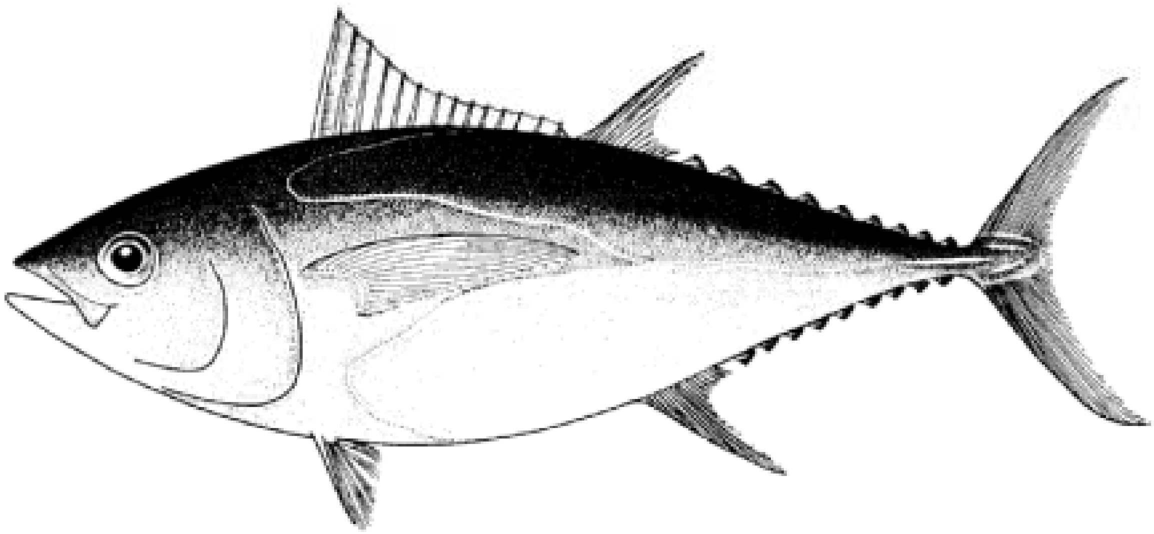
SCTB14 Working Paper

SWG–8

Estimation of bigeye catches by purse seiners



Keith Bigelow

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

1 Introduction

The purse seine fishery in the western and central Pacific Ocean is largely composed of vessels from Japan, Korea, Taiwan and the United States. These fleets catch a variety of tunas, such as skipjack (*Katsuwonus pelamis*), yellowfin (*Thunnus albacares*), and bigeye (*Thunnus obesus*). Reported catch estimates from logsheets are biased for yellowfin and bigeye because bigeye are rarely distinguished from yellowfin. This results in an over-estimate of yellowfin and a corresponding under-estimate of bigeye in the fishery.

Bigeye catch estimates are determined for the Japanese and US fleets by port sampling. Estimates for the remaining fleets are based on the assumption that the proportions of bigeye in associated and unassociated sets are similar to those for the US fleet, though the procedure may be biased because bigeye proportions in unassociated and associated sets conducted by the other fleets may differ due to various factors (e.g. time and space variability, fleet behavior, gear modifications).

Estimates of yellowfin and bigeye catches in the purse seine fishery have been modified by the NMFS and SPC to account for misidentification (Coan et al. 2001, Lawson 2000). Modifications are largely dependent on a comprehensive NMFS port sampling programme that monitors purse seine landings in Pago Pago, American Samoa where 85% of the US catch was landed in 2000 (Coan et al. 2001). Through port sampling, bigeye catches for the entire US fishery are produced by estimating the proportion of bigeye in the combined catches of yellowfin and bigeye. Annual estimates for the remaining fleets are produced by applying the proportion of bigeye in the US catch (described in Lawson 2000). The proportions are weighted by school type because the composition of yellowfin and bigeye is higher in associated purse seine sets (e.g. drifting rafts, logs and FADs) than in unassociated or free swimming sets.

A tree-based regression method was presented at SCTB12 (Bigelow 1999) as an alternative to extrapolation to estimate purse-seine bigeye catch.

The objectives of this paper are to:

1. Briefly compare the composition of bigeye in the yellowfin+bigeye estimates from port and at-sea observer sampling programmes.
2. Consider other models (additive and delta) to examine factors affecting the composition of bigeye in the purse-seine fishery.
3. Provide an update of bigeye catches using the tree-based regression approach that considers all available port and at-sea observer data.

2 Data

Species composition data by length were available from port and observer sampling (Table 1). Length data were converted into weight ($W\ (kg) = 8.6388E\text{-}06 * L\ (cm)^{3.2174}$). A total of 5,840 samples could be related to other additional factors such as time (year and month), location (5° square), set type and vessel flag. There were 4,599 observer and 1,241 port samples. Observer data were compiled from the USMLT, SPC and Micronesian Fisheries Authority (MFA) programs. Overall, the US fleet was the most sampled (n=2,627).

The types of purse seine sets sampled by port samplers and observers were: 1) free swimming school, 2) feeding on baitfish, 3) drifting log/debris, 4) drifting FAD, 5) anchored FAD, 6) whale, 7) whale shark and 8) boat. These set types were reduced into three set types for further statistical analysis: 1) unassociated (free swimming school and feeding on baitfish), 2) drifting log and 3) FAD (drifting, anchored, whale, whale shark and boat). In general, the spatial distribution of samples from the individual fleets followed the distribution of actual fleet effort (Figures 1–4). One exception is the Philippines domestic purse-seine fishery where no species composition samples were available.

Prediction models for bigeye composition were constructed using variables in the port-sampling and observer data that were similar to variables in the corresponding logbook data in order to estimate catches. Variables considered were: 1) year, 2) month, 3) set type (1=unassociated, 2=log and 3=FAD sets), 4) fleet, 5) latitude and 6) longitude.

