Standing Committee on Tuna and Billfish

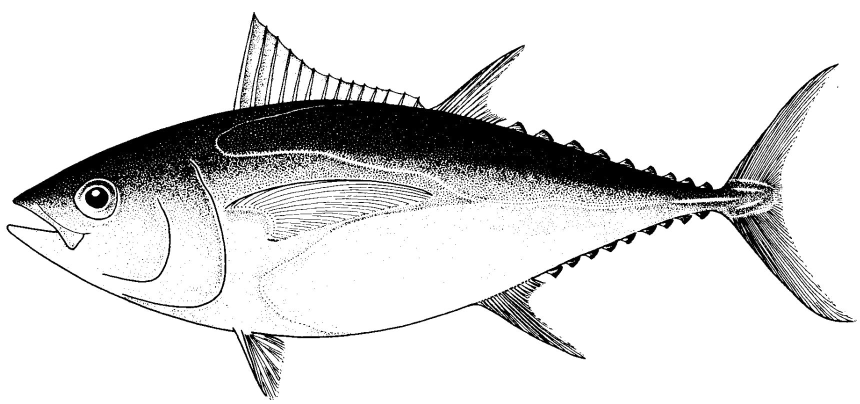
SCTB12 • 16-23 juin 1999 • TAHITI

Working Paper

SWG–3

Estimating bigeye composition in the purse seine fishery

Keith A. Bigelow



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

Estimating bigeye composition in the purse seine fishery

Keith Bigelow
Secretariat of the Pacific Community
Noumea, New Caledonia

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.

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. 1998, Lawson 1998). Modifications are largely dependent on a comprehensive NMFS port sampling programme that monitors purse seine landings in Pago Pago, American Samoa where most (82% in 1997) of the US catch is landed (Coan et al. 1998). 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 1998). 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. For example, the proportion of bigeye in the bigeye+yellowfin averages ~1% in unassociated sets and 13% in associated sets (Lawson 1999).

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. In 1998, the Yellowfin Research Group noted concern in the application of the estimating catches of non-US fleets (Anonymous 1998). Essentially, the procedure may be inappropriate 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, gear modifications).

Given the potential bias in bigeye catch estimation, the objectives of this paper are to:

1. Statistically compare bigeye composition between purse seine fleets
2. Consider an alternative statistical method (regression trees) to analyze factors affecting the composition of bigeye and estimate total catch

Species composition data in the purse seine fishery

Species composition data by weight was available from port and at-sea observer sampling (Table 1). A total of 1,543 samples could be related to other additional factors such as time (year and month), location (5° square), set type (associated or unassociated) and vessel flag.

- Port sampling data were available only for the US fleet sampled at Pago Pago. From 1988 to 1998, 475 samples were taken of which 80 and 395 were from unassociated and associated sets, respectively.
- Observer data were compiled from the SPC and Micronesian Maritime Authority (MMA) programs. Observer data obtained on US vessels under the USMLT were not used because species composition data taken with port sampling are more reliable. From 1993 to 1998, 1,063 purse seine sets were monitored of which 356 and 707 were from unassociated and associated sets, respectively. Monitoring occurred on vessels from 10 different fleets. Three fleets (Japan, Korea and Taiwan) had over 250 species composition samples; however, sampling was low for the remaining fleets.

Statistical comparison of bigeye composition in the purse seine fishery

Comparisons between fleets were only possible during the six-year period from 1993 to 1998. Since the spatial distribution of fishing activities has varied considerably over the six years, subsamples of the data were taken according to the following criteria:

1. **Set type –** Samples were stratified by set type (unassociated and associated)
2. **Time -** Samples had to occur within a seven month period. Seven months was chosen to use most of the available data.
3. **Space –** Samples had to be within an identical area of 10° of latitude and 40° of longitude
4. **Sample size –** For each set type, time and space strata, a fleet was only considered if at least five samples were taken

These criteria resulted in statistical comparisons for five unassociated set periods (Table 2) and seven associated set periods (Table 3). The analysis used 55% of the available unassociated samples and 72% of the associated samples.

Traditional parametric statistical methods (ANOVA) were not employed because the distribution of the data was not normal due to the number of zeros in the percentage of bigeye in the bigeye+yellowfin estimates. Instead, a Kruskal-Wallis *H* test, a non-parametric analog of one-way analysis of variance was used to detect differences in the distribution of the data.

For unassociated sets, comparisons were made for six fleets, but due to the paucity of data, only one comparison was conducted with the US fleet. For all five comparisons, there were no statistically significant differences ($p > 0.05$) in the proportion of bigeye in the bigeye+yellowfin estimates (Table 2). The lack of significant differences is because the proportion of bigeye is relatively low (~1%) in unassociated sets.

Nine fleets comprised comparisons for the associated sets (Table 3). The US fleet was included in each of the seven comparisons. In six of the seven comparisons, there were statistically significant differences ($p < 0.05$) in the proportion of bigeye in the bigeye+yellowfin estimates (Tables 3–4, Figure 1). There was only one period with non-significance ($p = 0.173$). The US fleet had higher bigeye proportions in most of the analyses (Figure 1). Bigeye composition in associated sets was characterized by high variability between sets as standard deviations were high (Table 4).

Regression trees as an alternative to analyze factors affecting the composition of bigeye and estimating total catch

Tree-based regression models are an alternative to linear and additive models. The 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 (year, month, latitude, longitude, school association and fleet) that could influence the proportion of bigeye in the purse seine catch. Data consisted of all observer and port sampling estimates of bigeye composition (N=1,543 observations).

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 to reduce overfitting. Pruning was accomplished through cross-validation procedures which determined the appropriate number of nodes (similar to tree branches). Cross-validation indicated a tree with approximately six branches (nodes) would be sufficient (Figure 2). The final tree had four predictors as latitude and fleet were excluded (Figure 1). The order of relative importance was 1) school association, 2) longitude, 3) year, 4) month. Values at the ends of a terminal node indicate the bigeye composition (percent) in the bigeye+yellowfin estimates. The residuals were not normally distributed in the initial tree-based regression (Figure 2) because bigeye proportion data contained a large amount of zero observations.

