



**SCIENTIFIC COMMITTEE
FIFTEENTH REGULAR SESSION**
Pohnpei, Federated States of Micronesia
12–20 August 2019

**Historical catch reconstruction and CPUE standardization for the stock
assessment of oceanic whitetip shark in the Western and Central Pacific Ocean**

WCPFC-SC15-2019/SA-IP-17

Laura Tremblay-Boyer¹ and Philipp Neubauer¹

¹Dragonfly Data Science, Wellington, New Zealand



Historical catch reconstruction and CPUE standardization for the stock assessment of oceanic whitetip shark in the Western and Central Pacific Ocean

Authors:

Laura Tremblay - Boyer
Philipp Neubauer



PO Box 27535, Wellington 6141
New Zealand
dragonfly.co.nz

Cover Notes

To be cited as:

Tremblay-Boyer, Laura; Philipp Neubauer (2019). Historical catch reconstruction and CPUE standardization for the stock assessment of oceanic whitetip shark in the Western and Central Pacific Ocean, 75 pages. WCPFC-SC15-2019/SA-IP-17. Report to the WCPFC Scientific Committee. Fifteenth Regular Session, 12–20 August 2018, Pohnpei, Federated States of Micronesia.

EXECUTIVE SUMMARY

This is a companion paper to the 2019 stock assessment for oceanic whitetip shark (*Carcharhinus longimanus*) (Tremblay-Boyer et al. 2019) detailing the historical catch reconstructions and CPUE standardization approaches used to create the data inputs to the assessment. We repeat here the methods from the main assessment report, provide a discussion of key results and present detailed model diagnostics on the observer data-based catch reconstructions and CPUE standardization. Other inputs to the assessment were described in Tremblay-Boyer et al. (2019).

Historical catches were reconstructed based on observer catch rates as logbook-reported catches of oceanic whitetip shark were considered unreliable over the assessment period of 1995–2016. We developed a prediction-model from observer catch rates to apply to known longline and purse-seine effort across the Western and Central Pacific Ocean. Catches aggregated across fleets were predicted to peak in 2001 and decline thereafter. The catches for the longline bycatch fleet are predicted to be the highest of all fleets, with catches for the other fleets negligible in comparison. This allocation probably reflects in part the reliability of catch rates recorded by the observer programmes assigned to each fleet, as well as the effort information used to extrapolate catches. The catches predicted were generally higher than those estimated from the same dataset by previous authors but predictions were also made here for CCMs that are typically excluded from these analyses due to a poor characterization of effort. The catch predictions otherwise showed good agreement for key years with an alternative time series that was developed based on global fin trade statistics.

A single standardized index was developed for this assessment based on the longline bycatch fleet following discussions at the 2019 Pre-Assessment Workshop (PAW) (Pilling & Brouwer 2019). This fleet was chosen as the underlying data were deemed the most reliable of those available, and also covered the greatest extent within the assessment region. The standardized year effects resulting from the CPUE model were highly variable at the start of the time series and started steadily declining from 1999 onwards. Influential covariates in the model included sea surface temperature, hooks-between-float as a proxy for set depth, species targeting cluster and observer programme.

The analyses presented here included years from the earlier period of observer programmes in the WCPO. Observer reporting for sharks for this period is known to be unreliable as well as highly variable given the sparse coverage. Any results for this early time period must be interpreted carefully. In addition, we found evidence of observer programmes having an impact on recorded catch rates independent from vessel flag, such that observer training, as would be expected, is a likely influence on the catch rates recorded for sharks across time and programmes. As this is a variable that is challenging to quantify and so cannot be standardized for, the results presented here should be interpreted in light of this caveat.

CONTENTS

1	INTRODUCTION	5
2	METHODS	6
2.1	Description of datasets	6
2.2	Catch reconstruction	7
2.2.1	Prediction of catch rates from observed sets	7
2.2.2	Extrapolation of observed catch rates to WCPO-wide effort	11
2.3	CPUE standardisation	13
3	RESULTS	14
3.1	Species targeting clusters	14
3.2	Models of catch rates based on observer data	14
3.2.1	Longline bycatch fleet	14
3.2.2	Longline target fleet	15
3.2.3	Purse seine fleets	16
3.3	Classification of deep vs. shallow set for WCPO longline effort	17
3.4	Historical catch reconstructions	17
3.5	CPUE standardization for the longline bycatch fleet	18
4	DISCUSSION	20
5	REFERENCES	22
6	TABLES	24
7	FIGURES	28
A	GLMMS DIAGNOSTICS	44
A.1	Diagnostics for the observer catch rates model for the longline bycatch fleet	44
A.2	Diagnostics for the observer catch rates model for the longline target fleet	51
A.3	Diagnostics for the observer catch rates model for the purse seine fleet	58
A.4	Diagnostics for the standardized CPUE for the longline bycatch fleet	65
B	ESTIMATED OCEANIC WHITETIP SHARK CATCH BY FLEET	72

1. INTRODUCTION

This is a companion paper to the 2019 stock assessment for oceanic whitetip shark (*Carcharhinus longimanus*) (Tremblay-Boyer et al. 2019) detailing the historical catch reconstructions and CPUE standardization approaches used to create the data inputs.

Historical catches for sharks in the Western and Central Pacific Ocean were poorly recorded until 2010, when CMM2010-07 became active, mandating the reporting of catches of key shark species, including oceanic whitetip shark. Nevertheless, reported shark catch was likely underestimated even after CMM2010-07 came into effect. As logbook-reported catches of oceanic whitetip shark were considered unreliable for the present assessment period, we applied two strategies to reconstruct catches. In the first instance, we created a prediction-model from observer catch rates to apply this model to known longline and purse-seine effort across the Western and Central Pacific Ocean. This approach is similar to that used for the 2012 assessment (Rice 2012) but we implemented it in different modelling framework. In the second instance, we applied the trade-based approach from Clarke (2018) to predict global catches of oceanic whitetip shark based on fin trade statistics. The latter were apportioned to the Western and Central Pacific Ocean using a set of alternative scaling methods. We describe and discuss the approach relying on observer catch rates models here and compare them to other estimates, including those for the trade-based approach (described in Tremblay-Boyer et al. 2019).

The previous assessment for this species (Rice & Harley 2012) included standardised indices for each of the four fishing fleets included in the stock assessment model, but only the index for the longline bycatch fleet was used in the reference case or the structural uncertainty grid. While the coverage rates of observed longline effort were considerably lower than for purse seine, the reliability of catch numbers estimated by longline sets is considered much higher as observers report every hook with a positive catch event. At the SPC 2019 Pre-Assessment Workshop (PAW) (Pilling & Brouwer 2019) it was agreed to focus on developing standardised CPUE indices for the longline bycatch fleet, as the underlying data were deemed the most reliable of those available, and also covered the greatest extent within the assessment region.

We provide here a discussion of key results and present model diagnostics on the observer data-based catch reconstructions and CPUE standardization used as input for the oceanic whitetip shark assessment. The methods outlined here were repeated in Tremblay-Boyer et al. (2019) for context. We also refer the reader to the main assessment report for a survey of current knowledge of the biology of the species, the processing of the catch-at-length data, as well as the discard mortality models that were applied to the catch reconstructions described here.

2. METHODS

2.1 Description of datasets

Datasets from the database of the Pacific Community (SPC) included catch, effort and observer data.

- **L_BEST:** SPC's best estimates of longline catch and effort (in hooks) for fleets in the WCPFC Convention Area (WCPFC-CA), available at the $5^\circ \times \text{month} \times \text{year} \times \text{flag} \times \text{fleet}$ resolution for key species of tuna and billfish, and sharks in some years. A version of this database (**L_BEST.HBF**) was available with an additional strata for hooks-between-floats (HBF), but effort coverage is uneven over fleets and years.
- **S_BEST:** SPC's best estimates of purse-seine catch and effort (sets and days) for fleets in the WCPFC-CA, available at the $1^\circ \times \text{set type} \times \text{month} \times \text{year} \times \text{flag} \times \text{fleet}$ resolution for key species of tuna and billfish, and sharks in some years.
- **Observer programmes for longline and purse-seine fleets:** The full observer dataset for longline and purse-seine fleets available to SPC was used for the analysis, including data from the SPC's Regional Observer Programme and national observer programmes. Records collected by longline observers that are relevant to this assessment are key gear and attributes (including date and time, location, HBF) and, for each observed hook with a positive catch event, the species, the fate of the catch (e.g., discarded or retained), the condition, the length and the sex of the individual. The quality and coverage for most variables changes over time and between programmes. For observed purse-seine sets, observers estimate the number of sharks caught of a given species from the brail net and, when possible, measure their length.

Data preparation

Extracts from SPC's databases were obtained in April 2019. All datasets were filtered to retain records within the Western and Central Pacific Ocean area only (Figure 1), over the period of the stock assessment from 1995 to 2016. For the longline observer datasets, number of hooks observed, when missing, was estimated from the product of hooks-between-floats and the number of baskets observed. Sets were classified as shallow when the number of HBFs was lower or equal to 10, following Peatman et al. (2018b). Oceanography covariates (sea surface temperature, chlorophyll-*a*, bathymetry and distance from the coast) were extracted at the lowest resolution possible and aggregated to match the resolution of each dataset. Longline sets occurring in sea surface temperatures below 16°C were removed as they were considered to be outside of oceanic whitetip shark habitat. When relevant, extreme values of oceanography covariates were bounded or filtered out ($> 99.5^{\text{th}}$ quantile). Records from implausible locations (e.g., on land) were omitted. Longline records without HBF information and purse-seine records without set-association information were also omitted.

2.2 Catch reconstruction

2.2.1 Prediction of catch rates from observed sets

Previous approaches to reconstruct catches for this species have also been based on observer catch data (see Lawson 2011, Rice 2012, Peatman et al. 2018b). The basis for these methods is similar: a model of catch-per-unit-effort is built based on observed sets and relevant covariates, and the model is then used to predict catches based on a reliable measure of total effort by fleet across the assessment region. The previous approaches differ in the modelling framework used to build the catch rate model, the covariates considered and the treatment of uncertainty. Lawson (2011) and Rice (2012) both used Generalised Linear Models (GLMs), assuming delta-log-normal error distributions (i.e., two-stage or hurdle model), but Rice (2012) filtered the data more extensively (e.g., only sets at SST $\geq 25^{\circ}\text{C}$ were retained) and permitted extra variability around the year effects. Uncertainty around model predictions of catches was not explicitly considered. Peatman et al. (2018b) used Generalised Estimating Equations (GEEs) to model catch rates, also with a delta-log-normal model structure. The GEE framework allows for the correlation between observed sets in the same observer trips to be accounted for. Catch predictions and uncertainty were estimated with a Monte Carlo simulation approach drawing samples from modelled catch distributions.

Here, we used Markov chain Monte Carlo (MCMC) methods to model catch rates in number of individuals for the observer longline and purse-seine datasets, assuming a negative binomial error distribution. Negative binomial error distributions are well suited at representing catch rates, as they can naturally account for high proportions of zero in the response variable without the requirement of a parallel model (such as required when assuming delta-log-normal error distribution). Negative binomial error distributions can also predict infrequent but high catch events. An advantage of the MCMC approach to fitting GLMs is that the uncertainty for any estimated parameter or derived quantity can be easily estimated by drawing from the posterior samples of the converged MCMC chains. Because of this, the scale of alternative catch scenarios can be informed by model-derived uncertainty instead of user-defined multipliers.

We used the R package “brms” (Bürkner 2017) to implement our approach. This package provides an efficient interface for fitting GLMs in the Stan language for Bayesian statistics (Carpenter et al. 2015). In addition, the brms package allows the user to customise probability distributions to improve their suitability for representing some features of the response variables and, therefore, improve the quality of the fit. Although the GLMs were fitted within a Bayesian framework, we did not use informative priors for any of the models. Finally, all models included a random effect for the vessel flag (Table 1), allowing the prediction of a distribution for flag effects, which can then be used to predict catches for countries without any observer coverage.

There was a high degree of overdispersion in the response variable, with most fishing sets reporting no captures ($\geq 90\%$ and $\geq 99\%$ zero sets in the longline and purse-seine datasets, respectively). The impact of varying effort by fishing record on the probability of a positive

capture event was accounted for by parameterising the negative binomial distribution by the number of “trials”, defined by the number of fishing hooks or sets used in the fishing set group. In addition, a new parameter ν was added to the parameterisation of the negative binomial distribution to allow more flexibility in how the overdispersion behaves as mean catch rates increase.

In the model, catches, c_i , in a longline or purse-seine set group, i , were thus modelled as samples from a negative-binomial distribution:

$$c_i \sim \text{NegativeBinomial}(\text{mean} = \mu_i n_i, \text{shape} = \theta n_i), \quad (1)$$

where n_i is the number of hooks or sets. The shape parameter, θ , allows for extra dispersion in the number of captures relative to a Poisson distribution. The negative binomial distribution has the property that the mean of n samples from a negative binomial distribution ($\text{NegativeBinomial}(\mu, \theta)$) is itself negative binomially distributed, with mean μn and shape θn . For this reason, while c_i is the number of catches per group, μ_i needs to be interpreted as the mean catch rate per longline hook or purse-seine set. The custom distribution facility of brms was used to code the negative binomial distribution for aggregated data. The mean capture rate within each group was then estimated as the exponential of the linear predictor, which was the sum of fixed and random effects.

A novel configuration of the negative binomial distribution was trialled in model fitting. Under the usual approach to fit a negative binomial GLM, overdispersion compared to mean catch rates μ is determined by the estimate of a single parameter θ assumed for all observations:

$$\psi = \mu + \frac{\mu^2}{\theta}. \quad (2)$$

This aspect can be challenging when combinations of covariate levels have considerably higher catch rates than others, as high μ combined with low θ can result in error distributions predicting implausibly high values (i.e., very long tails) at times. Although estimating covariate effects on θ as part of the model is possible, the results can be difficult to interpret as μ and θ are often correlated. For this reason, any covariate effect attributed to θ might otherwise be confounded with a covariate effect on μ . We modified instead the definition of θ , so that it includes a new parameter, ν , scaling the extent of overdispersion as a function of μ :

$$\theta \rightarrow \mu^\nu \theta, \quad (3)$$

so that overdispersion to the negative binomial distribution becomes:

$$\psi = \mu + \frac{\mu^2}{\mu^\nu \theta}. \quad (4)$$

This configuration allows for the overdispersion parameter to change as a function of μ : as ν approaches 2, it cancels out μ^2 in the numerator, so that the negative binomial distribution effectively becomes a Poisson distribution; as ν approaches 0, the additional μ^ν term goes to 1 and the distribution behaves in the usual way. Therefore, adding this new term in the model allowed for additional flexibility in the realised error distribution between observations with the estimation of a single additional parameter.

Prior to fitting, all observed sets were first aggregated to a spatial resolution of 5° to match the resolution of the L_BEST datasets (S_BEST has a 1°-resolution, but a 5°-resolution was chosen for both datasets for consistency in predictions), and observer programme, flag, year, month and set depth (for longline, shallow or deep) or set type (for purse seine). Because of the low observer coverage for some fleets, year and locations, and the aggregation at the 5°-scale, a minimum number of records were removed for the catch reconstruction component of the analysis. For the longline fleet, aggregated records with less than 50 hooks observed in total were removed. Each of the remaining aggregated records was considered a “fishing set group” and catch rates of oceanic whitetip shark were calculated over all sets in the event to use as the response variable in the GLM. The number of hundred hooks of the sets of the fishing event was treated as the number of “trials” in the negative binomial distribution.

Candidate model covariates were selected to retain operational features of sets likely to impact catch rates and environmental variables that might be representative of oceanic whitetip shark habitat and, therefore, local abundance (Table 1). We were limited in the choice of covariates by their availability in the L_BEST and S_BEST datasets, as model predictions from these datasets require all model covariates to be available. All models were also trialled with and without the addition of ν to the negative binomial distribution.

Models were fitted with four separate chains and 2000 iterations, including a 1000 iterations burn-in period that was discarded from posterior samples. Best model selection was performed on the basis of model diagnostics (including chain convergence) and leave-one-out cross-validation (LOO; Vehtari et al. 2016). The LOO Information Criterion (LOOIC) is a Bayesian equivalent to the Aikake Information Criterion (AIC) metric that balances additional complexity in model structure against the improvement in model performance. Models using the same dataset and nested model structures can be directly compared, with lower LOOIC values indicating models that maximize fit and minimize complexity.

Three independent catch rate models were optimised for the catch reconstruction based on the fleet.

Longline bycatch fleet

The bycatch model for the longline bycatch fleet was fitted separately from the target model based on different assumptions underlying the response variable. Records belonging to vessel with flags from Papua New Guinea and Solomon Islands were removed from the analysis. In previous research by Rice (2012), sets with evidence of shark targeting were also removed from the analysis (e.g., the use of wire traces, shark lines or the specification by the observer that the

set was targeting shark). Owing to the poor coverage and reliability of these covariates over fleet and years, we were more conservative in retaining data, and did not filter sets based on these variables.

The best model for the longline bycatch fleet was:

$OCS.obs \mid trial(sets) = Year + s(SST, k=3) + HBF.cat + cluster + (1Flag) + (1yy:Flag),$

including the ν coefficient to scale overdispersion as a function of average catch rates.

Alternative models included observer programme instead of flag, different configurations for the oceanography covariates and the flag-year interaction, and also the modelling of the overdispersion.

Longline target fleet

Little is known about the longline target fleet, so that it was not included in recent catch reconstructions (e.g., Peatman et al. 2018b). An estimation of oceanic whitetip captures assumed that 5% of effort in the L_BEST database was targeting sharks, but apart from this aspect, predictions of catches for the targeting fleet were from the same model as that used for the bycatch longline fleet by Rice (2012).

In the current assessment, we created a model of catch rates for the target longline fishery. Given the scarcity of observer records for the countries with this fishery, a subset of representative flags from Pacific Island countries and territories was retained, in addition to Papua New Guinea and Solomon Islands sets (American Samoa, Kiribati, Cook Islands, Fiji, Federated States of Micronesia, Marshall Islands, Samoa and Tuvalu). Recent records from Papua New Guinea and Solomon Islands with observer programmes from distant-water nations were removed, as they were considered to be unlikely to be representative of domestic fisheries (but noting that there is some evidence of shark targeting by distant-water nations too). Observer records from Papua New Guinea for 1996, 2000 and 2003 were removed as the total catches of oceanic whitetip shark in these years were very low, indicating that observed catches might not have been recorded. There is evidence that the shark target fisheries in the Bismarck Sea region have stopped (White et al. 2018); however, based on the uncertainty of the timespan of target fisheries, we assumed that target fisheries were ongoing for the catch reconstruction.

The best model for the longline target fleet was:

$OCS.obs \mid trial(sets) = yy + s(SST, k=3) + hbf.cat + cluster + (1:Flag),$

including the ν coefficient to scale overdispersion as a function of average catch rates.

This model did not include a flag-year interaction by design, so that the overall temporal trend in catches would be informed by the catch observed in other Pacific country fleets, but the scale determined by the vessel flag. This approach was used to compensate for the lack of reliable records for the target fisheries for some years.

Purse-seine fleet

A single model of observed catch rates was built for the purse-seine fleet, including associated and unassociated sets, but with set type as a covariate to allow for reconstructed catches to be predicted for each fleet separately.

The best model for the purse-seine fleet was:

$OCS_{obs} | trials(sets) \sim yy + s(dist2coast, k=3) + (1 | Flag) + (1 | yy:Flag) + SetType.$

This model had a random effect for flag and a random-effect interaction for year and flag, and fixed effect for set type and the distance of sets to the nearest coast. The model including ν did not result in a considerable improvement to the fit according to the LOOIC metric, presumably because there was less variation between flags fishing the same area in purse-seine than for longline catch rates.

Alternative models considered observer programme instead of flag, different configurations for the oceanography covariates and the flag-year interaction, and also the modelling of the overdispersion.

2.2.2 Extrapolation of observed catch rates to WCPO-wide effort

Catch rates predicted from the observer models were projected at the scale of the Western and Central Pacific Ocean based on estimates of effort from the L_BEST and S_BEST datasets.

For L_BEST, species-targeting clusters were predicted from species proportion for each record (as described in Table 1). Hooks-between-float information was missing for numerous records in L_BEST, especially in earlier years. HBFs are a proxy for the depth of the longline set, and a key factor for predicting and extrapolating catches of oceanic whitetip shark, as this species primarily occupies surface waters.

CCMs recently started reporting longline catch and effort statistics disaggregated by HBF; however, coverage for many countries remains lacking over most or part of the time period of this assessment.

In the previous catch reconstruction by Peatman et al. (2018b), ratio estimators were used to classify L_BEST records that were missing HBF information. Here, we used a Random Forest model instead (Liaw & Wiener 2002), as it allows the inclusion of covariates to predict the likely depth of sets (instead of assuming that unclassified sets for a country are directly representative of classified sets). The Random Forest model also provides outputs for the probability of a record having a given classification, which can be used to propagate uncertainty about this step into catch estimates as needed.

We used the dataset of HBF-disaggregated L_BEST to train a Random Forest model to predict whether a record should be assigned to a shallow- (<10 HBF) or deep-effort category (≥ 10 HBF), assuming a binomial error distribution:

- We used the Random Forest algorithm provided in the R package “randomForest” (Liaw & Wiener 2002).
- Covariates used to build the tree were: year, month, targeting cluster, 5°-longitude cell, 5°-latitude cell, and catches for albacore, yellowfin, bigeye and bluefin tuna, swordfish and other billfish (in numbers). We did not use shark catches, even though these data are available in L_BEST and likely provide information about set depth since sets that target sharks tend to be shallow. These data were not included as low shark catches in the early part of the time series are misrepresented owing to the lack of reporting.
- The L_BEST.HBF dataset was split evenly between a training and a testing dataset. The training dataset was used to fit the Random Forest model and model performance was assessed by predicting HBF classification for the testing dataset.
- The Random Forest model was tuned by first running the model with a high number of trees (500), and verifying the Area-Under-the Curve score to assess the number of trees required to reach a plateau. Five covariates were randomly considered at each node for splitting. The optimal tree depth was assessed to be 200.

The Random Forest model with optimised parameters was used to assign shallow- or deep-effort depth to L_BEST records lacking HBF information, based on the probabilities estimated by the binomial error model. Where partial HBF information for a stratum was available, predictions were only made for the effort lacking HBF classification. An uncertainty distribution of the final predictions of HBF classification for L_BEST by member countries was estimated by drawing 1000 draws from a Bernoulli distribution for each L_BEST record, assuming the probability of success is the probability estimated by the Random Forest model of the record being classified as deep. Predictions of the proportion of deep sets over time by fleet is shown in Figure 14, including the fit to the training dataset and 95% credible intervals based on sampling of the error distributions.

Once all effort in L_BEST was assigned to a shallow- or deep-set category, we were able to make predictions of L_BEST-wide catch for the assessment region, based on the catch models developed for the longline bycatch and target fleets (Figures 15 and 16). Predictions for the target longline fleet were derived on effort assigned to Papua New Guinea and Solomon Islands only.

The S_BEST dataset already contained all the required covariates so we were able to proceed with the historical catch predictions for the associated and unassociated purse-seine fleets without further processing of covariates. For all fleets, predictions of catches over time by fleet and year were aggregated for each model over each posterior draw, and summary statistics extracted. The median prediction was used as a baseline catch scenario, and the 90th quantile of the predictions as a high catch scenario.

2.3 CPUE standardisation

A similar approach to fitting the CPUE models was used to that for the catch reconstruction model, including the use of a negative binomial error distribution with an additional ν parameter (Section 2.2.1). In practice, the year effects estimated from the catch reconstruction models can be considered to be standardised CPUE rates. Nevertheless, a constraint to this approach of fitting the catch reconstruction models was that total catch rates would be predicted from the L_BEST and S_BEST datasets, which meant that (1) sets had to be aggregated at a lower spatial resolution of 5 degrees (for the longline dataset), and (2) only covariates available in L_BEST and S_BEST could be considered. In addition, as we were predicting catches across all flags fishing in the Western and Central Pacific Ocean, we had to retain as many flags as possible in the analysis to inform the value of the flag effect in the random effect distribution. Otherwise, the flag effect for a country fishing in the Western and Central Pacific Ocean but excluded from the model fit would be randomly drawn from that distribution. For this reason, we retained as many different flags as possible in the model dataset. These constraints were not present for the CPUE analysis as the key result was the estimated year effects, independent from flags.

As a result, we filtered the observer dataset to retain only observer programmes that had consistent observer coverage over time across the spatial distribution of oceanic whitetip shark. These observer programmes were: American Samoa, Fiji, Federated States of Micronesia, Hawaii, Kiribati, Marshall Islands, New Caledonia and French Polynesia. Sets were filtered to retain only those occurring in $SST \geq 16^{\circ}C$; sets with catch rates higher than the 99.5th quantile of positive catch rates were excluded as they were considered to be active shark targeting. The remaining sets were aggregated over flag, programme code, HBF category (shallow or deep), year, month, and 1° cell.

A range of model structures and combinations of covariates were trialled with the objective of improving model diagnostics and minimising the LOOIC metric. Covariates are described in Table 1.

The best model was:

$OCS.obs \mid trial(sets) = Year + s(SST, k=3) + (1|Program) + s(HBF, k=4) + cluster + (1|Year:Program)$, including the ν coefficient to scale overdispersion as a function of average catch rates.

This model was similar to the catch reconstruction model for this fleet, but included hooks-between-floats as a continuous instead of categorical variable, and an effect on observer programme instead of flag. These changes were based on comparisons between alternative models using the LOOIC metric.

The first year of the CPUE time series (1995) was the model's intercept. Standardised year-effects for 1996 to 2016 were scaled according to the intercept. The MCMC draws were mean-standardised and back-transformed from the log-link, and summary statistics for each year (median, 2.5th and 97.5th) extracted to form the standardised index of abundance.

3. RESULTS

3.1 Species targeting clusters

The approach used to derive the species targeting cluster is described in Table 1. We identified 5 species targeting clusters reflecting targeting for either specific species of tuna or species assemblages (Figure 2). Three of the clusters were mainly composed of a single species of tuna (albacore tuna, bigeye tuna, yellowfin tuna). The other two clusters were mixed assemblage of swordfish and marlins (blue marlin, black marlin and striped marlin), and tuna (all three main species). Note that we could have used a higher number of clusters and improve the classification diagnostics for the *k-means* algorithm but we elected to retain broad cluster delineations, since we expected the targeting cluster to be somewhat confounded with gear and environment covariates used in the catch reconstruction and CPUE standardization models.

3.2 Models of catch rates based on observer data

3.2.1 Longline bycatch fleet

The selected model for observer catch rates for the longline bycatch fleet included a year effect, a non-continuous linear effect for SST, a categorical effect for set depth (shallow or deep), a categorical effect for the species targeting effect, and random effects for flags and the interaction between flag and year. The negative binomial distribution was parameterized with the addition of the ν parameter. We used flag instead of observer programme as a covariate in order to be able to use the L_BEST database to extrapolate catch from model predictions, and L_BEST only contains effort by flag.

This model had the best diagnostics compared to alternatives, and also the second lowest LOOIC (Table 2). Another model including an effect of chlorophyll-a in addition to that of SST had a slightly better LOOIC but the chlorophyll-a effect was very weak and thus estimated with little precision, which would have led to instability in the model predictions. We thus decided to retain SST as the sole oceanography covariate. The covariates that resulted in the most improvement in LOOIC were the ν parameter, the non-linear effect for SST, the set depth category and the year \times flag interaction. Effects of bathymetry and distance from the coast were not significantly different from 0. Note that some model structures did not converge to stable estimates so we did not consider them further.

The model predicted an overall decline in catch rates independent of other covariates, with catch rates in 2016 0.18 times what they were in 1995. There was a strong positive effect of SST on catch rates that plateaued at about 27°C (Figure 3). Catch rates for shallow sets were predicted to be 3.5 times that for deep sets, and species target clusters with higher proportions of swordfish or marlins in their catch had the effect positive effect on catch rates, followed by target clusters with high proportions of bigeye tuna and albacore tuna.

Key model diagnostics comparing the overall distribution of the observed *vs.* predicted response variable are shown in Figure 4 and 5. The overall fit across all covariates was good but

some trends appear when predictions are disaggregated by flag and year. Summary statistics for the proportion of zero catch, the mean and median catch by fishing event when catch > 0 and the 90th quantile for positive catches by fishing event when catch > 0 are shown in Figure A-24 to A-32. The former two diagnostics allow to see how well the model captures the mean or median catch rate for the data, whereas the latter shows how the overdispersion in the data is modelled. In general, the model performs better at predicting the mean or the median than the overdispersion. The proportion of zeroes was fitted well across most flag and year combinations (Figure A-24 to A-26). The mean number of individuals caught for fishing events where catch > 0 was generally well fitted except for the early period where both the predicted mean and median were slightly higher than the observed (with observed values still occurring within the interquartile bounds for the predictions, except for Taiwan where the expected value was greater) (Figure A-27 to A-29). The 90th quantile for the number of individuals caught for fishing events where catch > 0 was over-estimated by the model, but when this metric was disaggregated between years and flags the pattern was most apparent for the US in the earlier part of the time series (Figure A-30 to A-32). MCMC traces for key parameters did not show any convergence issues (Figure A-33).

3.2.2 Longline target fleet

The selected model for observer catch rates for the longline target fleet included a year effect, a non-continuous linear effect for SST, a categorical effect for set depth (shallow or deep), a categorical effect for the species targeting effect, and a random effect for flags. The negative binomial distribution was parameterized with the addition of the ν parameter. We specifically did not consider an interaction between year and flag as observer coverage was minimal for many years for the two flags assigned to the target fleet (Papua New Guinea and Solomon Islands), and we wanted to use the overall year effect estimated with the inclusion of the fleets of other CCMs in the same region to result in a more robust standardized CPUE for this fleet.