3 Comparison of bigeye composition estimates from port and at-sea observer sampling programmes

The annual proportions of bigeye in the bigeye+yellowfin estimates were compared for port and observer sampling (Figures 5–6). For unassociated sets, the bigeye composition was relatively low compared to associated sets (cf. Figures 5 & 6). From observer sampling, the median estimates were essentially zero (Figure 5A); however, bigeye comprised 5–10% of the bigeye+yellowfin composition in the initial years of the US port-sampling programme.

For associated sets (drifting logs and FADs), there were differences between the annual estimates based on sampling programmes for a given fleet (Figure 6). The composition of bigeye was consistently higher in port samples compared to observer samples for both the Japanese (Figure 5A & B) and US fleets (Figure 5C & D). The differences in annual estimates between sampling programmes could be related to vessel differences, spatial or temporal effects or reflect the high variability in bigeye composition in associated sets. There is a need to further compare the two sampling programmes for sources of bias.

4 Models for analyzing factors affecting the composition of bigeye and predicting total catch

4.1 Geneneralized additive (GAMs) and delta models

A generalized additive model (GAM) was fit to the six variables to determine the relationship between explanatory variables and bigeye composition (Mean bigeye composition in bigeye+yellowfin=*f*(*year, month, set type, fleet, latitude, longitude*) assuming a Poisson distribution. Four of the variables (*year, month, set type, fleet*) were considered as categorical, whereas *latitude* and *longitude* were considered as continuous. The continuous variables were modeled by a *loess* smoothing function.

All variables in the GAM were considered highly significant from a stepwise GAM process. The relative effects of the explanatory variables on bigeye composition is illustrated in Figure 7. The relative magnitude of the explanatory variables can be inferred from the relative *y*-axis ranges. The variable *latitude* has the largest effect (*y*-axis range of 4) and suggests that bigeye composition is low at high latitudes such as New Zealand. Though *latitude* has the largest effect it is relatively unimportant because there are few samples outside the traditional latitudinal range of the fishery (5°N–15°S).

The effects on bigeye composition are intermediate for *set type*, *year* and *flag*. As expected, unassociated sets have significantly less bigeye than log and FAD sets. The *year* effect indicates that bigeye proportion has been declining continuously since 1996 when all other variables are considered in the model. The effect of *flag* is lowest for the Solomon Islands fleet, but could result from a small sample size.

The effects on bigeye composition are smallest for *month* and *longitude*. The effect of *month* is varied without trend. The effect of *longitude* indicates that bigeye composition is low in the western Pacific (~130°E), increases steadily throughout the western Pacific, peaks at the dateline (180°) and declines again in the central Pacific from 175° to 155°W.

Preliminary delta models were also fit to the port and observer sampling data. The model is fit in a two stage process, with the first model fit to the entire data set with the bigeye composition as presence or absence and a second model that includes only samples where the bigeye composition was greater than zero. The first model assumed a binomial distribution whereas the second model used a normal, lognormal or gamma distribution. The overall mean bigeye composition is then the product of the probability of success catch from the binomial model coupled with the model on non-zero catches. The initial results were not encouraging with the delta approach as the binomial model produced unrealistic estimates of the probability of positive bigeye catches.

4.2 Tree-based regression

Tree-based regression models are used in classification problems, whereby data are split by binary partitioning into separate subgroups. The model continues to find splits until no further improvement or splitting is necessary. A regression tree was initially grown from the entire set of possible predictors. The tree was examined in terms of its predictors, residual mean deviance, residuals and normal probability plot of residuals. The tree was pruned and sniped through cross-validation procedures to reduce overfitting. Cross-validation indicated a tree with approximately 70 branches (nodes) would be sufficient (Figure 8).

The final tree included all predictor variables. The order of relative importance was 1) set type, 2) year, 3) month, 4) fleet, 5) longitude and 6) latitude. The residuals were not normally distributed in the tree-based regression similar to the GAM, because bigeye proportion data contained a large amount of zero observations. The regression explained slightly greater than 20% of the variance in the data. Predictions from the tree results were applied to the entire fishery dataset in order to estimate total bigeye catch by fleet. The SPC "Best database" was used in conjunction with the tree results to predict the bigeye proportion using the predictor variables. The total catch was then estimated by multiplying the bigeye proportion by the reported bigeye+yellowfin estimates.

A comparison of bigeye proportions by set types for the US fleet is provided in Table 2. In general, the regression estimates of the bigeye proportion in unassociated sets is consistently higher than the port-sampling data presented in Lawson (2000, proportions provided by NMFS), whereas the proportions in associated sets vary between years.

Total annual bigeye catch estimates for the US fleet are compared from the two different estimation techniques (Table 3). Estimates from 1988 to 1999 were 31% higher based on the extrapolation of port-sampling data than the regression technique. The discrepancy results from the inclusion of observer data in the regression estimation which has a lower proportion of bigeye than the port-sampling results.

Annual estimates for the western and central Pacific Ocean (WCPO) range from 7 to 36 thousand tonnes based on extrapolation of port-sampling data and 9 to 37 thousand tonnes based on the regression (Figure 9). Contrary to the US fleet estimates, the regression estimates for the entire WCPO were greater than the extrapolation method in all years except 1996. Since 1988, annual regression estimates averaged 20% greater than estimates made by extrapolation.

The annual predicted proportions of bigeye in the bigeye+yellowfin estimates are presented for each fleet by set type (unassociated, Log & FAD) in Table 4. In general the bigeye proportion in log and FAD sets ranges from 15–30% of the total yellowfin+ bigeye.

Using these proportions from the regression technique, the spatial distribution of bigeye catch was calculated for 1999 (Figure 11). The fleets of Japan, Taiwan, Philippines and the US had the highest catches of bigeye. The US fleet in particular has high catches of bigeye given the fleet behavior to conduct FAD associated sets and the distribution near the dateline where the proportion of bigeye appears higher than in the western Pacific.

