Availability: 11 October 2014
IOTC-2014-WPEB10-INF09

SCIENTIFIC COMMITTEE
SEVENTH REGULAR SESSION

9-17 August 2011
Pohnpei, Federated States of Micronesia

Estimation of Catch Rates and Catches of Key Shark Species in Tuna Fisheries of the Western and Central Pacific Ocean Using Observer Data

WCPFC–SC7–2011 / EB–IP–02

Timothy Lawson [1]

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

ESTIMATION OF CATCH RATES AND CATCHES OF KEY SHARK SPECIES IN TUNA FISHERIES OF THE WESTERN AND CENTRAL PACIFIC OCEAN USING OBSERVER DATA

Timothy Lawson

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

July 2011

ABSTRACT

Catch rates and catches of blue shark, mako sharks, oceanic whitetip shark, silky shark and thresher sharks in longline fisheries and oceanic whitetip shark and silky shark in purse-seine fisheries of the Western and Central Pacific Ocean were estimated using observer data. Catch rates were predicted with Delta-Lognormal models fitted to longline observer data collected during 1991–2011 and purse-seine observer data collected during 1994–2011. The covariates latitude and longitude were parameterised as a two-dimensional spline and heat maps were used to depict the effect of latitude and longitude on predicted catch rates. Parametric bootstraps were used to determine confidence intervals for the estimates of catch rates and catches. Generalised Estimating Equations were used to examine correlation and dispersion in longline catch rates. Trends in the estimates of annual catch rates and catches are discussed in Clarke (2011) along with other indicators of the status of shark populations.

INTRODUCTION

Under the *Agreement for the Provision of Scientific Services to the Commission*, the SPC Oceanic Fisheries Programme (OFP) has been contracted by WCPFC to conduct statistical analyses to estimate catches of non-target species; the primary data that the OFP uses to estimate catches of non-target species are collected by observers onboard the fishing vessels. This paper presents recent developments in the methods for estimating longline and purse-seine catches of non-target species from observer data and their application to five of the WCPFC's key shark species and genera: blue shark (*Prionace glauca*), silky shark (*Carcharhinus falciformis*), oceanic whitetip shark (*Carcharhinus longimanus*), mako sharks (*Isurus* spp.) and thresher sharks (*Alopias* spp.). [1] These

---

[1] At its Seventh Regular Session in December 2010, the WCPFC adopted CMM 2010–07, Conservation and Management Measure for Sharks, in which the key shark species and genera are identified as blue shark, silky shark, oceanic whitetip shark, mako sharks, thresher sharks, porbeagle shark (*Lamna nasus*), and the following hammerhead

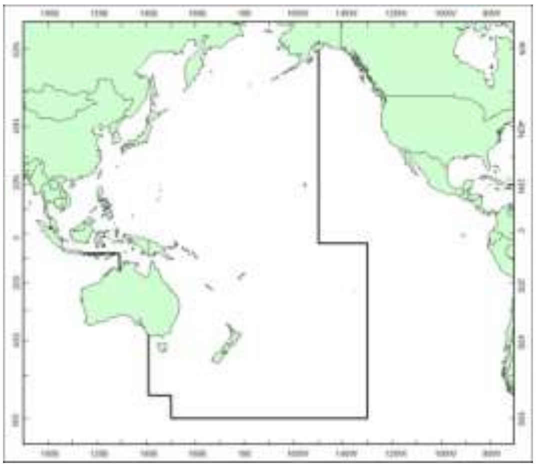
five species and genera are also the focus of the WCPFC Shark Research Plan (Clarke & Harley 2010).

Coverage of Longline Observer Data Held by the OFP

Table 1 presents the coverage of longline effort in the WCPFC Statistical Area (Figure 1) by observer data held by the OFP. Coverage from 1992 to 2009 has been 0.87%. Coverage of the distant-water longline fleets (other than the Japanese fleet fishing in the waters of Australia and New Zealand) by data held by the OFP is less than 0.1%; coverage of the Japanese fleet fishing in the waters of Australia and New Zealand has been 5.1% and 35.5% respectively. Coverage of the Hawaiian longline fleet has been 6.5%, while coverage of the New Zealand domestic fleet has been 3.6%. Coverage of the offshore longline fleets targeting yellowfin and bigeye, and albacore, have been 0.8% and 1.0% respectively. Coverage has thus been highly variable, ranging from negligible to moderate.

In addition to the negligible coverage for the distant-water fleets, the lack of consistent coverage through time for the Japanese fleet fishing in the Australian Fishing Zone (AFZ), due to the termination of fishing in 1998, and the Hawaiian fleet, due to the lack of data provided to SPC since 2004, has been problematic. Observer data covering the Hawaiian fleet from 2010 onwards may soon be provided to the WCPFC.

Figure 1. WCPFC Statistical Area



---

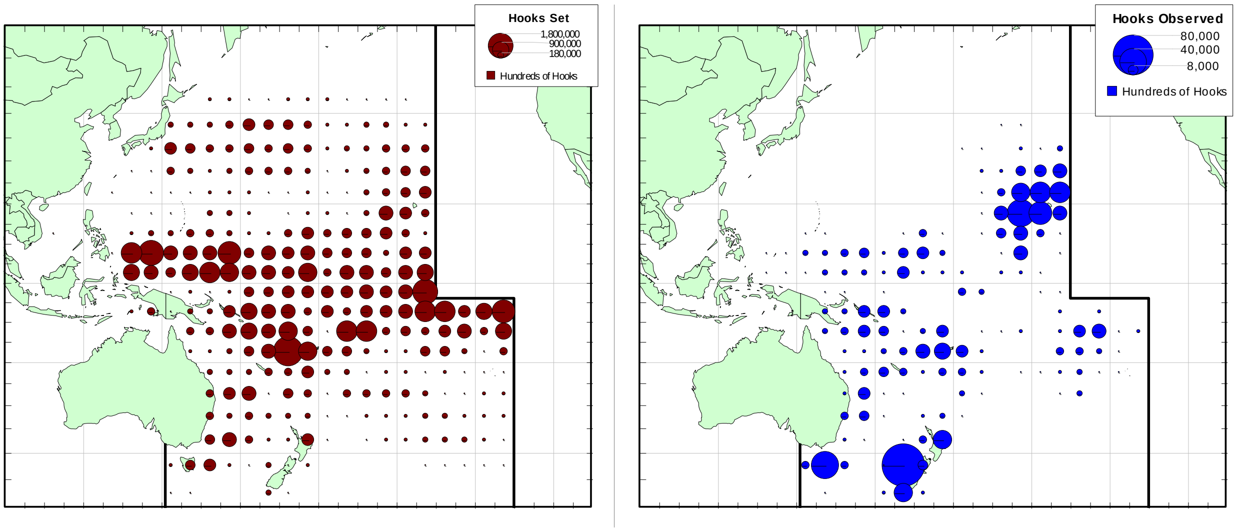
sharks: winghead (*Eusphyra blochii*), scalloped (*Sphyrna corona*), great (*Sphyrna mokarran*) and smooth (*Sphyrna zygaena*).

Table 1. Coverage of longline fishing effort by observer data held by the SPC Oceanic Fisheries Programme, by sector

Year	Australia: Japanese Fleet	Distant-Water Albacore	Distant-Water Yellowfin & Bigeye	Hawaii	New Zealand: Domestic Fleet	New Zealand: Japanese Fleet	Offshore Albacore	Offshore Tropical	Total
1992	17.124	0.000	0.000	0.000	0.530	6.225	0.000	0.083	0.574
1993	16.013	0.000	0.000	0.000	0.000	31.440	0.000	0.276	0.872
1994	10.149	0.000	0.000	4.330	0.555	46.101	0.000	0.309	0.681
1995	6.434	0.000	0.028	4.140	2.611	88.792	0.685	0.256	0.593
1996	8.793	0.264	0.000	5.043	4.846	0.000	1.126	0.269	0.644
1997	5.491	0.000	0.000	3.531	5.258	81.322	0.597	0.971	0.867
1998	0.732	0.165	0.061	3.991	3.534	46.710	0.392	0.675	0.658
1999	0.000	0.070	0.000	3.166	0.412	84.144	0.416	0.466	0.516
2000	0.000	0.000	0.018	8.695	0.206	76.290	0.166	0.660	0.664
2001	0.000	0.000	0.000	15.152	3.106	65.801	0.084	0.107	0.866
2002	0.000	0.000	0.185	23.897	1.441	100.000	0.529	1.371	1.630
2003	0.000	0.000	0.027	21.505	6.343	47.162	0.826	1.209	1.671
2004	0.000	0.000	0.000	16.522	13.133	0.000	1.067	1.049	1.361
2005	0.000	0.000	0.261	0.000	2.768	51.348	1.512	1.081	0.650
2006	0.000	0.000	0.296	0.000	2.258	100.000	1.943	1.287	0.872
2007	0.000	0.000	0.170	0.000	4.226	63.908	1.584	1.031	0.751
2008	0.000	0.000	0.000	0.000	4.073	16.017	1.348	0.849	0.597
2009	0.000	0.000	0.000	0.000	0.000	0.000	1.165	0.331	0.443
Total	5.083	0.027	0.061	6.456	3.569	35.510	1.034	0.763	0.868

The geographic coverage of the longline observer data is summarised in Figure 2 and Appendix Figure A1. The coverage is dominated primarily by the Hawaiian fleet, but also the Japanese fleet fishing in the AFZ and the Japanese and New Zealand fleets fishing in the waters of New Zealand. Large areas in the WCPFC Statistical Area — to the west of 130°E, the northwest and the southeast — have not been covered by observer data, which complicates the estimation of catch rates and catches of sharks and other non-target species.

Figure 2. Distribution of longline hooks set and hooks observed in the WCPFC Statistical Area, excluding Indonesia and the Philippines, 1992–2009



Coverage of Purse-Seine Observer Data Held by the OFP

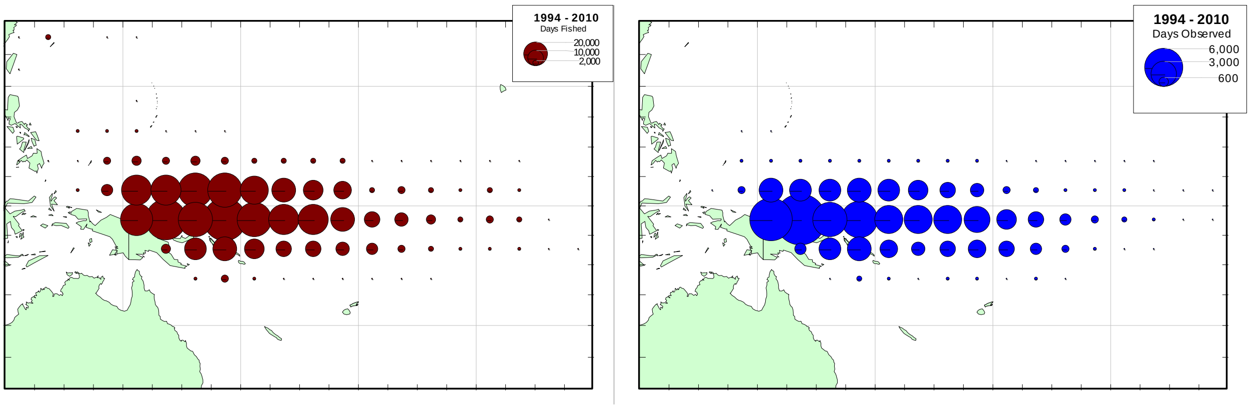
Table 2 presents the coverage of purse-seine effort in the WCPFC Statistical Area by observer data held by the OFP, excluding the domestic fleets of Indonesia and the Philippines. Coverage from 1995 to 2010 has been 11.05%. The coverage of sets on unassociated and associated school has been similar, 10.03% and 12.01% respectively.

The geographic coverage of the purse-seine observer data is summarised in Figure 3 and Appendix Figure A2. The coverage is dominated by the United States fleet from 1994 to 2001. In 2002, coverage of the Papua New Guinea fleet increased considerably and has remained high. In 2003, coverage under the FSM Arrangement increased. In 2010, observer coverage in the region increased to 100% as required by the WCPFC Conservation and Management Measure 2008–01; however, coverage of the data held by the OFP and available for analysis was only 21.6% due to lags in the provision and processing of the observer data. While coverage has not been representative in terms of the flag states prior to 2010, it has generally been much more representative in terms of the geographic distribution of fishing effort than for the longline fishery, since most purse-seine fishing takes place in a much smaller area than the area fished by longliners.

Table 2. Coverage of purse-seine fishing effort by observer data held by the SPC Oceanic Fisheries Programme

Year	Unassociated Schools	Associated Schools	Total
1995	4.16	2.83	3.56
1996	6.21	5.16	5.64
1997	5.65	6.45	6.13
1998	6.54	7.92	7.22
1999	2.46	4.27	3.62
2000	2.44	7.30	4.95
2001	5.07	7.40	6.13
2002	7.51	13.03	10.25
2003	7.92	13.72	10.83
2004	11.10	16.01	14.34
2005	12.03	18.86	15.51
2006	12.04	18.21	15.65
2007	10.79	15.37	13.07
2008	13.19	14.09	13.64
2009	17.71	11.60	14.51
2010	20.23	24.39	21.61
Total	10.03	12.01	11.05

Figure 3. Distribution of purse-seine days fished and days observed in the WCPFC Statistical Area, 1994–2010, excluding the domestic fleets of Indonesia and the Philippines



METHOD

The primary objective of this study was to estimate the catches of key shark species in the longline and purse-seine tuna fisheries in the WCPFC Statistical Area using observer data. The catches were estimated as the product of known longline and purse-seine effort in the region and catch rates predicted from models fit to observer data. Longline effort data held by the OFP are stratified by 5° latitude by 5° longitude (5x5), month and *hooks between floats* (HBF, a proxy for depth). Purse-seine effort data are stratified by 1° latitude by 1° longitude (1x1), month and school association.

A secondary objective was to standardise catch per unit of effort (CPUE); however, there are important differences between that objective and the estimation of catches. When estimating catches, nominal catch rates must be predicted by the model for all strata covered by the effort data, including those strata that are not covered by the observer data. For example, while there are broad geographic areas in the region that have not been covered by longline observer data, effort data exist for those areas and therefore the nominal catch rates must be predicted by the model so that the catches can be estimated.

Also, the covariates (independent variables) used in the model must be available both in the observer data, so that the model parameters can be fit, and in the effort data, so that the nominal catch rates can be predicted. Thus, certain covariates that are available in the observer data, but which are not available in the effort data — such as the type of bait used by a longliner or the use of a helicopter by a purse seiner — cannot be included in the model.

When standardising CPUE, however, there is no need to predict nominal catch rates; hence, there is no concern about strata not being covered by the observer data. [2] And all covariates available in the observer data can be used to fit the model, including those that may not be available in the effort data.

This distinction has implications for the manner in which the covariates are parameterised in the model. When estimating catches, the covariates must be parameterised such that once the model has been fit to the observer data, it can then be used to predict nominal catch rates in strata not covered by the observer data. The use of splines in this regard is discussed below.

Another important difference is that when the objective is to estimate catches, the focus is on predicting *nominal* catch rates using the values of the covariates for each stratum of the effort data.

[2] That is, no concern apart from the general question of whether the observer data are representative of the fishery.

In contrast, when standardising CPUE, the values of all covariates except *year* are set to *fixed* values and then the model is used to predict CPUE for all values of year. The values to which the other covariates are fixed are arbitrary, although, if a covariate is continuous, it is good practice to use values in the middle of the covariate's range.

Though not the primary objective of this study, catch rates standardised for *year* will still be examined.

Structure of the Delta-Lognormal Model of CPUE

The catches of non-target species are often zero and under these circumstances, Delta-Lognormal (DLN) models of catch rates are appropriate. [3] As mentioned above, when estimating catches, the covariates must be parameterised such that once the model has been fit to the observer data, it can then be used to predict nominal catch rates in strata not covered by the observer data. For longline, nominal catch rates must be predicted for strata of time period, geographic area and HBF. For purse-seine, they must be predicted for strata of time period, geographic area and school association. Stratifying the observer data at a relatively high level of resolution of time and area, such as 5x5 by month or 10x10 by quarter, and modelling the time-area strata as categorical variables, is not appropriate, since there will be strata in which fishing took place that are not covered by the observer data and for which catch rates cannot be predicted. Stratifying the observer data at a low resolution of geographic area — such as five or six broad areas that together cover the entire WCPFC Statistical Area, which are typically used for tuna stock assessments — would allow the prediction of catch rates for areas not covered at a higher resolution by the observer data, but at the cost of loss of precision in the predicted catch rates.

Rather than parameterising the covariates as categorical variables, a solution to this problem is to model the relationship between the response (dependent variable) and a covariate as a *piecewise polynomial*, also termed a *spline*. In a piecewise polynomial, the range of values of a covariate is separated into regions and the effect of each value of the covariate within a region is modelled as a polynomial (Chambers & Hastie 1992). The advantage of splines is that this highly *nonlinear* relationship can be transformed into a *linear* relationship between the response and the values of the *basis functions* determined for each value of the covariates. See the appendix for notes on the use of basis functions.

[3] Other model structures — such as Zero-Inflated Lognormal, Tweedy or Quasi-Poisson models — are also appropriate; however, the choice of model structure is probably less important to this analysis than the use of splines, and the other model structures were not examined in this study.

In the splines used below, the degree of the polynomials, *d*, in each region into which the range of the covariate has been divided is fixed at three — thus *cubic splines* — and the regions are usually quantiles. When using R functions to determine the basis values, such as **bs** for B-splines or **ns** for natural splines, the number of quantiles is usually specified by the number of *knots*, *k*, such that there are $k + 1$ quantiles or regions; that is, $k = 0$ results in one quantile of 100%, $k = 1$ in two quantiles of 50%, $k = 2$ in three quantiles of 33.3%, etc. The number of knots, and thus quantiles, is usually determined by choosing the number that minimises a model selection criterion, such as the Bayesian Information Criterion (BIC, Schwarz 1978).

The structure of the DLN model of CPUE is given by

$$CPUE = F_1(...) \cdot F_2(...) \tag{2}$$

where F_1 is the probability that the catch rate in a stratum is positive and F_2 is the catch rate in a stratum if the catch rate is positive, and $(\ldots)$ stands for whatever covariates that F_1 and F_2 might depend on (and which may differ); F_1 is usually referred to as the *logistic* part, and F_2 to the *lognormal* part, of the DLN model.

For the logistic part, the observed values of the response are assigned to be 1 if the catch rate in the stratum is positive and 0 otherwise; they are thus binomial random variables. Predictions of F_1 are only meaningful if they lie within [0,1]; however, this restriction is not incorporated in a simple linear regression of the observed values of the dependent variable on the basis functions of a spline. To ensure that the predicted values of F_1 lie within [0,1], a generalised linear model (GLM) with a *logit* link function — the *logistic regression model* — is used to estimate the regression coefficients. The logit link is given by:

$$\eta = \ln\left(\frac{\mu}{1-\mu}\right) \tag{3}$$

where μ is the mean of a binomial random variable, which, in this case, is the probability that the catch rate in a stratum is positive. We express F_1 in terms of the logit link as follows:

$$F_1(\ldots) = \frac{e^{\eta(\ldots)}}{1+e^{\eta(\ldots)}}. \tag{4}$$

Expressing $\eta(...)$ as a function of the basis functions of splines and regression coefficients, we have

$$\eta(\ldots) = \alpha_0 + \sum_{i=1}^{V} \sum_{j=1}^{DF_i} \alpha_{ij} B_{ij}(x_i) \tag{5}$$

where V is the number of covariates in the model; DF_i is the number of degrees of freedom of the spline for the i^th covariate; $B_{ij}(x_i)$ are the values of the basis functions for the i^th covariate, x_i, and α_0 and α_{ij} are the regression coefficients.

For the lognormal part, the regression coefficients are estimated with a simple linear regression of the natural logarithm of CPUE on the basis functions. Expressing F_2 in terms of log CPUE, we have

$$F_2(\ldots) = e^{\ln CPUE(\ldots) + \frac{S^2}{2}} . \tag{6}$$

where S^2 is the residual variance of the linear regression. The expected value of the exponent of a normal random variable of mean zero and variance σ^2 is $e^{-\frac{\sigma^2}{2}}$; to remove this bias, we include $e^{+\frac{S^2}{2}}$ in equation (6).

Finally, expressing *ln[CPUE]* as a function of the basis functions of splines and regression coefficients, we have

$$\ln CPUE(\ldots) = \alpha'_0 + \sum_{i=1}^{V'} \sum_{j=1}^{DF'_i} \alpha'_{ij} B'_{ij}(x'_i) \tag{7}$$

where *ln[CPUE]* is the natural log of CPUE in strata for which CPUE is positive, and the primes in the right-hand side of equation (7) indicate that the number of covariates, degrees of freedom, values of basis functions and regression coefficients are for the lognormal part of the DLN model.

Estimation of Catches in the WCPFC Statistical Area and Confidence Intervals

Longline catches were estimated using effort data covering the WCPFC Statistical Area, east of 130°E, stratified by year, month, 5x5 and two categories of HBF: shallow (< 10 HBF) and deep (≥ 10 HBF). Purse-seine catches were estimated using effort data covering the area from 20°N to 20°S and 130°E to 150°W, stratified by year, month, areas of 2° of latitude and 5° of longlitude, and school association (unassociated and associated). For each stratum of effort data, nominal CPUE was predicted with the DLN model fitted to the observer data. Catches for each stratum were estimated as the product of the predicted CPUE and the known effort. Estimates of annual catches were determined by summing the estimated catches over strata and grouping by year.

Confidence intervals for estimates of catches were constructed through the use of multinomial normal distributions of the regression coefficients; this technique is sometimes called a *parametric bootstrap*. Multinomial distributions for each of the logistic and lognormal models were parameterised with mean vectors equal to the point estimates of the regression coefficients, and covariance matrices determined from the correlation matrices and the standard error vectors. The multinomial distributions were used to generate 1000 sets of the regression coefficients for each of the logistic and lognormal models. When predicting CPUE for each stratum of effort, the 1000 sets of regression coefficients were used to generate 1000 estimates of CPUE for each stratum, which in turn were used to generate 1000 estimates of the catch for each stratum by multiplying by the known effort for the stratum. Summing the catches over strata and grouping by year for each set resulted in 1000 estimates of the catch for each year. The median of the 1000 estimates was used as the point estimate of the catch rate and annual catch. Confidence intervals for estimates of catch rates and annual catches were taken to be the 2.5% and 97.5% quantiles of the 1000 estimates.

APPLICATION TO LONGLINE

Definition of Replicates and Responses

For longline, the response in the logistic part of the DLN model of CPUE is 1 or 0 depending on whether the catch rate during a trip was positive or zero, while the reponse for the lognormal part is the the natural logarithm of the average catch rate in units of number of sharks per hundred hooks during a trip. Trips were used as the replicate, rather than longline sets, since sets tend not to be independent of one another; sets made during a fishing trip tend to catch similar species at similar rates because they usually occur within similar strata of time period, geographic area and depth, and therefore do not provide much additional information to the average catch rate for the trip.

Longline trips may also lack independence, but to a lesser degree than sets. The lack of independence among trips in the various longline sectors listed in Table 2 below were briefly examined with General Estimating Equations (GEE, Liang & Zeger 1986), which allow for correlation among the observations within sectors.

At the time of the analysis, there were a total of 3,405 longline trips from 1991 to 2011 in the observer database. Only trips during which at least five sets were made, and at least 2000 hooks were set, were used in the analyses; 286 trips were not used for this reason. There were 61 trips for which shark or swordfish was suspected of being the target species on the basis of the catch composition; these trips were not used since the shark catch rates may not be representative of the vast majority of longline effort. Table 3 presents the number of trips by year and sector; data covering a total of 3,058 trips were used. The Hawaiian fleet represents 42.1% of the total number of trips, followed by the offshore sectors targeting yellowfin and bigeye in tropical waters, 20.2%, and albacore in sub-tropical and temperate waters, 19.6%. No observer data are available for the domestic fisheries of Indonesia, the Philippines and Chinese Taipei; observer data covering the domestic longline fleet of Australia have not yet been imported into the OFP observer database.

Table 3. Number of trips taken by observers on longliners in the WCPO and used in the analysis

Year	Australia: Japanese Fleet	Distant-Water Albacore	Distant-Water Yellowfin & Bigeye	Hawaii	New Zealand: Domestic Fleet	New Zealand: Japanese Fleet	Offshore Albacore	Offshore Yellowfin & Bigeye	Total
1991	56	0	0	0	0	3	0	0	59
1992	54	0	0	0	2	6	0	1	63
1993	74	0	0	0	0	17	0	8	99
1994	54	0	0	45	1	7	0	18	125
1995	32	0	1	42	3	8	7	22	115
1996	28	1	0	50	5	0	11	15	110
1997	25	0	0	33	6	8	6	37	115
1998	2	2	1	44	9	5	5	31	99
1999	0	1	0	35	2	6	11	24	79
2000	0	0	1	98	3	4	5	31	142
2001	0	0	0	202	18	4	4	7	235
2002	0	0	2	273	9	4	27	73	388
2003	0	0	1	259	5	4	42	55	366
2004	0	0	0	205	14	0	50	59	328
2005	0	0	4	0	9	2	52	40	107
2006	0	0	3	0	10	3	62	76	154
2007	0	0	1	0	14	3	47	68	133
2008	0	0	0	0	15	2	72	31	120
2009	0	0	0	0	0	0	104	20	124
2010	0	0	0	0	0	0	88	3	91
2011	0	0	0	0	0	0	6	0	6
Total	325	4	14	1,286	125	86	599	619	3,058
%	10.63%	0.13%	0.46%	42.05%	4.09%	2.81%	19.59%	20.24%	100.00%

The replicates were screened to eliminate outliers, with the definition of outliers determined by inspection. Table 4 presents the level of CPUE that defines the outliers, the number of trips for which CPUE was greater than or equal to the level, the percentage of trips that were screened for outliers and the number of trips remaining and used in the analyses.

Table 4. Level of longline CPUE (sharks per hundred hooks) defining outliers and the number of outliers out of a total of 3,058 trips

Species or Group	Level of CPUE	Number of Outlier Trips	% of Total	Trips Remaining
Blue Shark	2.00	44	1.44%	3,014
Mako Sharks	0.20	18	0.59%	3,040
Oceanic Whitetip Shark	0.25	13	0.43%	3,045
Silky Shark	0.40	28	0.92%	3,030
Thresher Sharks	0.20	28	0.92%	3,030

Parameterisation of the Covariates

For each trip, the covariates were assigned as follows. The *latitude* and *longitude* assigned to the trip were the average latitude and longitude of the locations of the sets, weighted by the number of hooks per set. Most trips were of less than one month in duration; the month during which the largest number of days on which a set was made was assigned as the *year* and *month* for the trip. The number of *hooks between floats* for the trip was calculated as the average number of hooks between floats per set, weighted by the number of hooks per set. The catch rate for the trip was calculated as the total number of sharks caught divided by the total number of hooks set and expressed as the number of sharks per 100 hooks.

Year and month

In the exploratory phase of the analysis, *year* was initially parameterised as a spline and various attempts were made to parameterise *month* in order to incorporate seasonality into the model. If the analysis was confined to either the northern hemisphere or the southern hemisphere, parameterising *month* as a spline would suffice to capture any seasonality in catch rates; however, when both the northern and southern hemispheres are included in the analysis, *month* will not have the same effect and other approaches must be considered.

In the first attempt, month was parameterised as two variables: (i) a spline of *month* nested within the northern hemisphere and (ii) a spline of *month* nested within the southern hemisphere. The degrees of freedom of the splines was set to three. The values of the basis functions for each of the two variables are the same, except that the values for *month* nested in the northern hemisphere are set to zero for trips in the southern hemisphere and the values for *month* nested in the southern hemisphere are set to zero for trips in the northern hemisphere. However, the results showed that *month* parameterised in this way was confounded with latitude, such that the effect of latitude on predicted CPUE was jagged and not smooth at the equator.

A second attempt was made by parameterising month as one variable that had positive values in the northern hemisphere and negative values in the southern hemisphere, e.g., January–December ranged from 1 to 12 in the northern hemisphere and from –1 to –12 in the southern hemisphere. The values of the basis functions were determined with a knot at zero, such that a separate cubic spline would be fit in each hemisphere. Again, however, the effect of latitude on predicted CPUE was jagged and not smooth at the equator.

In the third attempt, the effect of *month* was forced to be symmetrical in the northern and southern hemispheres by parameterising *month* in the southern hemisphere as *(month + 5) modula 12 + 1*. Thus January in the northern hemisphere is 1, while January in the southern hemisphere is 7; February in the north is 2, while February in the south is 8, etc. Again, the effect of latitude was jagged.

The conclusion, perhaps to have been expected, is that the structure of the model is such that seasonality is confounded with latitude and cannot be separated. Therefore, rather than including *month* as a separate variable, it was decided to include *year* and *month* in a single variable as *year + (month - 0.5) / 12*, and parameterise the combined *year_month* variable as a spline. Thus January 1992 is 1992.004, February 1992 is 1992.125, etc. While no longer modelling seasonality, this parameterisation allows *month* to be used to more precisely model time trends in CPUE.

Latitude and longitude

In the exploratory phase, *latitude* and *longitude* were first parameterised as univariate splines; that is, *latitude* was parameterised as a spline and *longitude* was parameterised as a separate spline. However, the results of this parameterisation were unreasonable and suggested that latitude and longitude were confounded in the data for the fleets with the greatest coverage. The data covering the Japanese fleet in the Australian Fishing Zone, which range from 145°E to 160°E, are primarily south of 25°S. The data covering the fleets in New Zealand, which range further to the east, from

165°E to 180°, are primarily south of 30°S. The data covering the Hawaiian fleet, which range still further to the east, from 180° to 150°W, are primarily north of 15°N. While not a major fleet in terms of coverage, but the only fleet operating in its longitudinal band, the data covering the French Polynesian fleet, which range from 150°W to 140°W, are primarily from 10°S to 20°S. Thus the observer data are concentrated at certain latitudes consistently across the region, from west to east.

To eliminate this problem, *longitude* was constrained to be a linear variable, rather than a spline, given that *latitude* is more important in terms of explaining variation in shark CPUE. The results were reasonable and the effect of *latitude* on predicted CPUE correctly reflected our knowledge about the distribution of the key shark species, with the latitude effect for tropical sharks (oceanic whitetip shark and silky shark) being high in the tropics and lower at higher latitudes, and the converse true for the other sharks (blue shark, mako sharks and thresher sharks).

However, this paramerisation does not account for interactions between *latitude* and *longitude*, which are known to be important. Therefore, *latitude* and *longitude* were finally parameterised as a *multivariate* spline; that is, *latitude* and *longitude* were considered as a single variable, *lat_lon*, having two dimensions, i.e., a surface. This is somewhat similar to parameterising latitude and longitude as categorical variables, such as 5x5 areas, which can be thought of as a two-dimensional step function, except that the multivariate spline is continuous and thus allows much greater precision, as will be seen below in the maps of the *lat_lon* effect on predicted CPUE.

Hooks between floats

Hooks between floats (HBF), a proxy for depth, was parameterised as a spline. During the exploratory phase, it was not found necessary to consider alternative parameterisations.

Degrees of Freedom for the DLN Models

A search was conducted over values of the degrees of freedom of each covariate to identify the combination of degrees of freedom that minimised the BIC for each of the logistic and lognormal parts of the DLN. Table 5 presents statistics on the DLN models that were subsequently used to predict shark CPUE. The total number of trips and the number of trips with a positive catch are shown in the columns on the left-hand side, while the number of degrees of freedom of splines and deviance explained are shown on the right-hand side. Variables for which the degrees of freedom that minimised the BIC is one were included in the model as a linear variable and not as a spline; variables for which the degrees of freedom is zero were not included in the model. The deviance explained by each variable in isolation of the other variables is given under each variable; the deviance explained by all variables together is given under *Total*.

For blue sharks, of the 3,014 trips used in the analysis, 2,456 or 81.5% had positive catch rates. This value is much higher than for the other key shark species, which ranged from 58.6% for mako sharks down to 31.3% for silky sharks.

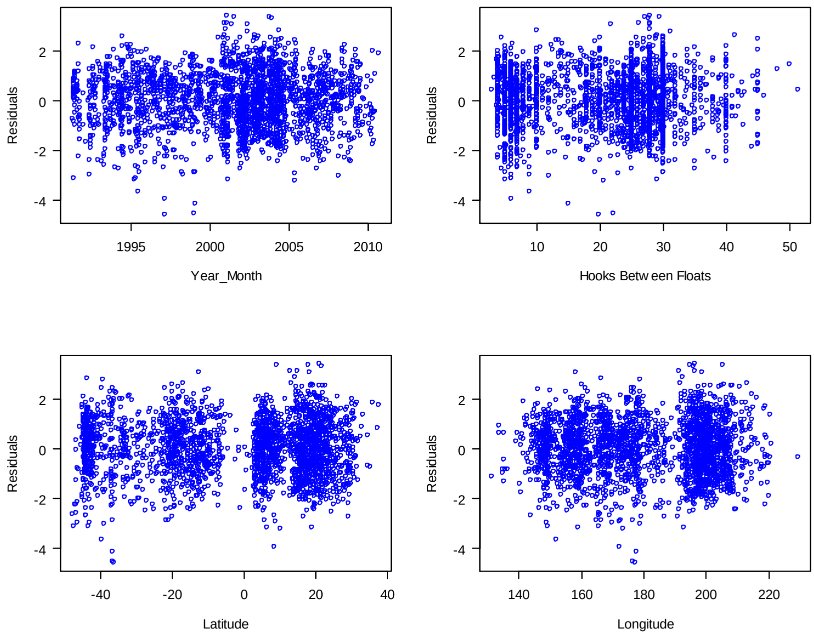
Table 5. Statistics on DLN models of longline CPUE for key shark species and genera

Species or Genera	Observed Trips			Model	Year + Month		Latitude x Longitude		Hooks Between Floats		Total	
	All	Non-Zero	%		Degrees of Freedom	Deviance Explained	Degrees of Freedom	Deviance Explained	Degrees of Freedom	Deviance Explained	Degrees of Freedom	Deviance Explained
Blue Shark	3,014	2,456	81.5%	Logistic	3	9.2%	25	17.6%	1	7.6%	29	24.9%
				Lognormal	8	37.4%	25	31.2%	1	26.8%	34	51.9%
Mako Sharks	3,040	1,781	58.6%	Logistic	1	1.0%	21	13.5%	3	1.6%	25	14.5%
				Lognormal	5	11.1%	11	28.1%	1	9.6%	17	36.8%
Oceanic Whitetip Shark	3,045	1,303	42.8%	Logistic	4	4.7%	13	4.7%	0	---	17	27.9%
				Lognormal	8	20.3%	19	15.6%	3	12.6%	30	41.5%
Silky Shark	3,030	949	31.3%	Logistic	8	12.8%	16	42.5%	0	---	24	46.2%
				Lognormal	4	8.1%	6	32.4%	4	19.9%	14	43.4%
Thresher Sharks	3,030	1,481	48.9%	Logistic	0	---	18	13.4%	0	---	18	14.0%
				Lognormal	0	---	22	42.0%	1	0.3%	23	42.3%

Residuals

The residuals for the lognormal part of the DLN model of blue shark CPUE are plotted in Figure 4; each residual represents one trip. The residuals tend not to exhibit lack of fit. This is typical of these models and plots of residuals will not be shown further.

Figure 4. Plots of residuals for the lognormal part of the DLN model of blue shark CPUE for longline



Effects of the Covariates on Predicted CPUE

The effect of each covariate on predictions of catch rates from the DLN model was examined by fixing the other covariates at a pre-determined value and varying the covariate being examined. (See the appendix for an explanation of how the basis functions were determined for the predictions.) For example, to examine the effect of *year_month* on CPUE, the other independent variables — *latitude, longitude* and *HBF* — were held at fixed values, while CPUE was predicted for values over the range of *year_month*. The fixed value for *latitude* was set to zero (the equator); the fixed value for *longitude* was set to 180; and the fixed value for *HBF* was set to 10. When examining the effects of the other covariates, the fixed value for *year_month* was set to 2000 + (6 – 0.5) / 12 = 2000.458, i.e. June 2000.

Latitude and longitude

To examine the effect of *latitude* and *longitude*, the variables *year_month* and *HBF* were held at their fixed values, while CPUE was predicted for the central point of all 1x1 grids in the *lat_lon* surface. The values of predicted CPUE were plotted for each 1x1 grid in the *heat maps* shown in Figure 5. In a heat map, the colour red indicates low values, white indicates high values and yellow indicates intermediate values. The scale of the contours is approximately logarithmic, rather than linear, to highlight the differences at small values of CPUE.

