ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Fisheries Research 70 (2004) 265–274

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data

Tom Nishida[a,*], Ding-Geng Chen[b]

[a] *National Research Institute of Far Seas Fisheries, 5-7-1 Shimizu-Orido, Shizuoka-City, Shizuoka 424-8633, Japan*
[b] *Department of Mathematics and Statistics, Portland State University, P.O. Box 751, Portland, OR 97207, USA*

Abstract

Catch-per-unit-effort (CPUE) data have often been used to obtain a relative index of the abundance of a fish stock by standardizing nominal CPUE using various statistical methods. The theory underlying most of these methods assumes the independence of the observed CPUEs. This assumption is invalid for a fish population because of their spatial autocorrelation. To overcome this problem, we incorporated spatial autocorrelation into the standard general linear model (GLM). We also incorporated into it a habitat-based model (HBM), to reflect, more effectively, the vertical distributions of tuna. As a case study, we fitted both the standard-GLM and spatial-GLM (with or without HBM) to the yellowfin tuna CPUE data of the Japanese longline fisheries in the Indian Ocean. Four distance models (Gaussian, exponential, linear and spherical) were examined for spatial autocorrelation. We found that the spatial-GLMs always produced the best goodness-of-fit to the data and gave more realistic estimates of the variances of the parameters, and that HBM-based GLMs always produced better goodness-of-fit to the data than those without. Of the four distance models, the Gaussian model performed the best. The point estimates of the relative indices of the abundance of yellowfin tuna differed slightly between standard and spatial GLMs, while their 95% confidence intervals from the spatial-GLMs were larger than those from the standard-GLM. Therefore, spatial-GLMs yield more robust estimates of the relative indices of the abundance of yellowfin tuna, especially when the nominal CPUEs are strongly spatially autocorrelated.

Keywords: CPUE standardization; Spatial autocorrelation; Spatial-GLM; Tuna longline fisheries; Variogram; Variance–covariance matrix; Yellowfin tuna

1. Introduction

Catch-per-unit-effort (CPUE) data have often been utilized to obtain a relative index of the abundance of a fish stock. The nominal (observed) CPUEs are affected by changes of year, season, area of fishing and various environmental factors. Many statistical methods have been used to ‘standardize’ them to account for such variations. These include the general or generalized linear models (hereafter referred to as the standard-GLM), general additive models (GAM), neural networks (NN), regression trees (RT), and others (ICCAT,

* Corresponding author. Tel.: +81 543 36 6037; fax: +81 543 35 9642.
E-mail address: tnishida@affrc.go.jp (T. Nishida).

0165-7836/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2004.08.008

Table 1
A brief summary of the data (for details, see Nishida et al. (2003))

Type of data	Variables and resolutions	Source
Japanese (yellowfin) tuna longline commercial fisheries data in the Indian Ocean (1958–2001)	Catch, effort and number of hooks between floats (HBF) by month and 5° × 5° area	National Research Institute of Far Seas Fisheries (NRIFSF), Japan
Marine environmental data	Sea surface temperature (SST) by month and 5° × 5° area (1958–2001) Thermocline depth (at 20 °C) by month and 5° × 5° area (1958–2001)	Sub-arctic Gyre Experiment (SAGE) compiled by Japan Meteorological Agency, Japan JEDAC data set (Scripps Institution of Oceanography) and GAO data set (Gestionnaire d'Applications Océanographiques) compiled by Institut de Recherche pour le Développement (IRD), France

2003). Of these, the standard-GLM is the most commonly used. The statistical theory underlying these methods assumes that the observed CPUE data are independent. This assumption is invalid for a fish population, because many species of fish live and move together: the more closely in space and/or time[1] the observations of fish abundance are made, the more similar they are. Thus, spatial autocorrelation brings a potentially major problem in standardizing the CPUEs by use of the standard-GLM, GAM, NN or RT.

Habitat-based models (HBM) have been incorporated into the standard-GLM in recent CPUE standardizations (known below as the standard-GLM/HBM). Specifically, GLM/HBMs have been used to analyze data on the CPUEs of some species of billfish and tropical tuna (Hinton and Nakano, 1996; Bigelow et al., 2002; Yokawa and Takeuchi, 2002; Bigelow et al., 2003). The HBM uses the effective and nominal tuna longline fishing effort by accounting for both the swimming depths of the fish and the gear depths in the CPUE standardization. The standard-GLM/HBM produced more accurate standardized CPUEs and relative indices of the abundance of the fish.

In this paper, we incorporate spatial autocorrelation into the standard-GLM (hereafter called the spatial-GLM) and also into the spatial-GLM/HBM. As a case study, we use these models to analyze data on the CPUE of yellowfin tuna (*Thunnus albacares*) of the Japanese longline fisheries in the Indian Ocean, including two environmental factors – thermocline depths (TD) and sea surface temperature (SST). Finally, we evaluate the results of the spatial-GLMs, discuss the biases in the results of CPUE analyses, and outline our future approach to analyzing spatially structured CPUE data.

[1] Autocorrelation by time was not examined in this paper.

2. Methods

Two methods (standard-GLM and spatial-GLM) were used to analyze the Japanese yellowfin tuna longline CPUE data in the Indian Ocean. Each method was employed with and without HBM. The data were detailed in Nishida et al. (2003) and summarized in Table 1. The theory behind these methods is explained below. The Statistical Analyses System (SAS) package was utilized in all calculations: PROC MIXED (Littell et al., 1996) for the spatial-GLMs, and PROC GLM for the standard-GLMs. Both SAS procedures are available in the SAS/STAT module (SAS Institute, 1999).

2.1. Standard-GLM

The standard-GLM without HBM is of the form:

$$
\begin{aligned}
\ln(N\_\mathrm{CPUE}_{ijkl} + \mathrm{constant}) &= \mathrm{INTERCEPT} + \mathrm{YR}_i + M_j + A_k + G_l \\
&\quad + \mathrm{SST}_{ijk} + \mathrm{TD}_{ijk} + (\mathrm{YR}\times M)_{ij} \\
&\quad + (\mathrm{YR}\times A)_{ik} + (M\times A)_{jk} + (M\times G)_{jl} \\
&\quad + (A\times G)_{kl} + (\mathrm{SST}\times M)_j + (\mathrm{SST}\times A)_k \\
&\quad + (\mathrm{TD}\times M)_j + (\mathrm{TD}\times A)_k + \varepsilon_{ijkl} \qquad (1)
\end{aligned}
$$

where *N*_CPUE is the nominal CPUE (i.e. the number of yellowfin tuna caught per 1000 hooks); "constant" is 10% of the global mean of the nominal *N*_CPUE

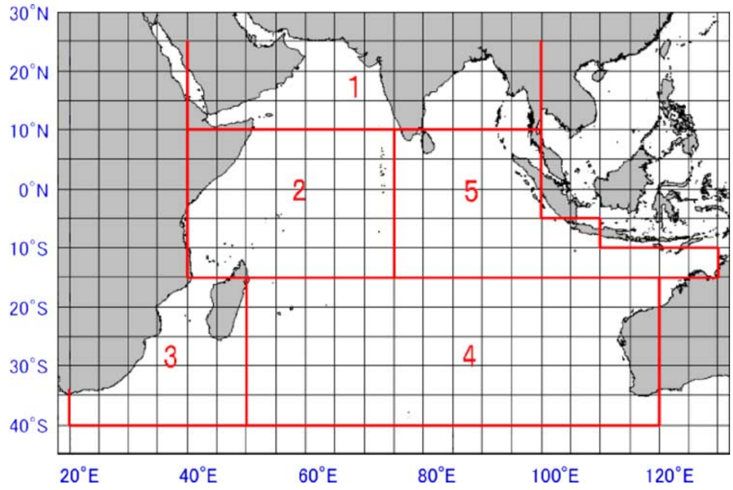

Fig. 1. Five sub-areas adopted by the IOTC (2002) for standardizing yellowfin tuna longline CPUE data in the Indian Ocean.

(0.87193 in this case) in order to mitigate the problem of zero catch (Campbell et al., 1996); INTERCEPT the intercept (mean *N*_CPUE); YR_i (i = 1 to I) is the effect of year from 1958 to 2001; M_j (j = 1 to J) the effect of month (January to December); A_k (k = 1 to K) the effect of sub-area (1–5) adopted at the IOTC Working Party of Tropical Tuna (WPTT) meeting in 2002 (IOTC, 2002) (Fig. 1); SST the effect of sea surface temperature (°C) (a continuous variable); TD the thermocline depth (m) (a continuous variable); ε_{ijkl} the error term, assumed to be independently, identically distributed (i.i.d.) with $N(0, \sigma^2)$ for all i, j, k and l; and G_l the effect of the number of hooks between floats (HBF). Following Bigelow et al. (2002) and Nishida et al. (2003), the number of hooks between two floats (HBF) was divided into six classes: class 1, 5–6; class 2, 7–9; class 3, 10–11; class 4, 12–15; class 5, 16–20; class 6, 21–25. As the Japanese tuna longline fisheries with 3–4 hooks between floats targeted swordfish at night, the data from this type of gear were excluded in subsequent analyses.

Data on the number of hooks between two floats are available from 1966 to 1975–2001; those from 1958 to 1965 and from 1967 to 1974 are missing. To ensure the continuity of the data for a comprehensive analysis, the HBF for the period 1958–1965 was estimated using its spatial pattern in 1966 and that for the period 1967–1974 was estimated from its mean patterns in 1966 and 1975. We also added the effects of sea surface temperature (SST) and thermocline depth (TD) to the model, as both environmental factors may have affected the distribution and abundance of yellowfin tuna.

The standard-GLM model, Eq. (1), is a sub-model of the full model that includes other single and interaction terms (Nishida et al., 2003). Because some single terms in the full model were statistically insignificant and some interaction terms caused it not to converge, as a result of a lack of data, these terms were excluded in the reduced standard-GLM.

