SCIENTIFIC COMMITTEE
TENTH REGULAR SESSION

Majuro, Republic of the Marshall Islands
6-14 August 2014

Analysis of tagging data for the 2014 tropical tuna assessments: data quality rules, tagger effects, and reporting rates

WCPFC-SC10-2014/SA-IP-06

A M Berger[1], S McKechnie[1], F Abascal[1], B Kumasi[2], T Usu[2], S J Nicol[1]

[1]Oceanic Fisheries Programme (OFP)
Secretariat of the Pacific Community, Noumea, New Caledonia

[2]National Fisheries Authority
Papua New Guinea

Analysis of tagging data for the 2014 tropical tuna assessments: data quality rules, tagger effects, and reporting rates

A M Berger, S McKechnie, F Abascal, B Kumasi, T Usu, S J Nicol

Oceanic Fisheries Programme, SPC
National Fisheries Authority, Papua New Guinea

Executive Summary

Tuna tagging in the western and central Pacific Ocean has occurred periodically over the past 35 years, which has resulted in nearly 800,000 tag release and over 100,000 tag recovery records catalogued. Tagging has been carried out predominantly through pole-and-line fishing events from specifically chartered vessels, where fish are captured, tagged with a conventional dart tag in a fish cradle, and released back into the water in just a few seconds. Information is recorded at the time of tagging such as date, location, fish size, and species. Tag release events used in the 2014 tropical tuna base assessments were associated with the Pacific Tuna Tagging Programme (PTTP; 2006–2012), the Regional Tuna Tagging Project (RTTP; 1989 – 1992), the Skipjack Survey and Assessment Programme (SSAP; 1977 – 1982), the Commonwealth Scientific and Industrial Research Organisation bigeye tagging project (CSIRO; 1995 – 2001), and the Japanese skipjack tagging programme (JP; 1988 – 2012).

Information on critical demographic and fishery rates can be obtained from tagging experiments, which can help inform estimates generated in the MULTIFAN-CL population assessment model. These experiments provide information on tuna survival, mortality components (e.g., fishing, natural, and handling), population abundance, growth, and movement. However, such information can also be misleading and contribute to biased results and ill-informed management decisions if not properly considered. As such, three different analyses were conducted in an attempt to maximize the representativeness and potential accuracy of population-level inferences from tagging experiments used in the 2014 stock assessments. These included:

1. correcting releases for recovery data quality,
2. estimating tagger and other effects to supplement release corrections for base levels of tag shedding and tag-related mortality, and
3. estimating reporting rates (prior distributions and penalty weights) to account for under-reporting of tag recoveries.

The above analyses and data corrections were performed for each species (bigeye, skipjack, and yellowfin) and tagging programme where data were available (Table 1). A brief synopsis of each follows.

1. Data extractions and corrections for recovery data quality

Missing data fields, incorrectly specified recovery data, or recoveries outside of model regions can result in physically recovered tags being unusable for MULTIFAN-CL stock assessments. This presents a problem preserving observed recovery-release ratios that are critical for informing estimates of fishing mortality and population abundance. Thus, the number of releases was adjusted to account for unusable recoveries to preserve population inferences. In order for recoveries to be considered 'usable' in MULTIFAN-CL stock assessments, information on tag recovery date (year and month), location (latitude and longitude), fleet (gear and flag), and tag number must be available. Overall, forty-two percent of all tag recoveries were deemed not usable as a result of missing one or more entries of key information. Release corrections were applied at the individual length bin level for each tagging event and species. Corrections were applied to the tagging data prior to input into the assessment model.

2. Tagger effects and correction factors

Individual tagger experience and skill can have an influence on how well tags are implanted into fish and the ability of fish to properly recover from the tagging process (catch, handle, tag, and release). The probability of fish being exposed to conditions that relate to tag shedding (e.g., imprecise tag insertion) and tag-related mortality (e.g., less efficient tagging and handling techniques) would be expected to be lower for more experienced taggers than for taggers with less experience. Thus, statistical models were used to correct the tagging data for potential biases associated with individual tagger effects before these data were used to inform the stock assessment.

Tagger effects were only estimated for tag release events associated with the PTTP and RTTP. Generalized additive models (GAMs) and generalized linear models (GLMs) were used to evaluate which explanatory variables significantly influenced recovery rates and estimate the mean effect size (change in recovery rate) resulting from the particular set of observed conditions relative to optimal conditions. Explanatory variables included tagging *event*, *tagger*, *fish length*, *fish condition*, *tagging quality*, and *tagging station*. A single best model was selected for identifying significant variables, making statistical inferences, and developing correction factors.

Corrections adjusted the number of releases downwards to account for 1) base levels of tag shedding and tag-related mortality (that which would occur even under 'optimal' tagging conditions) and 2) additional levels as a result of the particular tagging conditions present at each tagging event. Corrections do not account for all tagging induced mortality and tag shedding, but it does remove the effects of conditions that can be measured and controlled for (e.g. taggers with low skill levels, fish that were released in suboptimal condition, etc.), thereby improving resulting demographic estimates. Estimated correction factors (Figure 1) resulted in the number of releases being lower on average than the total number of releases. This has the net effect of supplying the stock assessment model with information suggesting higher fishing mortality rates and lower biomass levels compared to using uncorrected tagging data. Revised estimates should better reflect true fishing mortalities and biomasses as a result of improved separation of key mortality components: fishing-related mortality, tagging-related mortality, and natural mortality. Corrections were applied to the tagging data prior to input into the assessment model.

3. Reporting rates

Tag return rates that are used to aid the estimation of exploitation rates and population biomass in MULTIFAN-CL assessment models must be corrected to account for the number of tagged fish that are recovered but not reported. Estimates of tag reporting rates are used to correct (or adjust) tag return rates to avoid systematically under estimating fishing mortality rates and over estimating fish stock biomass. These corrections need to also take into account that some tagged fish are recovered but not identified as such, due to tag shedding, and the mortality that occurs as a result of the physical tagging process. Tag seeding experiments were analysed to produce reporting rate prior distributions by gear, flag, tagging programme and species. Correcting for reporting rates occurred during assessment model fitting and optimization procedures through the use of the informative prior distributions. In the absence of other sources of information, reporting rates act as a multiplier on stock biomass and thus influence estimates of fishing mortality.

The number of releases from seeding trials has nearly tripled (from 1,156 to 3,368) since the 2011 assessments and now broadly covers the spatial extent of the main purse seine fleets in the WCPFC-CA (Table 5). More data and regional coverage has improved estimates, particularly for the Korean, Chinese-Taipei and Japanese fleets, which together account for about 40% of the total purse seine catch in the WCPFC convention area. Estimates for the Japanese fleet, while improved, were still based on very few trials and were inconsistent in some cases with empirical data. Therefore, the Japanese fleet was allocated a reporting rate equal to that for Chinese-Taipei, with a standard error equal to the maximum among all the flags. Reporting rates for Indonesian and Spanish fleets were set to that estimated for the Philippines and Ecuador, respectively, given the similarities in fishing grounds and offloading ports as a result of insufficient data for those fleets.

Tag reporting rates for the PTTP (Table 4) were modelled by using a generalized linear model (GLM) with vessel flag and tag type as explanatory variables. Point (or central tendency) estimates for each flag were extracted directly from the GLM. Prior distributions were estimated using Monte Carlo simulations (Figure 2 and 3), from which variances were estimated. Resulting distributions were then averaged for each region weighted by the percentage of the catch by each flag and input into the assessment models. Tag reporting rates for the RTTP (Table 4) remained the same as those previously estimated. For the Coral Sea tagging programme, the tag reporting prior for the Australian longline fleet was based on expert knowledge of the fishery. Tag reporting rate priors were not estimated for other fisheries and programmes (Skipjack Survey and Assessment Programme and the Japanese Tagging Programme), due to a lack of information to conduct suitable analyses. In such cases, uninformative priors were specified in the assessment models.

The current estimates indicate a reporting rate of approximately 10% below previous estimates, on average. Due to the increase in the number of releases, the priors in the current assessments are also significantly more informative (i.e., tighter distribution), as reflected by the high penalties in the regions where the seeding trials have provided more precise reporting rate estimates for the main fishing fleets (Table 4).

The process of analysing and adjusting tagging data for use in MULTIFAN-CL tuna stock assessments is imperative to maintain the reliability of model results. Several important changes have been made to this process from previous assessments that warrant recognition. First, release numbers were adjusted on data aggregated at the finest scale possible to account for unusable recoveries to preserve population inferences (Section 1). Second, tag shedding and tag-related mortality (combined base levels and additions from tagger-effects) were corrected for by adjusting tag releases downward by an estimated correction factor (Section 2) rather than incorporating them into the reporting rate prior distribution (i.e., the loss of these tags from the tagged population treated as tags not reported). Third, sample sizes from tag seeding experiments have increased, allowing improved precision of tag reporting rate estimates (Section 3). The exploration of candidate models using redefined regions in the skipjack assessment (from 3 to 5 regions) and the yellowfin and bigeye assessments (from 6 to 9 regions) also signify a significant change in preparing tagging data from previous years.

Introduction

Information on critical demographic and fishery rates can be obtained from tagging (release-recovery) experiments, which can help inform integrated statistical catch-at-age population dynamic models such as those used in stock assessment (e.g., MULTIFAN-CL; Fournier et al. 1998; Maunder and Punt 2013). These experiments, depending on the experimental design, can provide information on survival, mortality components (e.g., fishing, natural, and discard), population abundance, movement, and growth (Hilborn 1990; Sibert et al. 1999). However, estimates of "nuisance" rates such as tag shedding, tag-related mortality and tag reporting are also needed to alleviate potential biases that may be present in the tagging data when used in population models.

Tags that are shed (detached or expelled from the fish) are no longer susceptible to recovery and thus can incorrectly inform mortality rates when not accounted for in the stock assessment. There are many factors associated with the physical tagging process that can be a proximate cause of undesired short-term delayed mortality of released fish. The tagging process can increase susceptibility to disease from wound infection and capture stress, predation from lethargy, and various other behavioural stress reactions that can affect post-release behaviour and survival (Skomal 2007). Recovered tags that go unreported, either deliberately or unintentionally are treated as tags that remain at liberty. Thus, adjustments need to be made to the data (or in the population model) to correct for non-reporting so that 'true' recovery-to-release ratios are maintained.

