

## Generalized linear Bayesian models for standardizing CPUE: an application to a squid-jigging fishery in the northwest Pacific Ocean

JIE CAO<sup>1</sup>, XINJUN CHEN<sup>1, 2, 3</sup>, YONG CHEN<sup>4, 3</sup>, BILIN LIU<sup>1, 2, 3</sup>,  
JIN MA<sup>1</sup> and SILIANG LI<sup>1</sup>

<sup>1</sup> College of Marine Sciences, Shanghai Ocean University, 999 Hucheng Ring Road, Shanghai 201306, China.  
E-mail: xjchen@shou.edu.cn

<sup>2</sup> The Key Laboratory of Shanghai Education Commission for Oceanic Fisheries Resources Exploitation, 999 Hucheng Ring Road, Shanghai 201306, China.

<sup>3</sup> The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, 999 Hucheng Ring Road, Shanghai 201306, China.

<sup>4</sup> School of Marine Sciences, University of Maine, Orono, Maine 04469, USA.

**SUMMARY:** Generalized linear Bayesian (GLBM) non-hierarchical and hierarchical models were developed for standardization of catch per unit effort (CPUE). The GLBM containing the covariates of month, latitude, sea surface temperature (SST), sea surface salinity (SSS) and sea level height (SLH) had the best fit for the Chinese squid-jigging fishery of *Ommastrephes bartramii* in the northwest Pacific Ocean based on deviance information criteria. This best-fitting model tends to be more ecologically sound than other CPUE standardization models, such as generalized linear models and generalized additive models. GLBM was also used to deal with the problems of estimating stock abundance index (i.e. standardized CPUE) resulting from increased spatial heterogeneity of spatial dynamics of fishing efforts in the squid fishery by predicting the standardized CPUE for unfished areas. The standardized CPUE based on data including predicted CPUE of unfished areas was lower than the derived CPUE based on data with observed CPUE alone, in particular during the fishing peak of August to October. This study indicates that it is more appropriate to use the standardized CPUE derived from data including both predicted CPUE of unfished areas and observed CPUE of fished area as a stock abundance index. We suggest that the proposed method be used in CPUE standardization to account for impacts of large spatial heterogeneity of fishing efforts in fisheries.

**Keywords:** generalized linear Bayesian models, CPUE standardization, *Ommastrephes bartramii*, Chinese squid-jigging fishery, northwest Pacific Ocean.

**RESUMEN:** MODELOS LINEALES GENERALIZADOS BAYESIANOS PARA LA ESTANDARIZACIÓN DE CPUE: APLICACIÓN A LA PESQUERÍA DE CALAMAR MEDIANTE JIGGING EN EL PACÍFICO NOROCCIDENTAL. – Se desarrollan modelos lineales generalizados bayesianos (GLBM) jerárquicos y no-jerárquicos para la estandarización de captura por unidad de esfuerzo (CPUE). El modelo GLBM seleccionado para la pesquería del calamar *Ommastrephes bartramii* mediante jigging en el Pacífico noroccidental incorporó las variables explicativas mes, latitud, temperatura superficial del mar (SST), salinidad superficial del mar (SSS) y altura del nivel del mar (SLH). La selección del modelo se basó en el Criterio de Información de la Desviación (DIC). El modelo que mejor se ajustó a los datos tiene más sentido ecológico comparado con modelos de estandarización de CPUE basado en modelos lineales generalizados y modelos aditivos generalizados. Se utilizó también el GLBM para tratar el problema de la estimación de un índice de abundancia del stock (es decir, CPUE estandarizada) frente a la elevada heterogeneidad espacial en la dinámica del esfuerzo en la pesquería del calamar mediante la predicción de la CPUE estandarizada en áreas no pescadas. La CPUE estandarizada en base a los datos que incluyen la CPUE predicha en áreas no pescadas fue inferior a la CPUE derivada en base solamente a la CPUE observada, especialmente durante el pico de pesca de Agosto a Octubre. Este estudio muestra que es más apropiado usar la CPUE estandarizada derivada de datos que incluyen al mismo tiempo la CPUE predicha de las áreas no pescadas y la CPUE observada en el área pescada como índice de abundancia del stock. Se sugiere que se use el método propuesto para la estandarización de CPUE teniendo en cuenta la gran heterogeneidad espacial del esfuerzo pesquero.

**Palabras clave:** modelos lineales generalizados bayesianos, estandarización CPUE, *Ommastrephes bartramii*, pesquería china de calamar mediante jigging, Pacífico noroccidental.

## INTRODUCTION

Catch per unit effort (CPUE) of a fishery is often used as an abundance index which is usually assumed to be proportional to the stock abundance (Hilborn and Walters, 1992). For many fish species, such as those with short life spans and weak relationships between spawning stock biomass and recruitment (Yatsu *et al.*, 1998; Ichii, 2003; Ichii and Mahapatra, 2004), it is often difficult to conduct a traditional stock assessment based only on age- or length-structured data. Thus, developing a reliable time series of abundance indices from fisheries data is critical to monitoring the population dynamics of these fish species.

Observed fisheries CPUE data are, however, influenced by many factors, in addition to fish population abundance, including spatial-temporal factors such as area and season and environmental factors such as sea surface temperature (SST) and sea surface salinity (SSS; Rodhouse, 2001). The impacts of these factors on CPUE may shift the assumed proportionality between observed CPUE and stock abundance. Thus, CPUE standardization is needed to remove the impacts of factors other than population abundance (Gavaris, 1980; Quinn and Deriso, 1999; Tian *et al.*, 2009). Many statistical models have been developed for CPUE standardization: the generalized linear model (GLM) (Nelder and Wedderburn, 1972) is the most commonly used method in CPUE standardization, and its extension the generalized linear mixed model is also commonly used in evaluating the random effect of interactions (Pinheiro and Bates, 2000). Because of the limitations of the GLM in dealing with nonlinearity between dependent variables and multiple predictive variables, the generalized additive model (GAM) model is increasingly used recently for CPUE standardization (Bellido *et al.*, 2001; Tian *et al.*, 2009). However, these approaches may not work properly for some fisheries with unique characteristics, such as the neon flying squid (*Ommastrephes bartramii*) fishery in the north Pacific Ocean.

The neon flying squid is an oceanic cephalopod widely distributed in subtropical and temperate waters of the world (Roper *et al.*, 1984). It supports a major commercial squid fishery in the northwest Pacific Ocean. Japan began the commercial fishery in 1974 after catches of the Japanese common squid (*Todarodes pacificus*) dropped sharply during the early 1970s, and was followed by Korea and the Taiwan province of China. In the late 1980s, with the development of the driftnet fishery, annual catch increased rapidly. Mainland China started a small-scale commercial neon flying squid fishery in 1994, and since then the fishery has expanded greatly (Chen *et al.*, 2008).

