et al. 2019). A number of processes are undertaken prior to their inclusion into a stock assessment, including the 'correction' of biases in the apparent observed numbers of tag releases due to tagging-related mortality. The effect of covariates on the probability of recapture have been shown to include differing tagger experience, imprecise tag placement, and fish injured prior to tagging (Hoyle et al. 2015). The correction factors used to adjust the tag release for these mortalities are derived from statistical models to estimate these effects. Failing to do this may lead to estimated fishing mortality being biased low if left uncorrected (Vincent et al. 2019). This Information Paper describes the estimation of these models for the 2020 stock assessments of yellowfin and bigeye tuna.

2 Methods

The procedure for creating correction factors involves the following steps:

1. Extraction and filtering of tag release and recapture data
2. Examination of spatial and temporal representation and correlation across factors in the data
3. Estimation of, and selection from, a suite of potential statistical models
4. Final estimation of chosen model and generation of correction factors

2.1 Extraction, filtering and preparation of tagging data

Data extractions were made from the Skipjack Survey and Assessment Programme (SSAP, 1978-1982), Regional Tuna Tagging Programme (RTTP, 1990-1996) and Pacific Tuna Tagging Programme (PTTP, 2006-present) databases to extract individual tag release information, including whether a tag was recovered. A number of filtering steps were undertaken to remove tags that were considered inappropriate and likely to include significantly different mechanisms driving their recapture probability. Tags removed were:

- Tags released after 2017 (to allow adequate time for recoveries to have been reported, and for recovery data to have been processed)
- Tags from tag seeding experiments used to estimate reporting rates (Peatman et al. 2019)
- Tags released as part of double tagging experiments to estimate tag-shedding (Hampton 1997)
- Any tags associated with an electronic tagging experiment
- Tags released under the following PTTP projects: Japanese Tagging Trial Project, Purse-seine trial tagging, Longline tagging projects, albacore tagging projects
- Tags released as part of IATTC tagging cruises during the RTTP (*proj_id* IA)
- Tags released under the following RTTP and SSAP projects:
    - Tags released east of 150°W (i.e. tags released in the EPO)

Where possible, given slight differences in data recording across the four decades of these programmes, the following data were collected for each individual tag release:

- The name of the cruise leg (*cruise_id*)

- The school/event number (*tag_sch_id*)
- Species (*sp_id*)
- The fish length at release (*len*)
- The recorded tagging quality (*t_qual_id*)
- The recorded fish condition at release (*t_cond_id*)
- The association behaviour of the tagged school (*assoc_id*)
- The tagger performing the tagging operation (*tagger*)
- The station at which the tagging was performed (*cradle*)
- The gear used to catch the fish prior to tagging (*gr_id*)
- Whether the tag was eventually recovered and returned (*recovered*)

From these data, experience was calculated for each individual tagger through time, updated as numbers of tuna (all species) tagged by the end of each individual tagging event. This experience metric was used to create experience categories by binning tagger experience based on evenly dividing up the experience observed across all tag releases, or *a priori* selected categories based on the previous identified effect of experience from Hoyle et al. (2015). Thus, experience data available were:

- The experience of the tagger at time of tagging, in numbers of tagged fish deployed of any tuna species prior to the tagging event (*experience*)
- The experience quantile of the tagger at time of tagging, grouped as quantiles defined by the distribution of raw experience data above as 1) less than 3,000; 2) 3,000-10,000; 3) 10,000-20,000; and 4) more than 20,000 tags deployed (*XP_quant*)
- The experience group of the tagger, this time grouped to represent pre-perceived beginner and differing levels of experience: beginner with less than 500 tags deployed, intermediate with 500-2,500 tags deployed, competent with 2,500-10,000 tags deployed, and advanced with more than 10,000 tags deployed (*XP_Group*)

2.2 Examination of correlation across model terms

In an idealised experiment to model the effect of the above terms on tagging related mortality and recapture probability, tags would be released in an even spread across factors as possible. This would include a mix of gears being used in both central and western Pacific cruises, different taggers systematically switching tagging stations, as well as working a variety of different cruises, and all of the above being changed within the same tagging event.

In reality, these tagging programmes have been chiefly designed to maximise the effective releases of mark-recapture tags for inclusion in stock assessment or other population models, and are subject to many financial and logistic constraints. For these reasons, many of the potential terms that could be used in tagging effects model are problematic because there are insufficient observations across important factors to appropriately estimate statistical models.