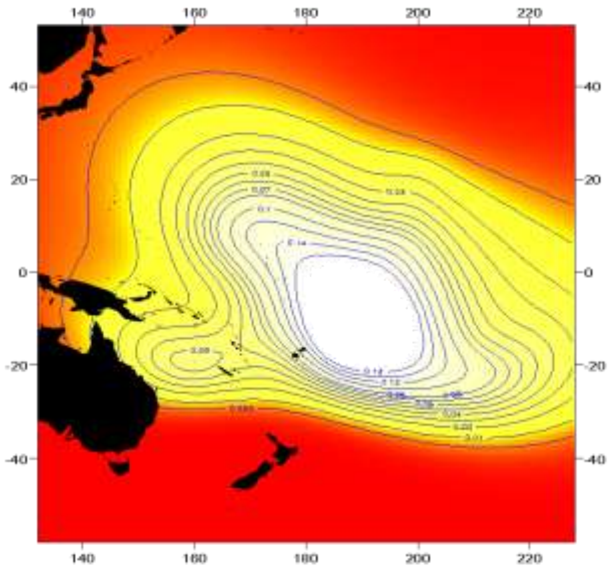
The two heat maps in the top row of Figure 5 are for the two species with higher catch rates in tropical waters: oceanic whitetip shark and silky shark. Catch rates for oceanic whitetip sharks are centred between 10°S and 20°S and appear to have an asymmetric, northwest to southeast, distribution. Catch rates for silky sharks are centred on about the equator and have a more symmetric distribution.

The heat maps in the second and third rows of Figure 5 are for three species and genera with higher catch rates in sub-tropical and temperate waters: blue sharks, thresher sharks and mako sharks. Catch rates for blue sharks appear to be high in both the northern and southern hemispheres. The high values north of 40°N are the result of result of large observed catches during a small number of trips by Hawaiian longliners; since there are few data for those latitudes, they have considerable influence on the *lat_lon* surface.

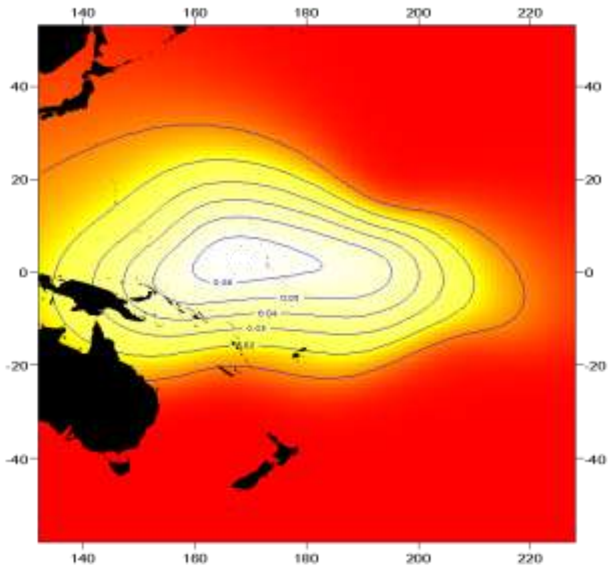
Catch rates for thresher sharks appear to be higher in the northern hemisphere. The observer data are dominated by bigeye threshers (*Alopias superciliosus*) and common threshers (*Alopias vulpinus*), while pelagic threshers (*Alopias pelagicus*) are less common. Catch rates for mako sharks are higher in the southern hemisphere. This group is overwhelmingly dominated by the shortfin makos (*Isurus oxyrinchus*)[4]; longfin makos (*Isurus paucus*) may inhabit tropical waters to a greater extent than shortfin makos

Figure 5. Effect of *latitude* and *longitude* on catch rates (sharks per 100 hooks) of key shark species and genera

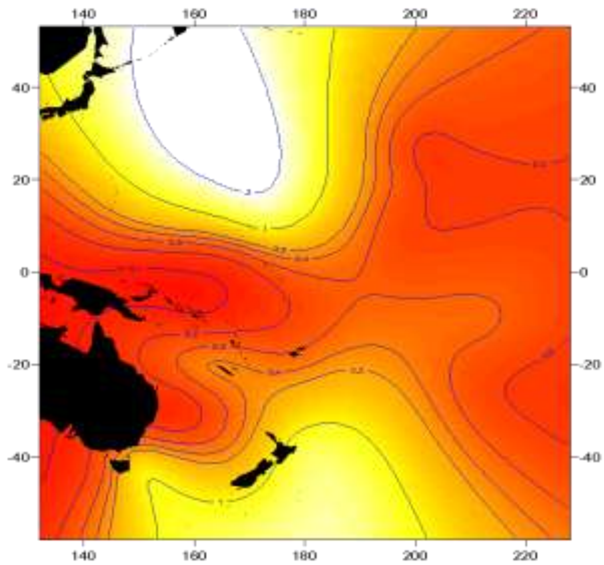
Oceanic Whitetip Shark



Silky Sharks



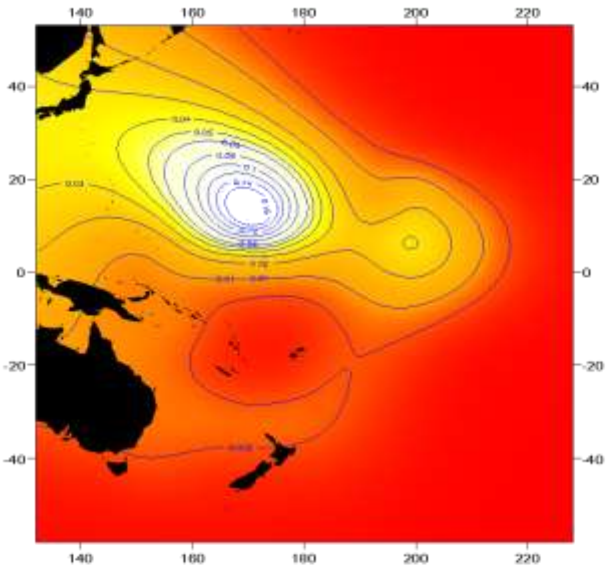
Blue Shark



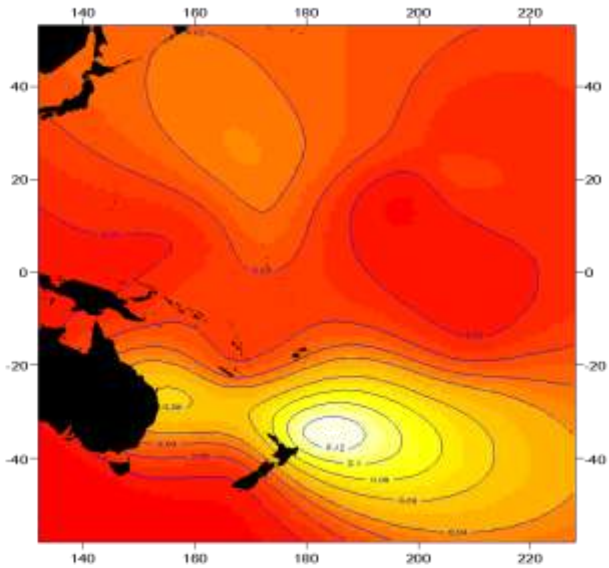
[4] Also, longfin makos may potentially be mis-identified as shortfin makos.

Figure 5 (continued)

Thresher Sharks



Mako Sharks

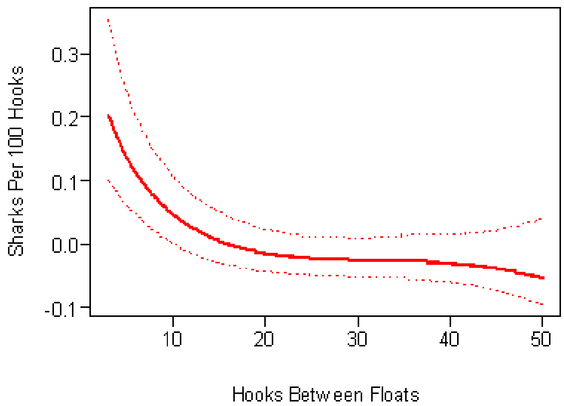


Hooks between floats

To examine the effect of *hooks between floats*, the variables *year_month*, *latitude* and *longitude* were held at their fixed values, while CPUE was predicted for 400 equally spaced values of *hooks between floats* ranging from the minimum to the maximum observed values; the mean of the predictions was then subtracted from the predictions to show the relative effect. The values of predicted CPUE are shown in Figure 6, with 95% confidence intervals. The decline in CPUE with depth is particularly steep for the tropical species, oceanic whitetip and silky sharks.

Figure 6. Effect of *hooks between floats* on catch rates of key shark species and genera

Oceanic Whitetip Shark



Silky Sharks

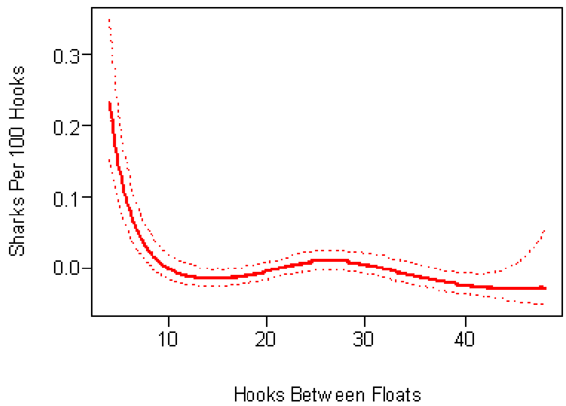
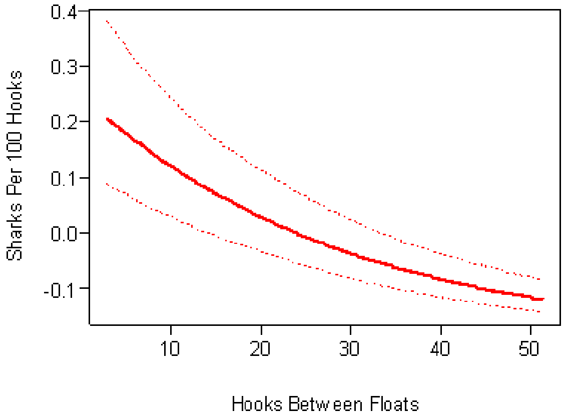
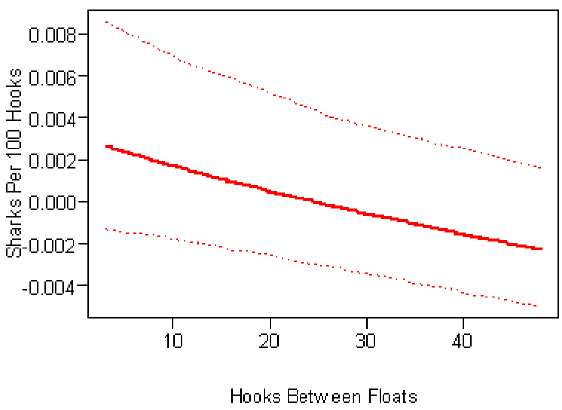


Figure 6 (continued)

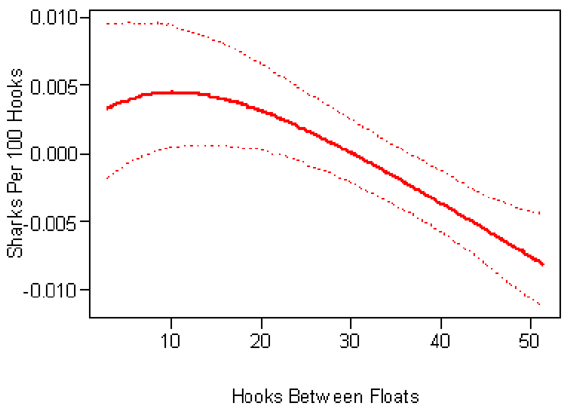
Blue Shark



Thresher Sharks



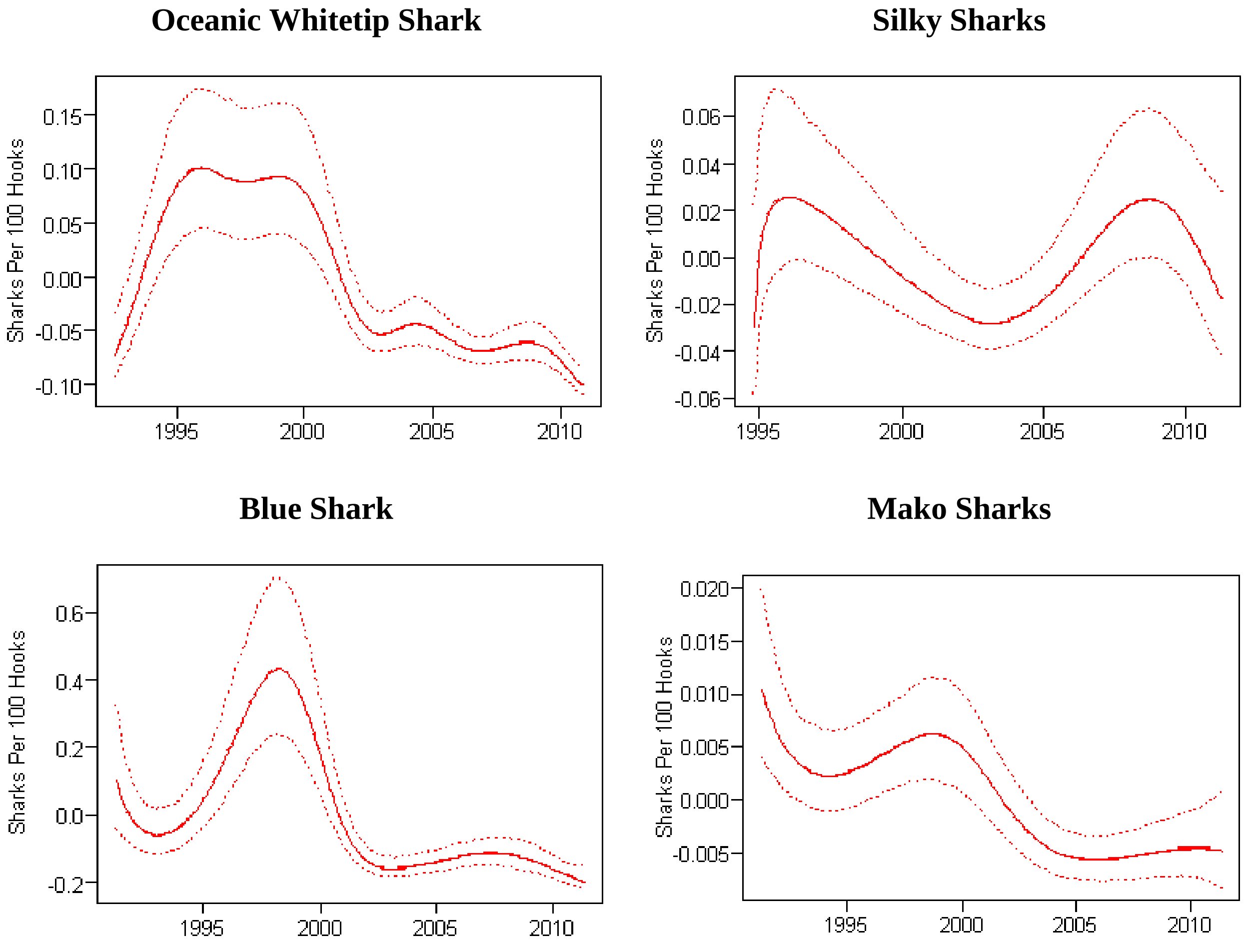
Mako Sharks



Year and month

To examine the effect of *year_month*, *hooks between floats*, *latitude* and *longitude* were held at their fixed values, while CPUE was predicted for 400 equally spaced values of *year_month* ranging from the minimum to the maximum observed values; the mean of the predictions was then subtracted from the predictions to show the relative effect. The values of predicted CPUE are shown in Figure 7, with 95% confidence intervals. For thresher sharks, *year_month* was not included in either the logistic or lognormal parts of the DLN model and so there is no effect.

Figure 7. Effect of *year_month* on longline catch rates of key shark species and genera



The interpretation of the *year_month* effect as an index of population abundance is complicated by (i) under-reporting of sharks by observers, reporting "sharks" without recording the species or genus, and possibly errors in species identification, in the early years of the time series, (ii) operational changes in the fishery, and (iii) possible targeting of sharks.

During the period 1992–1994, the quality of observer data collected in the region was less than in subsequent years. Under the South Pacific Regional Tuna Resource Assessment and Monitoring Project (SPRTRAMP), which was implemented by SPC in 1995, the training of observers improved considerably and the debriefing of observers was introduced. The complete lack of silky sharks in the observed catch during 1992–1993 and a low observed catch in 1994 (see also Table 5 and Figure 8 below) are due to reporting errors, and such errors may have affected the estimates of catch rates of other shark species early in the time series.

The following operation changes in longline fishing are known to have affected shark catch rates (Clarke et al. 2010):

- Japan longline fishing in the AFZ ceased in 1997.
- A trip limit for sharks was imposed in Australia in 2000.
- Shark finning was banned in Hawaii in 2000.
- The shallow set longline fishery in Hawaii was closed from 2001 to 2004.
- The use of wire traces generally has declined since 2004.
- Wire traces were banned in Australia in 2005.

Targeting of sharks by the Japanese offshore and distant-water longline fleet (LLL) in the North Pacific was examined by Clarke et al. (2011). They conclude that the "[c]alculation of concentration indices for the LLL fleet provides some evidence for increasing targeting of blue sharks, and perhaps makos, within the main longline fishing grounds in the North Pacific (i.e. Region 1) since the late 1990s. Other information on total catches and catch rates (nominal and standardized), as well as indications from target species information recorded on logsheets, are consistent with this trend." While the logsheet data covering the Japanese longline fleet in the North Pacific that were examined by Clarke et al. (2011) were not available for this analysis, observer data covering the Japanese fleet in the AFZ in the 1990s were included (Table 1). The increases in standardised (Figure 7) and nominal catch rates (Figure 8) for blue shark during the mid-1990s — and, to a lesser extent, for mako sharks — may therefore be due to increased targeting by Japanese vessels fishing in the AFZ and the subsequent decline in blue shark catch rates shown in Figures 7 and 8 may be due, in part, to the decrease in the amount of observer data covering the Japanese longline fleet, following its cessation of fishing in the AFZ in 1997.

Estimates of Shark Catch Rates and Catches

Shark catches were estimated from longline effort data stratified by year, month, 5x5 area and two categories of *hooks between floats*: shallow (< 10 HBF) and deep (≥ 10 HBF). The effort data cover the WCPFC Statistical Area, east of 130°E; catches by the fleets of Indonesia and the Philippines were ignored because no observer data nor effort data are available for these fleets. Table 6 presents annual shark catches estimated using the method described above, while Figure 8 shows plots of the time series of estimates of annual catch rates and catches, with 95% confidence intervals. The point estimate of each annual catch in Figure 8 is the median of the set of 1000 parametric bootstrap estimates of the annual catch (see Method), while each catch rate is the median of the set of 1000 estimates of the annual catch divided by the known annual effort. The estimates of catch rates are thus nominal and so differ from the plots in Figure 7.