2.2. Standard-GLM/HBM

Because gear type G was used in calculating the effective CPUE (ICCAT, 2003), the standard-GLM/HBM excluded any terms related to it to avoid double standardization, and is of the form:

$$\begin{aligned}\ln(E\_\mathrm{CPUE}_{ijkl} + \mathrm{constant}) &= \mathrm{INTERCEPT} + Y_i + M_j + A_k + \mathrm{SST}_{ijk} \\ &\quad + \mathrm{TD}_{ijk} + (\mathrm{YR} \times M)_{ij} + (\mathrm{YR} \times A)_{ik} \\ &\quad + (M \times A)_{jk} + (\mathrm{SST} \times M)_j + (\mathrm{SST} \times A)_k \\ &\quad + (\mathrm{TD} \times M)_j + (\mathrm{TD} \times A)_k + \varepsilon_{ijkl} \qquad (2)\end{aligned}$$

where all notations are the same as in Eq. (1), except that *E*_CPUE is the effective CPUE (i.e. the number of yellowfin tuna caught per 1000 effective hooks), "constant" is 1.78559, INTERCEPT the intercept (mean *E*_CPUE), and ε_{ijk} is assumed to be i.i.d., with $\varepsilon_{ijk} \sim N(0, \sigma^2)$ for all i, j, k and l.

2.3. Spatial-GLM

Eq. (1) can be written as:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{3}$$

where $\mathbf{Y}$ = (ln(CPUE$_{1111}$ + const), ..., ln(CPUE$_{ijkl}$ + const)) is the vectorized CPUE from the standard-GLM, $\mathbf{X}$ the corresponding design matrix, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{ijkl})$ is the corresponding $n \times 1$ vector, assumed to be distributed as $N(0, \sigma^2\mathbf{I})$, as ε is i.i.d.

The assumption of independent CPUE or ε in the standard-GLM, Eq. (3), is obviously violated for a fish population, for it ignores the spatial autocorrelation in the many features of the population. After all, fish move together with a positive spatial dependency. The more closely in space the observations are made, the more similar they are. Therefore, this positive spatial correlation should be included in the CPUE standardization by extending the standard-GLM approach.

This can be solved by resorting to geostatistical models (Cressie, 1991). The spatial dependency is justified in the geostatistical variogram model. With the inherited positive spatial dependency, the error term $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ in Eq. (3) is no longer i.i.d., for $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij} \geq 0, \quad i \neq j$, so that

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{4}$$

where $\varepsilon \sim N(0, V)$, and $V = (\mathrm{cov}(\varepsilon_i, \varepsilon_j))$ is the $n \times n$ variance–covariance matrix with non-negative off-diagonal elements. This is the spatial-GLM.

The covariance $\mathrm{cov}(\varepsilon_i, \varepsilon_j)$ between ε_i and ε_j is usually a function of the distance d_{ij} from the observational locations i and j, as specified by the spatial coordinates of longitudes and latitudes. The most commonly used models are the exponential model, spherical model, Gaussian model and the linear model, defined as:

$$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = \sigma_1^2 \times \begin{cases} 1 - \exp\left(-\dfrac{d_{ij}}{r}\right) & \text{(exponential model)} \\ 1 - \dfrac{3d_{ij}}{2r} + \dfrac{d_{ij}^3}{2r^3} & \text{(spherical model)} \\ 1 - \exp\left(-\dfrac{d_{ij}^2}{r^2}\right) & \text{(Gaussian model)} \\ 1 - rd_{ij} & \text{(linear model)} \end{cases} \tag{5}$$

for $0 < d_{ij} \leq r$, and $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = \sigma_0^2 + \sigma_1^2$ for $d_{ij} = 0$, where σ_0^2 is the nugget effect (which represents a discontinuity at the origin of the variogram with amplitude σ_0^2 and corresponds to the unobserved small-scale variance and observation error), σ_1^2 the partial sill or maximum level of heterogeneity, and r the range of influence (i.e., the distance beyond which there is practically no spatial correlation between data points). Consequently, the expression levels off to a constant magnitude of $\sigma_0^2 + \sigma_1^2$ at a distance greater than r, and is generally of the same magnitude as the statistical variance of the sample population.

The parameters in Eq. (5), along with the parameter vector β in the spatial-GLM (Eq. (4)), can be estimated simultaneously using the maximum likelihood estimation (ML) or the restricted maximum likelihood estimation (REML), with the log-likelihood function:

$$\begin{aligned} \text{ML}: \ & ll(\beta, \sigma_0^2, \sigma^2, r) \\ & = -0.5\, n \ln(|V|) - 0.5n \times \ln(D'V^{-1}D) \\ & \quad - 0.5\left[1 + \ln\left(\frac{2\pi}{n}\right)\right], \\ \text{REML}: \ & ll(\beta, \sigma_0^2, \sigma^2, r) \\ & = -0.5n \ln(|V|) - 0.5n \ln(|X'V^{-1}X) - 0.5(n-p) \\ & \quad \times \ln(D'V^{-1}D) - 0.5(n-p)\left\{1 + \ln\left[\frac{2\pi}{n-p}\right]\right\} \end{aligned} \tag{6}$$

where $D = Y - X(X'V^{-1}X)^{-}X'V^{-1}Y$, and p the rank of X. An optimization procedure can be used to get the estimates $\hat{\beta}, \hat{\sigma}_0^2, \hat{\sigma}^2$ and $\hat{r}$ of the parameters, and statistical inference can be made, based on the estimated parameters and the likelihood ratio test (LRT).

For example, the LRT can be used to test whether there is spatial dependency for H_0: $r = 0$. If the null hypothesis is rejected, there is a significant spatial dependency in the data. The LRT can also be used to test the nugget effect as $H_0: \ \sigma_0^2 = 0$ for the unobserved small-scale variance and observation error.

2.4. Spatial-GLM/HBM

The same procedure used for the spatial-GLM can be applied to the effective CPUE, as defined in Eq. (2), to yield the spatial-GLM/HBM. For both spatial-

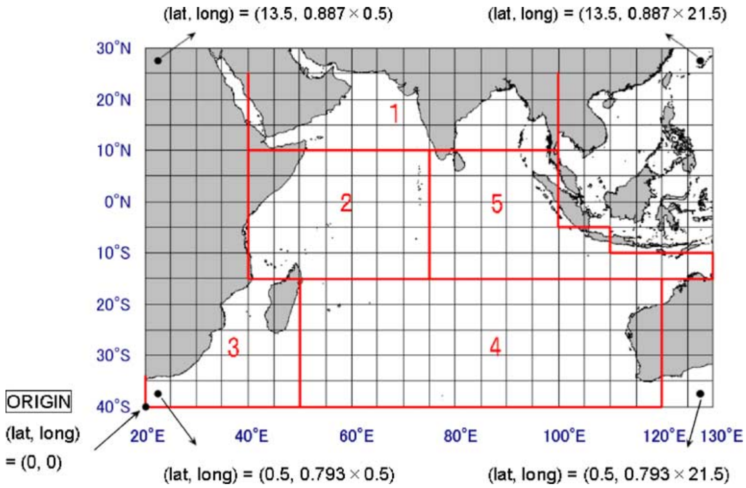

Fig. 2. Definition of the coordinate system for computing the distances between two 5° × 5° areas in the distance-based models in the spatial-GLM (the distance of 5° latitudes on the Equator is set to 1).

GLM and spatial-GLM/HBM, the spatial autocorrelation structure was modeled as covariograms, as defined in Eq. (5), which require data on the distances between observations. In the spatial-GLM, three distance-related parameters (sill, range and nuggets) were estimated, along with those in the standard-GLM. The spatial-GLM was carried out by PROC MIXED in the SAS package (SAS Institute, 1999).

2.5. Abundance indices

The estimated abundance indices in year *i* in all the GLM models were calculated by exponentiating the least squared means of the year effect from the GLM model (as part of the SAS outputs) and then subtracting from it a constant (ICCAT, 2003).

2.6. A case study

Four GLMs (standard-GLM, standard-GLM/HBM, spatial-GLM and spatial-GLM/HBM) were used to analyze yellowfin tuna CPUE data (1958–2001) of the Japanese longline fisheries in the Indian Ocean, together with some environmental factors. For the spatial-GLM, the distances were computed by setting up a coordinate system (Fig. 2). In this coordinate system, the lower left corner (20°E, 40°S) was treated as the origin, i.e. (lat, long) = (0, 0). By setting that the distance of the 5° latitudes on the Equator is unity (=1), the coordinates of all the central points of 5° × 5° areas were computed. The coordinates of the four corners are also given, in Fig. 2. For this computation, the spherical model for Earth was applied, then the *x*- and *y*-distances from the origin to the central point of the 5° × 5° areas were computed, based on the distances between 5° latitudes (Table 2) (note that the distances between 5° longitudes = 1 at any latitude).

In the spatial-GLM and spatial-GLM/HBM, four distance models (Gaussian, exponential, linear, and spherical) were also examined. Of the 10 models fitted, two were standard-GLMs, and eight spatial-GLMs. We evaluated their results by AIC, R^2 and graphically.

Table 2
Distance of 5° latitudes when the distance of 5° latitudes on the Equator is unity

Latitude range by 5°	Latitude at the central point	Distance of the 5° latitudes at the central point
0–5	2.5	0.9990
5–10	7.5	0.9914
10–15	12.5	0.9763
15–20	17.5	0.9537
20–25	22.5	0.9239
25–30	27.5	0.8870
30–35	32.5	0.8434
35–40	37.5	0.7934

3. Results

The results of fitting the 10 GLMs are summarized in Table 3, which gives AIC, R^2 and the results of the test of the spatial independence in the data. The spatial-GLMs always gave the best goodness-of-fit to the data. The HBM-based GLMs always produced better goodness-of-fit than those without the

Table 3
Summary of the results of fitting the 10 GLMs and the results of the spatial independency (r) test

Type of model	R^2	AIC	AIC (rank)	A, $-2LL_{spatial}$	B, $-2LL_{standard}$	LRT = (A) − (B)	Test (rank)
Standard-GLM	0.585	75780	(10)	–	75778	–	–
Standard-GLM/HBM	0.602	75041	(9)	–	75039	–	–
Spatial-GLM							
Exponential	0.749	68676	(7)	68670	75778	7108	(7)***
Gaussian	0.755	68281	(5)	68275	75778	7503	(3)***
Linear	0.711	70331	(8)	70325	75778	5453	(8)***
Spherical	0.754	68435	(6)	68429	75778	7349	(5)***
Spatial-GLM/HBM							
Exponential	0.761	67649	(3)	67643	75039	7396	(4)***
Gaussian	0.767	67267	(1)	67261	75039	7778	(1)***
Linear	0.763	67876	(4)	67870	75039	7169	(6)***
Spherical	0.768	67430	(2)	67424	75039	7615	(2)***

The likelihood ratio test (LRT) statistics for H_0: spatially independent ($r = 0$) is calculated by (A) − (B) = ($-2LL_{spatial}$) − ($-2LL_{standard}$) (LL is the value of the log-likelihood function) which is asymptotically χ^2 distributed.
*** Highly significant at $\chi^2_{1,0.001} = 10.83$.