Tuna tagging in the western and central Pacific Ocean has occurred periodically over the past 35 years, and has resulted in nearly 800,000 tag release and over 100,000 tag recovery records catalogued. Tagging has been carried out predominantly through pole-and-line fishing events from specifically chartered vessels, where fish are captured, tagged with a conventional dart tag in a fish cradle, and released back into the water in just a few seconds. Information is recorded at the time of tagging such as date, location, fish size, and species. Precautions have been taken (e.g., established tag mixing periods and variance inflation) and continue to be taken (e.g., newly redefined regional structure) in an attempt to best meet key assumptions associated with release-recovery tagging experiments, such as the assumption for equal probability of recovery (Kolody and Hoyle 2014) and independence (Hampton and Fournier 2001).

Tagging data preparation and analyses

Tagging experiments can provide significant information about the dynamics of harvested fish populations that would otherwise be unavailable or less precise. However, such information can also be misleading and contribute to biased results and ill-informed management decisions if not properly considered. As such, three different analyses were conducted in an attempt to maximize the representativeness and potential accuracy of population-level inferences from tagging experiments used in the 2014 bigeye, skipjack, and yellowfin stock assessment. These included:

1. correcting releases for recovery data quality,
2. estimating tagger effects to supplement release corrections for base levels of tag shedding and tag-related mortality, and

3. estimating reporting rates (prior distributions and penalty weights) to account for under-reporting of tag recoveries.

Analyses 1 and 2 were conducted as part of the data preparation phase, where corrections were applied directly to the raw release data, aggregated by tagging event and length bin, prior to use in the assessment model. Corrections were applied to the number of released tags to account for 'unusable' recoveries as some recovered tags were missing key pieces of supporting information (see Section 1 below). Releases were also corrected for base levels of tag shedding (rate = 0.059 from Hampton et al. 1997), base levels of tag-related mortality (rate = 0.07 from SPC OFP), and a 'tagger effect' component (*sensu* Hoyle et al. 2014) to capture any additional tag shedding or mortality arising from circumstances at the time of tagging (see Section 2 below). Analysis 3 was conducted outside the assessment model (see Section 3 below), but results were used as input prior information or fixed values in the assessment model. Tagging analyses and data corrections were performed for each species and tagging programme where data were available (Table 1).

Table 1. Overview of tagging data corrections and analyses conducted for each species and tagging programme. Missing fields or incorrectly specified recovery data can result in physically recovered tags being unusable in MULTIFAN-CL stock assessments.

Species	Programme	Data Correction Analyses
Bigeye	PTTP	Recovery data quality, tagging mortality, shed tags, tagger effects, reporting rate
	RTTP	Recovery data quality, tagging mortality, shed tags, reporting rate
	CSIRO	
Skipjack	PTTP	Recovery data quality, tagging mortality, shed tags, tagger effects, reporting rate
	RTTP	Recovery data quality, tagging mortality, shed tags, tagger effects, reporting rate
	SSAP	Recovery data quality, tagging mortality, shed tags
	JP*	
Yellowfin	PTTP	Recovery data quality, tagging mortality, shed tags, tagger effects, reporting rate
	RTTP	Recovery data quality, tagging mortality, shed tags, tagger effects, reporting rate
	CSIRO	

* Refer to WCPFC-SC10-2014/SA-WP-05 for details.

1. Data extractions and corrections for recovery data quality

Tag release and recovery information were extracted from databases held and managed by the Secretariat of the Pacific Community on April 28th, 2014, or were obtained with permission from cooperating WCPFC member nations. Tag release events used in the 2014 tropical tuna base assessments were associated with the Pacific Tuna Tagging Programme (PTTP; 2006–2012), the Regional Tuna Tagging Project (RTTP; 1989 – 1992), the Skipjack Survey and Assessment Programme (SSAP; 1977 – 1982), the Commonwealth Scientific and Industrial Research Organisation bigeye tagging project (CSIRO; 1995 – 2001), and the Japanese skipjack tagging programme (JP; 1988 – 2012). The use of tagging data from the Japanese tagging program is not discussed further here; refer to WCPFC-SC10-2014/SA-WP-05 for more details.

Tagged fish that were missing key release information (date, latitude, longitude, species, tag identifier, fish length) were removed from the tagging dataset (<1% of all records). Events that contained less than 10 releases in any one time period (quarter/year) and assessment region were removed due to the propensity for unidentifiable parameters or spurious estimates with low sample sizes. The resulting total number of tagging release events and tags released varied by species and tagging programme (Table 2).

Missing fields, incorrectly specified recovery data, or recoveries outside of model regions can result in physically recovered tags being unusable for MULTIFAN-CL stock assessments. This presents a problem preserving observed recovery-release ratios that are critical for informing estimates of fishing mortality and population abundance. Thus, the number of releases was adjusted to account for unusable recoveries to preserve population inferences. In order for recoveries to be considered 'usable' in MULTIFAN-CL stock assessments, information on tag recovery date (year and quarter), location (MULTIFAN-CL region), fleet (gear and flag), and tag number must be available. If any one piece of information was missing or could not be interpolated with high confidence, the tag could not be assigned to the appropriate assessment region, year, quarter, fishing fleet, and length bin in the stock assessment. Release length bins (2 cm intervals) coincided with those used for length-frequency data in MULTIFAN-CL: range for bigeye = 10 – 200cm; skipjack = 2 – 110cm; yellowfin = 10 – 200cm. Recoveries outside of assessment model regions (e.g., bigeye in the Eastern Pacific Ocean) were considered unusable at this time, and releases were also adjusted to account for these tags.

For each release event (e), the number of releases in each length bin (b) was adjusted to account for unusable recoveries. Correction factors ($CF_{e,b}^{Rel}$) were calculated according to one of the following three scenarios:

$$CF_{e,b}^{Rel} = \begin{cases} Rec_{e,b}^{T} > 1 \; and \; Rec_{e,b}^{U} > 1 & \dfrac{Rec_{e,b}^{U}}{Rec_{e,b}^{T}} \\ Rec_{e,b}^{T} > 1 \; and \; Rec_{e,b}^{U} = 0 & median\left(\dfrac{Rec_{e}^{U}}{Rec_{e}^{T}}\right) \\ Rec_{e,b}^{T} = 0 \; and \; Rec_{e,b}^{U} = 0 & 1, \end{cases}$$

where Rec^{T} was the total recoveries and Rec^{U} was the usable recoveries. These correction factors were used along with corrections for base levels of tag shedding, base levels of tag-related mortality, and tagger effects (see Section 2) to correct the overall number of tag releases for use in MULTIFAN-CL stock assessments.

Table 2. Summary of tag releases and recoveries by tagging programme and species.

Programme	Total Releases	Proportions Bigeye	Skipjack	Yellowfin	Total Recoveries	Usable Recoveries	Adjusted Releases
PTTP	356,132	0.09	0.63	0.28	48,164	24,058	177,891
RTTP	140,605	0.06	0.66	0.28	15,496	11,817	107,220
SSAP	83,905	0	1	0	4,585	3,426	62,693
CSIRO*	974	0.74	0.26	0	91	91	974

* Note that correction factors were not calculated for CSIRO releases.

2. Tagger effects and correction factors

Individual tagger experience and skill can have an influence on how well tags are implanted into fish and the ability of fish to properly recover from the tagging process (catch, handle, tag, and release). The probability of fish being exposed to conditions that relate to tag shedding (e.g., imprecise tag insertion) and tag-related mortality (e.g., less efficient tagging and handling techniques) would be expected to be lower for more experienced taggers than for taggers with less experience. This occurrence results in the probability of a tag being recovered being co-dependent upon the individual who tagged the fish, which is not a desirable feature when recovery rates are used to inform population parameters such as fishing mortality and biomass in the stock assessment model. Thus, statistical models were used to correct the tagging data for potential biases associated with individual tagger effects before these data were used to inform the stock assessment.

Methods

Analyses were conducted separately for each species (bigeye, skipjack, and yellowfin) because of inherent differences in physiology and behaviour that would likely result in different tagger effect estimates and correction factors (*sensu* Hoyle *et al.* 2014). Only PTTP release events that occurred before January 2013 were included in these analyses as recovery rates from more recent events have the potential to be biased low due to recovery data time lags associated with reporting and data quality control.

Each species-specific dataset was filtered to ensure stability with fitted statistical models while maintaining the maximum amount of tagging events possible. Incomplete records, small sample sizes, and extreme outliers were removed to further promote model stabilization and interpretation of the results. For example, tagging events that did not meet the minimum threshold for number of releases (PTTP and RTTP: SKJ=30, YFT=20, BET=15) were removed from consideration. Individual taggers that did not meet the minimum threshold for number of fish tagged and released (PTTP: SKJ=200, YFT=200, BET=100; RTTP: SKJ=200, YFT=100, BET=30) were also removed. Levels of other categorical variables that were hypothesized to influence tagger effects (e.g., tag quality, fish condition, and tagging station) were removed if the threshold release numbers were not met (PTTP: SKJ=200, YFT=200, BET=100; RTTP: SKJ=200, YFT=100, BET=100). The proportion of releases remaining after filtering was generally high (88-97%) with the only exception being for PTTP bigeye where the low number of fish tagged for many tagging events resulted in 68% of releases remaining.

Tagger effects

Tagger effects were only estimated for tag release events associated with the PTTP and RTTP. Other tagging programs were not evaluated for lack of sufficient or contrasting supporting data (i.e., tagger name, fish condition, tagging location, etc.) to model hypothesized effects. For example, a few release events that occurred in the Central Pacific that were undertaken by the IATTC where all tagging was undertaken by two taggers who were not present in any other tagging events. In this case, there was insufficient contrast in the data, raising concerns about the reliability of estimated coefficients, so they were excluded from the analysis.