A key example of this is the use of individual tagging events/schools in the model (*tag_sch_id*). The most important factor in the probability of a given tag being caught and returned is the spatiotemporal distribution of the fishing effort around the location of tagging, the behaviour of the tagged fish in relation to this effort, and the effect of the gear used to catch fish prior to tagging. As most of these dynamics cannot be fully known, the individual tagging event has previously been used as a catch-all model term that attempts to account for this combined variability, and is typically responsible for explaining the largest proportion of the observed variance in tag returns. However, in most cases there are no more than four individual taggers operating during a single event, each assigned to a single station.

Poor representation and multicollinearity in statistical model predictors can lead to potentially inflated variance of estimated slopes, biased parameter estimates and misidentification of significance (Guisan et al. 2002, Freckleton 2011). The lack of data across levels of categorical model terms was visualised using burning embers plots. For both yellowfin and bigeye tuna, very large gaps in observations exist when using both individual tagging event and individual taggers (figure 1). Similarly, the degree to which tagger, station, and gear are confounded can be seen in burning embers plots of these data (figure 2). To overcome these low numbers of observations, alternative model terms that capture differences in spatiotemporal fishing effort and environmentally-driven behaviour around tagging events were explored.

The assumption that, although consisting of several distinct tagging events, tag releases within a single cruise leg (typically two weeks in length) are more similar to each other than between cruises was applied. In addition, combining releases at the cruise leg level allowed multiple observations of other category levels which were fixed at the tagging event level, including gear and the school association behaviour.

While including an individual tagger term aims to capture the innate skill and experience of each tagger, very few taggers consistently work across events or even cruise legs (figure 1.). Similarly, many taggers may be less effective during the initial stage of their participation in tagging cruises than later on once they have built up more recent experience. Excluding individual tagging events, individual taggers, and tagging station from the analysis and switching to broader experience groups and cruise legs, increased the representation in observed categorical levels across all three programmes (figure 3).

Initial tag event-level exploratory model runs for bigeye, fitted to PTTP data, indicated that models without tagger effects had more support than models including tagger effects, based on comparison of information criteria. This is likely due to the ability of the model to adjust tag-event effects to explain some of the variation that would otherwise be explained by tagger, given the imbalanced nature of the dataset. Additionally, the tagger effects were counter-intuitive, with the most experienced taggers having worse estimated effects than taggers with far less experience. This presents difficulties when selecting a defensible ‘optimum’ tagger for estimation of correction factors.

Given this poor representation and multicollinearity in the data, tagging event, individual tagger, and station were explored as terms in alternative models, but excluded from the final selection process.

2.3 Estimation and selection of statistical models

Multiple statistical models were estimated using the combined tag releases from all three tagging programmes, separated by species to account for potential biological differences in tagging-induced mortality. Several filters were applied to the data before assembling the final aggregated dataset for model estimation. These broadly followed those outlined in Berger et al. (2014), excluding tagging cruise legs with less than 20 (yellowfin) or 15 (bigeye) tag releases, and any other category levels required at least 100 observations in total to be included.

A minimal dataset for modelling was constructed by applying the data filters above for all potential model terms. This allowed information criteria to be appropriately compared between alternative model structures during model selection, as the data on which models were estimated was identical. The Generalized Additive Models (GAMs) R package *mgcv* was used to estimate model parameters, assuming a binomial response for the number of recoveries and non-recoveries for each strata in the aggregated data.

The simplest model was assumed to require at least cruise leg, fish length, fish condition at release, and the recorded quality of the tagging procedure (though this variable only varied for yellowfin tuna releases). Alternative models included combinations of the gear used prior to tagging, the school association behaviour at time of tagging, and a single metric capturing the experience of the tagger (one of either total tags deployed, experience group, or experience quantile as described above).

As the models were built on datasets with relatively large numbers of observations, model selection was based on Bayesian Information Criteria to select the most parsimonious model given the data. Residual diagnostics were also examined using quantile residuals to ensure there was no violation of distributional assumptions.

2.4 Final model estimation and correction factor calculation

Once a final model had been selected, parameters were re-estimated using a full dataset, with data filtering rules only applied for terms selected for inclusion in the model. The estimated model was then used to calculate the ratio between the predicted probability of recapture under the observed conditions and the predicted probability of recapture under 'optimal' conditions controlling for idealised fish condition, tagging quality, and experience of the tagger, if those terms were included in the final, chosen model.

The approach follows that of Berger et al. (2014) and Vincent et al. (2019), controlling for terms included in the final model. Previous correction factors included a correction for an 'optimal' station, assuming a simplified "*bow*" position. However, tagging stations are not necessarily comparable between different tagging platforms, which tend to operate in different areas with different spatiotemporal fishing effort. Additionally, and particularly for pole and line-based cruises, more highly experienced taggers tend to work at the bow stations, with less experienced taggers working elsewhere. This may lead to further over corrections for the large number of stern released fish on central Pacific cruises, particularly for bigeye (figure 2). As station was not considered for inclusion in the models, this control was therefore not used when estimating correction factors.