Table 6. Estimates of longline shark catches (thousands of sharks) in the WCPFC Statistical Area east of 130°E

Year	Oceanic Whitetip	Silky Shark	Blue Shark	Thresher Sharks	Mako Sharks	Total
1992	39	0	1,351	58	86	1,534
1993	85	0	1,333	64	71	1,552
1994	184	16	1,662	70	75	2,007
1995	236	161	2,350	75	73	2,896
1996	196	140	3,050	68	72	3,527
1997	186	135	3,587	57	76	4,040
1998	249	165	4,049	62	90	4,615
1999	223	167	3,683	74	100	4,247
2000	186	163	2,124	70	91	2,635
2001	122	149	1,033	71	84	1,459
2002	110	142	627	80	79	1,038
2003	88	97	574	76	74	909
2004	100	103	639	75	65	983
2005	74	114	671	71	55	985
2006	46	133	642	64	47	932
2007	51	167	672	72	44	1,006
2008	55	185	588	71	47	946
2009	53	189	358	61	53	715
Average	127	124	1,611	69	71	2,001
%	6.34%	6.18%	80.48%	3.44%	3.55%	100.00%

Figure 8. Estimates of longline catch rates (left) and catches (right) of sharks in the WCPFC Statistical Area east of 130°E

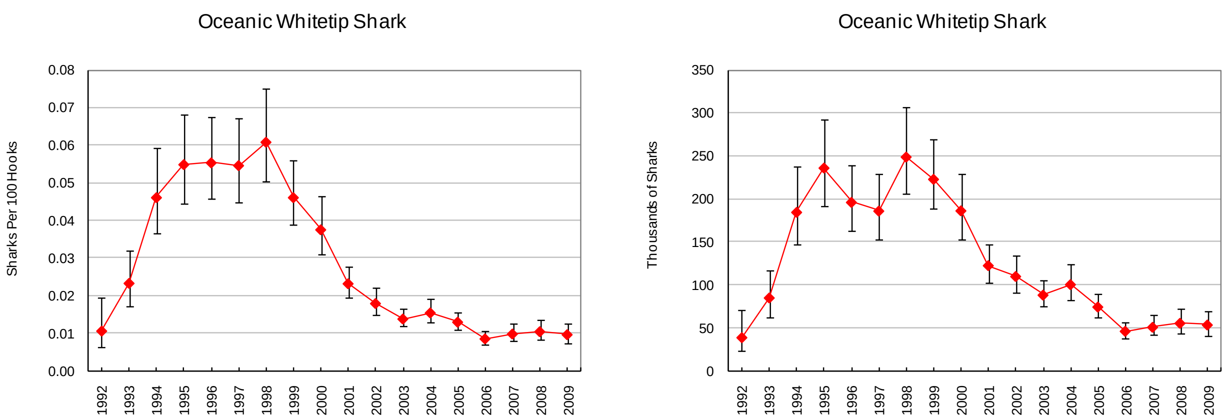


Figure 8 (continued)

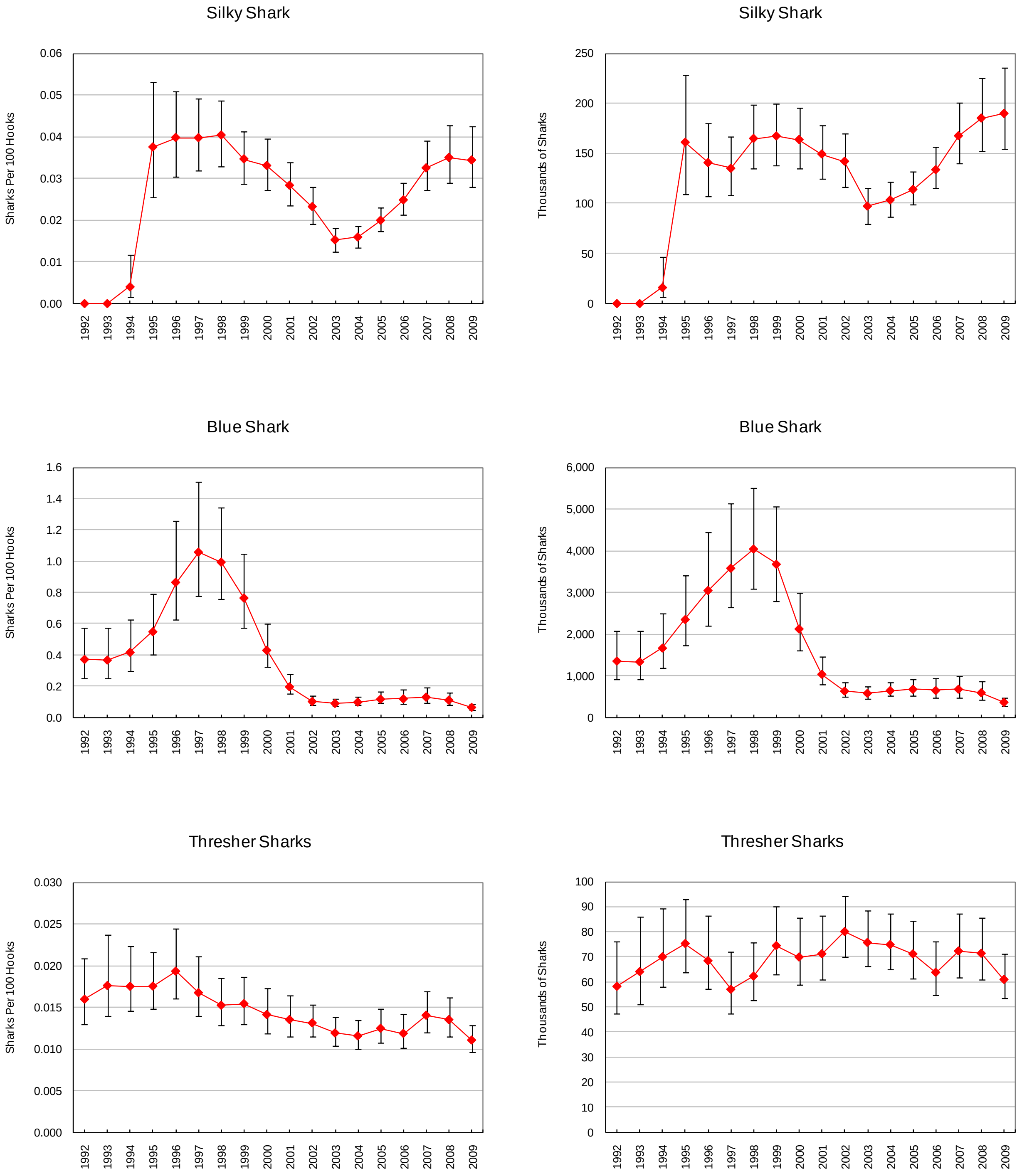
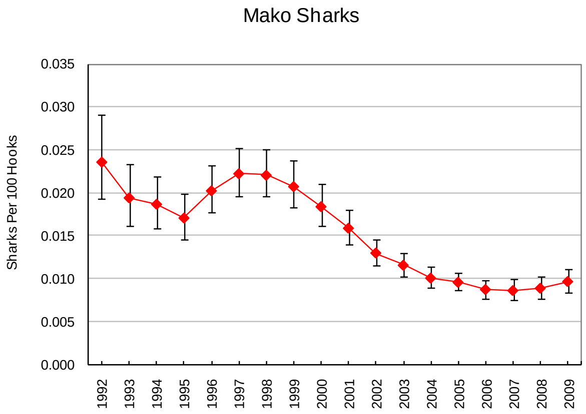
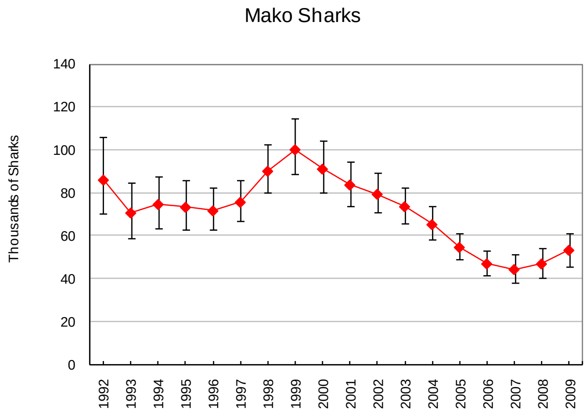


Figure 8 (continued)

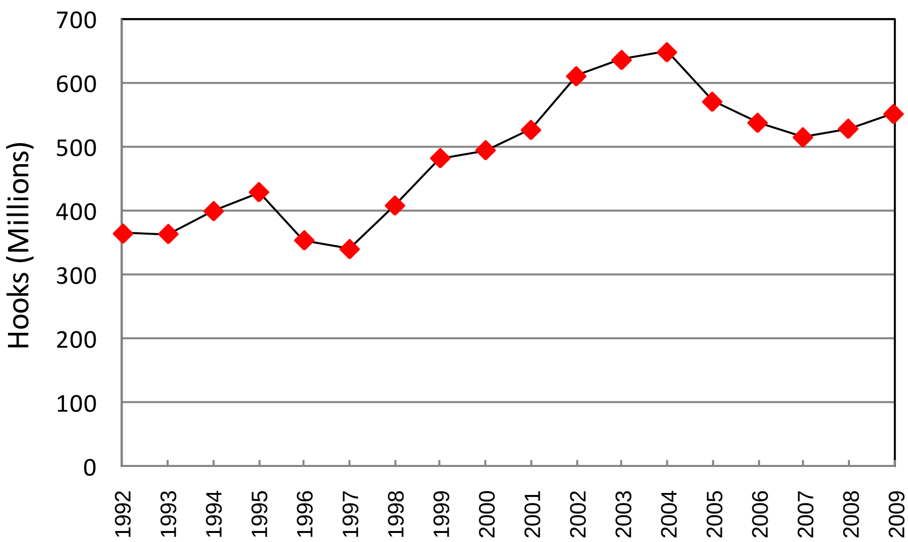




As discussed above, the accuracy of the estimates of catch rates shown in Figure 8 may be affected by reporting errors early in the time series, particularly for silky sharks, and possibly by the targeting of sharks. The trends in estimates of annual catch rates and catches are discussed in Clarke (2011) along with other indicators of the status of shark populations.

The time series of estimates of shark catches depend on longline effort; Figure 8 shows longline effort east of 130°E, excluding the fleets of Indonesia and the Philippines. Since peaking in 2004, longline effort in the region has declined.

Figure 9. Longline effort in the WCPFC Statistical Area east of 130°E



Application of Generalised Estimating Equations (GEE)

Generalised Estimating Equations are a general method for analysing data collected in *clusters* where observations within a cluster may be correlated and observations in separate clusters are independent. The **geeglm** function in the R package **geepack** (Halekoh et al. 2006) was used to apply GEEs to the logistic and lognormal parts of the DLN models of shark catch rates for longline trips described above. The clusters were defined as the eight longline sectors listed in Table 2.

Four working correlation structures are available in **geeglm**: (i) *independence*, in which the observations within a cluster are independent; (ii) *exchangeable*, in which all observations in a cluster have the same correlation; (iii) *ar1*, in which the correlations are auto-regressive; and (iv) *unstructured*, in which the correlation between each and every pair of observations in a cluster is distinct. The *exchangeable* correlation structure was used for both the logistic and lognormal parts of the DLN models of shark catch rates.

With the *exchangeable* correlation structure, the correlation of observations within sectors is the same for each sector and is estimated by the correlation parameter *alpha*. To examine the effect of the covariates on the estimate of the correlation within sectors, **geeglm** was first used to fit the response variables of the logistic and lognormal parts of the DLN models with only the intercept as a predictor, without the covariates; it was then used to fit the response variables with the full models, i.e., with the covariates parameterised as splines as listed in Table 5. The estimates of the correlation parameter *alpha* for the logistic and the lognormal responses are presented in Table 7 for both cases. With only the intercept, the estimates of correlation parameters are all small to moderate positive values, as might be expected, with the exception of the lognormal response for oceanic whitetip, which was much smaller than the others. With the full models, the estimates of *alpha* are all negligible. These results indicate that the full models capture all of the correlation among the responses and suggest that the use of the catch rate per trip as replicates in the DLN models is appropriate; however, it is not clear why the sign of the correlations for the full models are all negative.

Table 7. Estimates of the correlation parameter *alpha* for the logistic and lognormal responses of DLN models of shark catch rates

Species	Intercept Only		Full Model	
	Logistic	Lognormal	Logistic	Lognormal
Blue Shark	0.0813	0.2641	-0.0011	-0.0015
Mako Sharks	0.1375	0.1067	-0.0010	-0.0023
Oceanic Whitetip	0.0709	0.0122	-0.0010	-0.0022
Silky Shark	0.2758	0.4408	-0.0011	-0.0024
Thresher Sharks	0.0751	0.3340	-0.0007	-0.0017

GEEs also allow for overdispersion, i.e., a higher variance in the response variable than that assumed by the probability distribution used to model the response. Overdispersion is estimated in **geeglm** by the *scale* parameter, with values greater than 1.0 indicating overdispersion; estimates of the *scale* parameter for the full DLN models of shark catch rates are given in Table 8. Overdispersion is considerable for blue shark and thresher sharks, indicating that the variances of the estimates of the DLN parameters for these species from standard GLMs are being under-estimated.

Table 8. Estimates of the *scale* parameter for the logistic and lognormal parts of DLN models of shark catch rates

Species	Logistic	Lognormal
Blue Shark	1.0855	1.2247
Mako Sharks	0.9849	0.6062
Oceanic Whitetip	1.1328	0.9841
Silky Shark	0.8780	0.9397
Thresher Sharks	1.1615	1.4521

Overdispersion is accounted for by **geeglm** when estimating the covariance matrix of the estimates of the model parameter with the *sandwich* variance estimate (Halekoh et al. 2006). The results of using *sandwich* variance estimates for the parameters of the logistic and lognormal parts of the full DLN model of blue shark catch rates are shown in Figures 10 and 11. For GEEs with the *independent* or *exchangeable* correlation structures, it is always the case that the model parameter estimates are no different from a standard GLM; hence the point estimates of the effects of *year_month* and *hooks between floats* in Figure 9 are no different from those for blue shark shown in Figures 5 and 6, while the point estimates of the catch rates and catches in Figure 10 are no different from those for blue shark shown in Figure 7. However, in each of the plots shown in Figures 10 and 11, the 95% confidence intervals are much greater than for the standard GLM.

Figure 10. **Effect of *year_month* and *hooks between floats* on blue catch rates determined with Generalized Estimating Equations**

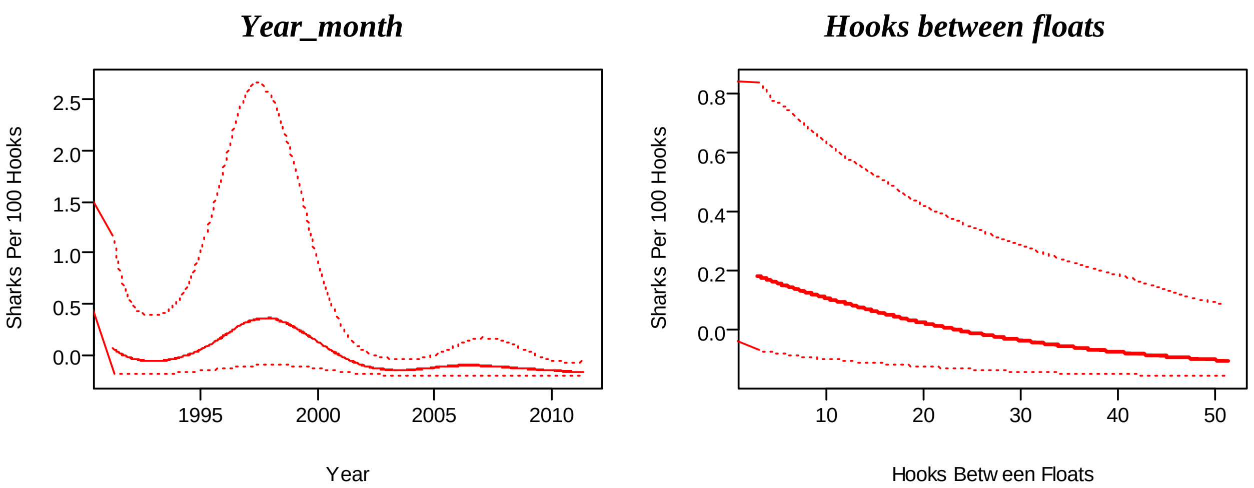
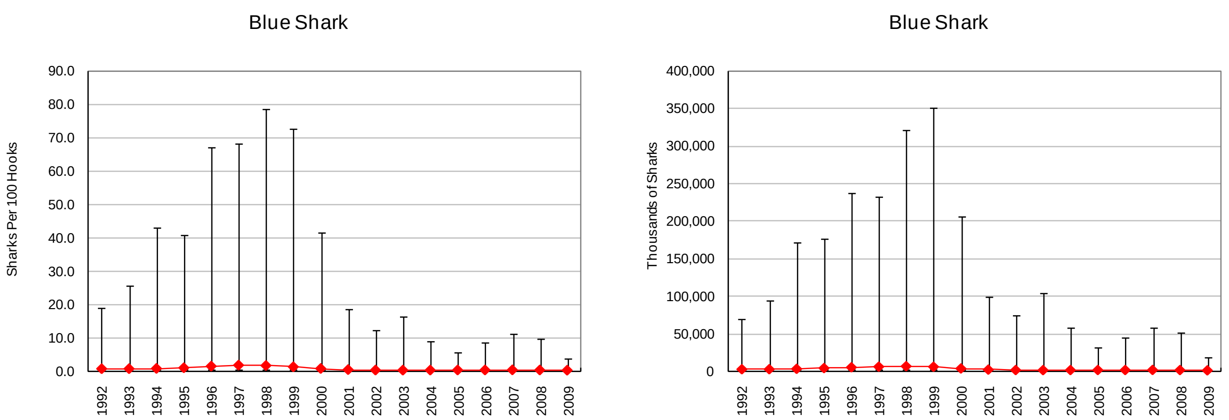


Figure 11. **Estimates of longline catch rates (left) and catches (right) of blue shark in the WCPFC Statistical Area east of 130°E determined from Generalised Estimating Equations**



APPLICATION TO PURSE-SEINE

Definition of Replicates and Responses

For purse seine, the response variable in the logistic part of the DLN model of CPUE is 1 or 0 depending on whether the catch rate in a stratum of trip and school association was positive or zero, while the reponse for the lognormal part is the natural logarithm of the average catch rate in units of number of sharks per day fished or searched in a stratum of trip and school association. If all sets were on schools of the same association (unassociated or associated), an observed trip will have one stratum of trip and school association; otherwise, it will have two.