Conclusions & Recommendations

- The Statistics Working Group is invited to note the content of the paper as an alternative method to estimate bigeye catches by purse-seine vessels in the WCPO.
- The GAM and tree-based regressions had similar results on the factors influencing the proportion of bigeye in purse-seine catches.
- The tree-based method has an advantage over the extrapolation from port-sampling by statistically incorporating differences in fleet behaviour and temporal and spatial variability which are not inherent in the estimates by extrapolation.
- The discrepancy between the port-sampling and observer estimates requires further attention. Bigeye composition in associated set types is characterised by high variability between sets; however, there may be continued identification difficulties in the observer sampling that leads to a downward bias or factors in the port-sampling that represent an upward bias.

5 Acknowledgements

We are grateful to Naozumi Miyabe of the Japan National Research Institute of Far Seas Fisheries, who compiled the port-sampling data for the Japanese purse-seine fishery.

6 References

Bigelow, K. 1999. Estimating bigeye composition in the purse seine fishery. Working Paper SWG–3, Twelth Meeting of the Standing Committee on Tuna and Billfish, Papeete, Tahiti, Fench Polynesia. 16–23 June 1999.

Coan, A., J. Childers, R. Ito, B. Kikkawa & D. Hamm. Summary of U.S. fisheries statistics for highly migratory species in the central-western Pacific, 1996-2000. National Fisheries Report–USA, Fourteenth Meeting of the Standing Committee on Tuna and Billfish, Noumea, New Caledonia. 9–16 August 2000.

Lawson, T.A. 2000. SPC Tuna Fishery Yearbook, 1999. Secretariat of the Pacific Community, Noumea, New Caledonia. 155 pp.

Table 1. Number of bigeye composition samples from the purse-seine fishery (1988–2000).

Fleet	Number of observer samples		Number of port samples	
	Unassociated	Associated	Unassociated	Associated
FSM	35	95		
Japan	85	227	74	246
Korea	396	241		
PNG	26	86		
Philippines	2	225		
Solomon Is.	4	51		
Taiwan	598	745		
USA	246	1460	112	809
Vanuatu	45	32		
Total	1437	3162	186	1055

Table 2. Comparison of bigeye proportions by set type for the US purse seine fishery based on port-sampling data (Lawson 2000) and tree-based regression.

Year	US Port sampling (Lawson 2000) Proportion of bigeye in YFT+BET		Tree-based estimates Proportion of bigeye in YFT+BET	
	Unassociated	Associated	Unassociated	Associated
1988	0.030	0.089	0.050	0.127
1989	0.005	0.160	0.033	0.154
1990	0.003	0.116	0.027	0.228
1991	0.010	0.101	0.028	0.212
1992	0.007	0.139	0.043	0.132
1993	0.009	0.129	0.028	0.125
1994	0.004	0.116	0.028	0.225
1995	0.009	0.165	0.033	0.156
1996	0.013	0.407	0.027	0.243
1997			0.033	0.245
1998			0.029	0.243

Table 3. Comparison of annual catches (metric tonnes) of bigeye by purse seiners in the WCPO based on port-sampling data (Lawson 2000) and tree-based regression.

Year	US fleet			All fleets		
	Port-sampling	Regression	% difference	Port-sampling	Regression	% difference
1988	1,948	1,078	81	7,305	8,921	-18
1989	2,421	2,888	-16	12,651	16,640	-24
1990	1,762	3,522	-50	12,143	24,221	-50
1991	1,550	2,748	-44	13,406	23,391	-43
1992	3,480	3,479	0	19,384	19,610	-1
1993	3,731	3,539	5	14,286	16,267	-12
1994	1,711	3,101	-45	11,178	21,848	-49
1995	3,190	2,413	32	14,222	16,046	-11
1996	9,860	3,591	175	18,244	17,945	2
1997	10,058	8,875	13	31,637	36,901	-14
1998	5,561	3,954	41	18,342	23,830	-23
1999	17,403	8,678	101	35,667	36,082	-1

Table 4. Proportions of bigeye tuna in the yellowfin+bigeye by set type (unassociated, log and FAD) and fleet estimated by tree-based regression.

	Fleet				Fleet				Fleet			
	AU				ES				FM			
Year	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled
1988	0.037	0.336		0.236								
1989	0.037	0.325		0.152								
1990	0.045	0.271	0.261	0.171								
1991	0.347	0.283	0.449	0.299					0.037	0.091	0.091	0.076
1992	0.035	0.278	0.111	0.187					0.036	0.102	0.096	0.074
1993	0.027	0.449	0.449	0.164					0.030	0.098	0.093	0.071
1994									0.033	0.161	0.172	0.103
1995									0.036	0.173	0.111	0.091
1996									0.020	0.154	0.144	0.137
1997									0.035	0.182	0.121	0.117
1998									0.033	0.110	0.123	0.057
1999					0.037		0.202	0.201	0.032	0.115	0.139	0.107
Avg	0.088	0.324	0.317	0.201	0.037		0.202	0.201	0.032	0.132	0.121	0.093

	Fleet				Fleet				Fleet			
	ID				JP				KI			
Year	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled
1988	0.085	0.098	0.100	0.096	0.040	0.100	0.100	0.092				
1989	0.098	0.131	0.169	0.125	0.038	0.139	0.117	0.107				
1990	0.086	0.211	0.231	0.150	0.042	0.188	0.156	0.112				
1991	0.071	0.158	0.154	0.114	0.039	0.170	0.131	0.094				
1992	0.102	0.100	0.100	0.101	0.051	0.099	0.100	0.072				
1993	0.088	0.097	0.099	0.092	0.032	0.098	0.099	0.059				
1994	0.093	0.171	0.172	0.123	0.031	0.195	0.202	0.110	0.017	0.276		0.273
1995	0.107	0.123	0.133	0.116	0.038	0.126	0.122	0.075	0.037	0.166	0.111	0.077
1996	0.091	0.146	0.197	0.142	0.042	0.180	0.095	0.124	0.037	0.289	0.460	0.208
1997	0.107	0.184	0.182	0.154	0.041	0.139	0.108	0.100	0.041	0.282	0.354	0.128
1998	0.094	0.182	0.204	0.124	0.033	0.127	0.143	0.061	0.033	0.319		0.114
1999	0.100	0.162	0.192	0.158	0.037	0.172	0.184	0.142	0.025	0.127	0.192	0.097
Avg	0.094	0.147	0.161	0.125	0.039	0.144	0.130	0.096	0.032	0.243	0.279	0.150