For the small number of tags that contained observed categorical variable levels that were removed during model estimation process and thus not modelled (i.e. cruise legs filtered due to very few tag releases), either appropriately similar levels or a level estimated to have the median coefficient for that term was chosen.

Finally, the corrections for all observed tags during an assessment release group (currently grouped by quarter and assessment region) were averaged to provide the correction factor for that release group. A suite of correction factors using different tagging effects models, each estimated using alternative combinations of the available variables (described in 2.1 above), were provided for various stock assessment diagnostic scenarios.

3 Results

3.1 Yellowfin tuna

For tags deployed in yellowfin tuna, all explored model structures included cruise leg, fish length at release, fish condition at release, and tagging quality. Alternative models were also estimated with all combinations of school association at release, fishing gear, and one of either absolute tagging experience, tagging experience group, or tagging experience quantile. The dataset contained 57,875 observations in the minimal aggregated dataset for BIC comparison.

The model selection process described above selected for a model containing cruise leg, fish length, fish condition, tagging quality, school association, and experience group of the tagger. The final model was a logistic regression estimated on *k* aggregated groups of tag releases with shared set of covariate values. The response y_k was the number of tags recovered from n_k tag releases, and related via a logit link function to the linear predictors, such that:

$$y_k \sim \text{Binomial}(n_k, p_k)$$

$$\log\left(\frac{p_k}{1-p_k}\right) = \beta_0 + \beta_{cruise\_leg[k]} + \beta_{schass[k]} + \beta_{condition[k]} + \beta_{quality[k]} + \beta_{XP\_group[k]} + f(length[k])$$

All variables were categorical, except the continuous variable *length*, which was modelled using a thin-plate regression spline denoted *f()*.

The effects plots for the terms in the selected yellowfin model are shown in figure 4. For one covariate, fish condition, the level with the most positive effect, *shark bite*, was slightly above that of the agreed reference level, *good*. In this case, the parameter for this level was capped at that of *good* when calculating correction factors.

Probability of recapture was highest for fish released at around 110cm, though was consistently high for fish of this length down to 50cm. Bleeding fish, those dropped on deck or with eye damage had consistently lower probabilities of survival, as did fish badly tagged or tagged too slowly. Beginner taggers (those who had deployed 500 or less tags) were associated with a lower and more variable probability of tag recapture, with little effect between the other experience group levels. The largest effects on probability of capture was the cruise leg on which the tag was deployed. For observed school association behaviours not in the modelled dataset, these were set to *No Info*. Un-modelled cruise legs were set to the cruise leg with the median, estimated model coefficient.

Final correction factors (using an *advanced* experience group, *good* fish condition and *good* tagging quality for control) for all tag releases ranged between 1 and 0.221 (for a 47cm fish caught on a *seamount* and recorded as *dropped on deck* and tagged *too slowly* by a *competent* tagger). Once aggregated to assessment release group, correction factors ranged between 1 and 0.742 for single year, quarter, and assessment region group.

3.2 Bigeye tuna

Similar to models of yellowfin tuna tag releases, all model structures for bigeye tuna tag deployment included cruise leg, fish length at release, and fish condition at release, with alternative models including combinations of school association at release, fishing gear, and one of either absolute tagging experience, tagging experience group, or tagging experience quantile. Tagging quality was not included as a model term, as the modelled dataset included only 'good' tagging quality observations. The dataset contained 18,817 observations in the minimal aggregated dataset for BIC comparison.

The model selection process described above selected for the minimal model, containing only cruise leg, fish length, and fish condition. The final model can therefore be described by:

$$\log\left(\frac{p_k}{1-p_k}\right) = \beta_0 + \beta_{cruise\_leg[k]} + \beta_{condition[k]} + f(length[k])$$

All variables were categorical, except the continuous variable *length*, which was modelled using a thin-plate regression spline denoted *f()*.

The effects plots for the terms in the selected bigeye tuna model are shown in figure 5. Similar to the models for yellowfin tuna, one level of *fish condition*, *mouth damage*, was estimated to have a slightly more positive effect than the reference level, *good*. In this case, the parameter for this level was capped at that of *good* when calculating correction factors.

Tags deployed in fish of around length 55cm had consistently higher probabilities of recovery, while fish dropped on deck were predicted to cause the most negative change due to condition in probability of recapture. Differences between *good, mouth damaged,* and *shark bite* fish conditions appeared to be negligible in effect for tag recapture probability. Un-modelled cruise legs were, again, set to the cruise leg with the median, estimated model coefficient.