As an alternative to the initial regression, data were square-root transformed in an attempt to normalize the distribution. A tree-based regression based on transformed data included all the predictor variables except latitude (Figure 3). Cross-validation on the second tree suggested that more branches (~22) were necessary to estimate the proportion of bigeye (Figure 2). Residuals had a more normal distribution when the data were transformed.

Application of regression tree to catch data grouped by 5°x5°

The tree-based regressions were based on a subset of fishery data which are considered to be the most accurate (i.e. port and at-sea observer sampling). Predictions from the tree results were applied to the entire fishery dataset in order to estimate total bigeye catch by fleet. The SPC Regional Database was used in conjunction with the tree results to predict the bigeye proportion by fleet, year, month, 5° square and set type. The total catch was then estimated by multiplying the bigeye proportion by the reported bigeye+yellowfin estimates.

Total annual bigeye catch estimates for each fleet are compared from the three different estimation techniques: 1) original method of extrapolating the bigeye proportion in the US fleet to other fleets (Table 5), 2) tree-based predictions based on un-transformed data (Table 6) and 3) tree-based predictions based on square-root transformed data and back-transformed into bigeye proportions (Table 7). For both the US and all fleets, predicted annual bigeye catches were usually much lower for the tree-based techniques based on transformed data (Figures 4–5). Annual estimates for the western and central Pacific Ocean range from 7 to 37 thousand tonnes based on un-transformed data and 2 to 26 thousand tonnes based on transformed data (Tables 6–7). Presently, the different results due to no transformation or square root transformation cannot be explained. It should be noted that conflicting results from tree-based regressions have previously occurred due to data transformation (P. Kleiber pers. Comm.).

For the US fleet, predicted estimates based on un-transformed data are similar to port sampling estimates. However, in 1996 the estimates differ by >4,000 tonnes. The discrepancy results because other fleets in 1996 had lower proportions of bigeye in the bigeye+yellowfin than the US fleet. Thus, lower catch estimates are predicted for the US fleet in the regression due to the year effect.

For all fleets, annual tree-based estimates based on un-transformed data were higher (average – 122%, range 71–263%) than the currently used extrapolation method over the ten-year time-series. In contrast, annual tree-based estimates based on transformed data were lower (average – 67%, range 16–140%). Fleet-wide catch estimates based on un-transformed data show good coherence with the current extrapolation method in the years from 1993 to 1997 (average – 96%, range 73–134%), but less coherence from 1988 to 1992 (average – 149%, range 78–263%). Higher variability in the predicted values prior to 1993 probably result from the fact that the tree-based results only use US fleet data, as no non-US data is available.

Conclusions

1. The most accurate species composition data from port and observer sampling suggests that there are no significant differences in bigeye composition in unassociated sets between fleets. However, there are significant differences in bigeye composition in associated sets between fleets. Bigeye composition in associated sets was characterized by high variability between sets.
2. Tree-based regressions offer an alternative method to quantify what factors influence the proportion of bigeye in purse seine catches.
3. Catch estimates from the tree-based regressions differed depending on if the initial data were un-transformed or square-root transformed. Transformation appears to have a major effect in reducing catch estimates. Differences based on data transformation require further investigation.

4. Tree-based regressions on un-transformed data show good coherence with the current extrapolation method in the years from 1993 to 1997, a time frame when sampling occurred on all fleets.
5. Tree-based methods have the advantage over the current extrapolation method in statistically incorporating sources of variability into the estimation process other than school association. The major disadvantage is in obtaining representative samples from these highly mobile fleets.

References

Anonymous 1998. Report of the Eleventh Meeting of the Standing Committee on Tuna and Billfish, 28 May – 6 June 1998, Honolulu, Hawaii. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia. 108 pp.

Coan, A.L., Ito, R. & B. Kikkawa 1998. U.S. fisheries for tropical tunas and billfish of the central-western Pacfic and south Pacific albacore, 1993–1997. Working paper 43, Eleventh Meeting of the Standing Committee on Tuna and Billfish, 28 May – 6 June 1998, Honolulu, Hawaii 21 pp.

Lawson, T.A. 1998. SPC Tuna Fishery Yearbook, 1997. Secretariat of the Pacific Community, Noumea, New Caledonia. 132 pp.

Lawson, T. A. 1999. Estimates of annual catches of target species in the western and central Pacific Ocean. Working Paper SWG–2. Twelfth Meeting of the Standing Committee on Tuna and Billfish. Papeete, Tahiti, French Polynesia. 16th–23rd June 1999.

Table 1. Number of bigeye composition samples in the purse seine fishery taken from 1988 to 1998.

Fleet	Number of observer samples		Number of port samples	
	Unassociated	Associated	Unassociated	Associated
Japan	89	212		
Korea	142	122		
Taiwan	76	259		
USA	0	0	80	395
FSM	19	52		
Kiribati	0	16		
Philippines	2	16		
PNG	15	22		
Vanuatu	13	8		
FSM arrangement	5	0		
Total	361	707	80	395

Table 2. Statistical results of the proportion of bigeye tuna in unassociated sets between purse seine fleets.