	Fleet				Fleet				Fleet			
	KR				PG				PH			
Year	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled
1988	0.041	0.098	0.106	0.085					0.080	0.128	0.122	0.118
1989	0.030	0.156	0.230	0.093					0.057	0.169	0.179	0.157
1990	0.028	0.272	0.354	0.157					0.120	0.244	0.240	0.232
1991	0.032	0.181	0.172	0.097					0.046	0.153	0.211	0.156
1992	0.040	0.098	0.100	0.068					0.106	0.123	0.134	0.126
1993	0.030	0.100	0.101	0.054					0.053	0.131	0.123	0.115
1994	0.029	0.217	0.168	0.095	0.063	0.291	0.234	0.118	0.093	0.242	0.226	0.180
1995	0.038	0.164	0.217	0.073	0.078	0.162	0.160	0.102	0.102	0.164	0.162	0.147
1996	0.041	0.234	0.239	0.099	0.055	0.222	0.222	0.101	0.084	0.242	0.236	0.223
1997	0.035	0.224	0.277	0.108	0.017	0.306	0.245	0.268	0.060	0.238	0.244	0.236
1998	0.031	0.231	0.252	0.060	0.029	0.237	0.244	0.233	0.017	0.247	0.237	0.219
1999	0.030	0.220	0.236	0.086	0.052	0.232	0.241	0.238	0.099	0.243	0.246	0.237
Avg	0.034	0.183	0.204	0.090	0.049	0.242	0.224	0.177	0.076	0.194	0.197	0.179

Table 4 con't. Proportions of bigeye tuna in the yellowfin+bigeye by set type (unassociated, log and FAD) and fleet estimated by tree-based regression.

	Fleet SB				Fleet SU				Fleet TW			
Year	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled
1988	0.029	0.135	0.141	0.131					0.061	0.126	0.136	0.122
1989	0.029	0.186	0.187	0.177					0.073	0.179	0.239	0.163
1990	0.026	0.244	0.230	0.208					0.045	0.248	0.232	0.233
1991	0.027	0.191	0.197	0.159					0.061	0.175	0.196	0.158
1992	0.029	0.125	0.130	0.104					0.046	0.134	0.140	0.094
1993	0.031	0.117	0.129	0.110	0.031	0.116	0.114	0.068	0.042	0.126	0.123	0.065
1994	0.033	0.236	0.244	0.197	0.025	0.222	0.222	0.133	0.039	0.254	0.266	0.101
1995	0.034	0.163	0.174	0.139					0.048	0.155	0.180	0.071
1996	0.035	0.252	0.257	0.204					0.071	0.236	0.235	0.117
1997	0.033	0.244	0.245	0.196					0.044	0.253	0.261	0.123
1998	0.021	0.231	0.245	0.187					0.031	0.231	0.234	0.051
1999	0.023	0.251	0.240	0.179					0.033	0.247	0.257	0.148
Avg	0.029	0.198	0.202	0.166	0.028	0.169	0.168	0.100	0.050	0.197	0.208	0.121

	Fleet US				Fleet VU			
Year	Unass	Log	FAD	Pooled	Unass	Log	FAD	Pooled
1988	0.050	0.127	0.114	0.088				
1989	0.033	0.153	0.262	0.065				
1990	0.027	0.228	0.244	0.062				
1991	0.028	0.212	0.121	0.069				
1992	0.043	0.132	0.157	0.081				
1993	0.028	0.125	0.092	0.071				
1994	0.028	0.225	0.229	0.051	0.037	0.348		
1995	0.033	0.150	0.278	0.072	0.035	0.171	0.282	0.113
1996	0.027	0.239	0.245	0.168	0.062	0.259	0.337	0.216
1997	0.033	0.251	0.234	0.152	0.034	0.259	0.231	0.157
1998	0.029	0.238	0.250	0.096	0.032	0.252	0.223	0.100
1999	0.033	0.229	0.247	0.240	0.033	0.263	0.274	0.180
Avg	0.033	0.192	0.206	0.101	0.039	0.259	0.269	0.153

Figure 1. Comparison of the spatial distribution of samples from the US observer program (left) and actual fleet distribution (1997–2000, right).

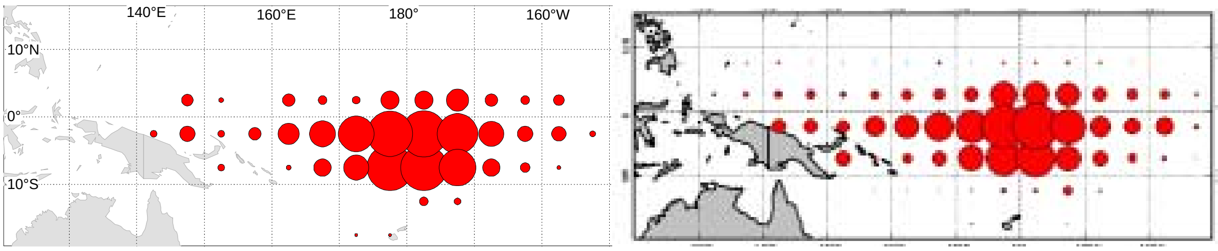


Figure 2. Comparison of the spatial distribution of samples from the US port-sampling program (left) and actual fleet distribution (1988–2000, right).

