Estimation of Standardized CPUE Series of Pacific Bluefin Tuna for Taiwanese Longline Fishery under Incomplete Data

Shui-Kai Chang[1], Hung-I Liu[2], and Yu-Wen Fu[1]

[1]*National Sun Yat-sen University, Kaohsiung, Taiwan*
[2]*Overseas Fisheries Development Council, ROC, Taipei, Taiwan*

November 2015

Information paper submitted to the ISC Pacific bluefin tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), 18-25 November 2015, Kaohsiung, Taiwan.

Summary

PBF was a seasonal target species to Taiwan offshore longline fishery. Since 2010, catch information of date and location and size information of length and weight, of each PBF could be obtained from a catch documentation scheme (CDS). Before that year, however, only market landing data with small coverage of logbooks were available. Therefore, several non-traditional procedures were performed to estimate standardized PBF CPUE series for 2001-2014. (1) Estimating PBF catch in number from landing weight for 2001-2003 of which years the information was not available, based on an MCMC simulation; (2) Deriving fishing days for 2007-2009 from data of vessel monitoring system (VMS) and voyage data recorder (VDR) based on a new developed algorithm; (3) Deriving fishing days for 2001-2006 from vessels trip information based on linear relationships between fishing days and at-sea days for a trip, by vessel size and fishing port, during 2007-2014; (4) Standardizing the CPUE for 2001-2014 using generalized linear models with delta lognormal and zero-inflated negative binomial assumptions. Results of both models showed a declining trend from 2001 to 2010 with annual fluctuations and starting to increase since 2013 after two years' low status.

Introduction

Pacific bluefin tuna (PBF) is an import seasonal target species for Taiwanese longline fishery. The catch has been as high as 3089 mt in 1999 but was continuously declined to the lowest record of 213 mt in 2012. Recently the catch has shown increasing sign to be 483 mt in 2014 and 492 mt (preliminary value) in 2015 (Fig. 1). The catch was composed mainly of 150-200 kg median size fish (>60%) in the early 2000s, but following the decrease of available median size fish, large fish of >200kg became the majority. Recently, however, more median size fish was observed in the catch and its ratio has reached 50% in 2015.

To enhance the management on PBF fishery, Taiwan implemented specific regulations on the fishery since 2010. Vessels intending to fish PBF are required to obtain a PBF fishing license from the authority in January. Vessels larger than 20 GRT are also required to install a functional vessel monitoring system (VMS). All the PBF vessels have to join the catch documentation scheme (CDS): when caught PBF, skipper has to attach a tag to each of the fish and report the information of catching date, location, tag number and weight estimate, to nearby fishery radio station; after returned port, skipper has to acquire a CDS document for his PBF catch; while landing, port inspector will examine the tags and CDS documents and measure length and weight of the fish.

With implementation of the CDS, detail catch information on each PBF were available since 2010. However, for the years before 2010, in addition to low coverage of logbooks, only daily market landing data since 2001 by vessel and by fishing port were available; the effort data that are needed for calculation of catch per unit of effort (CPUE) were very incomplete. Therefore, alternative procedures are necessary to obtain a standardized PBF CPUE series for Taiwanese longline fishery. This study presents several non-traditional procedures taking advantage of other sources of information to tackle the incomplete data situation: (1) Estimating PBF catch in number from landing weight for 2001-2003 of which years the information was incomplete, based on an MCMC simulation; (2) Deriving fishing days for 2007-2009 from data of vessel monitoring system (VMS) and voyage data recorder (VDR) based on a new developed algorithm; (3) Deriving fishing days for 2001-2006 from vessels trip information based on linear relationships between fishing days and at-sea days for a trip, by vessel size and fishing port, during 2007-2014; (4) Standardizing the CPUE for 2001-2014 using generalized linear models (GLMs) with delta lognormal and zero-inflated negative binomial assumptions.

Materials and Methods

The data

Six types of data were used in this study to estimate PBF CPUE for Taiwanese longline fishery (Table 1). The first type of data was a complete set of CDS data with detail information on PBF catch during 2010-2014. The second one was a series of market landing data with catch in weight and number of PBF by vessels, date and fishing port. PBF catch was all landed in the three major fishing ports (markets) of Taiwan, namely (in the order of landing amount) Tungkang, Suao, and Singang (Fig. 2). In terms of PBF catch, the CDS data was consistent with the market landing data and therefore the landing data with longer time series was served to provide basic catch information of PBF for calculation of CPUE, except that the data of number of fish landed for 2001-2003 was incomplete and needed to be estimated from landing weight. The third type of data was logbooks; however, the coverage was very low (<5%) until 2010 when the CDS was implemented. Logbook data was used only for verification purpose that was introduced in later section.