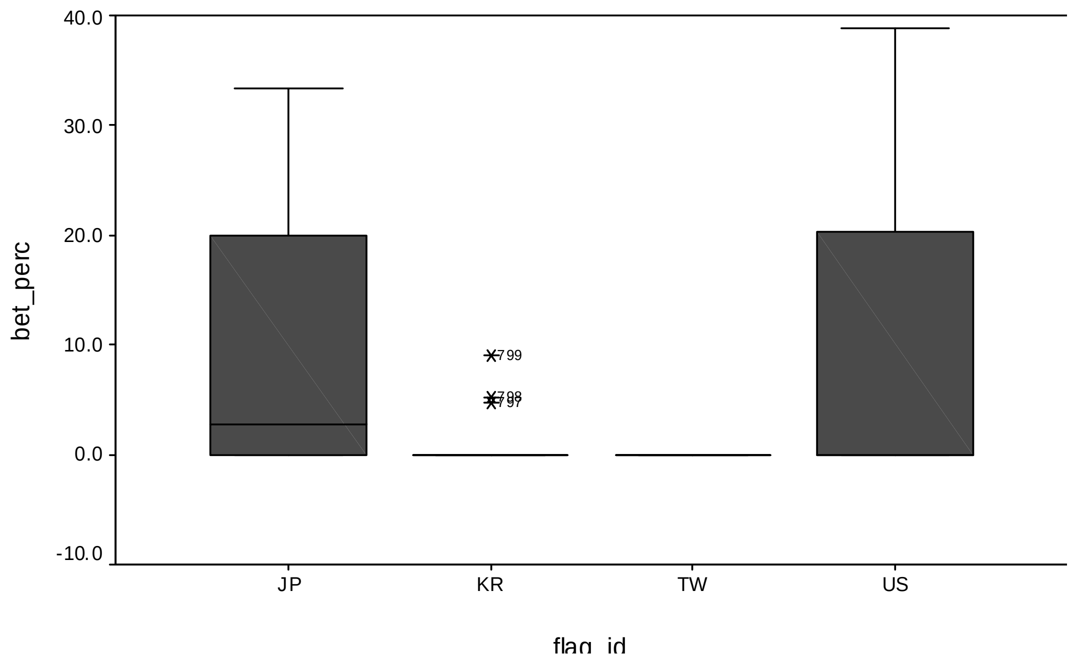
Time	Number of purse seine samples for statistical comparison							Significance
	Japan	Korea	Taiwan	USA	PNG	Vanuatu	Total	
5.93 - 11.93	12	14	9				35	0.181
8.94 - 2.95	18	16	12	11			57	0.150
12.95 - 5.96	16	27	19			6	68	1.000
2.97 - 7.97	10	41			4		55	0.843
2.98 - 7.98	14		13				27	0.563
Total	70	98	53	11	4	6	242	

Table 3. Statistical results of the proportion of bigeye tuna in associated sets between purse seine fleets.

Time	Number of purse seine samples for statistical comparison										Significance
	Japan	Korea	Taiwan	USA	FSM	Kiribati	Philippines	PNG	Vanuatu	Total	
6.93 - 12.93	25	14	5	5						49	0.032 *
1.94 - 7.94		34	26	16	23					99	<0.001*
8.94 - 2.95	30	6	23	19						78	0.173
3.95 - 9.95	50		49	5	15	16		12		147	0.006 *
1.96 - 7.96	23	28	15	38			15		7	126	<0.001 *
1.97 - 7.97	53	25	43	68				7		196	<0.001 *
10.97 - 4.98	14		66	25						105	0.039 *
Total	195	107	227	176	38	16	15	19	7	800	

Table 4/Figure 1. Statistical comparison between purse seine fleets of bigeye proportions in associated sets. STD – standard deviation.

Comparison #1, associated sets, 6.93 - 12.93				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
Japan	25	2.77	12.46	18.39
Korea	14	0.00	1.36	2.86
Taiwan	5	0.00	0.00	0.00
USA	5	0.00	11.82	17.48
Total	49			



Comparison #2, associated sets, 1.94 - 7.94				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
FSM	23	12.50	16.69	19.03
Korea	34	0.00	6.27	13.53
Taiwan	26	0.00	1.99	5.64
USA	16	8.76	23.38	29.26
Total	99			