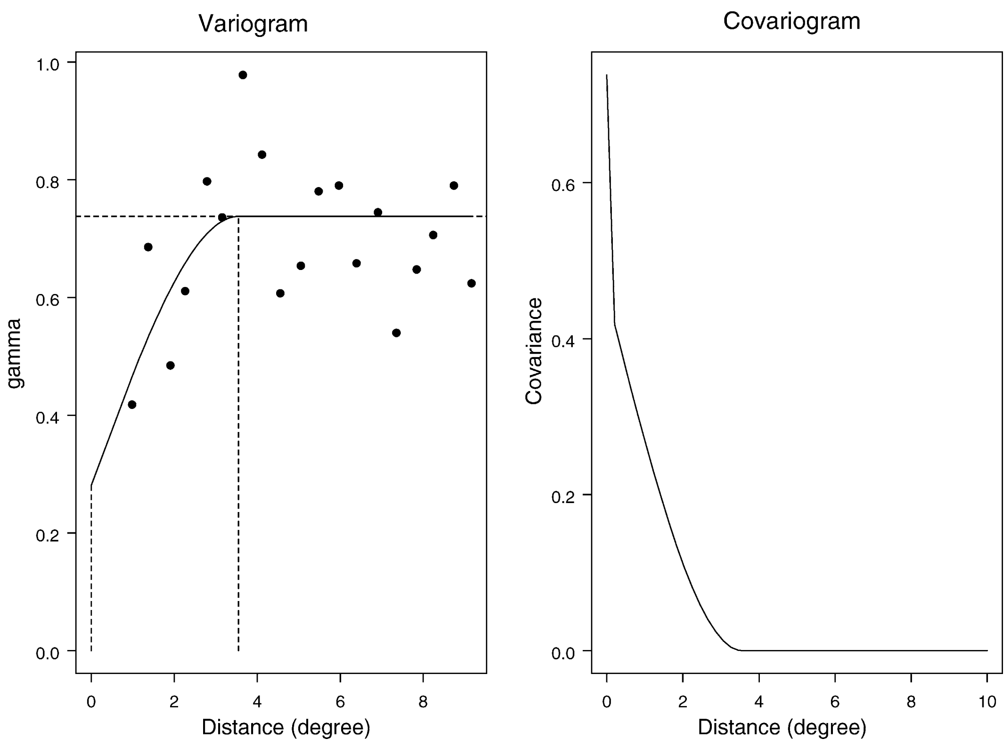

Fig. 3. Semivariogram (left panel) and covariogram (right panel) models in the spatial-GLM.

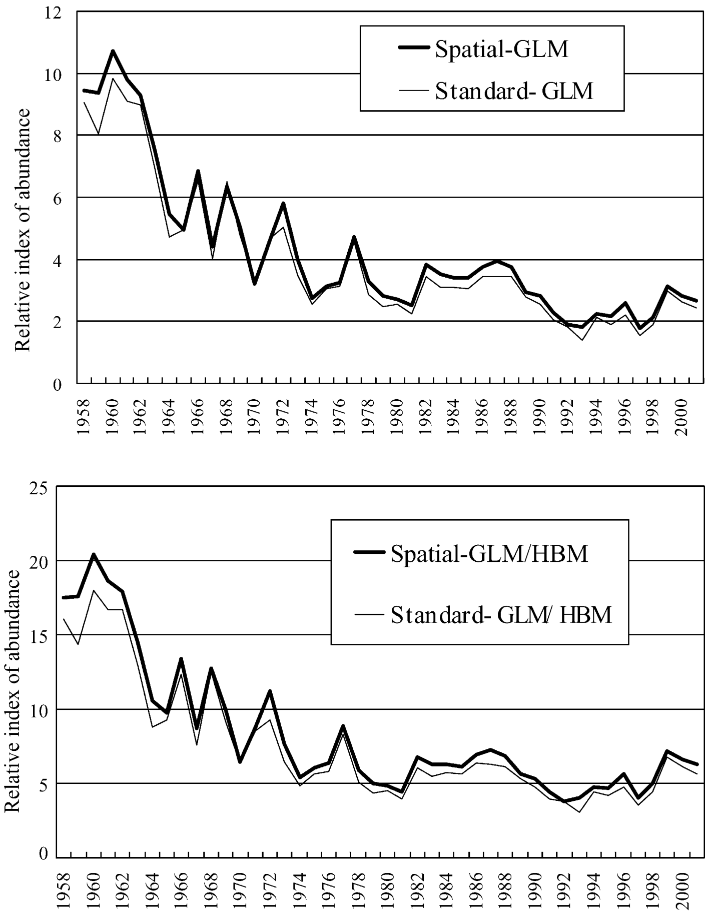

Fig. 4. A comparison of the relative index of abundance of yellowfin tuna from the spatial-GLM with that from the standard-GLM (upper panel), and from the spatial-GLM/HBM with that from the standard-GLM/HBM (lower panel).