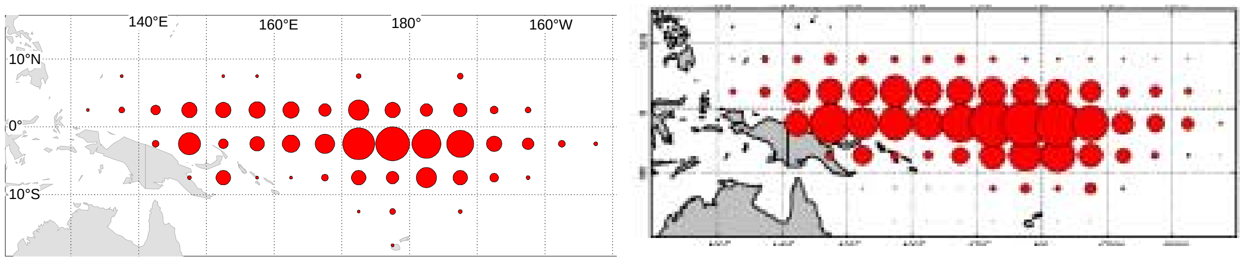


Figure 3. Comparison of the spatial distribution of the samples from observers on non-US vessels (left) and actual fleet distribution (1997–2000, right).

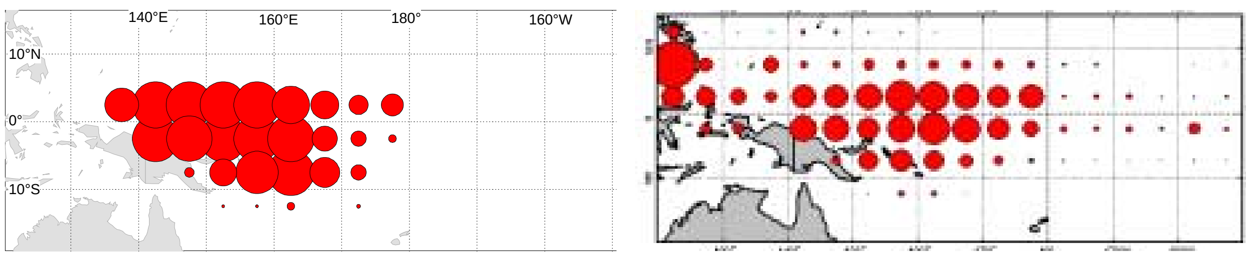


Figure 4. Comparison of the spatial distribution of the samples from the Japanese port-sampling program (left) and actual fleet distribution (1995–2000, right).

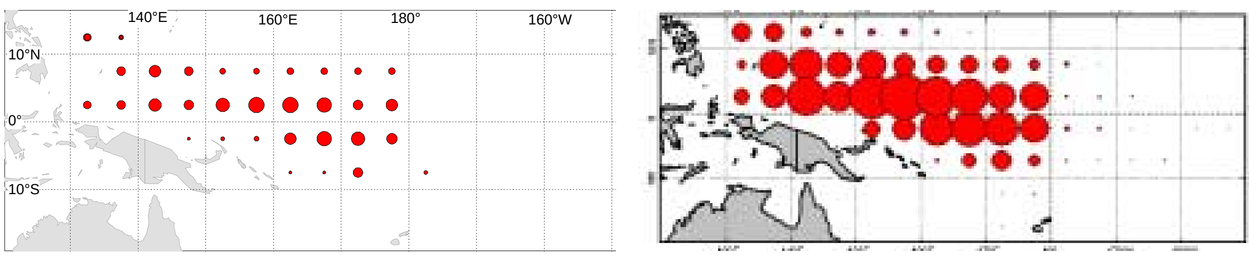


Figure 5. Annual proportions of bigeye tuna in yellowfin+bigeye by unassociated set type and observer sampling (A) and port-sampling of Japanese and US vessels (B). Histogram indicates 25th and 75th percentiles, circle indicates the median.