Final correction factors (using *good* fish condition for control) for all bigeye tuna tag releases ranged between 1 and 0.523 (for a 39cm fish, released with *eye damage*). Once aggregated to assessment release group, correction factors ranged from 1 to 0.946.

4 Discussion

The construction of tagging and conditions effects models is a standard auxiliary analysis supporting the stock assessment of skipjack, yellowfin and bigeye tuna. Here, we have re-examined some of the model terms previously used in these analyses, and described a model selection process for choosing a parsimonious model driven by the tagging data available. In addition, we have applied the same process to tag releases from all WCPO tagging programmes, now pooling the data for model estimation.

There are relatively low levels of overlap between tagger and individual tagging events for both bigeye and yellowfin. This effect was greatest in bigeye, where the pool of taggers is much smaller owing to the logistics of historically BET-focussed central Pacific cruises. For this reason, it is recommended that the simultaneous use of both these terms in tagging effects and conditions models is discontinued as models are unlikely to sensibly disentangle the effect between individual taggers and tagging events.

The results from this year's tagging effects models show broadly the same patterns from previous work on the subject (Berger et al. 2014, Hoyle et al. 2015), where condition, length and the spatiotemporal nature of tagging events are the most important factors in predicting probability of tag recapture. However, removal of individual tagger as a model term, with which tag releases are subsequently corrected for, can result in correction factors that scale the effective tag releases far less than in previous studies. This was particularly the case for bigeye tuna, where the estimated effect of experience group on recapture probability was very low. Such an effect is coherent with bigeye tuna being typically more docile during tagging procedures, and therefor easier to tag, but results in a much weaker correction than previously used where the reference tagger was generally an individual tagging in Western Pacific, Pole & Line cruises. For the information-criteria selected model described here, this resulted in a far lower level of release corrections for bigeye tuna than for previous assessments.

For the selected yellowfin tuna model, the inclusion of an experience group term shows that the greatest effect is for beginner taggers who have tagged less than 500 fish in total. This number is lower than that described in Hoyle et al. (2015), where it was summarised that beyond 2500 fish there was little effect in tagger experience. Here, all examined models that included experience terms for both species showed little difference between intermediate, competent and advanced tagger categories, with most actually estimating a slightly more positive effect for intermediate over competent taggers.

For both species, the estimated effect of fish condition was also similar to those of previous studies, with little apparent difference caused by *mouth damage* or cookie cutter *shark bite*s to those fish released in '*good'* condition. This is likely due to inconsistent recording of fish condition during fast tagging, as well as non-obvious indicators of poor condition being missed in many fish of apparently '*good*' condition.

While here we have demonstrated an objective method to model selection in tagging effects models, there remains several improvements that should be developed. In particular, the most consistently important term for all models is the 'catch-all' term of either individual tagging event or cruise leg, which due to the spatiotemporal distribution of fishing effort alongside the behaviour of the tagged school, predicts probability of recapture more than other covariates. Estimating tagging effects models on short term release-recapture events, and including the local known fishing effort at that time, may allow more precise estimation of the effects of condition and experience, and thus more appropriate corrections for all effective tag releases used in the stock assessment process. Similarly, pooling data across species, with a thorough examination of multicollinearity, multiple interactions between species and length, condition and other covariates, may also allow more accurate estimation of key model terms that are used to control for correction factor calculations.

Recommendations:

- Continue to apply this objective model selection framework to future tagging effects analyses prior to calculation of correction factors for future assessments
- Re-examine combining tags for skipjack, yellowfin and bigeye tuna in modelled datasets, with the inclusion of interaction terms between species and covariates that are likely to vary for each of the three tunas
- Explore potential terms with mechanisms that capture the spatiotemporal fishing effort at time of release, replacing the need for variance in recapture probability to be replicated by abstract tag event or cruise leg covariates

References

Berger A.M., S. McKechnie, F. Abascal, B. Kumasi, T. Usu & S.J. Nicol (2014). Analysis of tagging data for the 2014 tropical tuna assessments: data quality rules, tagger effects, and reporting rates. WCPFC-SC10-SA-IP-06.

Freckleton, R. P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology*, *65*(1), 91-101.

Guisan, A., Edwards Jr, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, *157*(2-3), 89-100.

Hampton, J. (1997). Estimates of tag-reporting and tag-shedding rates in a large-scale tuna tagging experiment in the western tropical Pacific Ocean. *Oceanographic Literature Review*, *11*(44), 1346.

Hoyle, S. D., Kolody, D. S., & Nicol, S. (2013). Analyses of tagging data for tropical tunas, with implications for the structure of WCPO bigeye stock assessments. In *WCPFC Scientific Committee, 9th Regular Session*.