The rest three types of data were used for estimation of fishing effort. According to logbooks, the number of hooks deployed per day for registered PBF longliners were different among vessel size categories (CT1 for 5 to less than 10 GRT, CT2 for 10 to less than 20 GRT, CT3 for 20 to less than 50 GRT, and CT4 for 50 to less than 100 GRT), but were similar within category (mean ± SD = 189 ± 59, 706 ± 181, 851 ± 287, 972 ± 353, for CT1 – CT4, respectively; cv = 26 – 36%). Considering the difficulties in estimating hooks information under incomplete data situation, this study used fishing days as indicator of fishing effort. Number of hooks could be easily calculated from the above information by vessel size, if necessary, but will introduce additional variations. The forth type of data was vessel trips data containing records of vessels leaving and entering the major fishing ports in Taiwan, by vessel and by port, during 1993-2014. The data was collected by the Coast Guard of Taiwan for security purpose under the special situation between Taiwan and China (Chang, 2014). Number of days at sea of a vessel could be calculated from the data. The fifth one was VDR data, containing per 3-min information of vessel position, speed and direction information for 2007-2014. The data is originally for offshore/coastal vessels to apply for oil subsidy from government based on the distance they travelled at sea, and so the coverage was high. The last one was VMS data, containing per 1 – 6 hours records of vessel position, speed and direction for 2007-2014. Number of installation by PBF vessels was low before 2010. These two datasets were combined as hourly VDR/VMS data to make the data more complete for this study. Actual fishing days could be calculated from CDS data for 2010 onwards. Fishing days for 2007-2009 could be estimated from the VDR/VMS data based on some algorithms verified by CDS and logbook data of 2010-2014. A relationship could be further established between estimated fishing days and at-sea days by vessel size and by port based to be applied for estimation of fishing days from vessel trips data for 2001-2006.

The methods

1. Estimating fish numbers from Markov Chain Monte Carlo (MCMC) simulation for 2001-2003

   For 2001-2003, only 20% - 36% landed number of fish (n) were recorded. All landing weights of PBF by vessel by day were available with maximum record of 8,428 kg, corresponding to about 30 fish from the average weight of 2004-2006. According to data of 2004 - 2006, weight of singe PBF was in the range of 80 – 350 kg with a normal distribution shape. Therefore, the study calculated the annual mean and SD and constructed a normal distribution for that year assuming that the weight record in the above range was for a single fish. A set of 10,000 accumulated weight samples was then randomly drawn from the distribution for each of the n = 1 to 50 conditions, e.g., for n = 3 condition, three weights were drawn from the distribution and the sum-up weight was treated as one of the 10,000 weight samples. Totally 500,000 simulated weights were obtained for all conditions. For the records without information of *n*, another distribution for *n* for each landing weight record by vessel and day could then be obtained by comparing the weight with the simulated weights data; and the final estimate of *n* was randomly selected from the distribution.

2. Deriving fishing days from VDR/VMS data for 2007-2009

High-tech data like VMS data with vessel positions, speeds and directions information have been used to derive fishing efforts for many fisheries (Lee et al., 2010). Chang and Yuan (2014) recommended two approaches to derive fishing days from VMS data of Taiwanese distant water longliners: the first one, the optimal-speed-time-ranges approach, was based on vessel speed in the afternoon (hook retrieval period) and the second one, the within-day distance approach, was based on vessel moving distance within a day. Almost all PBF was caught by offshore longliners which are more motive and may have different operation pattern than distant water longliners. Therefore, this study tested these two approaches following similar procedures in Chang and Yuan (2014) on the VDR/VMS data of selected PBF vessels that have corresponding logbooks and CDS for follow-up examination of performance measures. In addition, by experience, offshore longliners have clear pattern of change direction after completion of hook deployment operation preparing for hook retrieval operation. If the vessel is move heading to a location the change of direction in a day may be small or close to 0 (i.e., straight line). Therefore, this study developed and tested a third approach based on change of vessel direction.

(1) Vessel-speed approach: A pre-test on the optimal-speed-time-ranges approach suggested that PBF vessels, unlike the distant water longliners, did not have clear pattern of deploying hooks in the morning and retrieving them in the afternoon. However, the speed was still comparatively lower during retrieving operation for handling the catch. Therefore, this study renamed the approach as vessel-speed approach and tested the performances of the following criteria: a day with at least an instance of vessel speed at *x* knots is defined as a fishing day; *x* is 1 – 7 knots per one knot.

(2) Within-day-distance approach: Longliners move in shorter distance in a day while conducting fishing operation. The study tested the performance of criteria that defining a day as a fishing day if the within-day-distance if below *x* km; *x* is 70 – 190 km per 20 km.

(3) Direction-change approach: Change of vessel direction has been observed in fishing operation while the vessel completed the hook deployment and returning to either the start or the end positions of deployment after 2 – 4 hours preparing for retrieval operation. With this observation, the study tested the performance of criteria that defining a day as a fishing day if the angle of direction change is within *x* degree; *x* is 60° – 120° degree per 30°. In the test, the VDR/VMS data was firstly aggregated into one record per 2, 3, and 4 hours to show clear trend of direction change.

The performance of the criterion was assessed based on the ability to maximize agreement between the predicted fishing-day/non-fishing-day distribution from the VDR/VMS data and the observed distribution from available actual fishing days information from logbooks and CDS (Chang and Yuan, 2014). To have more complete data for assess performance, the tests were performed only on data since 2010. Elements of the confusion matrix were denoted as true positive (TP), false negative (FN), false positive (FP), and true negative (TN). The sensitivity of the criterion (or the true positive rate, TPR = TP/(TP + FN)) was measured as the ability to accurately predict fishing days; and, the specificity (or the true negative rate, TNR = TN/(TN + FP)) was measured as the ability to accurately predict non-fishing days (Fawcett 2006; Fukuda et al. 2011). Criterion that maximized the sum (SSS), or minimized the absolute difference value (DSS) when the SSS is similar to others, of sensitivity and specificity was considered optimal (Liu et al. 2005; Jiménez-Valverde and Lobo 2007; Chang and Yuan, 2014).

The selected final criterion was then applied to the VDR/VMS data of 2007-2009 to derive fishing days for each PBF vessel.

3. Deriving fishing days from vessels trip information for 2001 – 2006

The vessel trips data from the Coast Guard could be used to calculate at-sea days of each trip for each PBF vessel. A data exploration analysis on the data listed in Table 1 for the years after 2010 suggested that there were linear relationships between vessels' at-sea days (independent variable) and fishing days (dependent variable), by vessel size and by fishing ports. Owing to the gradual relaxation on port control during recent years, the at-sea days might be over-estimated due to lacks of interim port entrance records for some vessels. Therefore, the vessel trips data needed to be screened in advance before establishing the relationships from the data of 2007 – 2014. The linear relationships were constructed by vessel size (CT1 – CT4) and by fishing port (three ports) and were then applied to the trip data of 2001 – 2006 to estimate fishing days.

4. Standardizing CPUE for 2001 – 2014

Data for CPUE analyses were processed in advance to be in trip basis because there was no daily fishing information for data before 2007. Only data of the registered PBF vessels and in the major fishing season of May to July was used. A trip covered two months was allocated to (1) May if it covered April and May; (2) July if it covered July and August; (3) the specific month that have more fishing days occurred. Vessels with annual PBF catch less than 5 were excluded from that year data to avoid noise effect.

There were two major PBF fishing grounds with different size of fish for Taiwanese offshore longliners which could be split by the line of 24.3° (Fig. 3). Average size of PBF catch (2010 – 2014) in the northern fishing ground was 235 kg, about 25 kg smaller than the southern fishing ground. The northern ground was exploited almost by vessels leaving from Suao fishing port, while the southern ground was by vessels from Singang and Tungkang. Therefore, although there was no fishing location information for 2001 – 2006, separation of catch could be made based on leaving port of the vessels in the vessel trips data.

Covariates considered in the GLM model include: year (2001 – 2014), month (May – July), fishing area (northern and southern fishing ground), and vessel tonnage (CT1 – CT4). Two model assumptions were investigated to address the high percentage of zero catch in the data: delta lognormal assumption and zero-inflated negative binomial (ZINB) assumption. Descriptions and discussions on the two assumptions in the GLM could be found in MacNeil et al. (2009) and Brodziak and Walsh (2013). Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used for final model selection.

Results and Discussions

1. Estimating fish numbers from MCMC simulation for 2001 – 2003

The MCMC simulation has been applied to 2004 – 2006 first to examine the consistency of the result with the recorded fish number data (Fig. 4). Although the simulated size tended to be smaller than the actual size in the two tails, they were considered consistent. The method was then applied to the 2001 – 2003 data for estimation of catch in number.

2. Deriving fishing days from VDR/VMS data for 2007-2009

The performance statistics of the criteria for deriving fishing days applied to 2010 – 2014 data (Table 2) shows that an instance of 2 knots in the data was the optimal criterion for the vessel-speed approach, a daily movement distance of 150 km was the optimal choice for the within-day-distance approach, and a change of vessel direction of 100° occurred in a day for per 3-hour data was the optimal choice for the direction-change approach. Among them, the direction-change approach has the highest SSS and lowest DSS and therefore was recommended to be used as the optimal criterion for deriving fishing days from VDR/VMS data.

Fig. 5 shows navigating tracks of a vessel and the part that was classified as fishing. The non-fishing period was apparently occurred when the vessel navigating to and returning from the fishing ground. Fig. 6 presents the distributions of at-sea days and fishing days that classified by the optimal direction-change criterion, of PBF vessels for 2007 as an exmaple. The fishing-day classification has removed a lot of at-sea days not in the fishing ground.

3. Deriving fishing days from vessels trip information for 2001 – 2006

Linear relationships between the at-sea days calculated from vessel trips data and the fishing days estimated from VDR/VMS or CDS data were established by vessel size and by port for 2007 – 2014 (Table 3). The relationships were all statistically significant at 1% level with $R^2 > 90\%$. The coefficients were applied to the vessel trips data of 2001 – 2006 to estimate the fishing days from at-sea days by the category.

4. Standardizing CPUE for 2001 – 2014

From the above works, CPUE on trip basis were calculated for the whole series of 2001 – 2014. Performing GLM on the data, the best explanatory variable combinations under delta-lognormal assumption were year, month, fishing area, vessel tonnage and interaction between year and month for positive catch model and year, month and fishing area for proportion model. The diagnostic residual plots for this GLM run in Fig. 7 indicated the appropriateness of the two-stage delta lognormal model for evaluation of the factors that influence the PBF catch rate.

The GLM with ZINB assumption could not converge under the full set of covariates. The model could converge when the factor of vessel tonnage was removed. The best explanatory variable combination was year, month and fishing area, for both the count model and zero inflation model. Diagnostic plots for this run in Fig. 8 also indicated the appropriateness of this model for standardization of the PBF catch rate. However, due to vessel size was considered as an important factor influencing PBF CPUE but could not be properly considered in this model, the study recommended to take the delta lognormal model as the final model, although the two standardized CPUE series were very similar (Fig. 9).

References

Brodziak, J., and W. A. Walsh. 2013. Model selection and multimodel inference for standardizing catch rates of bycatch species: a case study of oceanic whitetip shark in the Hawaii-based longline fishery. *Can. J. Fish. Aquat. Sci.* 70: 1723-1740.

Chang, S.-K. 2014. Constructing logbook-like statistics for coastal fisheries using coastal surveillance radar and fish market data. *Marine Policy*, 43: 338–346.

Chang, S.-K., and T.-L. Yuan. 2014. Deriving high-resolution spatiotemporal fishing effort of large-scale longline fishery from vessel monitoring system (VMS) data and validated by observer data. *Can. J. Fish. Aquat. Sci.* 71: 1363-1370.

Jiménez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecol.* 31: 361-369.

Lee, J., A. B. South, and S. Jennings. 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES J. Mar. Sci.* 67: 1260-1271.

Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385-393.

MacNeil, M. A., J. K. Carlson, and L. R. Beerkircher. 2009. Shark depredation rates in pelagic longline fisheries: a case study from the Northwest Atlantic. *ICES J. Mar. Sci.* 66: 708-719.

Table 1. Types of data that were used in this study for estimation of PBF CPUE, with the available time periods and the remarks on data contents and limitations,

Data type	Period	Remarks
CDS data	2010-2014	Include complete information on catching date, position, length and weight of each PBF and supplemental information on vessels and auctions.
Market landing data	2001-2014	Include landing weight and number of PBF by vessel, landing date and fishing port (fish market). Fish number information was not available for 2001-2003.
Logbook data	2001~2014	Include information of operation date and location, fishing effort and catch by species. The coverage was very low until 2010.
Vessel trips data	1993-2014	Contain records of vessels leaving and entering the major fishing ports in Taiwan, by vessel and by port. The data was collected by the Coast Guard of Taiwan for security purpose.
Voyage Data Recorder (VDR) data	2007~2014	Contain per 3-min records of vessel position, speed and direction. The data is originally for offshore/coastal vessels to apply for oil subsidy from government.
Vessel Monitoring System (VMS) data	2007~2014	Contain per 1 – 6 hour records of vessel position, speed and direction. Number of installation was low before 2010.

Table 2. Performance statistics of fishing-day versus non-fishing-day binary classification criteria, based on (A) vessel-speed approach, (B) within-day-distance approach, and (C) direction-change approach. The optimal criterion of each approach was in bold.

		TP	FN	FP	TN	TPR	TNR	SSS	DSS
A. Vessel-speed approach									
1 knot		411	44	235	46	0.90	0.16	1.07	0.74
2 knots		**441**	**14**	**245**	**36**	**0.97**	**0.13**	**1.10**	**0.84**
3 knots		444	11	250	31	0.98	0.11	1.09	0.87
4 knots		450	5	252	29	0.99	0.10	1.09	0.89
5 knots		452	3	265	16	0.99	0.06	1.05	0.94
6 knots		455	0	277	4	1.00	0.01	1.01	0.99
7 knots		0	278	3	0	0.00	0.00	0.00	0.00
B. Within-day-distance approach									
70 km		308	146	136	143	0.68	0.51	1.19	0.17
90 km		350	104	146	133	0.77	0.48	1.25	0.29
110 km		374	80	161	118	0.82	0.42	1.25	0.40
130 km		401	53	172	107	0.88	0.38	1.27	0.50
150 km		**419**	**35**	**178**	**101**	**0.92**	**0.36**	**1.28**	**0.56**
170 km		428	26	191	88	0.94	0.32	1.26	0.63
190 km		434	20	203	76	0.96	0.27	1.23	0.68
C. Direction-change approach									
60°	2h	441	14	104	175	0.97	0.63	1.60	0.34
	3h	446	8	96	182	0.98	0.65	1.64	0.33
	4h	439	15	97	181	0.97	0.65	1.62	0.32
90°	2h	427	28	77	202	0.94	0.72	1.66	0.21
	3h	431	23	75	203	0.95	0.73	1.68	0.22
	4h	430	24	76	202	0.95	0.73	1.67	0.22
100°	2h	421	34	75	204	0.93	0.73	1.66	0.19
	3h	**423**	**31**	**69**	**209**	**0.93**	**0.75**	**1.68**	**0.18**
	4h	421	33	72	206	0.93	0.74	1.67	0.19
120°	2h	383	72	64	215	0.84	0.77	1.61	0.07
	3h	386	68	65	213	0.85	0.77	1.62	0.08
	4h	394	60	65	213	0.87	0.77	1.63	0.10

Note: The table shows true positive (TP), false negative (FN), false positive (FP), true negative (TN), True positive ratio (TPR), and true negative ratio (TNR), as well as their sensitivity and specificity. Sum of sensitivity and specificity (SSS) and absolute difference of sensitivity and specificity (DSS) are performance measures for the criteria.

Table 3. Statistics of linear relationships between at-sea days and fishing days established from data of 2007 – 2014. The relationships were all statistically significant at 1% level.

Ports	Vessel size	Coefficient	R^2
Suao	CT2	0.702	0.955
	CT3	0.707	0.955
	CT4	0.721	0.962
Singang	CT1	0.673	0.996
	CT2	0.759	0.947
	CT3	0.688	0.943
	CT4	0.726	0.968
Tungkang	CT1	0.634	0.922
	CT2	0.716	0.965
	CT3	0.637	0.942
	CT4	0.639	0.950

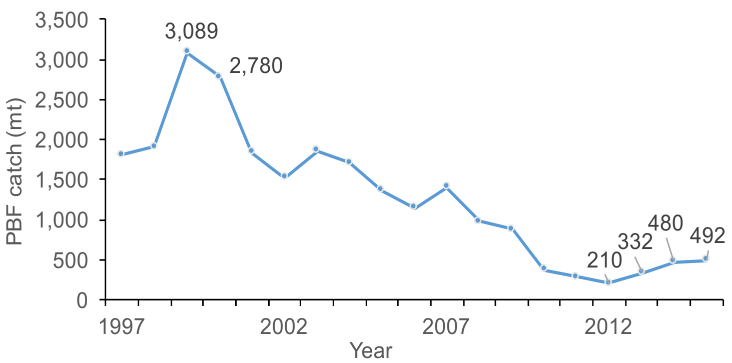

Fig. 1. PBF catch series by Taiwan longline fishery during 1997-2015. Data of 2015 is preliminary.

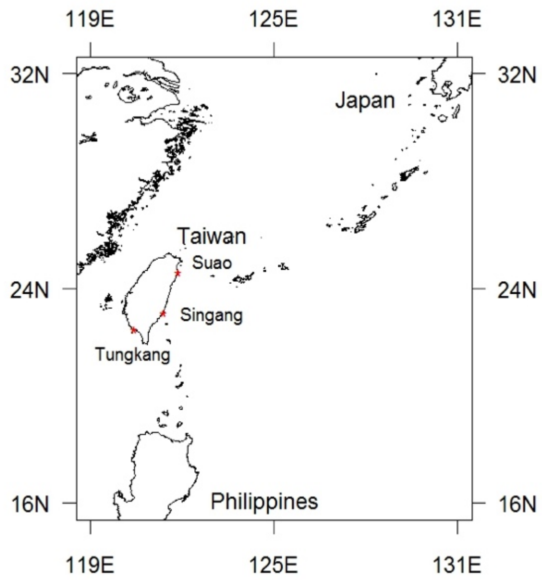

Fig. 2. Major PBF landing ports in Taiwan: Tungkang, Suao and Singang (in the order of landing amount).

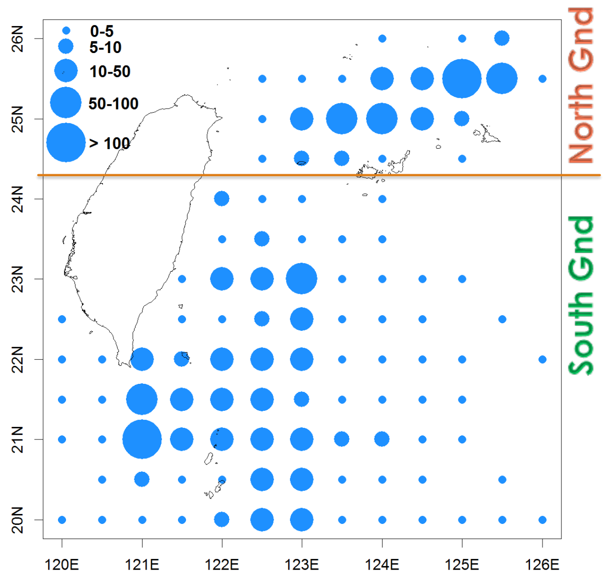

Fig. 3. Average PBF catch distribution off Taiwan for 2010 – 2015 by Taiwanese PBF longliners. The line splits the fishing grounds into southern ground and northern ground by 24.3°N.

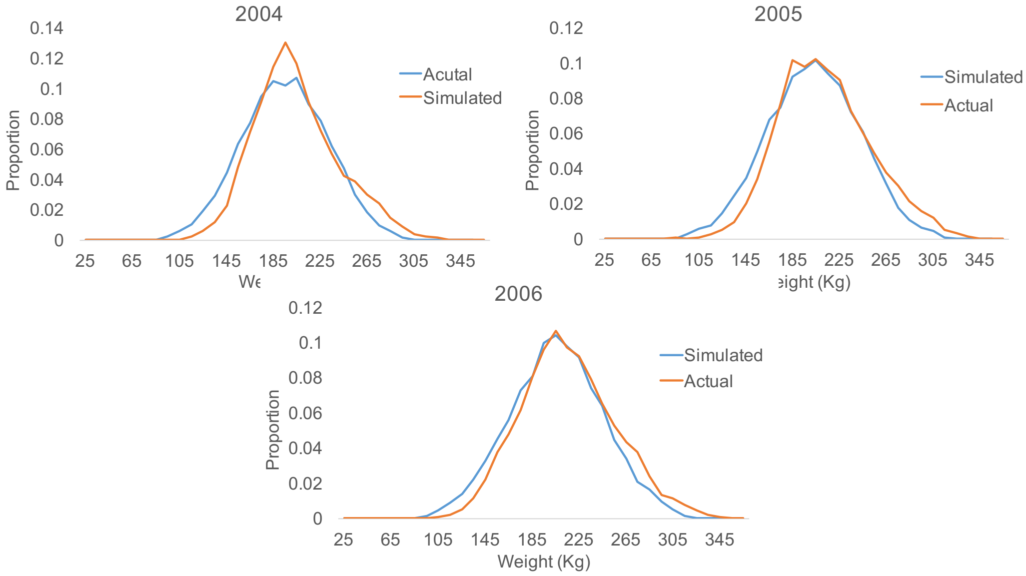

Fig. 4. PBF weight distributions calculated from actual fish numbers in market landing data and simulated fish numbers for 2004 – 2006.

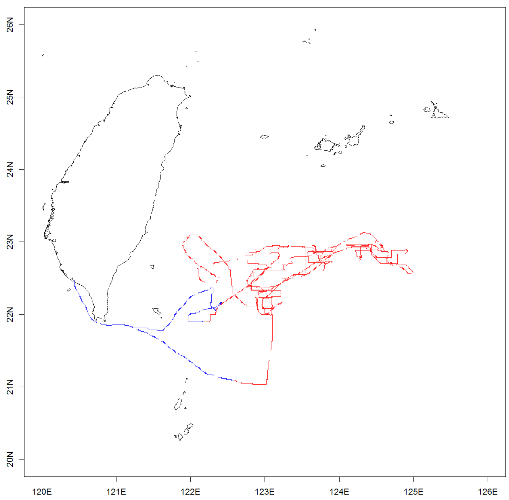

Fig. 5. Navigating tracks of a PBF vessel. Red line indicates a fishing status and blue line a non-fishing status, classified from the optimal direction-change criterion.

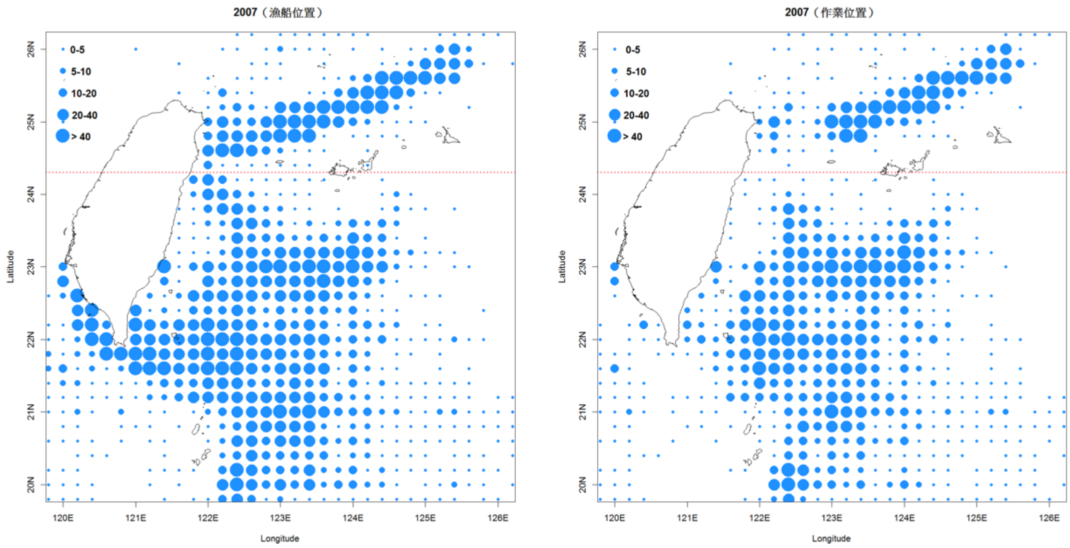

Fig. 6. Distributions of at-sea days (left panel) and fishing days (right panel) that classified by the optimal direction-change criterion, of PBF vessels for 2007.

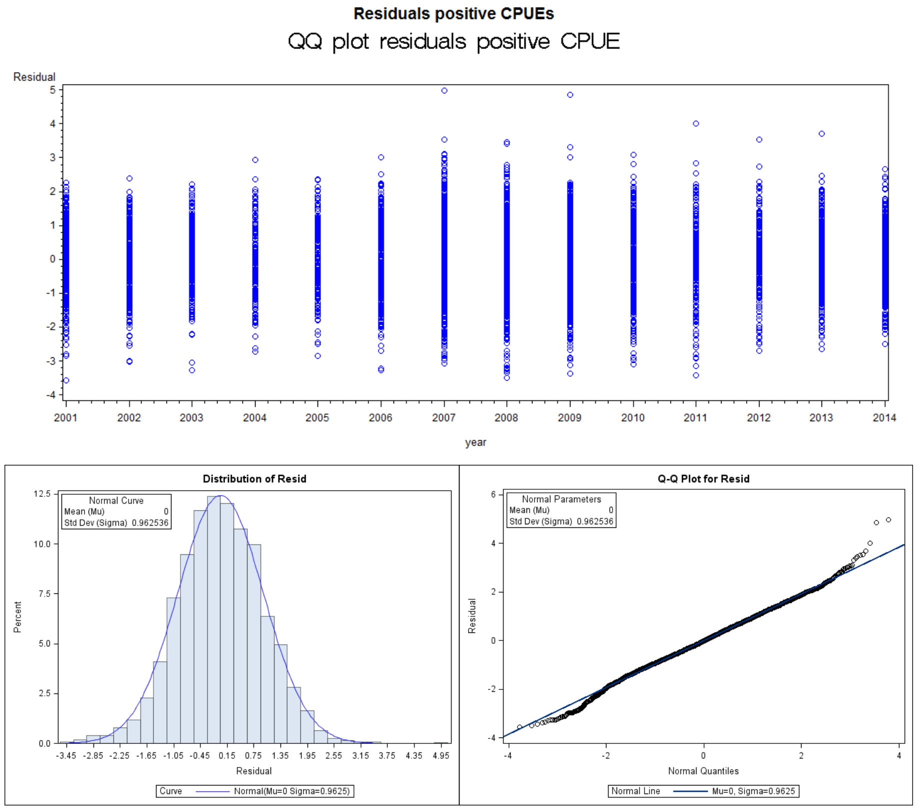

Fig. 7. Diagnostic residual plots for the GLM run with delta lognormal assumption for standardization of PBF CPUE.

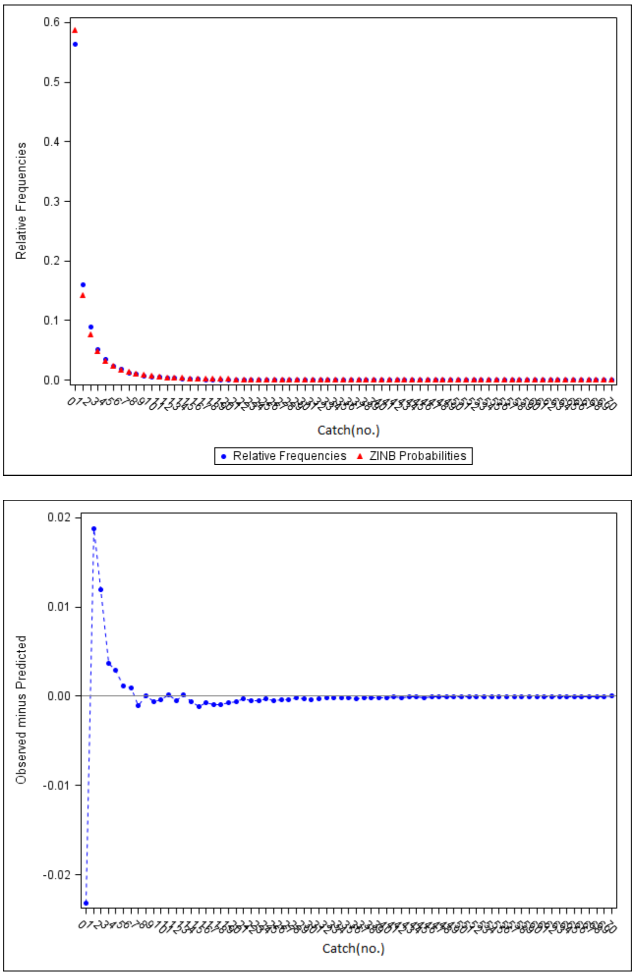

Fig. 8. Diagnostic plots for the GLM run with ZINB assumption for standardization of PBF CPUE.

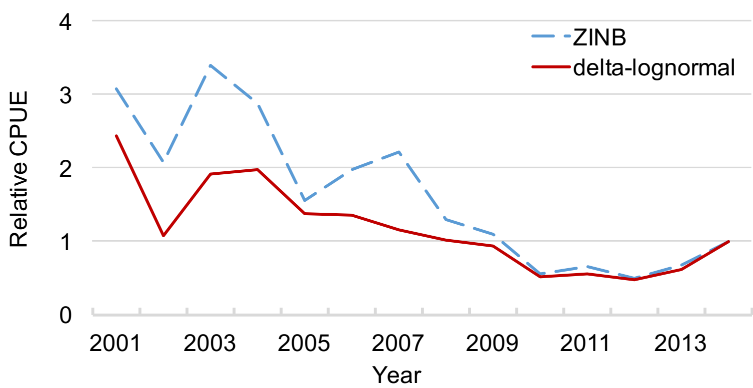

Fig. 9. Standardized CPUE series for Taiwanese PBF longline fishery by GLM with delta-lognormal and ZINB assumptions.