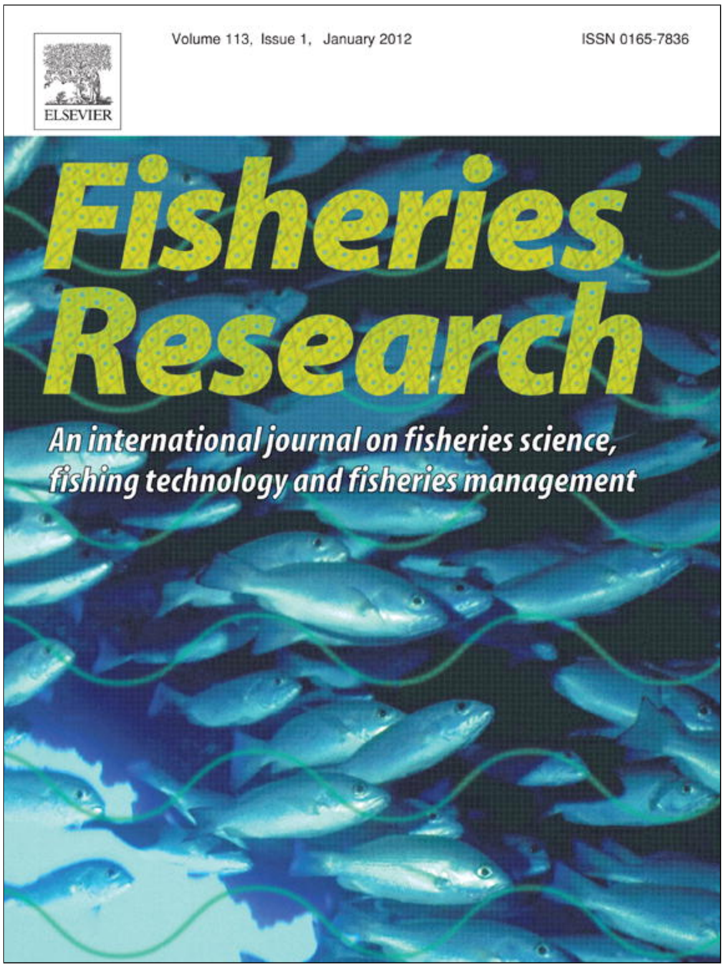

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Fisheries Research 121–122 (2012) 115–125

Contents lists available at SciVerse ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Comparison between univariate and bivariate geostatistical models for estimating catch per unit of effort (cpue): A simulation study

Julio César Pereira[a,*], Roseli Aparecida Leandro[b], Miguel Petrere Jr.[c], Tom Nishida[d]

[a] UFSCAR, Campus de Sorocaba, Rodovia João Leme dos Santos Km 110, SP 264, 18052-780 Sorocaba (SP), Brazil
[b] ESALQ/USP, Departamento de Ciências Exatas, CP 9, 13418-900 Piracicaba (SP), Brazil
[c] UNESP, Departamento de Ecologia, CP 199, 13506-900 Rio Claro (SP), Brazil
[d] National Research Institute of Far Seas Fisheries, Fisheries Research Agency, 5-7-1, Orido, Shimizu-Ward, Shizuoka 424-8633, Japan

ARTICLE INFO

Article history:
Received 2 March 2010
Received in revised form 17 January 2012
Accepted 27 January 2012

Keywords:
Bayesian inference
Geostatistical models
Cpue
Simulation

ABSTRACT

Data on catch and fishing effort are used for estimating the relative abundance of a fish stock. Geostatistical models have been used in this sort of analysis. Multivariate geostatistical models have been used in several areas but not often with fisheries data. In this paper a simulation study was carried out in order to compare an index of catch per unit of fishing effort (cpue) based on: (1) the adjustment of a univariate geostatistical model of the ratio between catch and effort; and (2) a bivariate model in which catch and effort are simultaneously modeled. The estimates obtained from both models presented close results suggesting that there is no advantage in using the bivariate model for estimating the index and indicating that the univariate model is preferred.

1. Introduction

Logbooks gathered from commercial fisheries provide huge data sets that are valuable for fish stock management (Vignaux, 1996; Walters, 2003). These data usually include: the yield or the number of individuals caught from a stock using one or more fishing gears; and the fishing effort—expressed, for instance, as days spent at sea, days spent fishing, or the number of hooks per time unit (Nadal-Egea, 1996). Logbooks for marine fisheries are usually georeferenced to predefined fishery sub-regions that are generally at a much coarser spatial scale than the fishing gear deployed. These catch (C) and effort (f) data are used in fish stock assessments (e.g., Vignaux, 1996) and are often used to calculate cpue as an index of stock abundance. Cpue may not be directly proportional to stock abundance in a given area. However, in many practical situations such an assumption is necessary. Provided catch and effort data are available for every sub-region within a fish species distribution, three cpue indices may be defined (cpue$_1$, cpue$_2$, cpue$_3$). Petrere et al. (2010) and Pereira et al. (2009) provide the following definitions: $\text{cpue}_1 = \left(\sum C_i/f_i\right)/n = \overline{(C/f)}$ which is the mean of ratios between catch and effort in different quadrats; $\text{cpue}_2 = \sum C_i/\sum f_i = \bar{C}/\bar{f}$ which is the ratio between the total catch across all quadrats and the total effort across all quadrats; and $\text{cpue}_3 = \sum C_i \cdot f_i/\sum f_i^2 = \overline{C \cdot f}/\overline{f^2}$ which is the ratio between the sum of the products of the catch by the effort in different quadrats, and the sum of the squares of the efforts.

Indices obtained by applying the above equations to fishery-dependent data should be analyzed with caution (Hilborn and Walters, 1992; Quinn and Deriso, 1999). Fisheries are known not to sample in a representative manner both spatially and temporally; and often, catch and effort data are not avaliable for all predefined areas within a stock's distributions. Walters (2003) indicates that missing cpue values must be imputed and it is critical to be explicit about how such values have been derived. In this paper, we focus on the spatial aspect of cpue imputation. We address the situation where fishing was not carried out in all quadrats in a region or where catch and effort data were not collected (or not properly recorded in logbooks). In this situation and in order to analyze data on catch and effort and calculate a cpue index for the entire region, we make assumptions based on what would have occurred in the quadrat that was not observed. Walters (2003) offers an alternative to estimating an index using only the observed data, for the same period of time, from quadrats neighboring those quadrats for which no data are available. His alternative is feasible where it is possible to show a spatial correlation structure between the catch ratios. He suggests that spatial statistics may be employed in order to interpolate data for the unobserved quadrats. He also suggests the use of covariates such as water temperature to assist the interpolation process.

Nóbrega (2008) identifies spatial dependence in catch–effort data derived from a small-scale fishery in northern and northeastern Brazil, using geostatistics for modeling and standardizing cpue.

* Corresponding author. Tel.: +55 15 3229 5973; fax: +55 15 3229 5940.
E-mail addresses: julio-pereira@ufscar.br (J.C. Pereira), raleandr@esalq.usp.br (R.A. Leandro), mpetrere@rc.unesp.br (M. Petrere Jr.), tnishida@affrc.go.jp (T. Nishida).

0165-7836/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2012.01.018

Nishida and Chen (2004) also noted the presence of spatial correlation of data for yellowfin tuna (*Thunnus albacore*) cpue, where they incorporated a spatial component in their cpue models. According to Nishida and Chen (2004), the spatial dependence occurs because for many fish species the individuals live and move together; therefore, the closer in space the observations are, the more alike they are.

In the above-cited references the authors propose the use of spatial models for the ratio between catch and effort, i.e. the quantity C/f. However, if the data available are not simply represented as C/f but can be georeferenced as catch and effort pairs (C, f), an alternative would be to model the pairs of variates (C, f), as proposed by Pereira et al. (2009). The current paper addresses this, namely: is it advantageous, when estimating an index of type cpue$_1$, to model the pair (C, f) instead of the ratio C/f when imputing values for quadrats that were not observed?

Multivariate geostatistical models have been used in several areas – for example, to model pollutants in the atmosphere (Schmidt and Gelfand, 2003), to model sales price and yield generated by real estate (Gelfand et al., 2004), and to model annual volumetric increment of forest trees (Bognola et al., 2008). Pereira et al. (2009) used a bivariate model for (C, f), which was compared with another bivariate model that did not consider the spatial dependence between C and f. Geostatistical multivariate models in such studies have demonstrated a superior performance when compared to models that do not consider the spatial correlation; and it has also been shown that such multivariate geostatistical models are superior to those geostatistical models where the variates are mutually independent. Therefore the advantage of using a multivariate geostatistical model may be justified because such a model accounts for the cross covariance structure as well as the spatial dependence for each variable (C and f). The multivariate geostatistical model takes into account the relationship between the catch in a given location (s) and the fishing effort at another location (s') and vice versa. However, the cpue$_1$ index is a function of C and f, in particular a function of C/f; and so it is questionable if it would still be advantageous to use a bivariate model for C and f instead of a geostatistical univariate model for the quantity C/f.

To examine these two analytical options, a geostatistical simulation was performed comparing two model possibilities: a univariate approach considering the ratio between catch and effort; and a bivariate approach (Gelfand et al., 2004) that modeled catch and effort simultaneously. In both approaches temperature is considered as a covariate. After adjusting the univariate model, interpolations for the unobserved quadrats were calculated. When interpolating the ratio between catch and effort it is only possible to estimate the first—that is cpue$_1$—of the three indices shown above. The bivariate model was also adjusted, interpolations were calculated for the catch and effort variates, and then cpue$_1$ was estimated.

The main purpose of this paper is to use a simulation study to perform comparisons based on different scenarios:

- the behavior of cpue$_1$ estimates, calculated after fitting both models; and
- cpue$_1$ estimates calculated after fitting the univariate geostatistical model using observed estimates that are obtained only through observed data, not predicting for unobserved/unreported quadrats.

The paper's presentation of a simulation study that addresses the respective outcomes of a univariate and a multivariate approach to cpue index imputation should provide a useful contribution to fishery science.

2. Material and methods

A simulation study was employed to compare a univariate model for the ratio between catch (C) and effort (f) and a bivariate model for catch and effort, for estimating the cpue$_1$ index for a given region.

2.1. Spatial univariate model (SUM) for C/f

The spatial univariate model (SUM) for the ratio between catch (C) and effort (f) is given by:

$$Y(s) = \mathbf{X}(s)\boldsymbol{\beta} + \sigma w(s) + \tau_2 u(s), \tag{1}$$

in which $Y(s) = \ln((C/f)(s))$, $\mathbf{X}$ is the design matrix containing the covariates effect, $w(\cdot)$ follows a Gaussian process with mean 0 and variance 1, and $u(s) \sim N(0,1)$, and the correlation function adopted was the exponential $\rho(d) = \exp(-\phi d)$. The ϕ parameter of this function determines how quickly the correlation drops to zero. Considering that $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))$ forms a partial random sample of the Gaussian process $\{Y(s) : s \in G\}$, the joint distribution of $Y(s_1), \ldots, Y(s_n)$ is n-variate normal, that is, $\mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\beta, \sigma^2\mathbf{R} + \tau^2\mathbf{I}_n)$, in which $R_{ii'} = \rho(s_i, s_{i'})$ and $\mathbf{I}_n$ is the n order identity matrix. We have also considered the temperature (*temp*) as a covariate; thus the ith line of matrix $\mathbf{X}\boldsymbol{\beta}$ is given by $\beta_0 + \beta_1\, temp(s_i)$. The joint distribution considered a function of the parameters through observation; thus the likelihood of the SUM model may be expressed as:

$$L(\boldsymbol{\beta}, \phi, \sigma^2, \tau^2|\mathbf{y}) = (2\pi)^{-n/2}|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-1/2} \times \exp\left(-\frac{1}{2}[\mathbf{y} - \mathbf{X}\beta]^T[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]^{-1}[\mathbf{y} - \mathbf{X}\beta]\right).$$

As shown in the geostatistical literature, it has been assumed that the SUM model parameters are prior independent; that is, the joint prior of the parameters is given by the product of the individual priors of each parameter and informative priors were adopted for σ^2 and τ^2, in other words, inverse gamma priors, $\sigma^2 \sim IG(a_\sigma, b_\sigma)$, $\tau^2 \sim IG(a_\tau, b_\tau)$. According to Banerjee et al. (2004), inverse gamma is a usual candidate for these parameters; in order to fix the hyperparameters of this prior their prior means were set in the minimum-squares estimates of the model variance ($\hat{\sigma}^2$) without the spatial component adjusted to the observed data with infinite variance. For example we fixed $E(\sigma^2) = \hat{\sigma}^2/2$ and $V(\sigma^2) = \infty$ as the mean and the variance of the inverse gamma distribution, with parameters a_σ and b_σ given by $E(\sigma^2) = b_\sigma/(a_\sigma - 1)$ and $V(\sigma^2) = b_\sigma^2/((a_\sigma - 1)^2(a_\sigma - 2))$, so we get the values of a_σ and b_σ. And for the ϕ parameter of the correlation function exponential, a gamma prior, frequently utilized, was adopted: $\phi \sim G(a_\phi, b_\phi)$, with high prior variance, and a mean such that when the correlation is 0.05, the d_0 value (effective range) is equal to half of the maximum distance (*max.dist*) between observed locations. This gives us $E(\phi) \approx 6/max.dist$, as shown in Appendix A. This prior reflects the belief that for distances larger than *max.dist*/2 the spatial correlation is close to zero. Parameter $\boldsymbol{\beta} = (\beta_{01}, \beta_{02})'$ received a normal bivariate prior with mean 0 and variances matrix $\sigma_\beta^2\mathbf{I}_2$, $\sigma_\beta^2 = 100$ (flat prior).

Following the Bayesian approach, the joint posterior distribution is proportional to the product of likelihood by the prior, that is:

$$\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \phi|\mathbf{y}) \propto |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-1/2} \times \exp\left(-\frac{1}{2}[\mathbf{y} - \mathbf{X}\beta]'[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]^{-1}[\mathbf{y} - \mathbf{X}\beta]\right) \times \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\phi) \tag{2}$$

As it is not possible to get posterior summaries directly from a joint posterior distribution (Eq. (2)) it is necessary to use Monte

Carlo Markov Chain (MCMC) algorithms (Gamerman and Lopes, 2006). Conditional complete distributions are necessary for the implementation of MCMC algorithms, as shown in Appendix B. The models were run using WinBUGS software (Spiegelhater et al., 2002).

2.2. Spatial bivariate model (SBM) for catch and effort

The following bivariate spatial model, proposed by Pereira et al. (2009) for the catch and effort joint modeling, was compared to a bivariate model without a spatial component.

The bivariate model, with vector $\mathbf{Y}(s)$ of $p=2$ dimension, was considered for joint modeling in which $Y_1=\ln(f)$ and $Y_2=\ln(C)$, where ln is the natural logarithm. Thus, it was assumed that the observation vectors at sample points $\mathbf{Y}(s_i)$, $i=1, 2, \ldots, n$ form a partial realization of the stochastic process $\{Y(s): s\in G\}$, $G\subset \Re^2$ and that the random vector $\mathbf{Y}(\cdot)$ follows a Gaussian process. A model for these data is described as:

$$\mathbf{Y}(s) = \mathbf{X}(s)\boldsymbol{\beta} + \mathbf{v}(s) + \varepsilon(s), \tag{3}$$

in which $\mathbf{v}(s)$ follows a normal bivariate distribution with a vector of means ($\mathbf{0}$), and a covariance matrix 2×2; $\varepsilon(s)$ is a white noise vector with normal distribution, covariances matrix $\mathbf{D}$, of dimension 2, diagonal, so that $D_{jj}=\tau_j^2$; and $\mathbf{X}(s)\boldsymbol{\beta}$ represents the trend of the process in which $\mathbf{X}(\cdot)$ is a design matrix, describing the possible covariate.

The model used for the covariance structure of the $\mathbf{v}(s)$ component of Eq. (3) was the linear coregionalization model proposed by Gelfand et al. (2004) and used by Pereira et al. (2009). For this model one has a range (i.e., the distance beyond which there is practically no spatial correlation between data points) associated with each component of $\mathbf{Y}(\cdot)$ if monotonic and isotropical correlation functions are used. The correlation function was an exponential function, both for catch and for effort: $\rho_j(d)=\exp(-|\phi_j d|)$, $j=1, 2$, in which d is the distance between any two points s, s'. The linear coregionalization model for the $\mathbf{v}(s)=\mathbf{Aw}(s)$ process may be re-parameterized according to a conditional approach. Gelfand et al. (2004) shows that there is an equivalence between model (3) written in the unconditional way and the conditional way. Thus, the model for the two variates, and taking into consideration the covariate, temperature (*temp*), written in the conditional way is given by:

$$\begin{array}{l} Y_1(s) = \beta_{01} + \beta_{11}\, temp(s) + \sigma_1 w_1(s) \\ Y_2(s)|Y_1(s) = \beta_{02} + \beta_{12}\, temp(s) + \alpha Y_1(s) + \sigma_2 w_2(s) + \tau_2 u_2(s) \end{array}. \tag{4}$$

From now on the model in Eq. (4) will be called the spatial bivariate model (SBM).

Using the conditional approach the likelihood may be shown as:

$$L(\mathbf{Y}|\boldsymbol{\theta}) = L(Y_1|\boldsymbol{\theta}_1)L(Y_2|Y_1, \boldsymbol{\theta}_2),$$

in which $\boldsymbol{\theta}$ is the parameter vector of the model; $\boldsymbol{\theta}_1=(\boldsymbol{\beta}_1, \sigma_1^2, \phi_1)'$, is the parameter vector referring to the first equation of model (SBM); and $\boldsymbol{\theta}_2=(\boldsymbol{\beta}_2, \sigma_2^2, \phi_2, \tau_2^2)'$ referring to the second line in Eq. (4). So,

$$L(\boldsymbol{\beta}_1, \phi_1, \sigma_1^2|y_1) = (2\pi)^{-n/2}|\sigma_1^2\mathbf{R}_1|^{-1/2}\exp\{-(1/2)[\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1]^T \times [\sigma_1^2\mathbf{R}_1]^{-1}[\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1]\},$$

in which $\mathbf{X}_1$ is the design matrix for $Y_1(s)$, $\boldsymbol{\beta}_1$ is the covariate coefficient vector and $(\mathbf{R}_1)_{ii'}=\rho_1(s_i, s_{i'})$.

$$L(\boldsymbol{\beta}_2, \phi_2, \sigma_2^2, \tau_2^2|y_2, y_1) = (2\pi)^{-n/2}|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-1/2} \times\exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2]^T[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2]\right\}$$

in which $\mathbf{X}_2$ is the design matrix for $Y_2(s)$, $\boldsymbol{\beta}_2$ is the covariate coefficients vector and $(\mathbf{R}_2)_{ii'}=\rho_2(s_i, s_{i'})$.

Using the Bayesian approach, it is necessary to incorporate relative uncertainty to the parameters of interest, thus assuming independence; the prior joint distribution is given by the product of the prior of each parameter. For $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ vectors, multivariate normal prior were attributed, with a vector of $\mathbf{0}$ mean and covariances matrix $\mathbf{I}\sigma_\beta^2$, with σ_β^2 fixed at a high value, $\sigma_\beta^2=100$ (flat prior). Inverse gamma prior were used in parameters σ_1^2 and σ_2^2, with averages equal to the minimum-squares estimates from a spatially independent model for each variate, $Y_1(s)$ and $Y_2(s)$, with infinite variance, that is, $\sigma_j^2 \sim IG(a_{\sigma_j}, b_{\sigma_j})$, with $a_{\sigma_j}=2$. An inverse gamma prior was also used for τ_2^2, $\tau_2^2\sim IG(a_{\tau_2}, b_{\tau_2})$, with infinite variance ($a_{\tau_2}=2$) and average (b_{r_2}) equal to the estimate of minimum squares. According to Schmidt and Gelfand (2003), Banerjee et al. (2004) and Paez et al. (2005), Gamma priors were associated with parameters ϕ_1 and ϕ_2, $\phi_j\sim G(a_{\phi_j}, b_{\phi_j})$, whose hyperparameters a_{ϕ_j} and b_{ϕ_j} were fixed in the same way as presented for the univariate model. Parameterization of prior distributions follows Gelman et al. (2004).

Dealing with the conditional model expressed by Eq. (4), and due to the independence of its likelihood function and with the use of the previously specified prior, the joint posterior distribution for the parameters is given by $\pi(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2)=\pi(\boldsymbol{\theta}_1|\mathbf{y}_1)\pi(\boldsymbol{\theta}_2|\mathbf{y}_2)$, in which

$$\pi(\boldsymbol{\theta}_1|\mathbf{y}_1) = \pi(\boldsymbol{\beta}_1, \sigma_1^2, \phi_1|\mathbf{y}_1) \propto L(\boldsymbol{\beta}_1, \sigma_1^2, \phi_1|\mathbf{y}_1)\pi(\boldsymbol{\beta}_1)\pi(\sigma_1^2)\pi(\phi_1), \tag{5}$$

and

$$\pi(\boldsymbol{\theta}_2|\mathbf{y}_2) = \pi(\boldsymbol{\beta}_2, \sigma_2^2, \tau_2^2, \phi_2|\mathbf{y}_2, \mathbf{y}_1) \propto L(\boldsymbol{\beta}_2, \sigma_2^2, \tau_2^2, \phi_2|\mathbf{y}_2, \mathbf{y}_1) \times \pi(\boldsymbol{\beta}_2)\pi(\sigma_2^2)\pi(\tau_2^2)\pi(\phi_2). \tag{6}$$

Since the posterior distributions (5) and (6) do not have a closed analytical form, MCMC methods (Gamerman and Lopes, 2006) were used in order to get a sample of the joint posterior distribution of the parameters. For computational implementation of these algorithms it is necessary to know the complete posterior conditional distributions of all the parameters of interest present in the model, and so Appendix C shows complete conditional distributions of each parameter (SBM).

While obtaining samples from the joint posterior distribution of the parameters of both the SBM and the SUM models, we used the computer program WinBugs (Spiegelhater et al., 2002) and we used chain sizes of 55,000 in which the first 5000 were discarded (burn-in); and following on from this on observations of 50 in 50 (thinning) were stored in order to minimize autocorrelation problems, yielding a sample size of 1000. The verification of the chains convergence was performed by graphical analysis of its trace.

Clues of convergence are obtained when, starting from a given number of iterations of the MCMC algorithm, the traces of the two chains (which are generated from different seed values) overlap and start to oscillate around a constant value.

2.3. Simulation

The simulation was performed in the same way as presented in Pereira et al. (2009). Firstly temperature values in a regular grid (10 × 10) were simulated taking into account a gradient that allowed the temperature to vary between 0 °C and 25 °C. Considering four different scenarios, fifty data sets simulating the logarithm of the fishing effort (Y_1) as well as those of the logarithm of the respective catches (Y_2) were performed. Temperature values were used in a Gaussian process mean, using Y_1 and Y_2 according to Eq. (4). The scenarios were:

(a) low correlation between catch and effort, and low spatial correlation;
(b) strong correlation between catch and effort, and low spatial correlation;
(c) low correlation between catch and effort, and strong spatial correlation; and
(d) strong correlation between catch and effort, and strong spatial correlation.

Each of the 100 points from the regular grid (10 × 10), which is shown in Pereira et al. (2009), represents a reference point within a quadrat where fishing had occurred (in other words, catch and effort). It should be noted that in a given region divided into quadrats and inhabited by a fish stock, fishing is not necessarily observed in all quadrats (Walters, 2003). In our study, data were simulated in all quadrats; so they represent the effort and its respective capture that would be observed had the fisheries been observed in all of them.

As catch and effort are considered random variates in each quadrat they might present different results; and for this purpose 50 data sets were generated to simulate different realizations of these variates in each quadrat. As a result, in each simulated run there were 100 pairs of effort and catch data, and from these a stock-abundance index was calculated, given by $cpue_1 = (1/100)\sum_{i=1}^{100} C_i/f_i$. The resulting value is the true value of the index for the present realization. Then, for each realization a sample size of 85 effort and catch pairs and a second sample size of 76 pairs were randomly selected. That is, two situations were considered: one in which 15% of the quadrats were considered as not observed and another in which 24% of the quadrats were considered as not observed. The simulated points considered as observed were pre-established: they were the same for the 50 realizations, and for this a Thomas process was used (Reis, 1998) to mimic a real situation in which the fishing spatial pattern is generally aggregated (Anganuzzi, 2004), in the same way as presented in Pereira et al. (2009). Note that when considering only at the observed locations (i.e., locations used for model fitting) in both situations, the maximum observed distance between the points remains approximately the same (14 units).

For each observed set (85 and 76 points) in each of the 50 realizations, two models were applied:

- SBM model from which data were simulated. The response variates are the logarithm of the fishing effort (Y_1) and the logarithm of the catch (Y_2); and
- SUM model. In this case the response variate is the logarithm of the ratio between catch and effort, and the generated data are from Y_1 and Y_2 in order to enable fitting of the SUM model. Thus, the difference $Y = Y_2 - Y_1$, was considered, which results in $Y = \ln(C/f)$ due to their logarithmic property. We are only able to generate Y_1 and Y_2 by simulation, whereas the variate of interest is ln(*C*/*f*). The model was fitted for this variate because several papers (such as Nishida and Chen, 2004; Matsunaga et al., 2006) use *Y* when modeling data on effort and catch.

For both models the Bayesian approach was used for the fitting (i.e. parameters estimation) via MCMC methods. From this, a sample of the joint posterior distribution of the parameters is obtained:

Table 1
Potential scale reduction (Gelman and Rubin, 1992) of two posterior chains of the parameters, starting from distinct initial values; using the univariate model, under a scenario of strong correlation between catch and effort, and strong spatial correlation; and for a sample size of 85 of the 'observed' quadrats.

Parameter	Point estimate ($\hat{R}$)	97.5% quantile
τ^2	1.0108	1.0204
σ^2	1.0651	1.0871
β_0	1.0002	1.0014
β_1	1.0026	1.0115
ϕ	1.0024	1.0025

- when fitting the SBM model, for the unobserved quadrats samples were obtained from the catch and effort predictive distribution; and
- when fitting the SUM model, for the unobserved quadrats samples of predictive distribution of the ratio between catch and effort were obtained.

For each predicted set in the unobserved sites, those values considered as observed (*C* and *f* simulated values used for model fitting) were added to its respective set, resulting in a data set composed of predicted and observed values, which were used in order to estimate $cpue_1$. So, for example, estimates for $cpue_1$ using the *SBM* were obtained by $cpue_{1\,SBM} = 1/(m+K)\sum_{i=1}^{m+K} C_i/f_i$, where $(C_i f_i)$, $i = 1, 2, \ldots, m$, $m = 76$ or 85, are the observed pairs; and $(C_i f_i)$, $i = m + 1, \ldots, m + K$, where $m + K = 100$, are the predicted pairs via the SBM model. Similarly estimates were obtained using the SUM model; however with that model we have predictions of *C*/*f* instead of pairs (*C*,*f*). A sample of estimates was obtained using each applied model, and the median of the sample gave a point estimate. Thus, for each one of the 50 realizations a point estimate for $cpue_1$ was obtained after application of the SBM model, and another was obtained after application of the SUM model. We refer to the $cpue_1$ estimates obtained after the fitting and predictions by the two models as 'adjusted $cpue_1$ estimates' and those obtained just from data considered as observed 'sampling-based estimates' (or sampling-based $cpue_1$), that is, $cpue_{1\,sampling\ based} = (1/m)\sum_{i=1}^{m} C_i/f_i$, $m = 85$ or 76; and they are further compared via calculation of the mean square error (MSE).

3. Results

Fig. 1 represents the traces of two posterior chains generated from distinct initial values for the parameters of the SUM model. Thus this model was implemented with a sample size of 85 of the 'observed' quadrats, under the fourth (d) of the scenarios of correlation-intensity combinations outlined above, namely strong correlation between catch and effort, and strong spatial correlation. The graphs allow us to see that there are convergence of the chains. The same was noted for the other correlation-intensity-combination scenarios, and also for the re-application of this model with a sample size of 76 for the four simulated scenarios (Pereira, 2009). For the SBM model, applied with a sample size of 85 and 76 of the 'observed' quadrats, a convergence of chains was also diagnosed for all of the scenarios; however, the chains in these cases present a similar behavior so are not shown here.

In addition to the graphical analysis of the traces of chains generated from the model input parameters, the Gelman and Rubin (1992) convergence criterion was also used for verification of convergence. This criterion is based on techniques of analysis of variance and consists of comparing the dispersion between and within the chains. Using at least two chains, one calculates the potential scale reduction ($\hat{R}$). The convergence can be evaluated by the proximity of $\hat{R}$ to 1. Gelman (1996) suggested accepting convergence when the value of $\hat{R}$ is below 1.2. Table 1 shows the

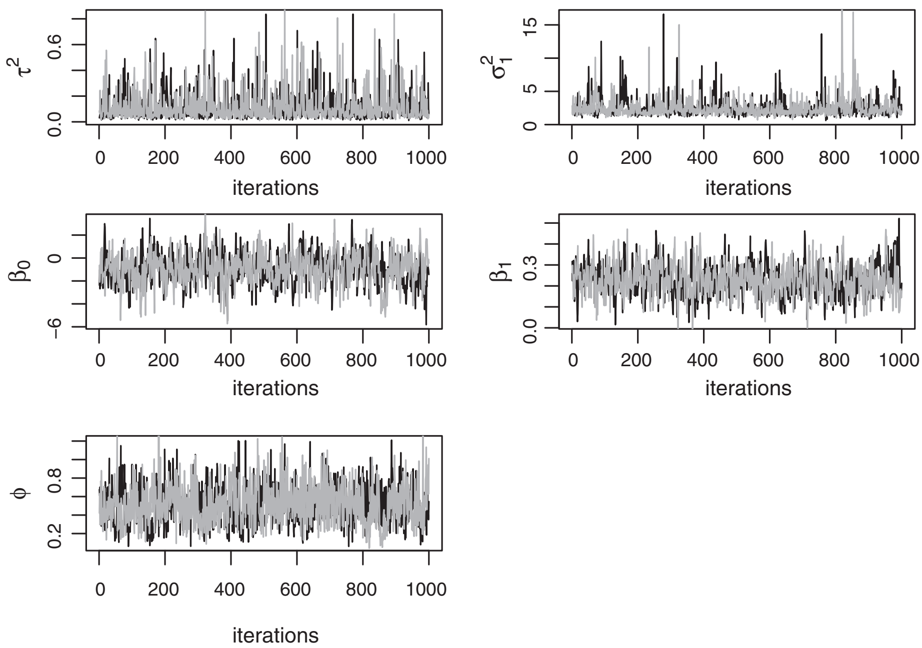

Fig. 1. Trace of two posterior chains of the parameters (represented in gray and black), starting from distinct initial values; using the univariate model, under a scenario of strong correlation between catch and effort, and strong spatial correlation; and for a sample size of 85 of the 'observed' quadrats.

Gelman and Rubin (1992) potential scale reduction of the SUM model parameters whose chains were presented in Fig. 1. According to this criterion the convergence is acceptable, since the $\hat{R}$ values are below 1.2. Using the same criterion, the convergence of the chains of the SUM model parameters was diagnosed in other simulated scenarios, and it was also diagnosed for the parameters of the SBM model in the four simulated scenarios.

Fig. A.1, Appendix D [for a sample size of 76 and scenario (d)] shows 95% credible intervals together with the values used in the simulation for the parameters of the SBM model. Almost all intervals include the values of the parameters used in the simulation. This indicates that the parameters are well estimated, which is important to verify since they are inferences from the estimated model. There are no true values of parameters for the SUM model because the data were generated from the SBM model and, thus, it is impossible to see if 95% intervals cover (enclose) the true values.

Fig. 2 presents a graph with values for cpue_1 shown relative to sampling-based estimates and to adjusted cpue_1 estimates from the univariate model, under the fourth correlation-intensity-combination scenario (i.e. d).

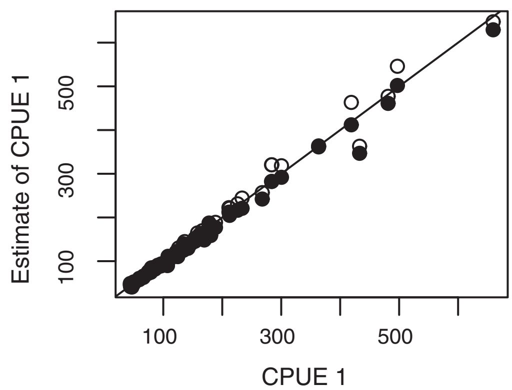

Fig. 2. The cpue_1 values relative to sampling-based estimates (in white) and adjusted estimates (in black) obtained using the univariate model; under a scenario of strong correlation between catch and effort, and strong spatial correlation; and for a sample size of 85 of the 'observed' quadrats.

Fig. 2 shows that the sampling based estimates are a little more scattered in relation to the true values for cpue_1 than are the adjusted estimates (which are closer to the straight line shown in the graph). This is reflected in Table 2 where, just for the fourth correlation-intensity scenario (i.e. d), the adjusted estimates using the SUM model are better than the sampling-based estimates.

Fig. A.2 in Appendix E shows histograms of the adjusted cpue_1 estimates and their respective medians arising from the application of the univariate model, together with the sampling-based estimates and the true value for cpue_1. The histograms provide a description of the uncertainty associated with the adjusted estimates as a representation of cpue_1.

Table 2 provides a comparison of the MSEs of the cpue_1, and the estimates that were calculated after fitting the univariate model [for which the response variate is the logarithm of the *C*/*f* (SUM model)], with the MSEs of estimates calculated after application of the SBM model [in which the response variate is the pair (*f*, *C*)]. The table shows that, despite the MSEs of the SBM model being lower for the 4 scenarios, the differences between the MSEs in estimates from both models are very small. This is also noted in Fig. 3

Table 2
Mean square error (MSE) of the adjusted estimates and sampling-based estimates for each one of the scenarios using the spatial bivariate model (SBM) and spatial univariate model (SUM) with a sample size of 85 of the 'observed' quadrats.

Scenarios	Adjusted		Sampling-based estimates
	SUM	SBM	
(a)	35.4559	35.1504	32.6330
(b)	475.4903	473.4891	426.4459
(c)	5.2845	5.2285	3.5032
(d)	245.6568	242.7253	283.2965

(a) low correlation between catch and effort and low spatial correlation;
(b) strong correlation between catch and effort and low spatial correlation;
(c) low correlation between catch and effort and strong spatial correlation; and
(d) strong correlation between catch and effort and strong spatial correlation.

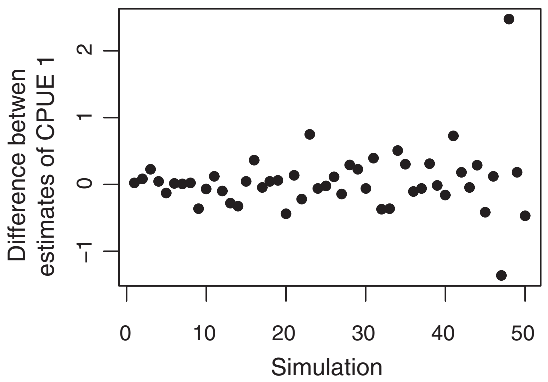

Fig. 3. Differences between the adjusted estimates of $cpue_1$ using univariate and bivariate models in each simulation, under a scenario of strong correlation between catch and effort, and strong spatial correlation, and for a sample size of 85 of the 'observed' quadrats.

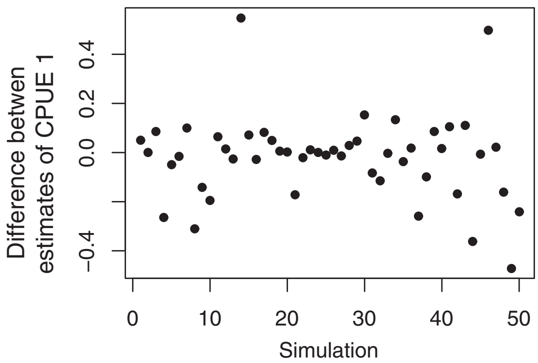

Fig. 4. Differences between the adjusted estimates of $cpue_1$ using univariate and bivariate models in each simulation, under a scenario of strong correlation between catch and effort, and strong spatial correlation, and for a sample size of 76 of the 'observed' quadrats.

which shows that the differences between the estimates obtained from both models are quite small and distributed around zero. Furthermore the SBM model is more complex, because in this case it involves the estimation of 10 parameters, while the SUM model has five parameters.

When more points were left for prediction (24 points; or, in other words, when the models were adjusted using a sample size of 76), the MSEs calculated from the estimates for $cpue_1$ (Table 3) increased in relation to those calculated from a sample size of 85 (Table 2).

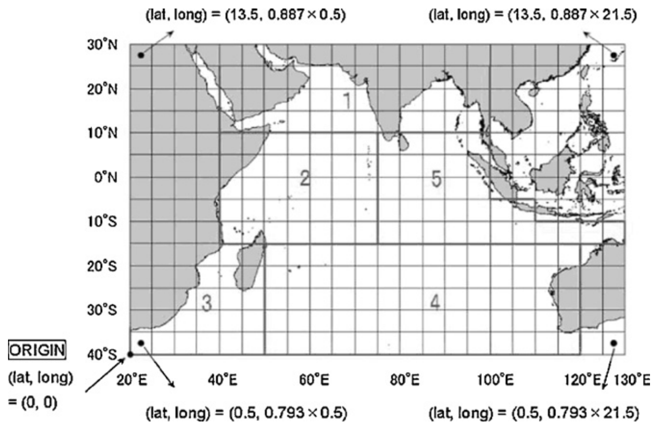

Fig. 5. Definition of the coordinate system for computing the distances between two 5° × 5° areas (the distance between 5 degrees of latitude at the Equator is set to 1). The five sub-areas were adopted by the Indian Ocean Tuna Commission (2002) for standardizing yellowfin-tuna longline-cpue data in the Indian Ocean.
(figure from Nishida and Chen, 2004).

Table 3
Mean square error (MSE) of the adjusted estimates and sampling-based estimates for $cpue_1$ for each one of the scenarios, through the use of the spatial bivariate model (SBM) and spatial univariate model (SUM) with 24 grid-data points left for prediction.

Scenarios	Adjusted		Sampling-based estimates
	SUM	SBM	
(a)	53.6231	53.5565	44.8562
(b)	1183.7940	918.5675	692.1623
(c)	24.8095	25.1901	17.0452
(d)	1397.8230	1322.8330	938.0567

(a) low correlation between catch and effort and low spatial correlation;
(b) strong correlation between catch and effort and low spatial correlation;
(c) low correlation between catch and effort and strong spatial correlation; and
(d) strong correlation between catch and effort and strong spatial correlation.

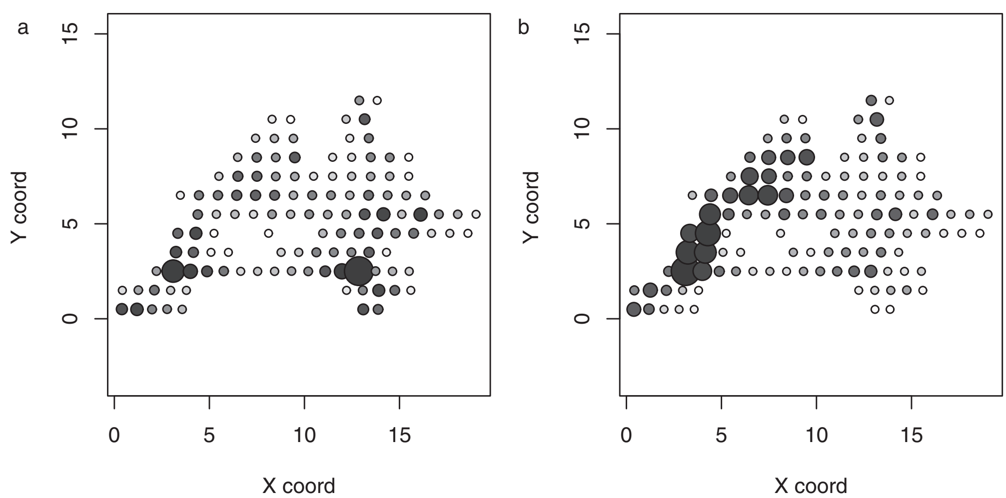

Fig. 6. Spatial distribution of the effort (a) and catch (b) data. The variate values are proportional to circle diameter, and the 4 different gray tonalities represent the division of the observations, sorted in ascending order, into four parts according to quartiles (1°, 2° and 3°) increasing in order from the lightest to the darkest.

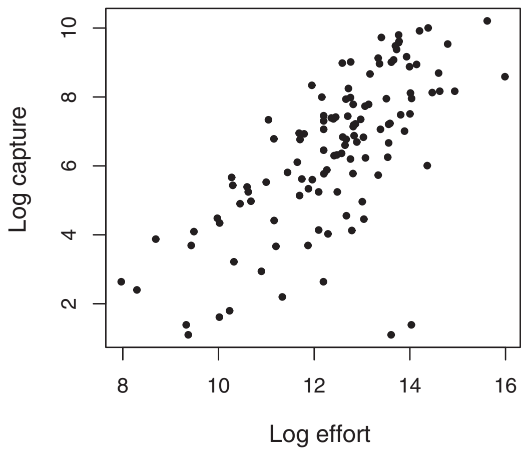

Fig. 7. Scatterplot of ln(catch) × ln(effort).

Table 2 shows that with a sample size of 85, in scenario (d), the adjusted estimates of $cpue_1$ using the SUM and SBM models were better than the sampling-based estimates. However, this did not occur with a sample size of 76 (Table 3), and for this sample size for all scenarios, the sampling-based estimates showed the lowest values of MSE compared with MSEs for the adjusted estimates via the SUM and SBM models.

Fig. 4 shows the difference between the adjusted estimates that were obtained using both models in a scenario of strong correlation between catch and effort, and strong spatial correlation, and with a sample size of 76 of the 'observed' grids. Adopting a sample size of 85 'observed' grids produced a similar result – that is, the difference between estimates from both models is very small, distributed around zero, producing MSEs that were very close (Table 3).

3.1. