SCIENTIFIC COMMITTEE
EIGHTH REGULAR SESSION

7-15 August 2012
Busan, Republic of Korea

Estimation of the species composition of the catch by purse seiners in the Western and Central Pacific Ocean using grab samples and spill samples collected by observers

WCPFC–SC8–2012 / ST–WP–03 (Rev. 1)

Timothy Lawson [1]

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

ESTIMATION OF THE SPECIES COMPOSITION OF THE CATCH BY PURSE SEINERS IN THE WESTERN AND CENTRAL PACIFIC OCEAN USING GRAB SAMPLES AND SPILL SAMPLES COLLECTED BY OBSERVERS

Timothy Lawson
Oceanic Fisheries Programme
Secretariat of the Pacific Community
Noumea, New Caledonia

Abstract

This paper (a) updates the estimation of the selectivity bias of grab samples collected by observers at sea with recent paired grab and spill sampling data; (b) considers the effect of layering by size during brailing on the selectivity bias; (c) corrects historical grab samples with new estimates of the selectivity bias; (d) further develops a model-based approach to estimate the species composition of purse-seine catches from grab samples corrected for selectivity bias and spill samples; and (e) uses the catches determined from the model-based estimates of the species composition to scale purse-seine length frequencies.

The increase in the number of paired samples from unassociated sets that are now available has allowed for more reliable estimates of the selectivity bias for larger fish. The use of splines, rather than categorical covariates, results in continuous estimates of the bias as a function of fish length. Layering by size during brailing is shown to occur and may be an important cause of the selectivity bias in grab samples.

The species compositions of purse-seine catches, 1967–2011, were estimated with models in which geographic area was included as either (a) the MFCL Skipjack Areas 2 and 3 (the ‗low resolution' models) or (b) a two-dimensional spline of latitude and longitude (the ‗high resolution' models). The heat maps of the effect of the latitude–longitude spline on the species composition shows that within each of the MFCL Skipjack Areas, the species composition varies considerably with location, which supports the use of the ‗high resolution' models to estimate the species composition.

Introduction

The estimation of the species composition of catches by purse seiners in the Western and Central Pacific Ocean is problematic because (a) catches are misreported on logsheets; (b) information on unloadings to canneries or during transshipment is unavailable; (c) port sampling during transshipment is biased because of well mixing and other factors; and (d) grab samples taken by observers at sea are subject to selectivity bias (Lawson 2009, Hampton & Williams 2011). The SPC Oceanic Fisheries Programme (OFP) has recently developed an observer sampling protocol — *the* *spill sample* — that avoids the selectivity bias (Lawson 2008). Paired grab and spill samples have been collected [1] and used to estimate the selectivity bias and correct the historical grab samples (Lawson 2010). This paper (a) updates the estimation of selectivity bias with recent paired sampling data; (b) considers the effect of layering by size during brailing on the selectivity bias; (c) corrects historical grab samples with the new estimates of the selectivity bias; (d) further develops a model-based approach to estimate the species composition of purse-seine catches from grab samples corrected for selectivity bias and spill samples; and (e) uses the catches determined from the model-based estimates of the species composition to scale purse-seine length frequencies.

For this study, lengths (cm) were converted to weights (kg) using the length-weight parameters below:

Species	**a**	**b**
Skipjack	0.8639E-05	3.2174
Yellowfin	2.5120E-05	2.9396
Bigeye	1.9729E-05	3.0247

Paired grab samples and spill samples, 2008–2012

Prior to 2008, it had been suspected that the species composition of the catch determined from grab samples collected by observers onboard purse seiners were biased, such that the proportion of skipjack was generally under-estimated and the proportion of yellowfin over-estimated. The protocol followed by observers when taking grab samples is to randomly select five fish per brail as the fish are transferred from the pursed net in the ocean to the holding wells on the wet deck. The cause of the bias was thought to be a tendency on the part of observers to non-randomly select fish; therefore, a new sampling protocol was developed to avoid this selectivity bias. The protocol for a spill sample is to spill the fish directly from the brail into a bin, such that the observer does not select the fish. A typical bin holds from about 100 to 400 fish depending on the size of the fish. Since a spill sample is much larger than a grab sample, and thus takes more time, spill samples are usually taken from every tenth brail. The starting brail for a given set is rotated to avoid possible effects of layering of fish by size (discussed below); thus, brail #1 is chosen as the first brail sampled from set #1, brail #2 as the first brail sampled from set #2, etc.

[1] The collection of paired grab and spill samples has been financed by the WCPFC under Project 60, ―The Collection and Evaluation of Purse-Seine Species Composition Data", since 2009, and by the New Zealand Aid Programme, ―Pacific Economic Growth Observer Programme", since 2011.

Spill sampling was conducted for the first time in March 2008, by OFP staff on a vessel operating in the waters of Papua New Guinea. [2] For each set, paired grab and spill samples were collected. Paired sampling was further conducted during three additional trips in Papua New Guinea, from June to August 2008, resulting in paired samples from a total of 66 sets. Lawson (2009) compared the paired grab and spill samples, and found that, as expected, the proportion of skipjack was higher, and the proportion of yellowfin lower, in species compositions determined from spill samples, compared to those determined from grab samples. Lawson (2010) obtained a similar result comparing paired samples collected during 17 trips from March 2008 to May 2010. At the time of writing (July 2012), paired sampling data are available covering 348 sets collected during 23 trips (Table 1). In Lawson (2010), data covering 254 sets were available, but only 11 of those (4.3%) were from unassociated schools. It is noteworthy that since then, the number of sets on unassociated schools has increased to 82 (23.6%).

Table 1. Date, location, catch and number of sets sampled for trips during which paired grab and spill samples were collected

Trip #	Date		Latitude		Longitude		Sampled Catch (Tonnes)	Number of Sets					
	Min	Max	Min	Max	Min	Max		Total	Anchored FADs	Drifting FADs	Logs	Unassoc	Other
1	23-Mar-08	27-Mar-2008	03S	01S	143E	146E	452	7	0	0	0	0	7
2	09-Jun-08	30-Jun-08	04S	00N	143E	149E	580	13	10	1	0	0	2
3	21-Jun-08	08-Aug-08	03S	00N	141E	150E	1,172	31	30	0	1	0	0
4	14-Jul-08	09-Aug-08	03S	02S	141E	146E	616	15	9	4	1	0	1
5	03-May-09	05-Jun-09	04S	02S	148E	151E	469	15	13	0	1	1	0
6	04-May-09	04-Jun-09	02S	01S	143E	146E	256	9	8	0	0	0	1
7	04-Jun-09	19-Jul-09	05S	02S	142E	151E	613	23	20	1	2	0	0
8	15-Jun-09	18-Jul-09	04S	01S	144E	148E	335	13	9	0	4	0	0
9	16-Jun-09	26-Jul-09	05S	02S	142E	150E	352	22	17	0	5	0	0
10	22-Aug-09	10-Sep-09	04S	04S	150E	151E	317	16	10	1	4	0	1
11	27-Sep-09	10-Oct-09	05S	02S	143E	150E	518	10	7	0	3	0	0
12	09-Oct-09	21-Oct-09	02S	02S	143E	144E	541	8	4	0	4	0	0
13	03-Nov-09	01-Dec-09	03S	01S	143E	146E	514	15	12	0	3	0	0
14	11-Nov-09	04-Dec-09	03S	02S	143E	146E	388	14	13	0	0	0	1
15	13-Nov-09	07-Dec-09	03S	02S	142E	142E	460	15	15	0	0	0	0
16	19-Mar-10	16-Apr-10	04S	00N	146E	165E	749	20	0	10	0	9	1
17	30-Apr-10	07-May-10	00N	01N	152E	154E	343	8	0	7	0	1	0
18	10-Dec-10	06-Jan-11	06S	01S	152E	160E	866	21	0	2	0	16	3
19	28-Nov-11	12-Dec-11	09S	08S	158E	159E	240	10	10	0	0	0	0
20	19-Jan-12	21-Feb-12	08S	02N	145E	162E	811	16	0	1	0	15	0
21	07-Feb-12	18-Feb-12	01N	02N	144E	150E	1,036	12	0	0	0	12	0
22	09-Mar-12	13-Apr-12	00N	02N	144E	155E	1,047	22	0	0	3	19	0
23	11-Mar-12	19-Apr-12	06S	02N	148E	161E	911	13	0	3	0	9	1
						Total	13,587	348	187	30	31	82	18

[2] Peter Sharples and Sifa Fukofuka took spill samples from seven sets from March 23 to 27, 2008, on a vessel belonging to the RD Fishing Co.

The location of the paired samples is shown in Figure 1. The samples have been taken primarily in the waters of Papua New Guinea, with those on unassociated schools located somewhat to the north of the majority of those on associated schools.

Figure 1. Location of sets from which paired spill and grab samples were collected, 2008–2012

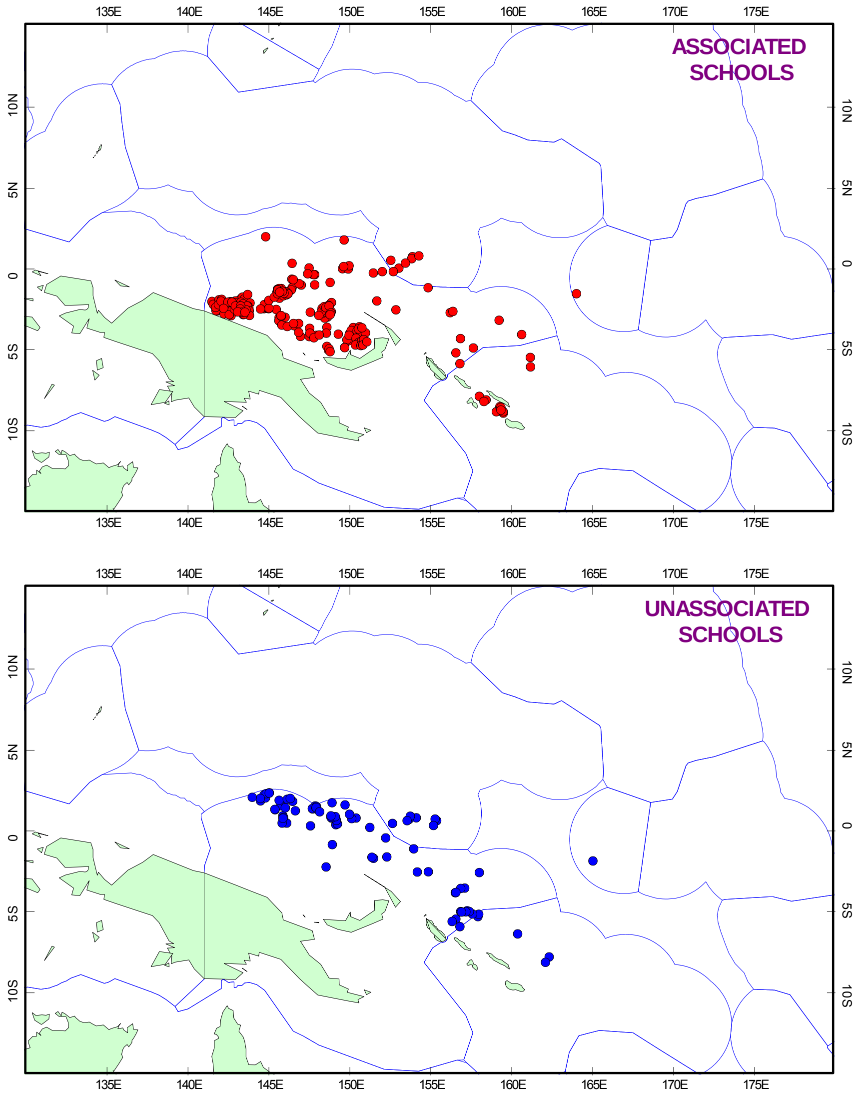
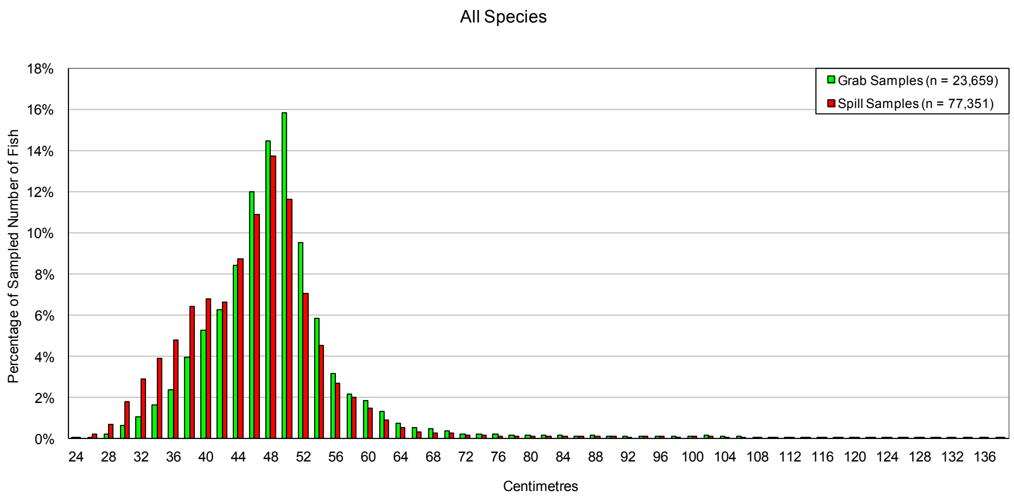
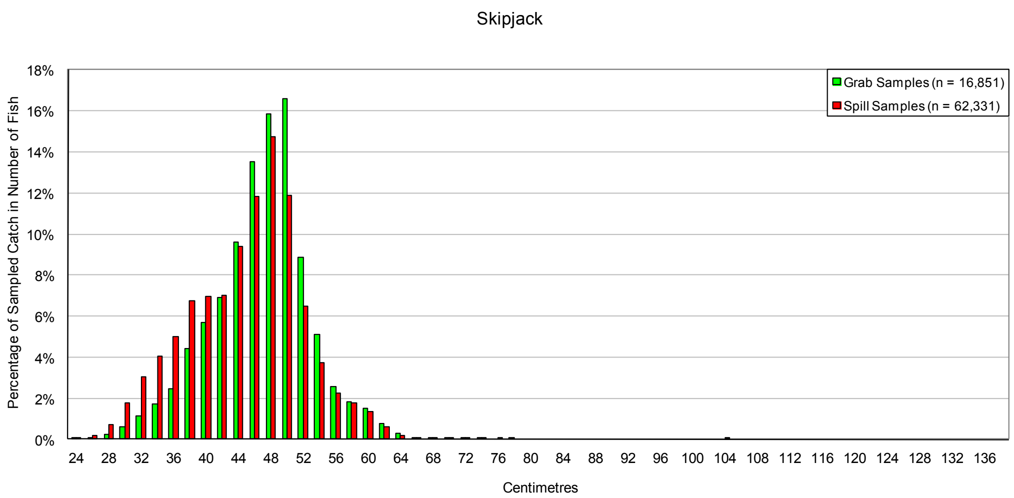


Figure 2 presents length frequencies (in terms of numbers of fish, rather than weight) determined from all 348 sets; the length frequencies for each set have been raised by the set weight. The length frequency at the top of Figure 2 is for skipjack, yellowfin and bigeye combined, and shows that for lengths less than 46 cm, there are greater proportions of fish in the spill samples than in the grab samples and vice versa for fish greater than or equal to 46 cm. Assuming that length frequencies determined from spill samples are unbiased, this is a clear indication of a size selection bias in the grab samples (although the bias may not necessarily be the result of non-random sampling by the grab samplers, as will be discussed below). The same pattern is observed in the length frequencies for each species separately, although the length at which the change occurs differs slightly. For bigeye, the pattern is somewhat less distinct because of the smaller number of fish in the samples. For all fish greater than about 70 cm, the pattern is indistinct for the same reason.

Figure 2. Length frequencies in terms of number of fish, determined from paired grab samples and spill samples





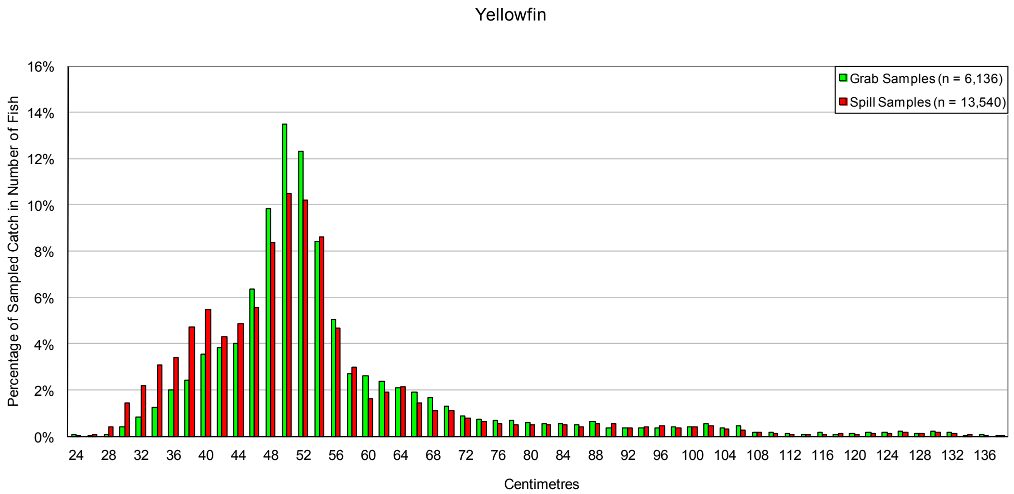
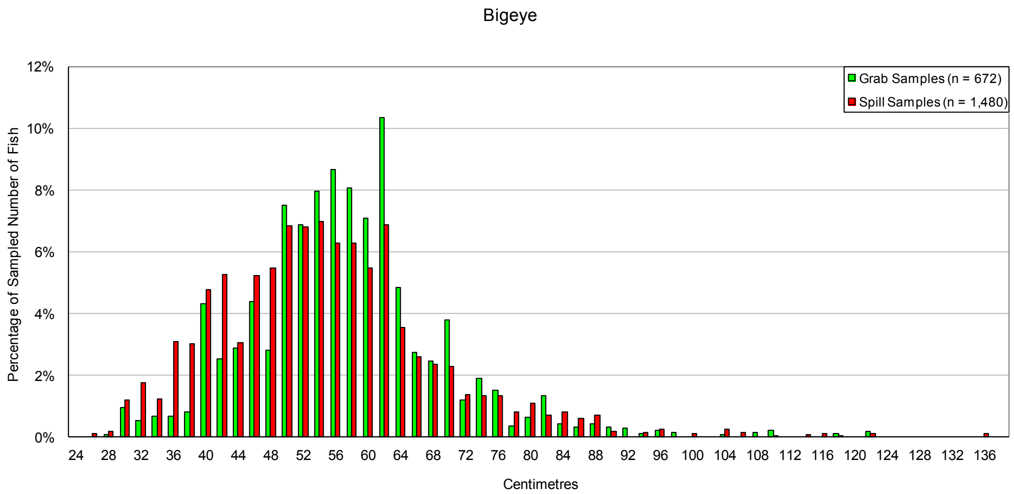


Figure 2 (continued)



Estimation of the selectivity bias in grab samples

The selectivity bias was estimated from the paired grab and spill samples using the model developed in Lawson (2009):

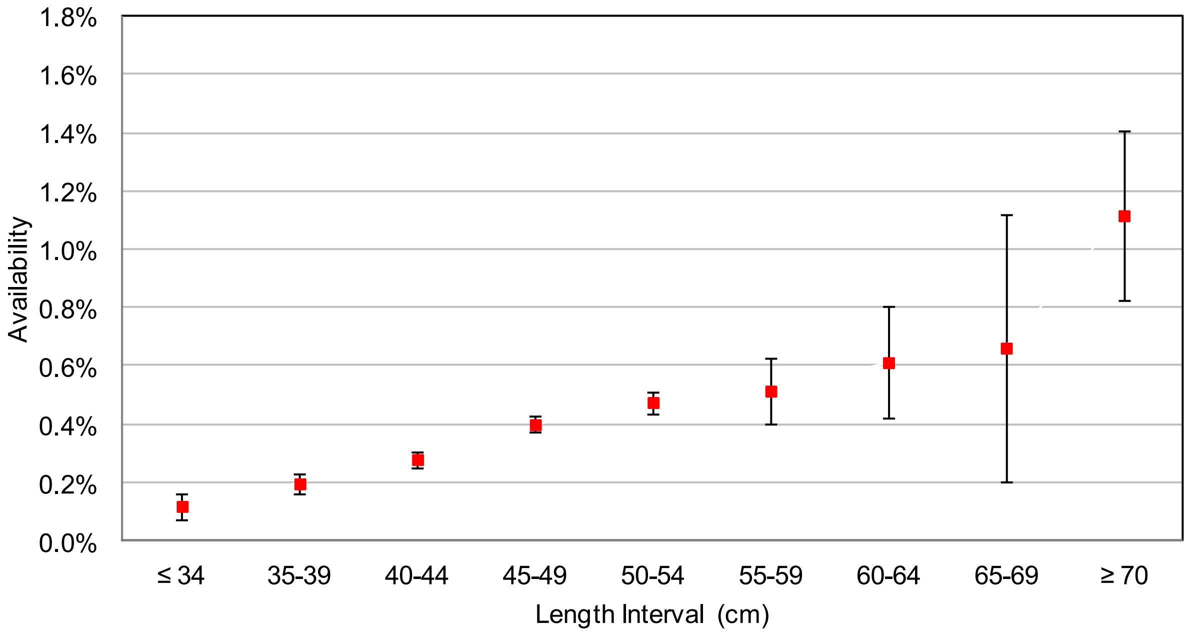
$$n_{jk} = N_{jk} \cdot A_j + \varepsilon \qquad (1)$$

$$= \frac{W_k \cdot T_{jk}}{\overline{w}_j} \cdot A_j + \varepsilon \qquad (2)$$

where n_{jk} is the number of fish in length interval j selected by a grab sampler from set k; N_{jk} is the "true" number of fish in length interval j in set k; A_j is the probability that a grab sampler will select a fish from among the N_{jk} fish, which can be considered as the *availability* of a fish to be selected; W_k is the total weight of set k; T_{jk} is the —true" proportion of fish of length interval *j* in set *k*, in terms of weight, determined from the spill sample taken from set k; $\overline{w}_j$ is the average weight of fish of length interval j; and ε is a random variable of mean zero. Note that when estimating *availability*, each length interval is considered independent and treated separately; the same approach is taken when correcting the historical grab samples with the estimates of *availability* using equations (5)–(10) below.

In Lawson (2010), the availability parameters, A_j in equations (1) and (2), were estimated for nine intervals of fish length: one interval for fish ≤ 34, seven intervals of 5 cm from 35 cm to 70 cm, and one interval for fish ≥ 70 cm. In the absence of selectivity bias, the A_j should be the same value for each length interval; Figure 3 shows that the estimates of availability increased with size, with smaller fish being under-selected relative to larger fish. However, the relationship was obscured by the wide error bars for fish ≥ 55 cm, which were due to the lack of sufficient data covering larger fish at that time.

Figure 3. Estimates of selectivity bias (availability) from Lawson (2010)



An artefact of estimating the selectivity bias for length intervals is that the length frequencies estimated from grab samples corrected for selectivity bias (Lawson 2011) are discontinuous at the boundaries of each interval. To allow for smoother length frequencies, the selectivity bias was estimated with a cubic spline.