Hoyle, S. D., Leroy, B. M., Nicol, S. J., & Hampton, W. J. (2015). Covariates of release mortality and tag loss in large-scale tuna tagging experiments. *Fisheries Research*, *163*, 106-118.

Paul, R. K. (2006). Multicollinearity: Causes, effects and remedies. *IASRI, New Delhi*, 58-65.

Peatman, T., Scutt Phillips, J., Roupsard, F., Sanchez, C., Leroy, B. & Smith, N. (2019). *Analysis of tag seeding data and reporting rates*. WCPFC-SC15-2019/SA-IP-06, Pohnpei, Federated States of Micronesia, 12–20 August.

Vincent, M.T., Y. Aoki , H. Kiyofuji, J Hampton, G.M. Pilling, 2019. Background analyses for the 2019 stock assessment of skipjack tuna. WCPFC-SC15-2019/SA-IP-04.

5 Figures

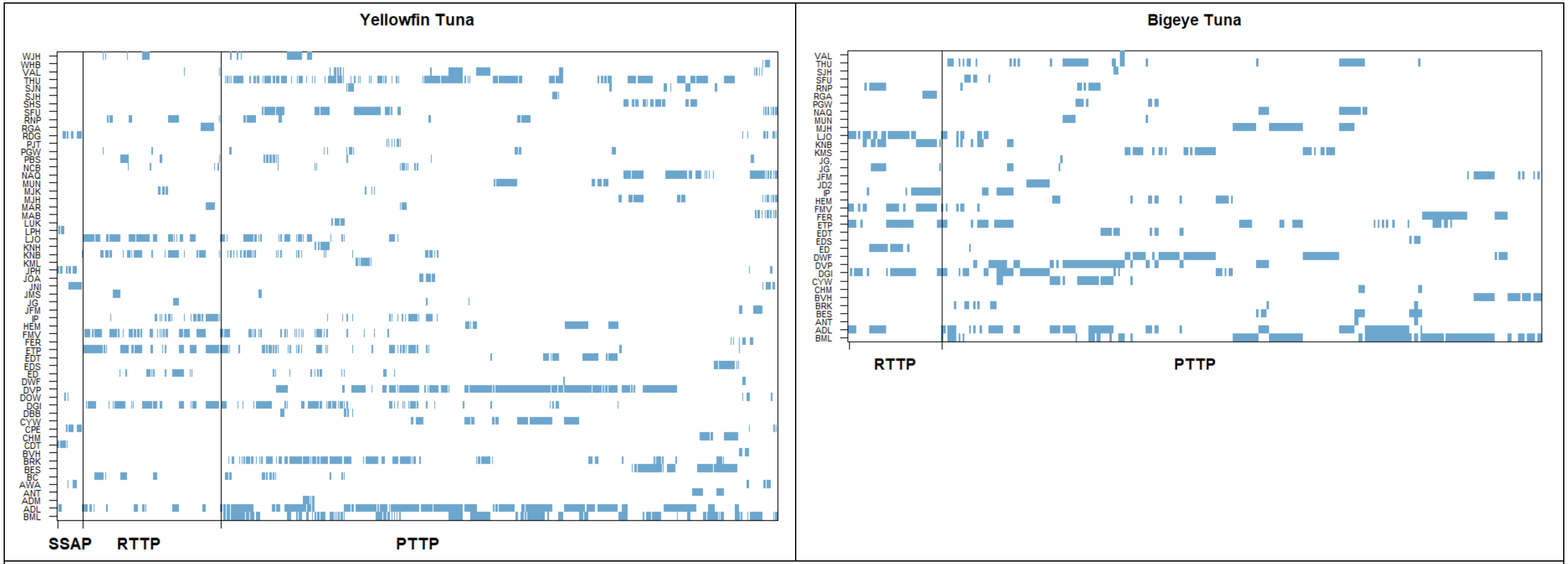

Figure 1. Distribution of observed tagging releases across individual events (x-axis) and individual taggers (y-axis). Blue marks show the presence of data for these combined levels of the two categorical model terms.