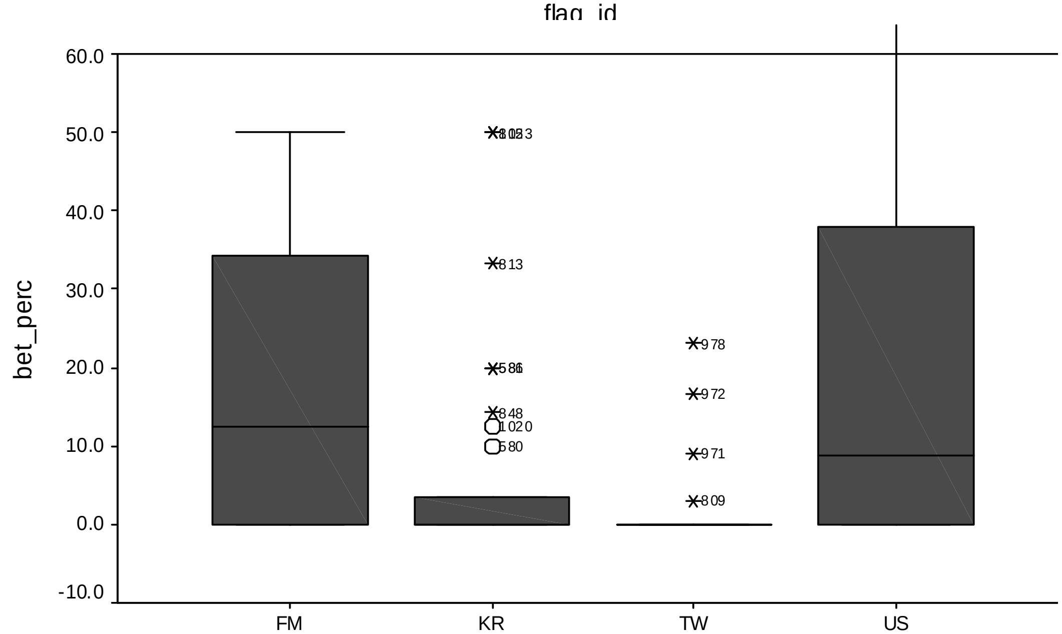
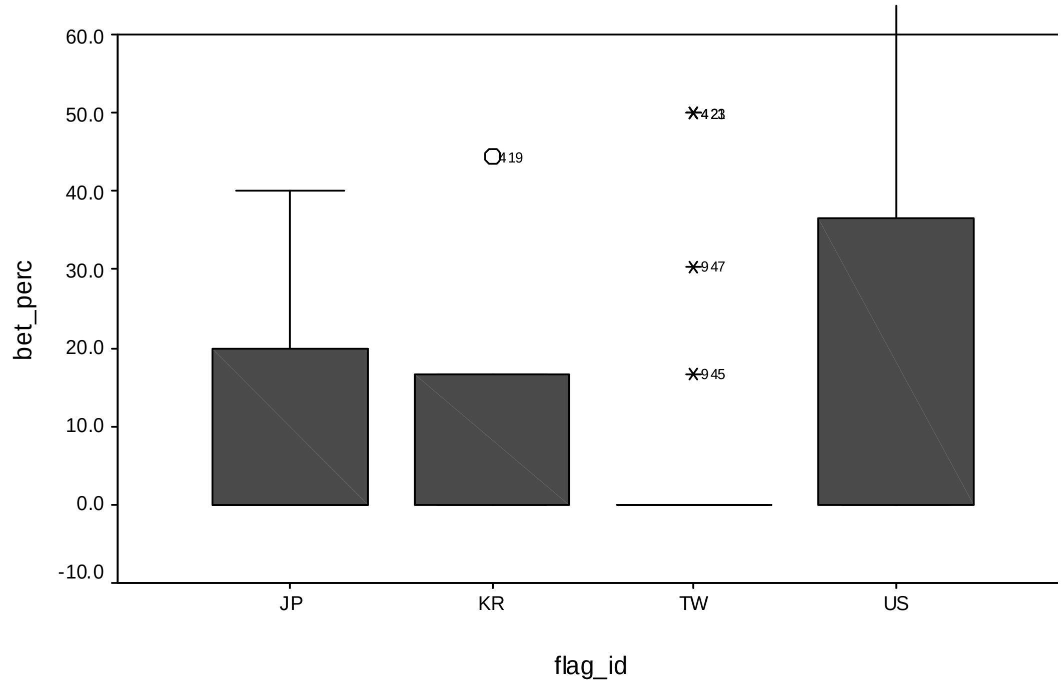


Table 4/Figure 1 con't. Statistical comparison between purse seine fleets of bigeye proportions in associated sets. STD – standard deviation.

Comparison #3, associated sets, 8.94 - 2.95				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
Japan	30	0.00	12.74	22.71
Korea	6	0.00	10.18	18.05
Taiwan	23	0.00	6.39	15.46
US	19	0.00	19.58	27.72
Total	78			



Comparison #4, associated sets, 3.95 - 9.95				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
FSM	15	0.00	8.00	15.81
Japan	50	0.11	10.00	14.07
Kirbati	16	0.00	1.56	6.25
PNG	12	22.61	28.43	27.92
Taiwan	49	0.00	9.44	16.31
USA	5	0.00	7.63	10.63
Total	147			

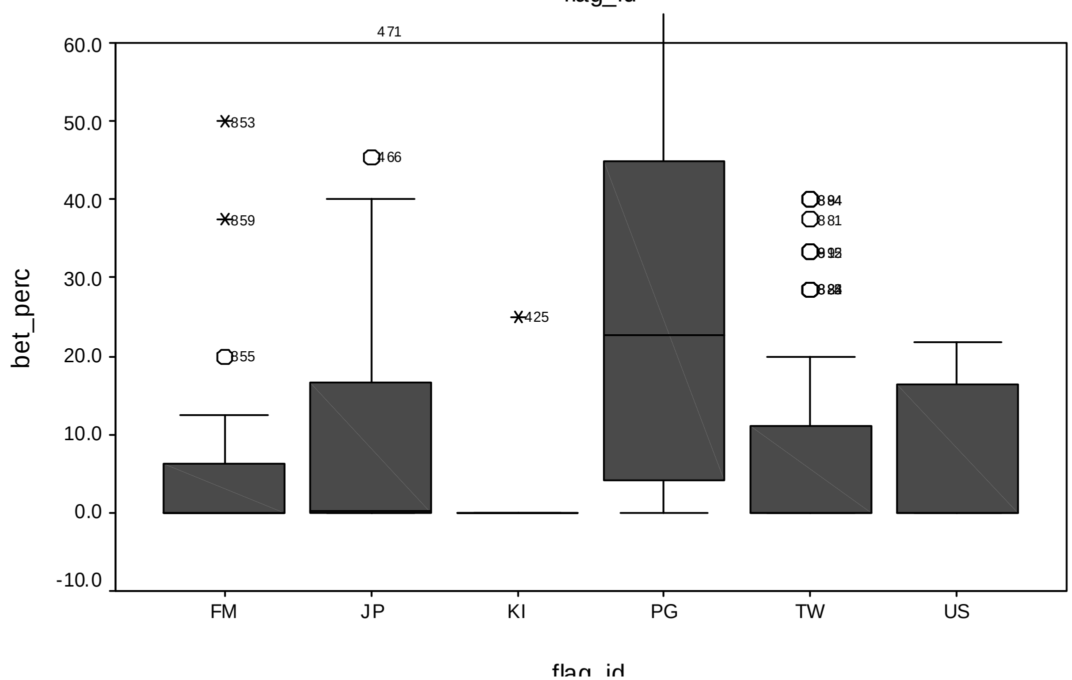
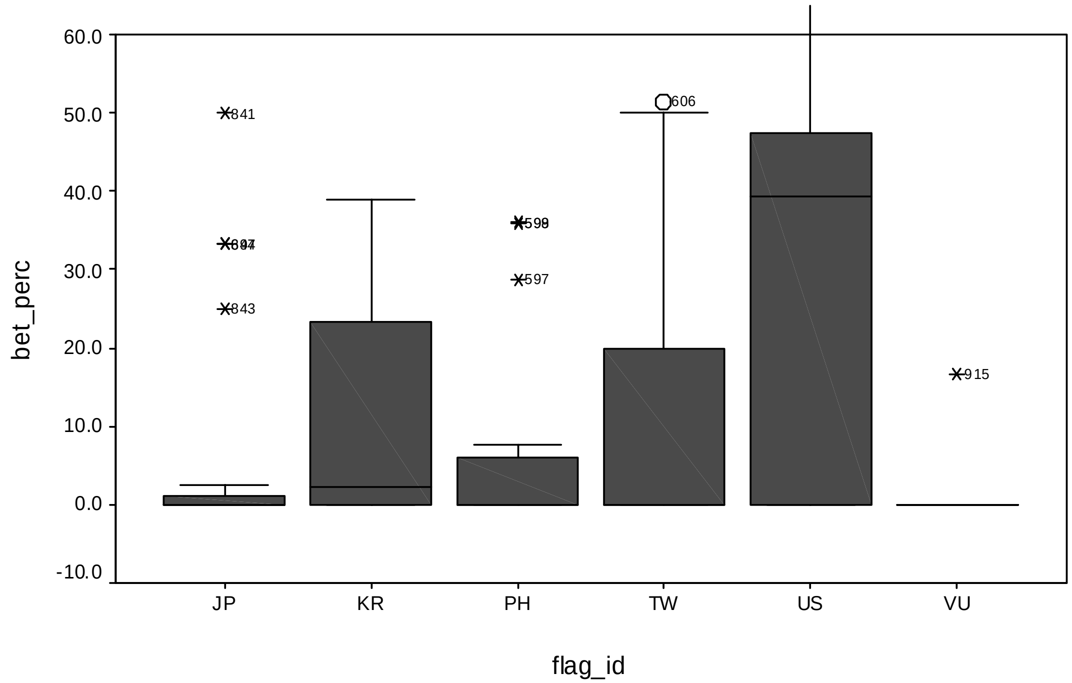