The North Pacific population of *O. bartramii* comprises two seasonal cohorts (autumn and winter-spring) and four stocks: (1) the central stock of the autumn cohort; (2) the east stock of the autumn cohort; (3) the west stock of the winter-spring cohort; and (4) the

central-east stock of the winter-spring cohort (Mori, 1997; Nagasawa *et al.*, 1998a, b; Yatsu *et al.*, 1998). One of the stocks for *O. bartramii*, the western stock of winter-spring cohort, is the main target of the Chinese jigging fishery (Chen *et al.*, 2007, 2008). This species plays an important role in the ecosystem of the North Pacific Ocean (Wang and Chen, 2005). Thus it is critical to monitor its population dynamics for the sustainable exploitation of this squid. However, because of its short life span and possible weak relationship between spawning stock biomass (SSB) and recruitment (Yatsu *et al.*, 1998; Ichii, 2003; Ichii and Mahapatra, 2004), it is difficult to conduct a traditional stock assessment based on age- or length-structured data. Current stock assessment methods for this squid including the depletion method (Chen *et al.*, 2008) and squid recruitment prediction (Cao *et al.*, 2009) are based on the availability of abundance indices. Therefore, developing a reliable abundance index from the fishery data is critically important for this squid fishery.

For the western stock of the winter-spring cohort of *O. bartramii*, previous studies suggest that the distribution of this squid and its fishery CPUE can be significantly influenced by many environmental variables on the fishing grounds. Chen (1999) showed that the distribution and abundance of neon flying squid in the central north Pacific Ocean were strongly affected by water temperature and salinity, with temperature having a higher predictive power for abundance. The relationships between CPUE of neon flying squid fishery and SST were established in several studies (Yatsu and Watanabe, 1996; Chen, 1997; Chen and Tian, 2005). However, the CPUE standardization for this squid fishery using the GLM model showed that environment variables (SST, SSS) had no significant impacts on CPUE (Tian *et al.*, 2009). Additionally, because of communication among fishing vessels and the use of SST information in squid searching, the monthly distribution of fishing effort is always concentrated on the highly productive areas perceived by the fishermen. Using such observed CPUE as squid abundance index may cause serious problems in monitoring the squid population. In order to overcome the problems, we explored the use of the generalized linear Bayesian model (GLBM) for the squid CPUE standardization in this study.

Bayesian statistical methods are widely used in fisheries stock assessment to provide a conceptually elegant approach for providing fishery management advice under uncertainty (McAllister and Kirkwood, 1998), but they have been rarely used for CPUE standardization. Bayesian methods are appealing as they can easily incorporate heterogeneous data and are flexible in allowing the use of new data (Zhang *et al.*, 2009). In this study we explored and developed alternative statistical methods for CPUE standardization and applied them to neon flying squid. Factors that might affect the CPUE of the neon flying squid fishery were identified. We determined whether our approach could deal

with the problems of developing an abundance index for squid in the presence of large heterogeneity in the spatial dynamics of fishing effort.

**MATERIALS AND METHODS**

**Data**

Commercial squid fisheries data on the western stock of the winter-spring cohort in the northwest Pacific Ocean west of 170°E were obtained from June to November 1995-2004 from the database of the Chinese Squid-Jigging Technology Working Group. The main fishing ground of Chinese squid-jigging vessels targeting the winter-spring cohort is shown in Figure 1. The Observer Program of Chinese squid-jigging fishery, coordinated by the Technology Working Group, covers more than 70% of Chinese squid fishing vessels in the northwestern Pacific Ocean. The fishery data include fishing date, fishing location (longitude, latitude), number of fishing vessels per day, and daily catch per fishing vessel. A spatial scale of 0.5°×0.5° (referred to as a fishing grid) was used in grouping the data. The CPUE of a fishing grid was calculated by dividing the total catch by the number of fishing vessels per day within the fishing grid. The mean observed CPUE<sub>i</sub> (metric tons per fishing day for one fishing vessel, t/d) for a fishing grid was calculated as:

$$CPUE_i = \frac{Catch_i}{Fishing\ days_i} \quad (1)$$

where *Catch<sub>i</sub>* is the sum of catch for all the fishing vessels within a fishing grid, and *Fishing days* is the sum of fishing days for all fishing vessels in the fishing grid. We chose month as the time-step in grouping observed CPUE for each fishing grid. In addition to catch and effort data, each monthly record also includes temporal information (year and month) and spatial information (latitude and longitude), and data on SST, SSS and sea level height (SLH). These environmental data were acquired from the website <http://iridl.ldeo.columbia.edu/SOURCES/CARTON-GIESE/> and grouped by the 0.5×0.5° grid for each month, consistent spatially and temporally with the above fisheries data.

**Covariates and error models**

The covariates considered in the analysis were selected if they were found to influence the squid fishery CPUE or squid habitats in previous studies (Chen, 1997, 1999; Murata and Nakamura, 1998; Chen and Tian, 2005). These covariates were divided into two groups: categorical covariates, including year, month, latitude and longitude (the centre point value of each fishing grid); and continuous covariates, including SST, SSS, and SLH. Moreover, as all Chinese squid-jigging vessels are modified from the same inshore bottom-trawlers and equipped with a main engine power

of 2×120 kW, squid-attracting lamp power of 112 kW and 16 squid-jigging machines, and thus almost identical in fishing power (Chen *et al.*, 2007), fishing power was not considered as a factor in modelling.

Two-way interactions among the main factors considered in this study were restricted to spatial effects versus temporal effects. The interactions between spatial-temporal factors and environmental variables were excluded in the GLM due to possible correlations between these terms (Bigelow *et al.*, 1999). Because the life-span of neon flying squid is only one year (Yatsu *et al.*, 1998) and this species is highly migratory, the interactive effects of spatial versus temporal factors on abundance are considered to be more important (Chen *et al.*, 2003). Correlation matrix between variables (including interaction terms) was calculated in order to avoid the high correlated terms in the GLBM model. All of the analyses were based on the assumption that CPUE data are log-normally distributed. An examination of plots of standardized residuals and quantile-quantile plots provided no evidence for violation of this assumption (results not shown because of space limitation).