Estimates of the parameters of the splines were obtained using the following relationship for data grouped by set and length interval, which is derived from equation (2):

$$\frac{n_{jk}}{\frac{W_k \cdot T_{jk}}{\overline{w}_j}} = A_j + \varepsilon \tag{3}$$

$$A_j = f\left(\overline{L}_{jk}, \beta\right) \tag{4}$$

where the left-hand side of equation (3) is determined for strata of length interval j in set k from the data, the function f is a cubic spline, $\overline{L}_{jk}$ is the average length in the stratum, and β is the vector of parameters to be estimated. The left-hand side of equation (3) is simply the number of fish in length interval j in the grab sample from set k, n_{jk}, divided by the ―true” number of fish in length interval j in set k; that is, the *availability*. The length of the vector β is equal to the degrees of freedom of the spline, which determines the complexity of the relationship between availability and length; the complexity increases with the degrees of freedom.

While equations (3) and (4) represent a continuous relationship between availability and length, rather than a step function between availability and length intervals, the model is fit to data points of availability and average length determined from length intervals. With the recently available data covering large fish from the additional unassociated schools, the length interval for fish ≥ 70 cm was replaced with three intervals — 70–79 cm, 80–99 cm and ≥ 100 cm — increasing the number of length intervals from nine to eleven; these intervals appear to be the most informative as determined by the confidence regions of the resulting estimates of availability.

The model was fit with values of the degrees of freedom ranging from 3 to 20; the model that minimised the Bayesian Information Criterion (BIC, Schwarz 1978) had 3 degrees of freedom and explained 24.6% of the deviance. See Appendix I for a summary of the fit of the model. Fitted availability is plotted with observed availability in Figure 4. The horizontal striations correspond to the length intervals, with the lowest band corresponding to the smallest length interval. Note that for all length intervals, there is a wide distribution of observed availability, although the distributions become more concentrated towards lower values of availability as the lengths of fish in the interval get smaller.

Figure 4. Observed vs fitted values of availability

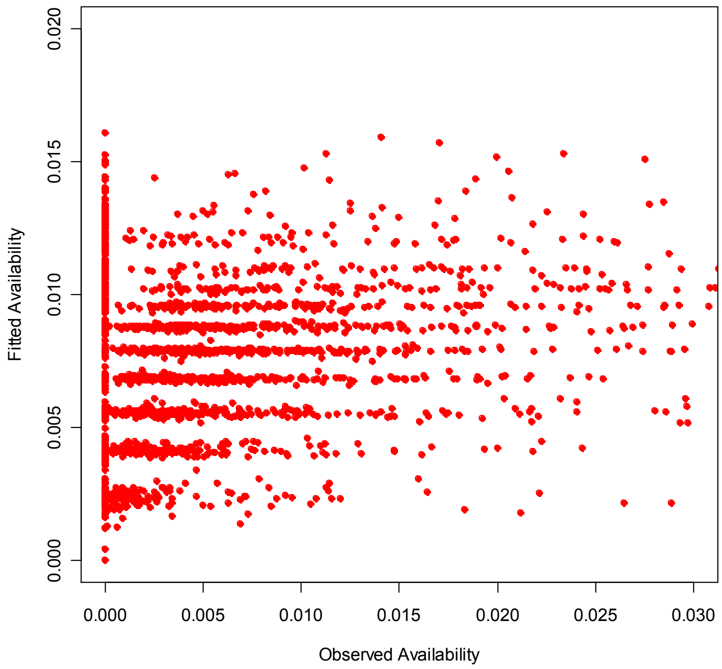
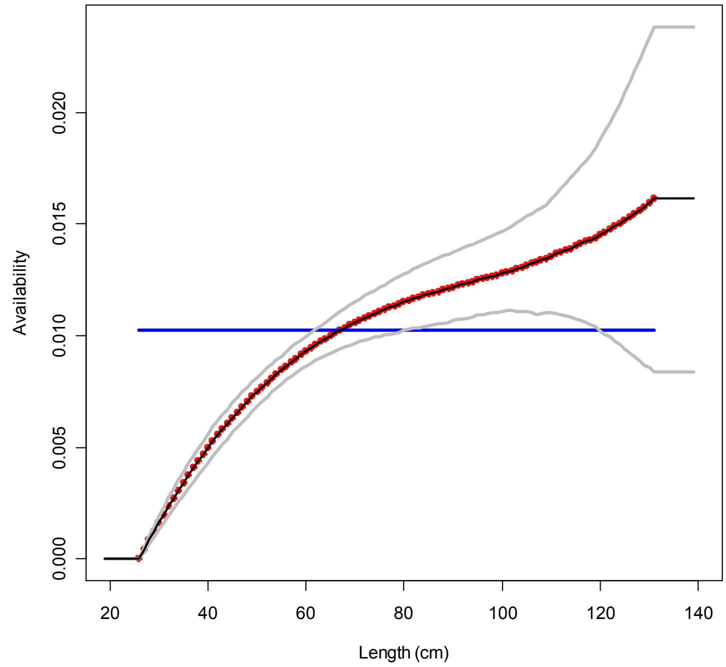


Figure 5 shows the fitted values (red dots) together with predicted values (black line) and upper and lower 95% confidence regions (grey lines). The fitted and predicted values of availability increase monotonically with length, increasing rapidly for smaller lengths, then less rapidly from about 60 cm to 100 cm, then somewhat more rapidly for larger lengths. The confidence regions are narrow for smaller fish, but increase progressively; the confidence regions for fish greater than about 110 cm are wide, which reflects the relative lack of data for large fish. The blue line in Figure 5 represents the average availability, 1.024%, which is reached at 67 cm.

Figure 5. Relationship between availability and length determined using a cubic spline



The absolute values of availability in Figure 5 are greater than those estimated in Lawson (2010). When nine length intervals are used (Figure 3), availability increases above 1% only for fish ≥ 70 cm. However, when splines are used, availability increases above 1% for fish ≥ 66 cm. In the latter case, availability increases faster for smaller fish, such that the availability for fish of 40 cm is about twice the level than when length intervals are used.

The discrepancy in the results for splines as compared to length intervals can be attributed to the noise in the data, as shown in Figure 4, such that a modification in the structure of the model may have a greater effect on the results where data have a high variance than otherwise. The results from using splines should be considered more accurate since the relationship is continuous and thus more highly resolved, compared to the step function obtained when length intervals are used. If so, the implication is that smaller fish are under-selected to a lesser extent than previously considered. Given the high level of noise in the data, the accuracy of the estimates of availability should increase as additional paired samples become available.

Selectivity bias and layering

Figure 5 indicates that observers have a tendency to over-select larger fish when taking grab samples. This is usually considered to be the result of observer behaviour; however, it can also occur as the result of layering in the set by size of fish. Consider the following thought experiment.

Imagine a set containing an equal weight of large yellowfin and small skipjack, such as a 100 tonne set containing 50 tonnes of yellowfin each weighting 20 kgs — that is, 2,500 fish — and 50 tonnes of skipjack each weighing 2 kgs — that is, 25,000 fish. Suppose also that there is complete layering in the set, such that all of the large yellowfin are brailed before all of the small skipjack. A brail typically contains about 5 tonnes of fish, so there will be 10 brails of yellowfin and 10 brails of skipjack. A brail of the yellowfin would include about 250 fish, while a brail of the skipjack would include about 2,500 fish. The protocol for grab sampling is to randomly select five fish per brail, so the observer will select a total of 50 yellowfin from 10 brails and 50 skipjack from 10 brails. The weight of the 50 yellowfin that were sampled is 1,000 kgs and the weight of the 50 skipjack is 100 kgs, so the species composition (in terms of weight) for the set will be estimated as 90.9% yellowfin and 9.1% skipjack, whereas the true species composition is 50% yellowfin and 50% skipjack. The problem is that, with layering by size and different average sizes for each of the species in the set, the probability of a fish being sampled — the availability — differs; for yellowfin the probability is 50 / 2,500 = 2.0%, while for skipjack the probability is 50 / 25,000 = 0.2%. Note that this result occurs even when the grab sampler *randomly* selects each fish.

This thought experiment is based on an extreme situation, but the effect will be similar for any level of layering by size and different average sizes for each of the species.

To determine the extent to which layering by size of fish occurs in grab samples, the sequential order in which fish are recorded on the PS–4 observer sampling form was examined. Only sets with at least a small amount of variation in size of fish were included; sets for which the coefficient of variation of the weight (kgs) of all fish in the sample (regardless of species) was less than 10% were therefore excluded. Also, sets for which there were less than 50 fish in the grab samples were excluded. Screening the data for these two conditions resulted in 16,678 sampled sets remaining for analysis, out of a total of 17,212, or 96.9% of all sets.

Observers are instructed to record species and length (cm) of the five fish grabbed from each brail in sequential order on the PS–4 observer form (Appendix II). However, the form has six columns in which data for 25 fish can be recorded, and rather than recording the fish in sequential order, observers sometimes record the data in columns according to species, e.g., all skipjack in column #1, all yellowfin in column #2, etc. Also, sometimes the observer may not have time to measure the fish between brails and so will put the selected fish to one side; when the observer finally measures the fish, they may be sorted by species first and so the data recorded on the form will not represent the sequential order of the fish as they were selected.

To identify sets for which the data were not recorded in sequential order, runs tests were conducted. A run is defined to be a maximal subsequence of like events; for example, the sequence SKJ, SKJ, SKJ, YFT, YFT, SKJ contains three runs, including maximal subsequences of three skipjack, two yellowfin and one skipjack. If observers record data by listing them in columns by species or after sorting them by species, then there will be a much smaller number of runs in the data than expected.

The runs tests for each of the 16,678 sets were conducted by (a) generating a probability distribution of the number of runs for the sample by simulating 1000 replicates of sampling, given the numbers of skipjack, yellowfin and bigeye in the sample, and (b) determining the cumulative probability of the number of runs in the actual data recorded by the observer from the generated probability distribution. Since 1000 replicates were simulated for each combination of the numbers of skipjack, yellowfin and bigeye in a set, the cumulative probabilities were multiples of 0.001. Visual inspection of the data on the PS–4 forms for those sets for which the cumulative probability was zero showed clearly that the data had not been recorded sequentially. Data for which the cumulative probability was 0.001 or 0.002 showed a small degree of non-sequential recording, possibly due to sorting of each of the five fish grabbed from each brail before recording of the data for the five fish on the PS–4 form. This level of sorting was considered acceptable given the purposes of the analysis, which was to determine the extent of layering throughout the brailing of the entire set. There were 2,852 sets that were excluded because the cumulative probability was zero, leaving 13,826 sets for analysis.

The sequentially-ordered data for each set were divided into ten quantiles of 10% of the number of fish sampled. For each set, the average size of fish (kg) was determined by species and for all three species combined. For each quantile in each set, the relative weight was determined; that is, the average weight of fish in the quantile divided by the average weight of fish in the set. Examining relative weights rather than absolute weights takes into account the variation in average weights among sets. The relative weights for each quantile were then averaged over all sets.

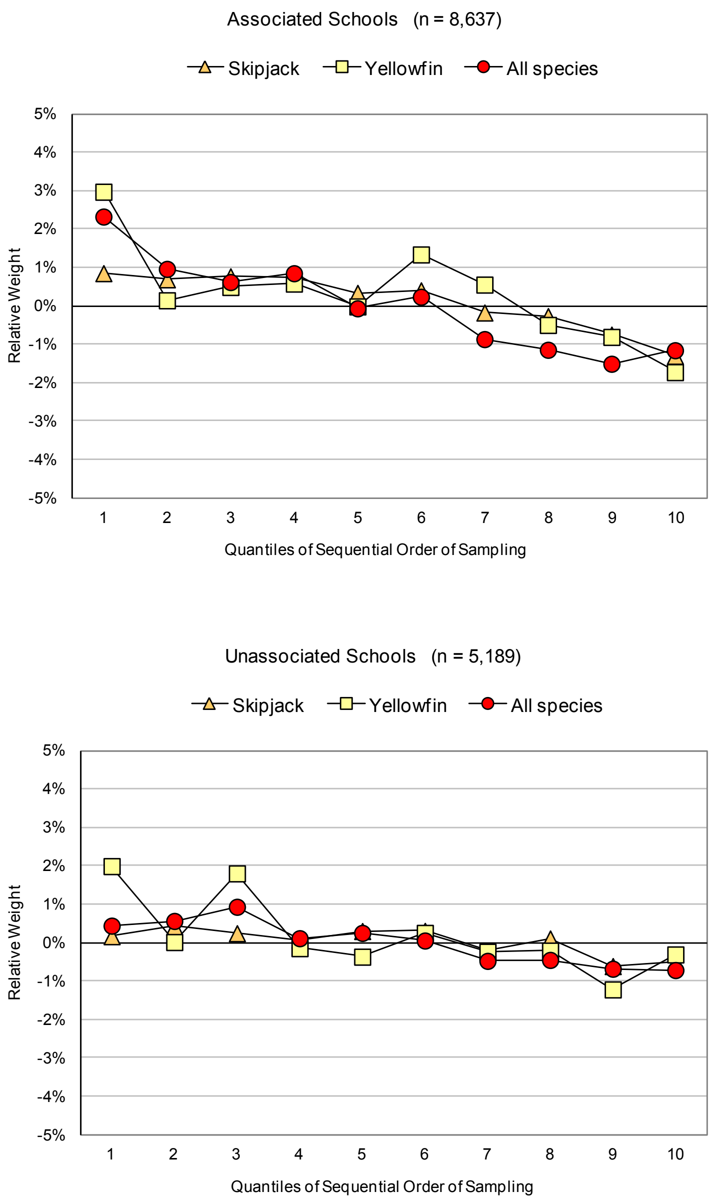
Figure 6 presents the results for associated and unassociated sets. The average relative weights are shown for skipjack, yellowfin and all three species combined; the average relative weights for bigeye are not shown separately since the number of bigeye in the samples was insufficient to obtain meaningful results. For both associated and unassociated sets, the relative weight tends to decline from the beginning of brailing to the end of brailing; the effect appears particularly strong for associated schools, declining, on average, from about +2% of the average weight of fish per set at the beginning of brailing to about –2% at the end.

This trend in relative weight during brailing suggests that layering is at least partly responsible for the selectivity bias. It remains to be determined the extent to which other factors — such as observer behaviour — may contribute to the bias.

Regarding the affect of layering on spill samples, it was noted above that the first brail chosen for a spill sample is rotated; thus, brail #1 is chosen as the first brail sampled from set #1, brail #2 as the first brail sampled from set #2, etc. Thereafter, if the set is large, every tenth brail is chosen for a spill sample. While the species composition for an individual set may be affected by layering, particularly if only one brail is chosen for a spill sample, the effect should be less important in species compositions determined from multiple spill samples, such as for a trip or for strata of time period and geographic area, due to averaging. It should also be noted that all fish in a spill sample

— or, more precisely, all kilograms of fish in a spill sample — have an equal probablility of being sampled, unlike a grab sample, for which large fish have a greater probability than small fish; thus, the effect of layering is less important for spill samples.

Figure 6. Average relative weight of fish in a set by sequential order of grab sampling by observers (see text)



Correction of historical grab samples for selectivity bias

The estimates of availability can be used to correct the species composition for an individual set determined from the grab samples as follows:

$$P_{ik} = \frac{\sum_j W_{ijk}}{\sum_i \sum_j W_{ijk}} \tag{5}$$

$$= \frac{\sum_j \left(N_{ijk} \cdot \overline{w}_{ij}\right)}{\sum_i \sum_j \left(N_{ijk} \cdot \overline{w}_{ij}\right)} \tag{6}$$

where P_{ik} is the proportion of species i in set k ; W_{ijk} is the weight of fish of species i and length j in set k ; N_{ijk} is the number of fish of species i and length j in set k ; and $\overline{w}_{ij}$ is the average weight of fish of species i and length j . From equation (1), we have

$$N_{ijk} = \frac{n_{ijk}}{A_j} \tag{7}$$

where n_{ijk} is the number of fish of species i and length j selected by a grab sampler from set k and A_j is the probability that a grab sampler will select a fish from among the N_{ijk} fish. Substituting equation (7) into equation (6), we obtain

$$\hat{P}_{ik} = \frac{\sum_j \left(\frac{n_{ijk}}{A_j} \cdot \overline{w}_{ij}\right)}{\sum_i \sum_j \left(\frac{n_{ijk}}{A_j} \cdot \overline{w}_{ij}\right)} \tag{8}$$

$$= \frac{\sum_j \frac{w_{ijk}}{A_j}}{\sum_i \sum_j \frac{w_{ijk}}{A_j}} \tag{9}$$

$$\overline{w}_{ij} = a_{ij} \cdot j^{b_{ij}} \tag{10}$$

where $\widehat{P}_{ik}$ is the estimated proportion of species i in set k and w_{ijk} is the weight of fish of species i and length j selected by a grab sampler from set k . The average weights, $\overline{w}_{ij}$, are determined from the length and the length-weight parameters a and b given in the Introduction.

Similar to the approach taken when estimating *availability*, the correction for each length j is considered independent and treated separately. The magnitude of the correction of the species composition for a particular set — that is, for all lengths combined — depends on the range of lengths of fish in the grab samples for the set. When the fish are all the same size, the *availability*, A_j, will be the same for each species and the magnitude of the correction to the species composition of the set will therefore be zero. With a greater range of sizes, the magnitude of the correction increases. Hence, even though lengths are considered independent and treated separately when estimating *availability* and correcting the grab samples, the effect of the range of lengths is still captured in the corrected species composition for each set.

Equation (8) and the estimates of availability (Figure 5) were used to correct the species compositions in grab samples collected during 1993–2012. Prior to correcting the grab samples, they were screened for data quality. Before screening, there were 55,651 sets (excluding those from which a spill sample was also taken). The results of the screening were as follows:

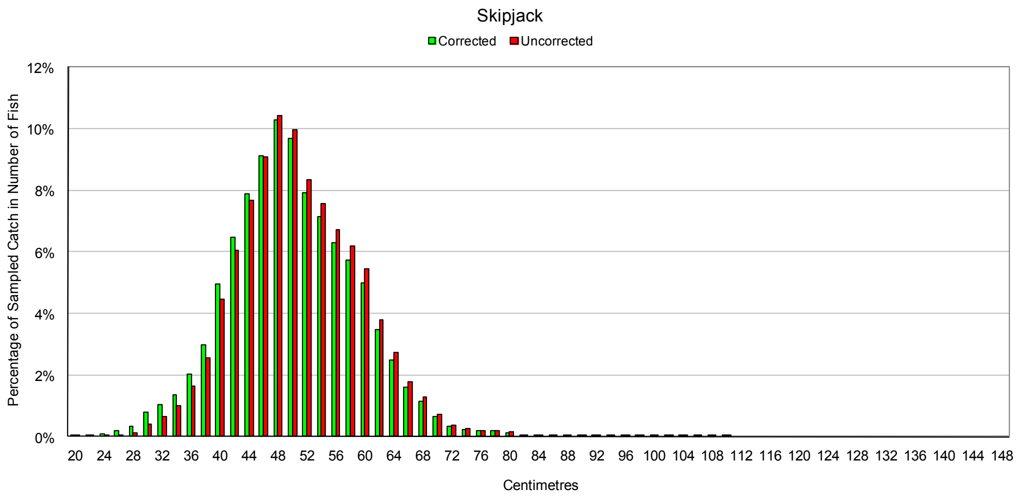
- The OFP has manually checked all samples collected during 1993–2002; there were 2,097 sets excluded because of poor data quality during this period. [3]
- The quality of data collected during an observer's first trip after completing the observer training course is typically inferior to that from subsequent trips; for data collected during 2003–2012, there were 6,202 sets excluded because they were collected during the observer's first trip.
- 1,676 sets were excluded because the school association was not recorded.
- 1,278 sets were excluded because the set was a skunk set, i.e., with a catch less than 2.5 tonnes.
- 707 sets were excluded because the sample size was too small, given the catch per set.

A total of 11,960 sets (21.5%) were excluded, leaving 43,601 sets. Together with 370 sets for which spill samples were taken, there was a total of 43,971 sets available for analyses of the species composition.

Figure 7 presents the length frequencies (in terms of number of fish) determined from the corrected and uncorrected grab samples. As expected, the length frequencies from the corrected samples contain more smaller fish, and less larger fish, than those from the uncorrected samples.

[3] Data for 1993 and 1994 were not used in Lawson (2010) because they had not been manually checked for data quality; these data have since been checked.

Figure 7. Length frequencies determined from grab samples, corrected and uncorrected for selectivity bias



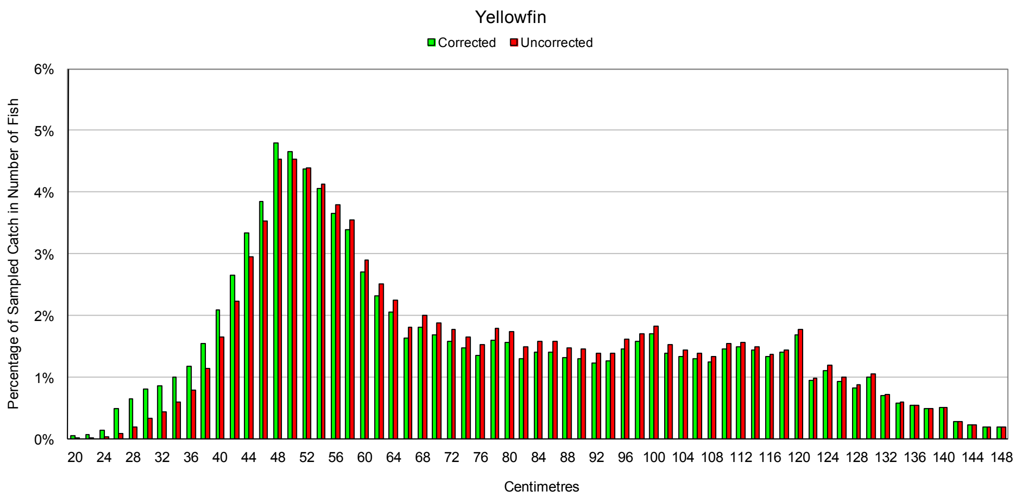
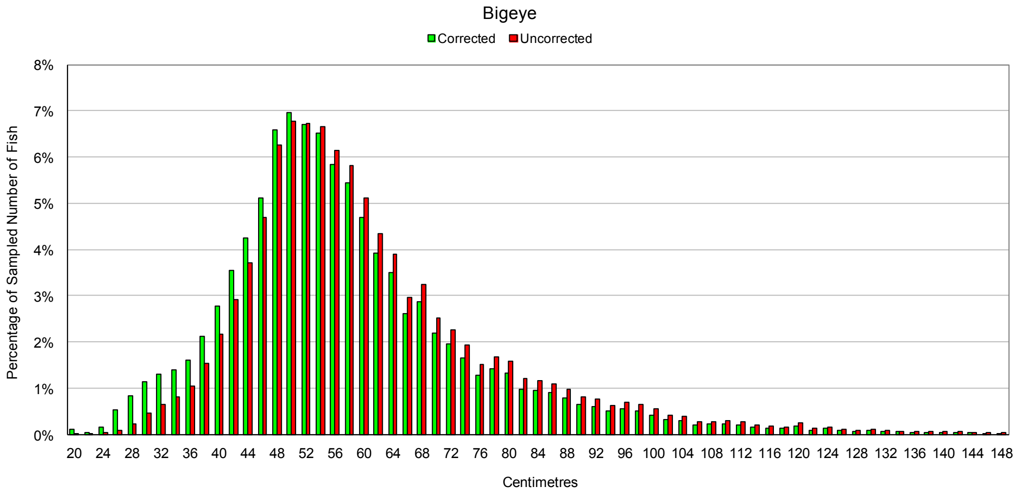


Figure 7 (continued)



Coverage of observer samples, 1993–2011

For unbiased estimates of the species composition, the time-area distribution of the grab samples collected by observers should be representative of the time-area distribution of the catch and fishing effort, either for the region as a whole or, if data are collected from and catches estimated for strata of area, within strata. Figure 8 compares the annual geographic distributions of days fished by all purse seiners and days monitored by observers. Until recently, observers have been placed onboard vessels purely on an opportunistic basis, and this is evident in Figure 8. Ignoring the small number of observed days during 1993–1994, the geographic distribution of observed effort is only somewhat different from that of the fleet as a whole for 1995–2001. However, they are considerably different for 2002–2009, primarily due to increased observer coverage in Papua New Guinea. With increased coverage in 2010 following the implementation of WCPFC Conservation and Management Measure (CMM) 2008–01, which increased the target rate of observer coverage to 100% as of the beginning of 2010, the distribution of observed effort is similar to the distribution of days fished by the fleet.