Table 4/Figure 1 con't. Statistical comparison between purse seine fleets of bigeye proportions in associated sets. STD – standard deviation.

Comparison #5, associated sets, 1.96 - 7.96				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
Japan	23	0.00	10.61	24.08
Korea	28	2.20	11.35	13.39
Philippines	15	0.00	7.53	13.75
Taiwan	15	0.00	15.11	24.21
USA	38	39.38	33.37	27.59
Vanuatu	7	0.00	2.38	6.29
Total	126			



Comparison #6, associated sets, 1.97 - 7.97				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
Japan	53	20.00	23.67	19.65
Korea	25	0.00	4.44	8.20
PNG	7	0.00	0.00	0.00
Taiwan	43	8.04	17.29	22.77
USA	68	20.33	26.38	28.06
Total	196			

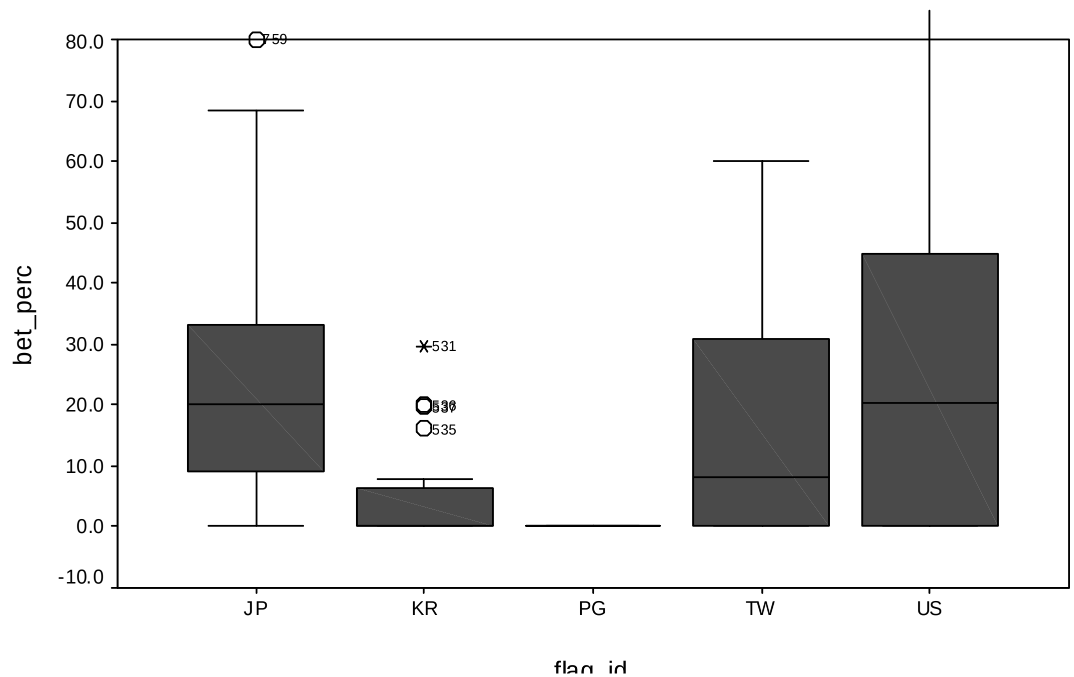


Table 4/Figure 1 con't. Statistical comparison between purse seine fleets of bigeye proportions in associated sets. STD – standard deviation.