**Generalized Linear Bayesian Models**

*Lognormal basic model*

The GLBM is based on the Bayesian estimator for the GLM. In this study we used lognormal GLM for the Bayesian estimation. Therefore, CPUEs were modelled using the log-normal distribution:

$$U_{i,j,k,l} \sim \text{Lognormal}(\bar{U}_{i,j,k,l}, \sigma^2) \quad (3)$$

where *U<sub>i,j,k,l</sub>* is the observed CPUE at *Latitude j* and *Longitude k* in *Month l* of Fishing Year *i*,  $\bar{U}_{i,j,k,l}$  is the mean of the distribution on the log scale for  $U_{i,j,k,l}$ , and  $\sigma$  is the standard deviation of the distribution on the log scale. Following Campbell and Tuck (1996), to overcome the problem of CPUE having zero values we added a small constant ( $\Delta$ ) equivalent to 10% of the overall CPUE (total catch divided by total effort over all the year) to each observed CPUE value:

$$U_{i,j,k,l} + \Delta \sim \text{Lognormal}(\bar{U}_{i,j,k,l}, \sigma^2) \quad (4)$$

The mean,  $\bar{U}_{i,j,k,l}$ , was estimated based on the effects of explanatory variables:

$$\bar{U}_{i,j,k,l} = k + year_i * Year + month_l * Month + lon_k * Longitude + lat_j * Latitude + SST * SST + SSS * SSS + SLH * SLH + Interactions. \quad (5)$$

where *k* is the intercept; *year<sub>i</sub>*, *month<sub>l</sub>*, *lat<sub>j</sub>*, and *lon<sub>k</sub>* are the effects of *Year i*, *Month l*, *Latitude j*, and *Longitude k*, respectively; and *interactions* are the effects between spatial covariates and temporal covariates. The abundance index at *Latitude j* and *Longitude k* in *Month l* of Fishing Year *i* was estimated as:

$$U_{i,j,k,l} = \exp(k + year_i * Year + month_l * Month + lon_k * Longitude + lat_j * Latitude + SST * SST + SSS * SSS + SLH * SLH + Interactions + \sigma^2/2) - \Delta \quad (6)$$

#### Non-hierarchical and hierarchical models

In this study we developed lognormal generalized linear Bayesian models of both hierarchical and non-hierarchical form for standardizing CPUE. Hierarchical generalized linear models are also known as Generalized Linear Mixed Models. For a non-hierarchical model, interaction between spatial covariates and temporal covariates has a fixed effect; while for a hierarchical model we can assume that the interaction has a random effect (i.e. use of nested parameters), which can lead a more reliable estimation of the interactions.

#### Prior distribution of parameters and model initialization

Bayesian analyses require all model parameters to have prior probability distributions. Because we only had limited knowledge about all the model parameters, we assumed that the priors for all the parameters were non-informative. We assigned a normal distribution with mean = 0 and a large variance (100000) for  $k$ ,  $year_i$ ,  $month_l$ ,  $lat_j$  and  $lon_k$ . The normal distribution with such a large variance is similar to a uniform distribution. For  $\sigma^2$  we assigned an inverse gamma distribution,  $1/\sigma^2 \sim Gamma(0.001, 0.0001)$ , where 0.001 and 0.0001 are the parameters of shape and rate (the inverse scale) of the gamma distribution (Ntzoufras, 2009). For the non-hierarchical lognormal model, the prior of each *Interaction* was assigned a normal distribution defined as  $N(0, 100000)$ . For the hierarchical models, the priors of *Interactions* were assumed to follow a normal distribution  $\sim N(U_c, \sigma_c^2)$ , the hyperparameter,  $U_c$ , was assigned a normal distribution,  $\sim N(0, 100000)$  and  $1/\sigma_c^2 \sim Gamma(0.001, 0.0001)$ . Sensitivity tests were conducted to evaluate the robustness of these priors especially for the shape of the variance parameter of the Gamma.

#### Estimated abundance index using predictive CPUE

In order to deal with the high spatial heterogeneity of fishing effort distribution in the squid fishery, we used our models to predict CPUE in unfished areas based on the estimated effects of the explanatory variables as long as these areas were fished in some of the other months. Therefore, the abundance index for *Year i* and *Month l*,  $A_{i,l}$ , was estimated in two different ways, depending on whether predicted CPUE values in the unfished areas were used in the calculation of the abundance index. If predicted catch rates were used,  $A_{i,l}$  was calculated as:

$$A_{i,l} = \frac{1}{TG} \sum_1^{TG} U_{i,j,k,l} \quad (7)$$

where TG is the total number of fishing grids fished in the time series. If predicted catch rates were not used, TG refers to the number of fished grids in *Year i* and *Month l*.

#### Model selection

We started the process by analyzing the GLBM based on Equation (5) without adding interaction covariates and selected the most fitted model using the stepwise approach based on deviance information criteria (DIC) (Spiegelhalter *et al.*, 2002). The smaller the DIC value is, the better the model is. We then processed it by adding interaction covariates ( $year_i * month_l$ ,  $year_i * lon_k$ ,  $year_i * lat_j$ ,  $month_l * lon_k$  and  $month_l * lat_j$ ) both non-hierarchically and hierarchically, and identified the best-fitting models by these two approaches based on the DIC. Uncorrelated terms in each model were ensured during the modelling process.

#### Convergence diagnostics

Convergence is a fundamental assumption for the application of any method for sampling posterior distribution (Cowles and Carlin, 1996; Brooks and Gelman, 1997). In this paper, a large number of MCMC “burn-in” and iterations were done in order to stabilize the central tendency of posterior distributions for all parameters. “Burn-in” was discarded before the analysis. Two chains were used for convergence checking by comparing the variances within the chains and across the chains (Gelman and Rubin, 1992). The within-chain variance was calculated by taking  $n$  draws of  $m$  parameters with Equation (8), and the estimated variance across the chains was computed with Equation (9).

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_j^i - \bar{\theta}_j)^2 \quad (8)$$

$$\hat{V}(\theta) = \left(1 - \frac{1}{n}\right)W + \frac{1}{n} \left(\frac{n}{m-1} \sum_{j=1}^m (\theta_j - \bar{\theta})^2\right) \quad (9)$$

Once convergence is reached,  $W$  and  $\hat{V}(\theta)$  should be almost equivalent

$$\text{(e.g. } \sqrt{R} = \sqrt{\frac{\hat{V}(\theta)}{W}} = 1; \text{ Gelman and Rubin, 1992).}$$

Alternate convergence diagnostic of Geweke (1992) was also applied in this study because we used the posterior mean as the estimated value.

#### Solving models

All of the Bayesian analyses and modelling were implemented with R version 2.10.0 and WinBUGS14

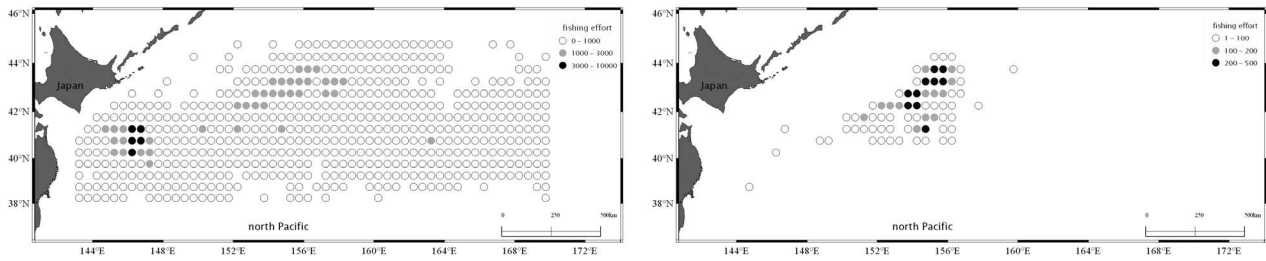


FIG. 1. – (a) fishing ground of the winter-spring cohort of *Ommastrephes bartramii* for the Chinese mainland squid jigging fishery and the spatial distribution of fishing efforts in 1995-2004; (b) spatial distribution of fishing effort in August 1998 (randomly selected from all the fishing months in 1995-2004).