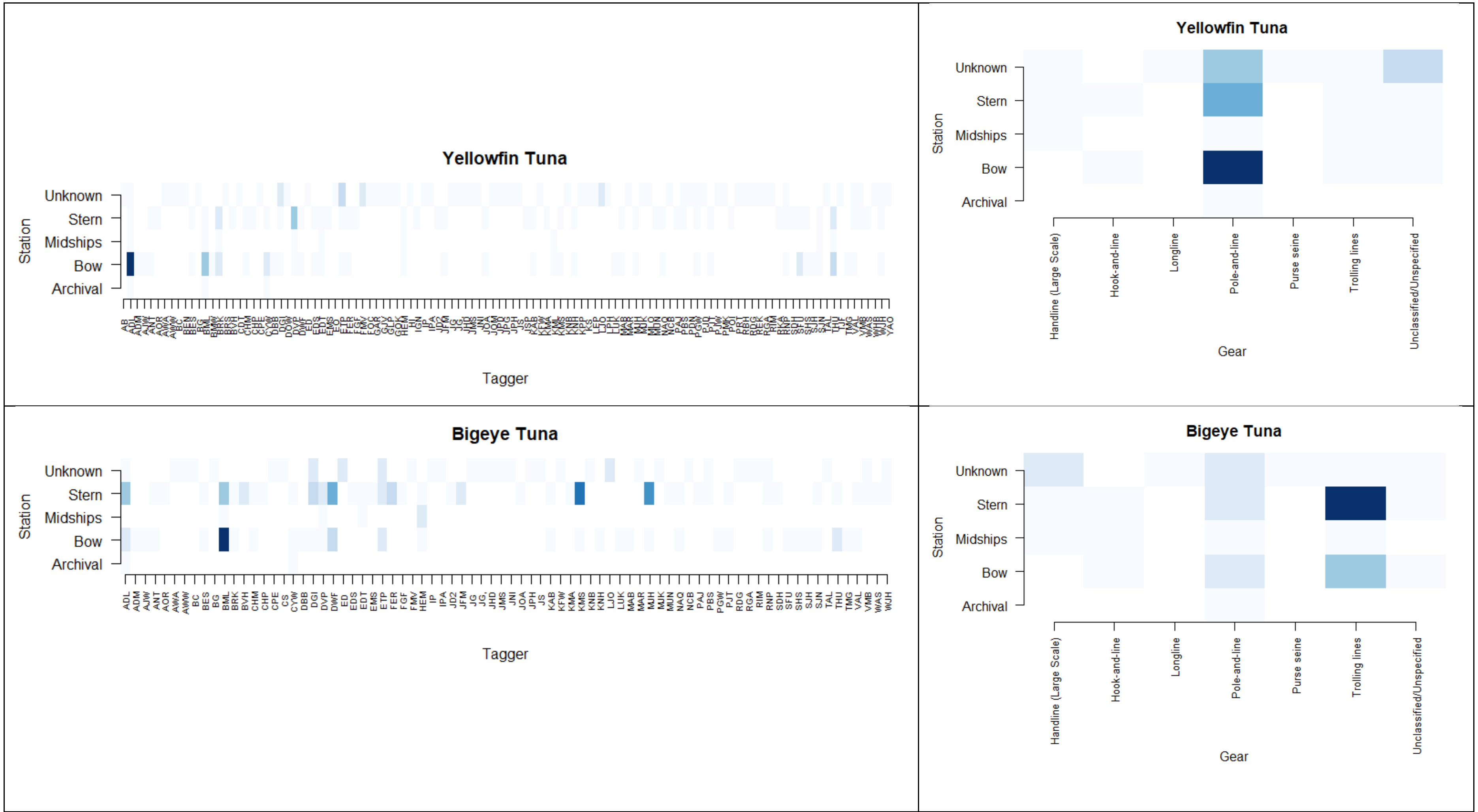

Figure 2. Distribution of relative numbers of observations across individual taggers and stations (left), and gear types and stations (right). Results are separated by species (top: yellowfin tuna, bottom: bigeye tuna).