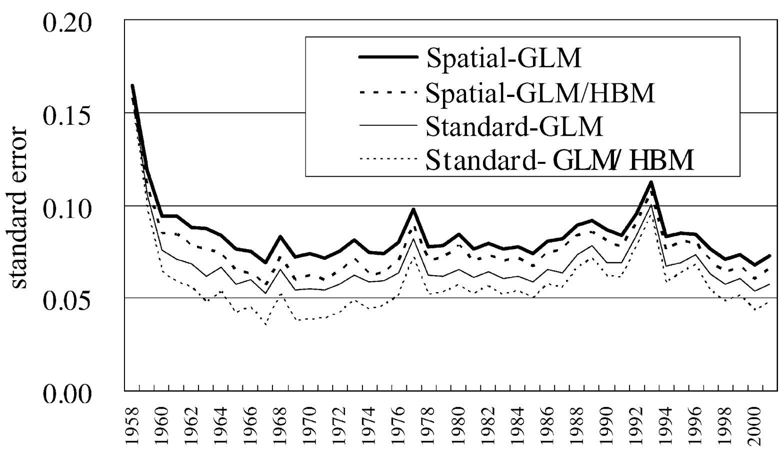

Fig. 5. Comparisons of the standard errors from the four GLMs.

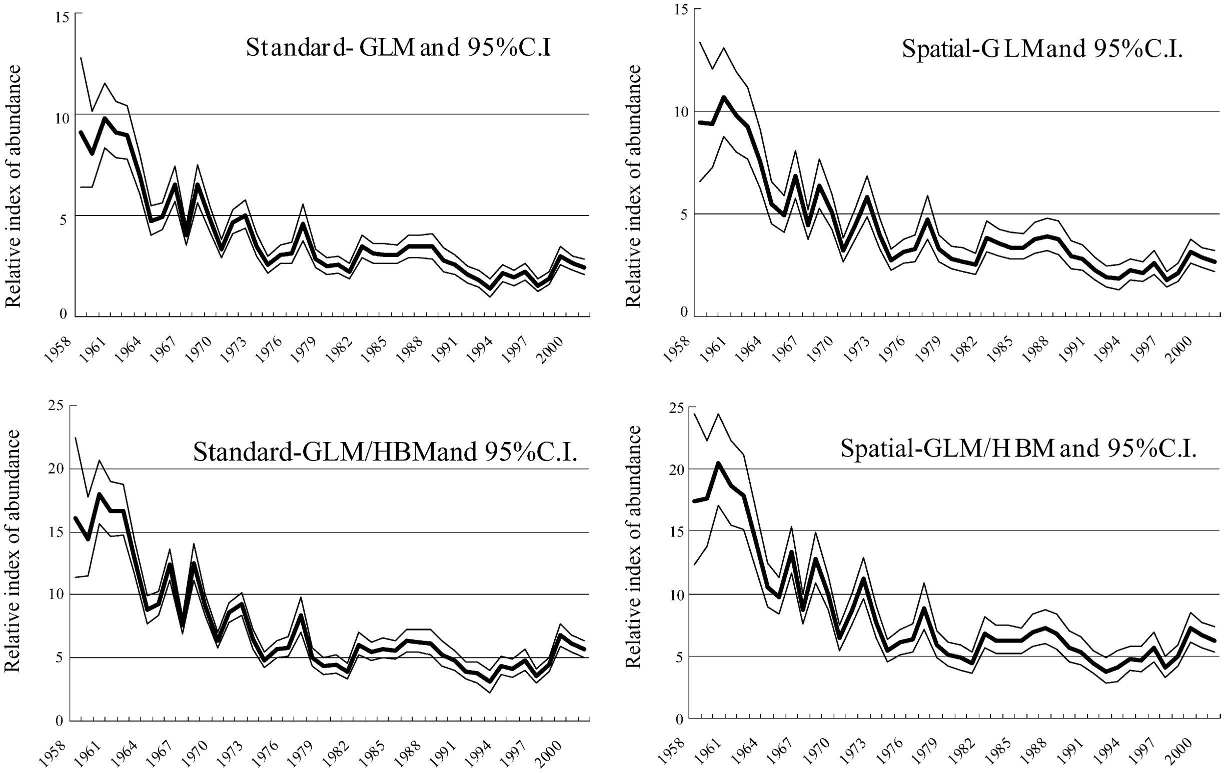

Fig. 6. A comparison of the relative index of abundance of yellowfin tuna and its 95% confidence intervals (C.I.) from the spatial-GLM with that from the standard-GLM (upper panel), and from the spatial-GLM/HBM with that from the standard-GLM/HBM (lower panel).

HBM. Of the four distance models, the Gaussian model had the best goodness-of-fit. Therefore, the spatial-GLM/HBM with the Gaussian distance model was the 'best' model in this case study.

The residuals from the standard-GLM were not i.i.d., but were autocorrelated within a distance of four $5^\circ \times 5^\circ$ units (Fig. 3), or about 20° latitude. The likelihood ratio test (LRT) between the spatial-GLM and standard-GLM for spatial independence of H_0: $r = 0$ showed that the LRT was highly significant in all cases (Table 3).