This model had the lowest LOOIC amongst the alternate structures tested (Table 3) but the impact of the addition of a SST covariate on model performance was not as important as for the bycatch fleet model as the dataset for the longline target fleet did not show as much contrast in this variable. Similarly, the effect of including chlorophyll-a, bathymetry or distance to coast was negligible. Otherwise, the two covariates that had the highest impact on model performance were set category (shallow *vs.* deep) and the addition of the ν parameter to the negative binomial distribution.

The model predicted variable catch rates throughout the time-series, and especially at the start which might have been due to observers not consistently recording shark catches at the start of the time series. There was a consistent decline in the estimated year effects from 2009. Set depth had a very strong effect on catch rates, with shallow sets predicted to have catch rates 10 times that of deep sets when other variables are accounted for. The effect of the species targeting cluster was only significant for the billfish targeting cluster showing a 1.6 times increase in catch rates compared to the intercept (the albacore targeting cluster). SST was estimated to

have a positive effect on catch rates (Figure 6) but was not estimated with much precision, especially for values of SST < 25°C.

Key model diagnostics comparing the overall distribution of the observed *vs.* predicted response variable are shown in Figure 7 and 8. The overall fit across all covariates was good but some trends appear when predictions are disaggregated by flag and year. Summary statistics for the proportion of zero catch, the mean and median catch by fishing event when catch > 0 and the 90th quantile for positive catches by fishing event when catch > 0 are shown in Figure A-34 to A-42. The proportion of zeroes was over-estimated for some flags but the fit was good for Papua New Guinea and Solomon Islands which had most of the fishing events for this model (Figure A-34 to A-36). Predicted values were less variable between years than observed since they were constrained by the random effect on the year × flag interaction (e.g. see Federated States of Micronesia). The mean number of individuals caught by fishing events where catch > 0 was well fitted when aggregating over years and flags but over-estimated for some flags when disaggregating by flag and year. The fit for Papua New Guinea and Solomon Islands was generally good but predictions for the earlier time period were uncertain. The 90th quantile for the number of individuals caught by fishing events where catch > 0 was relatively well estimated with no clear trend across flags or year, except for Federated States of Micronesia where this value was over-estimated for all years (Figure A-40 to A-42). MCMC traces for key parameters did not show any convergence issues (Figure A-43).

3.2.3 Purse seine fleets

The selected model for observer catch rates for the longline target fleet included a year effect, a non-continuous linear effect for the distance to the nearest coast, a categorical effect for set depth (shallow or deep), a categorical effect for the species targeting effect, and random effects for flags and the interaction between flags and years.

The selected model had a low LOOIC amongst the alternatives considered (Table 4). The covariate that had the most impact on model performance was the addition of the year × flag interaction. Lower LOOICs could have been achieved by including oceanography covariates like bathymetry and chlorophyll-a, but these were generally not well estimated by the model: either the effect was very small or the uncertainty bounds around the estimated spline would have resulted in unreliable predictions. We also did not consider SST as a covariate since there was minimal contrast in this variable 99% of fishing events occurred between SST of 27.5°C and 31°C. Early exploration showed that the addition of the ν parameter made little difference and it was not included when considering candidate model structures.

The model predicted very low catch rates in 1995 the intercept year (the lowest of the time-series), annual catch rates were then high up to 2002, and then declined from that point onwards. The set type used for the intercept was the anchored FAD. Unassociated sets had significantly lower catch rates in comparison, and drifting FADs had the highest effect on catch rates, with an increase of approximately 3 fold compared to the anchored FAD. The effect of set distance to coast was weak and estimated with low precision for distances beyond 400km,

but showed that in general, sets closer to coastlines were expected to have slightly higher catch rates (Figure 9).

Key model diagnostics comparing the overall distribution of the observed *vs.* predicted response variable are shown in Figure 10 and 11. The overall fit across all covariates was good. When predictions were disaggregated between years, flags and flags \times year (Figure A-44 to A-52) there was a clear pattern where the proportion of zero catch events was over-estimated by the model in the earlier time-period, which came primarily from the fit to US-flagged trips (Figure A-46). Otherwise, the model under-estimated the 90th quantile for some flags, e.g. Solomon Islands, and over-estimated this value for some years, e.g. 1996 and 1999 (Figure A-52). MCMC traces for key parameters did not show any convergence issues (Figure A-53).

3.3 Classification of deep vs. shallow set for WCPO longline effort

A random forest model was used to classify longline effort from the L_BEST dataset into shallow and deep categories to use as a covariate in the extrapolation of the longline catch rates model. The model was based on a subset of L_BEST for which hooks-between-floats information was provided (L_BEST.HBF). Figure 12 shows the proportion of classified effort by CCM over time that was present in L_BEST.HBF *vs.* L_BEST.HBF.

A number of key variables were used in the random forest model to capture spatial and temporal trends by flag as well as covariates that could be indicative of gear configuration such as the catch by species or the species targeting cluster. Of the provided covariates, targeting cluster and longitude (5°) were the most useful in improving classification accuracy, followed by latitude (5°) and year (Figure 13). We did not include flag as a covariate so that we could make predictions for the model for flags that were not present in L_BEST.HBF.

Figure 13 shows the predicted proportion of ‘deep’ effort by flag and year from the model, as well as the resulting uncertainty. Predicted trends tended to reflect those observed for most flags that had provided some HBF classified effort except for flags where the observed proportion varied abruptly between years (e.g. China, Papua New Guinea, Taiwan). As expected, uncertainty bounds were greater for CCMs and time periods where the proportion of observed classified effort was smaller compared to the total effort reported in L_BEST.

3.4 Historical catch reconstructions

Estimated historical catches for the longline bycatch fleet, the longline target fleet and the purse seine fleet split between associated and unassociated sets is shown in Figure 15 to 17. The catches by the longline bycatch fleet are estimated to be much higher than those for the longline target fleet and the purse seine fleets. According to the longline bycatch reconstruction (Figure 15), catches increase steadily from 1995 onwards, peak in 2001 (median: 563352 individuals) and declined steadily since. Catches are highly variable from year to year in the earlier time period. There are three alternative catch reconstructions these predictions can be compared to for part or all of the time span of the analysis. The predictions from

Rice (2012) and Peatman et al. (2018b) were based on the same observer dataset as the one we used. Predicted catches from Rice (2012) are slightly higher at the very start of the time series and lower thereafter (below the 10^{quantile} for the predictions). There is a closer match between the predictions of Peatman et al. (2018b) and ours. We included additional CCMs in our predictions for which longline effort is poorly known but assumed to be quite high, which could account for the difference in our estimates.

We also used an alternative approach to catch reconstructions based on estimates of the global fin trade (see Tremblay-Boyer et al. 2019). While this time series predicts catches for all fleets combined, most catches are predicted to occur in the longline bycatch fleet in our analysis so we compare them directly. The match with the catch reconstruction from trade-based estimated was close at the start of the time series but diverged from 2002 onwards. The match between the predictions was very close in 2000, which was also the year considered the most reliable for a trade-based catch reconstruction of oceanic whitetip sharks in the WCPO.

Reconstructed catches for the longline target fleet were overall estimated to be quite low (<5000 individuals for most years) (Figure 16). The predicted catches were much lower than those estimated by Rice (2012), but noting they used a different measure of effort for this fleet. There was no clear trend in catches except for a general increase from 1995 to 2000 and a decrease and stabilization thereafter. The highest predicted value was in 2010 (median: 9011 individuals) which strongly departed from values in neighbouring years. This high prediction did not come from a predicted high year effect for this year, but from a peak in the 2010 L_BEST effort reported for the fleets used to define shark targeting effort.

Reconstructed catches for the purse fleets were estimated to be higher than those for the longline target fleet, but still much lower than those for the longline bycatch fleet (Figure 17). The associated fleet was predicted to have higher catches than the unassociated fleet, with a steady decline in predicted catches from the pre-2000 period. The catches were predicted to be higher than those estimated by Rice (2012) especially in the earlier part of the time series. The catches were more similar to those estimated by Peatman et al. (2018a) (but still higher especially in 2003).

Predictions of annual catches by fleet are enumerated in Table B-6 to Table B-8 for reference. We also included predictions for the trade-based approach that was detailed in Tremblay-Boyer et al. (2019) (Table B-9).

3.5 CPUE standardization for the longline bycatch fleet

The CPUE standardization model was based on a subset of longline observer programmes having more consistent coverage over space and time (but still noting that longline observer coverage rates were quite low apart except for the US-based programmes). The selected model for CPUE standardization for the longline by-catch fleet included a year effect, a non-continuous linear effect for SST, a non-linear effect for set depth (shallow or deep), a categorical effect for the species targeting effect, and random effects for observer programme

and the interaction between observer programme and year. The negative binomial distribution was parameterized with the addition of the ν parameter. Note that, unlike for the catch reconstruction models, because we did not have to predict from the L_BEST dataset we were able to use observer programme instead of flag as a covariate. We were also able to model set depth as a continuous variable instead of a categorical shallow/deep effect. Both of these changes resulted in an improvement in model performance (see below).

The selected CPUE model had the lowest LOOIC of the model structures tested (Table 5). Based on the results from the related model for the catch reconstruction we only considered SST as an oceanography covariate. Using a continuous effect for HBF instead of category resulted in a slight improvement in model performance. Using an observer programme as a covariate instead of country flag considerably lowered the value for the LOOIC. Note that we did not use both flag and observer programme in the same model since they are confounded. Other covariates that had a strong positive impact on model performance were the species targeting cluster, the random interaction between year and observer programme, and the addition of ν to the parameterization of the negative binomial distribution.

The mean-standardised CPUE index resulting from the best model is shown in Figure 18. The year effects estimated from the model were highly variable at the start of the time series and started steadily declining from 1999 onwards. The lowest year effect estimated was for 2012 which predicted catch rates about 0.15 those of 1995 once other covariates were accounted for. The effect estimated for 2016 slightly higher at about 0.21 the 1995 catch rates. SST had a strong positive effects on catch rates especially up to 27°C (Figure 21). Catch rates declined steadily with increasing hooks-between-floats up to HBF~25. The billfish targeting cluster and the mixed tuna targeting cluster (albacore, bigeye and yellowfin tuna in even proportions) had a positive impact on catch rates compared to the intercept targeting cluster (bigeye tuna cluster).

The stepwise plot showing the change in standardized CPUE as each covariate was added to the model is shown in Figure 19. The biggest change was the addition of the targeting cluster which lowered the relative value of the index for the earlier time period and raised it slightly for the latter time period. However adding observer programme and the year–observer programme interaction somewhat reverted these changes, but resulted in a highly variable early CPUE. This indicates that the values for the targeting cluster and the observer programme covariates are correlated for at least some levels, as would be expected (since fleets tend to focus on targeting a given species). None of the covariates included removed the slight increase in CPUE present in the latter part of the time series. The results from the stepwise plot are also reflected in the influence plot showing the relative impact of the two categorical covariates, targeting cluster and observer programme, on the nominal CPUE (Figure 20). The observer programme covariate had a high negative influence on the nominal CPUE at the start of the time series while the cluster covariate had a moderate positive influence for that same time period.

Key model diagnostics comparing the overall distribution of the observed *vs.* predicted response variable were satisfactory (Figure 22 and 23). There was no noticeable trend in

the observed *vs.* predicted proportion of zeroes by fishing event when predictions were disaggregated between year, observer programme and year \times observer programme (Figure A-54 to Figure A-56). Similarly, there was no noticeable trend in the fit for the number of individuals sharks caught for positive catch events or the 90th quantile of individuals caught for positive catch events (Figure A-57 to A-62). MCMC traces for key parameters did not show any convergence issues (Figure A-63).

4. DISCUSSION

This report presented the historical catch reconstructions and CPUE standardization undertaken to support the 2019 stock assessment for oceanic whitetip shark in the WCPO (Tremblay-Boyer et al. 2019). The catches for the longline bycatch fleet are predicted to be the highest of all fleets. Catches for the other fleets are negligible in comparison. While we do expect most catches to occur in the longline bycatch fleet for this species, this allocation probably also reflects the reliability of catch rates recorded by the observer programs assigned to each fleet, as well as the effort information used to expolate catches. For the longline target fleet, there was sparse observer coverage and also years where oceanic whitetip shark catches appeared not to have been recorded. In addition the effort for that fleet is poorly known. The combination of these factors implies that catches for this fleet were likely under-estimated. In parallel, shark catches were poorly recorded by purse seine observers earlier in the time series and are known to be imprecise even in the more recent time period due to the logistics of sampling and the prioritization of tasks on purse seine sets. It is thus likely that catches by the longline target and purse seine fleets represent a higher proportion of the total WCPO catch for oceanic whitetip shark than what we report here.

Two new modelling approaches were used compared to previous versions of this work. First, all GLMMs models were fitted within a Bayesian framework *via* Markov Chain Monte Carlo algorithms (Bürkner 2017). A clear advantage to this approach was that it makes it straightforward to characterize the uncertainty around predictions, in contrast to previous approaches where uncertainty could not be estimated reliably (e.g. Rice 2012) or had to be estimated with a complicated extra step (Peatman et al. 2018b).

The other model development was to use a random forest model to classify WCPO-wide longline effort into shallow and deep categories. Set depth was an influential covariate in all longline models so this additional modelling step improved the quality of the estimated historical catches. We also estimated the uncertainty of the classification with a Monte Carlo simulation but this was not used further in the analysis beyond the interpretation of the results. Future iterations of this work should consider propagating this uncertainty in the historical catch reconstructions.

Another covariate that was obtained through modelling was the species targeting cluster, which was also an influential covariate in all longline models. We used a straight-forward *k-means* classification approach that had been used in the past for CPUE standardization in

WCPO tuna assessments (McKechnie et al. 2015). This simple algorithm remains a useful tool to capture key trends in how catches by species are distributed amongst sets.

A negative binomial distribution was assumed for all catch reconstructions and CPUE models. This error distribution has clear advantages in the modelling of catch rates given it is discrete, so directly represents the catch event, and can capture both high proportions of zeroes and rare high catch events. However it can be challenging to get good model fits when applied to a dataset where the behaviour of catch events varies greatly across covariates, as was the case here when modelling observer programmes across the WCPO, including variations in catch rates due to habitat, gear configuration, and observer training. The better diagnostics for the CPUE standardization than for the catch reconstructions underscore this, as the observer programmes had been filtered to retain only those having more reliable coverage. The ν parameter we added to the negative binomial parameterization improved model fits in most cases as it allowed the model to fine-tune the shape of the overdispersion to the training dataset. For all catch and CPUE models, the central tendency tended to be well captured but the tail end of catch events was often over-estimated. This informed our decision to use the 90th quantile (and not a higher one) to generate the ‘high’ catch scenario.

We modelled early years even though observer reporting for sharks for this period is known to be unreliable as well as highly variable given the very sparse coverage. We used the predictions as is¹ but would urge care in the interpretation and use of these predictions as they are likely under-estimates. For future iterations it might be good to decide upon a reliable start year (e.g. 2003, similar to Peatman et al. 2018b), and either change the time-span of the assessment accordingly or scale the catches before that year based on a set of scenario rules. Alternatively, we did not use informative priors in either the catch reconstructions or the CPUE standardization here, but we could have specified more precise priors for the early time period. However this approach would not be appropriate for the CPUE standardization in the absence of independent information (e.g. fishery-independent survey) about relative abundance. In that instance it might be useful to consider whether 1995 is in fact a suitable start year for the assessment.

Historical catch estimates for the purse seine fleets must also be interpreted carefully. There was no consistent instructions for sharks to be recorded for the earlier part of the time series. In parallel, sharks on purse seine sets can be challenging to identify and enumerate during the brailing and sorting process, even in the recent time period. Observers on purse seine vessels are responsible for a number of competing tasks and the recording of shark catches might not be prioritized. It thus seems likely that observer training and directives would strongly impact catch rates for sharks recorded by observers on purse seine fleets. For instance, the highest flag effect estimated for the purse seine model was for the US flag, and given key habitat and gear variables were accounted for in the model, this higher effect was probably due to observer training and not a real effect of this fleet on catch rates. Similarly, using observer programme instead of flag as a covariate in the CPUE model greatly improved model fit and performance

¹except for purse seine fleets in 1995 where the estimated value of almost zero catch deemed especially unreliable was set to the higher 1996 estimate

for the longline datasets, whereas one would *a priori* think that flag-specific gear configurations leading to changes in catch rates should be more informative. These examples underscore that observer training is likely to have a high impact on the recorded catch rates, especially in the earlier time period. For this reason, observer directives in the recording of shark catches should be as consistent as possible across programs, and the results presented should be interpreted here in light of this important caveat.

5. REFERENCES

- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. doi:10.18637/jss.v080.i01
- Carpenter, B.; Gelman, A.; Hoffman, M.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M. A.; Guo, J.; Li, P., & Riddell, A. (2015). Stan: A probabilistic programming language. *Journal of Statistical Software*. Retrieved May 18, 2016, from http://www.demonish.com/cracker/1431548798_9226234ebe/stan-resubmit-jss1293.pdf
- Clarke, S. (2018). *Historical Catch Estimate Reconstruction for the Pacific Ocean based on Shark Fin Trade Data (1980-2016)*. WCPFC-SC14-2018/SA-IP-09. Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Fourteenth Regular Session, 8–16 August 2018, Busan, Korea.
- Lawson, T. (2011). *Estimation of catch rates and catches of key shark species in tuna fisheries of the western and central pacific ocean using observer data*. WCPFC-SC12-2011/EB-IP-02. Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Seventh Regular Session, 9–17 August, Busan, Korea.
- Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2/3, 18–22.
- McKechnie, S.; Tremblay-Boyer, L., & Harley, S. (2015). *Longline cpue indices for bigeye tuna based on the analysis of pacific-wide operational catch and effort data*. WCPFC-SC11-2015/SA-WP-02. Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Eleventh Regular Session, 5–13 August 2015, Pohnpei, Federated States of Micronesia.
- Peatman, T.; Allain, V.; Caillot, S.; Park, T.; Williams, P.; Tuiloma, I.; Panizza, A.; Fukofuka, S., & Smith, N. (2018a). *Summary of purse seine fishery bycatch at a regional scale, 2003–2017*. WCPFC-SC14-2018/ST-IP-04 Rev 1 (24 July 2018). Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Fourteenth Regular Session, 8–16 August 2018, Busan, Korea.
- Peatman, T.; Bell, L.; Allain, V.; Caillot, S.; Williams, P.; Tuiloma, I.; Panizza, A.; Tremblay-Boyer, L.; Fukofuka, S., & Smith, N. (2018b). *Summary of longline bycatch at a regional scale, 2003–2017*. WCPFC-SC14-2018/ST-WP-03 Rev 3 (15 April 2019). Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Fourteenth Regular Session, 8–16 August 2018, Busan, Korea.
- Pilling, G. & Brouwer, S. (2019). *Report from the SPC Pre-Assessment Workshop, Nouméa, April 2019*. WCPFC-SC15/SA-IP-01. Report to the Western and Central Pacific Fisheries

- Commission Scientific Committee. Fifteenth Regular Session, 12–0 August 2019, Pohnpei, Federated States of Micronesia.
- Rice, J. (2012). *Alternate catch estimates for silky and oceanic whitetip sharks in Western and Central Pacific Ocean*, WCPFC-SC8/SA-IP-12. Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Eighth Regular Session, 7–15 August 2012, Busan, Korea.
- Rice, J. & Harley, S. (2012). *Stock assessment of oceanic whitetip sharks in the western and central Pacific Ocean*. WCPFC-SC8-2012/SA-WP-06 Rev 1. Report to the Western and Central Pacific Fisheries Commission Scientific Committee. Eighth Regular Session, 7–15 August 2012, Busan, Korea.
- Tremblay-Boyer, L.; Carvalho, F.; Neubauer, P., & Pilling, G. (2019). *Stock assessment for oceanic whitetip shark in the western and central pacific ocean*. WCPFC-SC15/SA-WP-06.
- Vehtari, A.; Gelman, A., & Gabry, J. (2016). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. R package version 0.1.6. Retrieved from <https://github.com/jgabry/loo>
- White, W.; Gisawa, L.; Baje, L.; Usu, T.; Yaman, L.; Sabub, B.; Appleyard, S.; Green, M.; Vieira, S.; Chin, A.; Smart, J.; Grant, M., & Simpfendorfer, C. (2018). *Sustainable management of the shark resources of Papua New Guinea: socioeconomic and biological characteristics of the fishery*. FR2018/20, Australia.

6. TABLES

Table 1: Model covariates of operational fishing features likely to influence catch rates of oceanic whitetip shark and environmental variables that may represent habitat of this species (LBEST and SBEST are databases of the SPC for longline and purse-seine fisheries, respectively).

Covariate	Description
Year	Year when the fishing set occurred, treated as categorical .
Flag	Country-assignment for the vessel performing the fishing set.
Programme	Country observer programme for fisheries observer observing the fishing sets.
HBF	Hooks-between-floats for the longline fishing set.
SetType	Set category for the purse-seine fishing set: anchored FADs (fishing aggregation devices), drifting FADS, whales, logs or floating objects, baited and unbaited free schools.
HBF.cat	Hooks-between-floats of the fishing set assigned to a categorical variable: shallow for ≤ 10 HBF, deep for > 10 HBF.
Cluster	Predicted targeting strategy for longline fishing set based on k-means clustering of the proportion in the total catch in number of albacore, bigeye, yellowfin and bluefin tuna, swordfish and other billfish. Cluster composition was predicted based on LBEST records and assuming 5 centres, resulting in a categorical variable with values from 1 to 5. Longline observed set targeting strategy was predicted according to the LBEST classification.
SST	Sea surface temperature aggregated at 5-degree scale for LBEST and 1-degree scale for SBEST, obtained from NOAA (https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html).
Chl- <i>a</i>	Sea surface chlorophyll- <i>a</i> concentration aggregated at 5-degree scale for LBEST and 1-degree scale for SBEST (https://coastwatch.pfeg.noaa.gov/erddap/griddap/erdMH1chlmday).
Bathymetry	Depth of the sea floor at the location where the fishing set occurred, aggregated at 5-degree scale for LBEST and 1-degree scale for SBEST (https://coastwatch.pfeg.noaa.gov/erddap/griddap/srtm15plus).
Dist2Coast	Distance of the set to the nearest coastline, aggregated at 5-degree scale for LBEST and 1-degree scale for SBEST.

Table 2: Values and effective number of parameters for the leave-one-out information criterion (LOOIC) for key model structures used to estimate oceanic whitetip shark catch rates from observed trips for the longline bycatch fleet. Models were ordered by LOOIC value.

Model	LOOIC	K	Δ
Year + s(SST, k=3) + s(chla, k = 3) + HBF + cluster + 1 Flag + 1 Year:Flag; + ν	27920.9	234.2	-11.3
Year + s(SST, k=3) + HBF + cluster + 1 Flag + 1 Year:Flag; + ν	27932.2	232.1	0.0
Year + s(SST, k=3) + HBF + cluster + 1 Flag + 1 Year:Flag	28570.7	236.3	638.5
Year + s(SST, k=3) + s(chla, k = 3) + HBF + cluster + 1 Flag; + ν	28692.8	77.0	760.6
Year + s(SST, k=3) + HBF + cluster + 1 Flag; + ν	28707.7	73.7	775.5
Year + s(SST, k=3) + cluster + 1 Flag; + ν	29026.5	71.4	1094.3
Year + s(SST, k=3) + HBF + cluster + 1 Flag	29440.6	90.7	1508.4
Year + HBF + cluster + 1 Flag; + ν	30256.7	67.7	2324.5

Table 3: Values and effective number of parameters for the leave-one-out information criterion (LOOIC) for key model structures used to estimate oceanic whitetip shark catch rates from observed trips for the longline target fleet. Models were ordered by LOOIC value.

Model	LOOIC	K	Δ
Year + s(SST, k=3) + HBF + cluster + 1 Flag; + ν	6368.20	54.00	0.00
Year + HBF + cluster + 1 Flag; + ν	6371.80	51.00	3.60
Year + s(SST, k=3) + HBF + 1 Flag; + ν	6385.10	48.40	16.90
Year + s(SST, k=3) + HBF + cluster + 1 Flag	6524.10	66.50	155.90
Year + s(SST, k=3) + cluster + 1 Flag; + ν	6554.60	50.30	186.40

Table 4: Values and effective number of parameters for the leave-one-out information criterion (LOOIC) for key model structures used to estimate oceanic whitetip shark catch rates from observed trips for the purse seine fleet. Models were ordered by LOOIC value.

Model	LOOIC	K	Δ
Year + s(dist2coast,k=3) + SetType + 1 Flag + 1 Year:Flag; $+\nu$	21237.0	236.0	-325.7
Year + s(chla,k=3) + SetType + 1 Flag + 1 Year:Flag	21545.2	267.1	-17.5
Year + s(SST,k=3) + SetType + 1 Flag + 1 Year:Flag	21546.8	270.3	-15.9
Year + s(dist2coast,k=3) + SetType + 1 Flag + 1 Year:Flag	21562.7	265.1	0.0
Year + s(Bathy,k=3) + SetType + 1 Flag + 1 Year:Flag	21564.0	270.5	1.3
Year + SetType + 1 Flag + 1 Year:Flag	21576.9	267.8	14.2
Year + s(chla,k=3) + SetType + 1 Flag	22304.7	134.3	742.0
Year + s(Bathy,k=3) + SetType + 1 Flag	22326.4	136.7	763.7
Year + s(dist2coast,k=3) + SetType + 1 Flag	22334.3	136.9	771.6
Year + SetType + 1 Flag	22342.6	136.3	779.9

Table 5: Values and effective number of parameters for the leave-one-out information criterion (LOOIC) for key model structures used to standardize CPUE for oceanic whitetip shark catch rates from key observed longline fleets. Models were ordered by LOOIC value.

Model	LOOIC	K	Δ
Year + s(SST, k=3) + (1 Program) + s(HBFnum, k=4) + cluster + (1 Year:Program); + ν	38686.0	128.9	0.0
Year + s(SST, k=3) + 1 Flag + s(HBFnum, k=4) + cluster + 1 Year:Flag; + ν	39188.7	158.2	502.7
Year + s(SST, k=3) + (1 Program) + s(HBFnum, k=4) + cluster + (1 Year:Program)	39231.8	155.7	545.8
Year + s(SST, k=3) + (1 Program) + s(HBFnum, k=4) + cluster; + ν	39593.1	51.7	907.1
Year + s(SST, k=3) + (1 Program) + s(HBFnum, k=4) + cluster	40239.8	64.2	1553.8
Year + s(SST, k=3) + 1 Flag + s(HBFnum, k=4) + cluster	40455.9	74.4	1769.9
Year + s(SST, k=3) + s(HBFnum, k=4) + cluster	41369.2	50.9	2683.2
Year + s(SST, k=3) + HBF + cluster	41403.5	48.0	2717.5

7. FIGURES

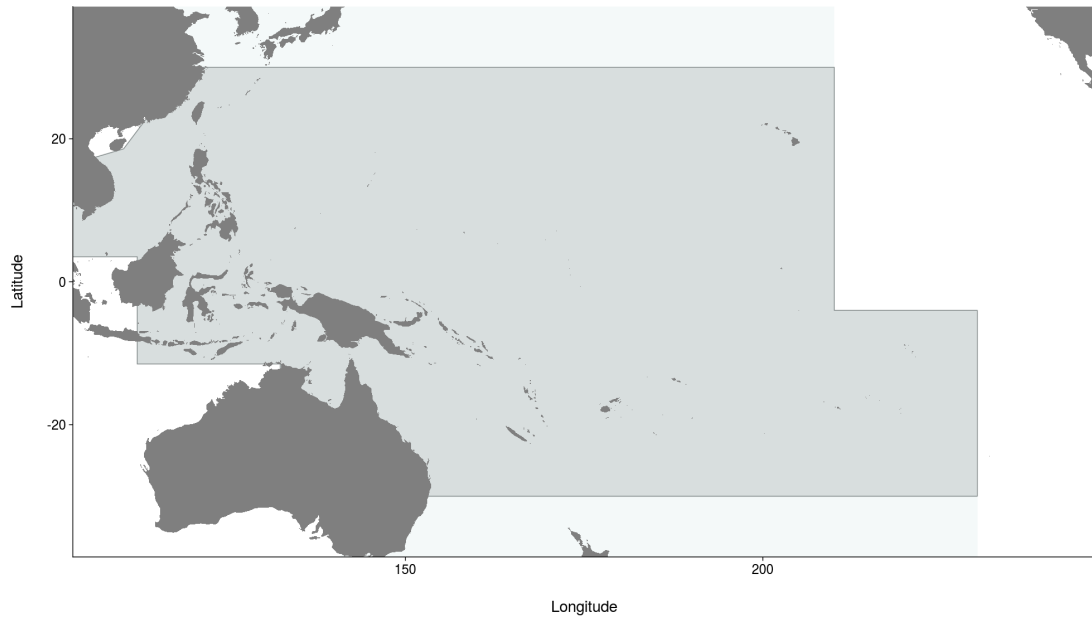


Figure 1: Western and Central Pacific Fisheries Commission convention area (light grey), including the stock assessment area for oceanic whitetip shark (dark grey), bounded by the 30°N and 30°S parallels

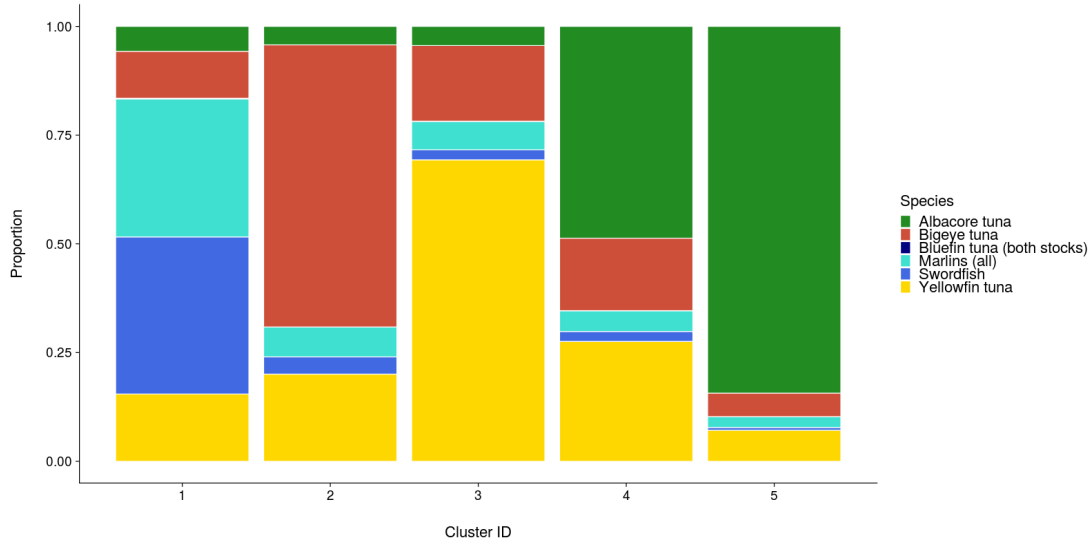


Figure 2: Average species composition for the 5 species targeting cluster identified from the L_BEST dataset aggregated over the 1995–2016 period.