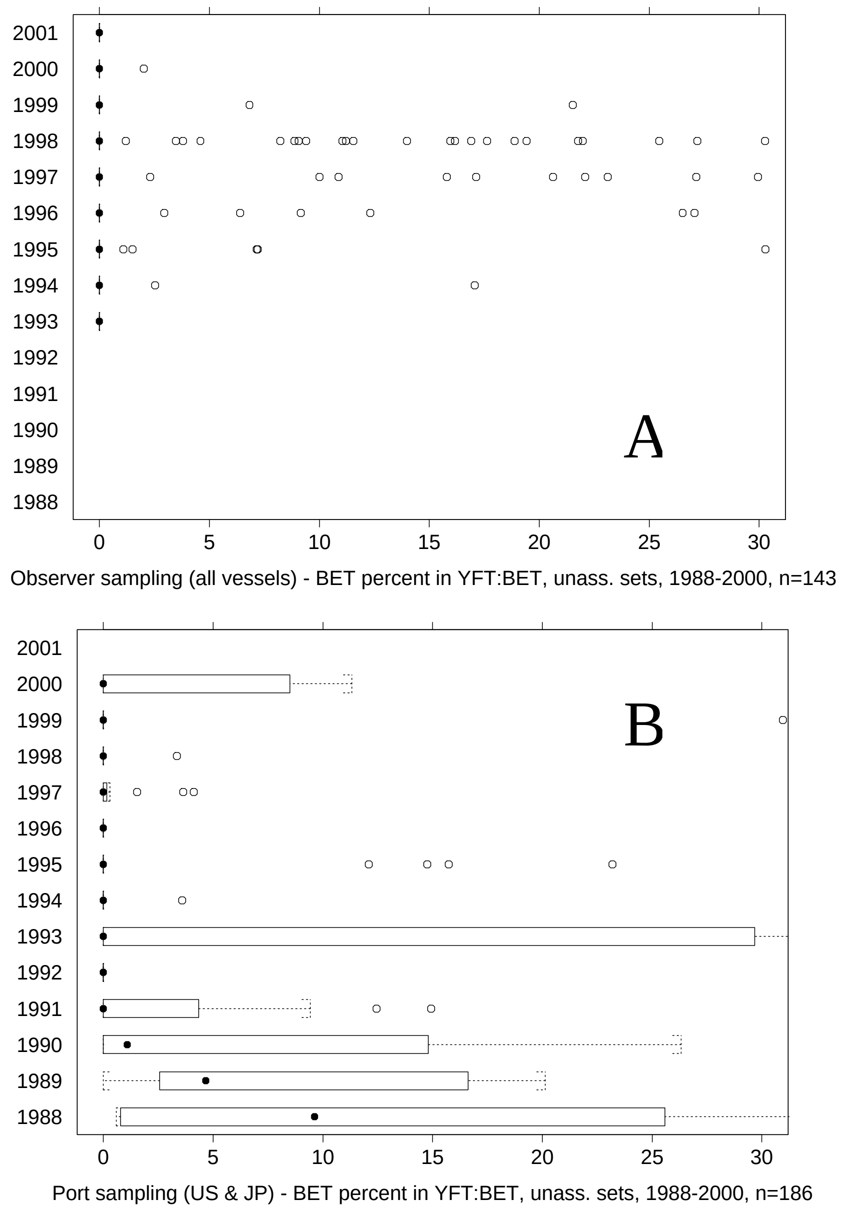


Figure 6. Annual proportions of bigeye tuna in yellowfin+bigeye by associated set type and sampling program (observer and port-sampling). Histogram indicates 25th and 75th percentiles, circle indicates the median.

A) Observer sampling (Japanese vessels), B) Japanese port-sampling, C) Observer sampling (US vessels) and D) US port-sampling

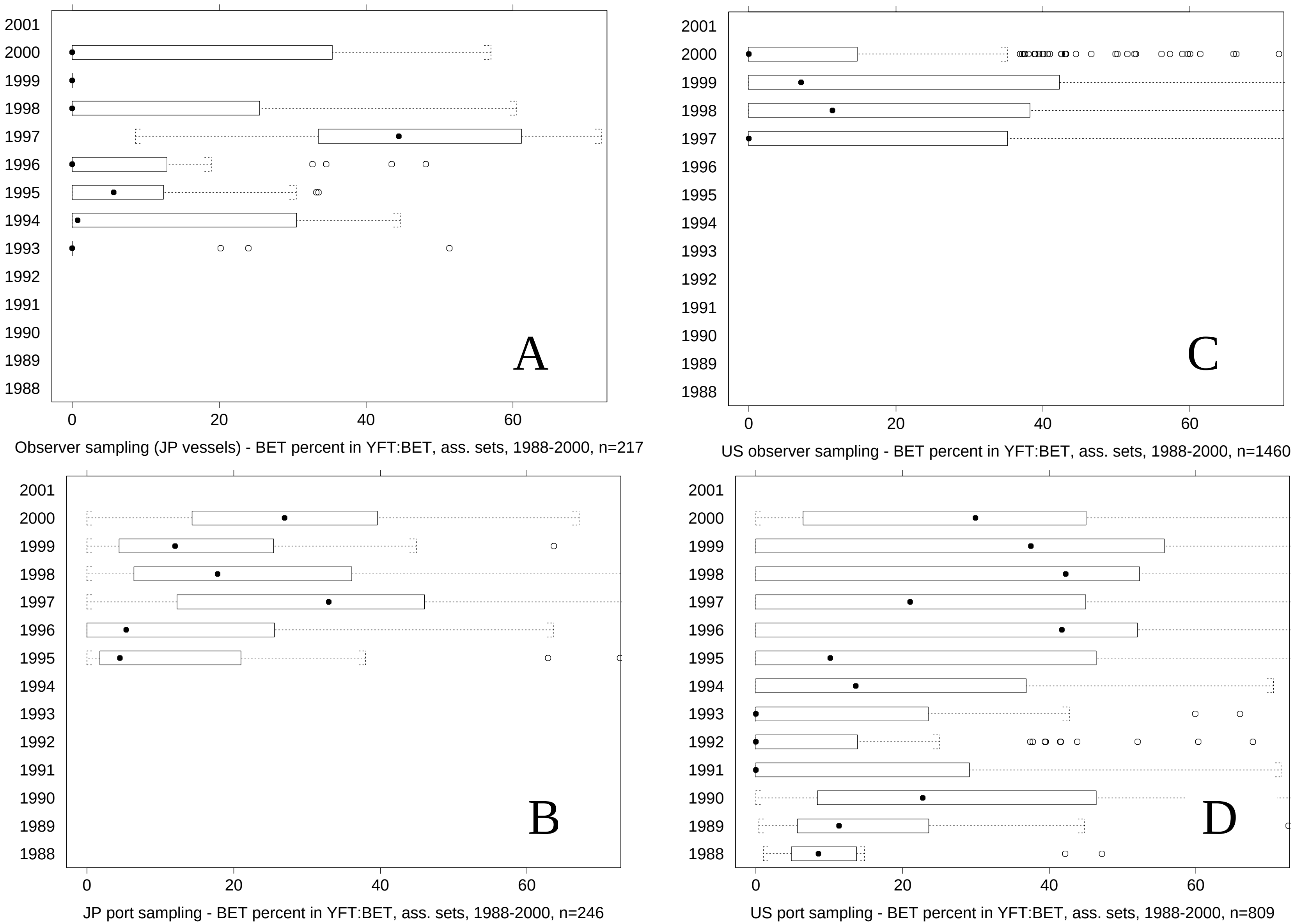
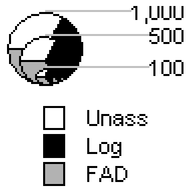


Figure 10. Spatial distribution of bigeye catch by set type (white – unassociated, black – log, grey – drifting FAD) for purse seine fleets in 1999. Bigeye catch was estimated by tree-based regression.