Generalized additive models (GAMs) and generalized linear models (GLMs) were used to evaluate which explanatory variables significantly influenced recovery rates and to estimate the mean effect size (change in recovery rate) resulting from the particular set of observed conditions. This approach allowed for explicit quantification of how the recovery rate would be expected to deviate, on average, as a result of the observed set of tagging conditions relative to a base set of conditions. Base conditions were set to mimic optimal conditions (e.g., fish tagged by a very experienced tagger, fish condition was good, etc.). The set of explanatory variables considered included: *Event*: the temporally and spatially unique tagging event during which the fish was tagged and released – such events essentially were equivalent to discrete tagging episodes on individual schools of tuna; *Tagger*: the individual who tagged the fish; *Length*: the fork length of fish measured in the cradle before release; *Condition*: the overall health or condition of the fish upon release categorized by 'good', 'bleeding', 'dropped on deck', 'eye damage', 'hit side of boat', 'long time on hook', 'mouth damage', 'shark bite' and 'tail damage'; *Quality*: the quality of tag placement categorized by 'good', 'badly placed' and 'too slow'; and *Station*: the location of the tagging station on the tagging vessel categorized by 'port bow', 'starboard bow', 'midships', and 'stern' (Table 3). Overall, the models used were similar to those of Hoyle *et al.* (2014); except that several explanatory variables used in their analysis were excluded in this analysis (e.g. tag type) because either those authors found them to be not significant or as a result of unbalanced data recording issues.

For the PTTP, the response variable (y_i) was binary (1 – recovered, 0 - not recovered for fish *i*) and the binomial GAM with the full set of explanatory variables was given by

$$y_i \sim \text{Bernoulli}(p_i)$$

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{Event,[i]} + f\left(Length_{[i]}\right) + \beta_{Tagger,[i]} + \beta_{Condition,[i]} + \beta_{Quality,[i]} + \beta_{Station,[i]},$$

where the $\beta_{x,[i]}$ indicates the coefficient for each named variable (x) at the observed level for fish *i*. The $f(Length_{[i]})$ term is a tensor product smooth function of the length of fish *i*. Due to the large number of skipjack releases, the amount of computing power necessary to fit the GAM was prohibitive, and so an equivalent GLM with a near identical structure to the GAM model was fitted. The sole difference was that instead of applying the tensor product smooth function a natural cubic spline with 4 degrees of freedom (as determined by comparing models that allowed alternative degrees of freedom using AIC) was used to fit length using the *ns* and *glm* functions in program R (2013).

In a similar fashion, a binomial GAM was used for the RTTP release events. The only exception being that the data were aggregated at the level of the individual explanatory variables, and so the number of successes (y_i recoveries) from the number of trials (n_i releases) was modelled as an indicator variable referencing groups of fish (rather than individual fish as with the PTTP due to data limitations). The full model was

$$y_i \sim \text{Binomial}(n_i, p_i)$$

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{Event,[i]} + f(Length_{[i]}) + \beta_{Tagger,[i]} + \beta_{Condition,[i]} + \beta_{Quality,[i]} + \beta_{Station,[i]},$$

where *i* now subscripts an aggregated group of fish for each unique set of categorical levels across the explanatory variables in the model.

Model selection for each of the six unique datasets (three species for each of the PTTP and RTTP) was undertaken using backwards elimination on the basis of AIC. Only the best model was used for identifying significant variables, making statistical inferences, and developing correction factors.

Table 3. Summary of tagging data attributes used in tagger effect models.

Variable	(type)	Metric/Level	PTTP BET	PTTP SKJ	PTTP YFT	RTTP BET	RTTP SKJ	RTTP YFT
Event	(numbers)	Count	147	888	719	67	561	356
Taggers	(numbers)	Count	18	39	29	15	21	23
Length	(cm)	Range	25-115	25-77	27-118	20-135	20-80	20-140
Quality	(proportions)	Good	>0.99	>0.99	>0.99	0.99	0.97	0.98
		Badly placed	<0.01	<0.01	<0.01	<0.01	0.03	0.02
		Too slow	<0.01	-	-	<0.01	-	<0.01
Condition	(proportions)	Good	0.89	0.95	0.94	0.94	0.93	0.93
		Bleeding	0.03	<0.01	<0.01	0.02	0.02	0.03
		Dropped	0.01	0.02	0.03	-	0.02	0.02
		Mouth damage	0.04	<0.01	0.01	-	<0.01	<0.01
		Shark bite	0.01	0.01	<0.01	0.03	0.02	0.01
		Eye damage	<0.01	-	<0.01	-	-	-
		Hit side boat	-	-	-	-	<0.01	-
Station	(proportions)	Port bow	0.08	0.37	0.38	0.21	0.54	0.44
		Mid-ship	0.08	<0.01	<0.01	0.25	<0.01	0.05
		Starboard bow	0.25	0.33	0.30	-	-	-
		Stern	0.59	0.29	0.32	0.54	0.45	0.52

Correction factors

Correction factors were developed for each species and tagging program (RTTP and PTTP) combination from the best tagger effect model results. These corrections adjust the number of releases downwards to account for 1) base levels of tag shedding and tag-related mortality (that which would occur even under ‘optimal’ tagging conditions) and 2) additional levels as a result of the particular tagging conditions present at each tagging event. Base level corrections are required to satisfy the assumption that survival of tagged and untagged fish remains equivalent. Corrections do not account for all tagging induced mortality and tag shedding, but it does remove the effects of conditions that can be measured and controlled for (e.g. taggers with low skill levels, fish that were released in suboptimal condition, etc.), thereby improving resulting demographic estimates.

Using the tagging events that were modelled, correction factors were estimated by first obtaining the fitted values for each tagged fish on the nominal scale (i.e., the probability of recovery given the observed values of each explanatory variable) for each dataset (species and tagging programme). Next, the predicted probability of recovery for each tagged individual under ‘optimal’ conditions was calculated by applying the modelled coefficients at the specified ‘optimal’ level for each factor. For example, the fitted values for skipjack in the PTTP would be calculated as

$$\mu_i^{fit} = logit^{-1}\left(\beta_0 + \beta_{Event,[i]} + f(Length_{[i]}) + \beta_{Tagger,[i]} + \beta_{Condition,[i]} + \beta_{Quality,[i]} + \beta_{Station,[i]}\right),$$

where the value of each $\beta_{x,[i]}$ was the coefficient for the observed factor level for fish *i*, and $f(Length_{[i]})$ was the value of the smoothing function at the observed length of fish *i*. The prediction under 'optimal' conditions would be calculated as

$$\mu_i^{opt} = logit^{-1}(\beta_0 + \beta_{Event,[i]} + f(Length_{[i]}) + \beta_{Tagger,["BML"]} + \beta_{Condition,["Good"]} + \beta_{Quality,["Good"]} + \beta_{Station,["\text{Port bow}"]}),$$

where $\beta_{Tagger,["BML"]}$, $\beta_{Condition,["Good"]}$, $\beta_{Quality,["Good"]}$ and $\beta_{Station,["\text{Port bow}"]}$ were the 'optimal' condition coefficients. The coefficient for *Event* was for the observed event for tagged fish *i*, and was not set to an 'optimal' condition in order to preserve the influence of this variable because it is related to spatial fishing dynamics which are explicitly dealt with in the stock assessment model proper. Similarly, the length of observed tagged fish ($Length_{[i]}$) was also preserved because the effect of fish length on recovery rates is characterized in the stock assessment model itself through selectivity parameters.

Correction factor calculations were carried out efficiently in program R (2013) using the *predict.glm* function. The correction factor for event *j* can then be calculated as the mean ratio of the predictions for the observed and 'optimal' conditions,

$$r_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\frac{\mu_{j,i}^{fit}}{\mu_{j,i}^{opt}},$$

where n_j is the number of tagged fish released in during release event *j*. The total number of releases for event *j* was then adjusted by the correction factor (r_j) for event *j*.

For events that were not modelled (generally due to sample sizes), releases were adjusted in a similar manner by replacing missing or unidentifiable coefficients with the median coefficient value across all other factor levels. In cases where a particular tagger was excluded because they didn't meet the minimum sample size (number of fish tagged) threshold, for example, that tagger was assigned the median coefficient across all taggers that did meet the minimum tagging threshold.

Results

The full model with all explanatory variables retained was selected as the best model for skipjack and yellowfin in the PTTP, while for bigeye the best model retained all variables except *Station*. For the RTTP, model selection procedures tended to pick more simplified models; the exception being skipjack where the full model was identified as the best fit. The best yellowfin model for this programme retained all variables except *Quality*, and the best bigeye model retained only the variable *Event*.

Estimated correction factors (Figure 1) resulted in the number of releases being lower on average than the total number of releases. This has the net effect of supplying the stock assessment model with information suggesting higher fishing mortality rates and lower biomass levels compared to using uncorrected tagging data. Revised estimates should better reflect true fishing mortalities and biomasses as a result of improved separation of key mortality components: fishing-related mortality, tagging-related mortality, and natural mortality. For the PTTP, the median correction factor was 0.68, 0.76 and 0.73 for bigeye, skipjack, and yellowfin respectively. For the RTTP, the median correction factor was 0.80 for skipjack and 0.76 for yellowfin. There was no available correction factor for bigeye in the RTTP because no correction variables were present in the selected best model (likely a result of sparse bigeye tagging data in the RTTP).

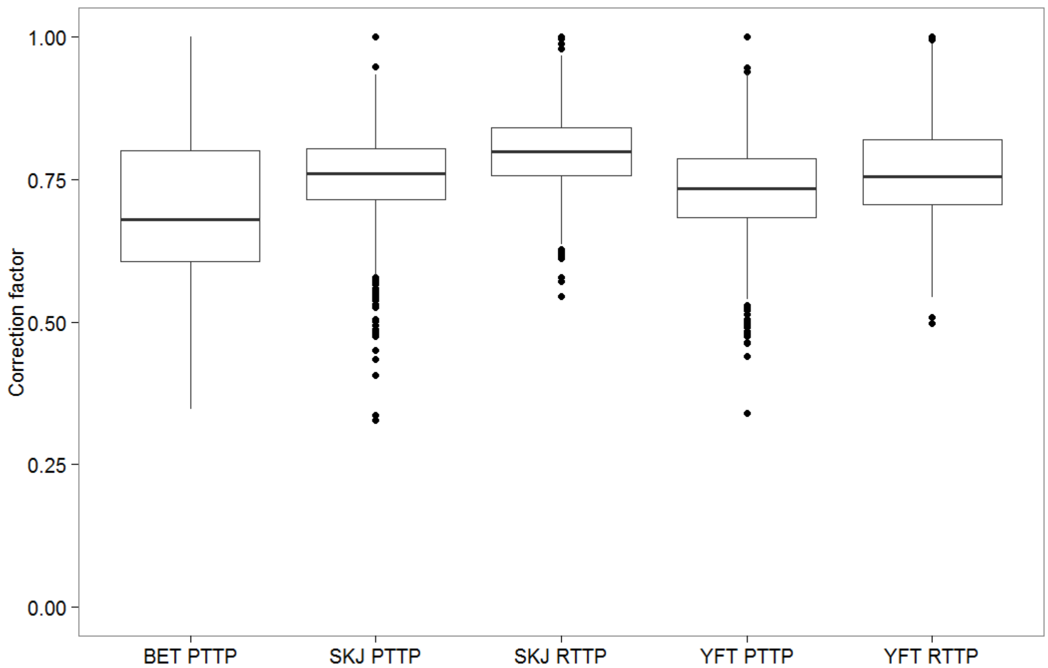

Figure 1. Boxplots showing the range of estimated correction factors that were applied to tag release events to adjust the number of tag releases for the influence of tagger effects on shedding and tag-related mortality. The central bar represents the overall median.