Comparison #7, associated sets, 10.97 - 4.98				
Fleet	**Number of samples**	**Median**	**Mean**	**STD**
Japan	14	0.00	8.46	17.33
Taiwan	66	4.87	14.15	17.74
USA	25	0.00	17.51	30.68
Total	105			

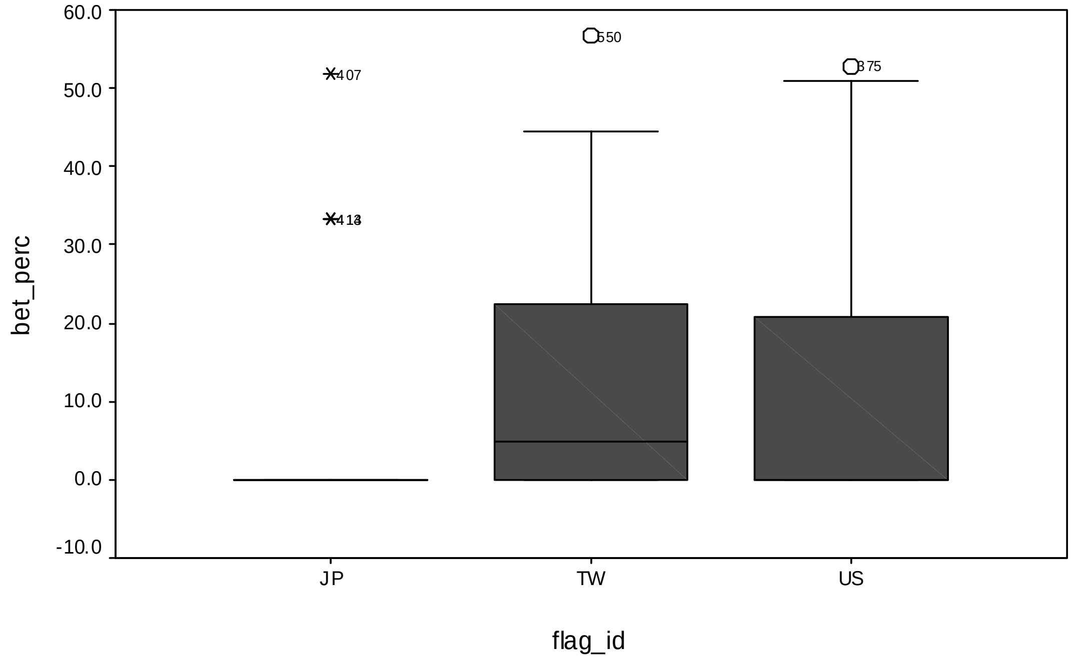


Table 5. Annual catches (metric tonnes) of bigeye by purse seine in the SPC statistical area. Estimates are based on extrapolating port sampling information except for Japan and the USA. Reproduced from Lawson (1998).

Year	Fleet					Total
	Japan	Korea	Taiwan	USA	Other	
1988	2,155	1,075	780	1,948	928	6,886
1989	4,025	2,060	2,268	2,421	2,325	13,099
1990	2,325	2,091	2,546	1,762	1,378	10,102
1991	2,499	2,604	3,175	1,550	1,672	11,500
1992	3,082	4,622	4,331	3,480	2,890	18,405
1993	3,169	2,586	2,733	3,731	2,379	14,598
1994	2,243	2,273	1,763	1,711	1,564	9,554
1995	2,787	2,313	1,387	3,190	3,074	12,751
1996	1,850	907	796	10,645	2,031	16,229
1997	6,155	3,042	3,020	9,499	4,103	25,819

Table 6. Annual catches (metric tonnes) of bigeye by purse seine in the SPC statistical area. Estimates are based on a tree-based regression of factors affecting the composition of bigeye in bigeye+yellowfin estimates. No transformation of bigeye composition.

Year	Fleet					Total
	Japan	Korea	Taiwan	USA	Other	
1988	2,492	1,411	978	1,474	927	7,282
1989	3,000	2,116	1,677	2,295	1,345	10,433
1990	6,685	6,662	6,700	3,811	3,584	27,442
1991	3,562	4,202	3,392	2,363	1,563	15,082
1992	6,187	8,622	8,718	3,300	5,556	32,383
1993	3,333	2,647	2,447	3,528	2,140	14,095
1994	2,712	2,298	2,156	1,947	1,356	10,469
1995	2,458	1,692	1,360	2,218	1,684	9,412
1996	1,965	1,249	644	6,152	1,941	11,951
1997	14,143	5,371	3,867	10,351	3,957	37,689

Table 7. Annual catches (metric tonnes) of bigeye by purse seine in the SPC statistical area. Estimates are based on a tree-based regression of factors affecting the composition of bigeye in bigeye+yellowfin estimates. Square-root transformation of bigeye composition.

Year	Fleet					Total
	Japan	Korea	Taiwan	USA	Other	
1988	2,908	1,844	1,138	1,755	1,531	9,176
1989	3,584	3,103	1,879	2,944	2,153	13,663
1990	3,524	3,498	2,712	3,188	1,696	14,618
1991	1,534	2,182	1,183	1,369	479	6,747
1992	818	1,226	1,004	722	563	4,333
1993	836	644	546	855	549	3,430
1994	716	1,771	490	285	318	3,580
1995	606	409	277	435	448	2,175
1996	1,056	861	212	4,051	1,908	8,088
1997	10,299	3,175	2,613	7,068	3,054	26,209