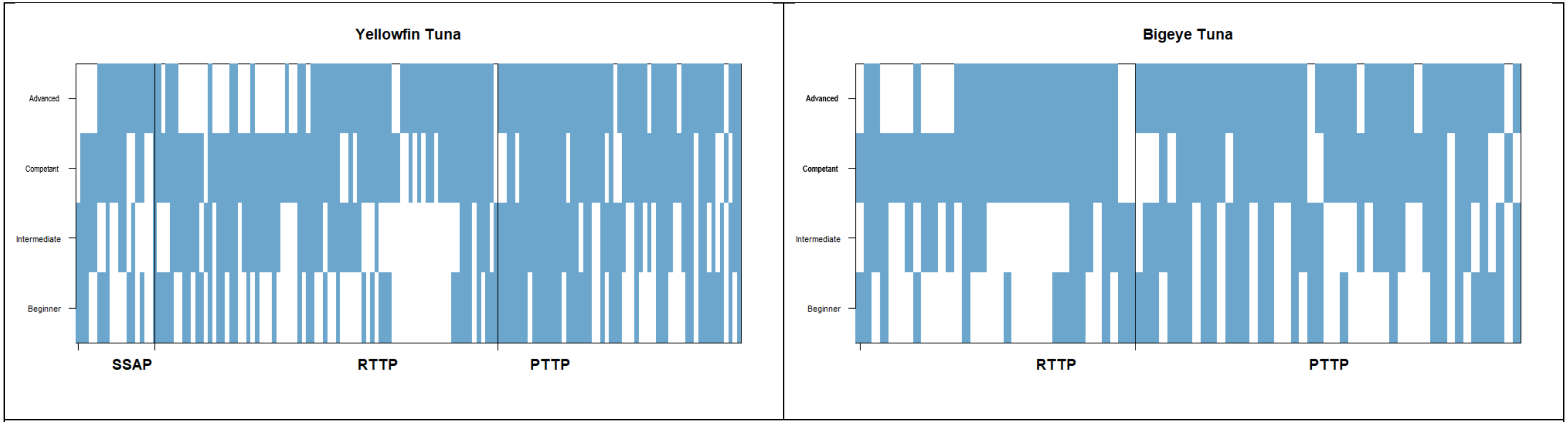

Figure 3. Distribution of observed tagging releases across cruise legs (x-axis) and the tagger experience group (y-axis). Blue marks show the presence of data for these combined levels of the two categorical model terms.

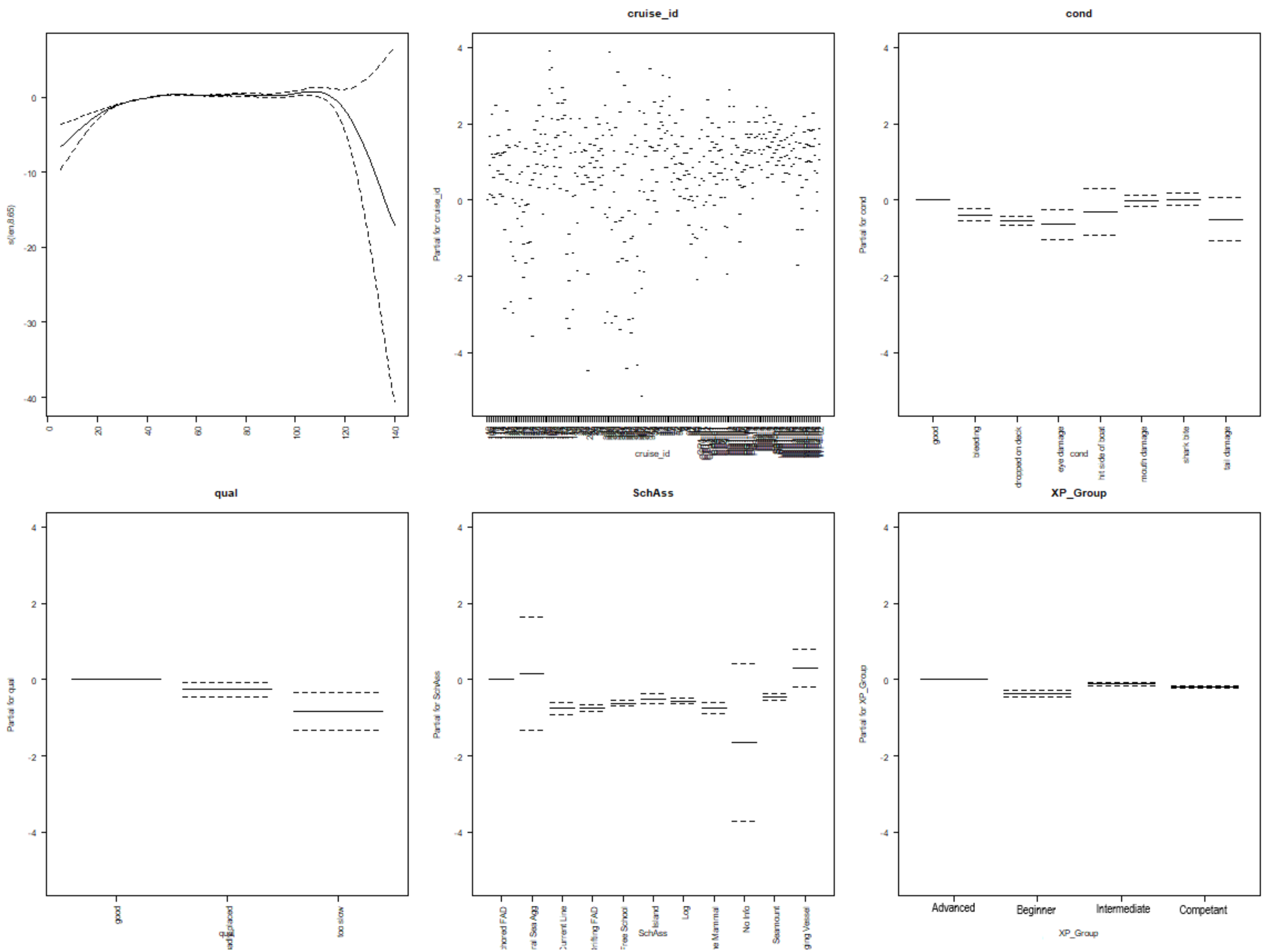

Figure 4. Partial term effects plots for the BIC-selected yellowfin tuna tagging effects model.