Only observed trips with at least five days fished or searched and for which information regarding the school associations was available were used. Table 9 presents the number of strata of purse-seine observer trip and school association (unassociated or associated) covered by data held by the OFP and used in the analysis. There are 4,460 strata from 1994 to 2011, including 2,004 (45%) strata of unassociated schools and 2,456 (55%) of associated schools.

Purse-seine catches were estimated only for oceanic whitetip shark and silky shark because the number of strata with positive catches was insufficient to estimate the DLN parameters for the other key shark species and genera. The numbers of strata with positive catches for blue shark, thresher sharks and mako sharks were 39, 64 and 79 respectively.

Table 9. Number of strata of purse-seine trip and school association covered by data held by the OFP and used in the analysis

Year	Unassociated Schools	Associated Schools	Total
1994	14	11	25
1995	32	30	62
1996	52	57	109
1997	48	62	110
1998	78	84	162
1999	18	50	68
2000	32	54	86
2001	56	70	126
2002	82	126	208
2003	90	142	232
2004	139	219	358
2005	166	245	411
2006	199	261	460
2007	179	253	432
2008	184	216	400
2009	240	226	466
2010	392	347	739
2011	3	3	6
Total	2,004	2,456	4,460
%	44.93%	55.07%	100.00%

The replicates were screened to eliminate outliers, with the definition of outliers determined by inspection. Table 10 presents the level of CPUE that defines the outliers, the number of strata for which CPUE was greater than or equal to the level, the percentage of strata that were screened for outliers and the number of strata remaining and used in the analyses.

Table 10. Level of purse-seine CPUE (sharks per day) defining outliers and the number of outliers out of a total of 4,460 strata

Species or Group	Level of CPUE	Number of Outlier Strata	% of Total	Strata Remaining
Oceanic Whitetip Shark	3.0	28	0.63%	4,432
Silky Shark	22.0	19	0.43%	4,441

Parameterisation of the Covariates

The number of days fished or searched per trip was allocated to the strata of unassociated and associated schools in proportion to the number of unassociated and associated schools fished during the trip. For each stratum of trip and school association, the independent variables were assigned as follows. The month during which the largest number of days were fished was assigned as the *year* and *month* for the stratum. The *latitude* and *longitude* assigned to each stratum were the average latitude and longitude of the locations of the sets. The catch rate for each stratum of trip and school association was calculated as the total number of sharks caught divided by the total number of days fished per stratum. As for longline, *year* and *month* were included as *year + (month – 0.5) / 12* and parameterised as a spline, while *latitude* and *longitude* were parameterised as a two-dimensional spline.

Degrees of Freedom for the DLN Models

A search was conducted over values of the degrees of freedom of each covariate to identify the combination of degrees of freedom that minimised the BIC for each of the logistic and lognormal parts of the DLN. Table 11 presents statistics on the DLN models that were subsequently used to predict shark CPUE. The total number of trips and the number of trips with a positive catch are shown in the columns on the left-hand side, while the number of degrees of freedom of splines and deviance explained are shown on the right-hand side. Variables for which the degrees of freedom that minimised the BIC is one were included in the model as a linear variable and not as a spline. The deviance explained by each variable in isolation of the other variables is given under each variable; the deviance explained by all variables together is given under *Total*.

The number of trip – association strata with non-zero catches was much greater for silky shark than for oceanic whitetip shark, 58.4% of all strata compared to 11.6%. The deviance explained for both species was considerably less than for the models of longline CPUE. For oceanic whitetip, none of the covariates explained more than 10% of the deviance, while for silky shark, only school association explained more than 10% of the deviance. The covariate *lat_lon* explained much less of the deviance than in the DLN models of longline CPUE, perhaps because purse-seine catch rates vary less than longline catch rates within the areas covered by the respective observer data.

Table 11. Statistics on DLN models of purse-seine CPUE for two key shark species

Species	Observed Strata			Model	Year + Month		Latitude x Longitude		School Association		Total	
	All	Non-Zero	%		Degrees of Freedom	Deviance Explained	Degrees of Freedom	Deviance Explained	Degrees of Freedom	Deviance Explained	Degrees of Freedom	Deviance Explained
Oceanic Whitetip Shark	4,432	516	11.6%	Logistic	5	5.7%	12	5.2%	1	1.6%	18	12.2%
				Lognormal	1	6.5%	10	3.2%	1	6.0%	12	13.7%
Silky Shark	4,441	2,595	58.4%	Logistic	3	5.9%	10	3.3%	1	10.5%	14	18.4%
				Lognormal	3	0.7%	10	3.8%	1	22.3%	14	27.6%

Effects of the Covariates on Predicted CPUE

The effect of *year_month* and *lat_lon* on predictions of catch rates from the DLN model was examined by fixing the other covariates at a pre-determined value and varying the covariate being examined. As for longline, the fixed value for *latitude* was set to zero (the equator); the fixed value for *longitude* was set to 180; the fixed value for *year_month* was set to 2000.458, i.e. June 2000; and the fixed value of *school association* was *associated*.

Latitude and longitude

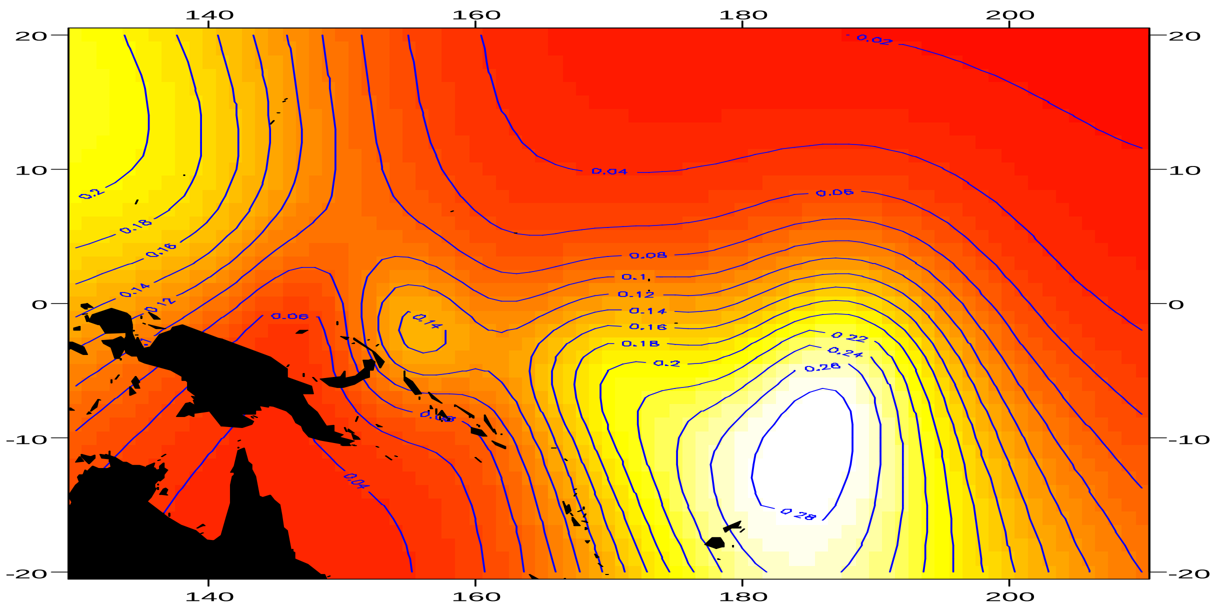
Figure 12 shows the CPUE heat maps for oceanic whitetip caught by purse seiners (top) and compares it to the heat map for longliners (bottom) for the same area. The two heat maps are somewhat similar, with high CPUE in the southeast part of the area and a diagonal axis from the northwest to the southeast. However, the heat map for purse seine also shows high CPUE in the northwest part of the area.

Figure 13 shows similar heat maps for silky shark. The heat maps are less similar than for oceanic whitetip, with the heat map for purse seine showing relatively high CPUE in the northeast part of the area.

The lack of symmetry and the edge effects in the purse-seine heat maps for both species suggest that they may be less informative than for longline, an observation that is consistent with the low level of deviance explained by the *lat_lon* covariate in the DLN models of purse-seine catch rates, compared to longline.

Figure 12. Effect of *latitude* and *longitude* on catch rates of oceanic whitetip shark

Purse seine (sharks per day)



Longline (sharks per 100 hooks)

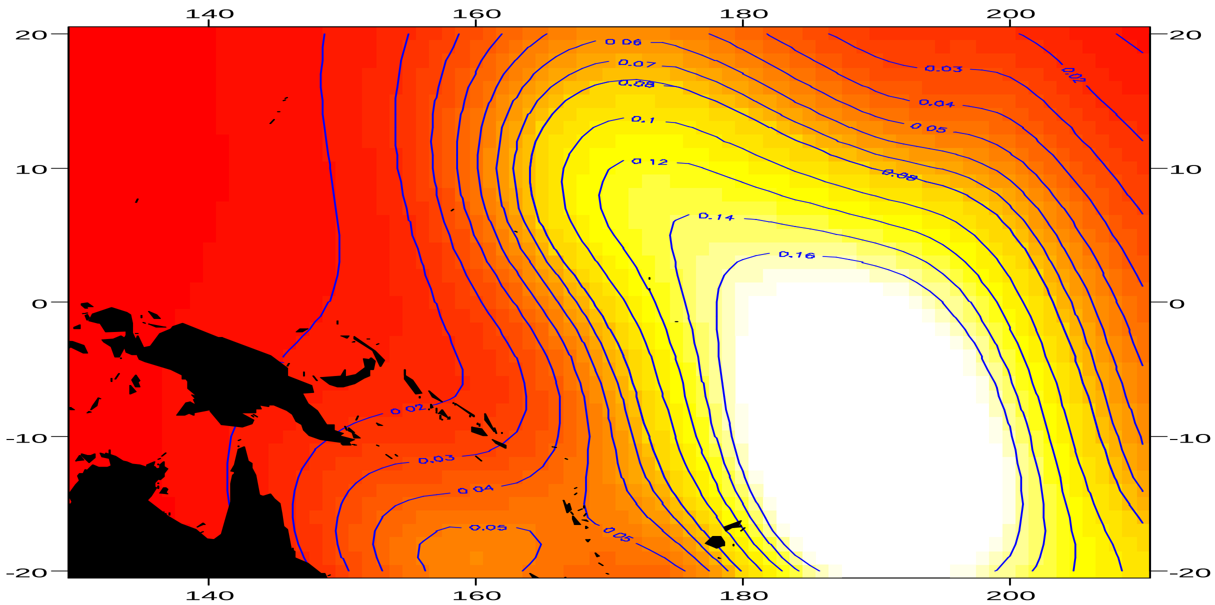
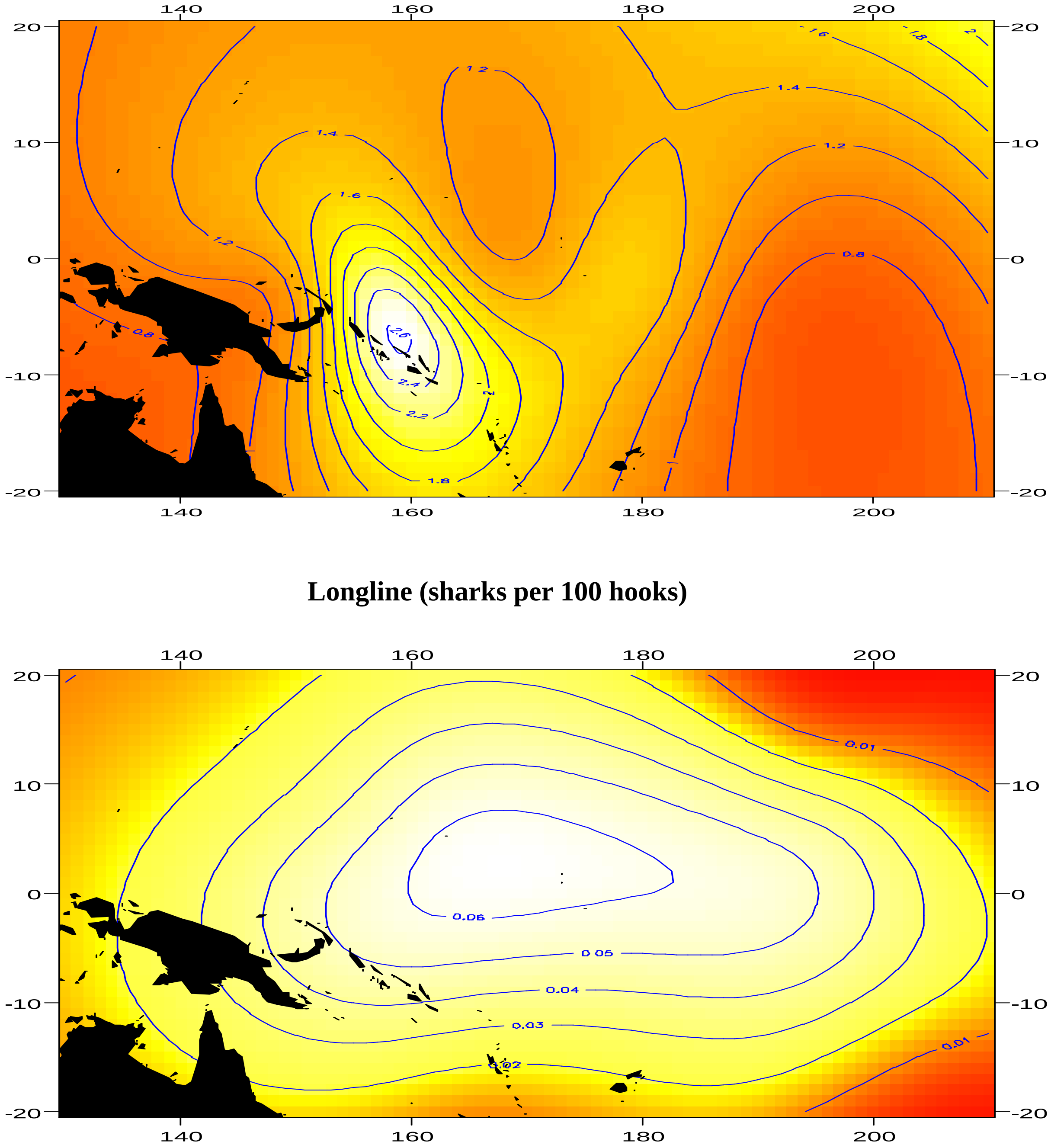


Figure 13. Effect of *latitude* and *longitude* on catch rates of silky shark

Purse seine (sharks per day)

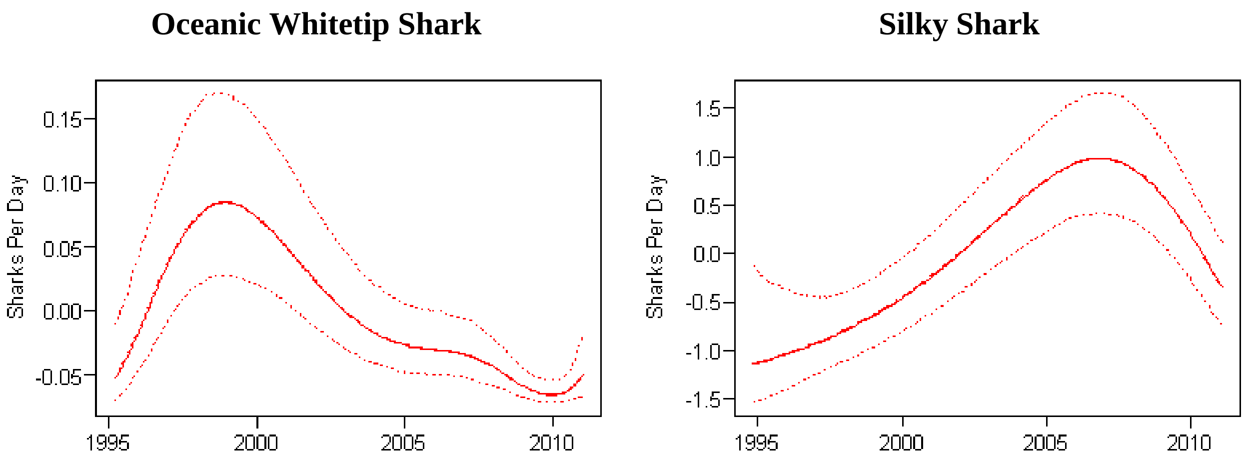
Longline (sharks per 100 hooks)



Year and month

To examine the effect of *year_month*, *latitude*, *longitude* and *school association* were held at their fixed values, while CPUE was predicted for 400 equally spaced values of *year_month* ranging from the minimum to the maximum observed values; the mean of the predictions was then subtracted from the predictions to show the relative effect. The values of predicted CPUE are shown in Figure 15, with 95% confidence intervals.

Figure 14. Effect of *year_month* on purse-seine catch rates of two key shark species



As for longline, the interpretation of the *year_month* effect as an index of population abundance is complicated by reporting errors in the early years of the time series.

Estimates of Shark Catch Rates and Catches

Shark catches were estimated from purse-seine effort data stratified by year, month, areas of 2° of latitude and 5° of longlitude, and school association (unassociated and associated). The effort data cover the area from 20°S to 20°N and 130°E to 210°W. Table 12 presents annual catch estimates, while Figure 16 shows plots of the time series of estimates of annual catch rates and catches, with 95% confidence intervals. As for longline, the point estimates are the median of the 1000 parametric bootstrap estimates and the catch rates in Figure 16 are nominal, rather than standardised, and so differ from the plots in Figure 15.