3. Reporting rates

Tag return rates that are used to aid the estimation of exploitation rates and population biomass in MULTIFAN-CL assessment models must be corrected to account for the number of animals that are recovered but not reported. Estimates of tag reporting rates are used to correct (or adjust) tag return rates to avoid systematically under estimating fishing mortality rates and over estimating fish stock biomass. These corrections need to also take into account that some tagged fish are recovered but not identified as such, due to tag shedding, and the mortality that occurs as a result of the physical tagging process.

In the previous tropical tuna assessment in the WCPFC-CA (2011), estimated tag reporting rates included corrections for tagging failure stemming from tagging-related mortality and tag shedding (Hoyle, 2011). For 2014 assessments, it was decided in accordance with the Pre-Assessment Workshop (WCPFC-SC10-2014/SA-IP-07) to correct the number of released tags to account for tagging failure (i.e., adjust the releases downwards; see Section 2 above) in an attempt to minimize the variance associated with the reporting rate estimates themselves (and hence tighten the reporting prior distributions). Excessively wide reporting rate prior distributions effectively give the assessment model leniency to make large model fitting adjustments that may not be related to tag reporting at all, but rather to other model inconsistencies. The approach used here along with more available data helped to alleviate some of the model instability encountered in previous assessments from reporting rate estimates being at or close to parameter boundaries. It also had the advantage of directly linking estimates of tagging failure to the tagging process itself (i.e., tags implanted into fish that die from physically being tagged and/or shed their tag should not be counted as a release available for recovery). The main disadvantage of this approach was that only point estimates of tagging failure were used to adjust tag releases.

There are various ways of estimating tag reporting rates. Some attempts have been carried out using empirical tagging data assuming a 100% reporting rate for a segment of the fleet (e.g. boats with observers on-board) or for high-reward tags (e.g., Brownie et al., 1985; Pollock et al. 2001, 2002). However, these methods rely on several assumptions, including time-invariant reporting, 100% detection, and reporting of all the tagged fish recovered in a component of the fishery. Alternatively, tag seeding experiments allow direct estimation of tag reporting rates for fishery components consistent with the experimental design (e.g. at the gear, flag and tagging program level) by secretly tagging fish once they are caught and then determining the number of those tags that are finally reported.

Methods

Tag reporting rates for the Pacific Tuna Tagging Programme (PTTP; Table 4) were estimated from tag seeding experiments carried out by trained observers aboard purse seiners of different nationalities throughout the WCPO between 2007 and 2012. Two types of tags were seeded in the WCPFC area: the conventional plastic anchor tags and stainless steel anchor tags. Both tag types had similar streamer and tethering; the only difference being the substance of the barbed anchor head. Data from 188 observer trips, with an average of 20.08 (± 7.56) tags seeded per trip, was available for the analyses (Table 5).

Conventional seeded tags were thought to have a lower reporting rate than conventional regularly deployed tags due to a lower retention of seeded tags inserted into dead and often frozen fish. Consequently, reporting rates were only estimated for steel anchor tags, under the assumption (based on experienced taggers advice) that they do not shed.

Reporting rates were modelled in the statistical package R (R Core Team, 2013) by updating the procedure described in Hoyle (2011). Briefly, reporting rates were modelled by using a generalized linear model (GLM) with a quasibinomial error distribution and vessel flag and tag type as explanatory variables. Point estimates for each flag were produced using the predict method for GLM fitted models using the "stats" R package. Variances were interpolated for each flag from reporting rate prior distributions that were estimated through Monte Carlo simulation by sampling from the probability distribution of the flag (Figure 2) and tag (Figure 3) terms. The distributions were then averaged for each region weighted by the percentage of the catch by each flag. The analysis was repeated for bigeye, skipjack, and yellowfin.

Tag reporting rates for the Regional Tuna Tagging Project (RTTP; Table 4) remained the same as those previously estimated (Hampton, 1997). For the Coral Sea tagging programme, the tag reporting prior for the Australian longline fleet in the yellowfin and bigeye assessment model

Table 4. Reporting rate prior distribution parameters for purse seine all fleets per species, region, and tagging program. The penalty term is used in MULTIFAN-CL and is inversely related to the variance of the distribution.

		PTTP		RTTP	
Species	Region	mean	penalty	Mean	penalty
Skipjack	2	0.62	43	0.59	84
	3	0.55	151	0.59	84
	5	0.68	82	0.59	84
Bigeye	3	0.59	89	0.59	84
	4	0.58	33	0.59	84
	7	0.59	89	0.59	84
	8	0.70	163	0.59	84
Yellowfin	3	0.62	70	0.59	84
	4	0.56	157	0.59	84
	7	0.62	70	0.59	84
	8	0.72	157	0.59	84

Table 5. Summary of seeding trials aboard purse seine vessels in the WCPFC-CA during the period 2007-2012 by flag and tag type (S13: steel anchor tag; Y13: plastic anchor tag).

			Releases		Recoveries		Nominal reporting rates		
Flag	% Catch	No. Kits	S13	Y13	S13	Y13	S13	Y13	Average
KR	15	73	474	373	169	169	0.36	0.45	0.40
JP	13.5	4	17	9	5	1	0.29	0.11	0.23
PG	12.5	8	25	90	16	77	0.64	0.86	0.81
US	12.5	68	823	260	533	112	0.65	0.43	0.60
TW	11.7	9	35	87	28	51	0.80	0.59	0.65
PH	10.9	40	302	305	250	234	0.83	0.77	0.80
ID	4	-	-	-	-	-	-	-	-
CN	3.6	6	39	41	20	14	0.51	0.34	0.43
MH	3.5	12	122	107	37	30	0.30	0.28	0.29
VU	2.2	14	118	137	87	89	0.74	0.65	0.69
ES	1.8	-	-	-	-	-	-	-	-
KI	1.6	16	87	109	55	54	0.63	0.50	0.56
NZ	1.6	3	-	58	-	1	-	0.02	0.02
FM	1.4	5	29	9	7	5	0.24	0.56	0.32
VN	1.3	-	-	-	-	-	-	-	-
SB	1.1	8	28	27	28	26	1.00	0.96	0.98
EC	1	2	10	15	10	11	1.00	0.73	0.84
SV	0.6	6	14	15	0	0	0.00	0.00	0.00
TV	0.3	-	-	-	-	-	-	-	-
Total	100	274	2123	1642	1245	874	0.59	0.53	0.56

regions 5 and 9 was set at 0.8 (penalty set to 50) based on expert knowledge of the fishery. Tag reporting rate priors were not estimated for other fisheries and programmes (Skipjack Survey and Assessment Programme and the Japanese Tagging Programme), due to a lack of information to conduct suitable analyses. In such cases, uninformative priors were specified in the assessment models.

Results

Since the previous tropical tuna assessments (2011), the number of releases from seeding trials has nearly tripled (from 1,156 to 3,368) and now broadly covers the spatial extent of the main purse seine fleets in the WCPFC-CA (Table 5). In the absence of other sources of information, reporting rates act as a multiplier on stock biomass and thus influence estimates of fishing mortality. Hence, it is important to have a good understanding of tag reporting to have an accurate evaluation of stock status. More data and better regional coverage have improved estimates, particularly for the Korean, Chinese-Taipei and Japanese fleets, which together account for about 40% of the total purse seine catch in the WCPFC convention area. Estimates for the Japanese fleet, while improved, were still based on very few trials and were inconsistent in some cases with empirical data (e.g., the relative number of tags returned by the Japanese fleet was at odds with others with a similar spatial distribution). Therefore, the Japanese fleet was allocated a reporting rate equal to that for Chinese-Taipei, with a standard error equal to the maximum among all the flags. Reporting rates for Indonesian and Spanish fleets were set to that estimated for the Philippines and Ecuador, respectively, given the similarities in fishing grounds and offloading ports as a result of insufficient data for those fleets.

Overall, current estimates range between 10-20% higher than the ones previously estimated by Hoyle (2011). This is predominantly an artefact of the different approaches taken and not the reporting rate itself. For instance, Hoyle's estimates included the effects of tag failure (current approach accounts for tag failure separate from tag reporting rates), which overall imply a reduction in the reporting rate of about 24% (6% base tag loss, 7% base mortality, and 13% average tagger effect). Therefore, the current estimates indicate a reporting rate of approximately 10% below previous estimates, on average. Due to the increase in the number of releases, the priors in the current assessments are also significantly more informative (i.e., tighter distribution), as reflected by the high penalties in the regions where the seeding trials have provided more precise reporting rate estimates for the main fishing fleets (Table 4).

The analysis assumes that the explanatory variable flag is the main factor affecting tag reporting rates, but it is possible that other factors related to fish processing (offloading port, fish processor, region, etc.) may also be influential factors. Although current estimates are considered a significant improvement in our knowledge of reporting rates, the significance of this parameter on assessment model outcomes supports the need for ongoing research, including continued tag seeding trials.