(Bayesian inference Using Gibbs Sampling; Spiegelhalter *et al.*, 2003). Appropriate MCMC iterations were used on the basis of the convergence assessment and a burn-in period with 100000 initial iterations being discarded was used in order to ensure that samples were drawn from a distribution close enough to the true stationary distribution to be usable for the estimation and inference. As initial values were part of the computing process, we assigned the covariates different initial values for two chains when we did MCMC runs. Autocorrelations among draws from posteriors were eliminated by thinning the samples for every 50 MCMC runs (Geweke, 1992).

RESULTS

Spatial distribution

The analyses were based on 3361 records, of which 3357 had positive catches and 4 had zero catch. The distribution of effort is shown in Figure 1a, indicating that there were 587 fishing grids fished from 1995 to 2004, while effort in a given month was only distributed in less than 100 fishing grids (Fig. 1b). Most fishing effort was deployed in August to October and effort data in June and November only consisted of a small proportion of the overall efforts.

Correlation

The correlations of all covariates including interaction terms are summarized in the matrix plot (Fig. 2). Non-significant correlations were observed between covariates without interaction terms. However, interaction terms of  $month_i * lon_k$  and  $month_i * lat_j$  showed high correlations with other interaction terms. Therefore, we excluded these two terms from the modelling process. There is no significant correlation between CPUE and any single factor alone considered in this study (Fig. 2).

Model selection and comparison

The best-fitting models derived from three approaches (i.e. interaction terms being excluded, in-

teraction terms of fixed effects being included and interaction terms of random effects being included) are listed in Table 1. The DIC results showed that the model selected in GLBM analyses and containing  $month_b$ ,  $lat_j$ ,  $SST$ ,  $SSS$  and  $SLH$  had the lowest value, suggesting that this model outperformed the other two best-fitting models which were derived from considering interaction terms of fixed and random effects (Table 1). The results of this model are shown in Table 2.

For the GLBM analyses from which interaction terms were excluded, the best-fitting model revealed that all the covariates except for  $year_i$  and  $lon_k$  considered in this study had significant effects (95% posterior interval does not cover zero) on CPUE and  $SSS$  had the most significant effect on CPUE (Fig. 3a). The covariates,  $month_i$ ,  $lat_j$  and  $SST$ , were the second, third and fourth most significant covariates, respectively (Fig. 3a), and  $SLH$  had a weaker impact on CPUE than the other four significant covariates (Fig. 3a). For the non-hierarchical GLBM analyses in

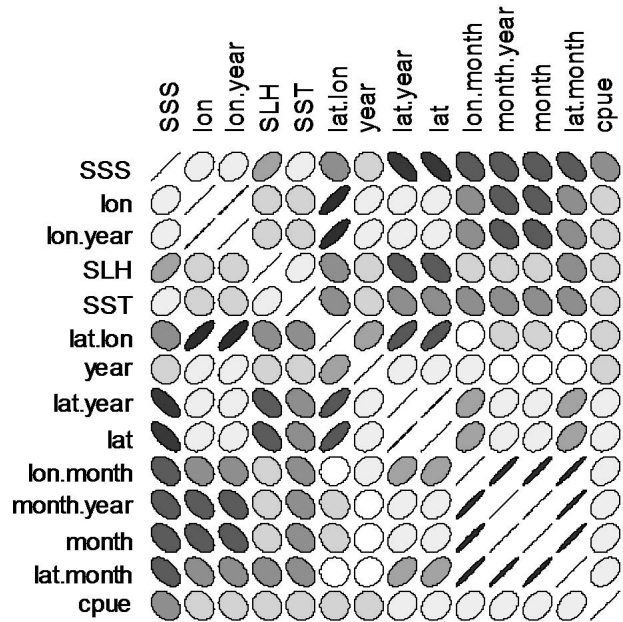


FIG. 2. – Correlation matrix of all covariates for standardizing CPUE of *Ommastrephes bartramii*. Ellipse shows the correlations (negative and positive), straight line represents the strongest correlation and circle represents the absence of correlation.

TABLE 1. – Results of the best-fitting models derived from the three models (i.e. interaction terms excluded, fixed effects interaction terms included, and random effects interaction term included)

Approach	model	DIC
Interaction terms excluded	$\bar{U}_{i,j,k,l} = month_l + lat_j + SST + SSS + SLH$	9458.5
Fixed effects interaction terms included	$\bar{U}_{i,j,k,l} = lat_j + SST + SSS + SLH + year_i * month_l + year_i * lon_k$	9462.2
Random effects interaction term included	$\bar{U}_{i,j,k,l} = lat_j + SST + SSS + SLH + year_i * month_l + year_i * lon_k$	9461.9

TABLE 2. – Posterior distribution of covariates and summary statistics for the selected models in which interaction terms were excluded.

Node	Mean	SD	MC error	2.50%	Median	97.50%	$\sqrt{R}$	n.eff
$month_l$	0.18200	0.00976	0.00025	0.16240	0.18230	0.20070	1.00	3300
$lat_j$	0.10650	0.00925	0.00071	0.08878	0.10640	0.12570	1.00	1200
$SST$	0.03955	0.00514	0.00016	0.02964	0.03952	0.04950	1.00	4000
$SSS$	-0.18770	0.01154	0.00088	-0.21770	-0.18730	-0.16570	1.00	3900
$SLH$	0.00235	0.00127	0.00004	0.00150	0.00235	0.00488	1.00	4000
$1/\sigma^2$	2.01500	0.04952	0.00008	1.91900	2.01600	2.11300	1.00	880
deviance	9452.0	3.67800	0.13290	9447.0	9452.0	9461.0	1.00	1200

MC refers to Monte Carlo; SD is the sample standard deviation; n.eff is a crude measure of effective sample size; 2 chains, each with 200000 iterations (first 100000 discarded, thinning sample = 50), 4000 iterations saved.