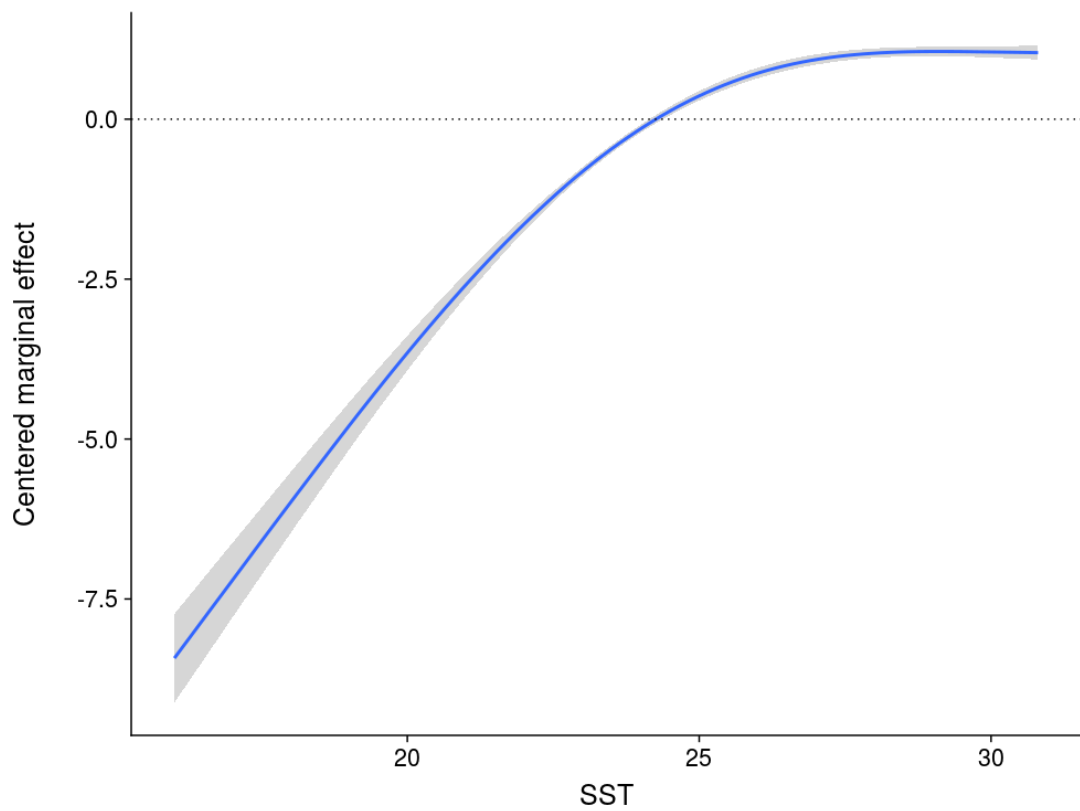


Figure 3: Predicted relationship between sea surface temperature (SST) and observed catch rates for the oceanic whitetip shark for the longline bycatch fleet, independent of other covariates.

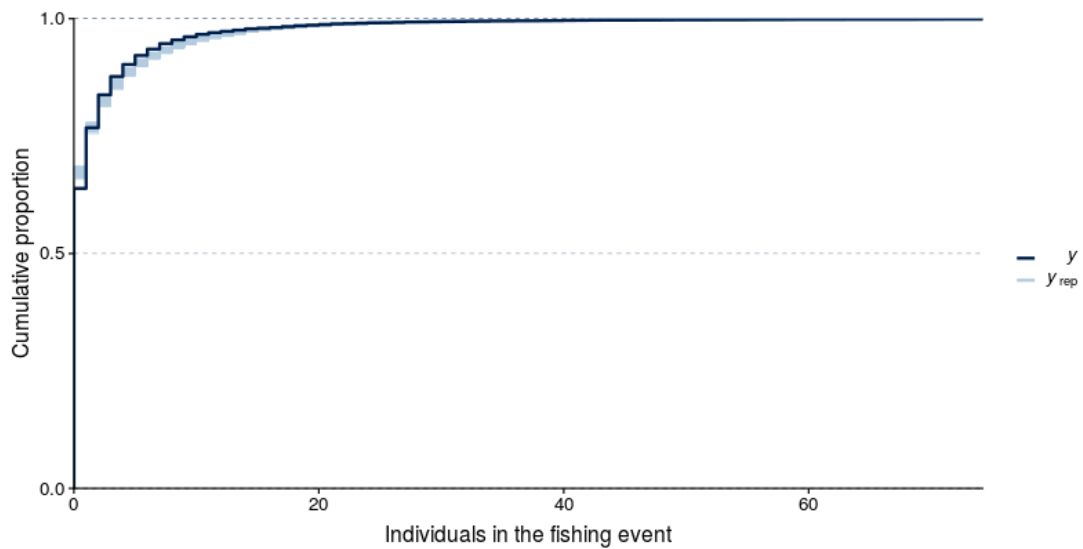


Figure 4: Posterior predictive check plot for the longline bycatch model: cumulative distribution function of the observations and draws from the model posterior showing whether the expected distribution of values for the response variable match between observations and predictions.

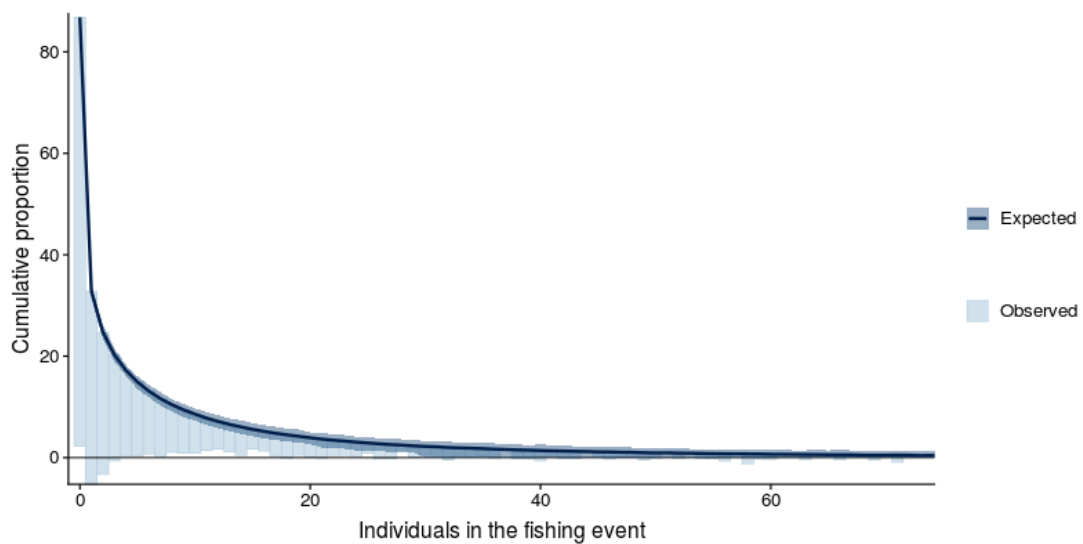


Figure 5: Rootogram diagnostic assessing the fit of a count regression model, applied to the observer catch rates model for the longline bycatch fleet. The y-axis shows the square-root rescaled expected and observed counts for ease of comparison. The black line shows the expected counts and the blue bars shows the observed counts. Observed count bar reaching below the x-axis indicate that the model under-represents the contribution of this bin in the dataset; observed count bars not reaching the x-axis indicate that the model over-represents the contribution of this bin in the dataset.

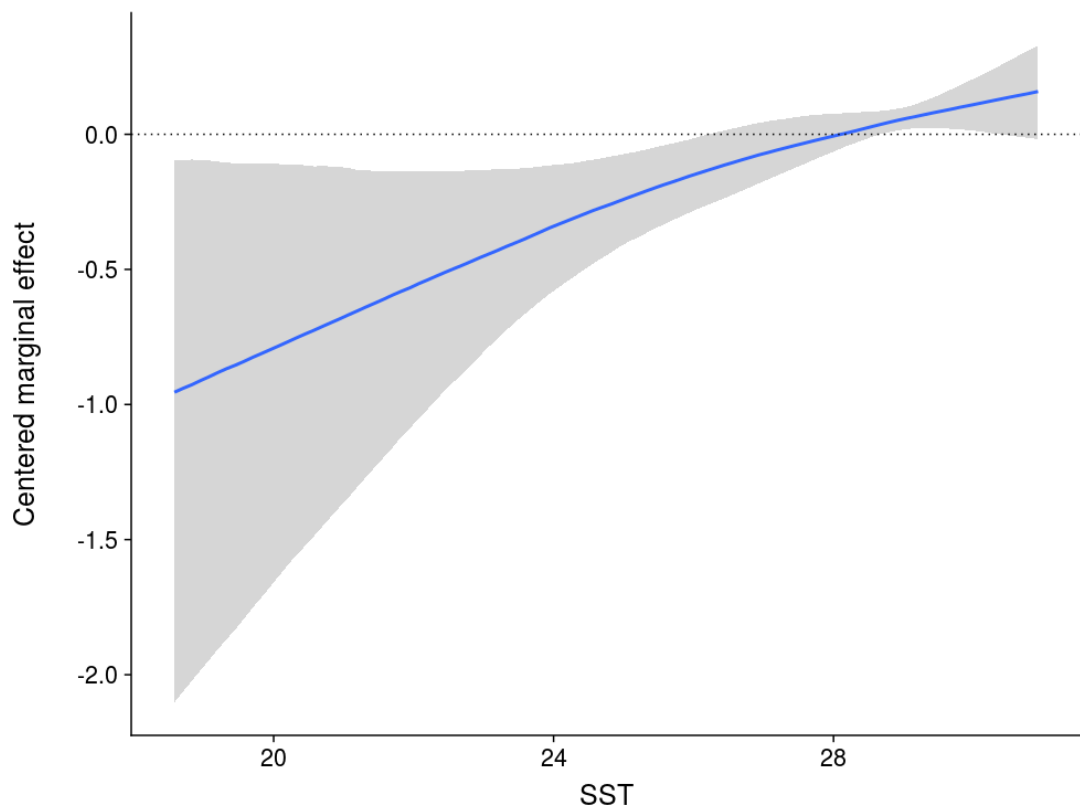


Figure 6: Predicted relationship between sea surface temperature (SST) and observed catch rates for the oceanic whitetip shark for the longline target fleet, independent of other covariates.

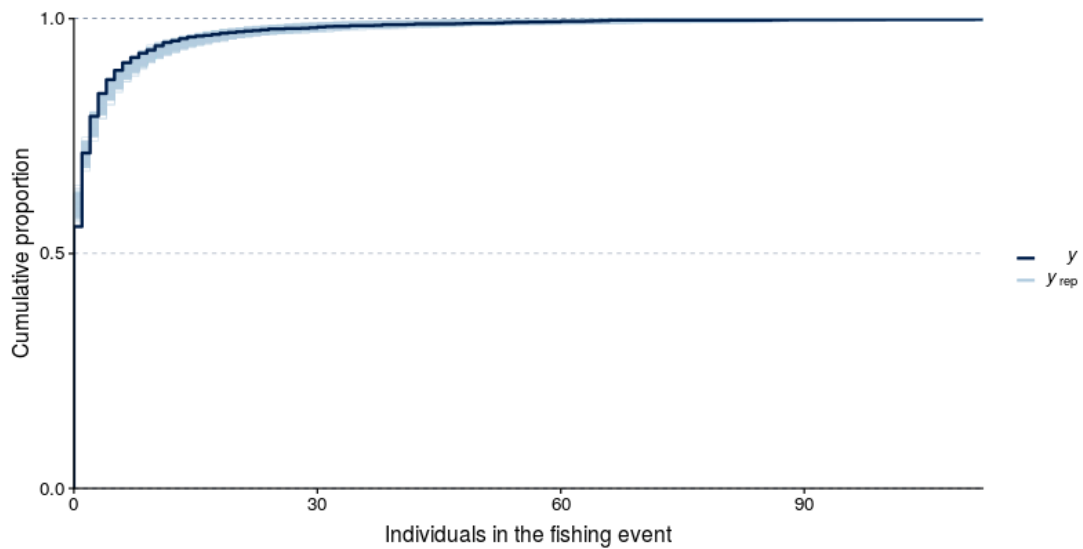


Figure 7: Posterior predictive check plot for the longline target fleet model: cumulative distribution function of the observations and draws from the model posterior showing whether the expected distribution of values for the response variable match between observations and predictions.

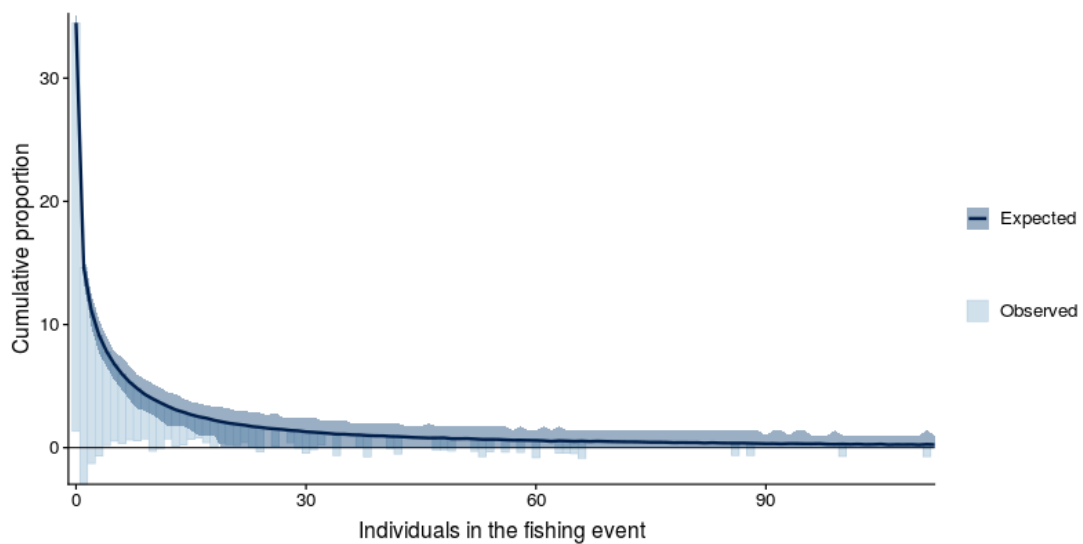


Figure 8: Rootogram diagnostic assessing the fit of a count regression model, applied to the observer catch rates model for the longline target fleet. The y-axis shows the square-root rescaled expected and observed counts for ease of comparison. The black line shows the expected counts and the blue bars shows the observed counts. Observed count bar reaching below the x-axis indicate that the model under-represents the contribution of this bin in the dataset; observed count bars not reaching the x-axis indicate that the model over-represents the contribution of this bin in the dataset.

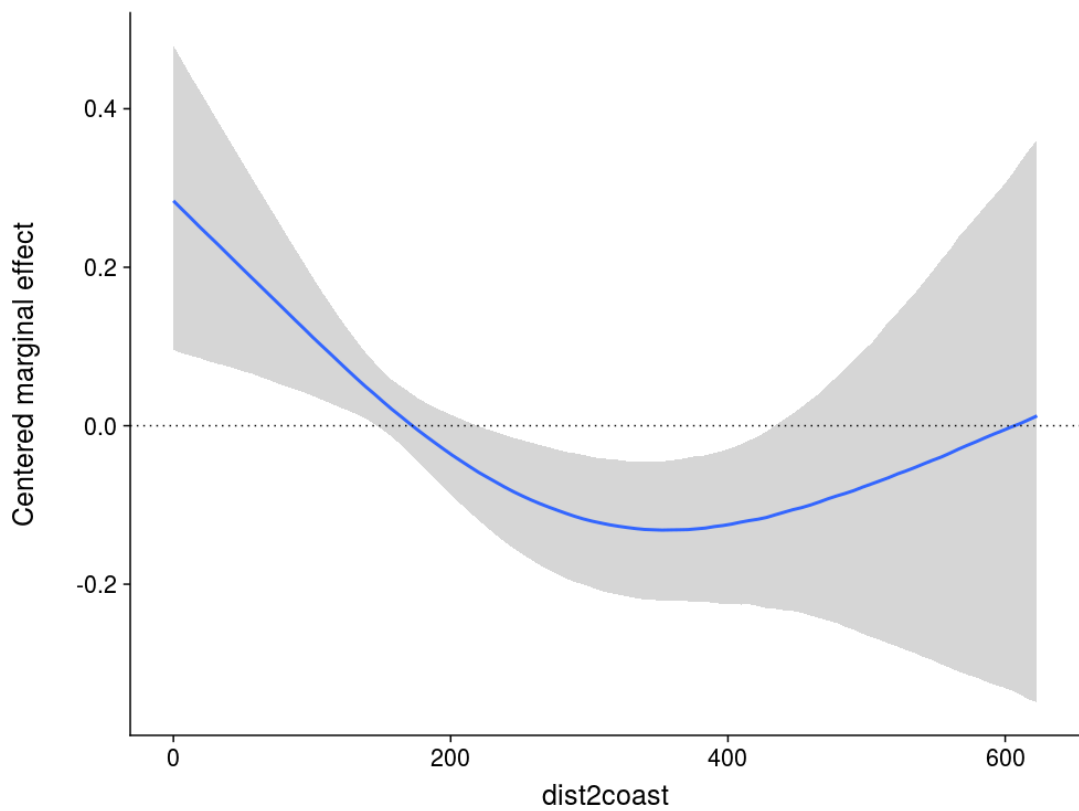


Figure 9: Predicted relationship between distance to the coastlines and observed catch rates for the oceanic whitetip shark in the model for purse seine fleets, independent of other covariates.

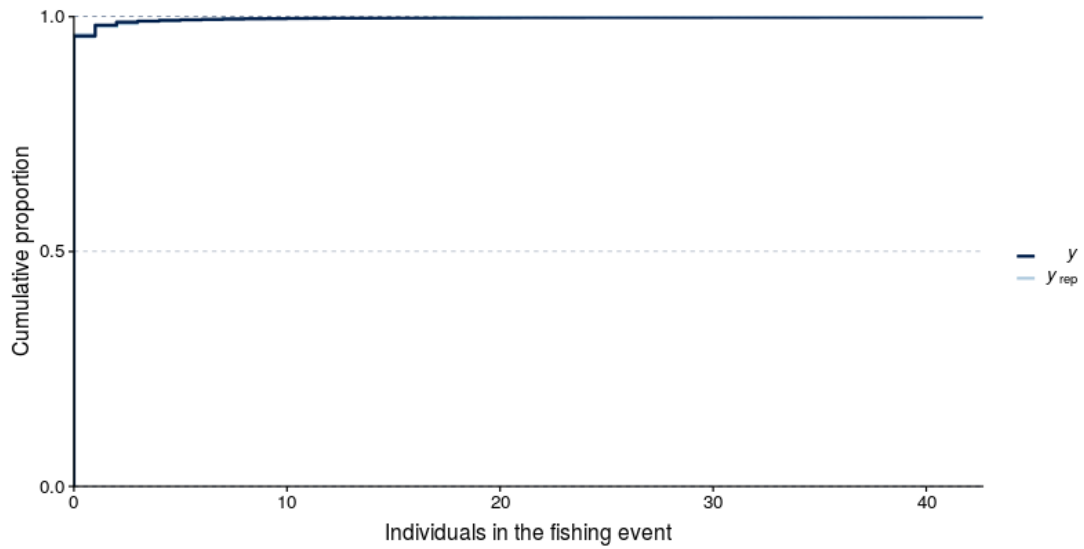


Figure 10: Posterior predictive check plot for the purse seine fleets model: cumulative distribution function of the observations and draws from the model posterior showing whether the expected distribution of values for the response variable match between observations and predictions.

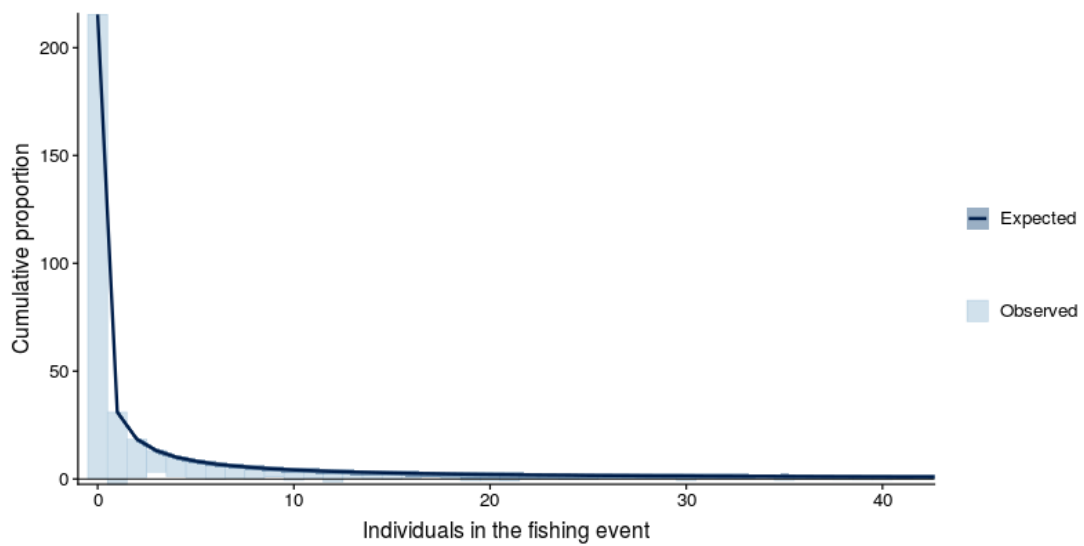


Figure 11: Rootogram diagnostic assessing the fit of a count regression model, applied to the observer catch rates model for the purse seine fleets. The y-axis shows the square-root rescaled expected and observed counts for ease of comparison. The black line shows the expected counts and the blue bars shows the observed counts. Observed count bar reaching below the x-axis indicate that the model under-represents the contribution of this bin in the dataset; observed count bars not reaching the x-axis indicate that the model over-represents the contribution of this bin in the dataset.



Figure 12: Proportion of LBEST effort provided with HBF information for key longline countries in LBEST active over the assessment's region.

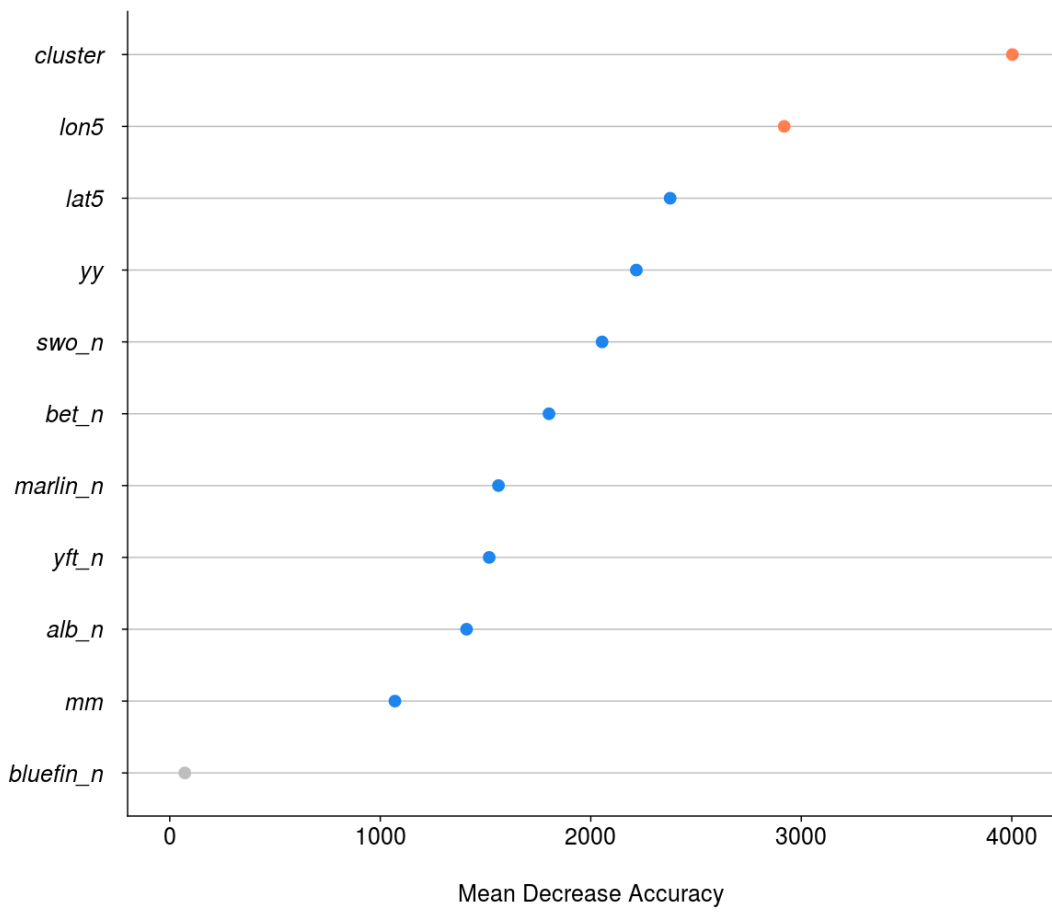


Figure 13: Importance of covariates for classification for the random forest model based on the mean decrease in the classification accuracy when the variable is excluded, ordered in decreasing abundance from top to bottom. Values highlighted in red or grey had especially high or low importance, respectively, in comparison to other covariates (blue).

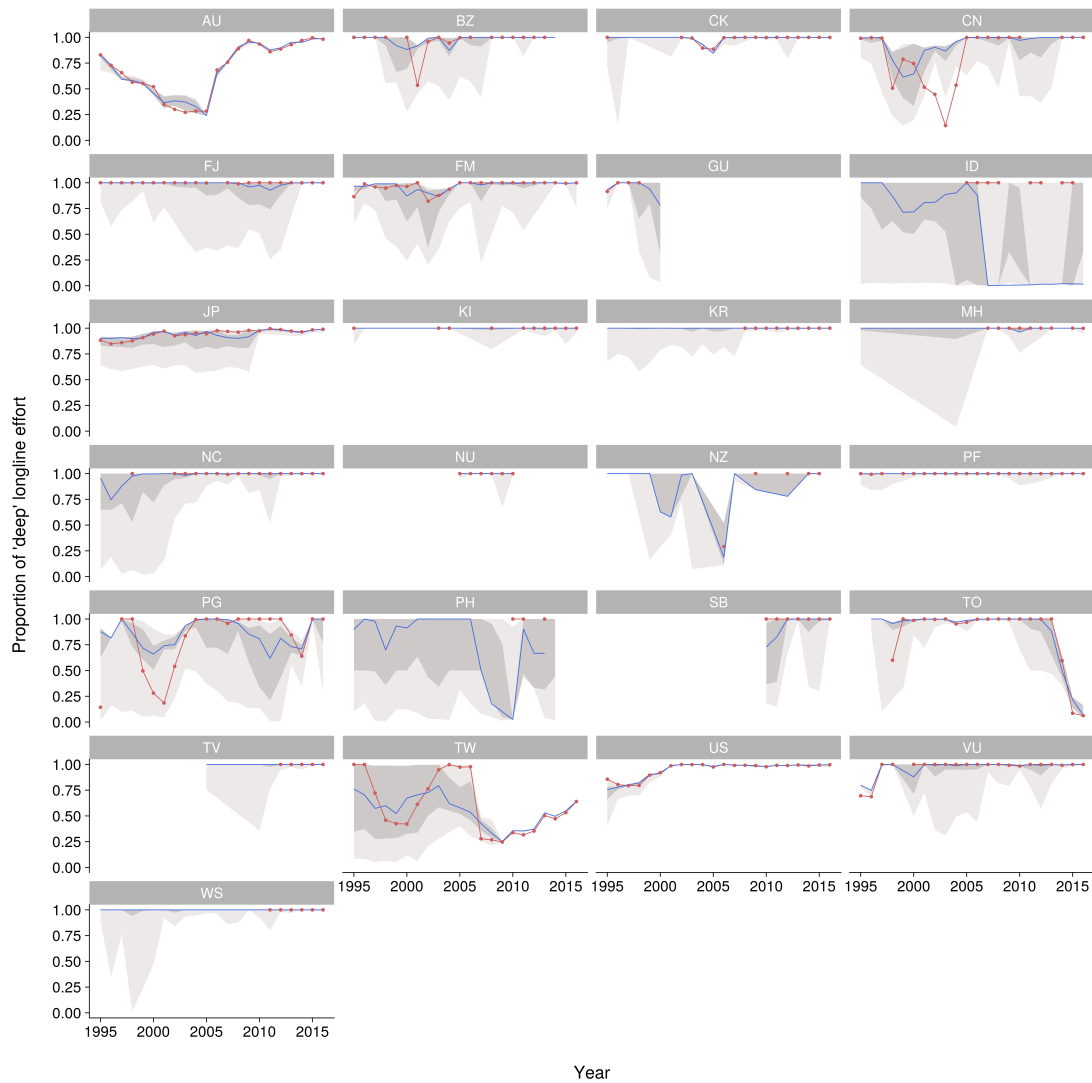


Figure 14: Prediction of the proportion of 'deep' LBEST effort over time for key longline countries in LBEST active over the assessment's region. The red line shows the observed proportion of 'deep' effort when provided and the blue line shows the prediction accounting for both observations and the random forest prediction for unobserved strata. The light and dark grey bounds show the 0.025th - 0.975th and 0.25th - 0.75th uncertainty bounds for the effort classification, also accounting for effort already provided with HBF resolution.