Table 12. Estimates of purse-seine catches (number of sharks) of two key shark species in the area from 20°S to 20°N and 130°E to 210°W

Year	Oceanic Whitetip	Silky Shark	Total
1995	997	23,800	24,797
1996	2,492	24,561	27,053
1997	3,677	28,102	31,779
1998	4,065	27,422	31,486
1999	4,302	35,172	39,474
2000	3,556	31,358	34,914
2001	3,003	35,069	38,072
2002	2,740	43,042	45,782
2003	2,076	56,544	58,620
2004	1,938	84,679	86,617
2005	1,747	78,976	80,723
2006	1,585	81,454	83,039
2007	1,392	78,999	80,391
2008	1,128	78,904	80,033
2009	711	69,790	70,501
2010	864	47,861	48,726
Average	2,267	51,608	53,875
%	4.21%	95.79%	100.00%

Figure 15. Estimates of purse-seine catch rates (left) and catches (right) of two key shark species in the area from 20°S to 20°N and 130°E to 210°W

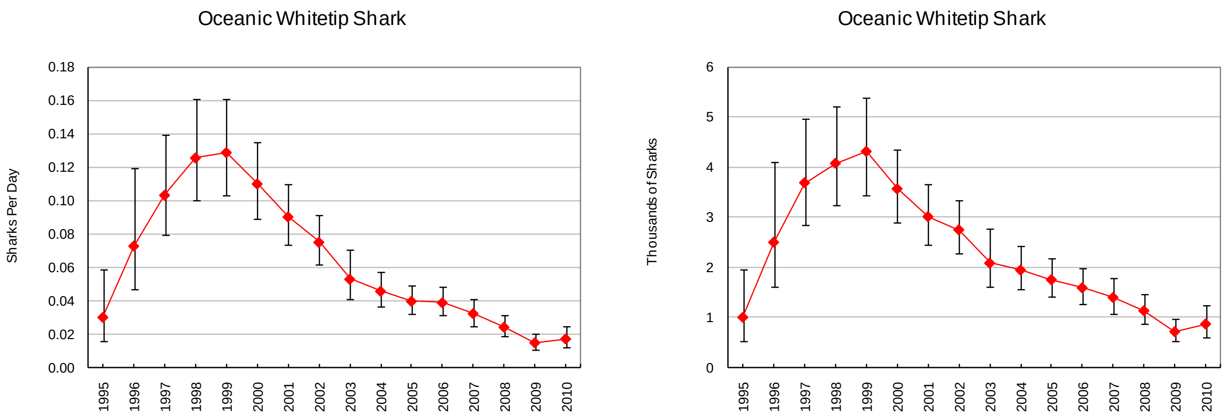
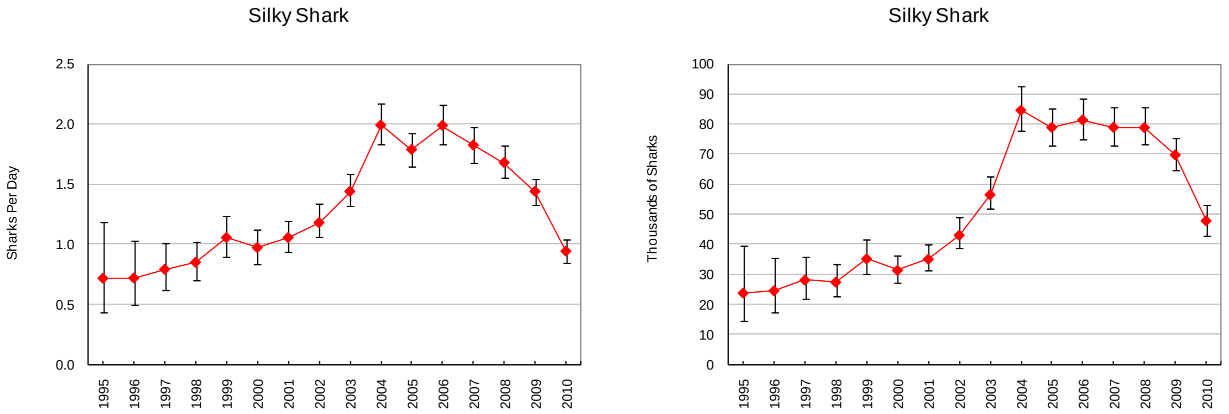


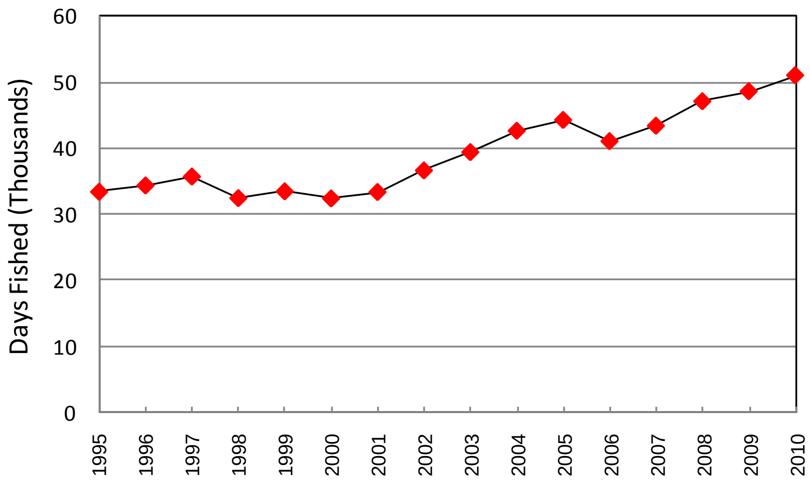
Figure 16 (continued)



As discussed above, the accuracy of the estimates of catch rates shown in Figure 16 may be affected by reporting errors early in the time series. The trends in oceanic whitetip catch rates and catches by purse seiners are similar to those for longline, while the trends for silky sharks are quite different than for longline. The trends in estimates of annual catch rates and catches are discussed in Clarke (2011) along with other indicators of the status of shark populations.

The time series of estimates of shark catches depend on purse-seine effort, which is shown in Figure 17.

Figure 16. Purse-seine effort from 20°S to 20°N and 130°E to 210°W



REFERENCES

Chambers, J.M. & T.J. Hastie. 1992. Statistical Models in S. Wadsworth & Brooks / Cole, Pacific Grove, California, United States of America.

Clarke, S.C. 2011. A status snapshot of key shark species in the Western and Central Pacific and potential mitigation options. WCPFC–SC7–2011/EB–WP–04. Seventh Regular Session of the WCPFC Scientific Committee, 9–17 August 2011, Pohnpei, Federated States of Micronesia. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.

Clarke, S.C. & S.J. Harley. 2010. A proposal for a research plan to determine the status of the key shark species. WCPFC–SC6–2010/EB–WP–01. Sixth Regular Session of the WCPFC Scientific Committee, Nuku'alofa, Tonga, 10–19 August 2010. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.

Clarke, S.C., T.A. Lawson, D. Bromhead & S.J. Harley. 2010. Progress toward shark assessments. WCPFC–2010–16. Seventh Regular Session of the Western and Central Pacific Fisheries Commission, 6–10 December 2010, Honolulu, Hawaii, United States of America. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia.

Clarke, S.C., K. Yokawa, H. Matsunaga & H. Nakano. 2011. Analysis of North Pacific shark data from Japanese commercial longline and research/training vessel records. WCPFC–SC7–2011/EB–WP–02. Seventh Regular Session of the WCPFC Scientific Committee, 9–17 August 2011, Pohnpei, Federated States of Micronesia. Oceanic Fisheries Programme, Secretariat of the Pacific Community, Noumea, New Caledonia and National Research Institute of Far Seas Fisheries, Shimizu, Japan.

Halekoh, U., S. Højsgaard & J. Yan. 2006. The R package **geepack** for Generalized Estimating Equations. *Journal of Statistical Software* 15 (2): 1–11.

Liang, K.Y. & S.L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73 (1): 13–22

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

Figure A1. Distribution of longline hooks set and hooks observed in the WCPFC Statistical Area, excluding the fleets of Indonesia and the Philippines

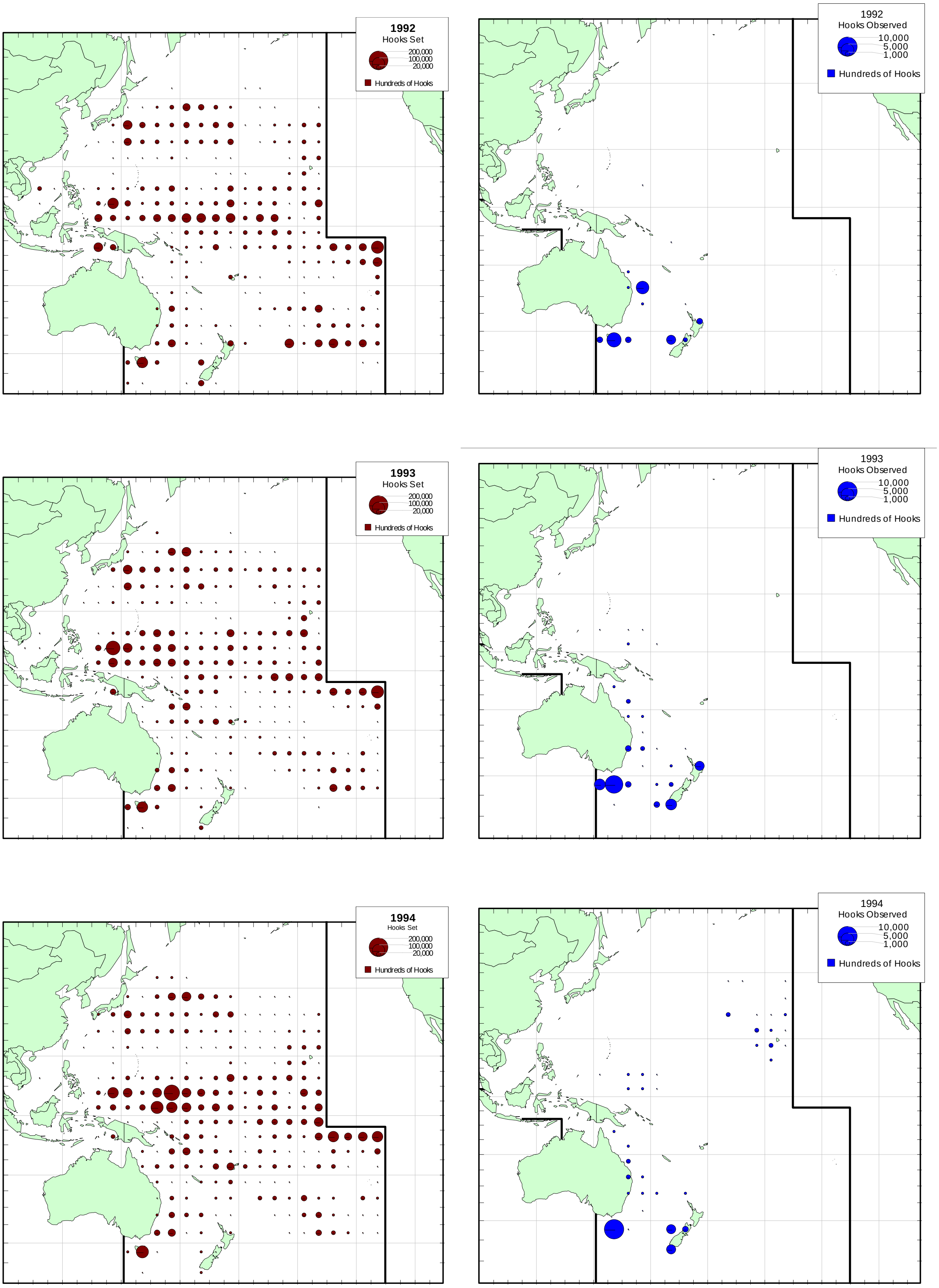


Figure A1 (continued)

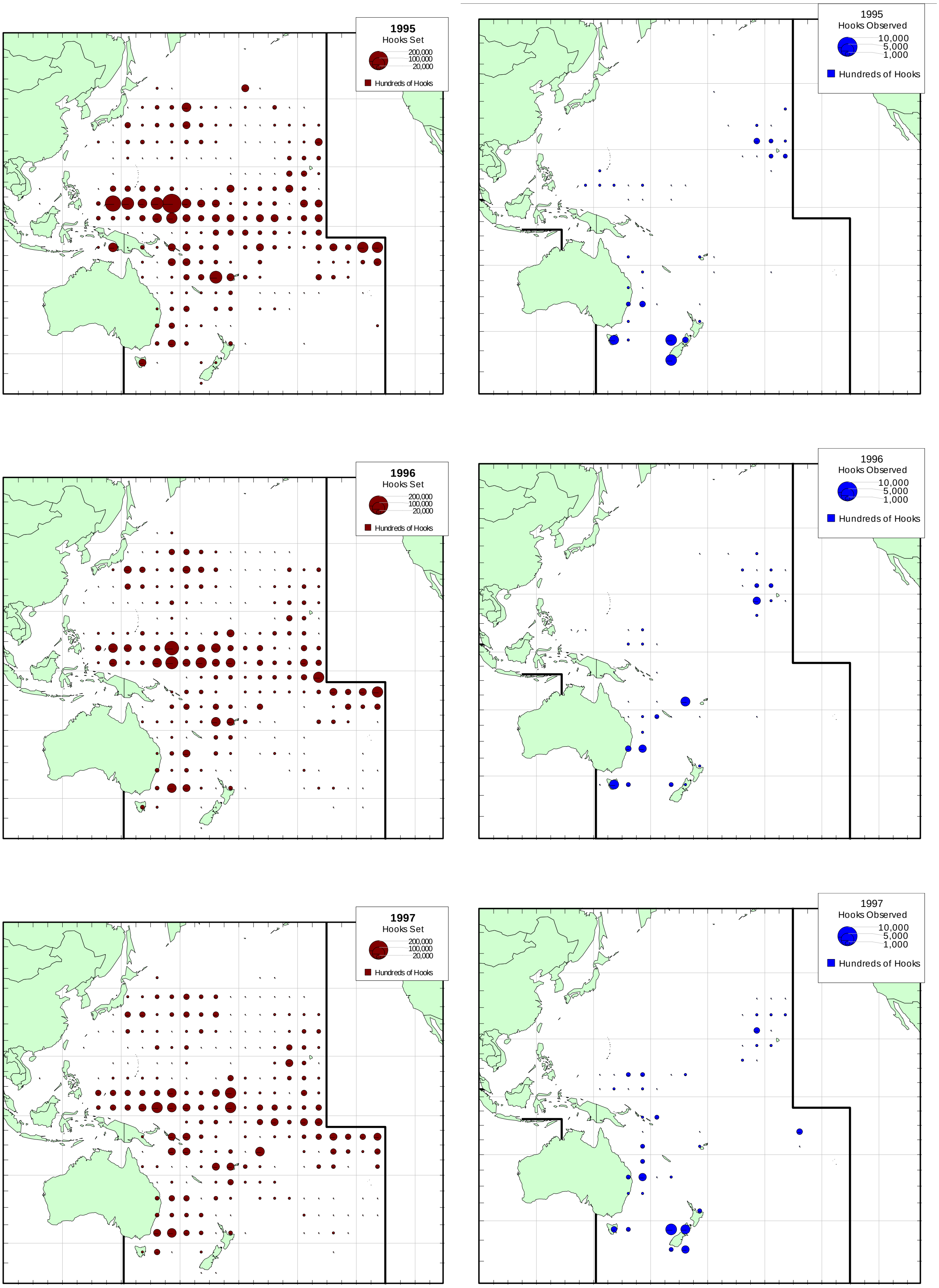
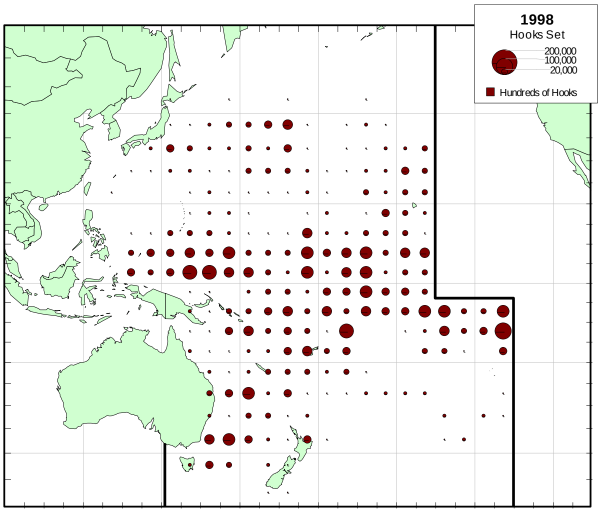
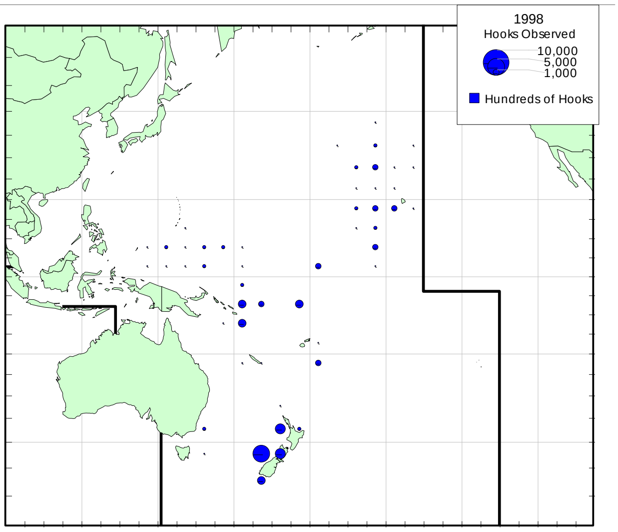
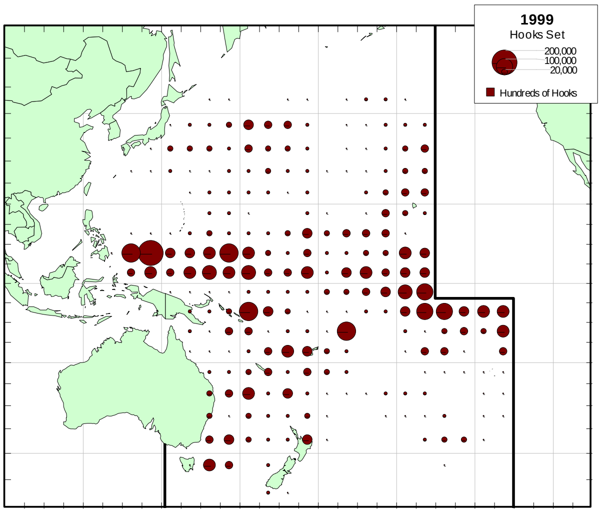
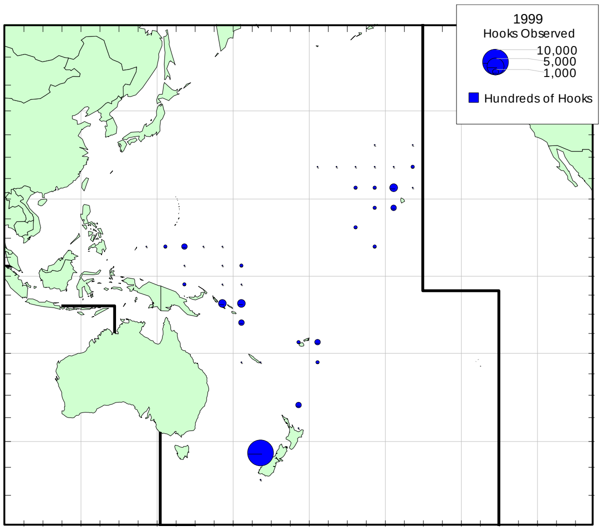