Discussion

The process of preparing tagging data for use in MULTIFAN-CL tuna stock assessments is imperative to maintain the reliability of model results. Several important changes have been made to this process from previous assessments that warrant recognition. First, release numbers

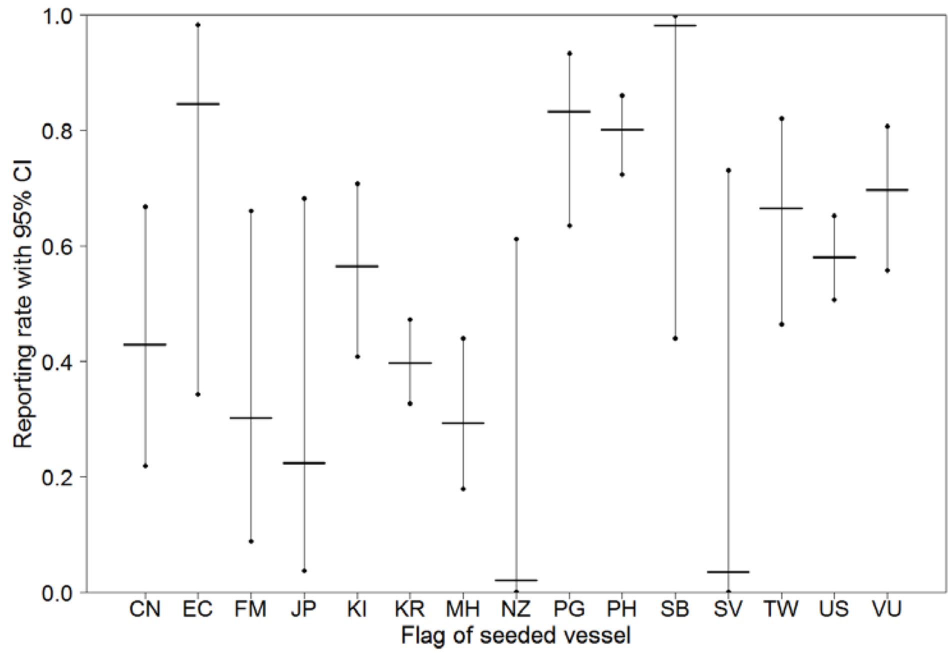

Figure 2. Reporting rate by flag, based on a quasibinomial generalized linear model with flag and tag type as explanatory variables.

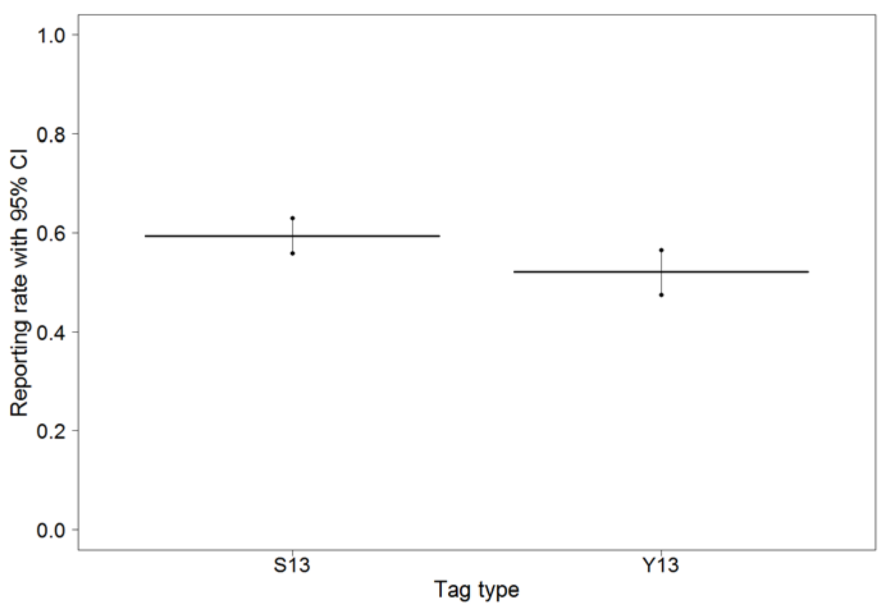

Figure 3. Reporting rate by tag type, based on a quasibinomial model with flag and tag type (S13: steel anchor tag; Y13: plastic anchor tag) as explanatory variables.

were adjusted to account for unusable recoveries to preserve population inferences (See Section 1). These adjustments were performed on data aggregated at the smallest scale possible (release event and length bin), while retaining the data input structure necessary for the assessment model. Previously, adjustments for unusable recoveries were done at the release event level. Finer scale adjustments require more recordkeeping, but act directly on the sampling unit of interest rather than applying a single adjustment to all fish by release event, regardless of size.

Second, tag shedding and tag-related mortality (combined base levels and additions from tagger-effects) were corrected for by adjusting tag releases downward by an estimated correction factor (see Section 2) rather than incorporating them into the reporting rate prior distribution (i.e., the loss of these tags from the tagged population treated as tags not reported). The latter has the effect of widening the reporting rate prior distribution, which gives the assessment model more freedom to adjust reporting rate parameters in an attempt to fit other data source discrepancies in the model (e.g., catch, CPUE trends) that are not related to tag reporting rates themselves. The former was the approach used in the 2014 assessments and, by accounting for the loss of tags more directly through adjusted releases, maintained the estimated reporting rate prior distributions at their most informative level (see Section 3). The disadvantage is that only the point estimates from tagger-effect models were used (in combination with base levels) to adjust releases, uncertainty in those effects were ignored.

Third, sample sizes from tag seeding experiments have increased, allowing improved precision of tag reporting rate estimates (Section 3). The more informative (less variance) prior distributions that result should help to improve model stability and limit model fitting adjustments that may not be related to tag reporting at all.

The exploration of candidate models using redefined regions in the skipjack assessment (from 3 to 5 regions) and the yellowfin and bigeye assessments (from 6 to 9 regions) also signify a significant change in preparing tagging data from previous years. The decision to split previously defined regions into smaller ones was, among other things, to better meet tag mixing assumptions and to improve fishing fleet definitions for assigning catchability and selectivity parameters. Tagging events – defined by tagging programme, year, quarter and region – increased by 43 to 65% as a result of moving to more regions. This resulted in more tagged populations to integrate into the assessment model, longer computing times on average to optimize models, and smaller sample sizes on average per tagging event.

References

Brownie C., Anderson D.R., Burnham K.P., Robson D.S. (1985). Statistical Inference from Band Recovery Data: a Handbook. U.S. Fish and Wildlife Service Resource Publication.

Fournier, D.A., J. Hampton, and J.R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences 55: 2105-2116.

Hampton, J., and Fournier, D.A. 2001. A spatially-disaggregated, length-based, age-structured population model of yellowfin tuna (Thunnus albacares) in the western and central Pacific Ocean. Marine and Freshwater Research 52:937−963.

Hilborn, R. 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. Canadian Journal of Fisheries and Aquatic Sciences 47: 635-643.

Hoyle S. (2011). Tag reporting rate prior distribution for the 2014 bigeye, yellowfin and skipjack stock assessments. WCPFC-SC7-2011/SA-IP-10.

Hoyle, S., B.M. Leroy, S.J. Nicol, and W.J. Hampton. 2014 - *In press*. Covariates of release mortality and tag loss in large-scale tuna tagging experiments. Fisheries Research. Available online 27 March 2014.

Hampton J. (1997). Estimates of tag-reporting and tag-shedding rates in a large-scale tuna tagging experiment in the western tropical Pacific Ocean. Fishery Bulletin 95, 68-79.

Kolody, D. and S. Hoyle. *In press* - 2014. Evaluation of tag mixing assumptions in western Pacific Ocean skipjack tuna stock assessment models. Fisheries Research. Available online 12 June 2014.

Maunder, M.N. and A.E. Punt. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research 142: 61-74.

Pollock K.H., Hoenig J.M., Hearn W.S., Calingert B. (2001). Tag reporting rate estimation: 1 An evaluation of the high-reward tagging method. N. Am. J. Fish. Manage. 21, 521–532.

Pollock K.H., Hearn W.S., Polacheck T. (2002). A general model for tagging on multiple component fisheries: an integration of age-dependent reporting rates and mortality estimation. Environ. Ecol. Stat. 9, 57–69.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Sibert, J.R., J. Hampton, D.A. Fournier, and P.J. Billis. 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (Katsuwonus pelamis). Canadian Journal of Fisheries and Aquatic Sciences 56: 925-938.

Skomal, G.B. 2007. Evaluating the physiological and physical consequences of capture on post-release survivorship in large pelagic fishes. Fisheries Management and Ecology 14: 81-89.

Appendices

Generalized and shortened versions of programming code are shown for the analysis that estimated the correction factors that accounted for individual tagger effects above base levels of tag shedding and short-term tag-related mortality (Appendix 1; referring to Section 2), and tag reporting rates by flag and region (Appendix 2; referring to Section 3). A request for the full program R code may be made by contacting simonn@spc.int.

Appendix 1. Program R code used to estimate tagger effect correction factors.

```
#####MODIFIED FROM S. HOYLE (2011)
#This script estimates relative differences in the probability of recovery associated with tagger
effects for three species and two tagging programmes (PTTP, RTTP).

#Load libraries
require(splines)
library(MASS)
library(xtable)
library(mgcv)
#_______________________________________________________________________________
##### EXTRACT AND STORE PTTP DATA
# Source support functions need for calculations
infile = ""  #data file inserted here
tag.all = infile  #specific tag data editing/cleaning omitted here for brevity

# Construct binary response variable
tag.all$relrec = ifelse(tag.all$recap == 'FALSE', 0, 1)

# Extract the year from the date string and use only up to 2012 for 2014 assessments
tag.all$rel_yr = as.numeric(substring(tag.all$rel_date,1,4))
tag.all = tag.all[tag.all$rel_yr < 2013,]

# Save for use later in the 'Correction of effective releases' section
save(tag.all, file='tag.dat_Uncleaned.RData')

#_______________________________________________________________________________
###### PROCESS DATA (shown for Skipjack only, but same general procedure for bigeye and yellowfin
load('tag.dat_Uncleaned.RData')
tag.spp = tag.all[tag.all$sp_id == 'S',]                    #extract for skipjack only ('S')
event.keep = table(tag.spp$tag_sch_id)                      # Table of number of releases by
tagging event
event.names = names(event.keep[event.keep > 29])            # Identify tagging events with >= 30
skj releases  (>=20 for YFT and >=15 for BET)
tag.spp = tag.spp[tag.spp$tag_sch_id in event.names,]      # Only keep those with >= 30 due to
sample size considerations

#Remove records with variables with low frequencies ~ <200, preliminary models suggested
parameter identitiy problems at low freequencies
table(tag.spp$cradle2)
table(tag.spp$Cond)
tag.spp = tag.spp[!(tag.spp$Cond in c('Eye damage','Hit side of boat','Tail damage')),]
table(tag.spp$Qual)
tag.spp = tag.spp[!tag.spp$Qual == 'Too slow',]

# Remove taggers with less than 200 releases
tagger.keep = table(tag.spp$tagger)                     # Number of releases by tagger
tagger.names = names(tagger.keep[tagger.keep > 199])    # Identify taggers with >= 200 skj
releases (>=200 for YFT and >=100 for BET)
tag.spp = tag.spp[tag.spp$tagger in tagger.names,]     # Only keep those with >= 200
tag.spp = tag.spp[!tag.spp$tagger in c('KMS','DWF'),]  # For BET only - IATTC taggers

event.keep = table(tag.spp$tag_sch_id)
event.names = names(event.keep[event.keep > 29])
tag.spp = tag.spp[tag.spp$tag_sch_id in event.names,]

# Set character strings to factors
tag.fct = sapply(tag.spp, is.character) # which variables are characters
tag.spp[tag.fct] = lapply(tag.spp[tag.fct], as.factor)

# Save it
write.csv(tag.spp, file='tag.dat_AsModelled_skj.csv')

#_______________________________________________________________________________
# Load PTTP data, run models, save model objects - SKJ
tag.spp = read.csv('tag.dat_AsModelled_skj.csv')
tag.spp$tag_sch_id = as.factor(tag.spp$tag_sch_id)   # Change from an integer to a factor

# Run GLMs with spline for length. Note: tested with dfs of 3 and 5 and confirmed that 4 was most
parsimonious (effective df of a prelim GAM was close to 4 too)
```

```
Bin.tag.correct = glm(relrec ~ tagger + ns(len5, df=4) + tag_sch_id + Cond + Qual + cradle2,
data=tag.spp, family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_full.RData')
Bin.tag.correct = glm(relrec ~ tagger + ns(len5, df=4) + tag_sch_id + Cond + Qual, data=tag.spp,
family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_cradle.RData')
Bin.tag.correct = glm(relrec ~ tagger + ns(len5, df=4) + tag_sch_id + Cond + cradle2,
data=tag.spp, family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_qual.RData')
Bin.tag.correct = glm(relrec ~ tagger + ns(len5, df=4) + tag_sch_id + Qual + cradle2,
data=tag.spp, family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_cond.RData')
Bin.tag.correct = glm(relrec ~ tagger + ns(len5, df=4) + Cond + Qual + cradle2, data=tag.spp,
family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_event.RData')
Bin.tag.correct = glm(relrec ~ tagger + tag_sch_id + Cond + Qual + cradle2, data=tag.spp,
family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_length.RData')
Bin.tag.correct = glm(relrec ~ ns(len5, df=4) + tag_sch_id + Cond + Qual + cradle2, data=tag.spp,
family=binomial("logit"))
  save(Bin.tag.correct, file='Bin.Mod.pttp_skj_tagger.RData')

#Get AIC values
load('Bin.Mod.pttp_skj_full.RData'); aic = AIC(Bin.tag.correct)
load('Bin.Mod.pttp_skj_cradle.RData'); aic = c(aic,AIC(Bin.tag.correct))
load('Bin.Mod.pttp_skj_cond.RData'); aic = c(aic,AIC(Bin.tag.correct))
load('Bin.Mod.pttp_skj_event.RData'); aic = c(aic,AIC(Bin.tag.correct))
load('Bin.Mod.pttp_skj_length.RData'); aic = c(aic,AIC(Bin.tag.correct))
load('Bin.Mod.pttp_skj_tagger.RData'); aic = c(aic,AIC(Bin.tag.correct))
load('Bin.Mod.pttp_skj_qual.RData'); aic = c(aic,AIC(Bin.tag.correct))

# Results table
Modl = c('Full','-Cradle','-Cond','-tag_sch_id','-Len','-Tagger','-Qual') # Model names
newtab = xtable(cbind(Modl,aic=round(aic,1),chngAIC=round(aic-min(aic),2))) # Construct xtable
for printing
print.xtable(newtab, type='html', file='GLM_Model_Selection_skj_PTTP.html')

#Examples of GAMS run for BET
#Bin.tag.correct = gam(relrec ~ tagger + te(len5) + tag_sch_id + Cond + cradle2, data=tag.spp,
family=binomial, control=list##(maxit=200)); r.sq = summary(Bin.tag.correct)$r.sq; dev.ex =
summary(Bin.tag.correct)$dev.expl; aic = AIC(Bin.tag.correct)

#Examples of GAMS run for YFT
#Bin.tag.correct = gam(relrec ~ tagger + te(len5) + tag_sch_id + Cond + Qual + cradle2,
data=tag.spp, family=binomial, control#=list(maxit=200))

#AIC diagnostics not shown here

#________________________________________________________________________________
#________________________________________________________________________________
#Extract and store RTTP data

# Load data
infile = ""         #data file inserted here
rtag.all = infile  #specific tag data editing\cleaning omitted here for brevity

# Check the total number of releases per tagging event
sort(tapply(rtag.all$totrel,rtag.all$tag_sch_id,sum))

# Save full cleaned data file
save(rtag.all, file='rtag.dat_AsModelled.RData')

#________________________________________________________________________________
###### PROCESS DATA (shown for Skipjack only, but same general procedure for bigeye and yellowfin
rtag.all.skj = rtag.all[rtag.all$sp == 'a.SKJ',]   # SKJ
sort(tapply(rtag.all.skj$totrel,rtag.all.skj$tag_sch_id,sum))

# Cull tagging events with less than 30 fish tagged (20 for YFT and 15 for BET)
event.keep = tapply(rtag.all.skj$totrel,rtag.all.skj$tag_sch_id,sum)
event.names = names(event.keep[event.keep > 29])
rtag.all.skj = rtag.all.skj[rtag.all.skj$tag_sch_id in event.names,]
```

```
# Cull taggers with less than 200 fish tagged  (100 for YFT and 30 for BET)
sort(tapply(rtag.all.skj$totrel,rtag.all.skj$tagger,sum))
tagger.keep = tapply(rtag.all.skj$totrel,rtag.all.skj$tagger,sum)
tagger.names = names(tagger.keep[tagger.keep > 199])
rtag.all.skj = rtag.all.skj[rtag.all.skj$tagger in tagger.names,]

# Get rid of data from other categories with less than X fish tagged
tapply(rtag.all.skj$totrel,rtag.all.skj$cradle2,sum)   # All good, no culling necessary for SKJ,
YFT, and BET
tapply(rtag.all.skj$totrel,rtag.all.skj$Cond,sum)
rtag.all.skj = rtag.all.skj[!rtag.all.skj$Cond == 'Tail damage',]   # < 200  (BET < 100; YFT <
100)
tapply(rtag.all.skj$totrel,rtag.all.skj$Qual,sum)
rtag.all.skj = rtag.all.skj[!rtag.all.skj$Qual == 'Too slow',]   # < 200 (BET - not available;
YFT all good so no cull)

# Ensure that tagging events with < 30 releases are removed   (redundant if no cleaning done made
above)
event.keep = tapply(rtag.all.skj$totrel,rtag.all.skj$tag_sch_id,sum)
event.names = names(event.keep[event.keep > 29])
rtag.all.skj = rtag.all.skj[rtag.all.skj$tag_sch_id in event.names,]
tag.spp = rtag.all.skj

save(tag.spp, file='rtag_AsModelled_SKJ.RData')

#________________________________________________________________________________
# Load RTTP data, run models, save model objects - SKJ

load('rtag_AsModelled_SKJ.RData')
tag.spp$tagger=as.factor(tag.spp$tagger);tag.spp$tag_sch_id=as.factor(tag.spp$tag_sch_id);tag.spp
$Cond=as.factor(tag.spp$Cond);tag.spp$Qual=as.factor(tag.spp$Qual);tag.spp$cradle2=as.factor(tag.
spp$cradle2)     # Set to factor

# Run GAMs
Bin.tag.correct = gam(relrec ~ tagger + te(len5) + tag_sch_id + Cond + Qual + cradle2,
data=tag.spp, family=binomial, control   =list(maxit=200)) r.sq = summary(Bin.tag.correct)$r.sq;
dev.ex = summary(Bin.tag.correct)$dev.expl; aic = AIC(Bin.tag.correct)
   save(Bin.tag.correct, file='Bin.Mod.rttp_skj.RData')

Bin.tag.correct = gam(relrec ~ as.factor(tagger) + te(len5) + as.factor(tag_sch_id) +
as.factor(Cond) + as.factor(Qual), data=tag.spp, family=binomial, control=list(maxit=200)); r.sq
= c(r.sq,summary(Bin.tag.correct)$r.sq); dev.ex = c(dev.ex,summary(Bin.tag.correct)$dev.expl);
aic = c(aic,AIC(Bin.tag.correct))

Bin.tag.correct = gam(relrec ~ as.factor(tagger) + te(len5) + as.factor(tag_sch_id) +
as.factor(Cond) + as.factor(cradle2), data=tag.spp, family=binomial, control=list(maxit=200));