The relative indices of the number of yellowfin tuna from standardizing their CPUEs are provided in Fig. 4. The point estimates, standard errors (S.E.) and 95% confidence intervals (C.I.) were also compared among the standard-GLM, spatial-GLM, standard-GLM/HBM and spatial-GLM/HBM (Figs. 4–7). The temporal trends of the point estimates differed slightly between the two methods in each comparison (Fig. 4), while the standard errors and 95% confidence intervals of the spatial-GLM were larger than those of the standard-GLM (Figs. 5–7). Of all the GLM models, the average ratios of the 95% C.I. of the spatial-GLM to the standard-GLM were 1.23 (upper C.I.) and 1.09 (lower C.I.), while those for the GLM/HBM were 1.46 (upper C.I.) and 1.35 (lower C.I.) (Fig. 7). Thus, the C.I.s from the GLM/HBM models were much larger.

4. Discussion

4.1. Evaluation of the spatial-GLM

Although the temporal trends in the CPUEs from the spatial-GLMs did not differ greatly from those from the standard-GLMs, the spatial-GLMs are preferred for analyzing the CPUE data on yellowfin tuna, especially if there is strong spatial autocorrelation among the data. This is because the spatial-GLMs took account of the

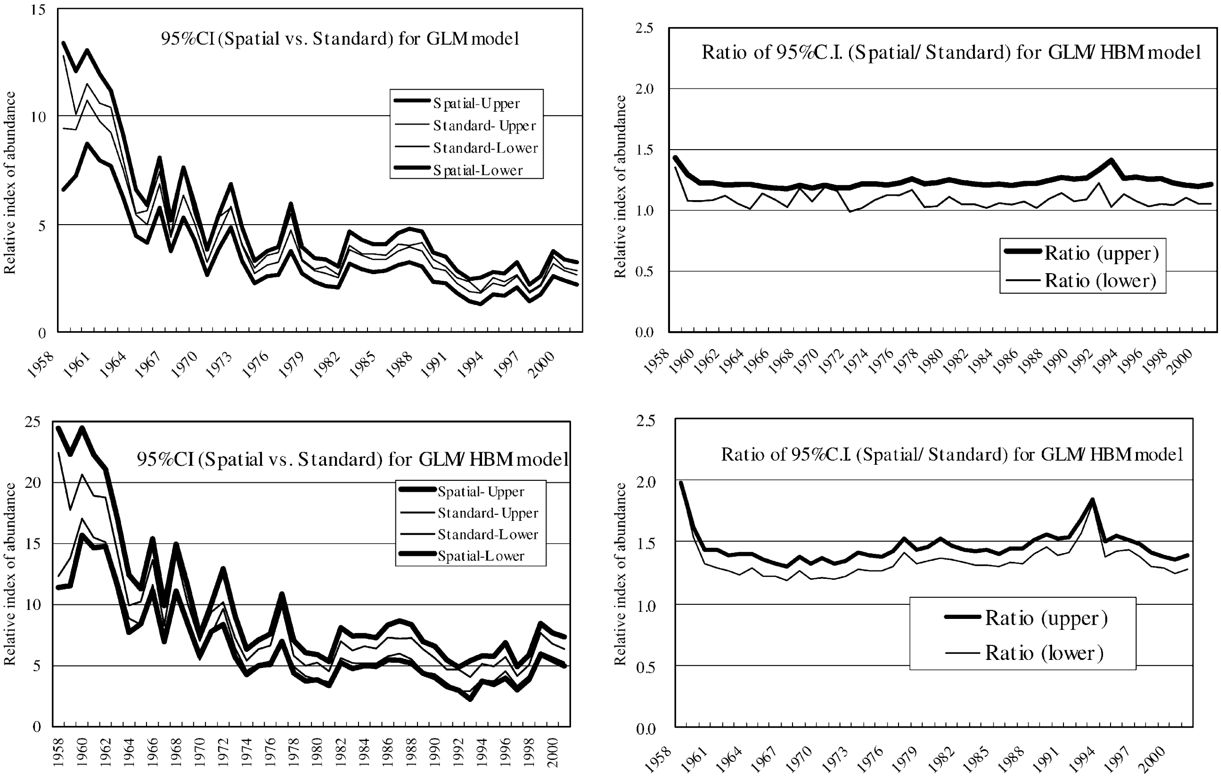

Fig. 7. A comparison of the 95% confidence intervals (C.I.) of the relative index of abundance of yellowfin tuna and ratios of 95% C.I. from the spatial-GLM with that from the standard-GLM (upper panel), and from the spatial-GLM/HBM with that from the standard-GLM/HBM (lower panel).

spatial autocorrelation effectively and yielded more realistic estimates of the variances. This is not surprising, especially considering the semivariograms and covariograms from the standard-GLMs (Fig. 3). A covariogram is a function of the distance between data points that measures how strong their spatial autocorrelation is. A positive spatial autocorrelation manifests itself in a decrease with distance to zero at some distance, where observations are no longer autocorrelated. Since the covariogram does not exist for some processes, a semivariogram is commonly used in geostatistics. It is calculated by summing up all the squared differences of the values between each pair of the points at different distances to measure the dissimilarity of the data points with distance. Their graphical representations can be used to examine the spatial correlation of the data points with their neighbors. The likelihood ratio tests (LRT) for spatial independence of H_0: $r = 0$ showed that the test statistics were highly significant in all cases (Table 3). Therefore, the spatial-GLMs perform better in analyzing yellowfin tuna CPUE data than the standard-GLMs, if the data exhibit strong spatial autocorrelation.