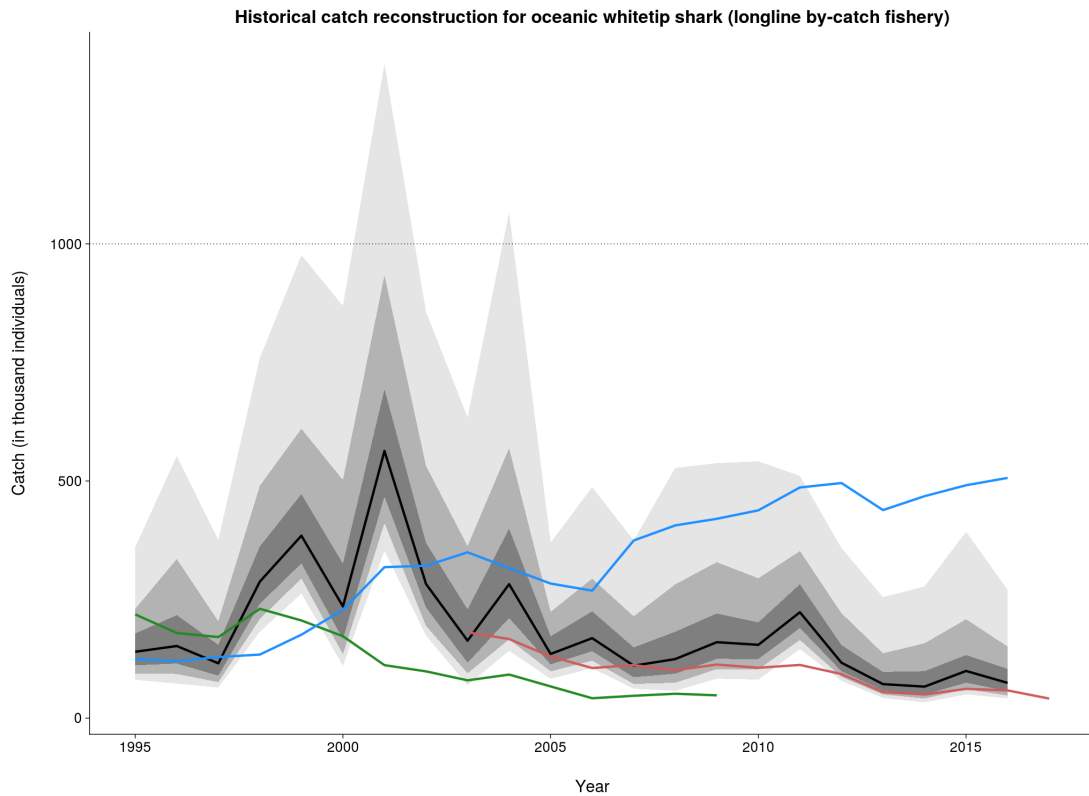


Figure 15: Median predictions of oceanic whitetip shark catch in the WCPO for the longline by-catch fleet based on a model of longline observed catch rates applied to LBEST effort. The light, dark and darker grey bounds show the 0.025th - 0.975th, 0.10th - 0.90th and 0.25th - 0.75th uncertainty bounds. For comparison with our estimates, the blue line shows the median prediction of historical catch based on global fin trade statistics, the red line shows the prediction of historical catch published in (Peatman et al. 2018b), and the green line shows the historical catches used for this fleet in the reference case for the 2012 assessment.

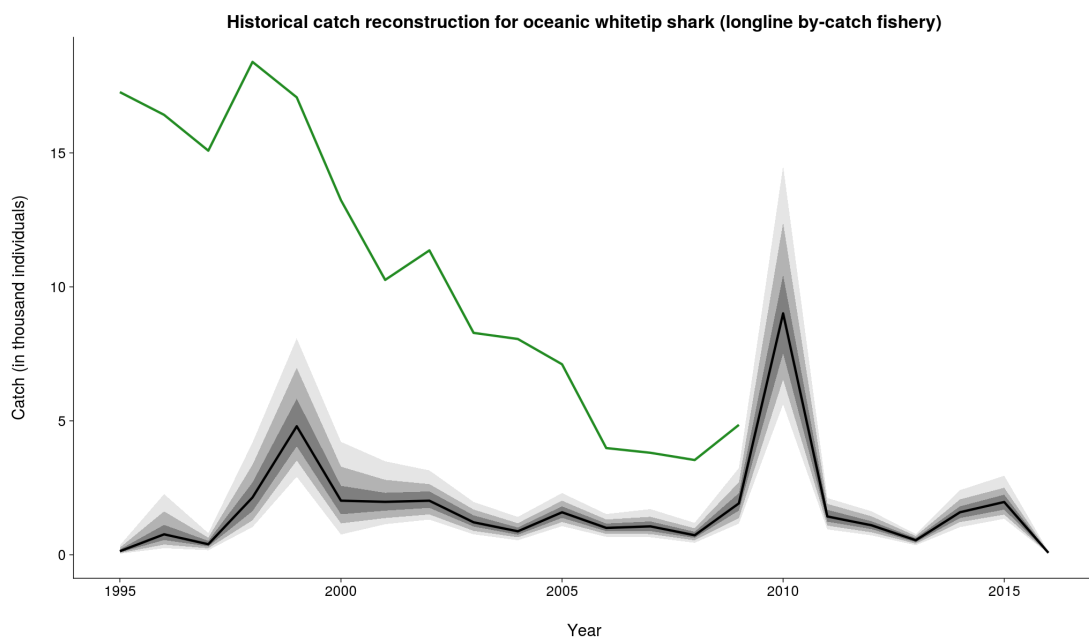


Figure 16: Median predictions of oceanic whitetip shark catch in the WCPO for the longline target fleet based on a model of longline observed catch rates applied to LBEST effort. The light, dark and darker grey bounds show the 0.025th - 0.975th, 0.10th - 0.90th and 0.25th - 0.75th uncertainty bounds. For comparison with our estimates, the green line shows the historical catches used for this fleet in the reference case for the 2012 assessment.

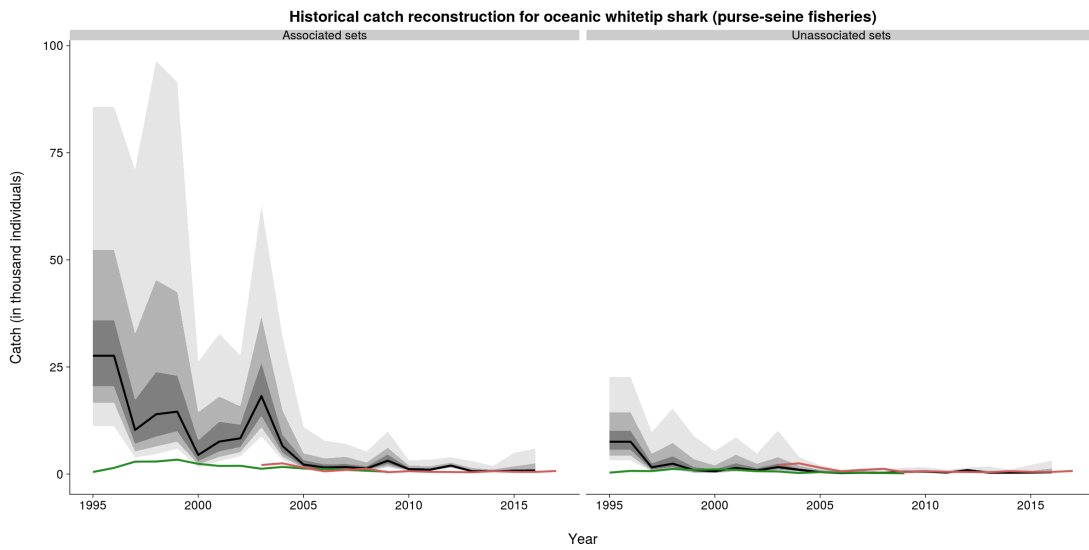


Figure 17: Median predictions of oceanic whitetip shark catch in the WCPO for the associated and unassociated purse-seine fleets based on a model of purse-seine observed catch rates applied to SBEST effort. The light, dark and darker grey bounds show the 0.025th-0.975th, 0.10th-0.90th and 0.25th-0.75th uncertainty bounds. For comparison with our estimates, the red line shows the prediction of historical catch published in (Peatman et al. 2018b) for these fleets and the green line shows the corresponding historical catches used in the reference case for the 2012 assessment.

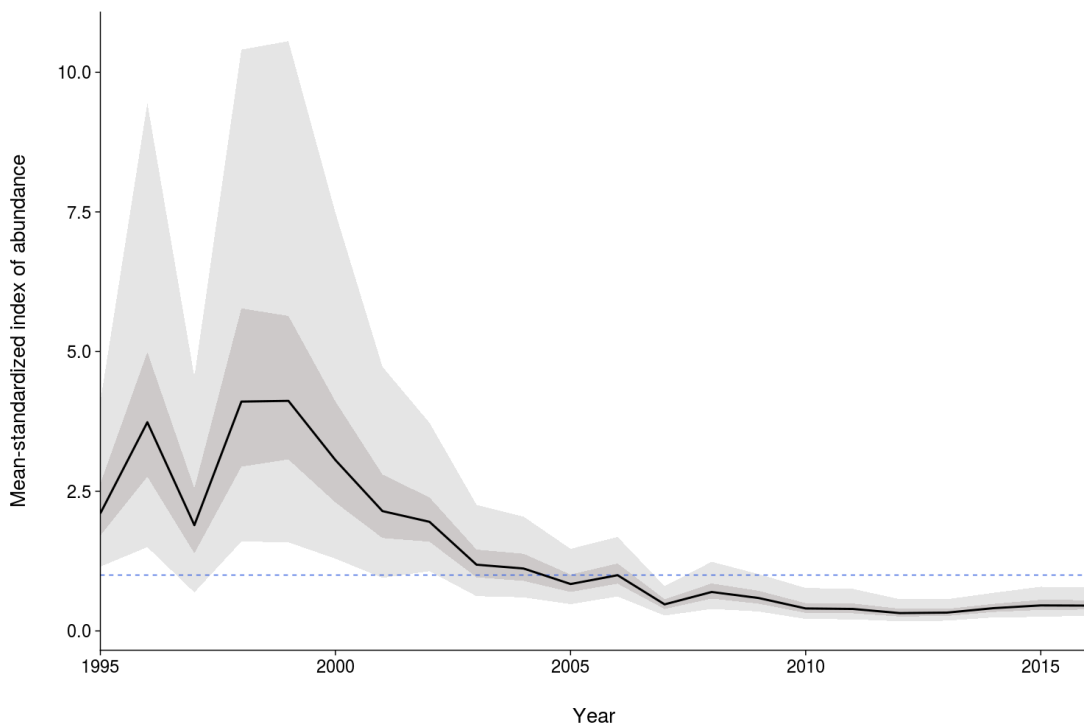


Figure 18: Mean-standardized CPUE index used in this assessment of oceanic whitetip shark. The light and dark grey bounds show the 0.025th-0.975th and 0.25th-0.75th uncertainty bounds around the year effects. The dashed line at 1 shows the reference mean value.

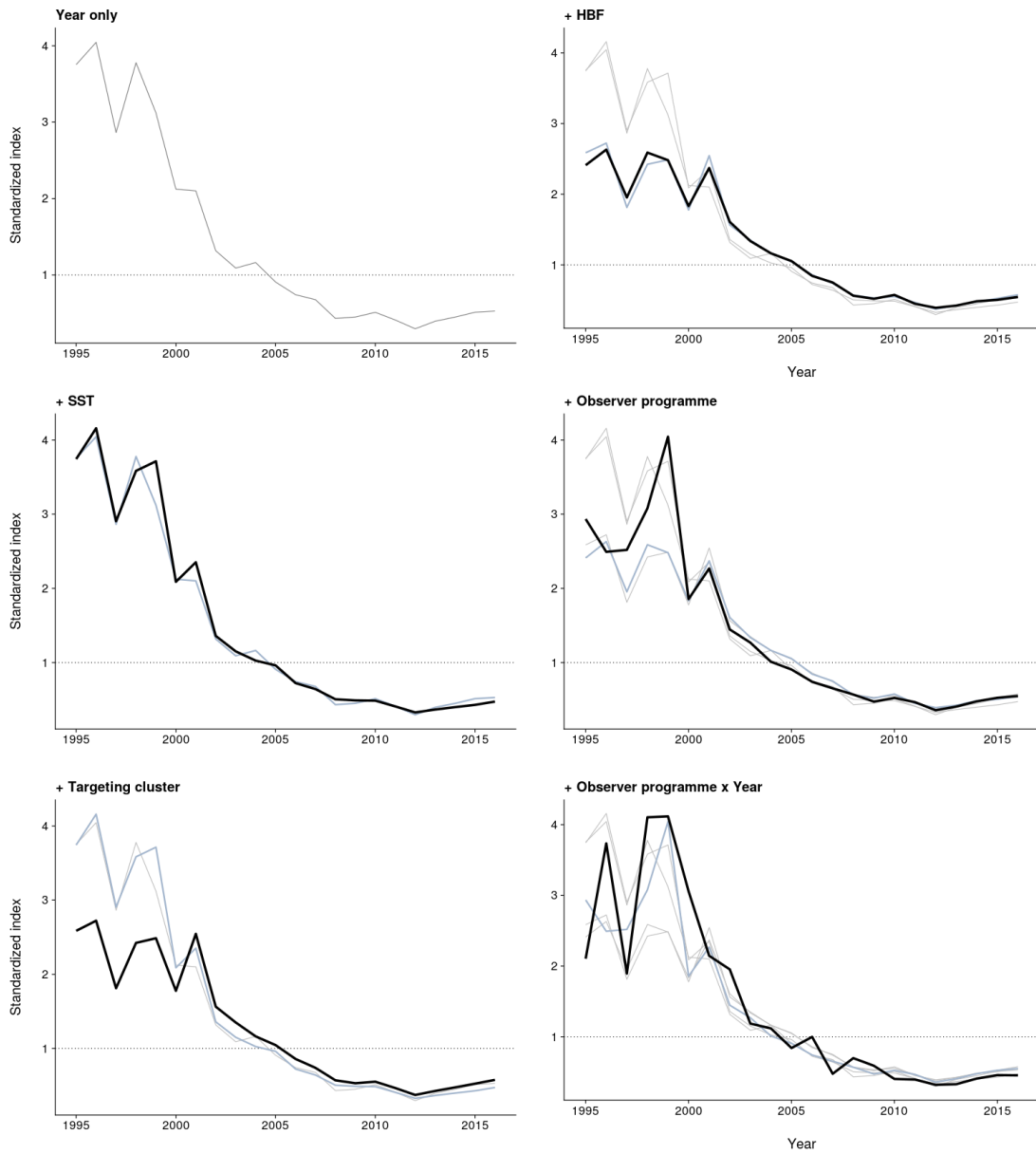


Figure 19: Stepwise plot for the CPUE standardization showing the impact on mean-centered standardized indices of gradually adding model covariates one at a time. New covariates are added from top to bottom, the updated CPUE index is shown in black and the index from the previous step is in blue. The dotted line at 1 shows the mean value for the index.

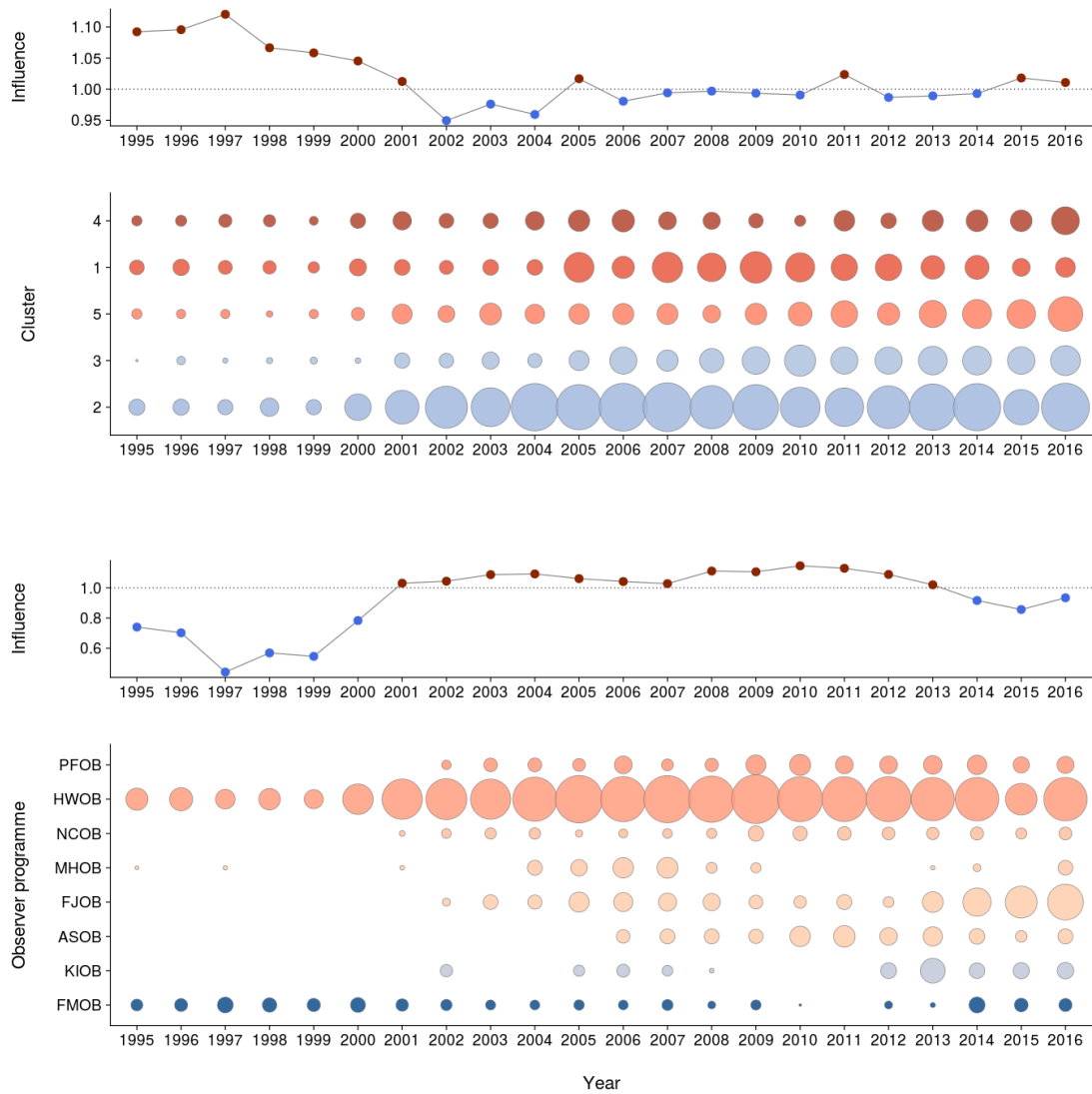


Figure 20: Influence plot for the two categorical covariates used in the CPUE standardization: observer programme (top) and targeting cluster (bottom). The influence (top panel) shows the relative impact on the standardized CPUE of covariates over time. A positive influence means that the nominal CPUE is higher because of that covariate during that time period, so that adding it to the standardization model results in a lower index overall. The dotted line at 1 shows the baseline influence. The bottom panel shows the relative number of fishing events for each covariate level across time, with the size of the circle scaling with n . Levels are placed in ascending order of the value estimated for their effect, from bottom to top. Points in reds are positive effects compared to the intercept, points in blue are negative effects compared to the intercept.

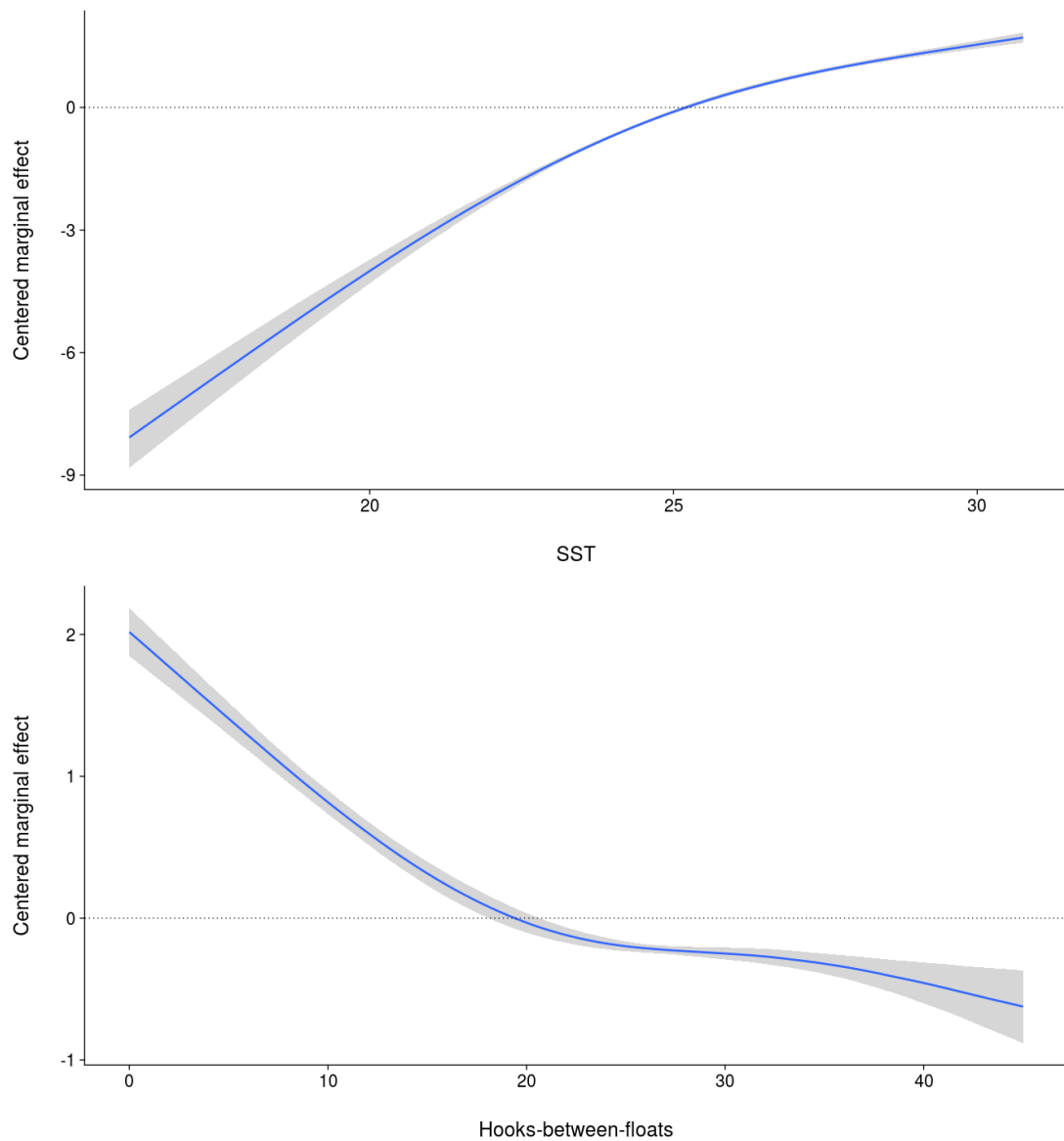


Figure 21: Predicted relationship for the observed catch rates for the oceanic whitetip shark vs. SST (top) and hooks - between - floats (bottom) in the model for the CPUE standardization based on the longline bycatch fleet, independent of other covariates.

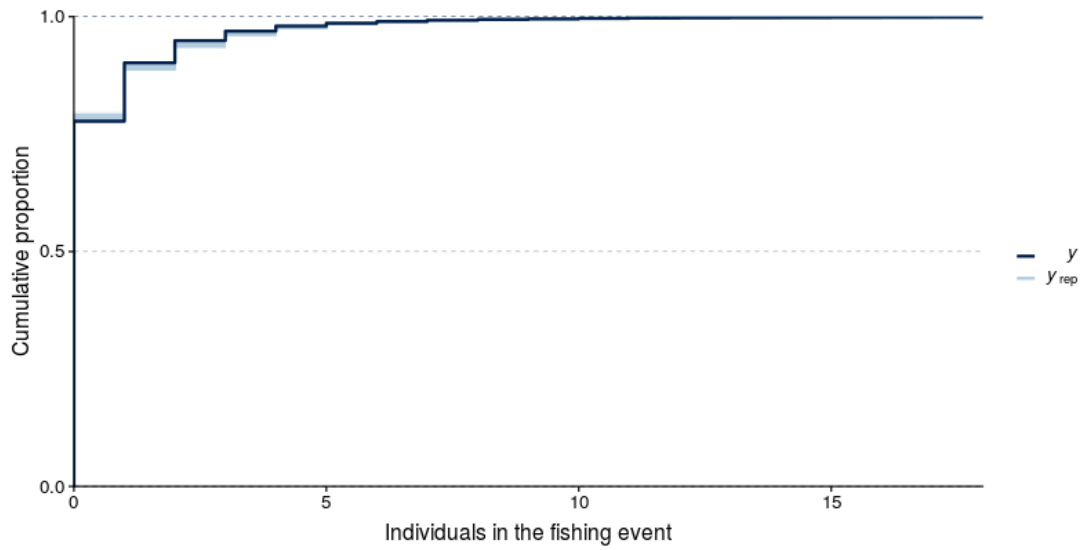


Figure 22: Posterior predictive check plot for the CPUE standardization model: cumulative distribution function of the observations and draws from the model posterior showing whether the expected distribution of values for the response variable match between observations and predictions.

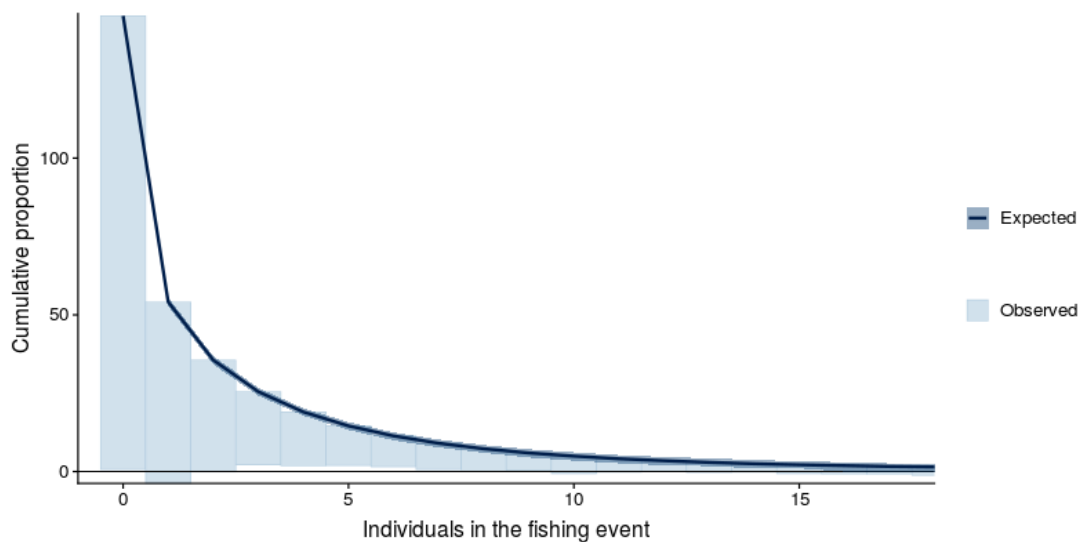


Figure 23: Rootogram diagnostic assessing the fit of a count regression model, applied to the CPUE standardization model. The y-axis shows the square-root rescaled expected and observed counts for ease of comparison. The black line shows the expected counts and the blue bars shows the observed counts. Observed count bar reaching below the x-axis indicate that the model under-represents the contribution of this bin in the dataset; observed count bars not reaching the x-axis indicate that the model over-represents the contribution of this bin in the dataset.