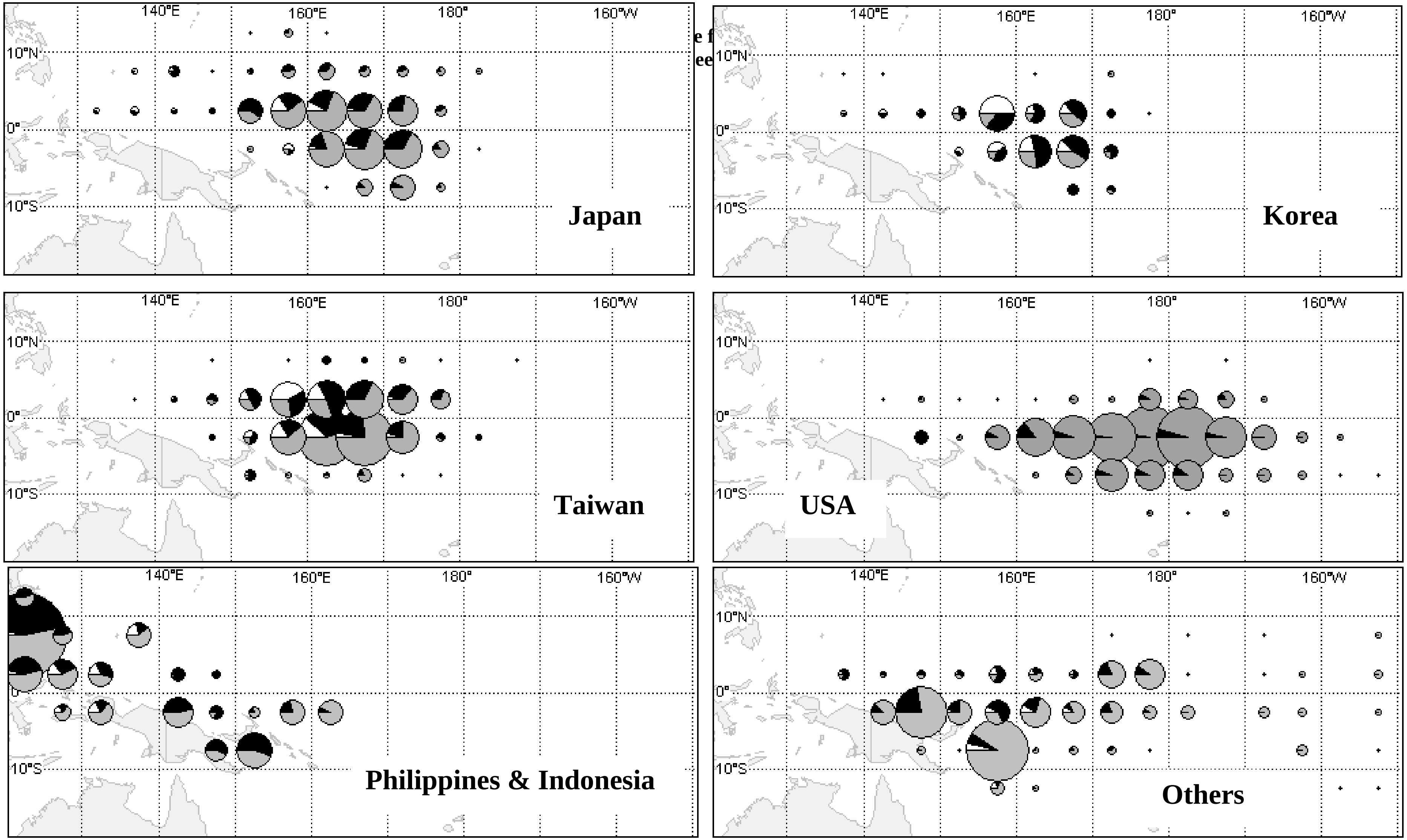


Figure 7 . Generalized additive model (GAM) derived effects of predictor variables, school type (A), year (B), month (C), fleet (D), latitude (E) and longitude (F) on the composition of bigeye in yellowfin+bigeye estimates. Relative density of data points is shown on the x-axis (rug-plot). All variables were categorical except latitude and longitude which were continuous.