r.sq = c(r.sq,summary(Bin.tag.correct)$r.sq); dev.ex =
c(dev.ex,summary(Bin.tag.correct)$dev.expl); aic = c(aic,AIC(Bin.tag.correct))

Bin.tag.correct = gam(relrec ~ as.factor(tagger) + te(len5) + as.factor(tag_sch_id) +
as.factor(Qual) + as.factor(cradle2), data=tag.spp, family=binomial, control=list(maxit=200));
r.sq = c(r.sq,summary(Bin.tag.correct)$r.sq); dev.ex =
c(dev.ex,summary(Bin.tag.correct)$dev.expl); aic = c(aic,AIC(Bin.tag.correct))

Bin.tag.correct = gam(relrec ~ as.factor(tagger) + te(len5) + as.factor(Cond) + as.factor(Qual) +
as.factor(cradle2), data=tag.spp, family=binomial, control=list(maxit=200)); r.sq =
c(r.sq,summary(Bin.tag.correct)$r.sq); dev.ex = c(dev.ex,summary(Bin.tag.correct)$dev.expl); aic
= c(aic,AIC(Bin.tag.correct))

Bin.tag.correct = gam(relrec ~ as.factor(tagger) + as.factor(tag_sch_id) + as.factor(Cond) +
as.factor(Qual) + as.factor(cradle2), data=tag.spp, family=binomial, control=list(maxit=200));
r.sq = c(r.sq,summary(Bin.tag.correct)$r.sq); dev.ex =
c(dev.ex,summary(Bin.tag.correct)$dev.expl); aic = c(aic,AIC(Bin.tag.correct))

Bin.tag.correct = gam(relrec ~ te(len5) + as.factor(tag_sch_id) + as.factor(Cond) +
as.factor(Qual) + as.factor(cradle2), data=tag.spp, family=binomial, control=list(maxit=200));
r.sq = c(r.sq,summary(Bin.tag.correct)$r.sq); dev.ex =
c(dev.ex,summary(Bin.tag.correct)$dev.expl); aic = c(aic,AIC(Bin.tag.correct))

# Results table
```

```
Modl = c('Full','-Cradle','-Qual','-Cond','-tag_sch_id','-Len','-Tagger') # Model names
newtab =
xtable(cbind(Modl,dev.ex=round(dev.ex,4),r.sq=round(r.sq,4),aic=round(aic,1),chngAIC=round(aic-
min(aic),2))) # Construct xtable for printing
print.xtable(newtab, type='html', file='GAM_Model_Selection_skj_RTTP.html')

#_______________________________________________________________________________
#_______________________________________________________________________________
# CORRECTION CALCULATIONS  - Do not want to take length into account in the correction since
that's mostly q, not mortality
# predict recoveries using existing model

# Function for calculating the correction ratio for the various modelled and unmodelled tagging
events
Correct.pttp =
function(dat.name,mod.name,correction.string,plot.main,plot.name,spp,rm.Cond,rm.Qual,stat.type,ta
g.proj)
{
    require(mgcv)
    if(tag.proj == 'pttp') tag.spp = read.csv(dat.name)
    if(tag.proj == 'rttp') load(dat.name)
    tag.spp$tag_sch_id = as.factor(tag.spp$tag_sch_id)
    load(mod.name)

    # Predict recapture rate based on observed values for covariates
    tag.spp$pred.obs = predict(Bin.tag.correct, newdata=tag.spp, type='response')
    # Set covariate values for the 'good' reference case e.g. good tagger, fish in good condition
etc.
    tag.spp.new = tag.spp
    tag.spp.new$tagger = correction.string[1]
    tag.spp.new$Cond = correction.string[2]
    tag.spp.new$Qual = correction.string[3]
    tag.spp.new$cradle2 = correction.string[4]

    # Predict recapture rate based on the reference case
    tag.spp$pred.reference = predict(Bin.tag.correct, newdata=tag.spp.new, type='response')

    # Calculate the ratio of observed to reference case recapture rate at the individual release
level
    tag.spp$correction = tag.spp$pred.obs/tag.spp$pred.reference

    # Aggregate to the tagging event level
    event.corrections = sort(tapply(tag.spp$correction, tag.spp$tag_sch_id, mean))

    # Transform into an exporting format
    correct.factors = data.frame(tag.event = names(event.corrections), correction.ratio =
event.corrections, correction.type =     'MODELLED')
    correct.factors =
correct.factors[order(as.numeric(as.character(correct.factors$tag.event))),]   # Reorder by
tagging event

    ### Correcting events not in model
    tag.fct = sapply(tag.spp, is.factor) # which variables are factors
    tag.spp[tag.fct] = lapply(tag.spp[tag.fct], as.character) # Change to character

    # Determine values to be used for taggers and tagging events that were not modelled but must
be corrected for
    N.taggers = length(unique(tag.spp$tagger))
    N.events = length(unique(tag.spp$tag_sch_id))

    # Identify a median relative tagger to be used in cases where unmodelled taggers are
encountered
    tagger.coefs = coef(Bin.tag.correct)[2:N.taggers]
    med.tagger = median(tagger.coefs)
    med.tagger = which(abs(tagger.coefs - med.tagger) == min(abs(tagger.coefs - med.tagger)))
    med.tagger = substring(names(med.tagger),7,9)   # Remove the 'tagger' part of the tagger
coefficient name

    # Identify a median tagging event in order to use predict function
    event.coefs = coef(Bin.tag.correct)[(N.taggers+1):(N.taggers+N.events-1)]
    med.event = median(event.coefs)   # Again ignores the intercept
```

```
    med.event = which(abs(event.coefs - med.event) == min(abs(event.coefs - med.event)))
    med.event = gsub("[^\\d]+", "", names(med.event), perl=TRUE)[1]   # Extract only the number

    # Extract skj data from the 'full' dataset
    if(tag.proj == 'pttp') load('tag.dat_Uncleaned.RData')   # Data.frame within this is tag.all
    if(tag.proj == 'rttp')
                {
                load('rtag.dat_AsModelled.RData')
                tag.all = rtag.all   # Data.frame within this is tag.all
                tag.all$sp_id = tag.all$sp
    }
    full.tags = tag.all[tag.all$sp_id == spp,]
    all.events = unique(full.tags$tag_sch_id)   # Identifies all tagging events available that
released fish

    # Identify the taggers for which model coefficients exist
    modelled.taggers = unique(tag.spp$tagger)

    # Determine unmodelled and modelled tagging events
    modelled.events = unique(tag.spp$tag_sch_id)   # Identifies all tagging events that were
modelled/corrected

    # Extract all releases for unmodelled/uncorrected tagging events
    unmodelled.tags = full.tags[!full.tags$tag_sch_id in modelled.events,]
    unmodelled.tags$tmp.tag_sch_id = unmodelled.tags$tag_sch_id # housekeeping

    # remove few releases where coefficients do not exist
    unmodelled.tags = unmodelled.tags[!unmodelled.tags$Cond in rm.Cond,] #c('Eye damage','Hit
side of boat','Tail damage'),]
    unmodelled.tags = unmodelled.tags[!unmodelled.tags$Qual in rm.Qual,] #== 'Too slow',] # Only
removes a total of 9 fish

    # Assign 'median' coefficients to releases where the tagger or event was unmodelled
    unmodelled.tags[!unmodelled.tags$tagger in c(modelled.taggers,'DWF','KMS'),]$tagger =
med.tagger
    if(tag.proj == 'pttp') unmodelled.tags[unmodelled.tags$tagger in c('DWF','KMS'),]$tagger =
'BML'  #set IATTC taggers to       base
    unmodelled.tags[!unmodelled.tags$tag_sch_id in modelled.events,]$tag_sch_id = med.event
# Median coefficients for      unmodelled events

    # Predict recapture rate based on 'observed' covariates
    unmodelled.tags$pred.obs = predict(Bin.tag.correct, newdata=unmodelled.tags, type='response')
    unmodelled.tags.new = unmodelled.tags
    unmodelled.tags.new$tagger = correction.string[1]
    unmodelled.tags.new$Cond = correction.string[2]
    unmodelled.tags.new$Qual = correction.string[3]
    unmodelled.tags.new$cradle2 = correction.string[4]

    # Predict recapture rate based on the reference case
    unmodelled.tags$pred.reference = predict(Bin.tag.correct, newdata=unmodelled.tags.new,
type='response')

    # Calculate the ratio of observed to reference case recapture rate at the individual release
level
    unmodelled.tags$correction = unmodelled.tags$pred.obs/unmodelled.tags$pred.reference

    # Aggregate to the tagging event level
    event.corrections_unmod = sort(tapply(unmodelled.tags$correction,
unmodelled.tags$tmp.tag_sch_id, mean))

    # Transform into an exporting format
    correct.factors_unmod = data.frame(tag.event = names(event.corrections_unmod),
correction.ratio = event.corrections_unmod       ,correction.type = 'UNMODELLED')
    correct.factors_unmod =
correct.factors_unmod[order(as.numeric(as.character(correct.factors_unmod$tag.event))),]   #
Reorder    by tagging event

    #Combine modelled and unmodelled corrections and save
    correct.factors = rbind(correct.factors, correct.factors_unmod)
    save(correct.factors, file=paste('correct.factors_',plot.main,'.RData',sep=''))
}
```

```r
#End function
#____________________________________________________________________________________
#____________________________________________________________________________________
# PTTP - SKJ correction   (similar for BET and YFT)
Correct.pttp(dat.name='tag.dat_AsModelled_skj.csv', mod.name='Bin.Mod.pttp_skj.RData',
correction.string=c('BML','Good','Good','BOW'),plot.main='SKJ - pttp',
plot.name='TaggerEfx_PTTP_SKJ_coefs', spp='S', rm.Cond=c('Eye damage','Hit side of boat','Tail
damage'),rm.Qual='Too slow', stat.type='GLM', tag.proj='pttp')

#____________________________________________________________________________________
# RTTP - SKJ correction   (similar for YFT; BET not done)
Correct.pttp(dat.name='rtag_AsModelled_SKJ.RData', mod.name='Bin.Mod.rttp_skj.RData',
correction.string=c('ETP','a.Good','a.Good','a.BOW'),plot.main='SKJ - rttp',
plot.name='TaggerEfx_RTTP_SKJ_coefs', spp='a.SKJ', rm.Cond='Tail damage',rm.Qual='Too slow',
stat.type='GAM', tag.proj='rttp')

#____________________________________________________________________________________
```