APPENDIX A: GLMMs diagnostics

A.1 Diagnostics for the observer catch rates model for the longline bycatch fleet

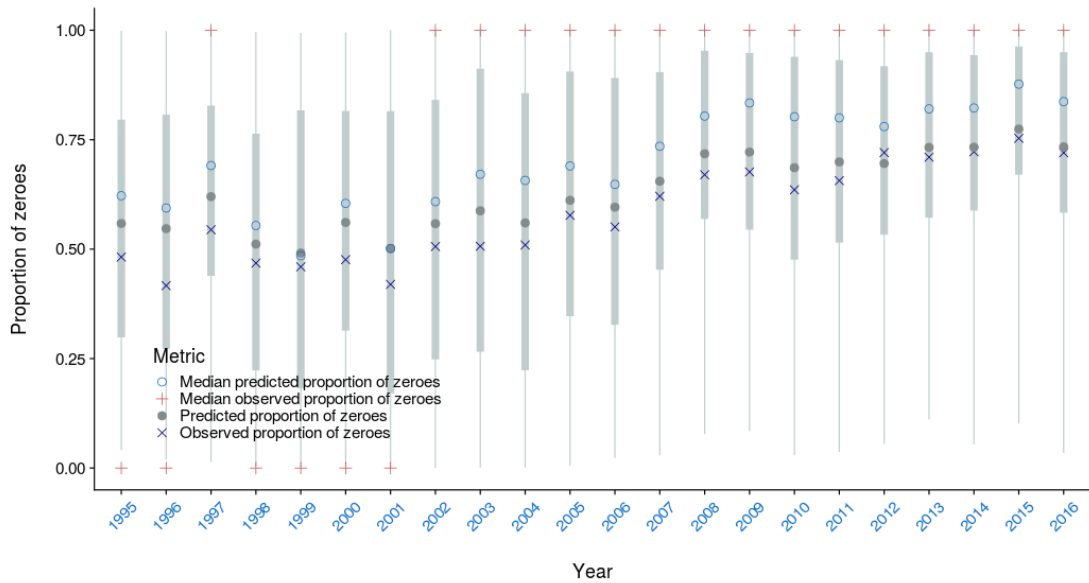


Figure A-24: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline bycatch fleet by year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

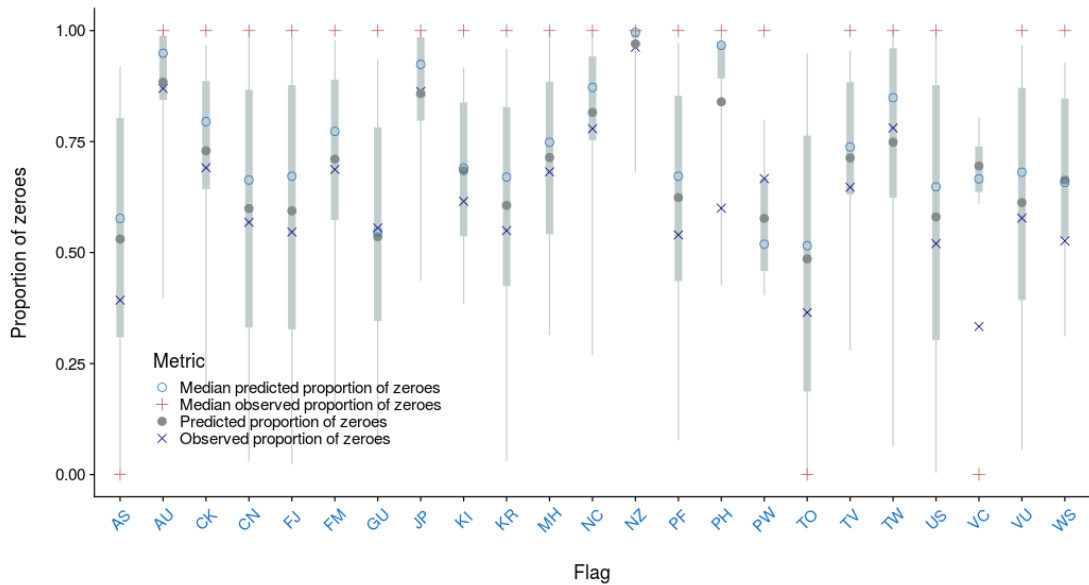


Figure A-25: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline bycatch fleet by flag. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

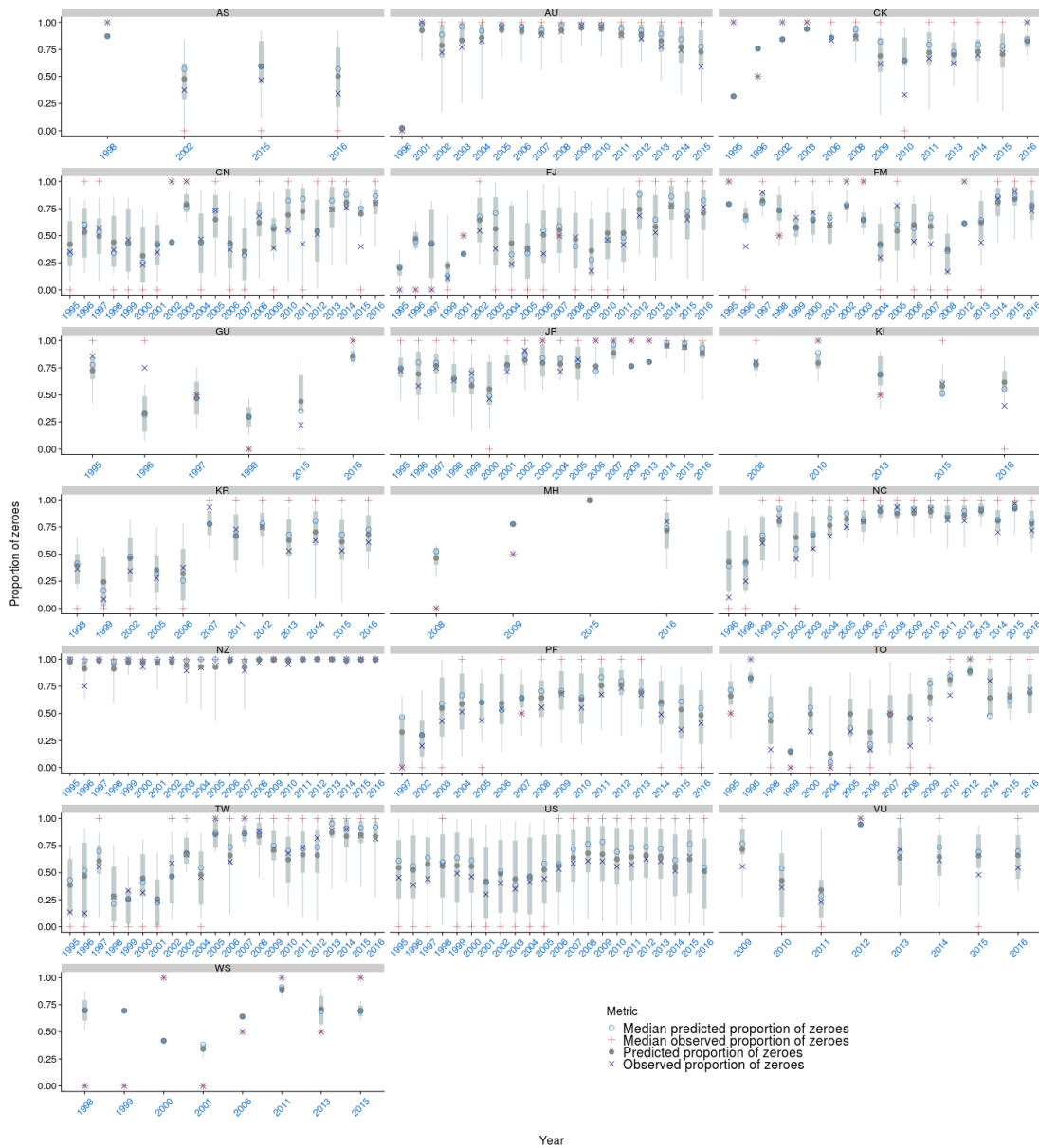


Figure A-26: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline bycatch fleet by flag and year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

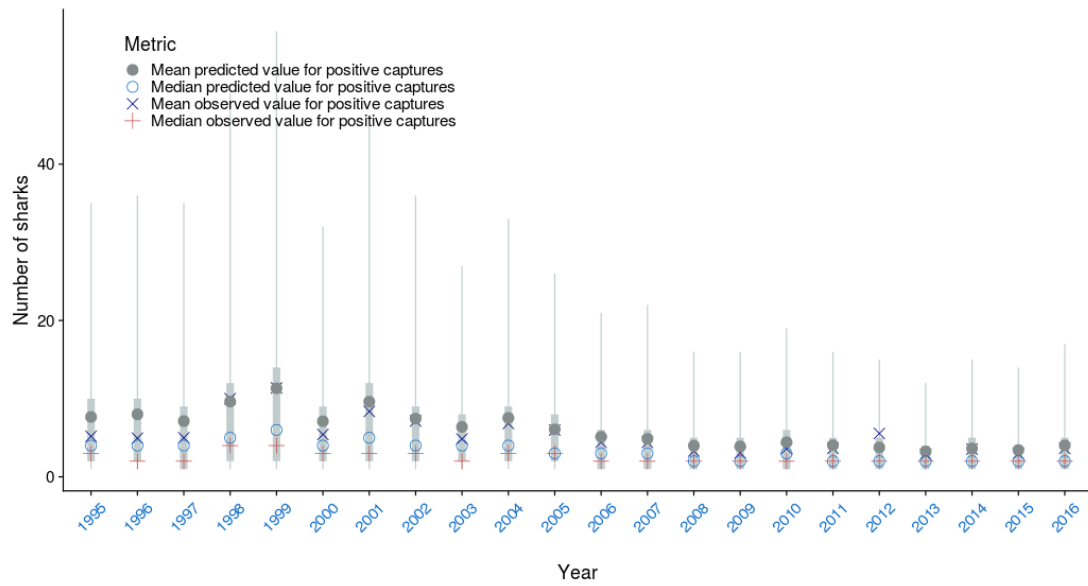


Figure A-27: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by year for the observer catch rate model for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

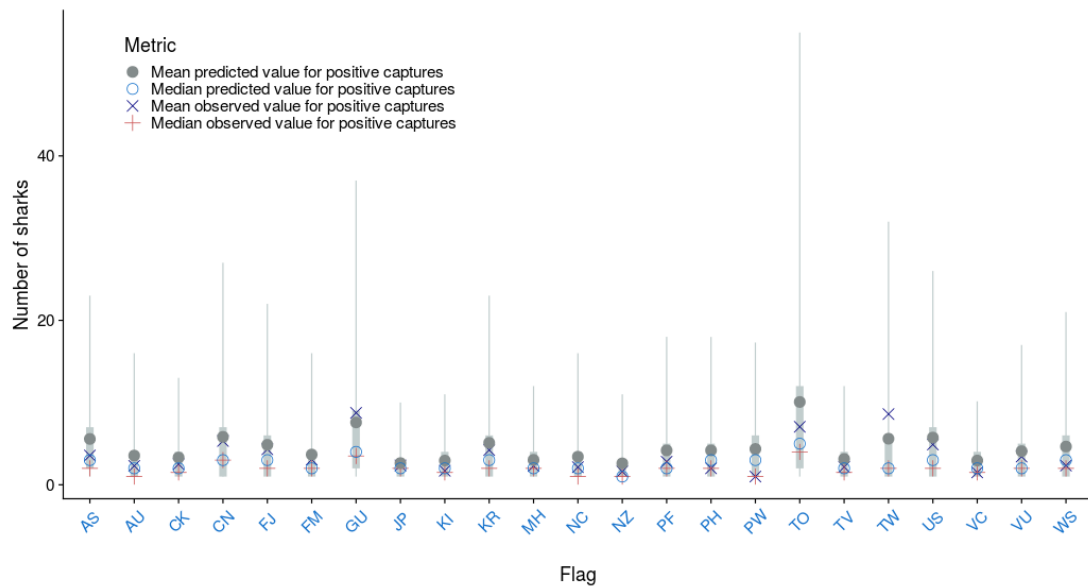


Figure A-28: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by flag for the observer catch rate model for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

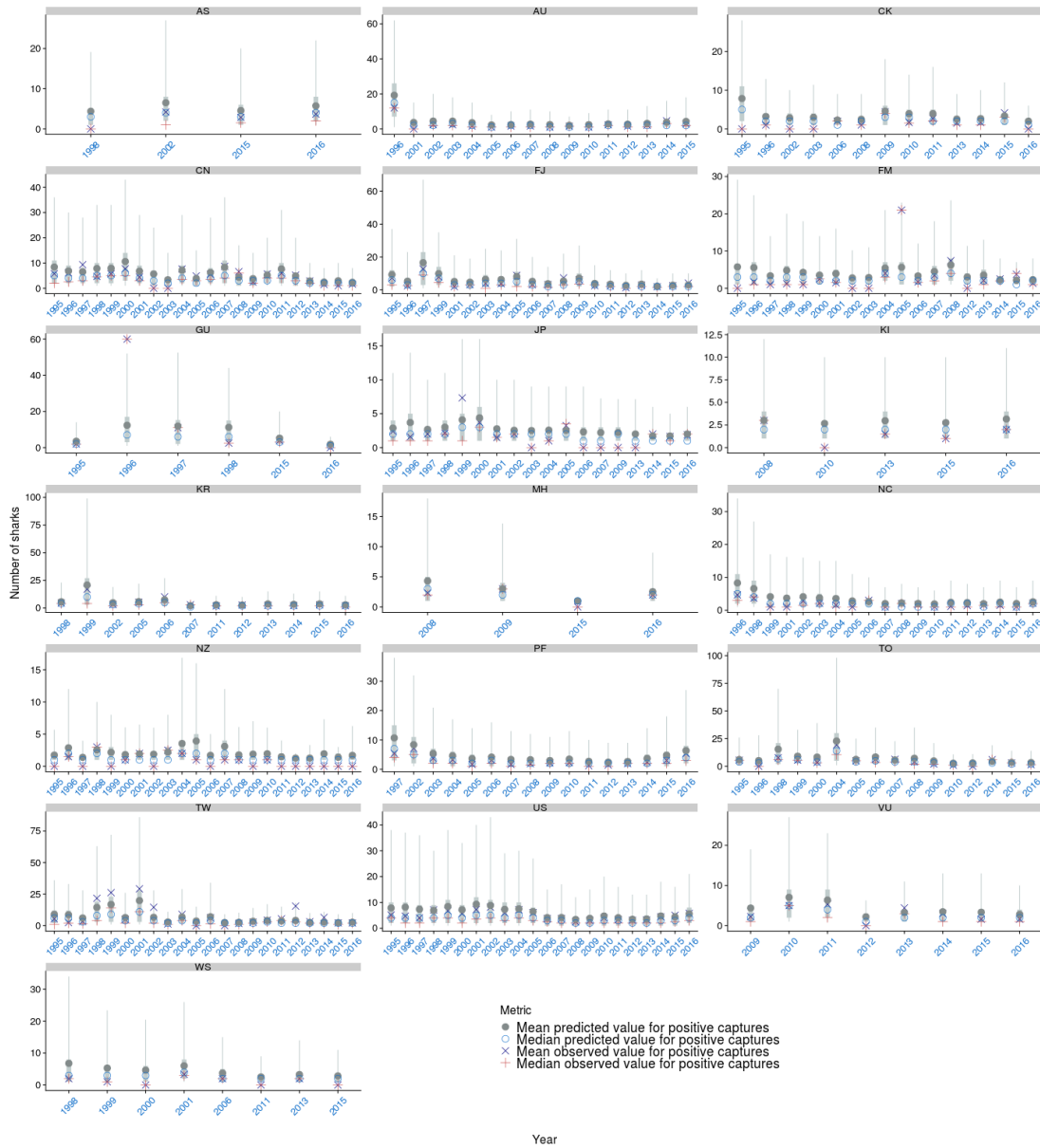


Figure A-29: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by flag and year for the observer catch rate model for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

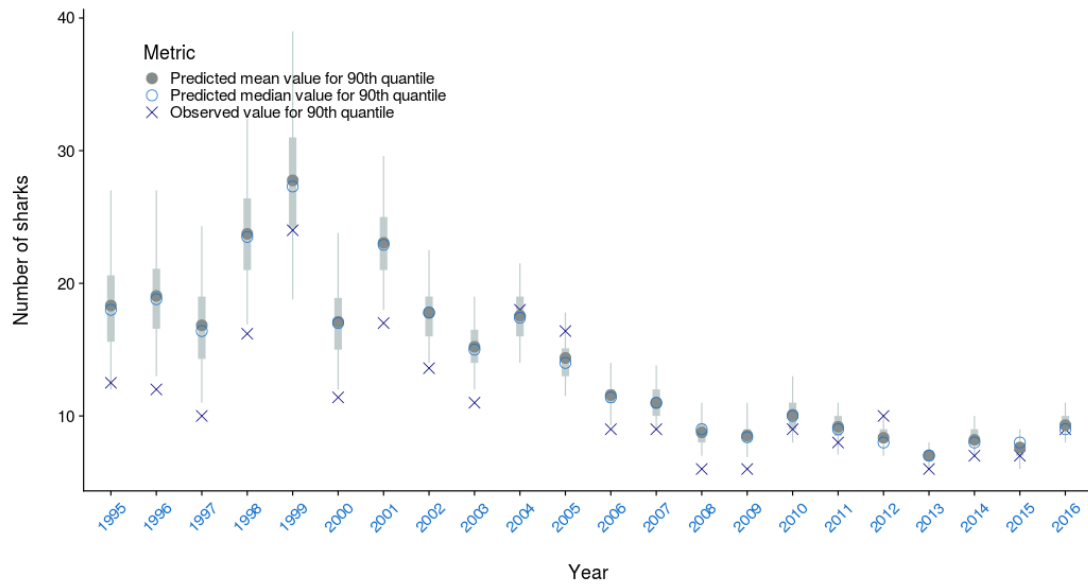


Figure A-30: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by year for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

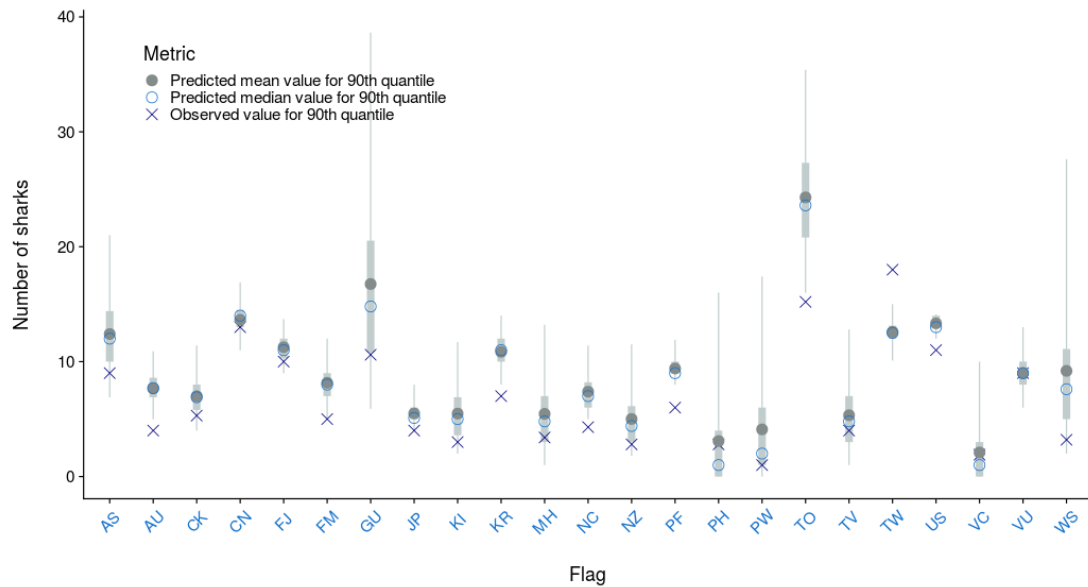


Figure A-31: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by flag for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

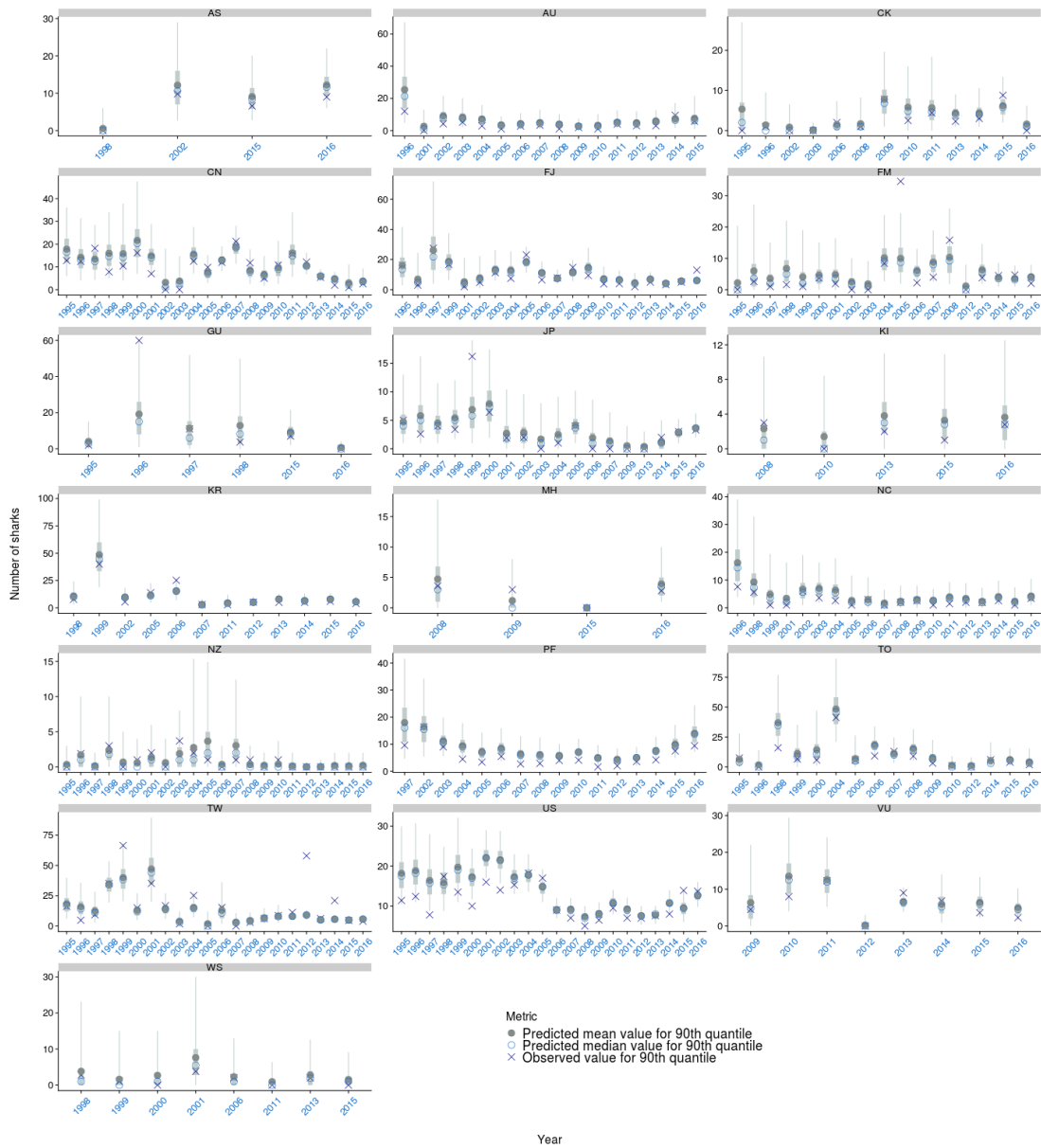


Figure A-32: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by flag and year for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

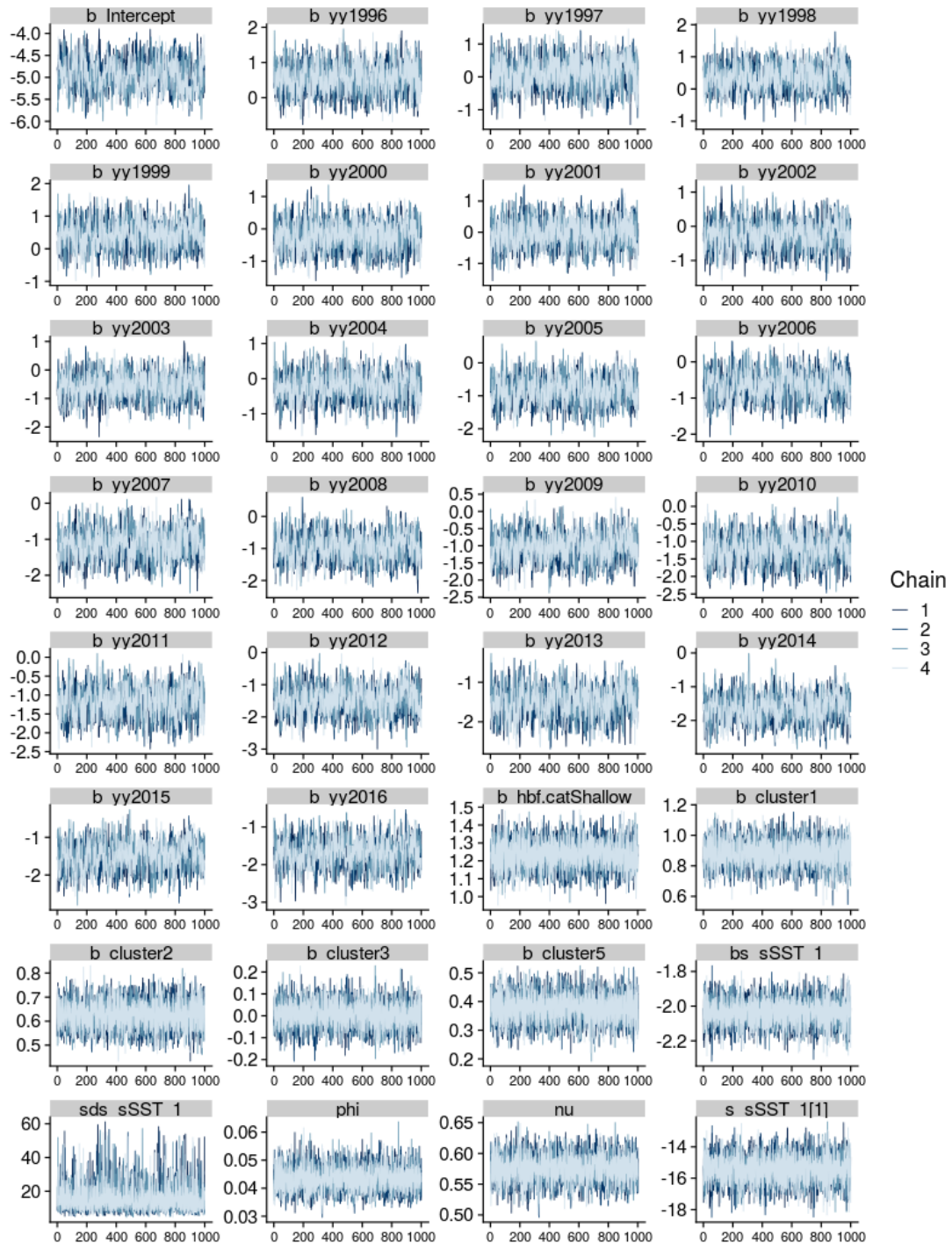


Figure A-33: Markov Chain Monte Carlo (MCMC) traces for key parameters of the observer catch rate model for the longline bycatch fleet. Chains (4) are shown in different colours.