Figure 8. Distribution of purse-seine days fished and days observed

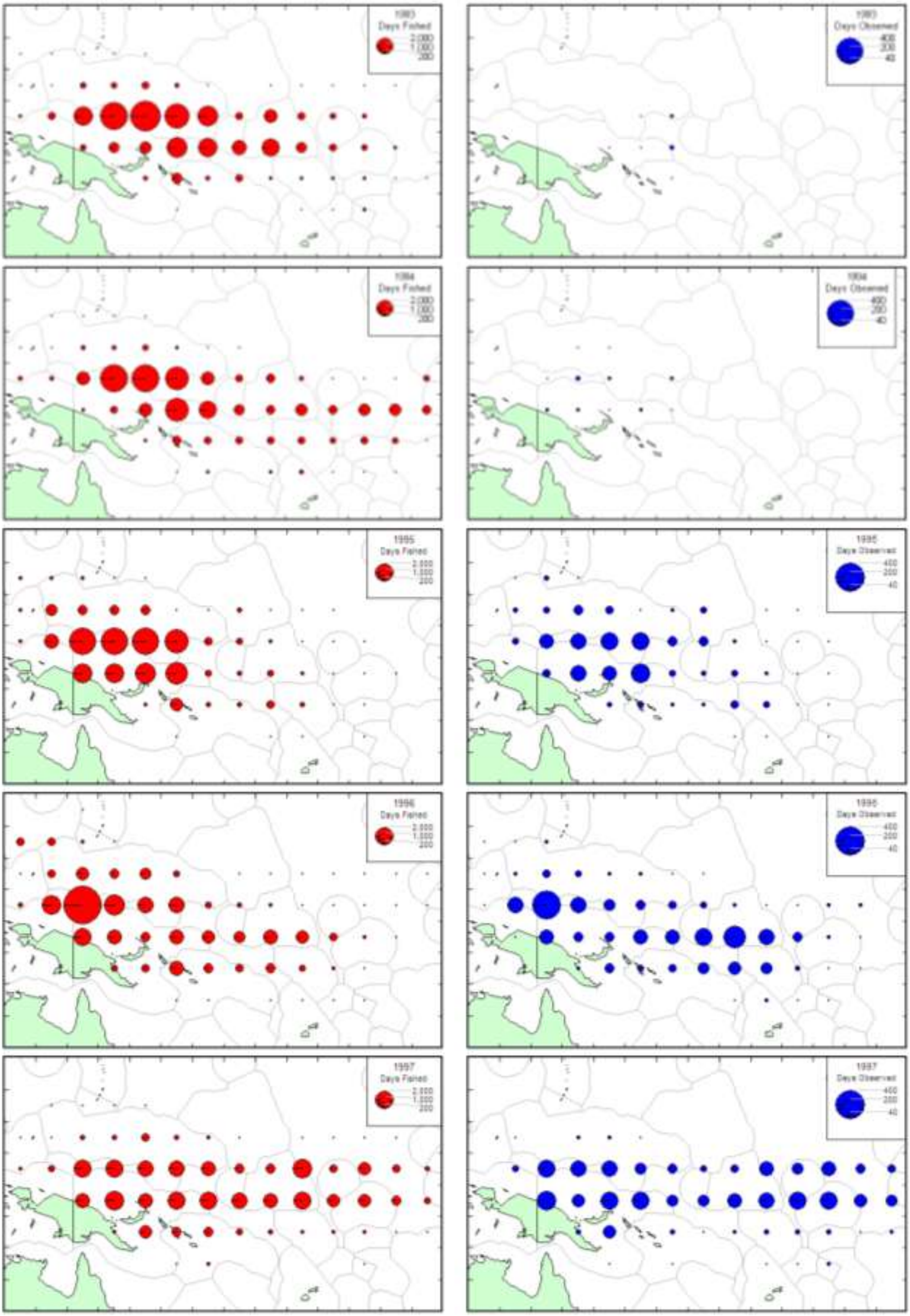


Figure 8 (continued)

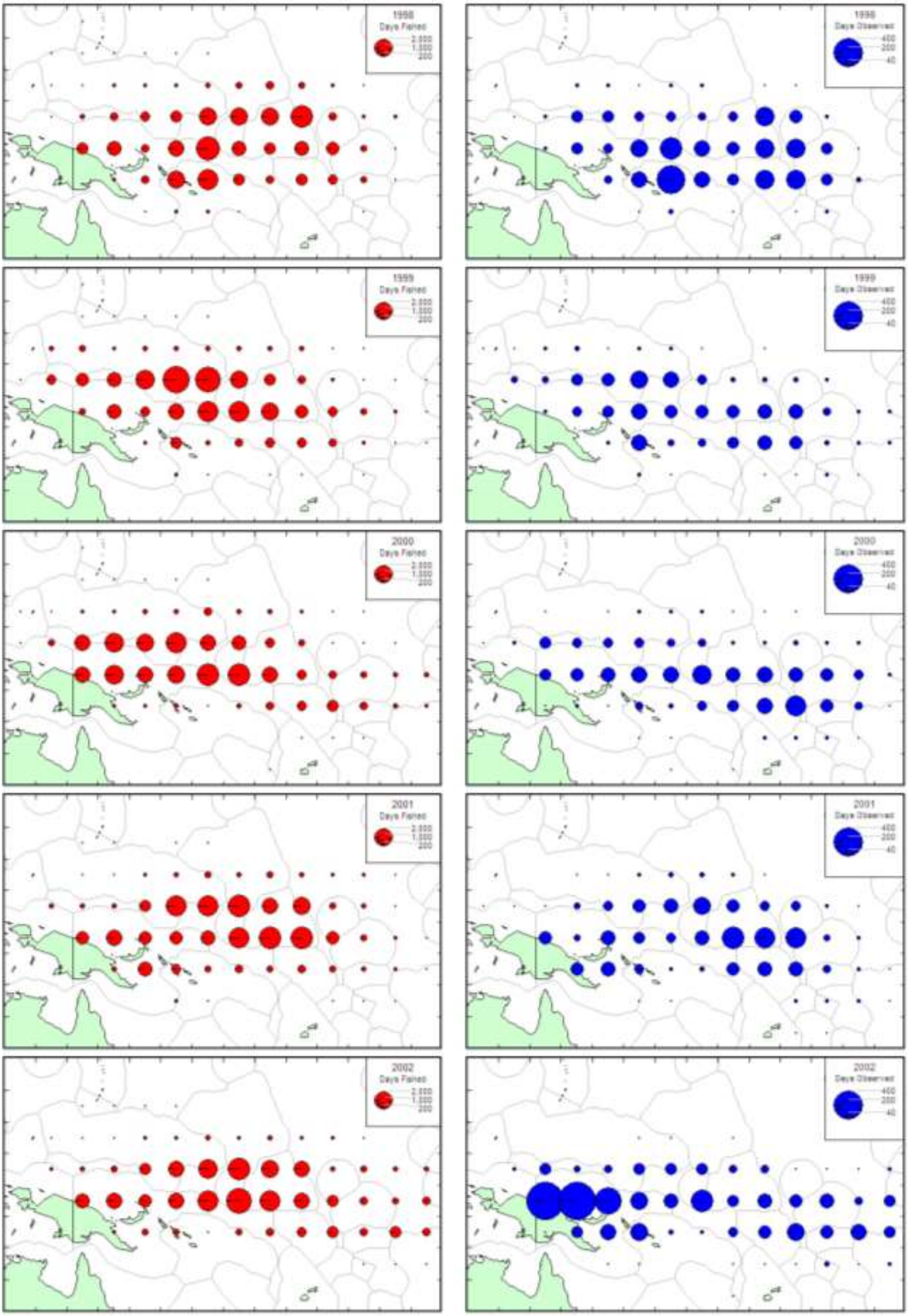


Figure 8 (continued)

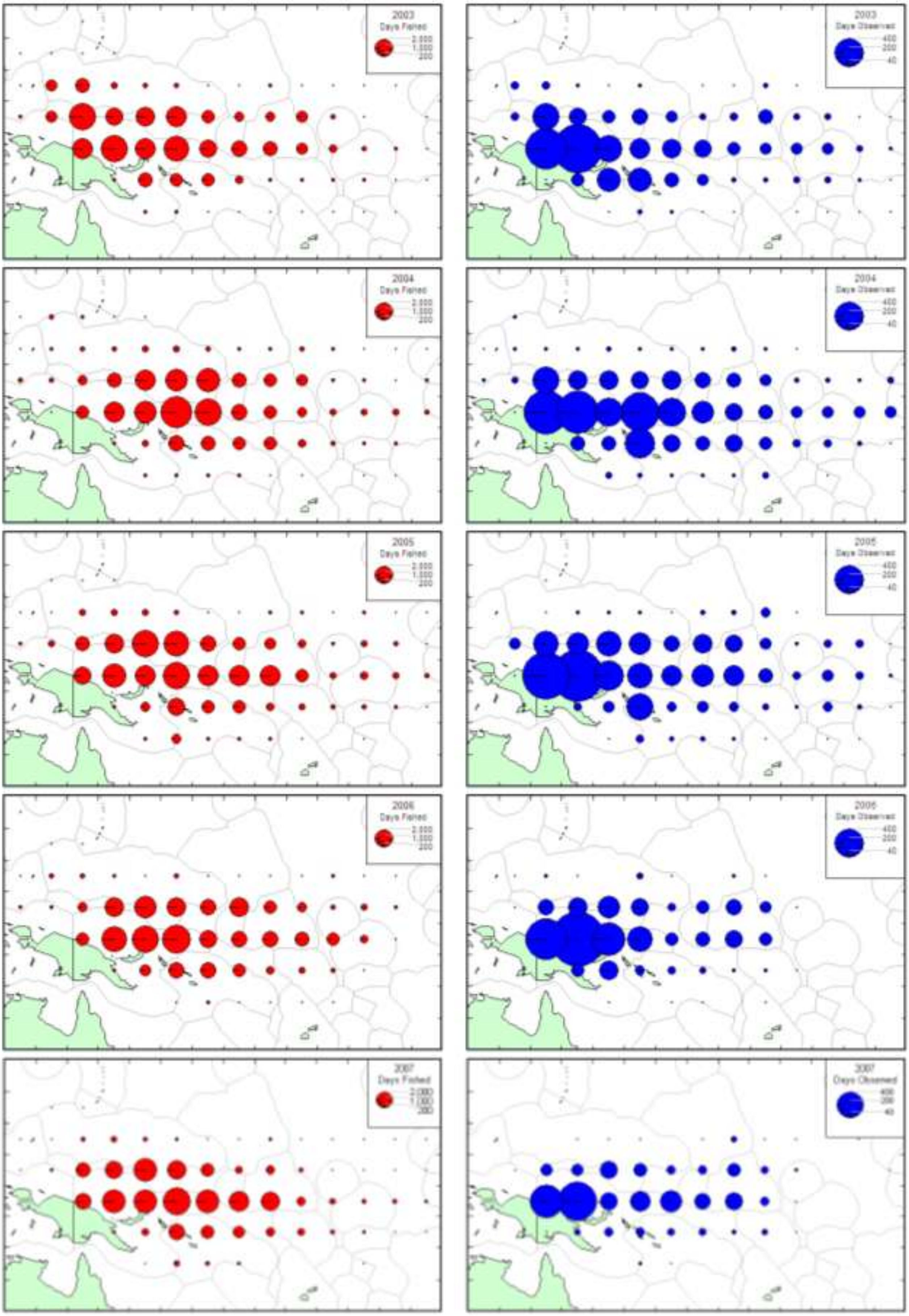
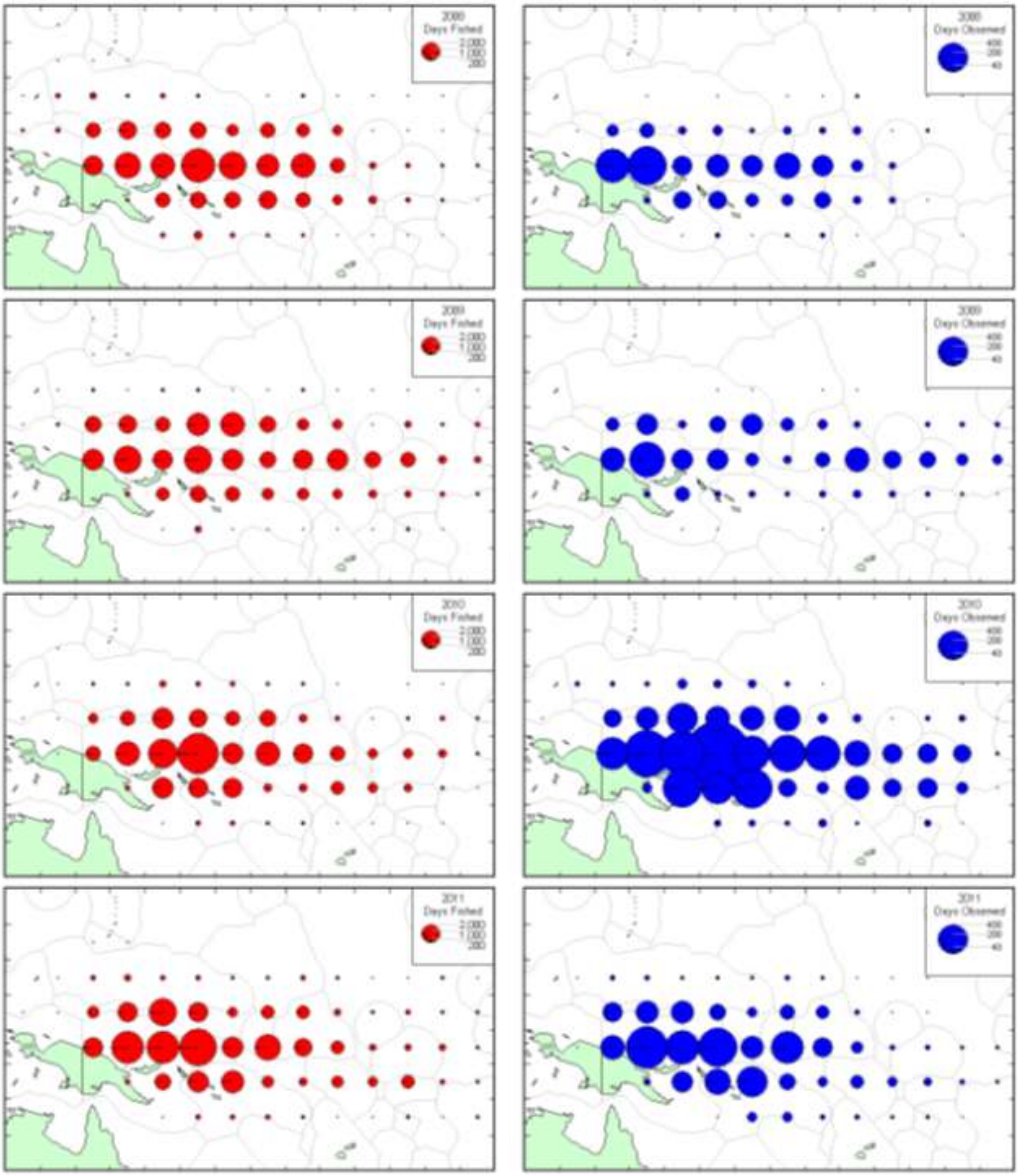


Figure 8 (continued)



The number of sets sampled by observers are summarised by vessel flag in Table 2. Coverage is dominated by the fleets of the United States (18.1%) and, since 2002, Papua New Guinea (31.3%).

Table 2. Number of sets sampled by observers, by vessel flag

From left to right: China, Ecuador, Spain, Federated States of Micronesia, Indonesia, Japan, Kiribati, Korea, Marshall Islands, New Zealand, Papua New Guinea, Philippines, Solomon Islands, San Salvador, Tuvalu, Chinese Taipei, United States of America and Vanuatu

Year	CN	EC	ES	FM	ID	JP	KI	KR	MH	NZ	PG	PH	SB	SV	TV	TW	US	VU	Total	%
1993	0	0	0	0	0	24	0	5	0	0	0	0	0	0	0	0	0	0	29	0.1
1994	0	0	0	0	0	18	0	19	0	0	0	0	0	0	0	23	0	0	60	0.1
1995	0	0	0	21	0	14	0	14	0	0	8	0	0	0	0	32	0	0	89	0.2
1996	0	0	0	8	0	13	0	90	0	0	8	17	0	0	0	228	0	12	376	0.9
1997	0	0	0	0	0	23	0	74	0	0	32	45	0	0	0	141	3	15	333	0.8
1998	0	0	0	5	0	28	0	154	0	0	86	0	26	0	0	393	190	16	898	2.0
1999	0	0	0	0	0	14	5	75	0	0	33	15	67	0	0	67	224	9	509	1.2
2000	0	0	0	28	0	29	6	78	0	0	66	0	15	0	0	112	333	0	667	1.5
2001	0	0	0	19	0	62	0	59	22	0	73	54	12	0	0	122	516	0	939	2.1
2002	0	0	0	62	0	50	28	22	10	0	699	373	186	0	0	74	402	0	1,906	4.3
2003	0	0	0	54	0	89	31	21	105	0	908	291	82	0	0	60	249	0	1,890	4.3
2004	0	0	0	74	0	60	0	192	126	22	1,388	433	108	0	0	166	189	84	2,842	6.5
2005	57	0	0	50	0	69	0	119	147	29	1,634	511	62	0	0	209	138	102	3,127	7.1
2006	35	0	0	97	0	62	42	217	304	2	1,891	543	0	0	0	152	95	55	3,495	7.9
2007	59	0	0	42	0	48	20	206	348	0	1,636	384	48	0	0	236	167	185	3,379	7.7
2008	24	0	56	114	0	55	35	192	206	26	1,191	469	123	0	0	208	689	42	3,430	7.8
2009	32	0	2	87	0	196	32	117	197	31	1,126	273	0	0	0	160	1,111	31	3,395	7.7
2010	490	26	158	361	14	1,081	46	1,530	378	0	2,084	585	0	24	146	1,041	2,730	155	10,849	24.7
2011	376	0	0	122	0	847	76	626	508	0	911	372	11	0	1	695	926	139	5,610	12.8
2012	0	0	0	0	0	109	0	30	14	0	0	0	0	0	0	0	0	0	153	0.3
Total	1,073	26	216	1,144	14	2,891	321	3,840	2,365	110	13,774	4,365	740	24	147	4,119	7,962	845	43,976	100.0
%	2.4	0.1	0.5	2.6	0.0	6.6	0.7	8.7	5.4	0.3	31.3	9.9	1.7	0.1	0.3	9.4	18.1	1.9	100.0	

Table 3 presents observer coverage of the catch in MULTIFAN-CL (MFCL) Skipjack Areas 2 and 3 (Figure 9), excluding the domestic fisheries of Indonesia and the Philippines, and subject to the screening for data quality discussed above. From 1993 to 1995, the coverage rate was less than 1%, and from 1996 to 1997, it was less than 2%. Coverage then increased to 3.8% in 1998. From 2009 to 2010, coverage increased considerably, from 10.4% to 38.1%, as a result of CMM 2008–01; coverage for 2010 and 2011 should increase further as additional data are provided.

Figure 9. MFCL Skipjack Areas 2 and 3

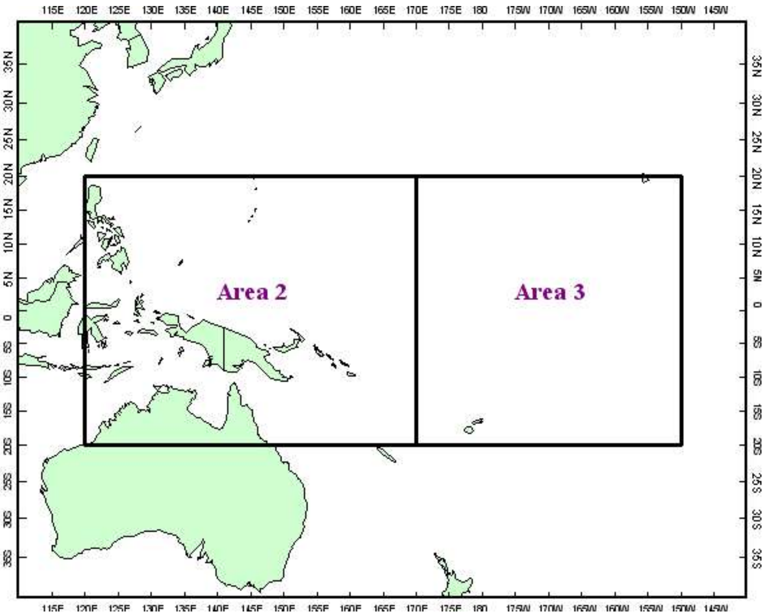


Table 3. Observer coverage of the catch (tonnes) in MFCL Skipjack Areas 2 and 3

Year	Associated Sets			Unassociated Sets			Total		
	Observed Catch	Total Catch	%	Observed Catch	Total Catch	%	Observed Catch	Total Catch	%
1993	411	354,480	0.1%	515	364,751	0.1%	926	719,231	0.1%
1994	1,462	397,856	0.4%	1,379	417,175	0.3%	2,841	815,031	0.3%
1995	3,890	336,877	1.2%	1,381	417,945	0.3%	5,271	754,822	0.7%
1996	6,828	427,167	1.6%	4,540	294,635	1.5%	11,368	721,802	1.6%
1997	9,310	442,607	2.1%	2,490	300,576	0.8%	11,800	743,183	1.6%
1998	22,330	510,380	4.4%	14,878	481,426	3.1%	37,208	991,806	3.8%
1999	20,556	630,203	3.3%	6,100	228,760	2.7%	26,656	858,963	3.1%
2000	24,270	503,407	4.8%	9,412	396,196	2.4%	33,682	899,603	3.7%
2001	23,532	408,205	5.8%	16,062	485,277	3.3%	39,594	893,482	4.4%
2002	49,327	559,821	8.8%	19,413	477,571	4.1%	68,740	1,037,392	6.6%
2003	48,241	463,941	10.4%	21,752	522,270	4.2%	69,993	986,211	7.1%
2004	86,988	798,712	10.9%	17,419	260,066	6.7%	104,407	1,058,778	9.9%
2005	77,436	637,005	12.2%	39,817	543,588	7.3%	117,253	1,180,593	9.9%
2006	108,855	786,487	13.8%	35,175	419,306	8.4%	144,030	1,205,793	11.9%
2007	107,729	773,489	13.9%	62,448	547,270	11.4%	170,177	1,320,759	12.9%
2008	93,090	748,763	12.4%	64,201	618,643	10.4%	157,291	1,367,406	11.5%
2009	94,670	923,976	10.2%	63,290	591,250	10.7%	157,960	1,515,226	10.4%
2010	231,014	572,923	40.3%	332,899	905,676	36.8%	563,913	1,478,599	38.1%
2011	169,524	841,669	20.1%	90,673	557,204	16.3%	260,197	1,398,873	18.6%
Total	1,179,463	11,117,968	10.6%	803,844	8,829,585	9.1%	1,983,307	19,947,553	9.9%

Estimation of the species composition from grab samples corrected for selectivity bias and spill samples

The 'two model' structure used in Lawson (2010)

Lawson (2010) used two models of the following form to predict the species composition of catches in the strata of year – quarter – area – school association that were not covered by samples collected by observers:

$$p_{ijkl}^{SKJ} = \beta_{Intercept}^{SKJ} + \beta_i^{SKJ} \cdot YY_i + \beta_j^{SKJ} \cdot QQ_j + \beta_k^{SKJ} \cdot AR_k + \beta_l^{SKJ} \cdot AS_l \quad (11)$$

$$p_{ijkl}^{YFT} = \beta_{Intercept}^{YFT} + \beta_i^{YFT} \cdot YY_i + \beta_j^{YFT} \cdot QQ_j + \beta_k^{YFT} \cdot AR_k + \beta_l^{YFT} \cdot AS_l \quad (12)$$

$$p_{ijkl}^{BET} = \beta_{Intercept}^{BET} + \beta_i^{BET} \cdot YY_i + \beta_j^{BET} \cdot QQ_j + \beta_k^{BET} \cdot AR_k + \beta_l^{BET} \cdot AS_l \quad (13)$$

where the proportion of each species — e.g., p_{ijkl}^{SKJ} , the proportion of skipjack in year *i*, quarter *j*, area *k* and school association *l* — is predicted as a function of year (YY_i), quarter (QQ_j), area (AR_k) and school association (AS_l). The *area* covariate has a low geographic resolution, with only two values, MFCL Skipjack Areas 2 and 3. The *school association* covariate has two values, representing associated schools and unassociated schools. The catches by species for each observed set — determined either from grab samples corrected for selectivity bias or spill samples — were summed over all observed sets within each stratum of *year – quarter – area – school association*, to give the total observed catch by species for each stratum. The proportions of each species in the catch were determined for each stratum and used as replicates to fit the model parameters, the β's. Each replicate was weighted by the number of sets for which the observed catches in the stratum were summed.