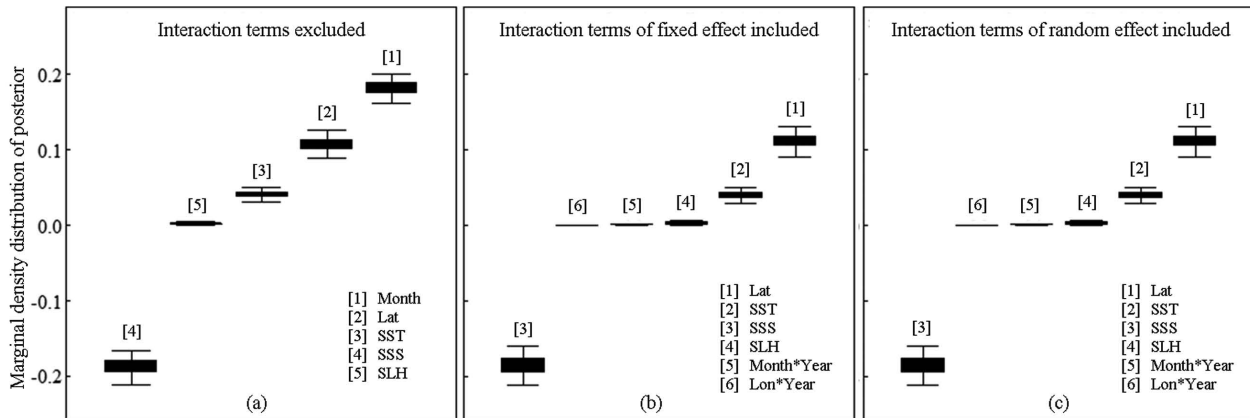


FIG. 3. – Impacts of covariates in three best-fitted models (a, interaction terms excluded; b, fixed effect interaction terms included; c, random effect interaction terms included) on the CPUE of *Ommastrephes bartramii* (y-axis shows the posterior intervals of parameters).

which interaction terms were assumed to have fixed effects on CPUE, the best-fitting model showed that the covariates of  $lat_j$ ,  $SST$ ,  $SSS$ ,  $SLH$ ,  $year_i \times month_l$  and  $year_i \times lon_k$  had significant effects on CPUE (Fig. 3b). Of the 6 significant covariates,  $SSS$ ,  $lat_j$ , and  $SST$  were the most, second and third significant covariates, respectively, and  $year_i \times month_l$ ,  $year_i \times lon_k$  and  $SLH$  had very weak impacts on CPUE, compared with the other 3 significant covariates (Fig. 3b). For the hierarchical GLBM analyses in which interaction terms were assumed to have random effects on CPUE, the best-fitting model was the same as the second approach but with lower DIC value. The impacts of covariates on CPUE were almost the same as the best-fitting model derived from the second approach (Fig. 3c).

The spatial-temporal interaction terms with both fixed and random effects were found to be less important in affecting CPUE according to the DIC results (Table 1), and the spatial and temporal factors tended

to influence CPUE independently as a covariate rather than together as a covariate.

### Convergence diagnostics

There is no evidence of non-convergence

$$(e.g. \sqrt{R} = \sqrt{\frac{\hat{V}(\theta)}{W}} = 1)$$

for posterior distributions of parameters for the best-fitting model based on the convergence diagnostic (Table 2). Marginal posterior density plots of parameters for the three best-fitting models also showed a relatively stable central tendency (Fig. 4). Alternate convergence diagnostic of Geweke (1992) indicated that there was no evidence of non-convergence ( $P > 0.05$ ) for the posterior distributions of parameters in the best-fitting model derived from the three approaches.

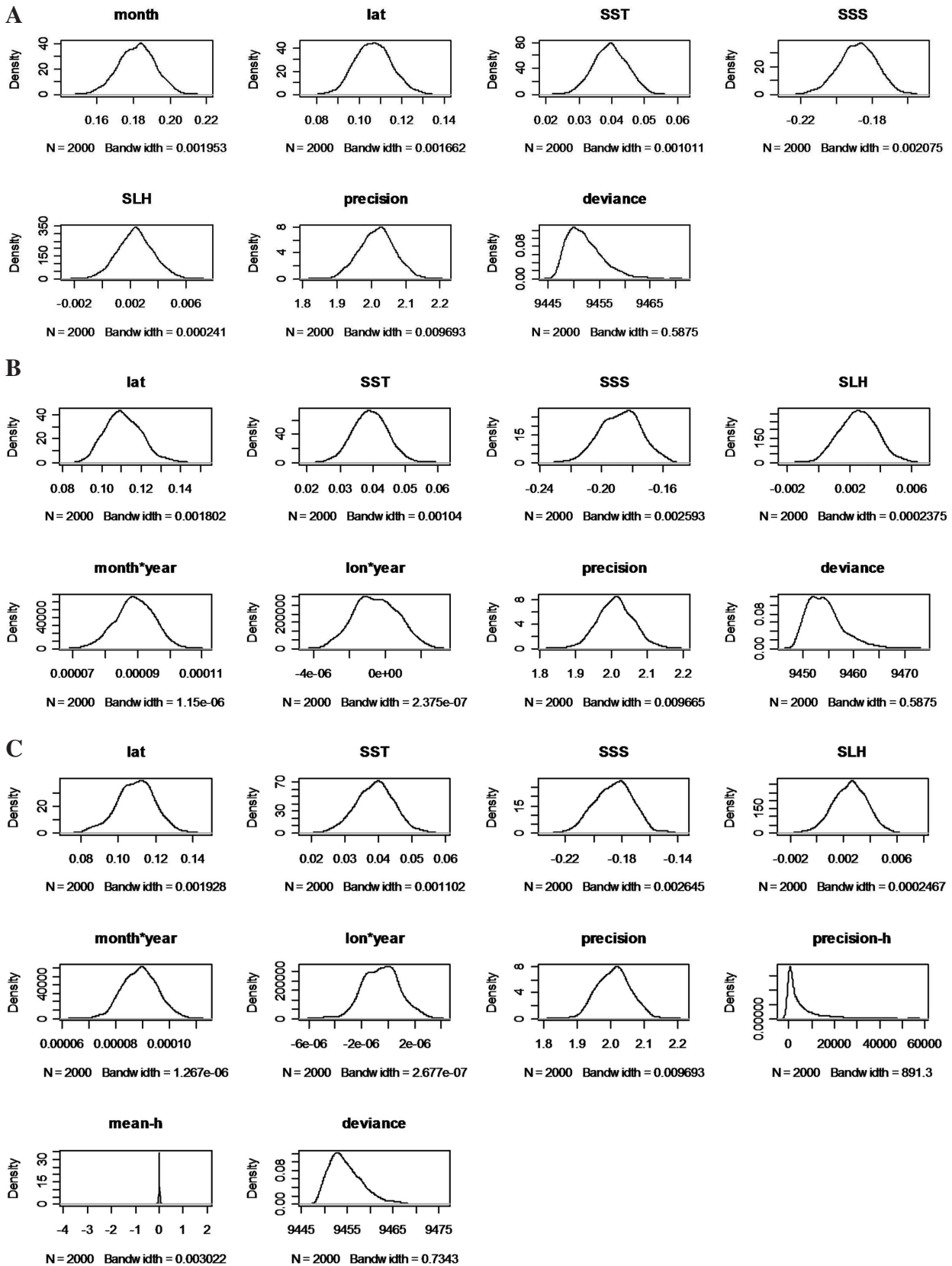


Fig. 4. – The marginal density distribution of posterior for the selected models in which (A) interaction terms were excluded, (B) fixed effect interaction terms were included and (C) random effect interaction terms were included.