A.2 Diagnostics for the observer catch rates model for the longline target fleet

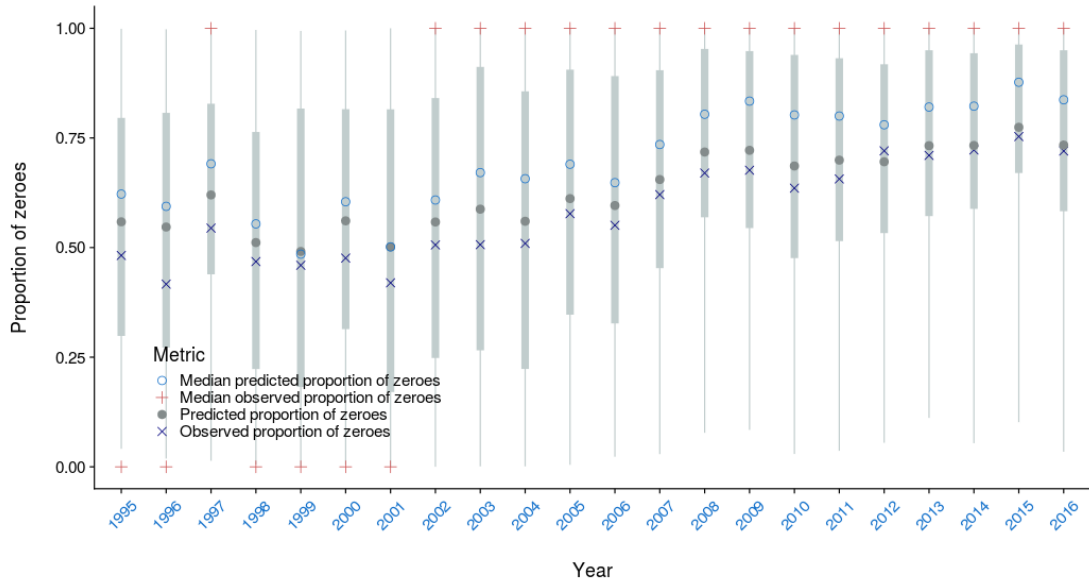


Figure A-34: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline target fleet by year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

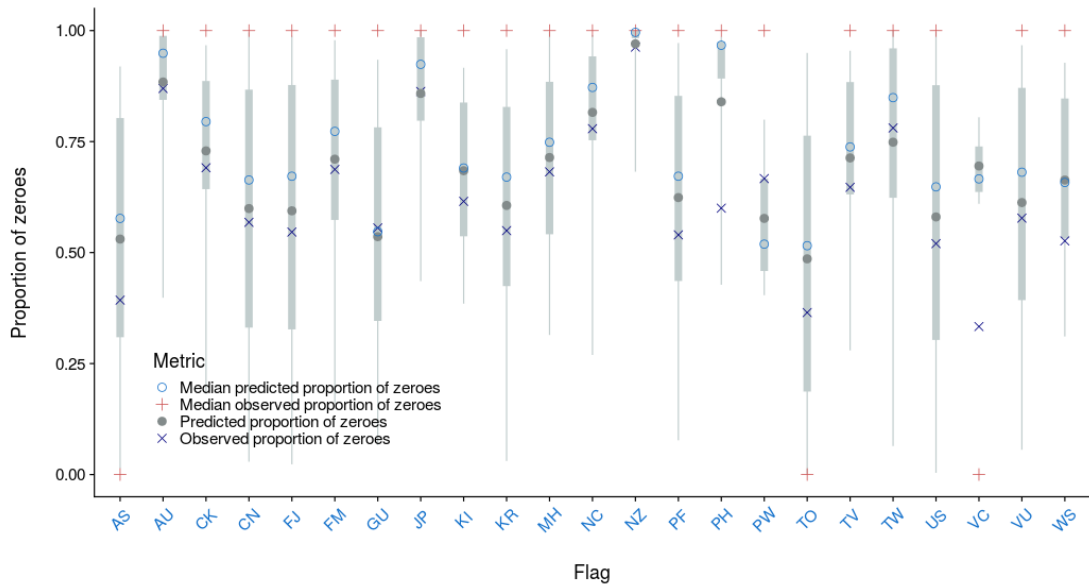


Figure A-35: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline target fleet by flag. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

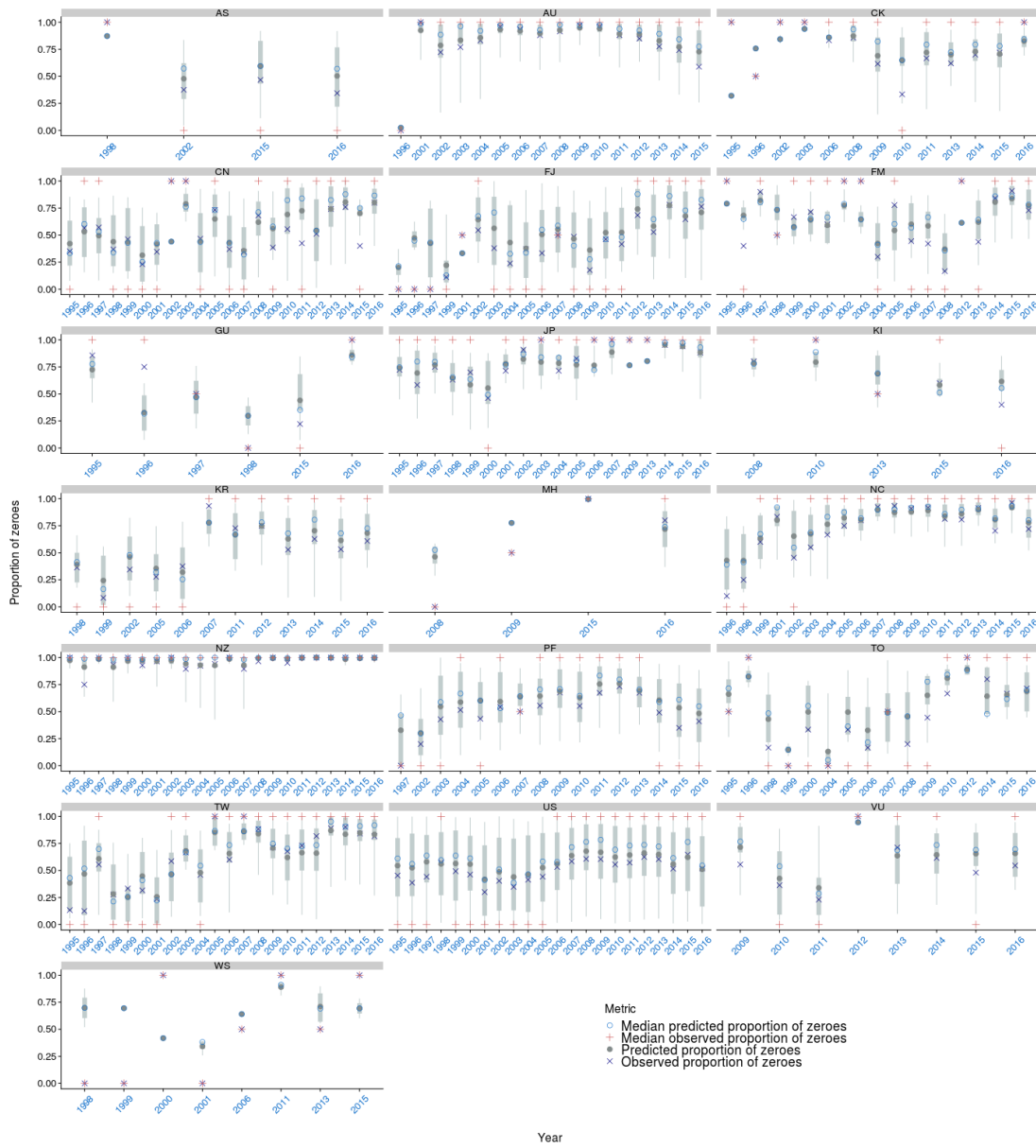


Figure A-36: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline target fleet by flag and year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

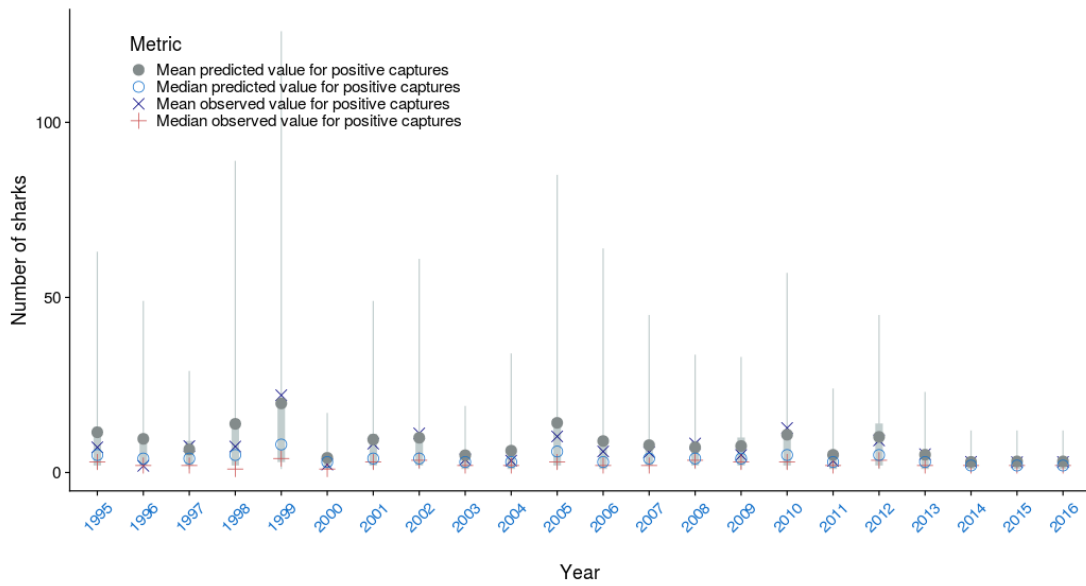


Figure A-37: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by year for the observer catch rate model for the longline target fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

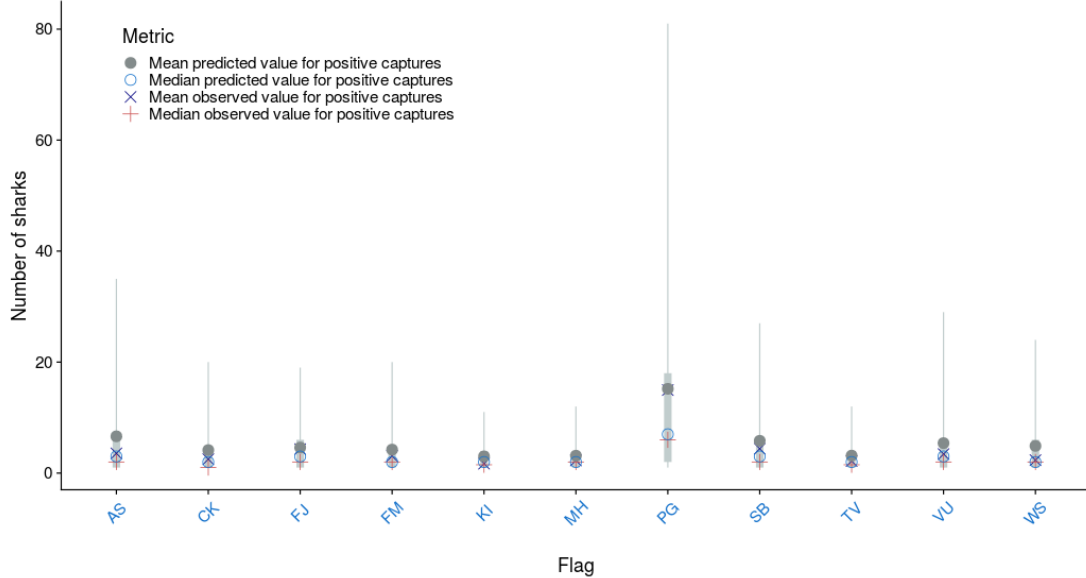


Figure A-38: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by flag for the observer catch rate model for the longline target fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

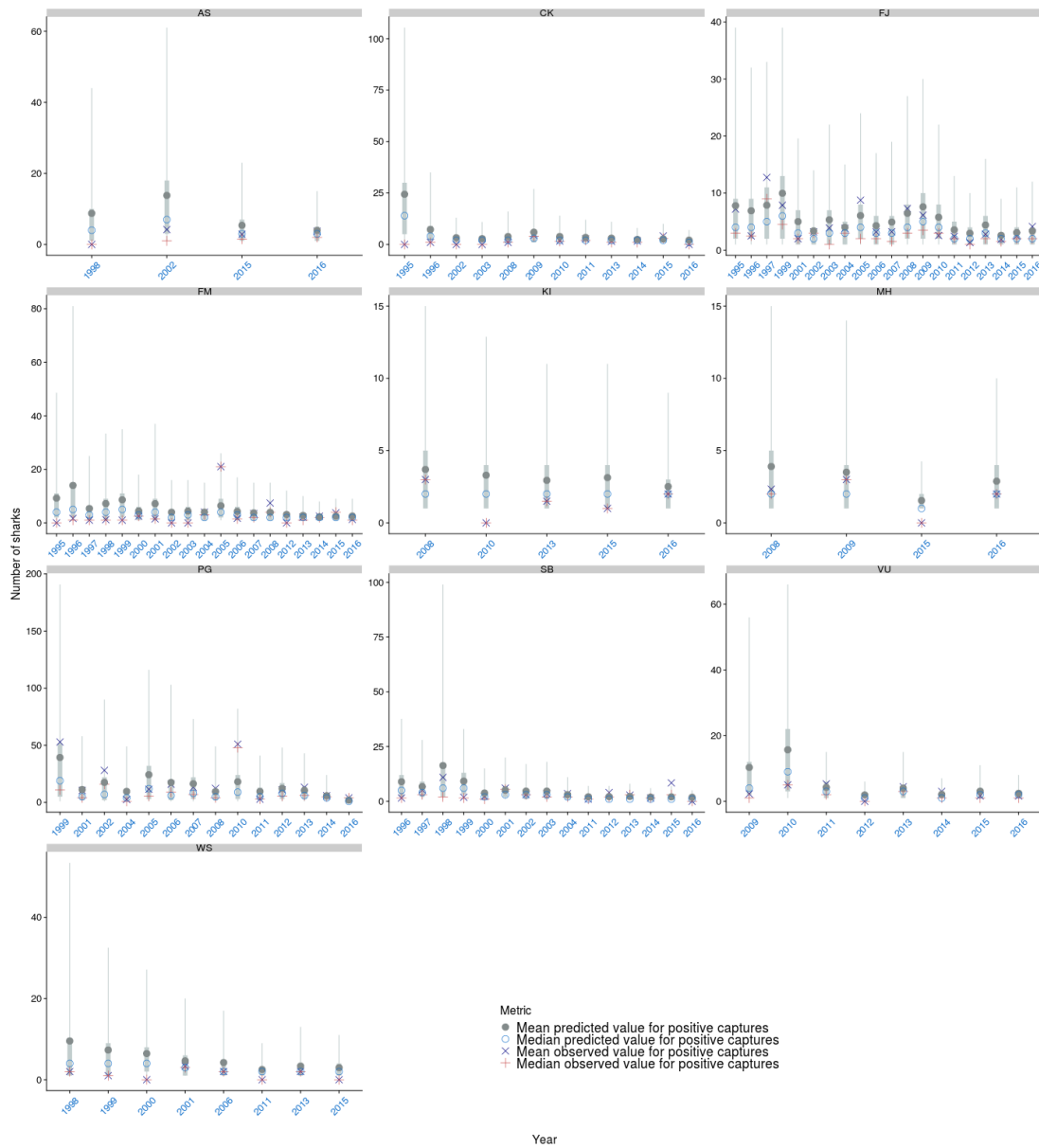


Figure A-39: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by flag and year for the observer catch rate model for the longline target fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

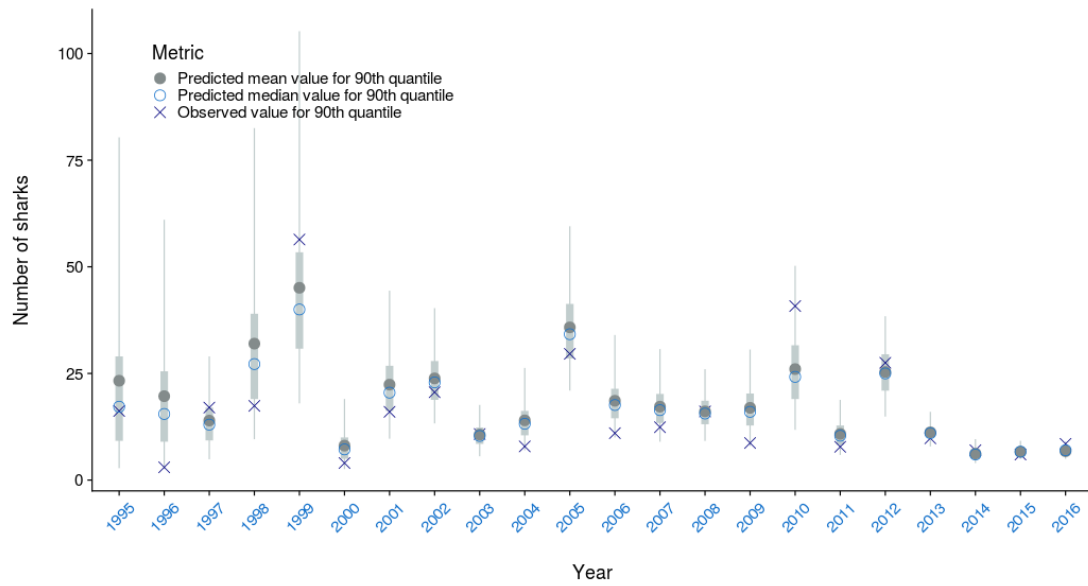


Figure A-40: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by year for the longline target fleet. The median ± 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

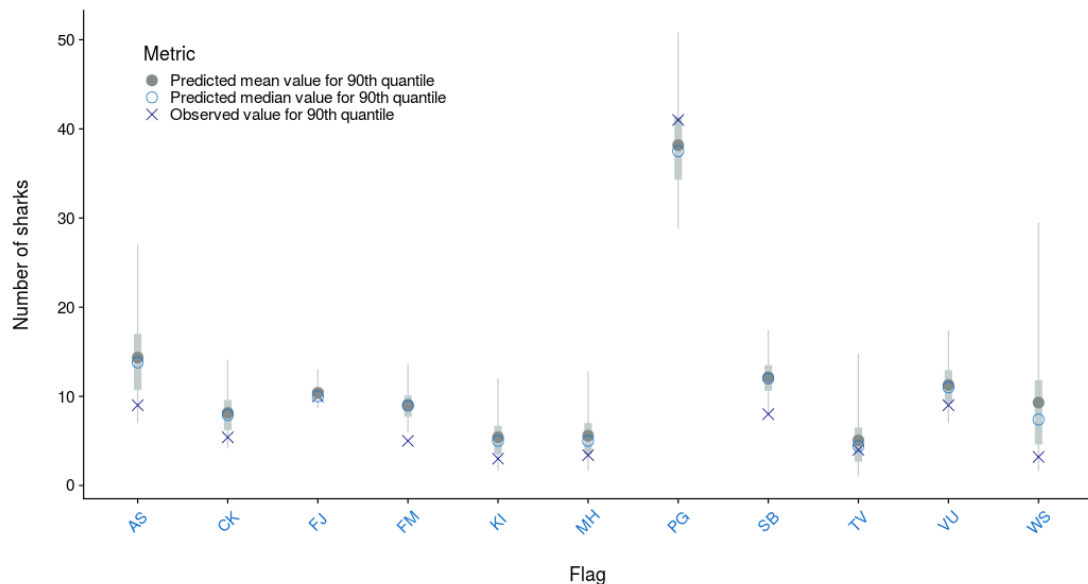


Figure A-41: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by flag for the longline target fleet. The median ± 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

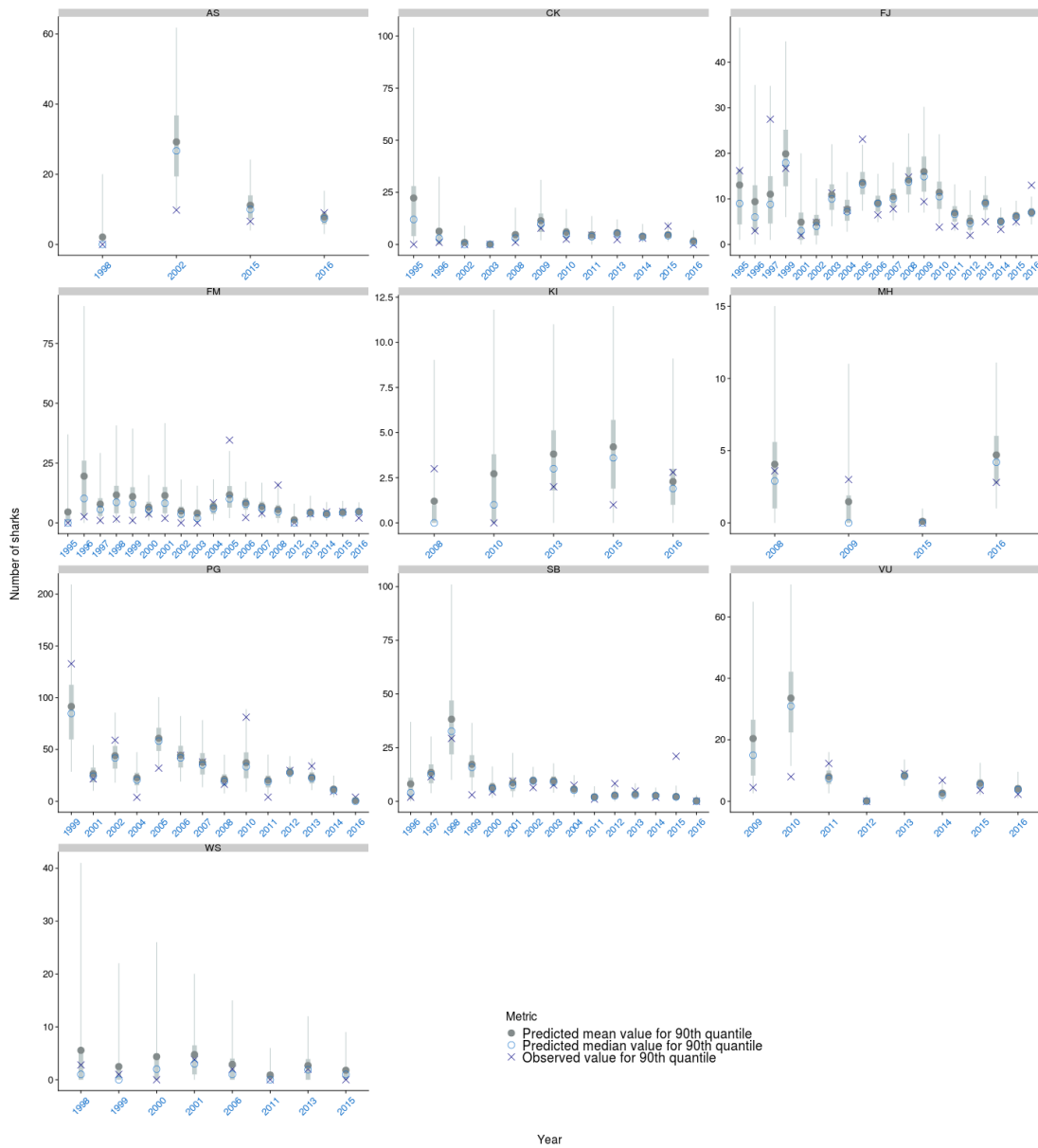


Figure A-42: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by flag and year for the longline target fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

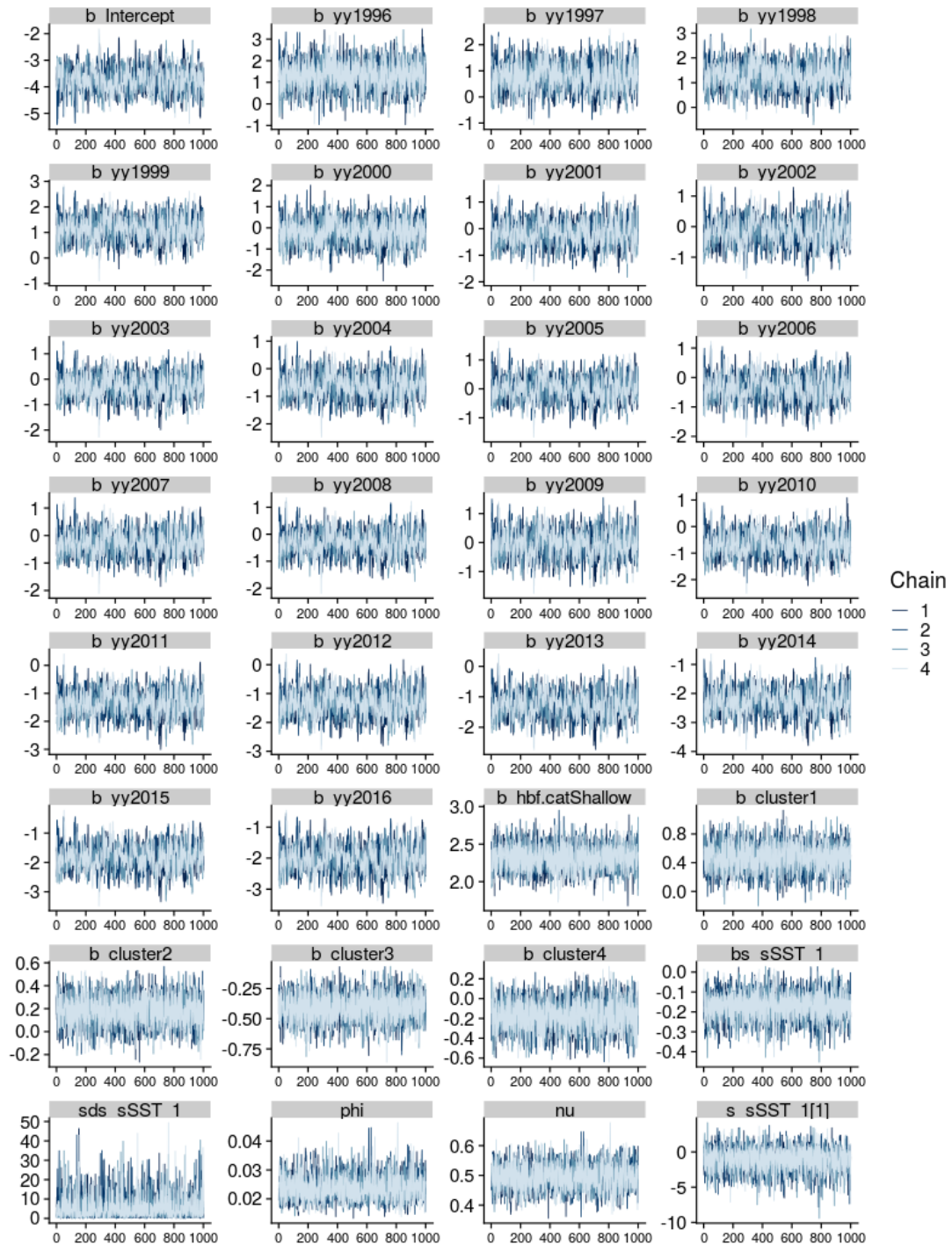


Figure A-43: Markov Chain Monte Carlo (MCMC) traces for key parameters of the observer catch rate model for the longline target fleet. Chains (4) are shown in different colours.

A.3 Diagnostics for the observer catch rates model for the purse seine fleet

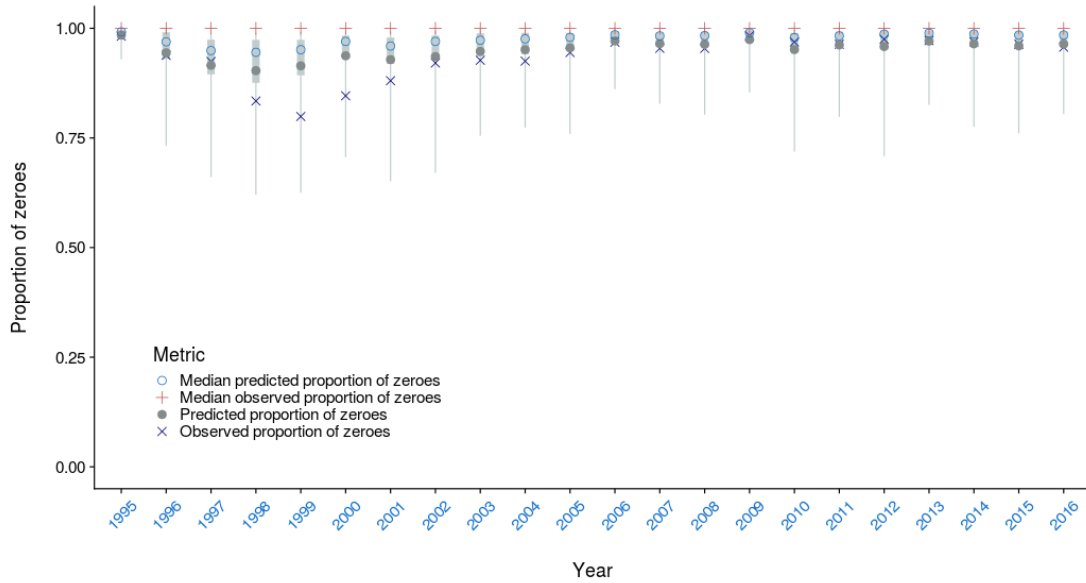


Figure A-44: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the purse seine fleet by year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

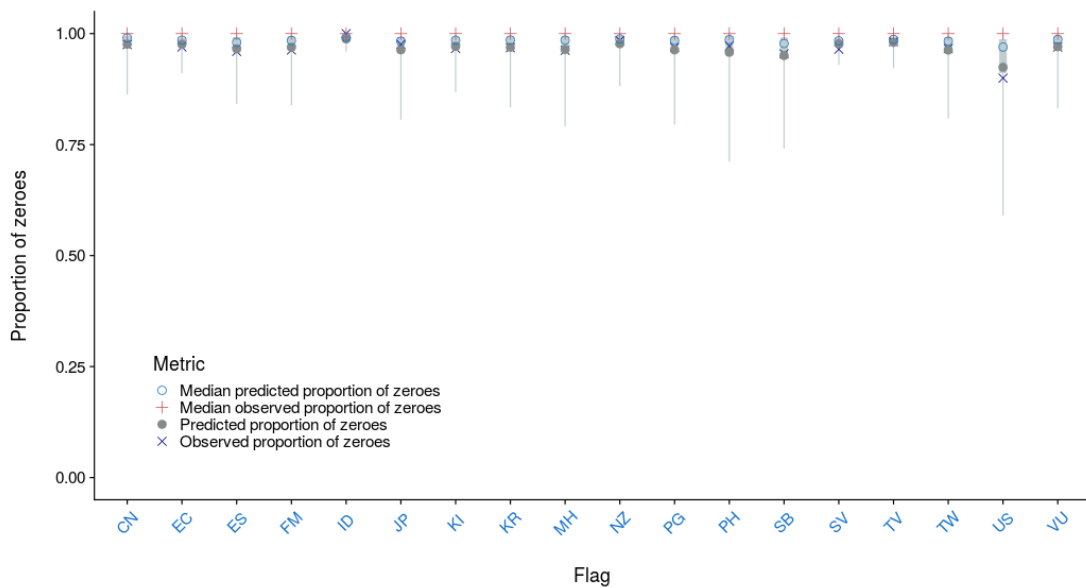


Figure A-45: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the purse seine fleet by flag. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