The models given by equations (11)–(13), with all first order interactions among all the covariates, except for the year : area interaction, were used to estimate the species composition for the period 1996–2009; the *year : area* interaction was excluded because of the lack of observer data for Area 3 prior to 1998.

The model given by equations (11)–(13) was not used to estimate the species composition for years prior to 1996 because of either zero observer coverage or a low level of coverage. Instead, a model without the year covariate was also fit to the replicates and used to estimate the species composition for 1967–1995; the model without the year covariate included all first order interactions among *quarter*, *area* and *school association*.

Models of the species composition in the form of equations (11)–(13), have the useful property that the predicted proportions of each species in a stratum always sum to unity when the same set of replicates — the model matrix — is used to fit each of the three equations. That is, if the same set of replicates are used to fit equation (11) for skipjack, equation (12) for yellowfin and equation (13) for bigeye — with or without a *year* covariate — then the predicted species composition for a given stratum will sum to unity. Such models are more statistically rigorous than the usual practice of *substituting* or *borrowing* the species compositions from neighbouring strata into strata with insufficient sampling data (e.g., Pianet et al., 2000). In fact, substituting from neighbouring strata can be considered as a less than rigorous method to predict the species composition using interactions among covariates. The model used here potentially allows for prediction of the species composition using information contained in the data in regard to the effect of *all covariates* and *all of their first order interactions*.

That the predicted proportions sum to unity is also useful when there is a need to maintain consistency between the estimated catches by species and the total catch for a given stratum of time period and geographic area. For example, the ‗s_best' catch and effort database maintained by the OFP — which is stratified by year, month, 1° x 1°, school association and vessel flag — is used for many purposes, such as determining the total catch in the MFCL Skipjack Areas. If the predicted proportions by species for a MFCL Skipjack Area in a given time period did not sum to unity, then in order for the sum of the estimated catches by species to be consistent with the total catch determined from ‗s_best', the predicted proportions by species would have to be ‗normalised' — that is, forced to sum to unity — which would be a less than rigorous procedure.

In equations (11)–(13), each of the covariates is categorical. As will be shown in the sections below on models with a high geographic resolution (determined from a two-dimensional spline of latitude and longitude), this property of the predicted species composition summing to unity also applies to models with continuous covariates.

Case A: a 'three model' structure with low geographic resolution

In Lawson (2010), the model used to estimate the species composition for 1996–2009 did not include the *year : area* interaction because the grab samples prior to 1998 only covered one of the two MFCL areas. (That is, an interaction can only be included if there is at least one datum available covering all combinations of the two covariates, otherwise the fitted model is indeterminate.) However, the lack of data for MFCL Skipjack Area 3 prior to 1998 is an indication of the more general problem of low coverage of the observer data in the early years of the time series; it was noted above that the observer coverage from 1993 to 1995 was less than 1%, and from 1996 to 1997, less than 2% (Table 3). A somewhat different approach has therefore been taken when developing species composition models in this study. For the period 1993–1995, it was considered that the observer coverage is inadequate for reliably estimating the species composition using a model that includes a *year* covariate; the species composition for this period was therefore estimated using a model without a *year* covariate, just as for the period 1967–1992, for which no observer data are available. For the period 1996–2001, for which the coverage ranges from 3.1% to 4.4%, it was considered that the species composition could be estimated using a model that includes a *year* covariate, but without including the *year* interactions terms, i.e., *year : quarter*, *year : area* and *year : school association*. For the period 2002–2011, during which the coverage increased from 6.6% in 2002 to 9.9% in 2005 and thereafter remained above 10%, the species composition was estimated using a model that includes a *year* covariate and all interactions.

Fits of the 'low resolution' models

The Case A models were fit to the observed species compositions for 212 weighted strata of *year – quarter – area – school association*. With three models and three species, there were nine fits to the observer data. As an example of the fit to the model with a *year* covariate and the *year* interactions, Appendix III gives a summary of the fit to the model of the proportion of skipjack.

Table 4 presents the deviance explained by the three models for each of the three species. The model without *year* does poorly, while the model with *year* and the *year* interactions does well. The model with *year* but without the *year* interactions does less than half as well as the model with *year* and the *year* interactions, which indicates the importance of the *year* interaction terms in explaining the species composition.

Table 4. Deviance explained by the three 'low resolution' models of the species composition

Model	Deviance explained		
	Skipjack	Yellowfin	Bigeye
1. Without *year*	13.9%	12.4%	23.7%
2. With *year* and without *year* interactions	31.8%	34.1%	35.9%
3. With *year* and with *year* interactions	71.1%	68.7%	75.2%

A Gaussian distribution of the errors was assumed in the fits; to check whether this assumption was appropriate, quantile-quantile plots of the residuals were examined (Figure 10). The linearity of the plots indicates that the assumption of normality is appropriate for the model with the *year* covariate and *year* interactions, but less so for the other models; other assumptions regarding the distribution of the errors were examined, but with no improvement. Figure 11 plots the observed proportions against the fitted proportions for the model with the *year* covariate and *year* interactions, and shows that the fit is relatively good.

The species composition of the catches in MFCL Skipjack Areas 2 and 3, and the annual catches by species, determined from the Case A models, are compared to those of the other cases following the presentation of those models below.

Figure 10. Q-Q plots for the models in Case A

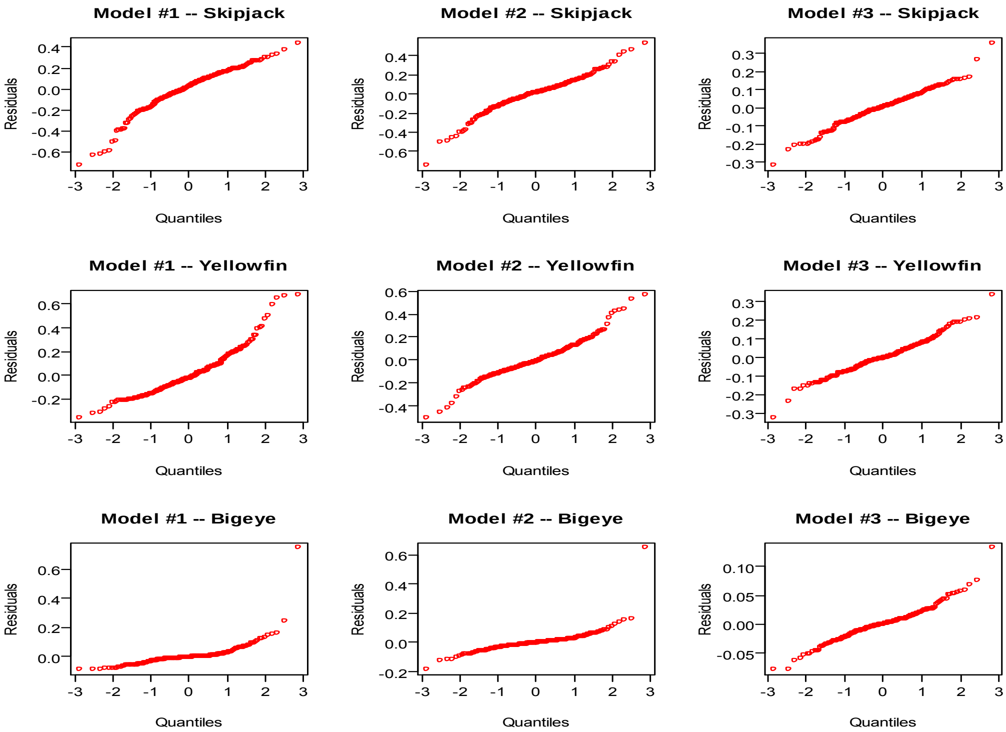
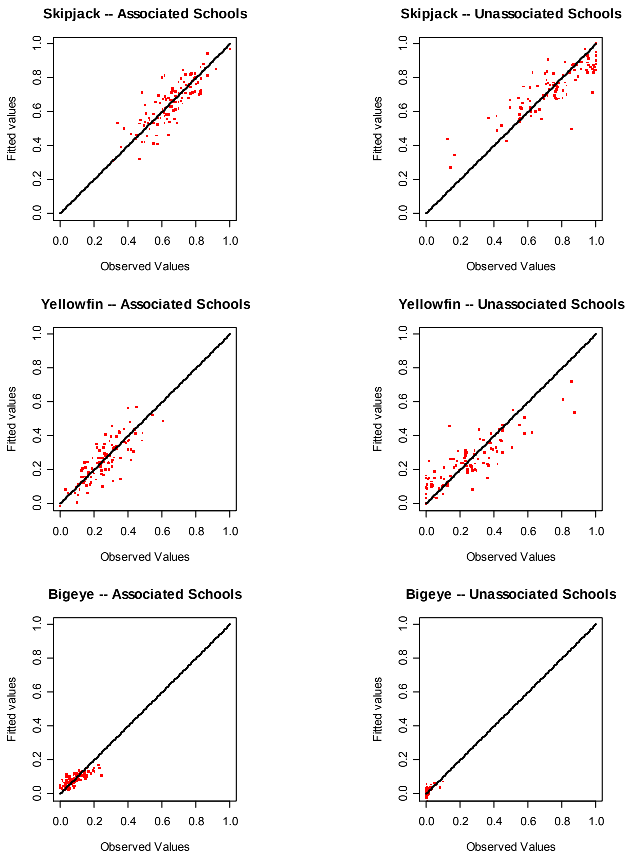


Figure 11. Observed vs fitted values of the proportion by species for the Case A models with the *year* covariate and *year* interactions



Cases B and C: a ‘two model’ and ‘three model’ structure with high geographic resolution

The low resolution models, in which the species composition is predicted on the basis of categorical covariates, are statistically rigorous in that predictions of the species composition for strata that are missing data are made on the basis of a model (and not by substituting the species composition from a neighbouring stratum according to an *ad hoc* procedure). However, the trade-off is that the number of geographic areas included in the model is necessarily small. In the Case A application discussed above, only two areas, covering almost the whole of the tropical purse-seine fishery, were used. The number of areas is necessarily small because an increase in the number of areas results in an increase of the number of *area* interaction terms in the model — *year : area*, *quarter : area* and *school association : area* — the parameters for each of which must be estimated from the observer data. Since, historically, the coverage of observer data has not been complete, there will be interaction terms that are not covered by the data if the number of areas is too large.

In the Case B and C models, rather than modelling geographic area as a categorical covariate limited to a small number of values, geographic area was modelled as a two-dimensional spline, herein termed the *lat_lon* covariate, based on *latitude* and *longitude*. Modelling the geographic area as a two-dimensional spline, or surface, allows the species composition to be predicted at any *point* of latitude and longitude.

As will be shown below, the effect of latitude and longitude on the species composition differs considerably between associated and unassociated schools. Therefore, instead of modelling *school association* as a categorical variable, the observer data for each school association were treated separately, with the two sets of data used to fit two distinct sets of models. Furthermore, for the models fit with observer data covering associated schools, school association was modelled at a higher resolution, with categorical values for *logs*, *drifting FADs*, *anchored FADs* and the combined *‘other’* type of associated schools.

The *flag* state of the fishing vessel was also included as a categorical covariate.

An advantage of fitting separate models with data covering associated and unassociated schools is that there is no need for interaction terms between *school association* and the other covariates. However, consideration was given to the inclusion of the *year* : *quarter* and *year* : *area* terms, given that the observer data may not be sufficiently informative to estimate the coefficients of a two-dimensional spline together with these interaction terms. Therefore, in Case B, models without the *year* covariate were used to estimate the species composition for 1967–1995, as in Case A, while models with the *year* covariate but with the *year : quarter* and *year : area* interactions were used to estimate the species composition for 1996–2011. Case C is similar to Case B, except that models with the *year : quarter* and *year : area* interactions were used to estimate the species composition for 2002–2011, as for Case A.

Other factors may affect the species composition; however, they were not considered in the models since the objective was to predict the species compositions of historical catches using the information associated with those catches that are available in the databases maintained by the OFP. The primary OFP database with historical catches is the ‗s_best' database, which contains information only on the year, month, 1° x 1° grid, school association and vessel flag.

Replicates

The replicates used to fit the models of Cases B and C were defined as strata of *trip – school association*; the response variables were the proportion of each species determined from the sum of the observed catches by species for the trip and school association, wherein the catches by species for each set were the product of the species composition determined from the samples and the set weight. There were a total of 4,862 strata of observed *trip – school association*. Strata of *trip – school association* were chosen as replicates, rather than individual sets, since the latter are not independent. In another context, General Estimating Equations have been used to show that *trip*, rather than *set*, is a more appropriate replicate when analysing observer data (Lawson 2011a).

Purse-seine trips typically take place over parts of two to three months; the year and month assigned to each stratum of *trip – school association* was the month for which the number of days in the stratum on which at least one set was made was greatest.

The *latitude* and *longitude* assigned to each stratum was the average latitude and longitude of the location of sets in the stratum, weighted by the catch.

For strata of *trip – associated schools*, the *school association* assigned to the stratum — logs, anchored FADs, drifting FADs or ‗other' — was the school association for which the total catch in the stratum was greatest.

When fitting the models, each stratum was weighted by the number of sets in the stratum.

The complexity of the *lat_lon* surface is determined in part by the *degrees of freedom* of the two-dimensional spline. However, for the predictions of the species compositions for each set of models to sum to unity, the model matrix — including the basis functions of the *lat_lon* spline — must be the same for each species; that is, the degrees of freedom must be the same for each species. To determine an appropriate value of the degrees of freedom for each set of models, the BIC was determined for values ranging from 3 to 29. For skipjack and yellowfin, the degrees of freedom that minimised the BIC was usually 12 or 13, while for bigeye, it was the lowest value, 3. For simplicity, a value of 13 was chosen for each of the models in Cases B and C. Using a value of 13 to fit the *lat_lon* surface for bigeye, for which the degrees of freedom that minimised the BIC was 3, may be considered over-fitting the model; however, in general, it appears that the estimated *lat_lon* surfaces reflect primarily the information in the observer data and are not particularly sensitive to the degrees of freedom, particularly when the degrees of freedom is greater than the value that minimises the BIC. It should also be noted how economical, in general, the *lat_lon* spline is with degrees of freedom; in comparison, categorical covariates for each of the 5° x 5° grids in MFCL Skipjack Areas 2 and 3 would use 143 degrees of freedom.

Fits of the Case B and C 'high resolution' models

For Case B, with (a) the observer data covering associated and unassociated schools modelled separately, (b) two models to estimate the species composition for 1967–1995 and 1996–2011, and (c) three species, there are a total of 12 fits. For Case C, there is a third model to estimate the species composition for 2002–2011 and so there are 18 fits.

The models were fit assuming Gaussian errors. Appendix IV presents, as an example, a summary of the fit of the model of the proportion of skipjack with the *year* covariate and the *year* interactions to observer data from associated schools.

Tables 5 shows the deviance explained by each of the covariates separately and for each of the three models in Cases B and C with all of the covariates combined. For associated schools, the *year : quarter* and *year : area* interaction terms are the most important, whereas for unassociated schools, the *lat_lon* surface is the most important. For the models with the *year* covariate and the *year : quarter* and *year : area* interactions, the deviance explained by all covariates combined — the bottom row of Table 5 — are moderate for skipjack and yellowfin, and less so for bigeye, particularly for unassociated schools. For the models without the *year : quarter* and *year : area* interactions — the second to bottom row — the deviance explained are somewhat lower, while for the models without the *year* covariate — the third to bottom row — the deviance explained is considerably lower.

Table 5. Deviance explained by separate covariates and all covariates combined for fits of the 'high resolution' models

Covariate	School Association					
	Associated			Unassociated		
	Skipjack	Yellowfin	Bigeye	Skipjack	Yellowfin	Bigeye
Year	20.1%	17.9%	8.6%	6.3%	6.4%	1.4%
Quarter	6.7%	6.2%	0.6%	1.1%	1.0%	0.3%
Year – Quarter Interaction	27.2%	24.7%	16.2%	17.9%	17.9%	4.1%
Latitude x Longitude	13.0%	20.4%	4.8%	22.1%	22.1%	1.1%
Flag	11.2%	16.8%	6.0%	10.2%	10.2%	1.1%
MFCL Area – Year Interaction	24.3%	23.0%	11.3%	9.9%	9.9%	2.4%
School Association Sub-Type	10.5%	15.9%	1.8%			
Case B and C: Model without *year* covariate	24.0%	32.9%	11.4%	24.7%	24.5%	2.4%
Case B and C: Model with year and without year interactions	35.4%	40.0%	18.5%	30.7%	30.7%	3.6%
Case C: Model with year and with year interactions	42.4%	46.6%	26.8%	40.5%	40.4%	6.9%

To check whether the assumption of Gaussian errors was appropriate, quantile-quantile plots of the residuals were examined. Figure 12 shows the Q–Q plots for the model in Cases B and C with the *year* covariate and without the *year* interactions. The linearity of the plots indicates that the assumption of normality is appropriate for skipjack and yellowfin, less so for bigeye in associated schools, and not for bigeye in unassociated schools. The Q-Q plots for the models with the *year* interactions (not shown) were similar to those without the *year* interactions. Other assumptions regarding the distribution of the errors were examined, but with no improvement.

Figure 12. Q-Q plots for the models in Case B and C with the *year* covariate and without *year* interactions

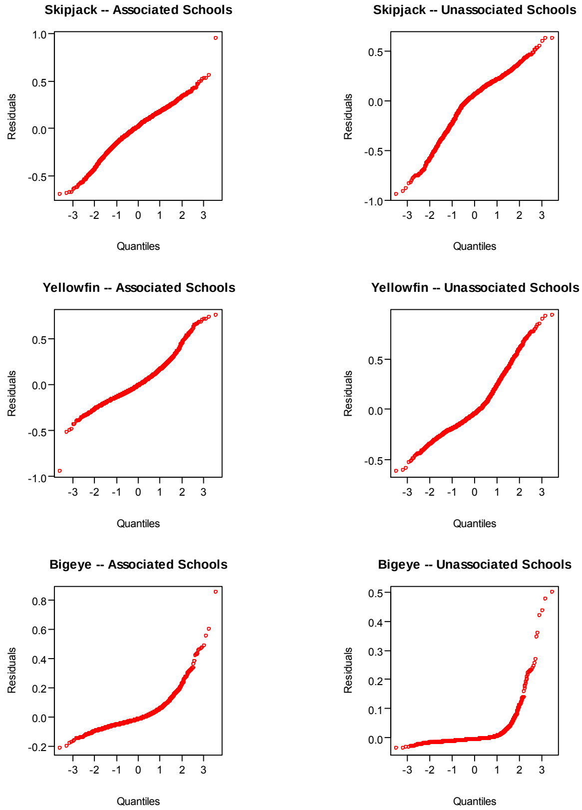


Figure 13 compares the observed values of the species composition for each replicate of *trip – school association* to the fitted values, for the models with the *year* covariate and without the *year* interactions. For intermediate values of the proportions of each species, the fit appears to be relatively good. For observed proportions close to 0.0 or 1.0, the fitted proportions cannot be distributed symmetrically above and below the diagonal lines in Figure 13; nevertheless, even taking that factor into account, there still appears to be lack of fit, particularly for low proportions of skipjack and high proportions of yellowfin.