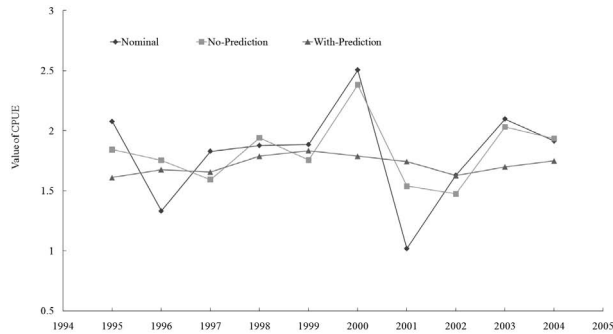


FIG. 5. – Comparison of yearly observed CPUE with standardized CPUE using GLBM based on data with and without predicted CPUEs in 1995-2004.

### Comparison of observed CPUE and standardized CPUEs

The CPUE values standardized using the best-fitting model with two different sets of data (with and without predicted abundance index) were compared with the observed CPUE under different spatial and temporal scales. For yearly CPUEs which were averaged across all fishing months, mean standardized CPUE estimates from the best-fitting model with the lowest DIC value not including predicted CPUE data in unfished grids in modelling had a similar trend to the observed CPUE, in particular for the later years, but tended to fluctuate less than the observed CPUE. Both showed that the highest stock size occurred in 2000. However, the derived abundance index not including the predicted CPUE data in unfished grids showed that the lowest abundance appeared in 2002, while for the observed CPUE it occurred in 2001 (Fig. 5). Yearly mean standardized

CPUE with the inclusion of predicted CPUE data in the unfished grids showed a stable trend in 1995 to 2004 with little fluctuation (Fig. 5). For monthly CPUE, the standardized CPUE with and without predicted CPUE data in the unfished grids showed the same trend, i.e. the stock size increased with the months, but in general the standardized CPUE with predicted CPUE data in the unfished grids was lower than the observed CPUE, in particular for the months when fishing activity peaked (August to October; Fig. 6).

### DISCUSSION

Here we developed Bayesian CPUE standardization models, non-hierarchical lognormal and hierarchical lognormal models, and applied these models to the CPUE standardization for the neon flying squid fishery to address the problems caused by the spatial heterogeneity of fishing effort distribution in the neon flying squid fisheries (Tian *et al.*, 2009). This study suggests that the approaches developed in this study can effectively deal with the CPUE standardization and estimation when spatial distribution of fishing efforts is highly heterogeneous, a phenomenon commonly observed in commercial fisheries (Hilborn and Walters, 1992).

In commercial fisheries for neon flying squid, fishing effort is usually concentrated in one small fishing area showing the highest returns because of high efficiency in searching fishing grounds and communications between vessels (Tian *et al.*, 2009). Our results show that monthly observed CPUEs appeared to be higher than the standardized CPUE, and the standardized CPUE based on data without predicted CPUE values in the unfished grids was lower than that with

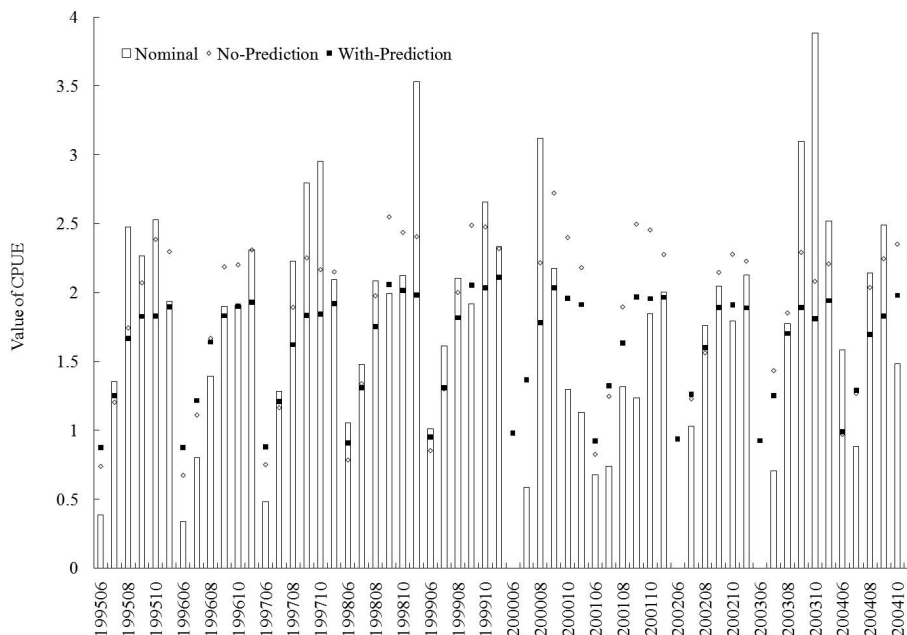


FIG. 6. – Comparison of monthly observed CPUE with standardized CPUE using GLBM based on data with and without predicted CPUEs in 1995-2004.



predicted CPUE values (Fig. 6), especially during the fishing peak of August to October (Chen, 1997). This indicates that high concentration of spatial distribution of fishing efforts may result in the observed CPUE and even the standardized CPUE without predicted CPUE values in the unfished grids having hyperstability relationships with abundance, which implies that CPUE stays high as abundance drops (Hilborn and Walters, 1992). Thus, previous studies for neon flying squid, which used observed CPUE as an abundance index directly in stock assessment (Chen *et al.*, 2003) and evaluating impacts of environmental variables on abundance (Cao *et al.*, 2009), may be biased. On the other hand, the models developed in this study have the capacity to predict abundance indices for unfished areas based on the estimated effects of the explanatory variables as long as these areas are fished in other months. Using such monthly CPUEs standardized with predicted abundance indices can deal with the spatial aggregation of fishing effort to a certain extent and reflect the abundance distribution more accurately. Previous studies also indicated that ignoring the unfished grids implicitly assumed that they behaved in the same way as the fished grids and could lead to severe hyper-depletion in abundance indices for fisheries, and several approaches were proposed to deal with the unfished grids (Walters, 2003; Campbell, 2004). However these approaches tend to bypass the unfished areas because of interpolation or weighting of the spatial regions. The GLBM developed in this study, on the other hand, aims to estimate the abundance index of unfished areas directly by sharing the information using the Bayesian statistical method.