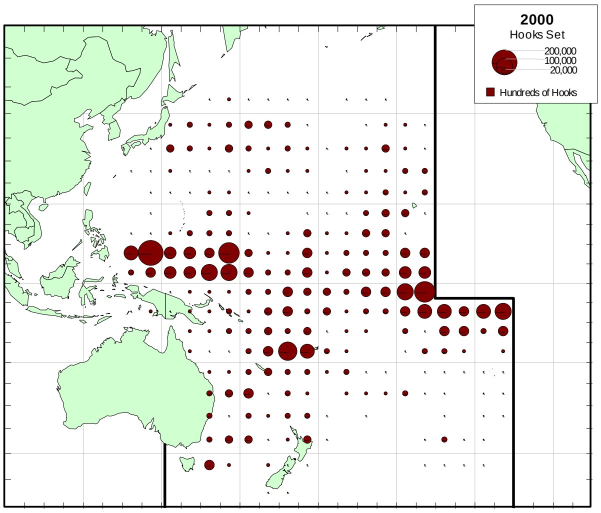
Figure A1 (continued)











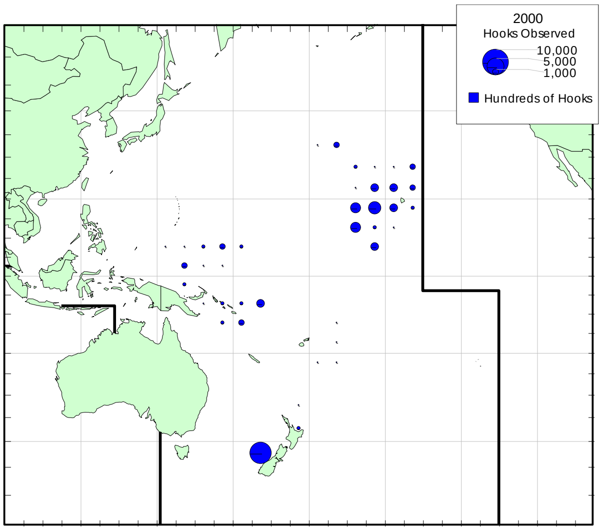
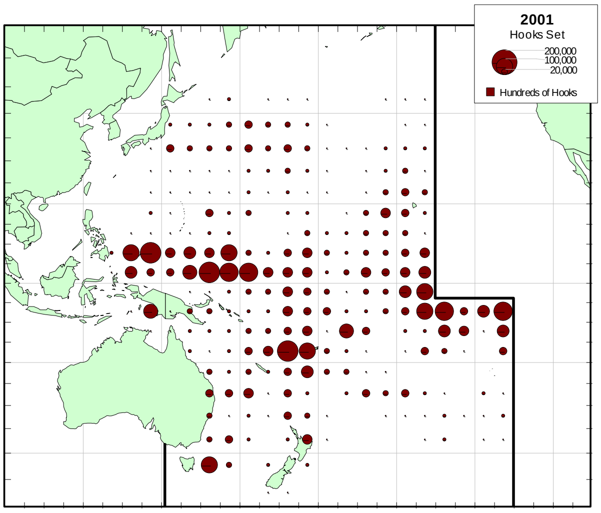
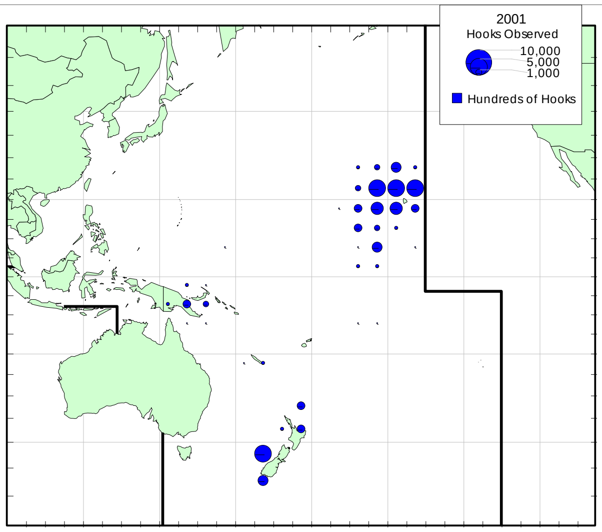
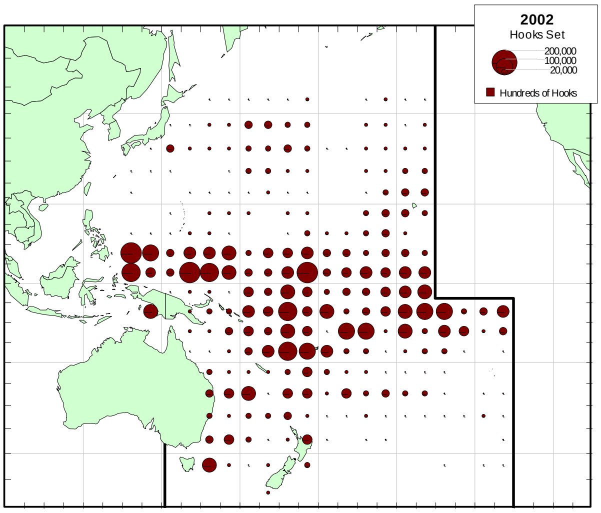
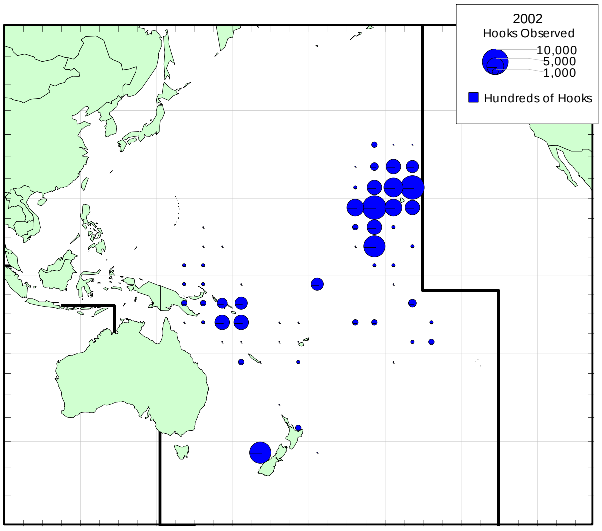


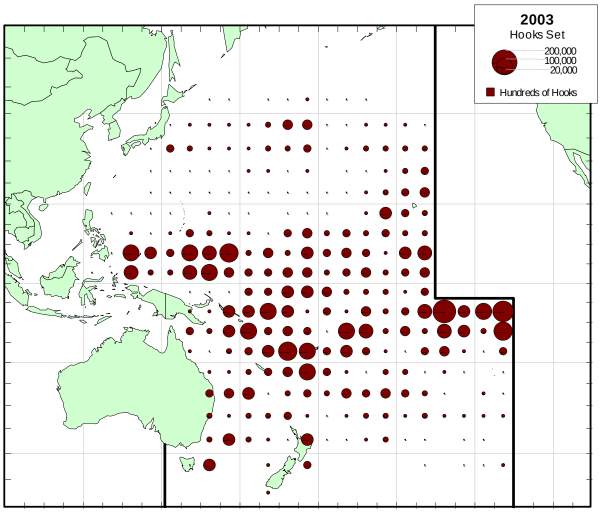
Figure A1 (continued)











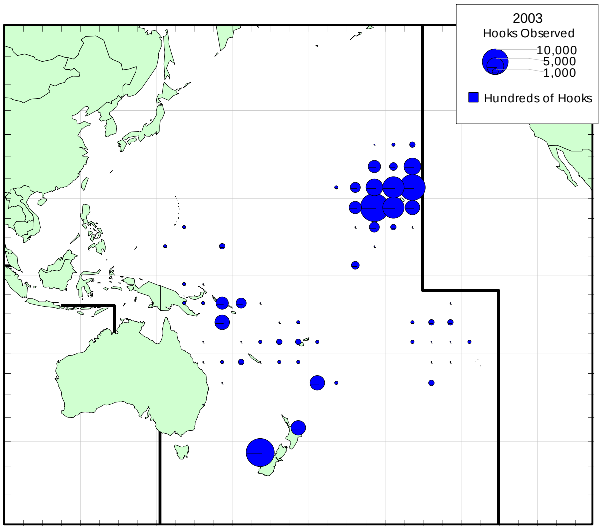


Figure A1 (continued)

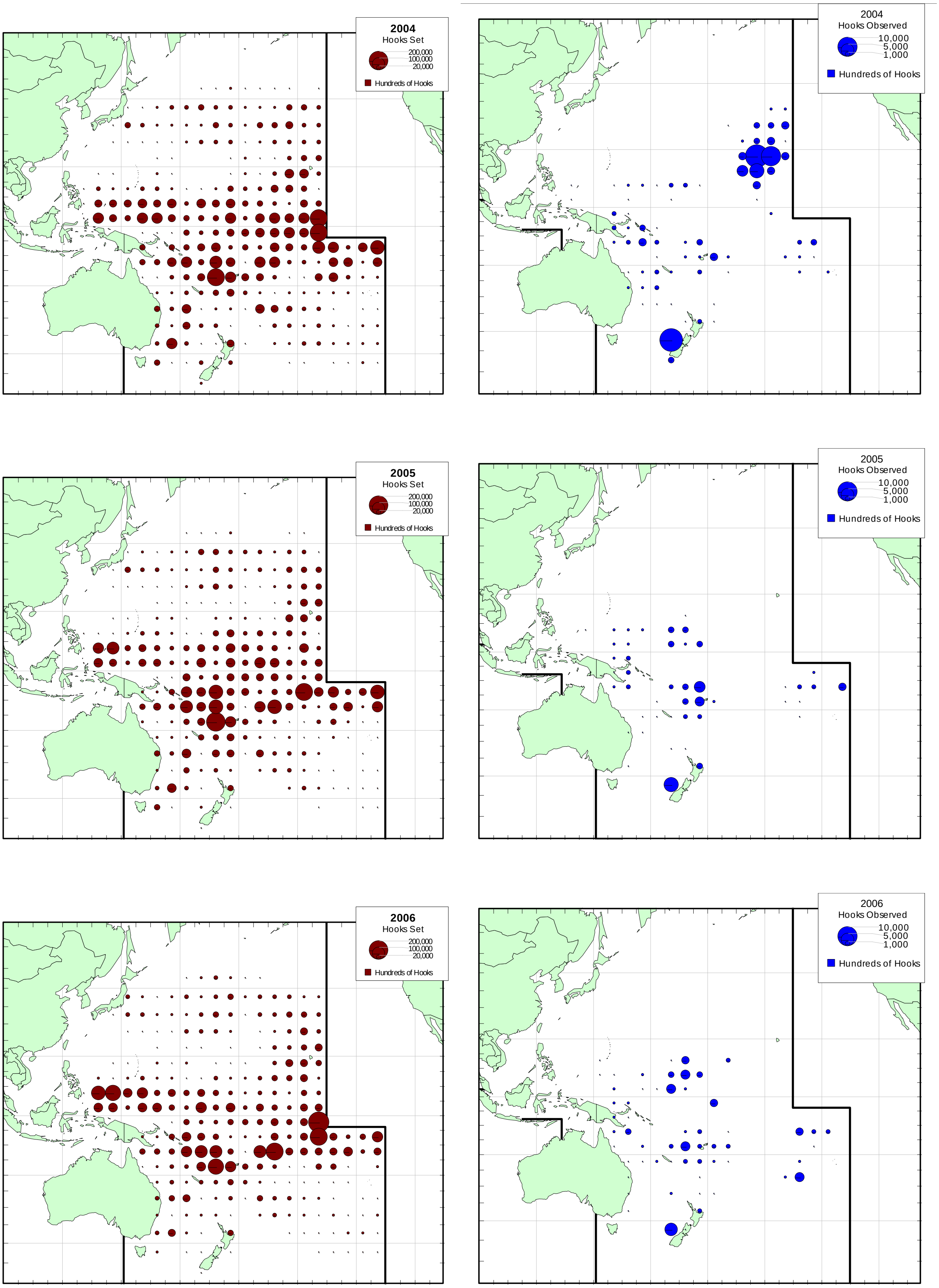


Figure A1 (continued)

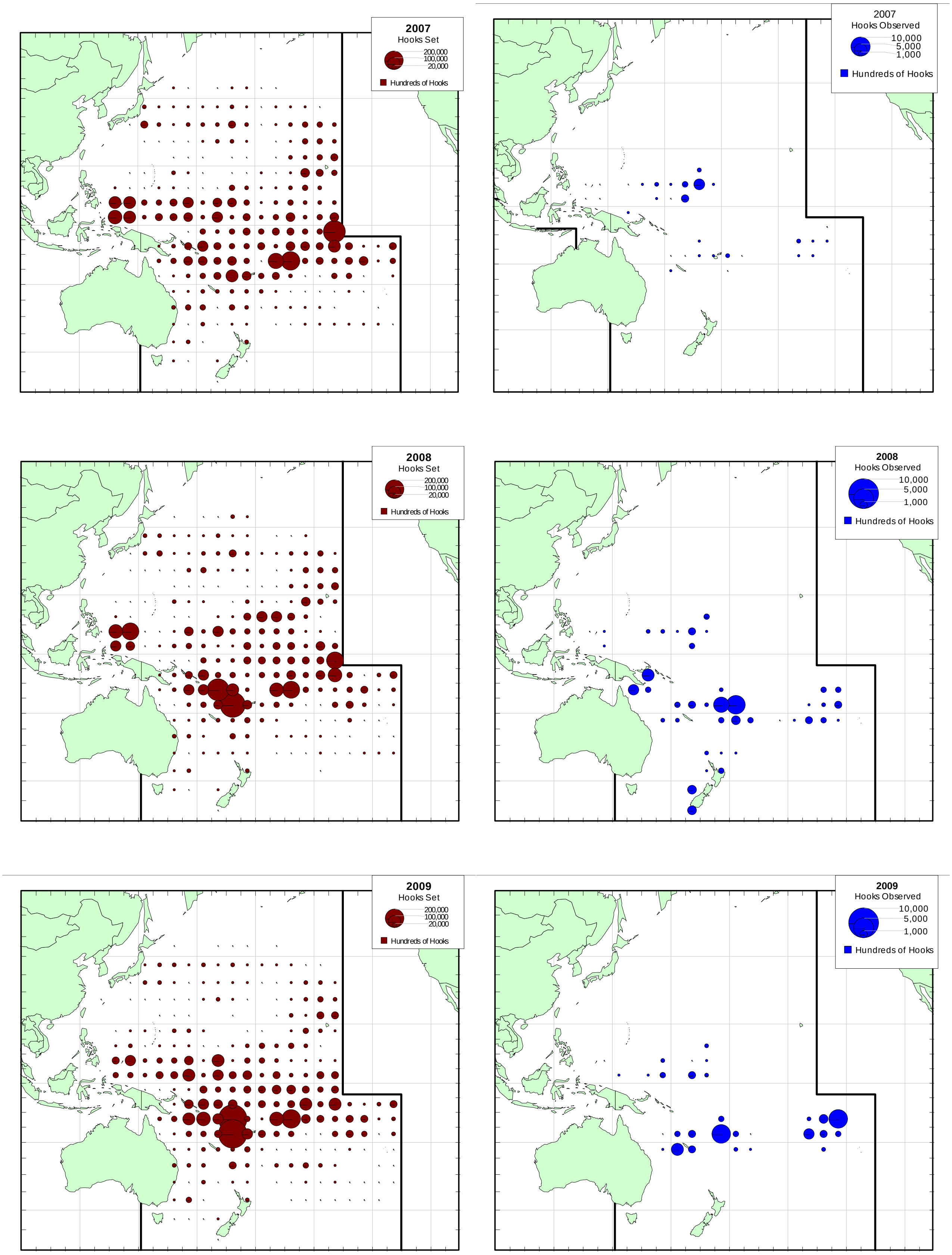


Figure A2. Distribution of purse-seine days fished (left) and days observed (right) in the Western and Central Pacific Ocean, excluding the domestic fleets of Indonesia and the Philippines

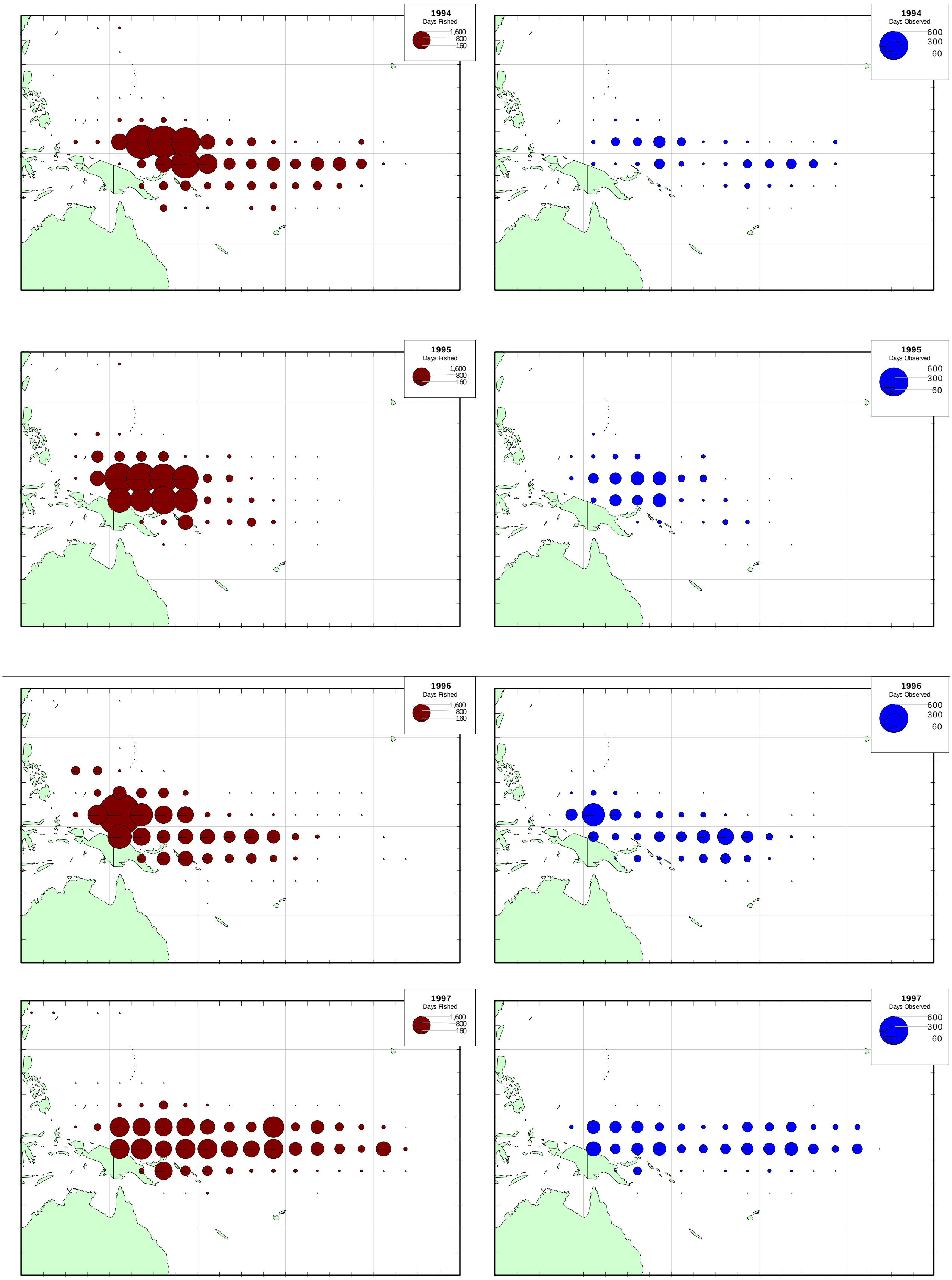


Figure A2 (continued)