Figure 13. Observed vs fitted values of the proportion by species for the Case B and C models with the *year* covariate and without *year* interactions

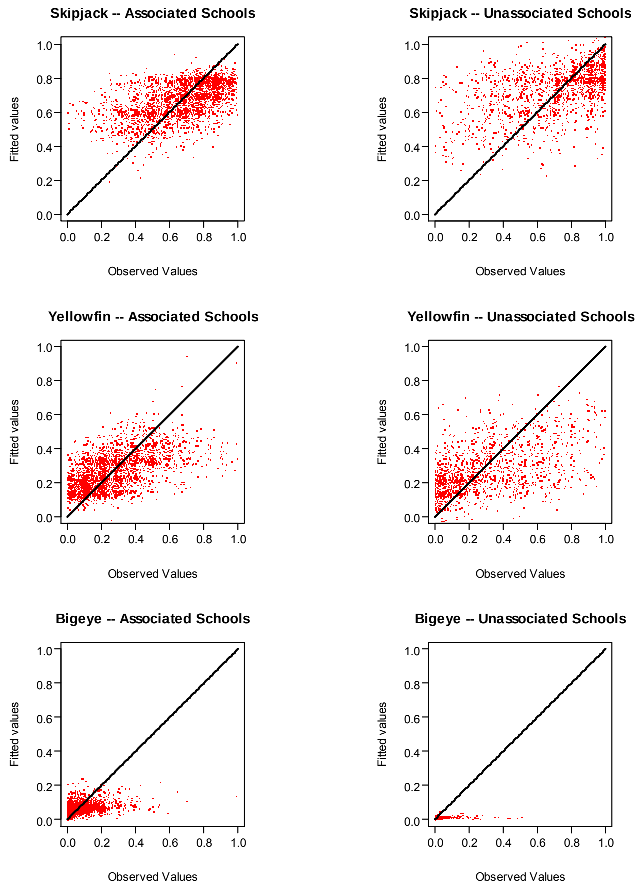


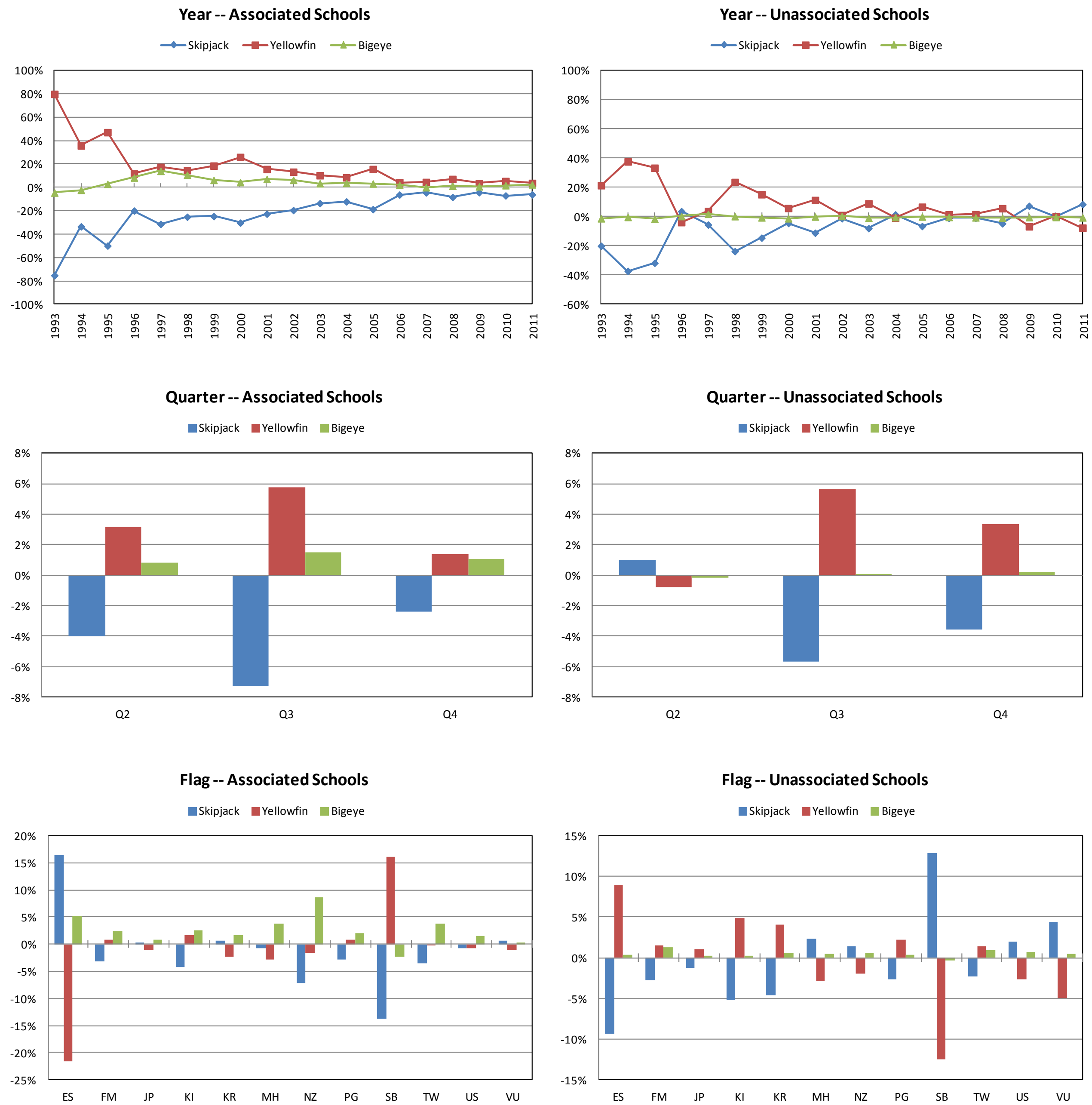
Figure 14 shows the value of the coefficients for *year*, *quarter*, *flag* and *associated school type* for fits of the Case B and C models with the *year* covariate and without the *year* interactions. The base values of these categorical covariates are 2012 for *year*, Q1 for *quarter*, _China‘ for *flag* and _anchored FADs‘ for *associated school type*.

For *year*, the somewhat extreme coefficients for 1993–1995 reflect observer coverage of less than 1% and should be considered unreliable. For 1996–2011, there appears to be an increasing trend in the coefficients for skipjack and a declining trend for yellowfin.

For *flag*, the values for _Spain‘ are consistent with the heat maps of the effect of latitude and longitude (Figure 15), given that the Spanish fleet tends to fish in the northeast part of the region. The values for _Solomon Islands‘ are also consistent with Figure 15, given the north-south gradients for the proportions of skipjack and yellowfin and the fact that the sets on unassociated schools by this fleet tend to be somewhat to the north of the sets on anchored FADs.

For *associated school types*, the coefficients for drifting FADs and logs are low, indicating little difference with the base value of _anchored FADs‘. For the combined _other‘ type of *associated school*, 53% of the observed sets are on live whales, 18% on live whale sharks and 28% are recorded as ‗other‘; thus, the proportions for ‗other‘ in Figure 14 represents primarily the effect of live whales and whale sharks.

Figure 14. Coefficients of the Case B and C models with the *year* covariate and without *year* interactions



ES = Spain, FM = Federated States of Micronesia, Micronesia, JP = Japan, KI = Kiribati, KR = Korea, MH = Marshall Islands, NZ = New Zealand, PG = Papua New Guinea, SB = Solomon Islands, TW = Chinese Taipei, US = United States of America, VU = Vanuatu

Figure 14 (continued)

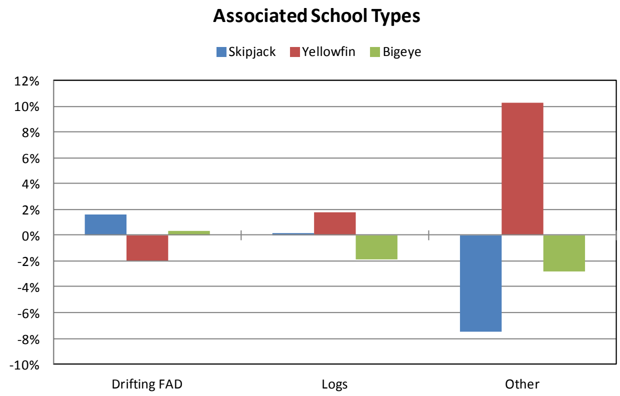
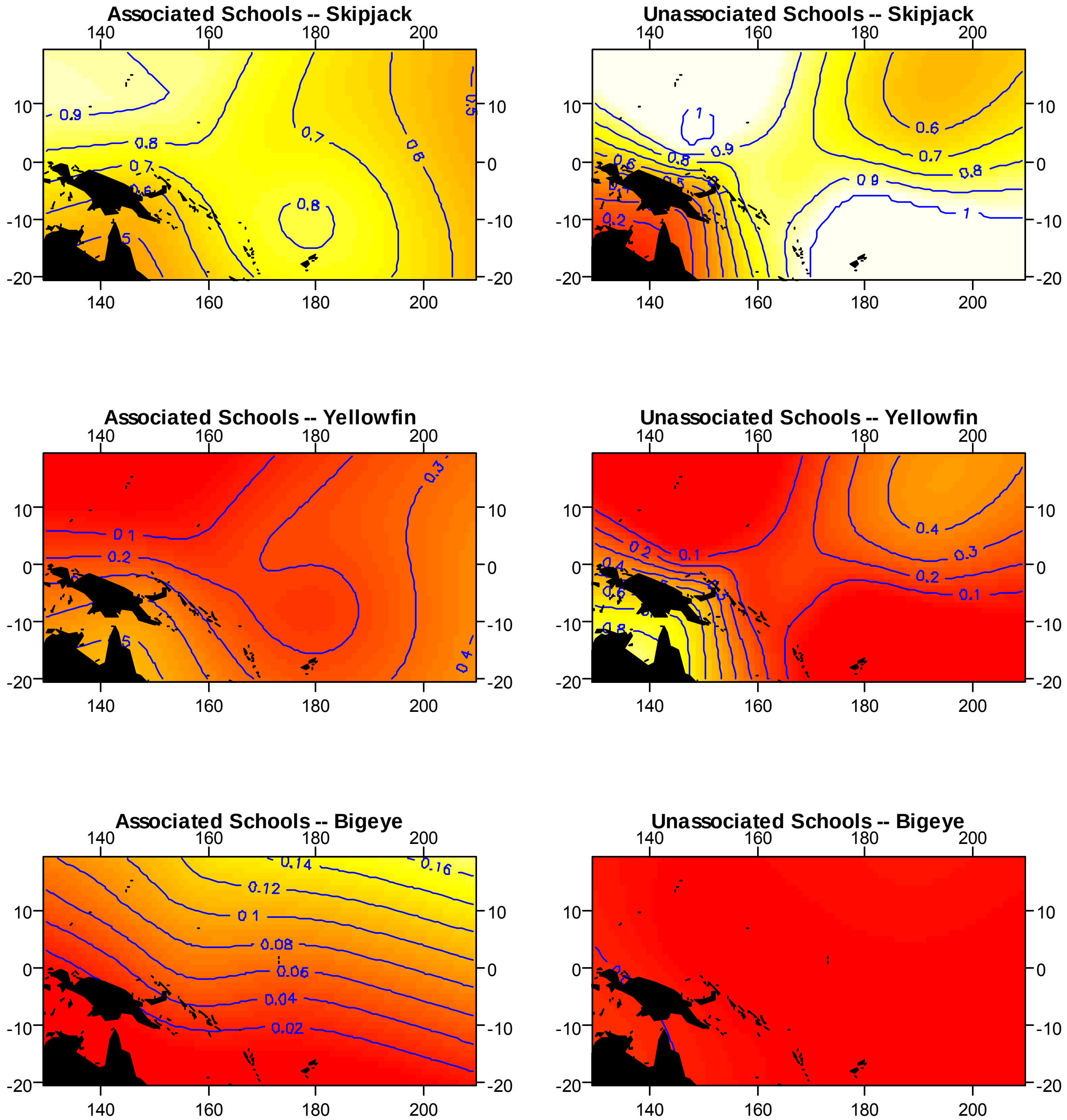


Figure 15 shows the effect of latitude and longitude on the purse-seine species composition for the model with the *year* covariate and without the *year* interactions; the colour scale and contour scale of the heat maps for skipjack and yellowfin are the same, while the scales for bigeye are lower so that the small proportions of bigeye can be distinguished. Base values for *year* of 2000, *month* of 6, and *flag* of ‗China' were used to construct Figure 15; other base values result in slightly different heat maps. The heat maps for skipjack and yellowfin are approximately mirror images for associated schools, and almost exactly mirror images for unassociated schools (for which the proportion of bigeye is usually negligible). The proportions of skipjack and yellowfin show northwest to southeast and southwest to northeast symmetries, particularly for unassociated schools. The proportions of bigeye in associated schools shows a southwest to northeast gradient.

Figure 15. Effect of latitude and longitude on purse-seine species composition (proportion) for the Case B and C models with the *year* covariate and without *year* interactions



Adjustment of catch data in 's_best' with the species compositions estimated for Cases B and C

The models in Cases B and C were used to predict the species composition for the catch data in _s_best‘, which are stratified by year, month, 1° x 1°, school association and vessel flag. All strata for MFCL Skipjack Areas 2 and 3 were adjusted, except those covering:

- the domestic fleets of Indonesia and the Philippines, which are included in the MFCL stock assessments of skipjack, yellowfin and bigeye as separate fisheries and for which little or no observer data are available;

- the fleets of Ecuador, San Salvador and Spain, which are considered to accurately report the species composition on logsheets;

- the Japanese fleet from 1996 onwards, for which the species compositions have already been adjusted based on port sampling of landing categories; and

- all catches to the west of 130°E, which are excluded from the main fisheries defined for MFCL Skipjack Areas 2 and 3, and MFCL Yellowfin – Bigeye Areas 3 and 4, in the assessments.

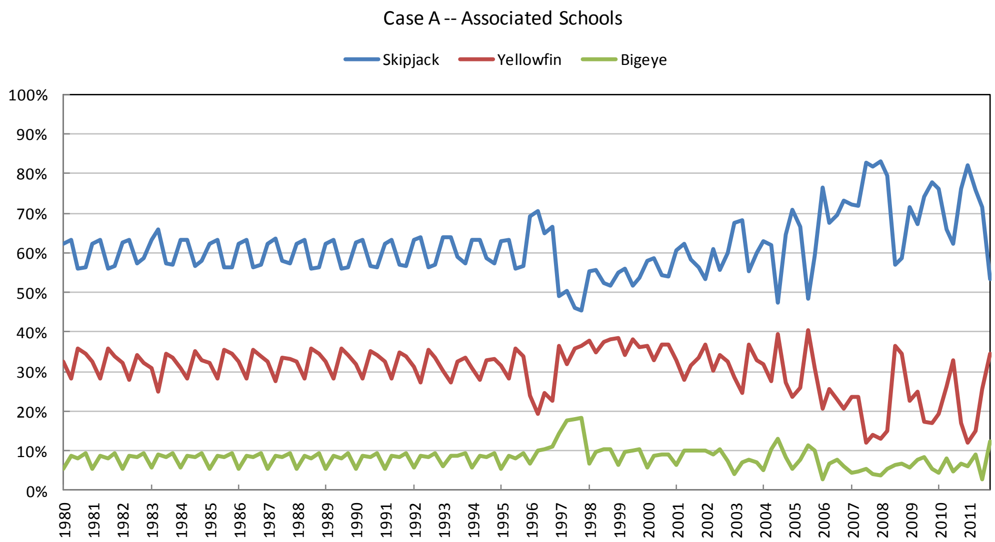
The number of strata that were adjusted are given in Table 6 by time period and school association; the total number of strata is large, 226,271. The catch in ‗s_best‘ covering Skipjack Areas 2 and 3 during 1967–2011 that were adjusted represent 87.8% of the catch in Skipjack Areas 2 and 3 that are used in the assessments.

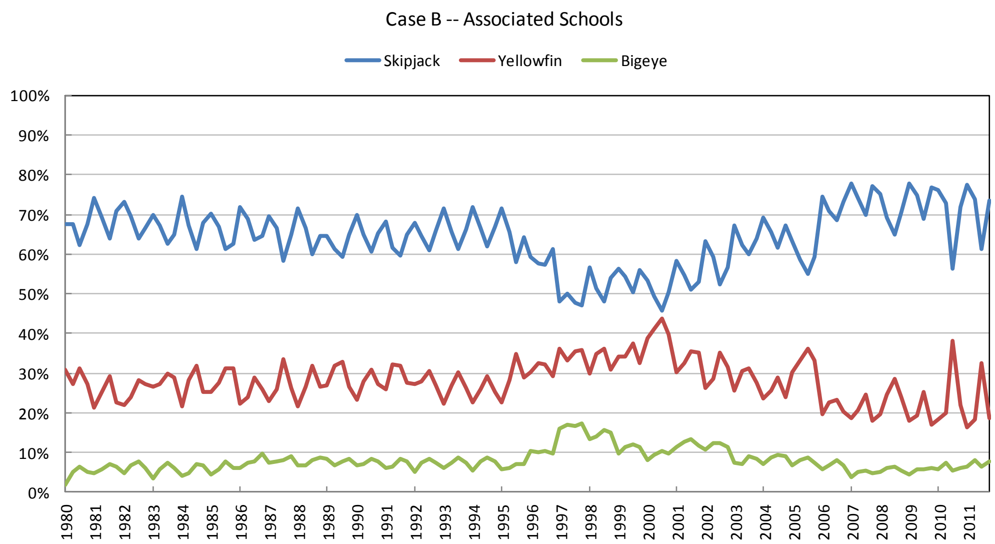
Table 6. Number of strata in 's_best' for which the catches were adjusted with species compositions estimated using the Case B and C models

Time Period	School Association	Number of Strata
1967 - 1995	Associated	55,973
	Unassociated	25,193
1996 - 2001	Associated	26,375
	Unassociated	12,648
2002 - 2011	Associated	66,640
	Unassociated	39,442
	Total	226,271

Figure 16 presents the species compositions of quarterly catches in MFCL Skipjack Areas 2 and 3, 1980–2011, that were adjusted with the ‗low resolution‘ models of Case A and the ‗high resolution‘ models of Cases B and C. Note that for 1967–1995, the models do not include the *year* covariate; for 1996–2001, the models include the *year* covariate but not the year interactions; and for 2002–2011, the models in Cases A and C include the *year* covariate and *year* interactions, while Case B includes the *year* covariate, but not the *year* interactions.

Figure 16. Purse-seine species compositions of catches in MFCL Skipjack Areas 2 and 3 that were adjusted with the 'low resolution' models of Case A and the 'high resolution' models of Cases B and C





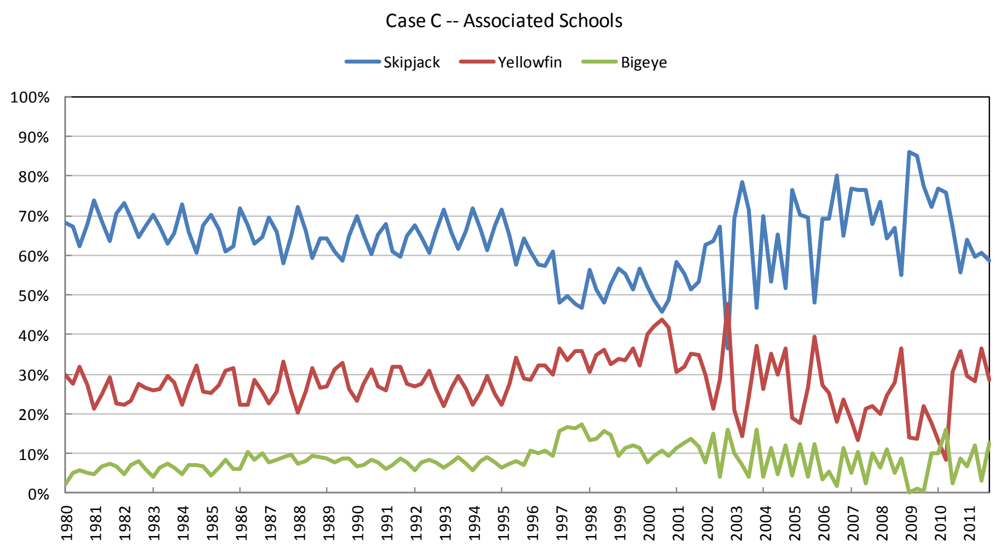
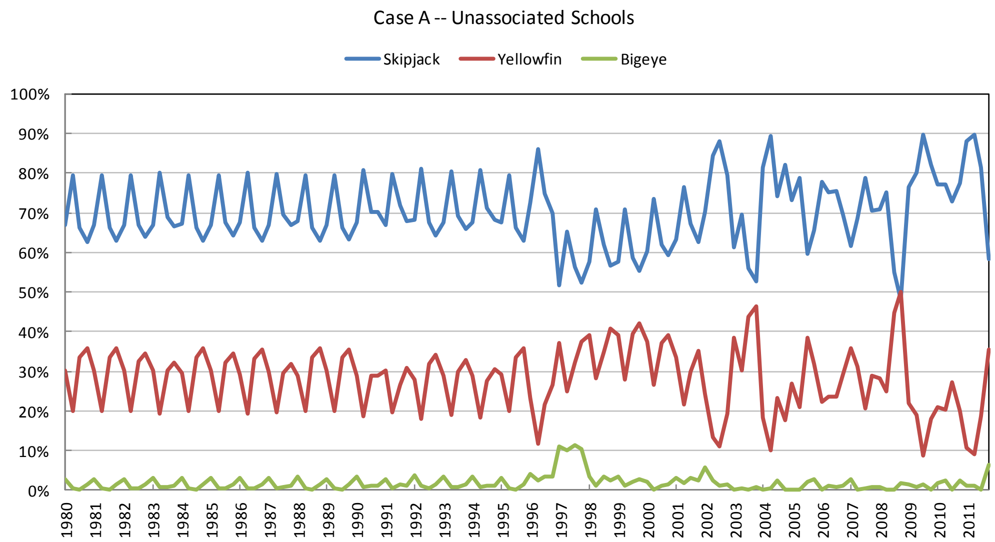
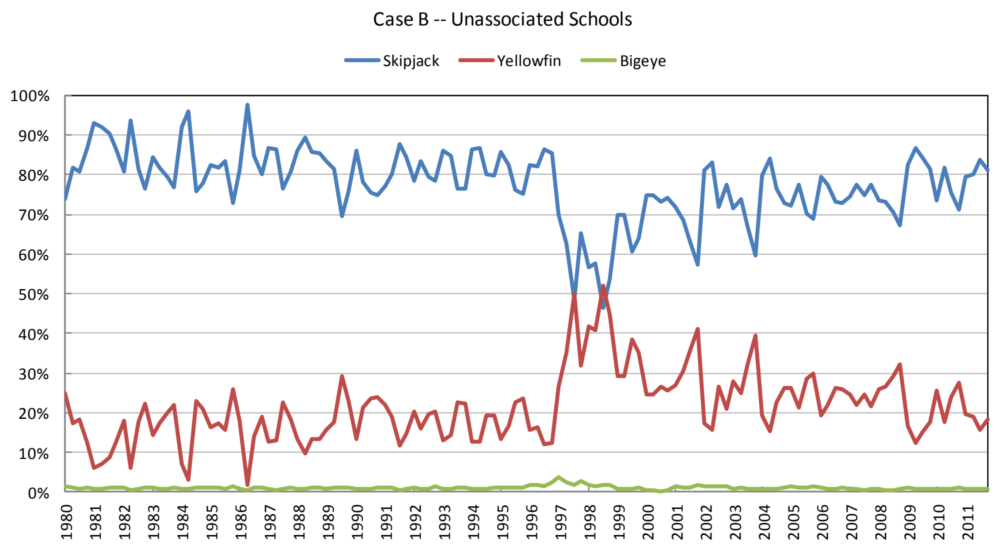
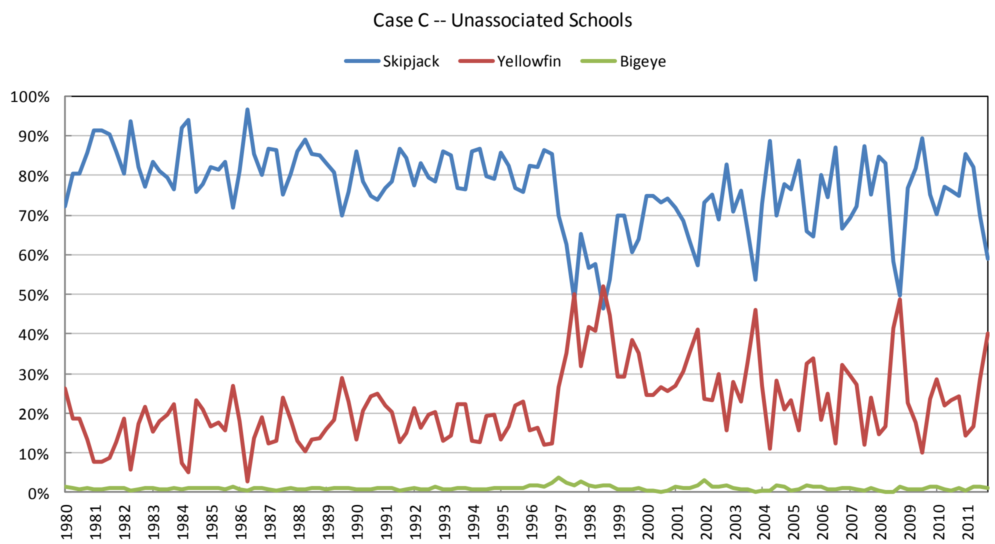


Figure 16 (continued)







The following points are of interest:

- The dominant feature of the trends during prior to 1996 is the effect of the *quarter* covariate. For associated schools, the range of the proportions of skipjack and yellowfin within a year are about 10%, while for unassociated schools they are about 18%. For the ‗high resolution' models, there is less regularity in the quarterly trends due to changes in the geographic distribution of fishing effort.