4.2. What are the appropriate approaches for spatial-CPUE standardization?

What are the appropriate approaches to standardize a set of spatial-CPUE data from a fish population? It is rather difficult to answer this question. In the present paper, we have only used the spatial-GLM in analyzing the data on yellowfin tuna in the Indian Ocean. Clearly, much work is needed on many other sets of data before drawing any further conclusions. Also, we coupled the spatial approach to GLMs only. Many common statistical methods, such as general additive models, regression trees, and neural networks, can be also made spatially. There is a need for searching the most appropriate method for the spatially structured CPUE data from among all possible statistical methods. This will be a very complex and difficult task, especially considering such structure and characteristics of the CPUE data considering the degree of spatial dependency and spatial patterns. Simulation may be an effective method for this evaluation. By simulating spatially structured CPUE data with varying degrees of spatial dependency and various spatial patterns and by analyzing them using various spatial and non-spatial statistical methods, we can estimate the relative indices of the abundance of the simulated population and examine the discrepancies between the true index of abundance and its estimates. In so doing, we may be able to establish some simple rules to select the most appropriate statistical model. We plan to investigate such aspects in the future.

4.3. Biases in analyzing the spatially distributed catch rates

In analyzing the spatial catch rates data of the Japanese tuna longline fisheries, Walters (2003) recognized that biases can arise from ignoring unfished areas. At the initial stage of the Japanese tuna longline fisheries in the Indian Ocean in the early 1950s, fishing grounds were limited to the tropical waters, where the abundance (CPUE) of yellowfin tuna was considerably high. By the early 1960s, the fisheries had expanded to the whole Indian Ocean, including the temperate waters, where their abundance (CPUE) was relatively low. Thus, the observed (nominal) CPUE temporal trends usually show the (apparent) biased sharp decrease during 1950s. In this connection, if we had analyzed the CPUE data from the entire period since 1952, our estimates of the relative indices of abundance, especially those in the early period of the fisheries, would have been seriously biased. Therefore, we excluded the data from the early years (1952–1957) in our analyses, to minimize the biases. In addition, further biases can also incur, if data on tuna catch rates are aggregated to the larger time-area units (Walters, 2003). This is because such aggregation masks and dilutes the real pattern of the original data. To minimize such biases, we used the optimum sampling unit ($5^\circ \times 5^\circ$ areas and month) considering the time span (1958–2001) and the area (whole Indian Ocean).

Acknowledgements

We thank *Norin Gijutsu Kaigi* (Agriculture, Forestry and Fisheries Research Council), Government of Japan, for funding the Pioneer Special Research Project (Development of the GIS-based sustainable yield techniques for tuna resources harmonized with the marine ecosystem) (2002–2004) that made this study possible. We also greatly appreciate Katsuyuki Izumi's (SAS consultant, SAS Japan) help with running the SAS PROC MIXED procedure. We thank three anonymous referees for their constructive comments on the paper.

References

Bigelow, K.A., Hampton, J., Miyabe, N., 2002. Application of a habitat-based model to estimate effective longline fishing effort and relative abundance of Pacific bigeye tuna (*Thunnus obesus*). Fish. Oceanogr. 11, 143–155.

Bigelow, K.A., Hampton, J., Miyabe, N., 2003. Sensitivity of temperature determinants in a habitat-based model for yellowfin in the western and central Pacific. ICCAT/SCRS/03/038, p. 16.

Campbell, R.A., Tuck, G., Tsuji, S., Nishida, T., 1996. Indices of abundance for southern bluefin tuna from analysis of fine-scale catch and effort data. CCSBT/SBFWS/96/16, p. 34.

Cressie, N.A.C., 1991. Statistics for Spatial Data. Wiley, New York.

Hinton, M.G., Nakano, H., 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from Japanese longline fisheries in the Pacific. Inter-Am. Trop. Tuna Comm. Bull. 21, 171–200.

ICCAT, 2003. Report of the ICCAT Working Group on Assessment Methods. ICCAT/2003/013, p. 13.

IOTC, 2002. Report of the Fourth Session of the IOTC Working Party on Tropical Tuna. IOTC-SC-02-05, p. 47.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute Inc., Cary, NC.

Nishida, T., Bigelow, K., Mohri, M., Marsac, F., 2003. Comparative study on Japanese tuna longline CPUE standardization of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean based on two methods: general linear model (GLM) and habitat-based model (HBM)/GLM combined (1958–2001). IOTC/WPTT-03-5, p. 25.

SAS Institute, 1999. SAS/STAT User's Guide, Release 8, 2nd ed. SAS Institute Inc., Cary, NC.

Walters, C., 2003. Folly and fantasy on the analysis of spatial catch rate data. Can. J. Fish. Aquat. Sci. 60, 1433–1436.

Yokawa, K., Takeuchi, Y., 2002. Estimation of abundance index of white marlin caught by Japanese longliners in the Atlantic Ocean. ICCAT/SCRS/02/060, p. 21.