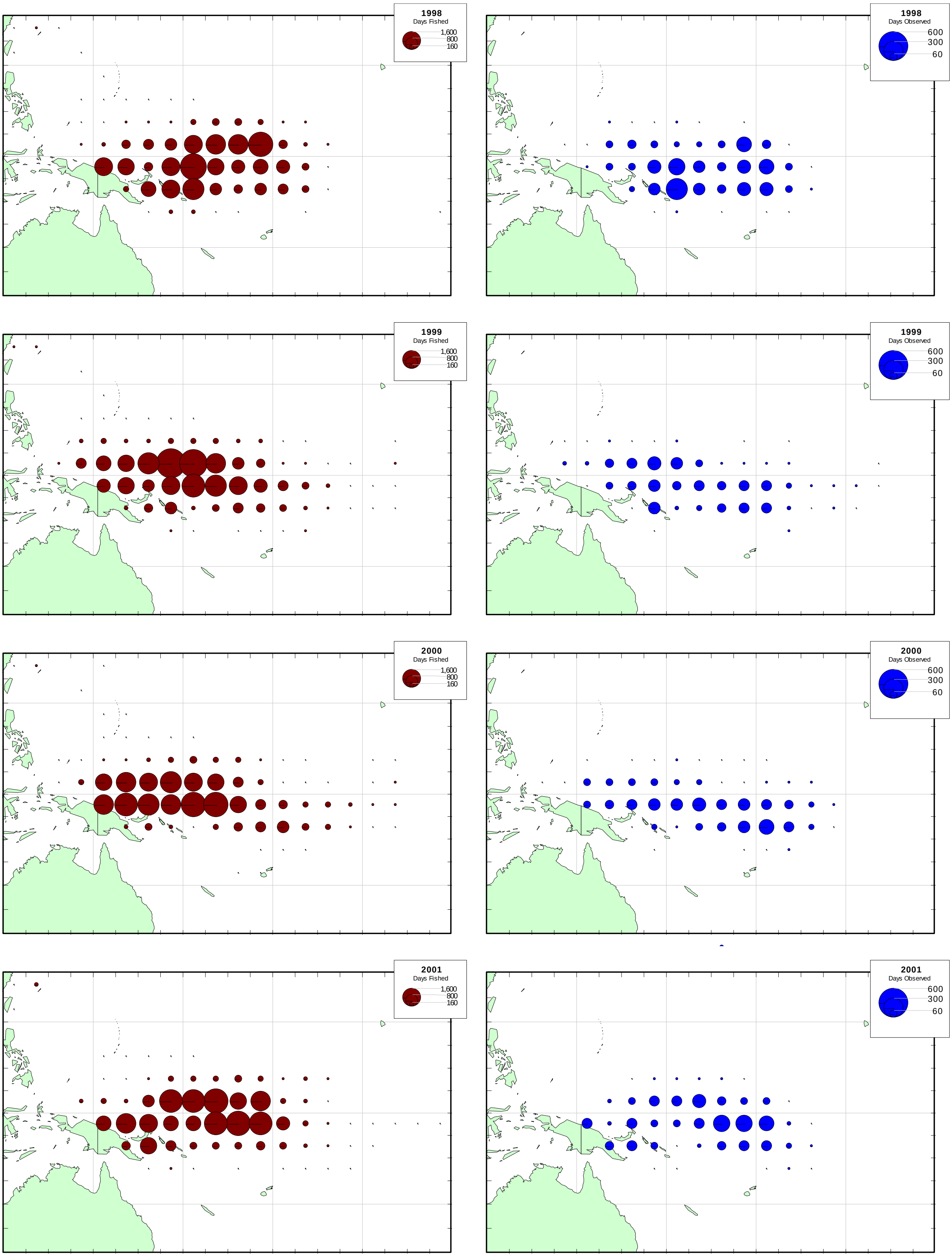


Figure A2 (continued)

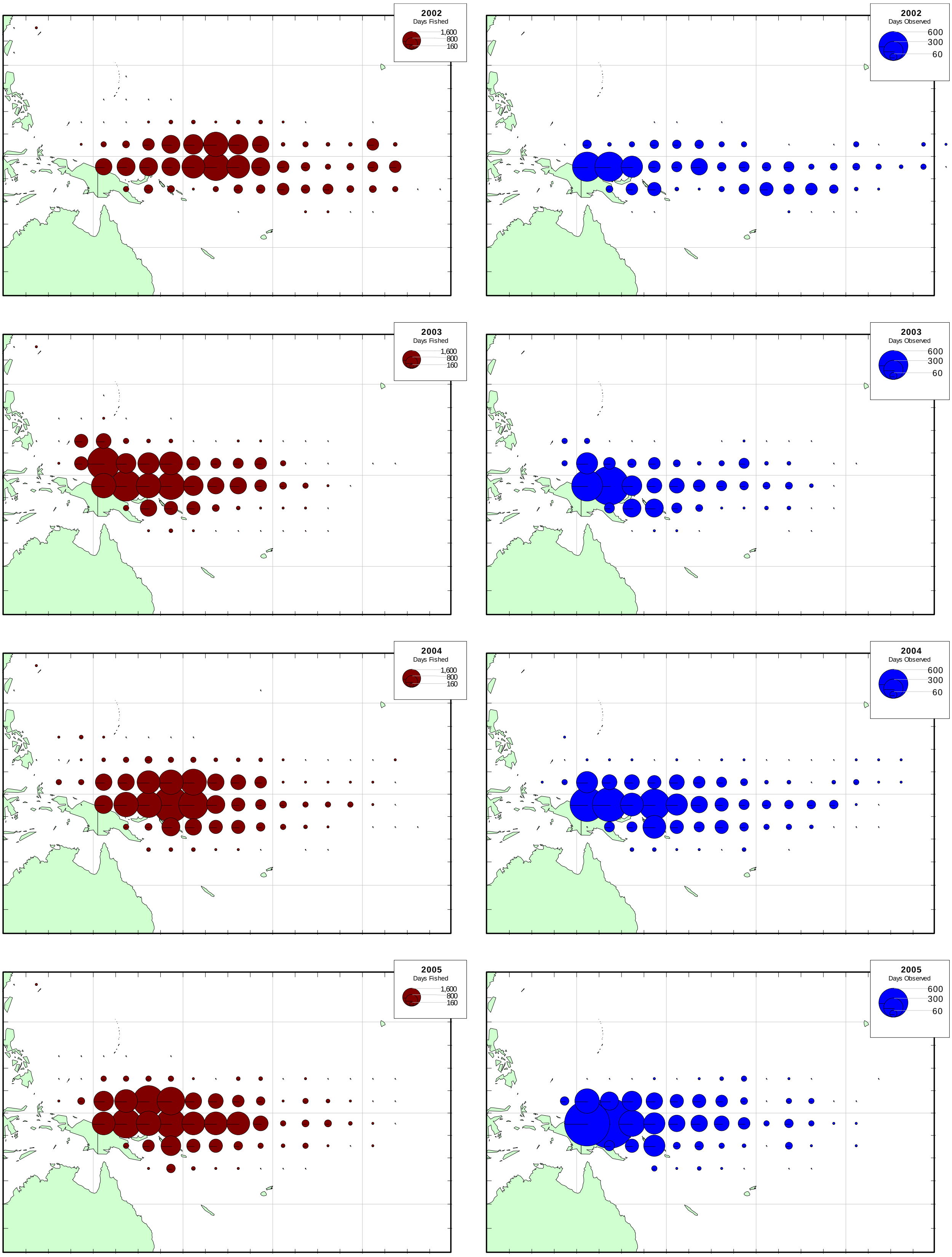


Figure A2 (continued)

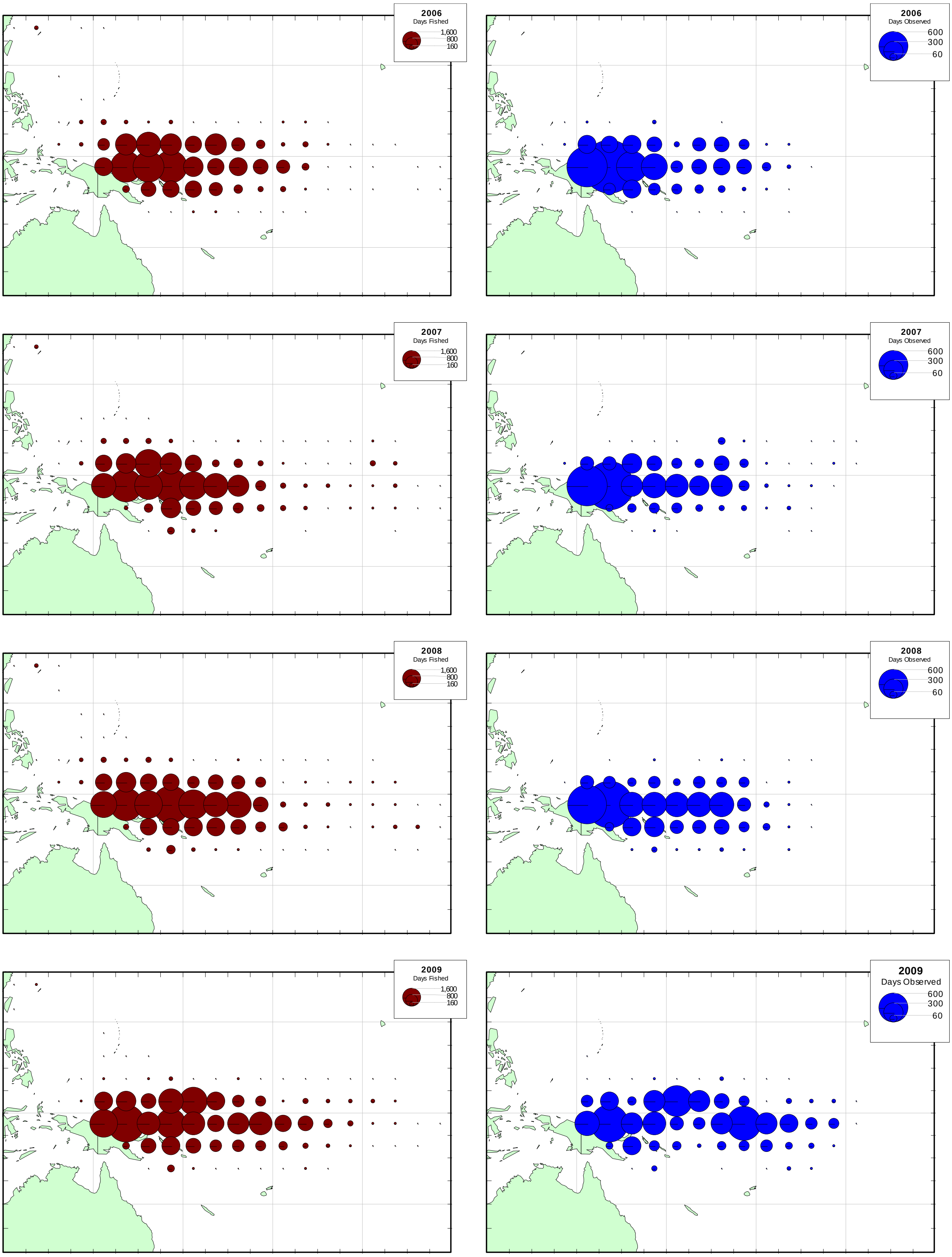
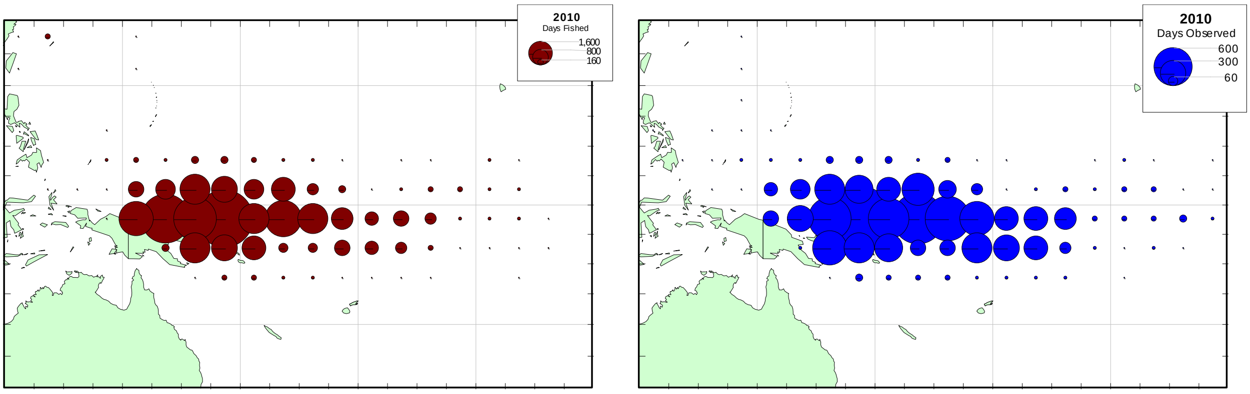


Figure A2 (continued)



APPENDIX. NOTES ON THE USE OF BASIS FUNCTIONS

Calculation of Basis Functions

The piecewise polynomial at a value *x* of a covariate is represented by

$$f(x) = \alpha_0 + \sum_{j=1}^{DF} \alpha_j B_j(x) . \qquad \text{(A1)}$$

where *f(x)* is the effect on the response variable of a value *x* of the covariate; *DF* is the degrees of freedom, which is equal to the number of knots plus the degree of the polynomial, *k* + *d*; $B_j(x)$ are the values of the basis functions and α_0 and α_j are the regression coefficients estimated in the DLN models.

For a spline of degree *d* with *k* knots (i.e., *k* + 1 quantiles), there are *k* + *d* basis functions if an intercept is excluded in the spline (and *k* + *d* + *1* basis functions if it is included). The *i* th basis function $B_i(x)$ of a cubic spline (*d* = 3) is defined recursively with de Boor's algorithm as follows:

$$B_i(x) = N_{i,d}(x), \; i = 1, k + d \qquad \text{(A2)}$$

where

$$N_{i,0}(x) = I(x_i \le x < x_{i+1}) \qquad \text{(A3)}$$

$$N_{i,j}(x) = \frac{x - x_i}{x_{i+j} - x_i} N_{i,j-1}(x) + \frac{x_{i+j+1} - x}{x_{i+j+1} - x_{i+1}} N_{i+1,j-1}(x) \qquad \text{(A4)}$$

where x_i and x_{i+1} are the *knots* defining the range of the *i* th quantile. First, the $N_{i,0}$ are calculated as either 0 or 1 depending on whether the value of *x* lies in the *i* th quantile. Then the $N_{i,1}$ are calculated. Then the $N_{i,2}$. When using cubic splines, the values of the basis functions at *x*, $B_i(x)$, are the $N_{i,3}$.

For example, a call to the **bs** function in the **splines** package in R, which generates the B-spline[5] basis matrix for a polynomial spline, with *df* = 7 and *intercept* = FALSE will result in a matrix with **length**(*x*) rows, i.e., one for each value of x, and seven columns, one for each basis function value.

---

[5] Short for "basis spline"

This is because **bs** *anchors* the B-spline basis by adding d + 1 *lower boundary knots* and d + 1 *upper boundary knots*, where the lower and upper boundary knots are equal to the minimum and maximum values of x respectively (unless specified otherwise). In a cubic spline without an intercept, a call to **bs** with df = 7 implies 7 – d = 4 *inner* knots. With an additional d + 1 = 4 lower boundary knots and d + 1 = 4 upper boundary knots, there are a total of 3 + 4 + 4 = 12 knots. When starting with 12 knots, it can be shown that there are 12 – (d + 1) = 8 functions of $N_{i,3}$, and when there is no intercept, **bs** deletes $N_{1,3}$, leaving 7 basis functions. The formula for determining the number of basis functions is therefore df – d + (d + 1) * 2 – (d + 1) – 1, which is equal to df.

Predictions with DLN models using splines

Predictions with DLN models using splines must take account of the fact that the values of the basis functions for a spline of a particular covariate depend on the entire set of values of the variable (see Chambers & Hastie 1992, pages 108, 241 and 288). This is because the basis functions depend on the knots, i.e., the values of x that define the ranges of the quantiles. If one set of values of x is used to determine the basis functions when fitting the DLN model and another set is used when predicting values of CPUE with the parameter estimates from the fitted model, the predicted values of CPUE will not make sense.

A common approach to prediction with splines is to combine the set of values of x used to fit the model and the set of values used to predict with the model, and then determine the basis functions using the combined set of data (see Chambers & Hastie 1992, page 289). However, this method will not be appropriate if the number of values used for predictions is large and they are not similarly distributed to the values used to fit the model, since the knots for the combined set of values may be quite different from the set of values used to fit the model.

A better approach is to proceed as before, first determining the basis functions using only the set of values used to fit the DLN model. Then, when predicting values of CPUE with the parameter estimates from the fitted DLN model, the basis functions are determined in one of three ways. First, if the value of x used for a prediction was also used for fitting the model, then the basis functions are simply those used when fitting the model and, thus, are already available. Second, if the value of x was not used for fitting the model, then the basis functions are determined from the **bs** object in R using the **predict** function; that is, the basis functions used for prediction are determined by interpolation of those used to fit the model. Finally, if a value of x used for prediction is beyond the range of values used to fit the model, then, usually, the basis functions are set to those of the minimum or maximum values used to fit the model, as appropriate, rather than using the **predict** function. For the multivariate *lat_lon* surface, however, it was found that more reasonable results were obtained by using **predict** for *all* values of latitude and longitude not used to fit the model; this is because the *lat_lon* effect on CPUE is generally well behaved, with the contours of CPUE in latitudes and longitudes beyond the range of those covered by the observer data being consistent with those that were.