The best-fitted model identified in the study includes the covariates of *month*, *lat<sub>j</sub>*, *SST*, *SSS* and *SLH*. These spatio-temporal factors were considered important in determining the spatial distribution of this squid fishery (Chen, 1997; Chen *et al.*, 2003; Cao *et al.*, 2009). The selection of the environmental factors in this study is consistent with the results of previous studies (Chen, 1997; Chen *et al.*, 2003; Cao *et al.*, 2009). In May or June when the fishing vessels of Chinese mainland started fishing operation, the squid gradually migrated northward from their spawning ground to the fishing ground for feeding and then southward back to spawning ground for spawning (Murata and Nakamura, 1998). The stock biomass would be closely related to the timing when peak recruitments migrated into the fishing ground and the latitudinal location where peak recruitments could migrate. Therefore *lat<sub>j</sub>*, which is the most significant covariate of all the spatio-temporal covariates of the best-fitted model (Fig. 3), is a critical factor for determining the distribution of peak recruitment. Our result indicates that *SST* and *SSS* were the main factors affecting CPUE. The other species of Ommastrephidae squids also have the similar linear relationships between abundance and *SST* or *SSS* (Rodhouse, 2001). However, the results of using traditional GLM models showed *SST* and *SSS* were

not significant in explaining CPUE (Tian *et al.*, 2009). The spatial-temporal interactions of both hierarchical and non-hierarchical models were identified as not significant in influencing CPUE as they were driven by environmental factors which were already included. For example, special oceanographic zones in the North Pacific such as transition zones, chlorophyll fronts (TZCF) and optimum-SST front zones for neon flying squid, which are closely related to the life history of neon flying squid, have a significant seasonal north-south movement (Ichii *et al.*, 2009). We also added the year effect (non-significant) for re-running the GLBM (Maunder and Punt, 2004), but found that it was hard to achieve convergence if we included the year effect.

The estimated posteriors were similar for different priors in the sensitivity analyses, indicating that the posteriors were determined by the data rather than by the priors. Compared with the observed CPUEs for neon flying squid stock, the standardized CPUEs tended to have smaller fluctuations and were in general lower than the observed CPUE, indicating that the impact of spatial, temporal and environmental factors on catch rates was largely removed through the standardization process.

There are also two issues that need to be addressed. Firstly, although we have a relatively large number of squid fishery data sets with a 10-year time span, these data lack spatial contrast. For example, there are almost no zero catch data because the fishing vessels would immediately depart the area when CPUEs dropped below economically viable levels. In a given time of the fishing season, the fishing vessels are usually concentrated at a few locations with similar environmental conditions, and there are small environmental gradients on the fishing grounds. This may result in loss of effective information for the Bayesian estimation. Additionally, the limited data of June and November could lead the results of monthly standardized CPUE for June and November to have large uncertainties. Future studies should compare the temporal trend of standardized CPUE with abundance indices derived from a fishery-independent scientific survey which is often considered as a more reliable abundance index because of its scientifically rigorous design (Hilborn and Walters, 1992).

Secondly, the basic model of GLBM developed in this study is GLM, which assumes linearity between the dependent and explanatory variables. However, many studies suggest that functional relationships between CPUE and environmental variables are likely to be non-linear (Bigelow *et al.*, 1999; Daniel and Michel, 2004). The general additive model (GAM), known as the extension of GLM with additive models, can analyze non-linear relationships (Hastie and Tibshirani, 1990), and Tian *et al.* (2009) showed that GAM models could explain more variations in CPUE standardization for neon flying squid with weak effects on CPUE. However, overfitting can be a problem with GAMs, which often make the fitted relationship perform less

well on a new data set than on the data set used for model fitting (Everitt, 2002). Bayesian models can easily incorporate heterogeneous data and are flexible in allowing the use of new data (Zhang *et al.*, 2009). Therefore, although we have to assume linearity of the relationship between neon flying squid CPUE and environmental variables (e.g. SST and SSS), the issue of formally detecting nonlinearity deserves further investigation before it is attempted to use GAM for neon flying squid CPUE standardization.

In summary, we explored a new modelling approach in this study to standardize fisheries CPUE, which tends to yield a more reliable abundance index and deals with the problem of spatial aggregations of fishing efforts by including predicted CPUEs of unfished grids in the CPUE standardization. However, we warn of potential problems associated with lack of contrast in the data used in this study, suggesting that the proposed modelling approach should be further tested using data from other fisheries.

#### ACKNOWLEDGEMENTS

We thank the Shanghai Ocean University for providing the CPUE data, and IRI/LDEO Climate Data Library of Columbia University for providing the environmental data. This study was supported by the Specialized Research Fund for the Doctoral Program of Higher Education (20093104110002), National Natural Science Foundation (NSFC40876090), “Shu Guang” project (08GG14) from Shanghai Municipal Education Commission, and Shanghai Leading Academic Discipline Project (Project S30702). Y. Chen’s involvement in the project was supported by the Shanghai Dongfeng Scholar Program.

#### REFERENCES

- Bellido, J.M., G.J. Pierce and J. Wang. – 2001. Modelling intra-annual variation in abundance of squid *Loligo forbesi* in Scottish waters using generalized additive models. *Fish. Res.*, 52(1): 23-29.
- Bigelow, K.A., C.H. Boggs and X. He. – 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fish. Oceanogr.*, 8(3): 178-198.
- Brooks, S. and A. Gelman. – 1997. General methods for monitoring convergence of iterative simulations. *JCGS*, 7: 434-455.
- Campbell, R.A. – 2004. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fish. Res.*, 70: 209-227.
- Campbell, R.A. and G. Tuck. – 1996. Spatial and temporal analyses of SBT fine-scale catch and effort data. Working Paper SBFWS/96/18 Presented at the Second CCSBT Scientific Meeting, Hobart, Australia, August 26-September 6 (1996). pp. 37.
- Cao, J., X.J. Chen and Y. Chen. – 2009. Influence of surface oceanographic variability on abundance of the western winter-spring cohort of neon flying squid *Ommastrephes bartramii* in the NW Pacific Ocean. *Mar. Ecol. Prog. Ser.*, 381: 119-127.
- Chen, X.J. – 1997. An analysis on marine environment factors of fishing ground of *Ommastrephes bartramii* in northwest Pacific. *J. Shanghai Fish. Univ.*, 6: 285-287 (in Chinese).
- Chen, X.J. – 1999. Study on the formation of fishing grounds of the large squid, *Ommastrephes bartramii* in the waters 160°E -170°E North Pacific Ocean. *J. Shanghai Fish. Univ.*, 8: 197-201 (in Chinese).
- Chen, X.J., Y. Chen, S.Q. Tian, B.L. Liu and W.G. Qian. – 2008. An assessment of the west winter-spring cohort of neon flying squid (*Ommastrephes bartramii*) in the Northwest Pacific Ocean. *Fish. Res.*, 92: 221-230.
- Chen, X.J. and S.Q. Tian. – 2005. Study on the catch distribution and relationship between fishing grounds and surface temperature for *Ommastrephes bartramii* in the northwestern Pacific Ocean. *Periodical of Ocean University of China*, 35: 101-107 (in Chinese).
- Chen, X.J., L.X. Xu and S.Q. Tian. – 2003. Spatial and temporal analysis of *Ommastrephes bartramii* resources and its fishing ground in North Pacific Ocean. *J. Fish. China*, 27(4): 334-342. (in Chinese)
- Chen, X.J., X.H. Zhao and Y. Chen. – 2007. El Niño/La Niña influence on the Western Winter-Spring Cohort of neon flying squid (*Ommastrephes bartarmii*) in the northwestern Pacific Ocean. *ICES J. Mar. Sci.*, 64: 1152-1160.
- Cowles, M.K. and B.P. Carlin. – 1996. Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91: 883-904.
- Daniel, G. and D. Michel. – 2004. Analysis of non-linear relationships between catch per unit effort and abundance in a tuna purse-seine fishery simulated with artificial neural networks. *ICES J. Mar. Sci.*, 61(5): 812-820.
- Everitt, B.S. – 2002. *The Cambridge Dictionary of Statistics*. 2nd edition. Cambridge University Press, Cambridge, UK.
- Gavaris, S. – 1980. Use of a multiplicative model to estimate catch rate and effort from commercial data. *Can. J. Fish. Aquat. Sci.*, 37(12): 2272-2275.
- Gelman, A. and D.B. Rubin. – 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7: 457-511.
- Geweke, J. – 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith (eds.), *Bayesian Statistics 4*, pp.169-193. Oxford University Press, Oxford.
- Hastie, T.J. and R.J. Tibshirani. – 1990. *Generalized Additive Models*. Chapman and Hall, Washington, DC. 352 pp.
- Hilborn, R. and C.J. Walters. – 1992. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman and Hall, New York. 570 pp.
- Ichii, T. – 2003. Akaika (Red flying squid). In: *The Current Status of International Fishery Stocks*, pp. 304-308. Japan Fisheries Agency/Fisheries Research Agency, (in Japanese)
- Ichii, T. and K. Mahapatra. – 2004. Stock assessment of the autumn cohort of neon flying squid (*Ommastrephes bartramii*) in the North Pacific based on the past driftnet fishery data. In: *Report of the 2004 Meeting on Squid Resources*, pp. 21-34. Japan Sea National Fisheries Research Institute, Niigata. (in Japanese).
- Ichii, T., K. Mahapatra, M. Sakai and Y. Okada. – 2009. Life history of the neon flying squid: effect of the oceanographic regime in the North Pacific Ocean. *Mar. Ecol. Prog. Ser.*, 378: 1-11.
- Maunder, M.N. and A.E. Punt. – 2004. Standardizing catch and effort data: a review of recent approaches. *Fish. Res.*, 70: 141-159.
- McAllister, M.K. and G.P. Kirkwood. – 1998. Bayesian stock assessment: a review and example application using logistic model. *ICES J. Mar. Sci.*, 55: 1031-1060.
- Mori, J. – 1997. Geographical differences between the parasites’ infection levels of the neon flying squid (*Ommastrephes bartramii*) from the North Pacific Ocean (Abstract). In: *Report of the 1995 Meeting on Squid Resources*, pp. 85-86. Contributions to the Fisheries Researches in the Japan Sea Block, N. 36. Japan Sea National Fisheries Research Institute, Niigata. (in Japanese)
- Murata, M. and Y. Nakamura. – 1998. Seasonal migration and diel vertical migration of the neon flying squid, *Ommastrephes bartramii*, in the North Pacific. In: T. Okutani (ed.), *Contributed Papers to the International Symposium on Large Pelagic Squids*, pp.13-30. Japan Marine Fishery Resources Research Center, Tokyo.
- Nagasawa, K., J. Mori and H. Okamura. – 1998a. Parasites as biological tags of stocks of neon flying squid (*Ommastrephes bartramii*) in the North Pacific Ocean. In: T. Okutani (ed.), *Contributed Papers to International Symposium on Large Pelagic Squids*, pp. 49-64. Japan Marine Fishery Resources Research Center, Tokyo.
- Nagasawa, K., J. Mori and H. Okamura. – 1998b. Parasites as

- biological tags of stocks of neon flying squid (*Ommastrephes bartramii*) in the North Pacific Ocean. In: *Report of the 1996 Meeting on Squid Resources*, pp. 120. National Research Institute of Far Seas Fisheries, Shimizu.
- Nelder, J.A. and R.W.M. Wedderburn. – 1972. Generalized linear models, *J. R. Statist. Soc. A.*, 137: 370–384.
- Ntzoufras, L. – 2009. *Bayesian Modeling Using Winbugs*. Wiley press. Hoboken, NJ.
- Pinheiro, J.C. and D.M. Bates. – 2000. *Mixed-Effects Models in S and S-plus*. Springer, New York.
- Quinn, T.J. and R.B. Deriso. – 1999. *Quantitative fish dynamics*. Oxford Univ. Press, New York.
- Rodhouse, P.G. – 2001. Managing and forecasting squid fisheries in variable environments. *Fish. Res.*, 54: 3-8.
- Roper, C.F.E., M.J. Sweeney and C.E. Nauen. – 1984. *FAO species catalogue. Vol 3: cephalopods of the world. An annotated and illustrated catalogue of species of interest to fisheries*. FAO Fish. Synop. 125, Rome.
- Spiegelhalter, D., N. Best, B. Carlin and A. Van der Linde. – 2002. Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc. B.*, 64: 583-639.
- Spiegelhalter, D., A. Thomas, N. Best and D. Lunn. – 2003. *WinBUGS Version 1.4 user manual*. MRC Biostatistics Unit, Cambridge.
- Tian, S.Q., X.J. Chen, Y. Chen, L.X. Xu and X.J. Dai. – 2009. Standardizing CPUE of *Ommastrephes bartramii* for Chinese squid-jigging fishery in Northwest Pacific Ocean. *Chin. J. Oceanol. Limnol.*, 27(4): 729-739.
- Wang, Y.G. and X.J. Chen. – 2005. *Oceanic Ommastrephidae Squids and their Fisheries in the World*. Ocean Press of China, Beijing. 124-155 pp. (in Chinese).
- Walters, C.J. – 2003. Folly and fantasy in the analysis of spatial catch rate data. *Can. J. Fish. Aquat. Sci.*, 60: 1433-1436.
- Yatsu, A., H. Tanaka and J. Mori. – 1998. Population structure of the neon flying squid, *Ommastrephes bartramii*, in the North Pacific. In: T. Okutani (ed.), *Contributed Papers to International Symposium on Large Pelagic Squid*, pp. 31-48. Japan Marine Fishery Resources Research Center, Tokyo.
- Yatsu, A. and T. Watanabe. – 1996. Interannual variability in neon flying squid abundance and oceanographic conditions in the central North Pacific, 1982–1992. *Bull. Natl. Res. Inst. Far Seas Fish.*, 33(3): 123-138.
- Zhang, Z., J. Lessard and A. Campbell. – 2009. Use of Bayesian hierarchical models to estimate northern abalone, *Haliotis kamtschatkana*, growth parameters from tag-recapture data. *Fish. Res.*, 95: 289-295.

Scient. ed.: F. Maynou.

Received December 9, 2010. Accepted March 2, 2011.

Published online August 22, 2011.