- For both associated and unassociated schools during 1980–1995, the ‗low resolution' models predict a lower proportion of skipjack and a higher proportion of yellowfin than the ‗high resolution' models. This indicates that during this period, fishing within each MFCL Area has tended to be concentrated in those locations where the proportion of skipjack is higher and yellowfin lower.

- The first year for which the species composition is estimated with a model including a *year* covariate is 1996; this explains the discontinuity between the periods before and after 1996.

- The long-term trends for each species are generally consistent among the three Cases.

- For both associated and unassociated schools, the direction of the quarterly changes for each Case are generally consistent, although their magnitudes differ considerably. For Cases A and C, inclusion of the *year* interactions results in a higher degree of volatility in the quarterly species compositions.

- But the trends among the Cases are not always consistent. For example, for associated schools in the last quarter of 2011, the final point in the time series, the proportion of skipjack declines and the proportion of yellowfin increases in Case A, the opposite occurs in Case B, while in Case C, the proportions of both skipjack and yellowfin decline. The proportion of bigeye, however, increases in all three Cases.

Table 7 presents the species compositions for 1972–2011, and Figure 17 presents the annual catch estimates, of the purse-seine catches in MFCL Skipjack Areas 2 and 3, excluding the domestic fisheries of Indonesia and the Philippines, for Cases A, B and C. Catches for those strata in ‗s_best' that were not adjusted have been included.

Table 7 indicates that for the entire time series combined, the species compositions do not differ greatly among the Cases; however, as suggested by Figure 16, Cases B and C have higher proportions of skipjack and lower proportions of yellowfin than Case A.

Table 7. Catches and species compositions in MFCL Skipjack Areas 2 and 3, 1972–2011

Case	School Association	Skipjack		Yellowfin		Bigeye		Total
		Tonnes	%	Tonnes	%	Tonnes	%	Tonnes
A	Associated	9,695,842	65.7%	3,941,384	26.7%	1,119,909	7.6%	14,757,136
	Unassociated	7,691,097	71.5%	2,897,925	26.9%	169,435	1.6%	10,758,457
	All	17,386,939	68.1%	6,839,310	26.8%	1,289,344	5.1%	25,515,593
B	Associated	9,935,637	67.3%	3,702,012	25.1%	1,119,487	7.6%	14,757,136
	Unassociated	8,168,526	75.9%	2,489,466	23.1%	100,465	0.9%	10,758,458
	All	18,104,164	71.0%	6,191,478	24.3%	1,219,952	4.8%	25,515,594
C	Associated	9,789,293	66.3%	3,774,837	25.6%	1,193,005	8.1%	14,757,136
	Unassociated	8,093,813	75.2%	2,560,052	23.8%	104,592	1.0%	10,758,457
	All	17,883,106	70.1%	6,334,890	24.8%	1,297,598	5.1%	25,515,593

Figure 17 shows that the annual catch estimates for skipjack vary to a lesser extent among the Cases than those for yellowfin and bigeye. Again, as suggested by Figure 16, for most, but not all years, Case A has a larger catch of yellowfin and bigeye than Cases B and C.

Figure 17. Annual purse-seine catches in MFCL Skipjack Areas 2 and 3 determined from species compositions estimated for Cases A, B and C

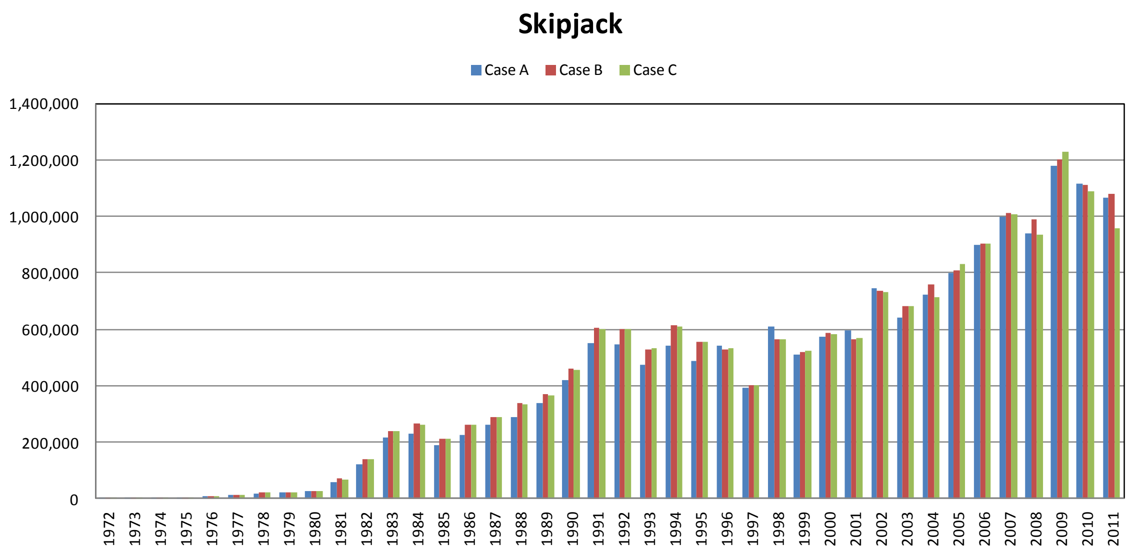
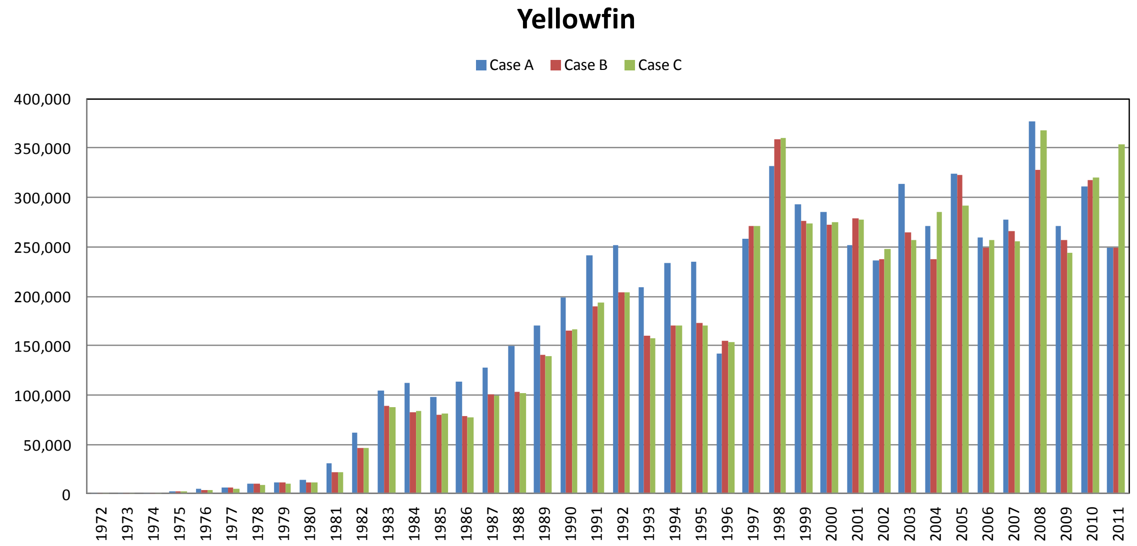
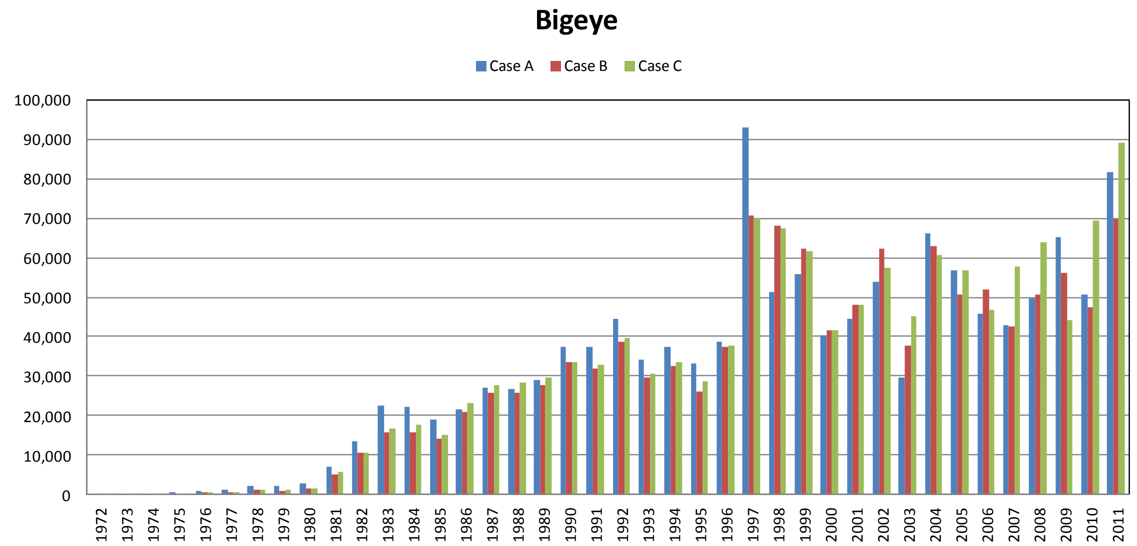


Figure 17 (continued)





Correction of length frequencies for size selectivity bias

The estimates of availability can be used to correct the length frequencies determined from the grab samples as follows (Lawson 2011b):

$$\hat{n}_{ijk} = \sum_i \sum_j n_{ijk} \cdot \frac{N_{ijk}}{\sum_i \sum_j N_{ijk}} \tag{14}$$

$$= \sum_i \sum_j n_{ijk} \cdot \frac{\frac{n_{ijk}}{A_j}}{\sum_i \sum_j \frac{n_{ijk}}{A_j}} \tag{15}$$

where $\hat{n}_{ijk}$ is the corrected number of fish of species i and length interval j in the samples from set k ; n_{ijk} is the uncorrected number of fish of species i and length interval j in the samples from set k ; N_{ijk} is the ―true" number of fish of species i in length interval j in the samples from set k ; and A_j is the probability that a grab sampler will select a fish in length interval j .

In equations (14) and (15), the total number of fish in the samples from a particular set, $\sum_i \sum_j n_{ijk}$, is applied to a corrected length frequency (in terms of proportions of numbers of fish) based on the estimates of availability, i.e., the right-hand part of the product in equations (14) and (15). Thus, the total of the corrected number of fish in the length frequency for a set is equal to the total of the uncorrected number of fish in the length frequency for that set. This has two effects. First, unlike the uncorrected numbers of fish, the corrected numbers of fish are not integers. Second, there is an effect on the species composition (in terms of numbers of fish) within a set, such that the total number of fish in the corrected length frequency for skipjack increases, while those for yellowfin and bigeye decrease; this is because the availability of smaller fish (primarily skipjack) is less than for larger fish (primarily yellowfin and bigeye).

Application of the 'high resolution' model to scaling of the length frequencies

In the past, length frequencies used in MFCL assessments of tuna have been determined simply by aggregating the samples for each set by strata of year, quarter, MFCL area and school association. No attempt was made to scale the samples by the catch by species at a finer resolution, e.g., 5° x 5° by quarter, prior to aggregating the samples by MFCL strata. Scaling of the length frequencies is not appropriate using catches estimated from species compositions determined from the ‗low resolution' models, since the covariate for geographic area is the MFCL area. However, since the covariate for geographic area in the ‗high resolution' models is a two-dimensional spline — i.e., the continuous latitude and longitude surface — scaling of the length frequencies is appropriate; the length frequencies were scaled with the catches determined from the Case B ‗high resolution' models.

The length frequencies for each species, determined from corrected grab samples and spill samples (see previous section), and raised by the set weight, were scaled by the catch in strata of *year – quarter – 5° x 5° – school association*, using the formulae derived in Appendix V. The procedure uses catches estimated only for those strata for which samples are available and ignores catches in other strata (Table 8). It also ignores samples in strata for which the catch determined from the ‗high resolution‘ model was zero; this can occasionally arise because the source for the location of the catch data, i.e., catch and effort logsheets, is different from the observer data. For skipjack, there were 22 out of 2,503 strata, 0.9%, covered by grab samples for which the catch was ―zero”. For yellowfin, there were 23 out of 2,151 strata, 1.1%, and for bigeye, there were 9 out of 1,626 strata, 0.6%.

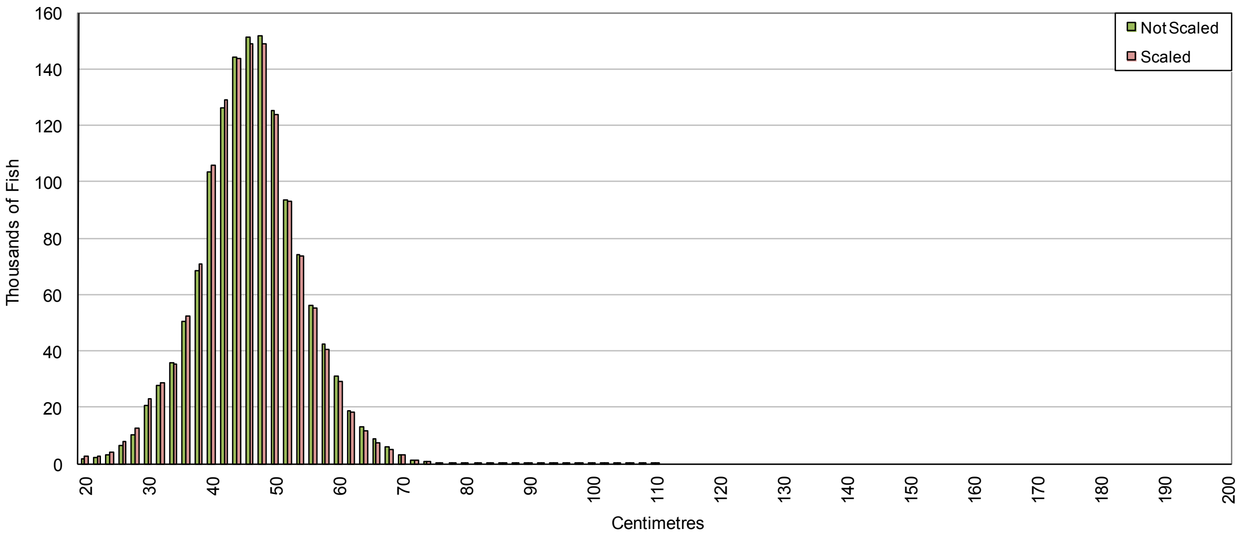
Table 8. Comparison of total catches to catches used for scaling length frequencies

Year	Total Catches				Catches Used for Scaling				Percentage			
	Skipjack	Yellowfin	Bigeye	Total	Skipjack	Yellowfin	Bigeye	Total	Skipjack	Yellowfin	Bigeye	Total
1993	594,592	167,672	30,408	792,672	40,248	12,962	0	53,210	6.77%	7.73%	0.00%	6.71%
1994	706,949	189,512	34,612	931,073	96,400	18,810	1,853	117,062	13.64%	9.93%	5.35%	12.57%
1995	669,022	192,952	28,258	890,232	175,868	65,828	7,921	249,618	26.29%	34.12%	28.03%	28.04%
1996	654,389	175,670	39,883	869,942	259,298	83,318	19,449	362,065	39.62%	47.43%	48.76%	41.62%
1997	521,227	300,952	76,408	898,587	148,208	95,531	30,262	274,001	28.43%	31.74%	39.61%	30.49%
1998	737,170	376,328	76,067	1,189,565	420,729	255,175	41,625	717,529	57.07%	67.81%	54.72%	60.32%
1999	716,912	296,081	63,926	1,076,919	310,732	151,235	25,618	487,585	43.34%	51.08%	40.07%	45.28%
2000	743,237	330,520	50,652	1,124,410	408,706	187,052	23,500	619,257	54.99%	56.59%	46.39%	55.07%
2001	715,403	334,400	54,891	1,104,695	451,901	216,108	35,174	703,183	63.17%	64.63%	64.08%	63.65%
2002	907,802	289,971	66,848	1,264,621	638,380	186,582	49,593	874,555	70.32%	64.34%	74.19%	69.16%
2003	873,252	327,852	43,764	1,244,867	641,071	242,254	28,734	912,059	73.41%	73.89%	65.66%	73.27%
2004	971,479	312,886	70,607	1,354,972	713,259	218,707	53,981	985,947	73.42%	69.90%	76.45%	72.77%
2005	1,016,311	417,326	62,398	1,496,035	762,889	302,422	44,797	1,110,108	75.06%	72.47%	71.79%	74.20%
2006	1,156,812	337,606	64,042	1,558,460	873,865	227,722	45,262	1,146,849	75.54%	67.45%	70.68%	73.59%
2007	1,278,718	368,686	50,948	1,698,351	926,956	239,129	32,488	1,198,572	72.49%	64.86%	63.77%	70.57%
2008	1,279,074	426,425	62,678	1,768,176	893,972	302,540	37,442	1,233,954	69.89%	70.95%	59.74%	69.79%
2009	1,486,861	315,721	67,114	1,869,696	1,093,075	217,667	43,670	1,354,411	73.52%	68.94%	65.07%	72.44%
2010	1,370,840	352,341	55,867	1,779,048	1,113,314	311,060	43,705	1,468,079	81.21%	88.28%	78.23%	82.52%
2011	1,299,849	280,132	76,770	1,656,751	955,205	219,940	54,780	1,229,925	73.49%	78.51%	71.36%	74.24%

Figure 18 compares the unscaled and scaled length frequencies by species and school association for samples collected during 1993–2011. The scaled length frequencies generally contain more small and fewer large fish; the effect is particularly notable for yellowfin caught in unassociated schools. This implies that fishing occurs more in 5° x 5° grids where the fish are smaller than in those where they are larger.

Figure 18. Unscaled and scaled length frequencies determined from grab samples corrected for selectivity bias and spill samples, 1993–2011

Skipjack -- Associated Schools



Skipjack -- Unassociated Schools

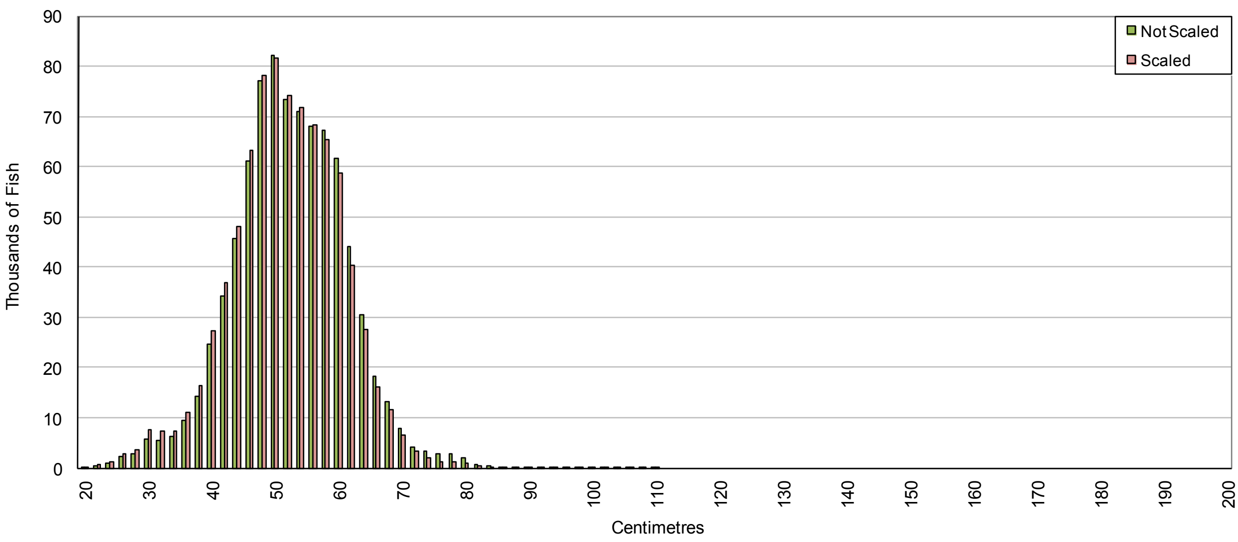
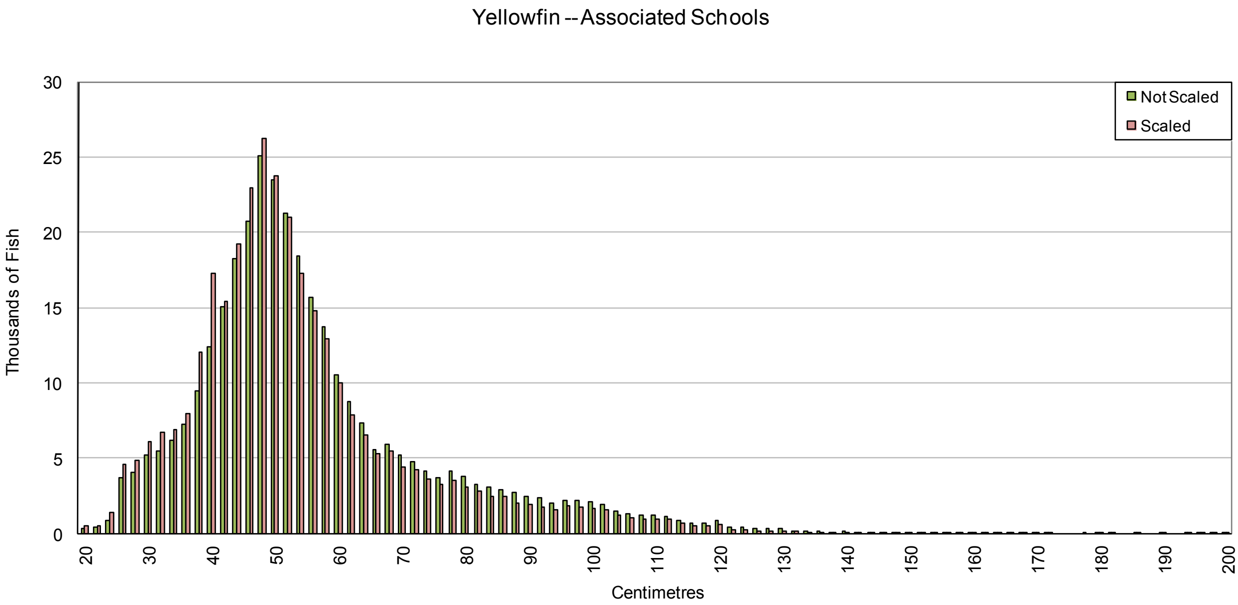


Figure 18 (continued)