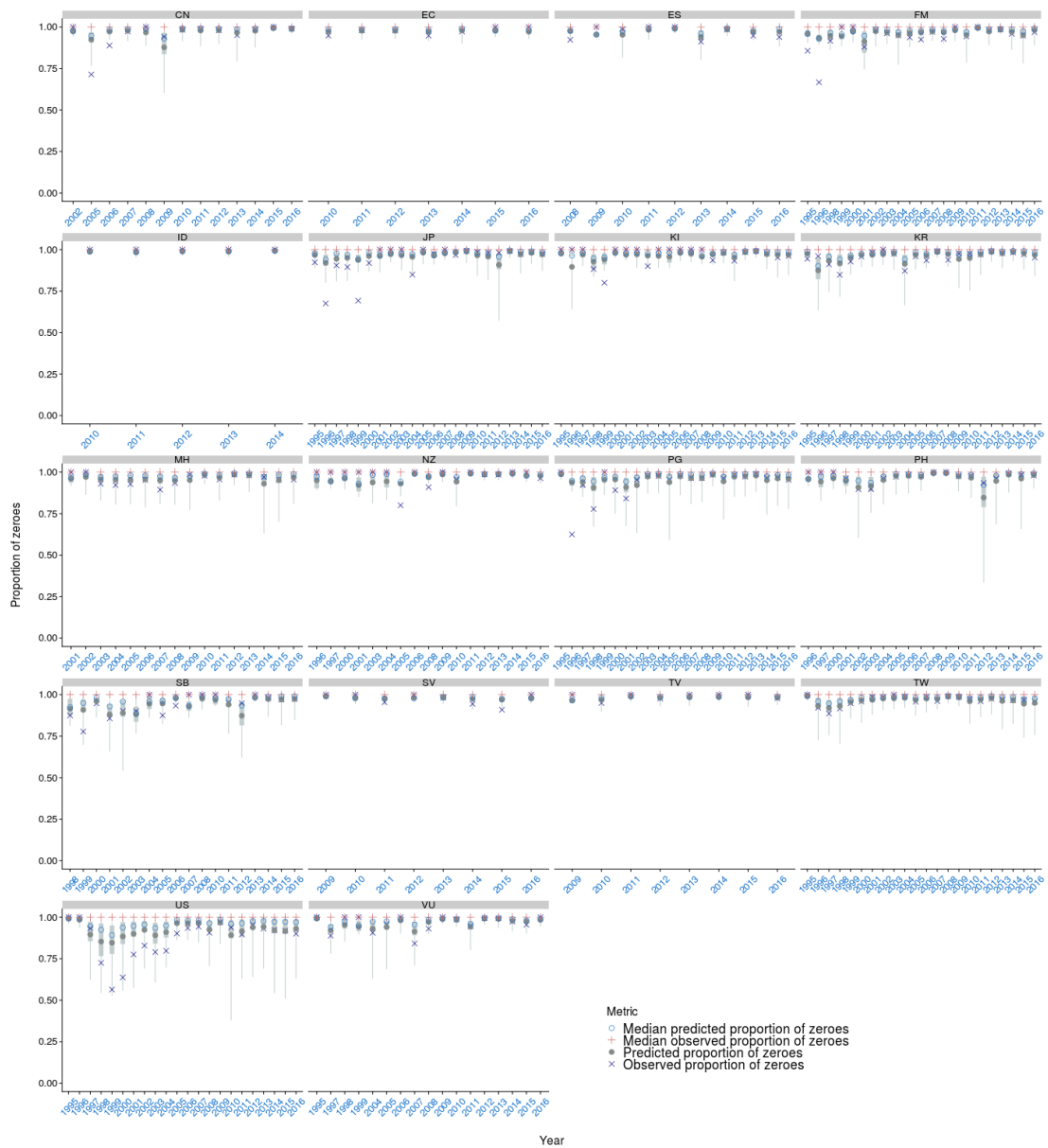


Figure A-46: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the purse seine fleet by flag and year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

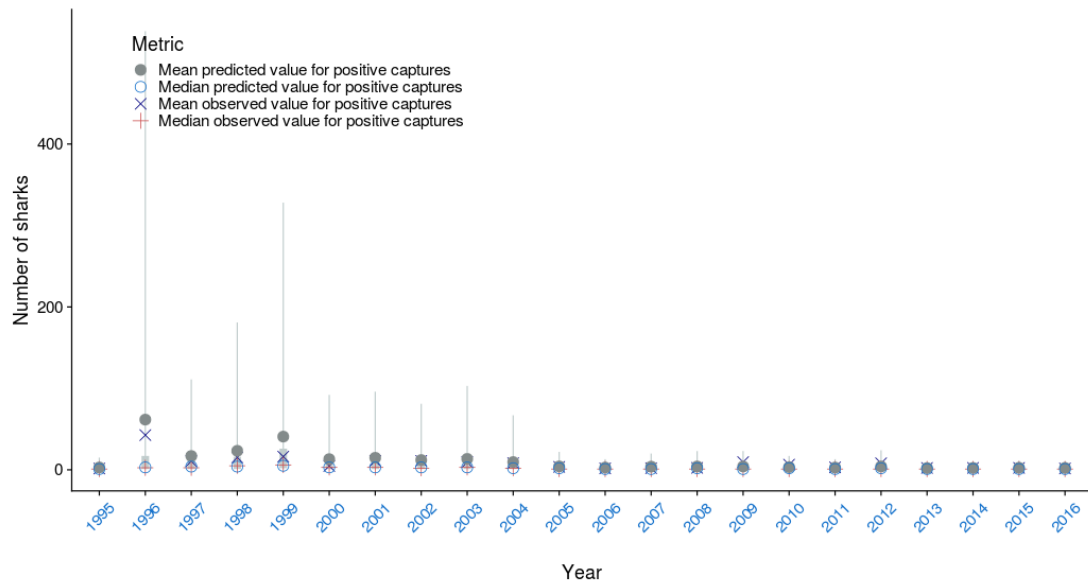


Figure A-47: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by year for the observer catch rate model for the purse seine fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

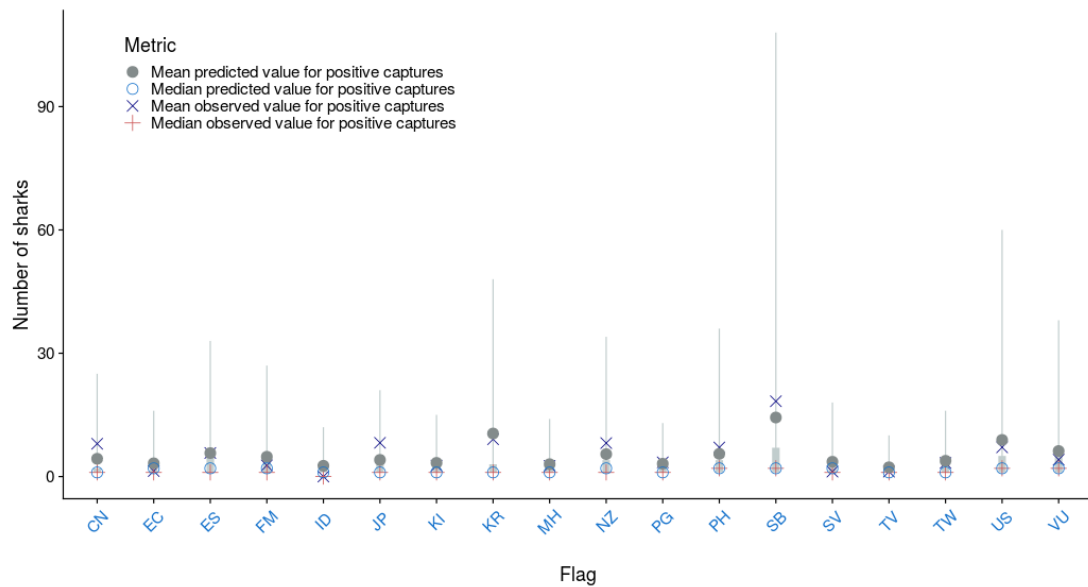


Figure A-48: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by flag for the observer catch rate model for the purse seine fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

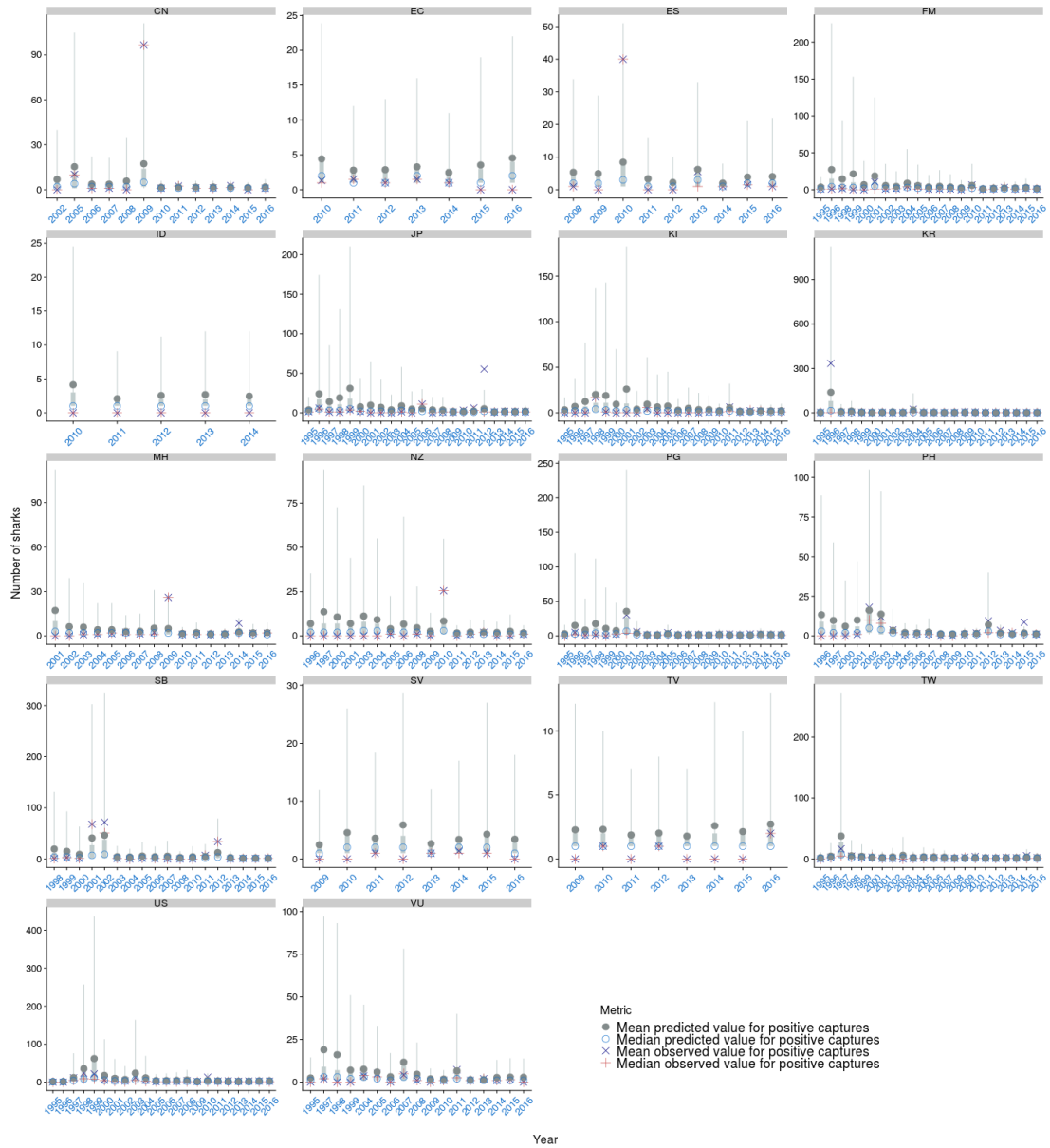


Figure A-49: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by flag and year for the observer catch rate model for the purse seine fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

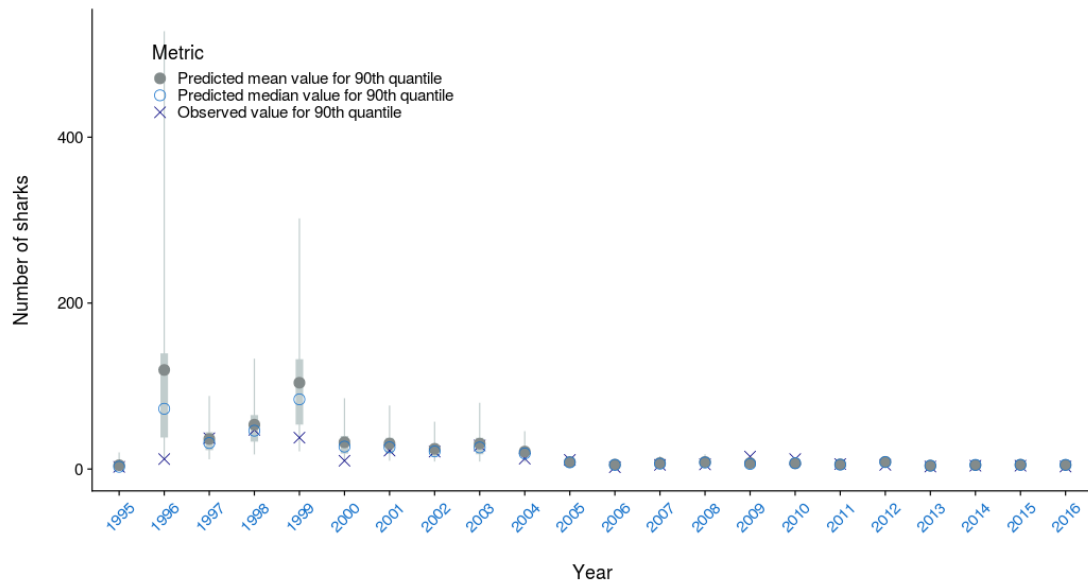


Figure A-50: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by year for the purse seine fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

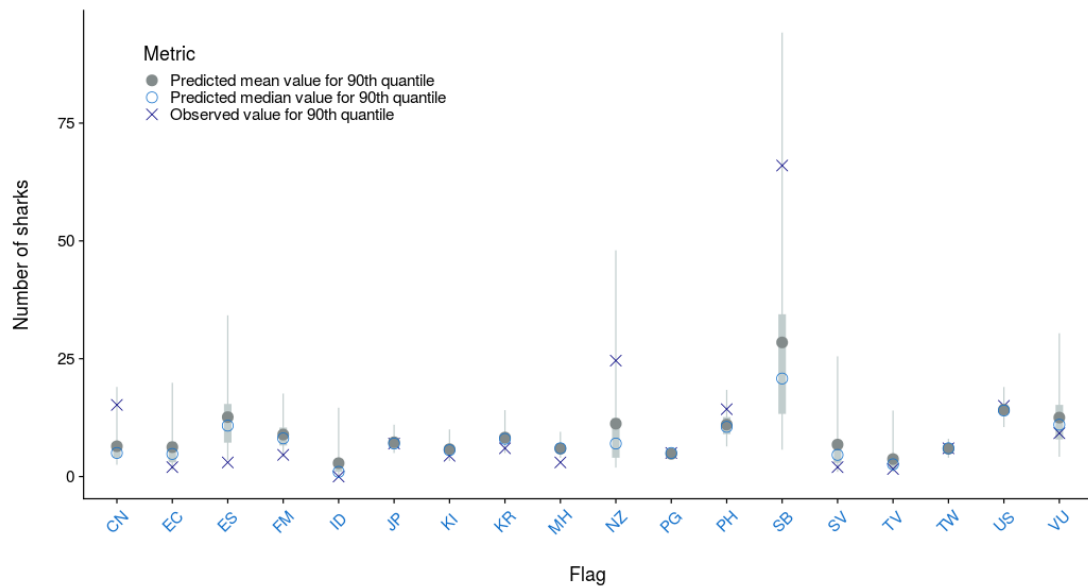


Figure A-51: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by flag for the purse seine fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

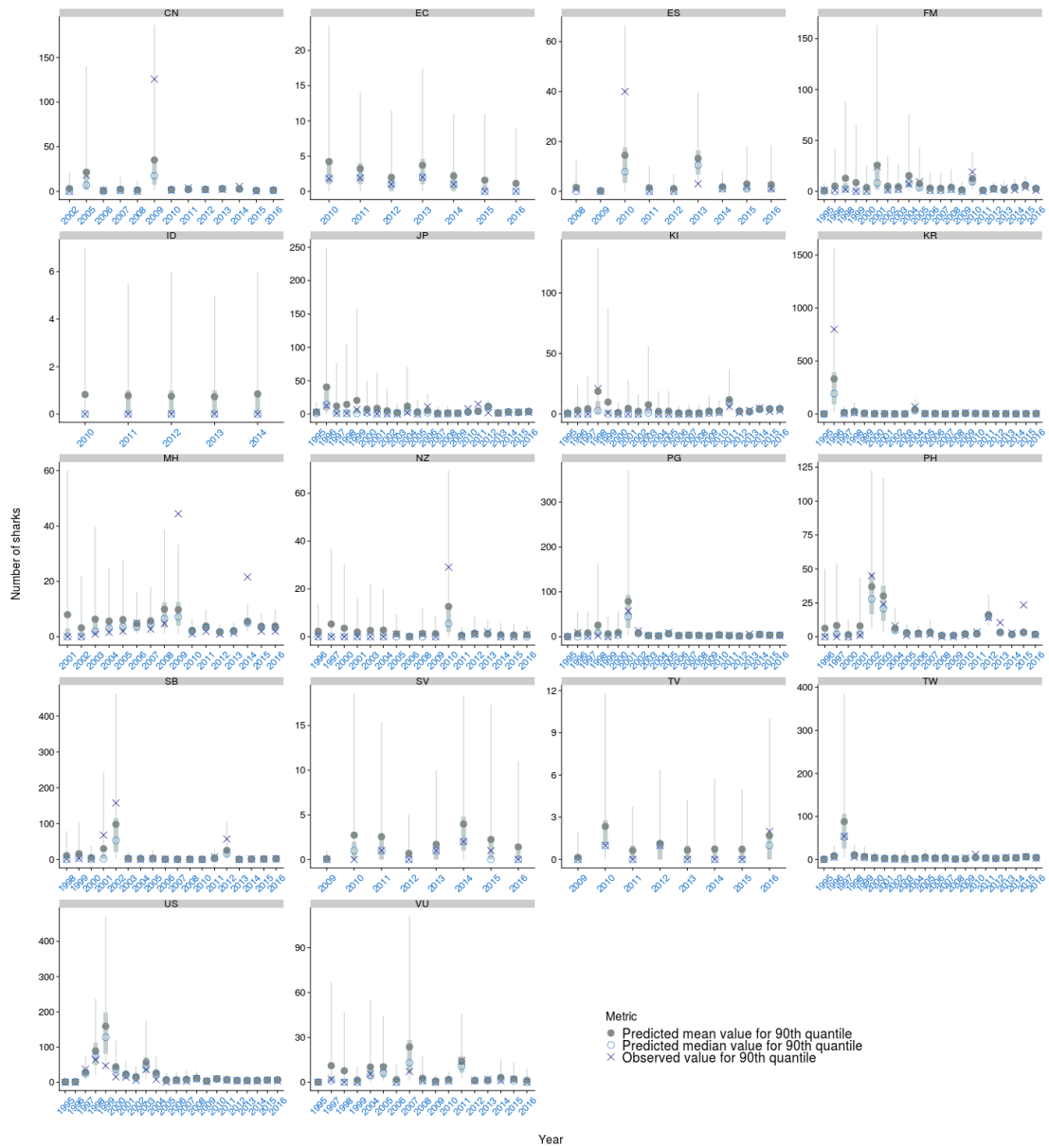


Figure A-52: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by flag and year for the purse seine fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

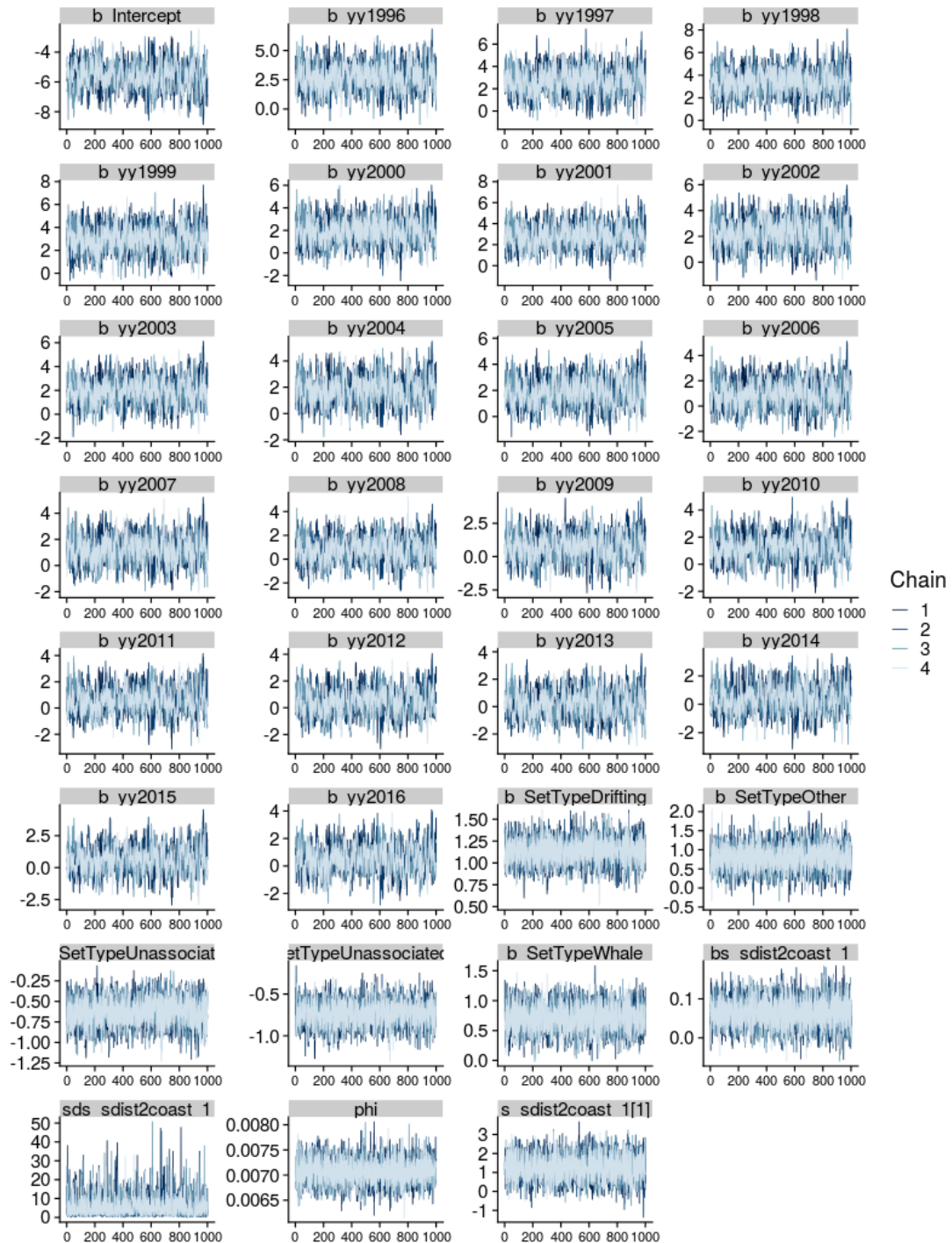


Figure A-53: Markov Chain Monte Carlo (MCMC) traces for key parameters of the observer catch rate model for the purse seine fleet. Chains (4) are shown in different colours.

A.4 Diagnostics for the standardized CPUE for the longline bycatch fleet

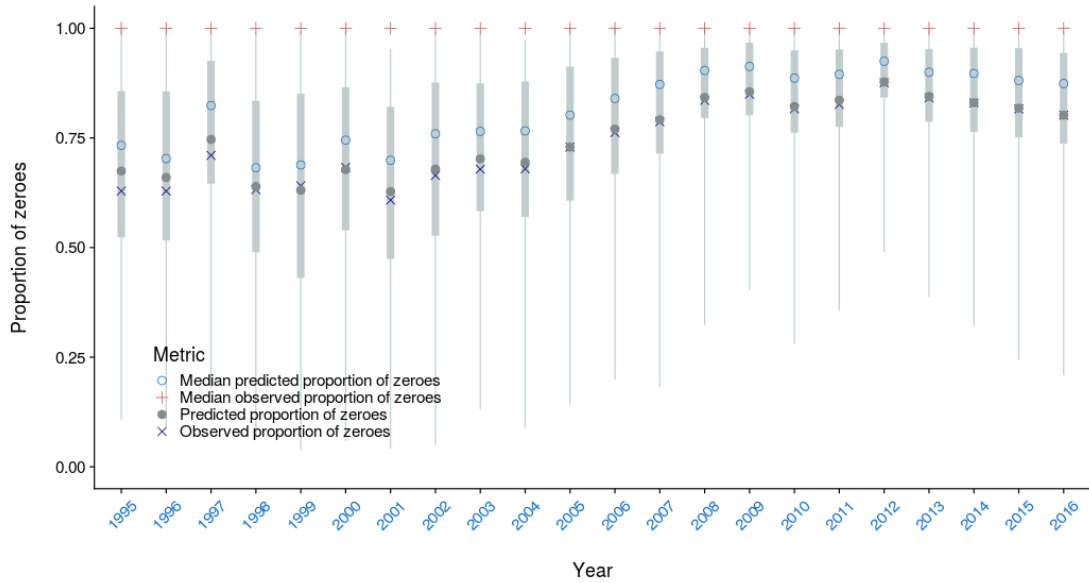


Figure A-54: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline bycatch fleet by year. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

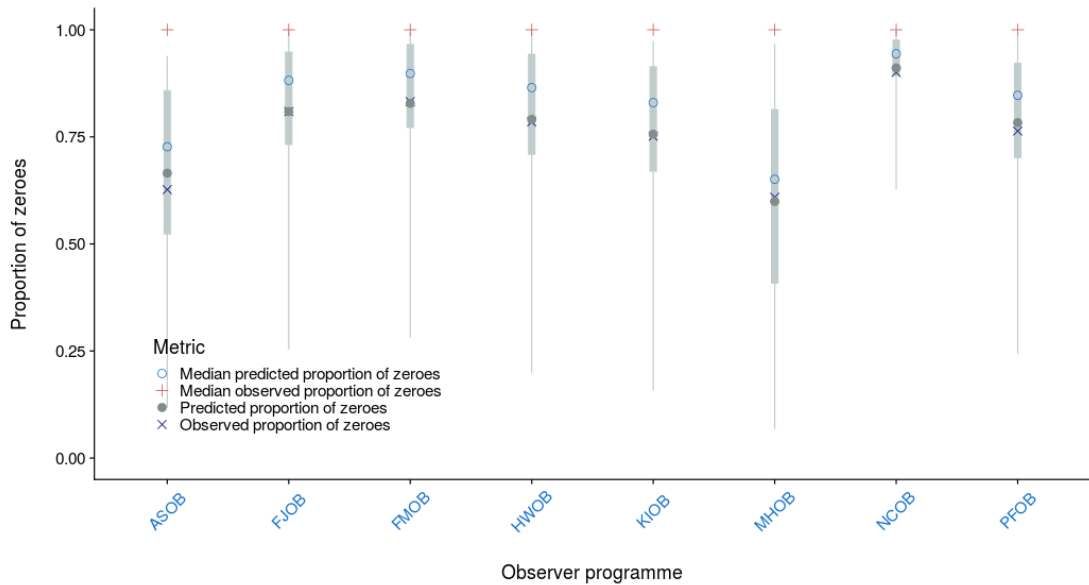


Figure A-55: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline bycatch fleet by observer programme. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

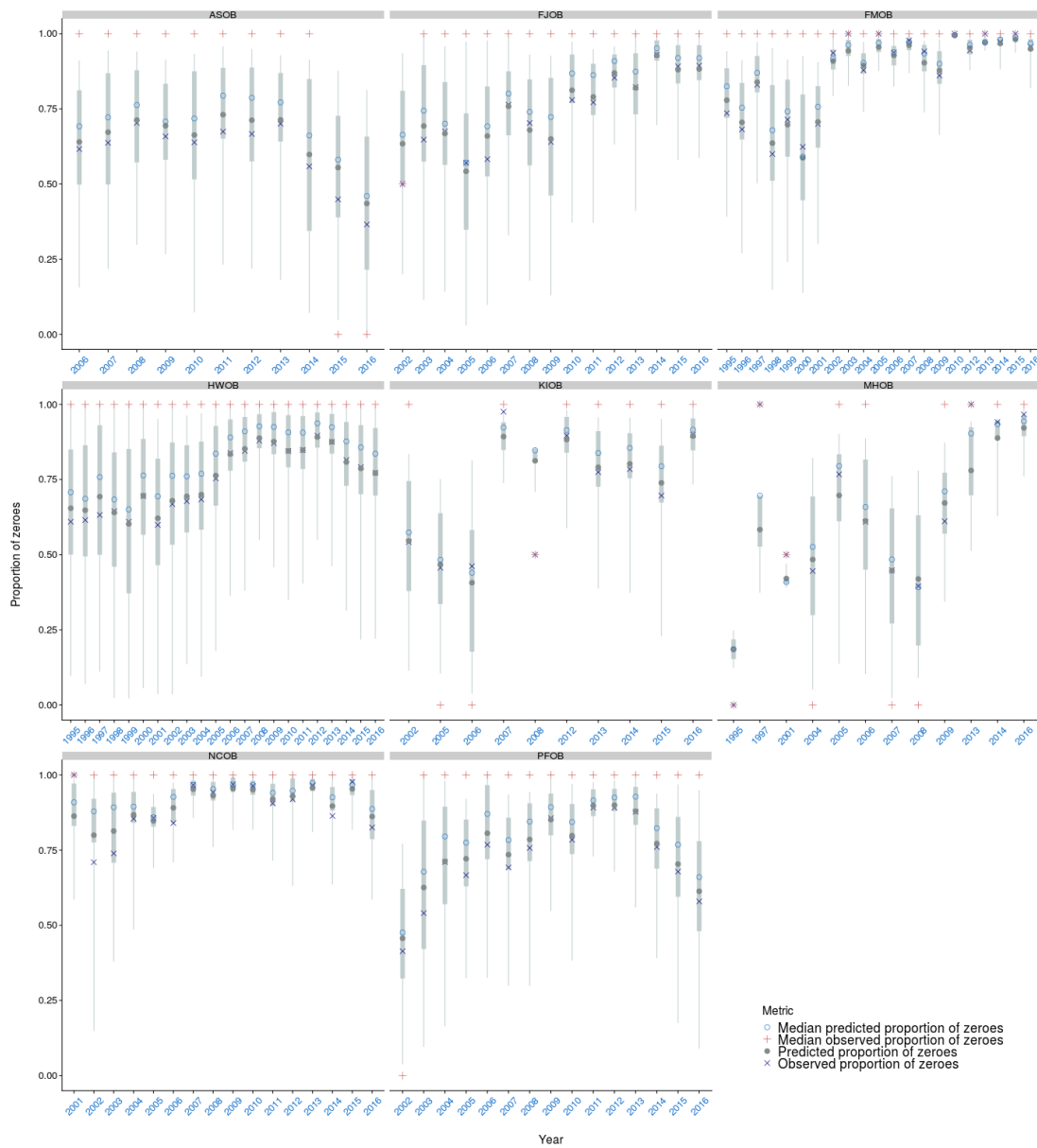


Figure A-56: Mean and median proportion of zero observed vs. predicted by fishing event for the observer catch rate model for the longline bycatch fleet by observer programme and year. The median $\pm 25^{\text{th}}$ quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