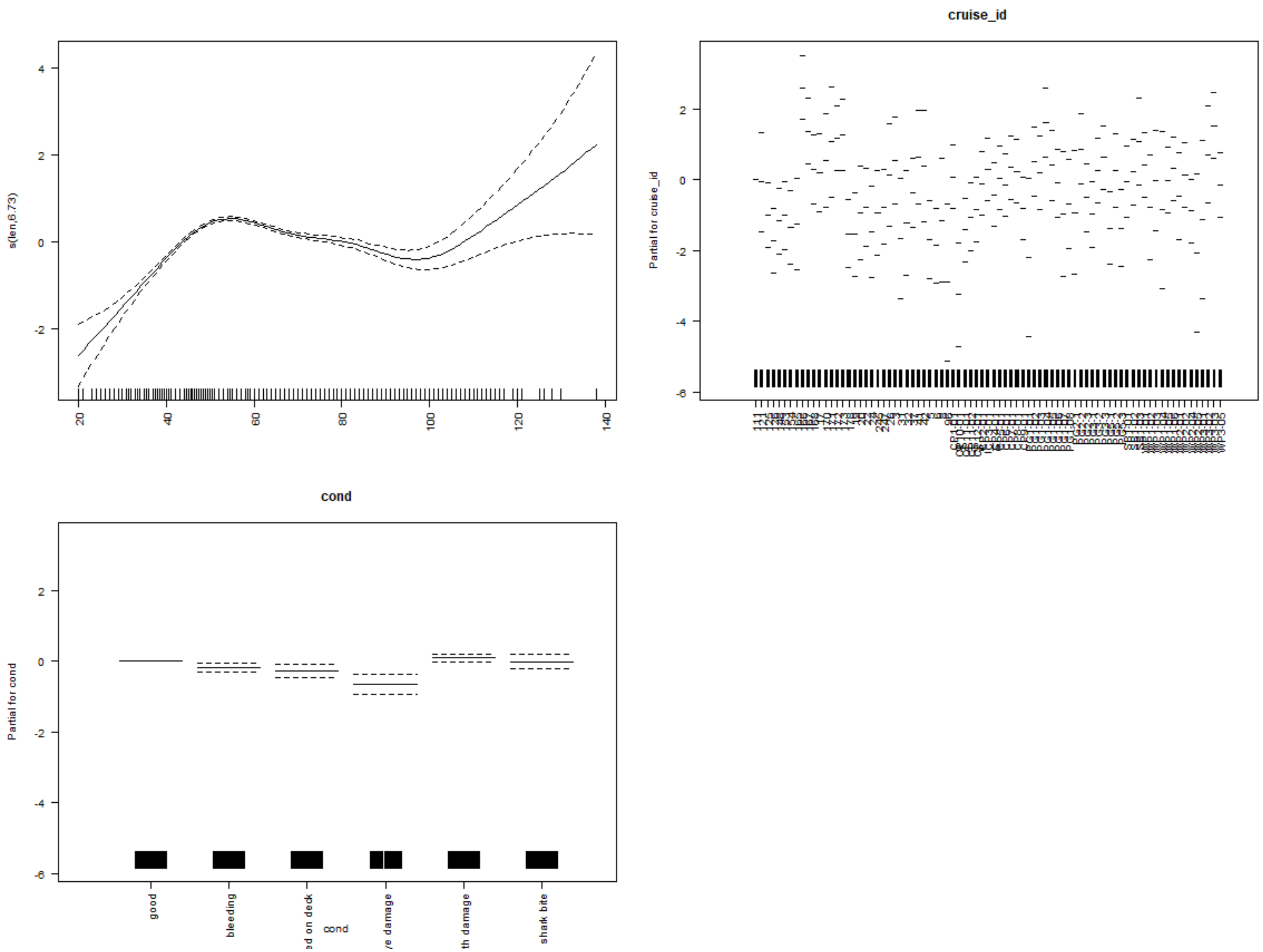

Figure 5. Partial term effects plots for the BIC-selected bigeye tuna tagging effects model.