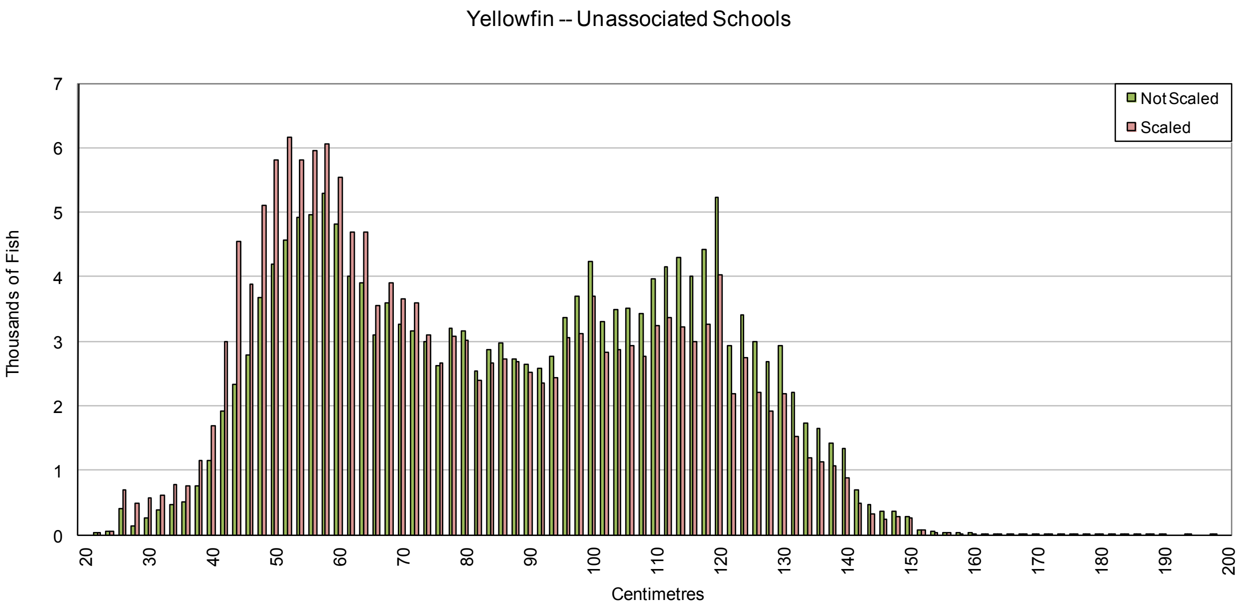
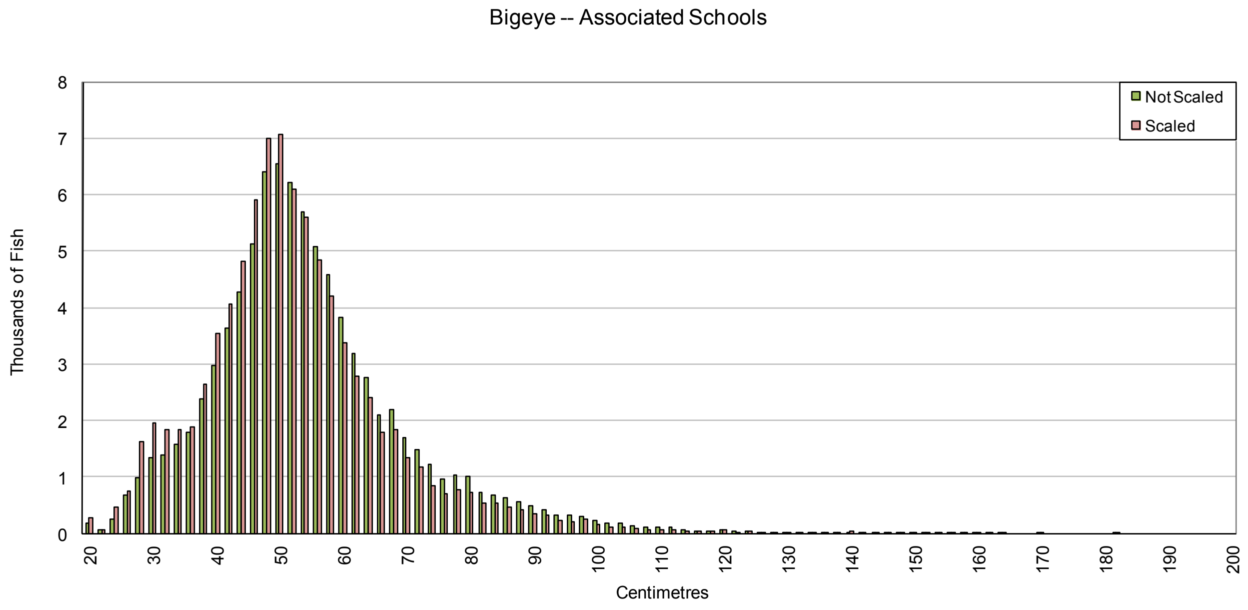
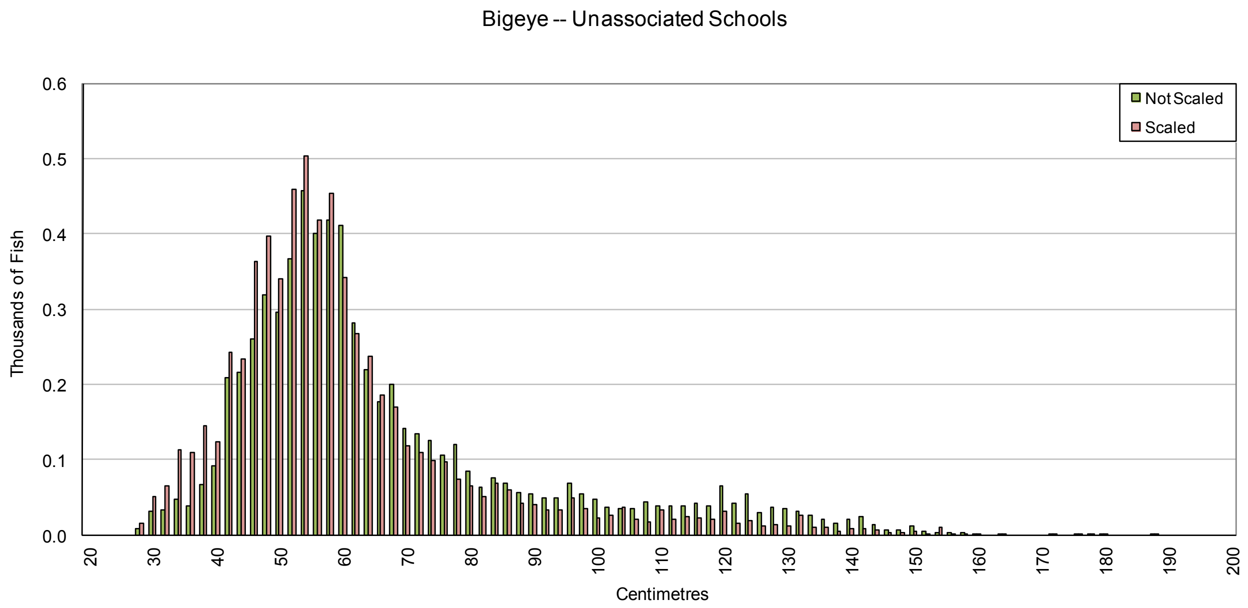


Figure 18 (continued)





Conclusions

Paired grab and spill samples, selectivity bias and layering

The increase in the number of paired samples from unassociated schools from 11 reported in Lawson (2010) to 82 has allowed for more reliable estimates of the selectivity bias for larger fish (Figure 5). Additional paired samples will be undertaken during the remainder of 2012 and in 2013.

The use of splines to estimate *availability* from paired samples, rather than categorical covariates for length intervals, results in continuous estimates of the bias as a function of fish length.

Layering by size during brailing has been shown to occur (Figure 6) and may be an important cause of the selectivity bias in grab samples.

Estimation of the species composition from grab samples corrected for selectivity bias and spill samples

Grab samples corrected for selectivity bias appear to contain useful information regarding the species composition of purse-seine catches, particularly in contrast to uncorrected grab samples.

Linear models of the species composition, with the proportion by species as a response variable, can be used to estimate the species composition for strata of time period, geographic area, school association and vessel flag. However, in their present form, the proportions predicted by the models are not constrained to the interval [0,1]. For Case A, the ‗low resolution‘ model, the estimated proportions by species fell outside [0,1] for at least one species in 4.4% of strata; for Case B, the ‗high resolution‘ model without *year* interactions, they were outside the interval in 3.2% of strata; and for Case C, the ‗high resolution‘ model with *year* interactions, they were outside for 6.2% of strata. The majority of these instances were for the proportion of bigeye in unassociated schools, for which the predicted proportions were slightly negative; catches of bigeye in unassociated schools are usually negligible. [4] The problem was resolved by normalising the species composition; [5] however, reformulating the model would be a more statistically rigorous solution. For example, the multinomial Poisson transformation has recently been used to estimate the species composition of the California groundfish fishery (Shelton et al. 2012). On the other hand, it is unlikely that a more rigorous model would result in large differences in the estimated species compositions and catches; the choice of the covariates included in the models and the time periods for which the various models are used to estimate the species compositions are more important.

The heat maps of the effect of the *lat_lon* spline on the species composition (Figure 15) shows that within each of the MFCL Skipjack Areas, the species composition varies considerably with location. This result supports the use of the ‗high resolution‘ models to estimate the species composition.

Including the *year* interaction terms in the ‗high resolution‘ model affects the estimates of the catches in the MFCL Areas (Figure 17), particularly for bigeye and, to a lesser extent, yellowfin, and introduces greater volatility in the quarterly species composition for 2002–2011 (Figure 16). It

[4] The proportions predicted by the models still sum to unity. If there are two species — skipjack and bigeye, for example —and the proportion of bigeye, p, is slightly negative, then the proportion of skipjack will be $1 - p$, i.e., slightly greater than 1. If there are three species, the proportions of skipjack and yellowfin will sum to $1 - p$.

[5] Proportions greater than 1.0 are set to 1.0 and proportions less than 0.0 are set to 0.0. The revised proportions for each of the three species are then divided by the sum of the revised proportions, such that the normalised proportions sum to unity.

is interesting to note that when similar plots of the quarterly species composition are done for the unadjusted catches in ‗s_best', which primarily reflect the catches recorded on logsheets, the level of quarterly volatility in the proportions of skipjack and yellowfin are equally high, which supports the inclusion of the *year* interaction terms.

The uncertainty of the estimates of the catches in the MFCL Areas has not yet been determined. One approach would be to use a parametric bootstrap to estimate confidence intervals. However, a sample of 1000 sets of coefficients from the prior distribution, which is the sample size typically used, applied to 226,271 strata in ‗s_best' (Table 6) for which the species composition must be estimated when using the ‗high resolution' models, for each of the three species, would require 678,813,000 estimates of the proportions by species. It remains to be seen whether this is computationally feasible.

In any case, the uncertainty can be evaluated subjectively by examining the coverage rates of the observer data upon which the estimates of the species composition are based (Table 3). No observer data are available for 1967–1992 and coverage is less than 2% for 1993–1997; hence, the uncertainty of the catch estimates for these years should be considered high. Coverage rises above 5% in 2002. The level of 5% was used to justify the inclusion of the *year* interaction terms in the models in cases A and C; other (higher) minimum levels of coverage could be examined in this regard.

The plots of the *year* coefficients of the Case B models (Figure 14) indicate trends in the proportions of skipjack and yellowfin. This implies that further consideration should be given to the temporal coverage of the observer data used to fit the models utilised to estimate the species composition for 1967–1995, in which a *year* covariate is not included. However, the observer coverage rates also need to be considered; excluding data for more recent years will result in a considerable decline in coverage.

Scaling the length frequencies by the catch

Scaling the length frequencies by the catch results in more small and fewer large fish, particularly for yellowfin and bigeye. However, the geographic coverage of the catch by the observer data is such that less than 65% of the catch is used to scale the length frequencies prior to 2002 (Table 8). Both the observer coverage rates and the percentage of the catch that is used to do the scaling affect the uncertainty of the scaled length frequencies.

References

Hampton, J. & P. Williams. 2011. Misreporting of purse seine catches of skipjack and yellowfin-bigeye on logsheets. Working Papers ST–WP–02. Seventh Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 9–17 August 2011, Pohnpei, Federated States of Micronesia. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia. http://www.wcpfc.int/doc/st-wp-02/misreporting-purse-seine-catches-skipjack-and-yellowfin-bigeye-logsheets

Lawson, T.A. 2008. Factors affecting the use of species composition data collected by observers and port samplers from purse seiners in the Western and Central Pacific Ocean. Working Paper SC4–ST–WP3. Fourth Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 11–22 August 2008, Port Moresby, Papua New Guinea. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia
http://www.wcpfc.int/sc4/pdf/sc4-st-wp3%20lawson.pdf

Lawson, T.A. 2009. Selectivity bias in grab samples and other factors affecting the analysis of species composition data collected by observers on purse seiners in the Western and Central Pacific Ocean. Working Paper SC5–ST–WP–03. Fifth Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 10–21 August 2009, Port Vila, Vanuatu. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.
http://www.wcpfc.int/system/files/documents/meetings/scientific-committee/5th-regular-session/statistics-swg/working-papers/SC5-ST-WP-03%20%5BSensitivity%20analysis%20-%20species%20coposition%5D.pdf

Lawson, T.A. 2010. Update on the estimation of selectivity bias based on paired spill and grab samples collected by observers on purse seiners in the Western and Central Pacific Ocean. Working Paper SC6–ST–WP–02. Sixth Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 10–19 August 2010, Nuku‘alofa, Tonga. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.
http://www.wcpfc.int/system/files/documents/meetings/scientific-committee/6th-regular-session/data-and-statistics-theme/working-papers/WCPFC-SC6-2010-ST-WP-02_Selectivity_bias.pdf

Lawson, T.A. 2011a. Estimation of catch rates and catches of key shark species in tuna fisheries of the Western and Central Pacific Ocean using observer data. Information Paper SC7–EB–IP–02. Seventh Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 9–17 August 2011, Pohnpei, Federated States of Micronesia. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.
http://www.wcpfc.int/system/files/documents/meetings/scientific-committee/7th-regular-session/ecosystems-and-bycatch-mitigation-theme/EB-IP-02%20%5BEstimation%20of%20Catch%20Rates%20and%20Catches%20of%20Key%20Shark%20Species%5D.pdf

Lawson, T.A. 2011b. Purse-Seine length frequencies corrected for selectivity bias in grab samples collected by observers. Working Paper SC7–ST–IP–02. Seventh Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission, 9–17 August 2011, Pohnpei, Federated States of Micronesia. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.
http://www.wcpfc.int/system/files/documents/meetings/scientific-committee/7th-regular-session/data-and-statistics-theme/information-pa/ST-IP-02%20%5BPS%20length%20freq%20corrected%20for%20selectivity%20bias%20in%20grab%20samples%5D.pdf

Pianet, R., P. Pallarés & C. Petit. 2000. New sampling and data processing strategy for estimating the composition of catches by species and sizes in the European purse seine tropical tuna fisheries. IOTC Proceedings No. 3 (2000) : 104-139. Indian Ocean Tuna Commission, Seychelles.
http://www.iotc.org/files/proceedings/2000/wpdcs/IOTC-2000-WPDCS-10.pdf

Schwarz, G.E. 1978. Estimating the dimension of a model. Annals of Statistics 6 (2): 461–464.

Shelton, A.O., E.J. Dick, D.E. Pearson, S. Ralston & M. Mangel. 2012. Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data. Can. J. Fish. Aquat. Sci. **69**: 231–246.
www.soe.ucsc.edu/~msmangel/Shelton%20et%20al%202012%20CJFAS.pdf

Tomlinson, P. 2002. Progress on sampling the Eastern Pacific Ocean tuna catch for species composition and length-frequency distributions. Stock Assessment Report 2: 339–365. Inter-American Tropical Tuna Commission, La Jolla, California, United States of America.
http://www.iattc.org/PDFFiles2/SAR2_sampling_ENG.pdf

Appendix I. Summary of the fit of *availability* using a spline

```
Call:
lm(formula = avail ~ -1 + bs(avg_length, df = mdf), data = obs_data)

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0160817 -0.0054039 -0.0027417  0.0008903  0.1949174 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
bs(avg_length, df = mdf)1 0.014871   0.001576   9.437  < 2e-16 ***
bs(avg_length, df = mdf)2 0.010087   0.003686   2.736  0.00626 ** 
bs(avg_length, df = mdf)3 0.016082   0.003741   4.299 1.78e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.01431 on 2532 degrees of freedom
Multiple R-Squared: 0.2457,     Adjusted R-squared: 0.2448 
F-statistic:   275 on 3 and 2532 DF,  p-value: < 2.2e-16
```

Appendix II. SPC/FFA Regional Purse-Seine Observer Form PS–4

SPC/FFA REGIONAL PURSE-SEINE OBSERVER
LENGTH MEASUREMENT — **FORM PS-4**

OBSERVER NAME	VESSEL NAME	OBSERVER TRIP ID NO.	PAGE OF

SAMPLING DETAILS

SAMPLE TYPE (fill only one of grab, spill, ...)	*if* GRAB: what is the Target No. of SAMPLES	*if* SPILL: brail # sampled	How many fish measured ?	*if* OTHER: What "Other" ? Use code (from back of form)	Which brail size was sampled ? (see PS-2) BRAIL 1 / BRAIL 2 (tick one)	BRAIL TIMES: start / end	PS-4 FORMS USED FOR THIS ONE SAMPLE: ___ of ___ (1,2,3) (total)	START OF SET DATE AND TIME (see PS-2): D D / M M / Y Y / h h / m m

MEASURING INSTRUMENT: Aluminium callipers / Measuring board / Deck tape / Other — Calibrated this set ? Y / N

COMMENTS ON SAMPLING PROTOCOL

MEASUREMENTS FOR EACH SPECIES
- UF - UPPER JAW TO FORK IN TAIL (TUNA, SHARK)
- LF - LOWER JAW TO FORK IN TAIL (BILLFISH)
- TW - TOTAL WIDTH (TIPS OF WINGS - RAYS)
- TL - TOTAL LENGTH - TO END OF TAIL (FISH WITH NO FORK IN TAIL)
- CL - CARAPACE LENGTH (TURTLES)

BRAIL

tallies	Full (1)	7/8 brails (2)	3/4 brails (3)	2/3 brails (4)	1/2 (5)	1/3 (6)	1/4 (7)	1/8 (8)	TOTAL BRAILS	SUM OF ALL BRAILS (see back of form)
	No.	No.	No.	No.	No.	No.	No.	No.		

Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
fullness																														
samples																														

SPECIES CODE	LENGTH (cm)	SPECIES CODE	LENGTH (cm)	SPECIES CODE	LENGTH (cm)	SPECIES CODE	LENGTH (cm)	SPECIES CODE	LENGTH (cm)	SPECIES CODE	LENGTH (cm)
1		21		41		61		81		101	
2		22		42		62		82		102	
3		23		43		63		83		103	
4		24		44		64		84		104	
5		25		45		65		85		105	
6		26		46		66		86		106	
7		27		47		67		87		107	
8		28		48		68		88		108	
9		29		49		69		89		109	
10		30		50		70		90		110	
11		31		51		71		91		111	
12		32		52		72		92		112	
13		33		53		73		93		113	
14		34		54		74		94		114	
15		35		55		75		95		115	
16		36		56		76		96		116	
17		37		57		77		97		117	
18		38		58		78		98		118	
19		39		59		79		99		119	
20		40		60		80		100		120	

column totals	No. / sum of lengths		No. / sum of lengths		No. / sum of lengths		No. / sum of lengths		No. / sum of lengths		No. / sum of lengths
		SKJ		Sub-totals		SKJ		Sub-totals		SKJ	
		YFT				YFT				YFT	
		BET				BET				BET	

	TARGET SPECIES			OTHER SPECIES								
	SKJ	YFT	BET									
Number sampled :												
Sum of lengths :												
Average length :												

Appendix III. Summary of the fit of the Case A ('low resolution') model of the proportion of skipjack, with a *year* covariate and *year* interactions

```
Call:
glm(formula = p_skj ~ (yy + qq + area_s + sch_id_a)^2, family = gaussian(),
    data = s_data, weights = s_data$sets)

Deviance Residuals:
      Min         1Q     Median         3Q        Max
-0.316277  -0.053041   0.003643   0.054975   0.358007

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.412195   0.078194   5.271 6.00e-07 ***
yy1999              0.155353   0.114127   1.361 0.175971
yy2000             -0.018681   0.106599  -0.175 0.861182
yy2001              0.248880   0.104797   2.375 0.019126 *
yy2002              0.059102   0.104797   0.564 0.573817
yy2003              0.262253   0.107729   2.434 0.016378 *
yy2004              0.210918   0.104797   2.013 0.046373 *
yy2005              0.279669   0.106542   2.625 0.009784 **
yy2006              0.350651   0.104797   3.346 0.001092 **
yy2007              0.307205   0.104797   2.931 0.004035 **
yy2008              0.407185   0.104797   3.885 0.000167 ***
yy2009              0.289490   0.106542   2.717 0.007551 **
yy2010              0.301636   0.104797   2.878 0.004729 **
yy2011              0.408342   0.106371   3.839 0.000198 ***
qq2                 0.264104   0.091139   2.898 0.004462 **
qq3                 0.303850   0.090816   3.346 0.001093 **
qq4                 0.367256   0.096624   3.801 0.000227 ***
area_s3            -0.088720   0.070022  -1.267 0.207574
sch_id_aU          -0.067092   0.070008  -0.958 0.339800
yy1999:qq2         -0.410528   0.128988  -3.183 0.001855 **
yy2000:qq2         -0.024747   0.125067  -0.198 0.843480
yy2001:qq2         -0.392095   0.118927  -3.297 0.001283 **
yy2002:qq2         -0.204333   0.118927  -1.718 0.088329 .
yy2003:qq2         -0.255893   0.128988  -1.984 0.049537 *
yy2004:qq2         -0.273397   0.118927  -2.299 0.023228 *
yy2005:qq2         -0.299220   0.125087  -2.392 0.018292 *
yy2006:qq2         -0.363816   0.118927  -3.059 0.002734 **
yy2007:qq2         -0.271176   0.118927  -2.280 0.024347 *
yy2008:qq2         -0.305323   0.118927  -2.567 0.011467 *
yy2009:qq2         -0.308351   0.125087  -2.465 0.015101 *
yy2010:qq2         -0.328181   0.118927  -2.760 0.006689 **
yy2011:qq2         -0.321222   0.118927  -2.701 0.007905 **
yy1999:qq3         -0.320052   0.128963  -2.482 0.014447 *
yy2000:qq3         -0.197454   0.118927  -1.660 0.099442 .
yy2001:qq3         -0.443370   0.118927  -3.728 0.000295 ***
yy2002:qq3         -0.312593   0.118927  -2.628 0.009689 **
yy2003:qq3         -0.424433   0.125346  -3.386 0.000957 ***
yy2004:qq3         -0.459068   0.118927  -3.860 0.000183 ***
yy2005:qq3         -0.537436   0.125087  -4.296 3.53e-05 ***
yy2006:qq3         -0.479969   0.118927  -4.036 9.57e-05 ***
yy2007:qq3         -0.208690   0.118927  -1.755 0.081827 .
yy2008:qq3         -0.582952   0.118927  -4.902 2.98e-06 ***
yy2009:qq3         -0.296069   0.125087  -2.367 0.019525 *
yy2010:qq3         -0.393680   0.118927  -3.310 0.001229 **
yy2011:qq3         -0.409710   0.118927  -3.445 0.000785 ***
yy1999:qq4         -0.497228   0.146059  -3.404 0.000900 ***
yy2000:qq4          0.043030   0.125069   0.344 0.731405
yy2001:qq4         -0.563764   0.125069  -4.508 1.53e-05 ***
yy2002:qq4         -0.297102   0.125069  -2.375 0.019096 *
yy2003:qq4         -0.444198   0.131058  -3.389 0.000946 ***
yy2004:qq4         -0.355711   0.125069  -2.844 0.005231 **
yy2005:qq4         -0.483597   0.130835  -3.696 0.000330 ***
yy2006:qq4         -0.472405   0.125069  -3.777 0.000247 ***
yy2007:qq4         -0.281994   0.125069  -2.255 0.025950 *
yy2008:qq4         -0.627786   0.125069  -5.020 1.80e-06 ***
```