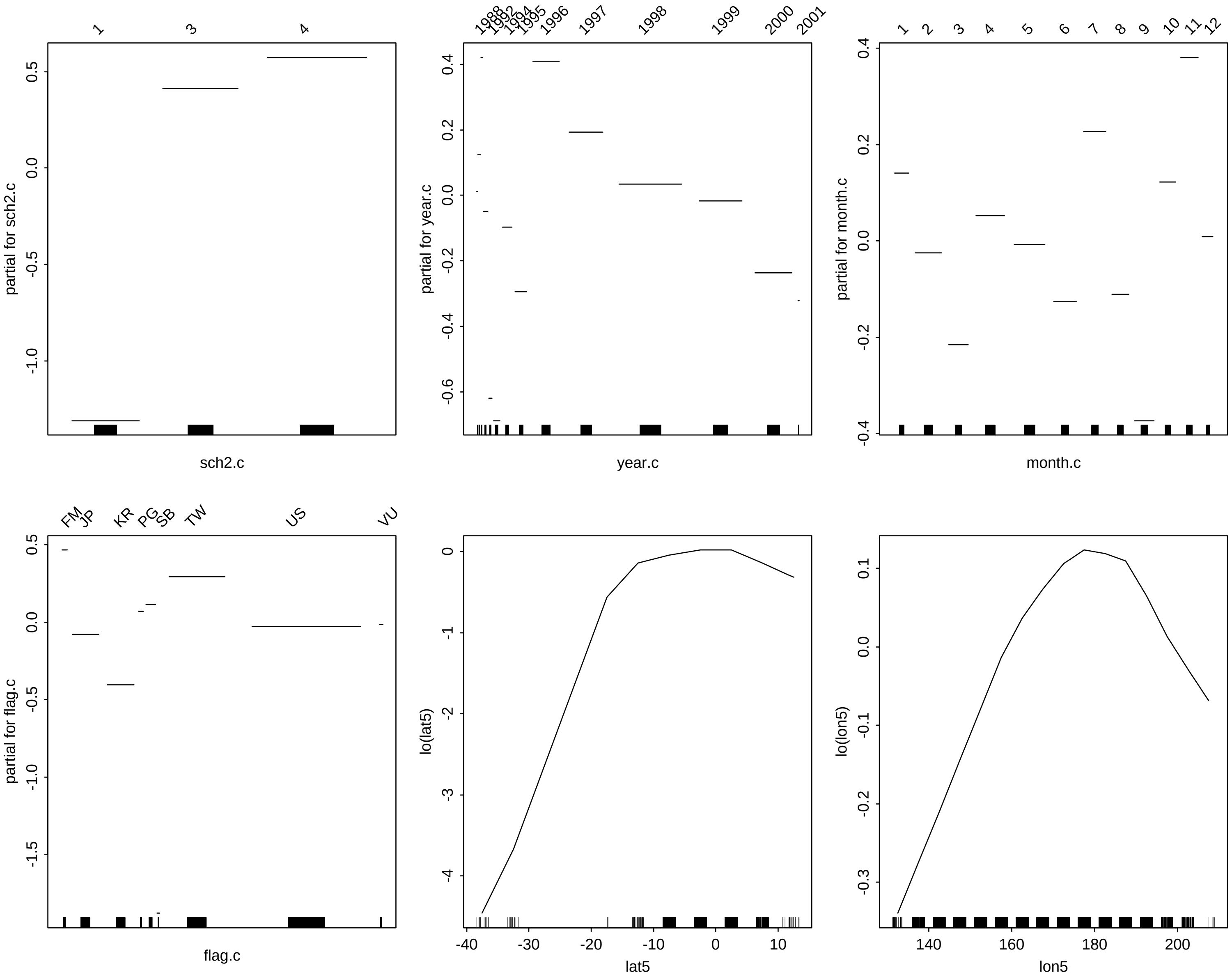


Figure 8. Pruned tree for the proportion of bigeye in the yellowfin+bigeye estimates for the purse seine fishery in the western and central Pacific Ocean. Initial 30 nodes illustrated.

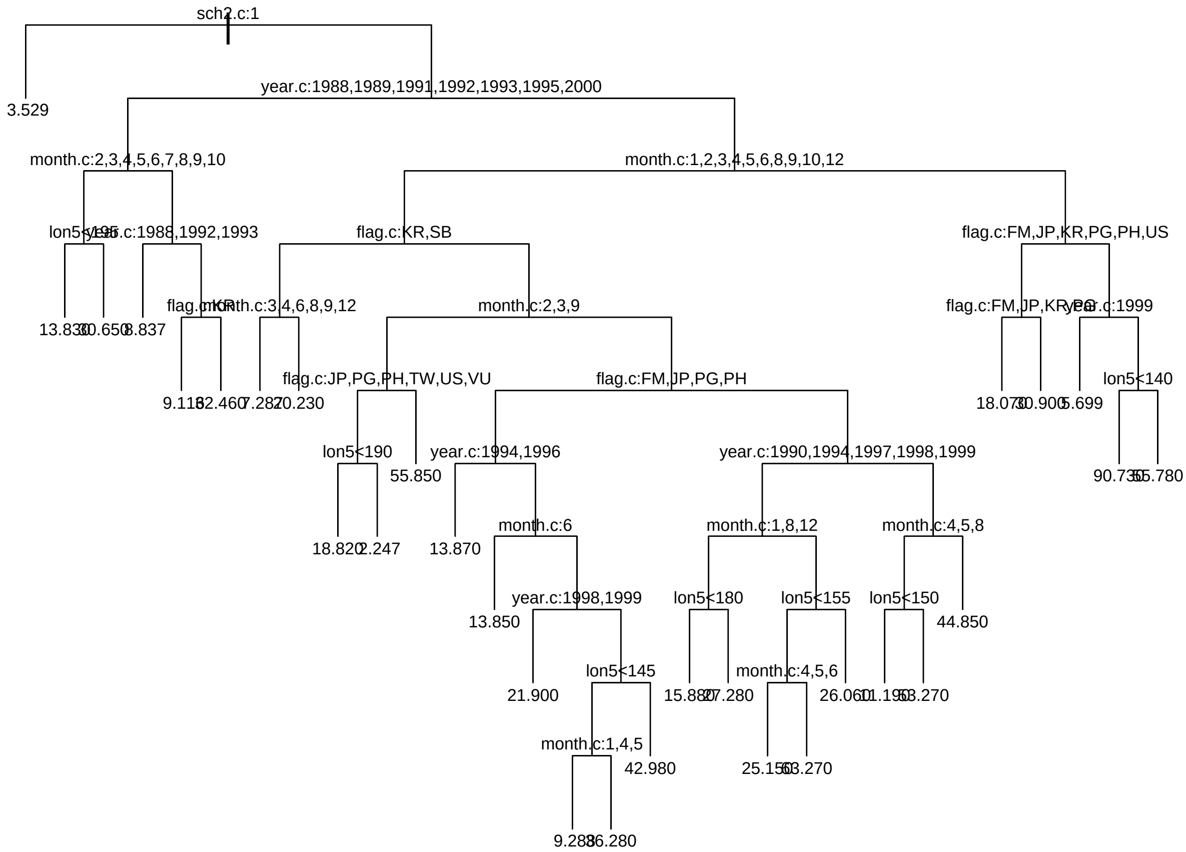


Figure 9. Annual bigeye catch by all purse seine fleets in the WCPO estimated from two methods (extrapolation of US port-sampling data and tree-based regression).