Appendix 2. Program R code used to estimate reporting rates.

```
#####MODIFIED FROM S. HOYLE (2011)

#This script estimates the Reporting Rates for the PS fleets according to the 2014 regional
structure and fleet definition for the three species.

require(boot)
require(RODBC)
require(car)

#First, query the database from CES: aggregate CE data raised for the PS fleet in the WCPFC.
Estimates are for the PTTP, so only select period 2007-2012.

CE <- read.csv('../Catch_WCPFC.TXT')

######################
####DEFINE REGIONS ###
######################
### DEFINE MFCL skj regions
CE$mfclskj <- rep(-9, length(CE$yy))
CE$mfclskj <- ifelse(CE$latd >  20 & CE$latd <  50 & CE$lond > 120 & CE$lond < 210, 1,
CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > 0 & CE$latd <  20 & CE$lond > 140 & CE$lond < 170, 2, CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > -5 & CE$latd <  0 & CE$lond > 155 & CE$lond < 160, 2, CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > -20 & CE$latd <  0 & CE$lond > 160 & CE$lond < 170, 2, CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > -20 & CE$latd <  20 & CE$lond > 170 & CE$lond < 210, 3,
CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > -20 & CE$latd <   20 & CE$lond > 110 & CE$lond < 140, 4,
CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > -20 & CE$latd <   0 & CE$lond > 140 & CE$lond < 155, 5,
CE$mfclskj)
CE$mfclskj <- ifelse(CE$latd > -20 & CE$latd <   -5 & CE$lond > 155 & CE$lond < 160, 5,
CE$mfclskj)
### END DEFINE MFCL skj regions

###DEFINE REGIONS BET/YFT
CE$mfclyft <- -9
CE$mfclyft[CE$latd >  20 & CE$latd <  50 & CE$lond > 120 & CE$lond < 170] <- 1
CE$mfclyft[CE$latd >  20 & CE$latd <  50 & CE$lond > 170 & CE$lond < 210] <- 2
CE$mfclyft[CE$latd > 0 & CE$latd <  20 & CE$lond > 140 & CE$lond < 170] <- 3
CE$mfclyft[CE$latd > -5 & CE$latd <  0 & CE$lond > 155 & CE$lond < 160] <- 3
CE$mfclyft[CE$latd > -10 & CE$latd <  0 & CE$lond > 160 & CE$lond < 170] <- 3
CE$mfclyft[CE$latd > -10 & CE$latd <  20 & CE$lond > 170 & CE$lond < 210] <- 4
CE$mfclyft[CE$latd > -40 & CE$latd < -10 & CE$lond > 140 & CE$lond < 170] <- 5
CE$mfclyft[CE$latd > -40 & CE$latd < -10 & CE$lond > 170 & CE$lond < 210] <- 6
CE$mfclyft[CE$latd > -10 & CE$latd < 20 & CE$lond > 110 & CE$lond < 140] <- 7
CE$mfclyft[CE$latd > -10 & CE$latd < 0 & CE$lond > 140 & CE$lond < 155] <- 8
CE$mfclyft[CE$latd > -10 & CE$latd < -5 & CE$lond > 155 & CE$lond < 160] <- 8
CE$mfclyft[CE$latd > -20 & CE$latd < -15 & CE$lond > 140 & CE$lond < 150] <- 9

###END DEFINE REGIONS

######################
####DEFINE FLAGS  ####
######################
#Only bother for those flags with at least 1% of the catch in the region.

SKJflags=NULL
for (SKJregion in 1:5) {
  a <- with(CE[CE$mfclskj==SKJregion,],tapply(skj_mt,flag,sum,na.rm=T)/sum(skj_mt,na.rm=T))
  a[is.na(a)]=0
  a=a[a>0.01]
  a=a/sum(a)
  SKJflags=rbind(SKJflags,data.frame(flag=names(a),region=SKJregion,catch=a))
}

BETflags=NULL
for (BETregion in 1:9) {
  a <- with(CE[CE$mfclyft==BETregion,],tapply(bet_mt,flag,sum,na.rm=T)/sum(bet_mt,na.rm=T))
  a[is.na(a)]=0
```

```
  a=a[a>0.01]
  a=a/sum(a)
  if (length(a)>0) {
    BETflags=rbind(BETflags,data.frame(flag=names(a),region=BETregion,catch=a))
  }
}

YFTflags=NULL
  for (YFTregion in 1:8) {
  a <- with(CE[CE$mfclyft==YFTregion,],tapply(yft_mt,flag,sum,na.rm=T)/sum(yft_mt,na.rm=T))
  a[is.na(a)]=0
  a=a[a>0.01]
  a=a/sum(a)
  YFTflags=rbind(YFTflags,data.frame(flag=names(a),region=YFTregion,catch=a))
}
##########################
####END DEFINE FLAGS  ####
##########################

##########################
## MODEL REPORTING RATES #
##########################
# Load tagging data from seeding trials from database.
channel <- odbcConnectAccess("File path")# File path not shown here
dat <- sqlQuery(channel, "SELECT tag_cruise.cruise_id, tag_cruise.cruise_start,
tag_cruise.cruise_end, tag_cruise.flag_id, tag_release.tag_type, Count(tag_release.tag_rel_id) AS
CountOftag_rel_id,
SUM(SWITCH(tag_recovery.tag_recovered='Y',1,tag_recovery.tag_recovered<>'Y',0)) AS
CountOfrecoveryyes
FROM ((tag_cruise INNER JOIN tag_event ON tag_cruise.tag_cru_id = tag_event.tag_cru_id) INNER
JOIN tag_release ON tag_event.tag_event_id = tag_release.tag_event_id) LEFT JOIN tag_recovery ON
tag_release.tag_rel_id = tag_recovery.tag_rel_id
WHERE (((tag_cruise.proj_id)=6))
GROUP BY tag_cruise.cruise_id, tag_cruise.cruise_start, tag_cruise.cruise_end,
tag_cruise.flag_id, tag_release.tag_type ORDER BY tag_cruise.cruise_id")
names(dat) <- c("cruise_id" ,"cruise_start","cruise_end","flag_id","tag_type","rel","recov")
close(channel)

dat$recov[is.na(dat$recov)]=0

###Eliminate trials after 1 Jan 2013. In addition, eliminate JP and TW trials after 30June2012
(at odds with the no of tags reported by these flags)
dat=dat[(dat$cruise_start<ISOdate(2013,1,1) & !dat$flag_id in c("JP","TW")) |
(dat$cruise_start<ISOdate(2012,6,30) & dat$flag_id in c("JP","TW")),]    #274 lines

#Model struggles to fit 100% or 0% recoveries (therefore a recovery in the case of SV is
assumed(this does not affect results)
dat[dat$flag_id=="SV" & dat$tag_type=="S13",7][1]=1

#Fit with binomial distribution. Flag and tag type as factors.
model1 <- glm(cbind(recov,rel-recov) ~ as.factor(flag_id) +
as.factor(tag_type),family=binomial,data=dat);
Anova(model1) #Flag_id and tag type significant
summary(model1)
#Residual deviance >> redisual degrees of freedom, so overdispersed. Use a quasibinomial model.

# Quasibinomial version
model1b <- glm(cbind(recov,rel-recov) ~ as.factor(flag_id) +
as.factor(tag_type),family=quasibinomial,data=dat); summary(model1b)  #disp 4.8419

# Simulate with model1b, flag effect and tag type (only S13 tags)
newdat <- expand.grid(flag_id=as.factor(sort(unique(dat$flag_id))),tag_type='S13')
dat.t <- predict.glm(model1b,newdat=newdat,type="terms",se.fit=T)
dat.r <- predict.glm(model1b,newdat=newdat,type="response",se.fit=T)
con <- attributes(dat.t[[1]])$constant
seedests <- cbind(newdat,con,dat.t[[1]],dat.t[[2]],dat.r[[1]],dat.r[[2]]) # get predictions
fl_mean <- con + seedests[,4]

#####################################
## add/Modify values for some flags##
#####################################
```

```
## It was decided that the estimate for JP is unrealistically low, taking into account the number
of tags they report from PTTP, and to assign the value of the TW fleet.
## There is no estimate for ES and ID. For ID, it was assumed the same RR as for PH; for ES, the
same as for EC.
## These were supplied a SD of 2.2 (maximum of the other flags to be conservative).

seedests=seedests[seedests$flag_id!='JP',]
tt=cbind(flag_id=c("JP","ES","ID","CO","GT","HN","NI","PA","VE"),tag_type=rep("S13",9),rbind(seed
ests[seedests$flag_id=="TW",3:9],

seedests[seedests$flag_id in c("EC","PH"),3:9],matrix(rep(apply(seedests[seedests$flag_id in
c("EC","SV","VU"),3:9],2,mean),6),byrow=T,nrow=6,dimnames=c(list(1:6),list(colnames(seedests[,3:9
]))))))
tt[,6]=max(seedests[,6])
names(tt)=names(seedests)
seedests=rbind(seedests,tt)
#####################################
## end add/Modify values          ##
#####################################

############################################
###FUNCTION FOR THE CALCULATION OF RR ######
############################################
make_pens <- function(dat,n,seedests,pl=T) {
   flag2=dat$flag
   meanRR=NA
   pen_RR=NA
   cv_RR=NA
   sd_RR=NA
   if (length(flag2)>0) {
     dat$mn <- seedests[match(flag2,seedests$flag_id),4]+ seedests$con[1]
     dat$stddv <- seedests[match(flag2,seedests$flag_id),6]
     dist_fl <- t(sapply(rep(length(dat$mn),n),rnorm,mean=dat$mn,sd=dat$stddv))
     dist_tt <- rnorm(n,seedests[1,5],seedests[1,7])
     dist_p <- inv.logit(dist_fl+dist_tt)
     wtd_p <- (rep(dat$catch,each=n) * dist_p)
     if (dim(wtd_p)[2] > dim(wtd_p)[1] ) {wtd_p=t(wtd_p)}
     RR_totdist <- apply(wtd_p,1,sum)
     meanRR <- mean(RR_totdist)
     if(pl) {
        windows()
        hist(RR_totdist,10,xlim=c(0,1))
     }
     cv_RR <- sd(RR_totdist)/mean(RR_totdist)   # CV
     pen_RR <- round(((1/cv_RR)^2)/2)
     sd_RR=sd(RR_totdist)
   }
  return(list(meanRR=meanRR,pen_RR=pen_RR,cv_RR=cv_RR,sd_RR=sd_RR))
  }
################################################
###END FUNCTION FOR THE CALCULATION OF RR ######
################################################

#########################
###RR PRIORS ESTIMATION##
#########################
#SKJ RR
SKJ_RR=NULL
for (i in 1:5) {
SKJ_RR=rbind(SKJ_RR,cbind(region=i,make_pens(SKJflags[SKJflags$region==i,],n=1000000,seedests=see
dests)))
}
graphics.off()
#YFT RR
YFT_RR=NULL
for (i in 1:8) {
YFT_RR=rbind(YFT_RR,cbind(region=i,make_pens(YFTflags[YFTflags$region==i,],n=1000000,seedests=see
dests)))
}
graphics.off()
#BET RR
```

```
BET_RR=NULL
for (i in 1:9) {
BET_RR=rbind(BET_RR,cbind(region=i,value=make_pens(BETflags[BETflags$region==i,],n=1000000,seedes
ts=seedests)))
}
graphics.off()

RESULT=rbind(cbind(rep('SKJ',5),1:5,matrix(SKJ_RR[,2],byrow=T,ncol=4)),
             cbind(rep('YFT',8),1:8,matrix(YFT_RR[,2],byrow=T,ncol=4)),
             cbind(rep('BET',9),1:9,matrix(BET_RR[,2],byrow=T,ncol=4)))

colnames(RESULT)=c("sp_id","region","RR_mean","RR_penalty","RR_cv","RR_sd")
#############################
###End RR PRIORS ESTIMATION##
#############################
```