Appendix III (continued)

```
yy2009:qq4        -0.346005   0.130835  -2.645 0.009264 **
yy2010:qq4        -0.330112   0.125069  -2.639 0.009398 **
yy2011:qq4        -0.655889   0.130278  -5.035 1.69e-06 ***
yy1999:area_s3     0.113137   0.114298   0.990 0.324225
yy2000:area_s3     0.013087   0.088473   0.148 0.882656
yy2001:area_s3     0.165477   0.086293   1.918 0.057517 .
yy2002:area_s3     0.311195   0.086293   3.606 0.000452 ***
yy2003:area_s3     0.102832   0.091679   1.122 0.264232
yy2004:area_s3     0.136353   0.086293   1.580 0.116690
yy2005:area_s3     0.172186   0.088483   1.946 0.053975 .
yy2006:area_s3     0.298739   0.086293   3.462 0.000742 ***
yy2007:area_s3     0.141719   0.086293   1.642 0.103123
yy2008:area_s3     0.216232   0.086293   2.506 0.013546 *
yy2009:area_s3     0.184964   0.088483   2.090 0.038677 *
yy2010:area_s3     0.172357   0.086293   1.997 0.048031 *
yy2011:area_s3     0.093716   0.088198   1.063 0.290098
yy1999:sch_id_aU   0.202837   0.105895   1.915 0.057796 .
yy2000:sch_id_aU   0.173879   0.088473   1.965 0.051666 .
yy2001:sch_id_aU   0.157466   0.086293   1.825 0.070501 .
yy2002:sch_id_aU   0.220207   0.086293   2.552 0.011960 *
yy2003:sch_id_aU   0.006760   0.091679   0.074 0.941344
yy2004:sch_id_aU   0.273724   0.086293   3.172 0.001919 **
yy2005:sch_id_aU   0.109873   0.088483   1.242 0.216735
yy2006:sch_id_aU   0.079250   0.086293   0.918 0.360244
yy2007:sch_id_aU  -0.037802   0.086293  -0.438 0.662117
yy2008:sch_id_aU  -0.049710   0.086293  -0.576 0.565642
yy2009:sch_id_aU   0.126645   0.088483   1.431 0.154924
yy2010:sch_id_aU   0.100452   0.086293   1.164 0.246683
yy2011:sch_id_aU   0.128753   0.088198   1.460 0.146932
qq2:area_s3       -0.023636   0.047380  -0.499 0.618787
qq3:area_s3       -0.010894   0.046755  -0.233 0.816149
qq4:area_s3       -0.004532   0.048345  -0.094 0.925461
qq2:sch_id_aU      0.073260   0.047380   1.546 0.124660
qq3:sch_id_aU      0.065594   0.046755   1.403 0.163200
qq4:sch_id_aU     -0.014995   0.047694  -0.314 0.753762
area_s3:sch_id_aU  0.016466   0.034029   0.484 0.629347
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.01414353)

    Null deviance: 5.9157  on 211  degrees of freedom
Residual deviance: 1.7114  on 121  degrees of freedom
AIC: -236.06
Number of Fisher Scoring iterations: 2
```

Appendix IV. Summary of the fit of the Case C ('high resolution') model of the proportion of skipjack, with the *year* covariate and *year : quarter* and *year : area* interaction, to observer data from associated schools

```
Call:
lm(formula = p_sp ~ fit_matrix, weights = this_sch$n_sets)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.36481 -0.28529  0.03171  0.31414  1.52682 

Coefficients: (43 not defined because of singularities)
                                    Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                        8.144e-01  9.404e-02    8.659  < 2e-16 ***
fit_matrixthis_sch$yy1993          2.527e-01  5.308e-01    0.476 0.634025    
fit_matrixthis_sch$yy1994         -4.697e-01  2.787e-01   -1.686 0.092011 .  
fit_matrixthis_sch$yy1995         -2.418e-01  2.601e-01   -0.929 0.352774    
fit_matrixthis_sch$yy1996         -1.867e-01  2.498e-01   -0.747 0.454903    
fit_matrixthis_sch$yy1997         -2.385e-01  2.450e-01   -0.973 0.330431    
fit_matrixthis_sch$yy1998         -5.445e-03  2.570e-01   -0.021 0.983099    
fit_matrixthis_sch$yy1999         -2.519e-01  3.259e-01   -0.773 0.439638    
fit_matrixthis_sch$yy2000         -1.219e-01  2.509e-01   -0.486 0.627209    
fit_matrixthis_sch$yy2001         -3.911e-02  2.517e-01   -0.155 0.876518    
fit_matrixthis_sch$yy2002          9.090e-02  2.470e-01    0.368 0.712850    
fit_matrixthis_sch$yy2003         -1.421e-03  2.471e-01   -0.006 0.995412    
fit_matrixthis_sch$yy2004         -2.121e-01  2.458e-01   -0.863 0.388436    
fit_matrixthis_sch$yy2005          4.608e-02  2.474e-01    0.186 0.852242    
fit_matrixthis_sch$yy2006          1.364e-01  2.439e-01    0.559 0.576071    
fit_matrixthis_sch$yy2007          7.312e-02  2.451e-01    0.298 0.765509    
fit_matrixthis_sch$yy2008         -2.543e-02  2.450e-01   -0.104 0.917332    
fit_matrixthis_sch$yy2009          1.559e-01  2.445e-01    0.638 0.523757    
fit_matrixthis_sch$yy2010          1.321e-01  2.433e-01    0.543 0.587357    
fit_matrixthis_sch$yy2011         -3.080e-02  2.399e-01   -0.128 0.897836    
fit_matrixthis_sch$qq2            -6.156e-02  2.540e-01   -0.242 0.808509    
fit_matrixthis_sch$qq3            -1.247e-01  2.559e-01   -0.487 0.626018    
fit_matrixthis_sch$qq4            -1.684e-01  2.571e-01   -0.655 0.512585    
fit_matrixbasis_lat_lon1           1.977e-02  9.162e-03    2.158 0.031024 *  
fit_matrixbasis_lat_lon2           7.726e-02  3.455e-02    2.236 0.025431 *  
fit_matrixbasis_lat_lon3          -4.967e-02  3.721e-02   -1.335 0.181994    
fit_matrixbasis_lat_lon4           1.522e-01  4.804e-02    3.167 0.001557 ** 
fit_matrixbasis_lat_lon5          -1.522e-01  4.658e-02   -3.268 0.001098 ** 
fit_matrixbasis_lat_lon6          -1.978e-01  5.808e-02   -3.406 0.000670 ***
fit_matrixbasis_lat_lon7           1.539e-01  5.605e-02    2.745 0.006085 ** 
fit_matrixbasis_lat_lon8           1.495e-02  9.418e-03    1.588 0.112509    
fit_matrixbasis_lat_lon9          -1.317e-01  4.098e-02   -3.213 0.001329 ** 
fit_matrixbasis_lat_lon10          1.278e-01  4.940e-02    2.587 0.009741 ** 
fit_matrixbasis_lat_lon11          8.314e-01  2.310e-01    3.599 0.000325 ***
fit_matrixbasis_lat_lon12          1.367e-02  8.023e-03    1.703 0.088608 .  
fit_matrixbasis_lat_lon13         -4.657e-02  1.743e-02   -2.672 0.007590 ** 
fit_matrixthis_sch$flag_idEC       1.798e-01  2.177e-01    0.826 0.408833    
fit_matrixthis_sch$flag_idES       1.315e-01  5.028e-02    2.615 0.008979 ** 
fit_matrixthis_sch$flag_idFM      -3.489e-02  2.493e-02   -1.400 0.161700    
fit_matrixthis_sch$flag_idID      -5.353e-01  1.694e-01   -3.160 0.001594 ** 
fit_matrixthis_sch$flag_idJP       3.535e-03  2.425e-02    0.146 0.884096    
fit_matrixthis_sch$flag_idKI      -1.823e-02  4.411e-02   -0.413 0.679401    
fit_matrixthis_sch$flag_idKR       2.106e-03  2.286e-02    0.092 0.926590    
fit_matrixthis_sch$flag_idMH      -5.647e-03  2.329e-02   -0.242 0.808463    
fit_matrixthis_sch$flag_idNZ      -7.284e-02  6.216e-02   -1.172 0.241368    
fit_matrixthis_sch$flag_idPG      -2.797e-02  2.089e-02   -1.339 0.180648    
fit_matrixthis_sch$flag_idPH      -1.332e-02  2.245e-02   -0.593 0.553147    
fit_matrixthis_sch$flag_idSB      -1.107e-01  3.213e-02   -3.446 0.000579 ***
fit_matrixthis_sch$flag_idSV      -3.658e-01  1.020e-01   -3.588 0.000340 ***
fit_matrixthis_sch$flag_idTV      -1.889e-01  1.010e-01   -1.870 0.061569 .  
fit_matrixthis_sch$flag_idTW      -3.042e-02  2.188e-02   -1.390 0.164690    
fit_matrixthis_sch$flag_idUS       4.915e-04  2.109e-02    0.023 0.981413    
fit_matrixthis_sch$flag_idVU       2.209e-02  2.924e-02    0.755 0.450183    
fit_matrixthis_sch$sch_id_xF       2.048e-02  1.092e-02    1.876 0.060793 .  
fit_matrixthis_sch$sch_id_xL       4.981e-03  9.432e-03    0.528 0.597460    
```

Appendix IV (continued)

```
fit_matrixthis_sch$sch_id_xO           -6.679e-02  1.667e-02    -4.008 6.31e-05 ***
fit_matrixthis_sch$yy_qqyy_qq_1993_4   -9.202e-01  4.890e-01    -1.882 0.059985 .
fit_matrixthis_sch$yy_qqyy_qq_1994_1    4.736e-02  3.165e-01     0.150 0.881068
fit_matrixthis_sch$yy_qqyy_qq_1994_2    7.122e-01  2.431e-01     2.930 0.003418 **
fit_matrixthis_sch$yy_qqyy_qq_1995_1   -6.280e-01  5.478e-01    -1.146 0.251710
fit_matrixthis_sch$yy_qqyy_qq_1995_2   -3.442e-01  1.383e-01    -2.489 0.012880 *
fit_matrixthis_sch$yy_qqyy_qq_1995_3   -1.672e-01  1.776e-01    -0.942 0.346414
fit_matrixthis_sch$yy_qqyy_qq_1996_1   -5.851e-02  2.823e-01    -0.207 0.835800
fit_matrixthis_sch$yy_qqyy_qq_1996_2    1.447e-01  9.924e-02     1.459 0.144822
fit_matrixthis_sch$yy_qqyy_qq_1996_3    1.790e-02  9.697e-02     0.185 0.853538
fit_matrixthis_sch$yy_qqyy_qq_1997_1   -1.120e-02  2.667e-01    -0.042 0.966498
fit_matrixthis_sch$yy_qqyy_qq_1997_2   -5.750e-02  8.303e-02    -0.693 0.488626
fit_matrixthis_sch$yy_qqyy_qq_1997_3    1.423e-01  1.633e-01     0.871 0.383883
fit_matrixthis_sch$yy_qqyy_qq_1998_1   -4.674e-01  2.733e-01    -1.710 0.087301 .
fit_matrixthis_sch$yy_qqyy_qq_1998_2   -2.173e-01  9.596e-02    -2.265 0.023625 *
fit_matrixthis_sch$yy_qqyy_qq_1998_3   -2.699e-01  1.066e-01    -2.533 0.011361 *
fit_matrixthis_sch$yy_qqyy_qq_1999_1   -1.266e-01  3.402e-01    -0.372 0.709838
fit_matrixthis_sch$yy_qqyy_qq_1999_2   -6.454e-02  2.214e-01    -0.291 0.770713
fit_matrixthis_sch$yy_qqyy_qq_1999_3    8.517e-02  2.221e-01     0.383 0.701414
fit_matrixthis_sch$yy_qqyy_qq_2000_1   -4.067e-01  2.677e-01    -1.520 0.128754
fit_matrixthis_sch$yy_qqyy_qq_2000_2   -2.105e-01  7.694e-02    -2.737 0.006249 **
fit_matrixthis_sch$yy_qqyy_qq_2000_3   -8.852e-02  8.425e-02    -1.051 0.293508
fit_matrixthis_sch$yy_qqyy_qq_2001_1   -1.496e-01  2.694e-01    -0.555 0.578633
fit_matrixthis_sch$yy_qqyy_qq_2001_2   -1.768e-01  8.044e-02    -2.198 0.028035 *
fit_matrixthis_sch$yy_qqyy_qq_2001_3   -1.250e-01  8.346e-02    -1.497 0.134428
fit_matrixthis_sch$yy_qqyy_qq_2002_1   -2.285e-01  2.600e-01    -0.879 0.379585
fit_matrixthis_sch$yy_qqyy_qq_2002_2   -1.572e-01  5.378e-02    -2.923 0.003495 **
fit_matrixthis_sch$yy_qqyy_qq_2002_3    8.415e-03  5.756e-02     0.146 0.883770
fit_matrixthis_sch$yy_qqyy_qq_2003_1   -1.903e-01  2.601e-01    -0.732 0.464394
fit_matrixthis_sch$yy_qqyy_qq_2003_2   -8.356e-02  5.884e-02    -1.420 0.155717
fit_matrixthis_sch$yy_qqyy_qq_2003_3   -1.688e-01  6.213e-02    -2.717 0.006629 **
fit_matrixthis_sch$yy_qqyy_qq_2004_1   -8.764e-02  2.589e-01    -0.338 0.735061
fit_matrixthis_sch$yy_qqyy_qq_2004_2   -5.281e-02  4.876e-02    -1.083 0.278904
fit_matrixthis_sch$yy_qqyy_qq_2004_3   -4.657e-02  5.465e-02    -0.852 0.394236
fit_matrixthis_sch$yy_qqyy_qq_2005_1   -1.557e-01  2.593e-01    -0.600 0.548346
fit_matrixthis_sch$yy_qqyy_qq_2005_2   -1.779e-01  4.855e-02    -3.665 0.000252 ***
fit_matrixthis_sch$yy_qqyy_qq_2005_3   -1.266e-01  5.345e-02    -2.369 0.017928 *
fit_matrixthis_sch$yy_qqyy_qq_2006_1   -8.883e-02  2.588e-01    -0.343 0.731422
fit_matrixthis_sch$yy_qqyy_qq_2006_2   -1.336e-01  4.792e-02    -2.787 0.005358 **
fit_matrixthis_sch$yy_qqyy_qq_2006_3   -7.853e-02  5.472e-02    -1.435 0.151358
fit_matrixthis_sch$yy_qqyy_qq_2007_1   -6.760e-02  2.589e-01    -0.261 0.794057
fit_matrixthis_sch$yy_qqyy_qq_2007_2   -9.207e-02  4.873e-02    -1.890 0.058918 .
fit_matrixthis_sch$yy_qqyy_qq_2007_3   -1.061e-01  5.511e-02    -1.925 0.054354 .
fit_matrixthis_sch$yy_qqyy_qq_2008_1   -8.642e-02  2.600e-01    -0.332 0.739603
fit_matrixthis_sch$yy_qqyy_qq_2008_2   -1.533e-02  5.250e-02    -0.292 0.770346
fit_matrixthis_sch$yy_qqyy_qq_2008_3   -2.018e-02  5.834e-02    -0.346 0.729414
fit_matrixthis_sch$yy_qqyy_qq_2009_1   -7.320e-02  2.597e-01    -0.282 0.778078
fit_matrixthis_sch$yy_qqyy_qq_2009_2   -1.848e-01  5.289e-02    -3.494 0.000485 ***
fit_matrixthis_sch$yy_qqyy_qq_2009_3   -1.257e-01  5.905e-02    -2.128 0.033413 *
fit_matrixthis_sch$yy_qqyy_qq_2010_1   -1.952e-01  2.579e-01    -0.757 0.449126
fit_matrixthis_sch$yy_qqyy_qq_2010_2   -1.801e-01  4.455e-02    -4.043 5.42e-05 ***
fit_matrixthis_sch$yy_qqyy_qq_2010_3   -1.684e-01  5.765e-02    -2.922 0.003509 **
fit_matrixthis_sch$yy_qqyy_qq_2011_1   -5.469e-02  2.546e-01    -0.215 0.829952
fit_matrixthis_sch$area_yyarea_yy_2_1998  8.659e-02  5.454e-02     1.588 0.112453
fit_matrixthis_sch$area_yyarea_yy_2_1999  7.378e-02  5.652e-02     1.305 0.191914
fit_matrixthis_sch$area_yyarea_yy_2_2000  8.268e-02  5.317e-02     1.555 0.120059
fit_matrixthis_sch$area_yyarea_yy_2_2001 -3.183e-02  4.749e-02    -0.670 0.502808
fit_matrixthis_sch$area_yyarea_yy_2_2002 -1.692e-01  4.849e-02    -3.490 0.000491 ***
fit_matrixthis_sch$area_yyarea_yy_2_2003  4.092e-02  5.307e-02     0.771 0.440740
fit_matrixthis_sch$area_yyarea_yy_2_2004  2.062e-01  4.433e-02     4.652 3.45e-06 ***
fit_matrixthis_sch$area_yyarea_yy_2_2005 -6.855e-02  5.292e-02    -1.295 0.195308
fit_matrixthis_sch$area_yyarea_yy_2_2006 -7.614e-02  3.945e-02    -1.930 0.053727 .
fit_matrixthis_sch$area_yyarea_yy_2_2007 -1.949e-05  4.038e-02 -0.000483 0.999615
fit_matrixthis_sch$area_yyarea_yy_2_2008  1.470e-02  4.008e-02     0.367 0.713766
fit_matrixthis_sch$area_yyarea_yy_2_2009 -6.886e-02  4.150e-02    -1.659 0.097200 .
fit_matrixthis_sch$area_yyarea_yy_2_2010 -3.248e-02  3.147e-02    -1.032 0.302035
fit_matrixthis_sch$area_yyarea_yy_2_2011  1.785e-02  3.292e-02     0.542 0.587771
```

Appendix IV (continued)

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4733 on 2564 degrees of freedom
Multiple R-Squared: 0.4239,     Adjusted R-squared: 0.3967
F-statistic: 15.59 on 121 and 2564 DF,  p-value: < 2.2e-16
```

Appendix V. Formulae for scaling length frequencies by the catch in strata of *year – quarter – 5° x 5° – school association* and strata of *year – quarter – MFCL Area – school association*

The length frequencies for each species, determined from corrected grab samples and spill samples, and raised by the set weight, were scaled by the catch using the formulae below. First, we derive the length frequencies in strata of *year – quarter – 5° x 5° – school association*, raised by the set weights but not scaled by the catch:

$$n'_{ijs} = N_{is} \cdot P_{ijs} \qquad (16)$$

where n'_{ijs} is the number of fish of species *i* and length *j* in stratum *s*, raised by the set weights, N_{is} is the number of fish of species *i* that were sampled in stratum *s* , and P_{ijs} is the proportion of fish of species *i* and length *j* in stratum *s*. We have

$$N_{is} = \sum_j \sum_k n_{ijk} \qquad (17)$$

$$P_{ijs} = \frac{\sum_k n'_{ijk}}{\sum_j \sum_k n'_{ijk}} \qquad (18)$$

where n_{ijk} is the number of fish of species *i* and length *j* that were sampled in set *k*, and n'_{ijk} is the number of fish of species *i* and length *j* in set *k*, raised by the set weight. The sums over sets is for all sets within stratum *s*. Further,

$$n'_{ijk} = \frac{w'_{ijk}}{\overline{w}_{ij}} \qquad (19)$$

$$w'_{ijk} = W_k \cdot P_{ijk} \qquad (20)$$

$$P_{ijk} = \frac{\frac{w_{ijk}}{A_j}}{\sum_i \sum_j \frac{w_{ijk}}{A_j}} \qquad (21)$$

$$= \frac{\frac{n_{ijk} \cdot \overline{w}_{ij}}{A_j}}{\sum_i \sum_j \frac{n_{ijk} \cdot \overline{w}_{ij}}{A_j}} \qquad (22)$$

where w'_{ijk} is the weight of fish of species *i* and length *j* that were sampled in set *k*, raised by the set weight; $\overline{w}_{ij}$ is the average weight of fish of species i and length j ; W_k is the weight of the catch from set k ; P_{ijk} is the proportion of species i and length j in the catch of set k ; w'_{ijk} is the weight of fish of species *i* and length *j* that were sampled in set *k*, raised by the set weight; w_{ijk} is the weight of fish of species *i* and length *j* that were sampled in set *k*, not raised by the set weight; and A_j is the *availability* of length *j*. All sets *k* are in stratum *s*.

Second, we take the length frequencies in the stratum, raised by the set weights, and scale them by the catch:

$$n''_{ijs} = \frac{c_{ijs}}{\overline{w}_{ij}} \tag{23}$$

where n''_{ijs} is the number of fish of species *i* and length *j* in stratum *s*, scaled by the catch; and c_{ijs} is the catch of fish of species *i* and length *j* in stratum *s*. We have

$$c_{ijs} = C_{is} \cdot \frac{w'_{ijs}}{\sum_j w'_{ijs}} \tag{24}$$

$$w'_{ijs} = \sum_k w'_{ijk} \tag{25}$$

where C_{is} is the catch fish of species *i* in stratum *s* and w'_{ijs} is the weight of sampled fish of species *i* and length *j* in stratum *s*, raised by the set weight. The sum over sets is for all sets within stratum *s*.

Finally, we derive scaled length frequencies for strata of *year – quarter – MFCL Area – school association* from the length frequencies in the stratum, raised by the set weights and scaled by the catch:

$$n''_{ijt} = N_{it} \cdot P_{ijt} \tag{26}$$

$$N_{it} = \sum_j \sum_k n_{ijk} \tag{27}$$

where n''_{ijt} is the number of fish of species *i* and length *j* in MFCL stratum *t*, scaled by the catch; N_{it} is the number of fish of species i that were sampled in MFCL stratum t ; n_{ijk} is the number of fish of species *i* and length *j* in set k that were sampled from sets within stratum t ; and P_{ijt} is the proportion of the numbers of fish of species *i* and length *j* in MFCL stratum *t*. We have

$$P_{ijt} = \frac{n''_{ijt}}{\sum_j n''_{ijt}} \tag{28}$$

$$n''_{ijt} = \sum_{s} n''_{ijs} \tag{29}$$

where n''_{ijs} is the number of fish of species *i* and length *j* in 5° x 5° stratum *s*, scaled by the catch. The sum over 5° x 5° strata is for all strata within MFCL stratum *t*.