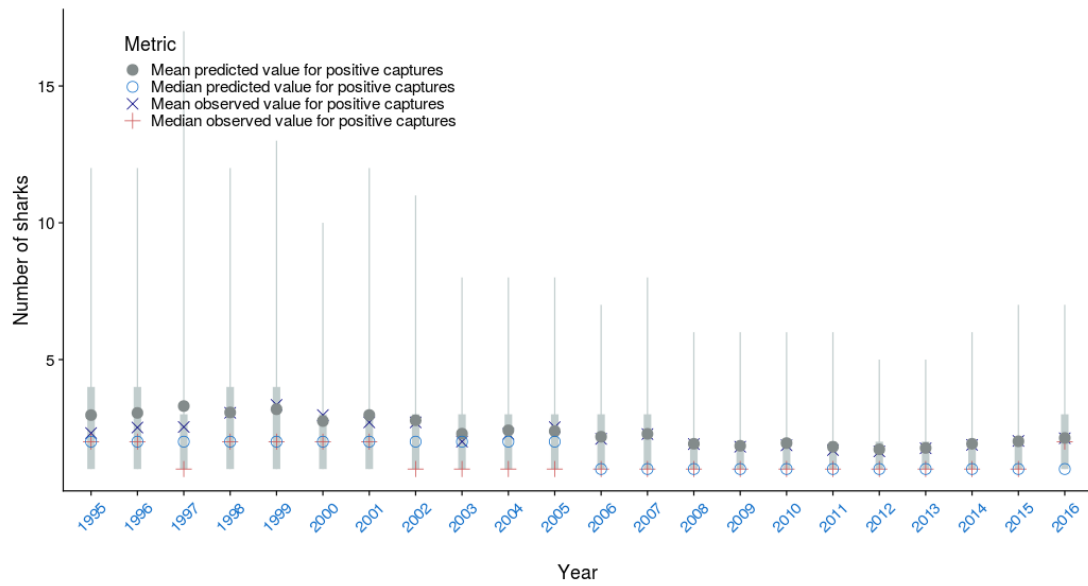


Figure A-57: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by year for the observer catch rate model for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

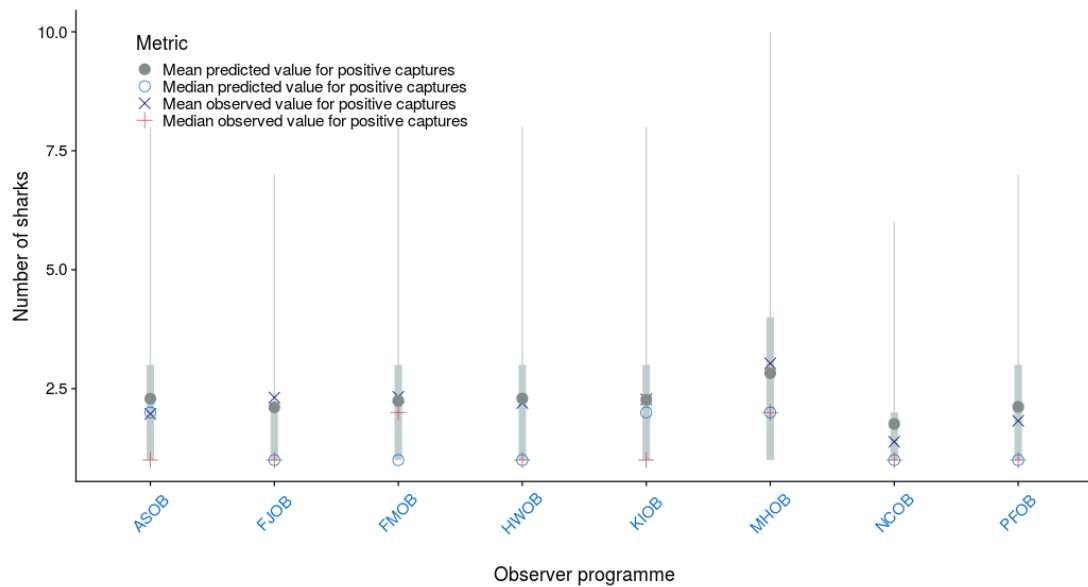


Figure A-58: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by observer programme for the observer catch rate model for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

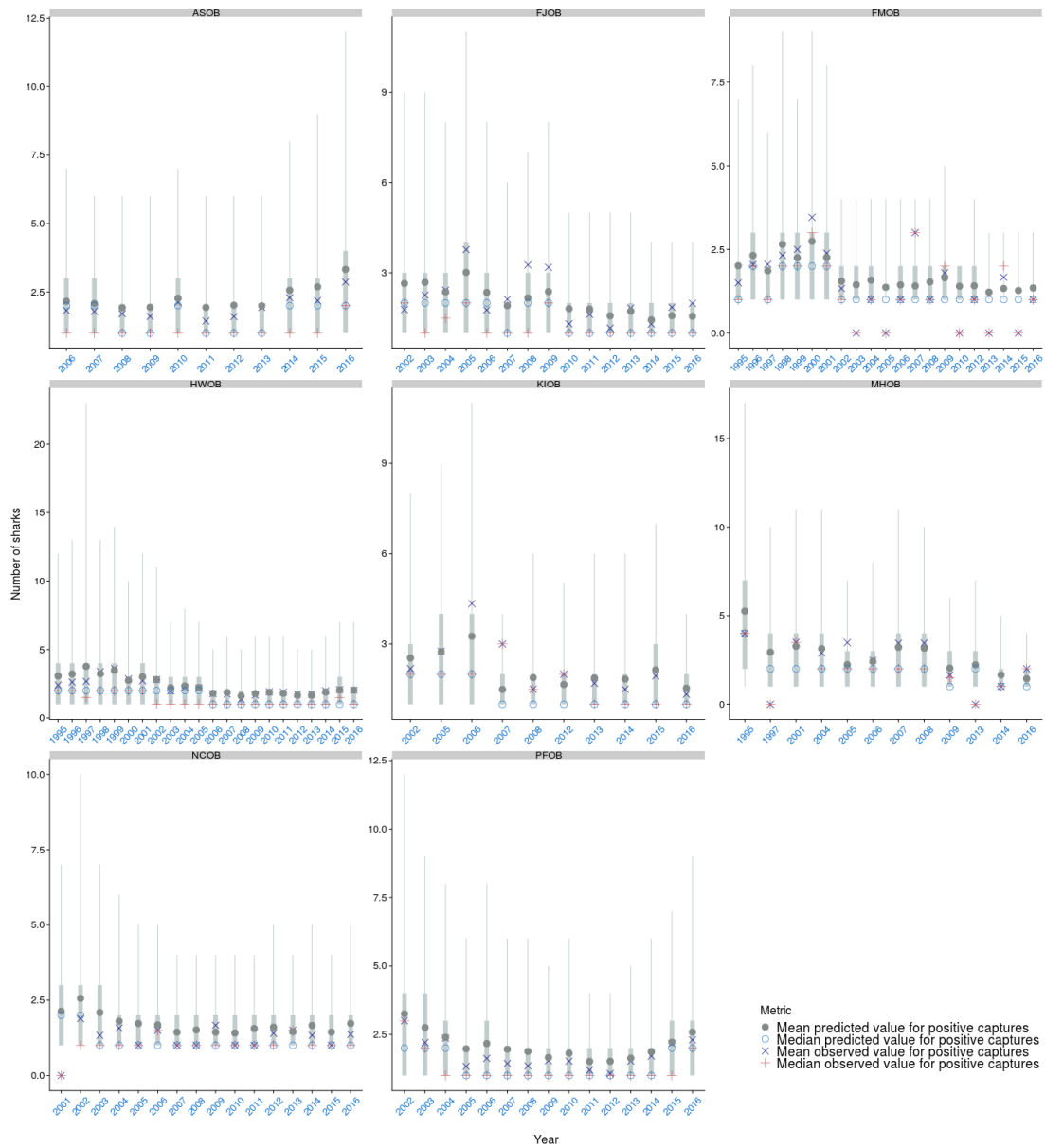


Figure A-59: Observed vs. predicted mean and median number of oceanic whitetip shark caught by fishing event where catch > 0 by observer programme and year for the observer catch rate model for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

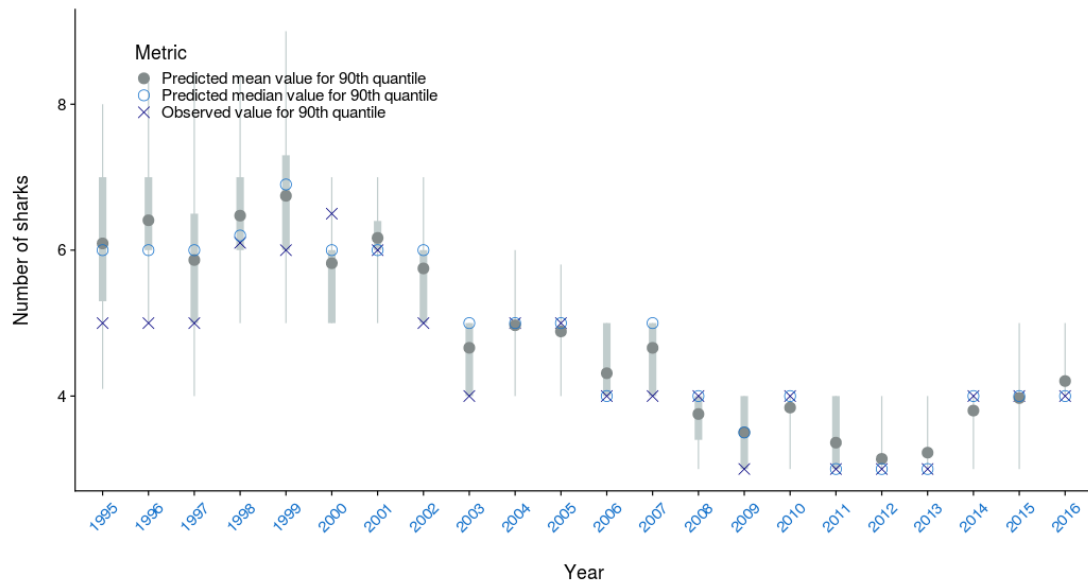


Figure A-60: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by year for the longline bycatch fleet. The median ± 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

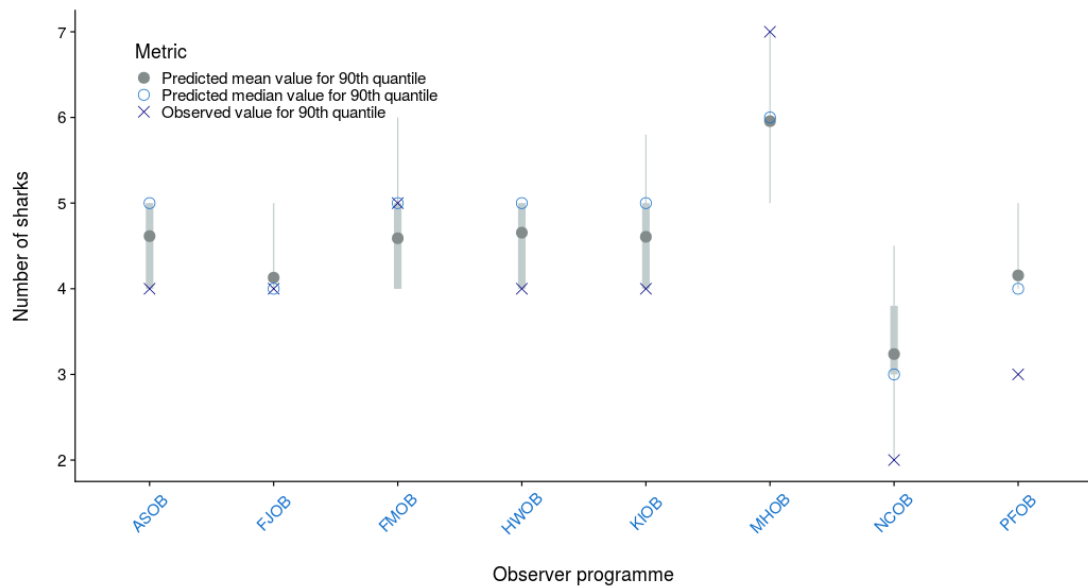


Figure A-61: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by observer programme for the longline bycatch fleet. The median ± 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5 - 97.5th quantile range.

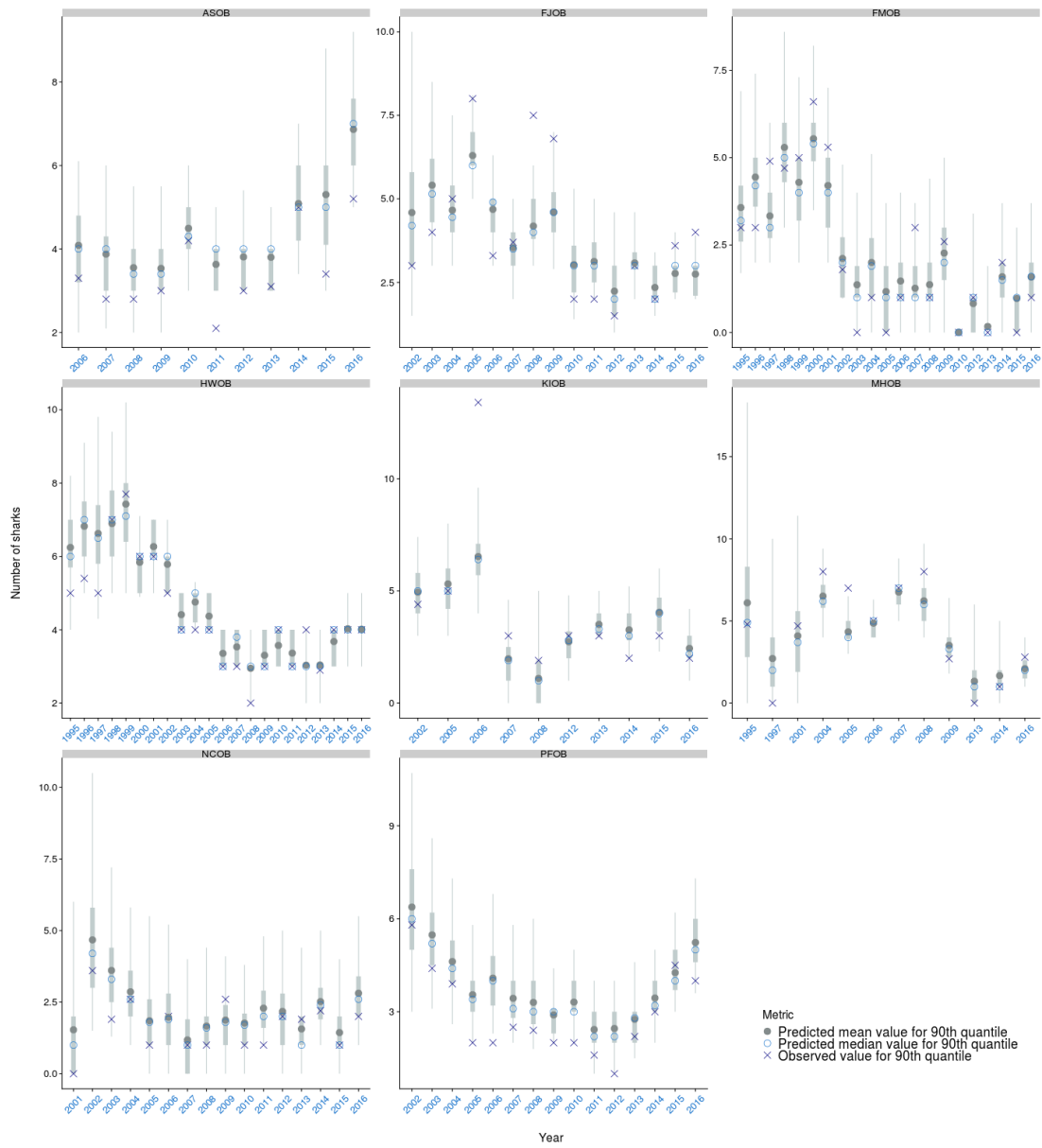


Figure A-62: Observed vs. predicted mean and median position of the 90th quantile of oceanic whitetip shark caught by fishing event where catch > 0 by observer programme and year for the longline bycatch fleet. The median \pm 25th quantiles for the predictions is shown in grey; the whiskers cover the 2.5-97.5th quantile range.

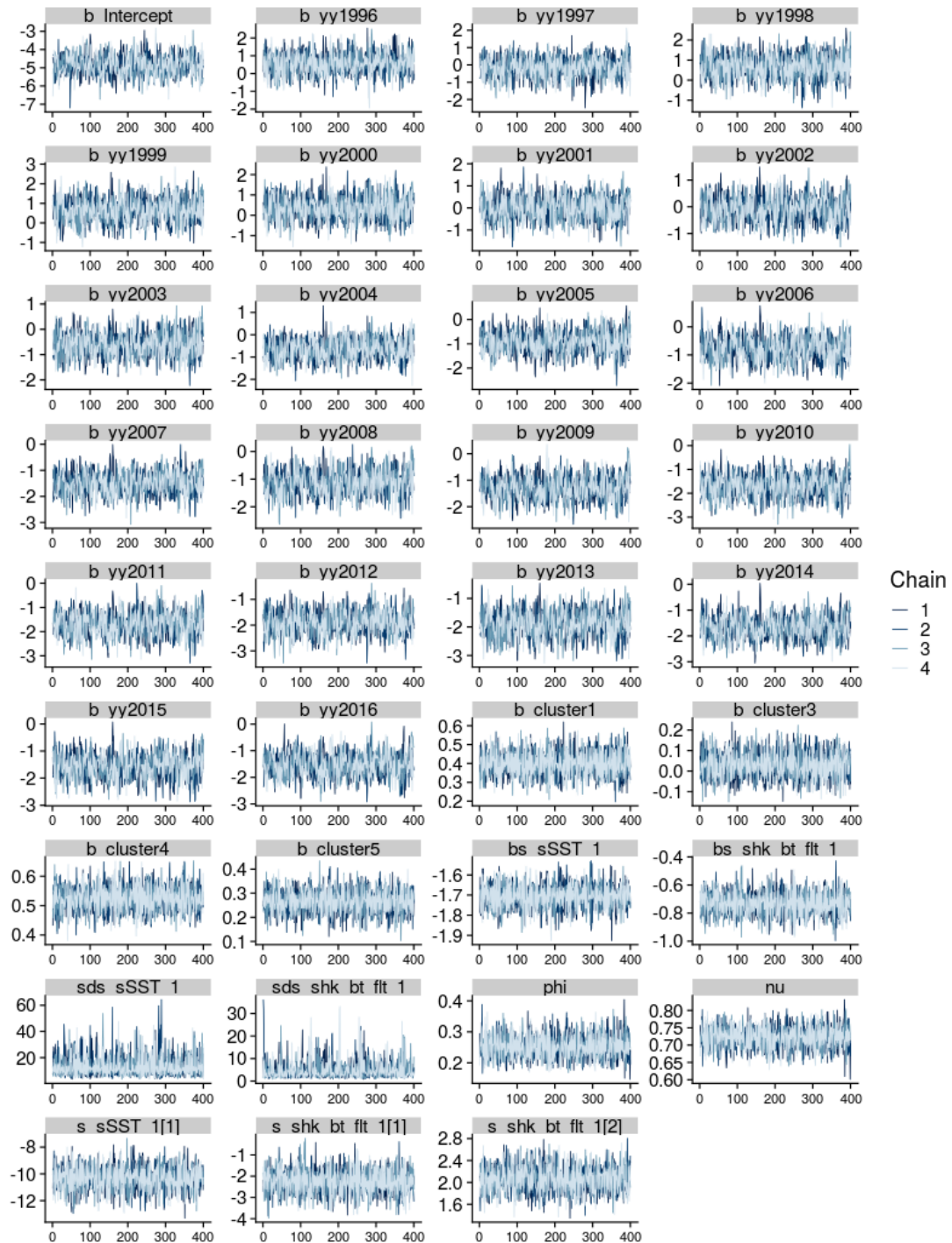


Figure A-63: Markov Chain Monte Carlo (MCMC) traces for key parameters of the observer catch rate model for the longline bycatch fleet. Chains (4) are shown in different colours.

APPENDIX B: Estimated oceanic whitetip shark catch by fleet

Table B-6: Predicted catches (in thousand individuals) and uncertainty from key quantiles of the posterior distribution of predicted catches for the longline bycatch fleet

Year	Median	10 th	25 th	75 th	90 th
1995	140.0	94.2	111.4	178.4	230.2
1996	152.2	93.2	115.1	217.1	335.5
1997	115.7	76.1	89.6	154.6	204.3
1998	287.9	209.2	241.1	362.1	489.9
1999	384.7	294.5	326.5	472.2	610.1
2000	235.5	134.9	166.7	325.5	501.9
2001	563.4	411.4	467.8	693.0	933.6
2002	282.9	194.4	233.6	369.4	531.7
2003	163.3	94.7	116.9	229.4	362.5
2004	282.4	166.2	211.2	399.1	569.0
2005	135.4	98.7	113.4	172.6	224.1
2006	168.6	121.2	141.2	225.1	294.1
2007	110.7	72.2	86.4	149.0	214.7
2008	124.6	74.7	93.8	182.2	282.1
2009	160.1	103.4	124.9	220.6	328.9
2010	154.6	103.0	124.8	201.8	294.8
2011	223.3	164.9	190.2	282.3	352.2
2012	116.9	87.1	99.0	153.9	221.2
2013	71.8	49.9	57.9	97.3	136.7
2014	66.5	41.5	49.7	99.0	157.8
2015	99.5	61.2	74.8	132.9	208.4
2016	74.4	47.8	57.1	103.7	151.1

Table B-7: Predicted catches (in thousand individuals) and uncertainty from key quantiles of the posterior distribution of predicted catches for the longline target fleet

Year	Median	10 th	25 th	75 th	90 th
1995	0.1	0.1	0.1	0.2	0.3
1996	0.8	0.4	0.6	1.1	1.6
1997	0.4	0.2	0.3	0.5	0.6
1998	2.1	1.3	1.6	2.7	3.4
1999	4.8	3.5	4.1	5.8	7.0
2000	2.0	1.2	1.5	2.6	3.3
2001	2.0	1.4	1.6	2.3	2.8
2002	2.0	1.5	1.8	2.4	2.6
2003	1.2	0.9	1.1	1.5	1.7
2004	0.9	0.7	0.8	1.0	1.2
2005	1.6	1.2	1.4	1.8	2.0
2006	1.0	0.8	0.9	1.2	1.3
2007	1.1	0.8	0.9	1.2	1.4
2008	0.7	0.5	0.6	0.9	1.0
2009	1.9	1.4	1.7	2.3	2.7
2010	9.0	6.5	7.5	10.4	12.4
2011	1.4	1.1	1.2	1.7	1.9
2012	1.1	0.8	1.0	1.3	1.4
2013	0.5	0.4	0.5	0.6	0.7
2014	1.6	1.2	1.4	1.8	2.1
2015	2.0	1.5	1.7	2.2	2.5
2016	0.1	0.1	0.1	0.1	0.1

Table B-8: Predicted catches (in thousand individuals) and uncertainty from key quantiles of the posterior distribution of predicted catches for the purse seine fleets

Year	Median	10 th	25 th	75 th	90 th
Associated sets					
1995	0.4	0.1	0.2	0.7	1.3
1996	27.6	16.7	20.5	35.9	52.3
1997	10.3	5.3	7.1	17.2	32.7
1998	14.0	6.5	8.8	23.8	45.3
1999	14.6	7.6	10.0	22.9	42.4
2000	4.5	2.3	3.0	7.8	14.4
2001	7.6	4.0	5.3	12.2	18.1
2002	8.3	5.1	6.4	11.5	15.8
2003	18.2	10.9	13.6	25.8	36.6
2004	6.6	4.1	5.0	9.1	14.7
2005	2.2	1.3	1.7	3.2	4.8
2006	1.5	0.9	1.1	2.2	3.6
2007	1.6	0.9	1.2	2.4	4.0
2008	1.3	0.8	1.0	1.8	2.7
2009	3.1	1.9	2.4	4.5	6.1
2010	1.1	0.8	1.0	1.4	2.0
2011	1.0	0.7	0.9	1.3	1.8
2012	2.0	1.5	1.7	2.3	2.7
2013	0.7	0.5	0.5	0.9	1.4
2014	0.7	0.5	0.6	0.8	1.0
2015	0.8	0.6	0.7	1.1	1.8
2016	0.8	0.5	0.6	1.2	2.5
Unassociated sets					
1995	0.1	0.0	0.1	0.2	0.3
1996	7.5	4.3	5.6	10.1	14.4
1997	1.6	0.8	1.1	2.5	4.7
1998	2.4	1.2	1.6	4.1	7.2
1999	0.9	0.4	0.6	1.6	3.5
2000	0.6	0.3	0.4	1.1	2.0
2001	1.4	0.7	1.0	2.5	4.5
2002	0.9	0.5	0.6	1.4	2.4
2003	1.6	0.9	1.2	2.4	3.9
2004	1.0	0.6	0.7	1.3	1.9
2005	0.5	0.3	0.3	0.7	1.0
2006	0.2	0.1	0.2	0.3	0.5
2007	0.3	0.2	0.2	0.5	0.8
2008	0.3	0.2	0.2	0.4	0.5
2009	0.5	0.3	0.4	0.7	1.0
2010	0.6	0.4	0.5	0.7	1.0
2011	0.3	0.2	0.3	0.4	0.6
2012	0.9	0.7	0.8	1.1	1.2
2013	0.3	0.2	0.2	0.4	0.8
2014	0.3	0.2	0.2	0.4	0.5
2015	0.3	0.2	0.3	0.4	0.7
2016	0.4	0.2	0.3	0.6	1.3

Table B-9: Predicted catches (in thousand individuals) and uncertainty from the 10th and 90th quantiles for the historical catch reconstruction based on global fin trade statistics, apportioned to the WCPO using three different methods

Year	Effort			Tuna catch			Area		
	Median	10 th	90 th	Median	10 th	90 th	Median	10 th	90 th
1995	123.1	75.1	183.4	198.8	121.2	296.2	144.0	87.8	214.5
1996	119.1	74.4	173.2	210.5	131.5	306.1	154.7	96.6	225.0
1997	128.0	79.9	186.1	229.9	143.6	334.3	166.9	104.2	242.7
1998	133.1	83.1	193.5	266.2	166.2	387.0	178.1	111.2	259.0
1999	174.7	109.1	254.0	285.3	178.2	414.9	199.6	124.7	290.3
2000	226.7	141.6	329.7	342.4	213.8	497.9	232.7	145.3	338.4
2001	315.0	189.8	483.3	409.0	246.4	627.4	285.1	171.8	437.3
2002	318.0	191.6	487.9	427.9	257.8	656.4	288.5	173.8	442.6
2003	346.0	208.5	530.8	438.8	264.4	673.2	308.3	185.8	473.0
2004	313.1	188.6	480.2	394.2	237.5	604.8	272.1	163.9	417.4
2005	280.8	169.2	430.8	375.6	226.3	576.2	260.8	157.1	400.1
2006	266.0	160.3	408.0	347.9	209.6	533.7	236.4	142.5	362.7
2007	367.7	221.4	562.0	512.7	308.7	783.5	315.0	189.6	481.3
2008	399.0	240.2	609.7	471.6	283.9	720.7	300.7	181.0	459.6
2009	412.6	248.4	630.6	465.0	279.9	710.6	302.0	181.8	461.5
2010	430.2	259.0	657.4	489.7	294.8	748.4	312.8	188.3	478.1
2011	477.4	287.4	729.6	501.2	301.8	766.0	335.4	201.9	512.6
2012	486.7	293.0	743.7	513.8	309.3	785.2	341.2	205.4	521.5
2013	430.6	259.2	658.1	508.1	305.9	776.6	336.8	202.8	514.8
2014	459.4	276.6	702.1	503.0	302.8	768.7	326.7	196.7	499.3
2015	482.0	290.2	736.6	471.2	283.7	720.1	317.7	191.3	485.6
2016	497.3	299.4	759.9	485.1	292.1	741.4	327.8	197.3	500.9