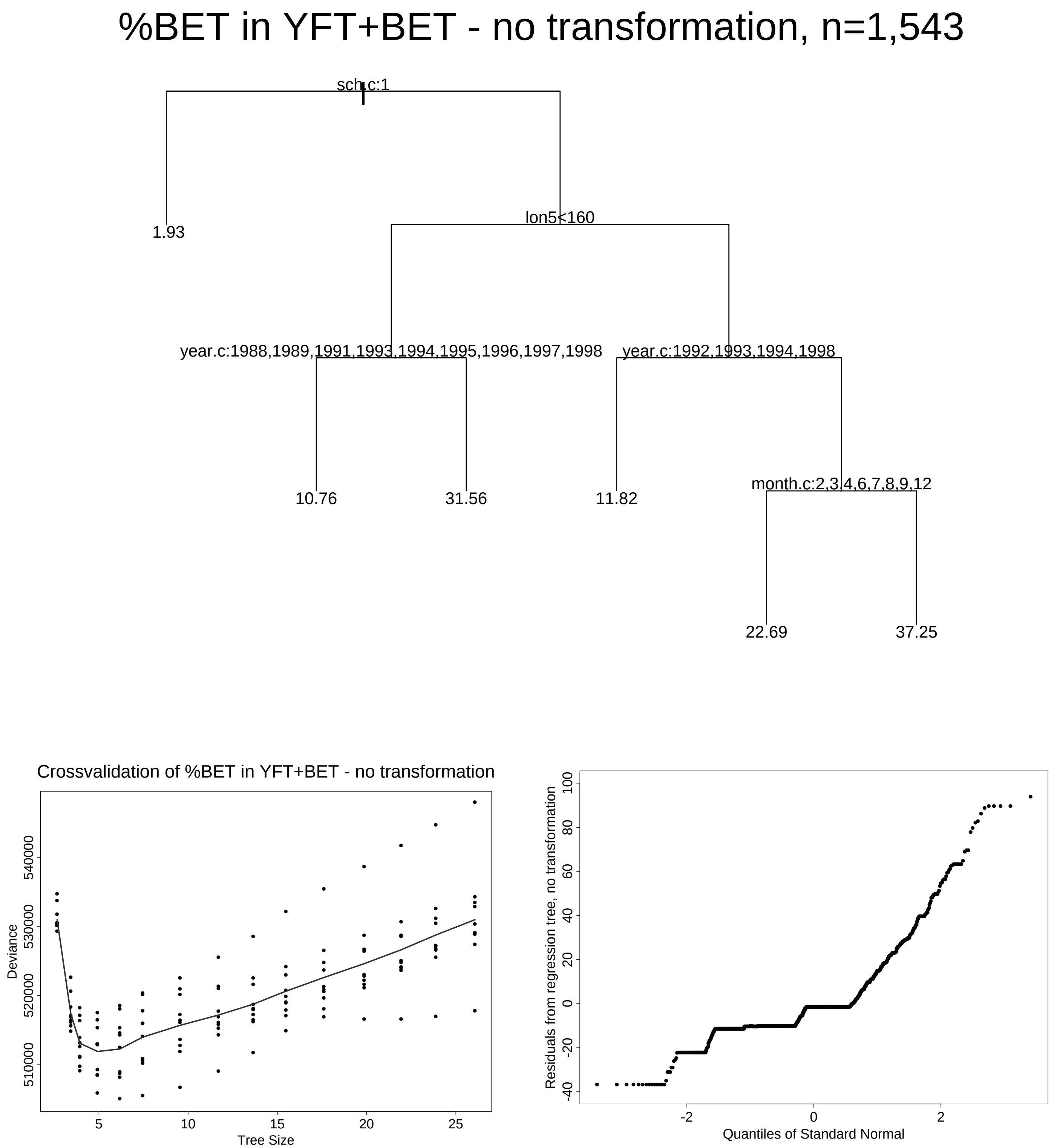


Figure 2. Top – Pruned tree for the proportion of bigeye in the bigeye+yellowfin estimates for the purse seine fishery in the western and central Pacific; un-transformed data. Left – Cross-validation results on the number of appropriate branches in the regression tree. Right – Quantile-quantile residual plots.

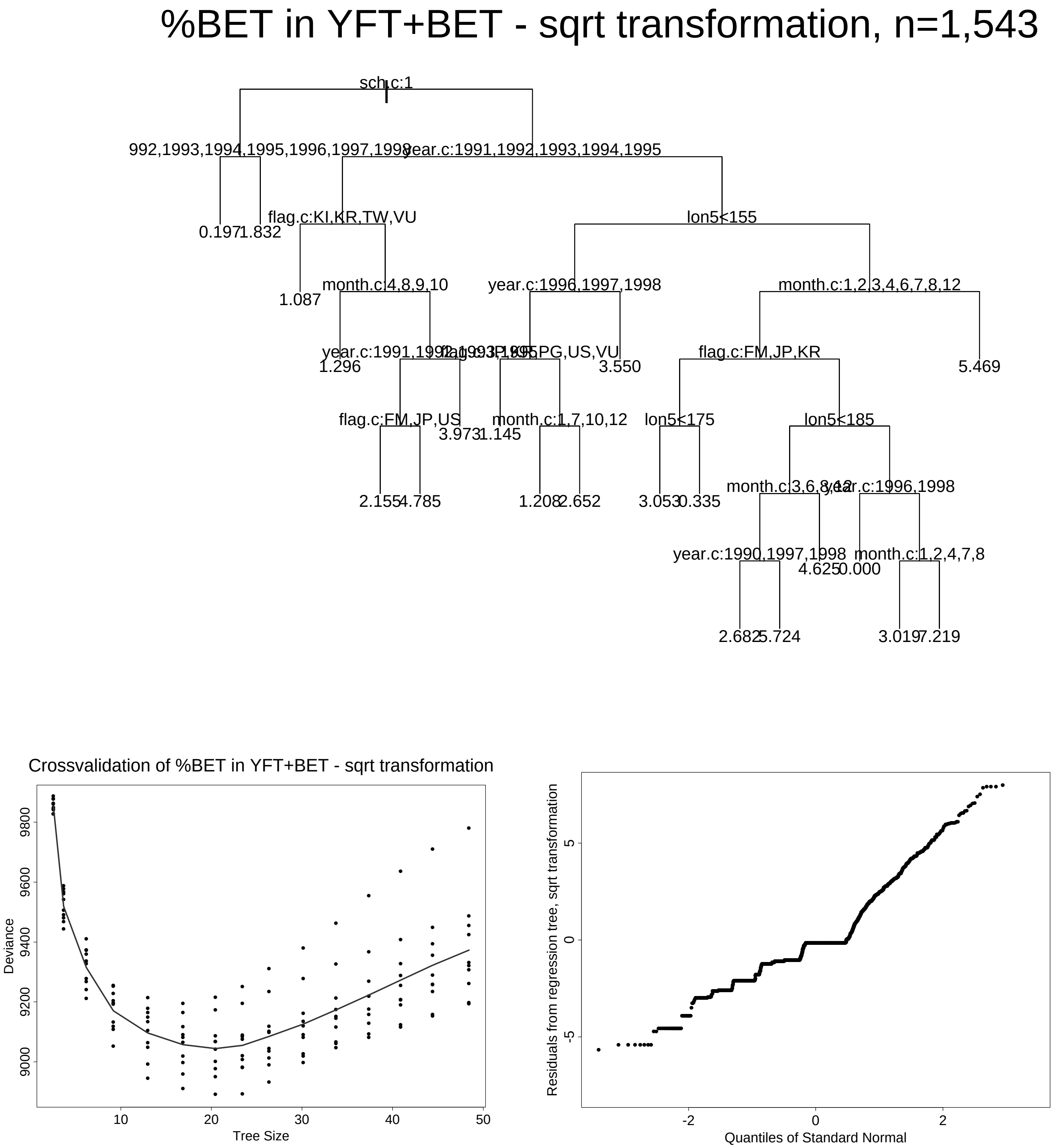


Figure 3. Top – Pruned tree for the proportion of bigeye in the bigeye+yellowfin estimates for the purse seine fishery in the western and central Pacific; square-root transformed data. Left – Cross-validation results on the number of appropriate branches in the regression tree. Right – Quantile-quantile residual plots.

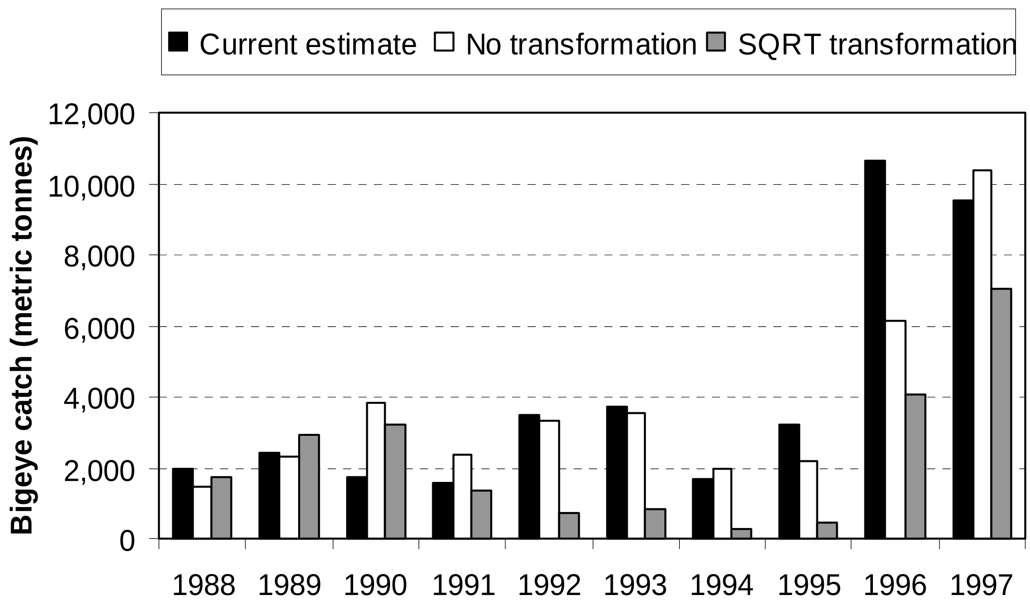


Figure 4. Annual bigeye catch by the US fleet in the western and central Pacific purse seine fishery estimated from three methods.

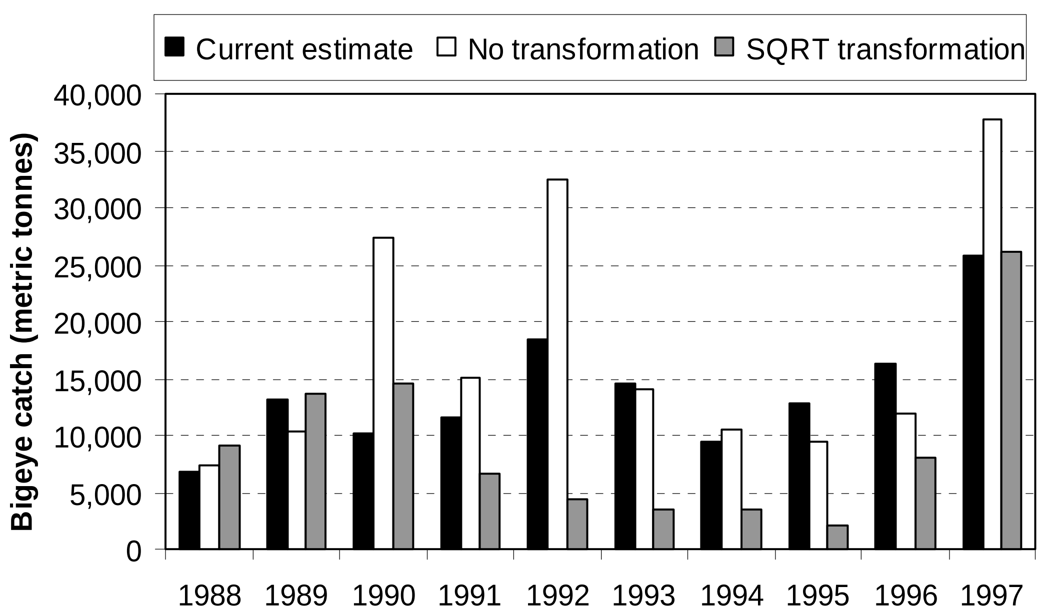


Figure 5. Annual bigeye catch by all fleets in the western and central Pacific purse seine fishery estimated from three methods.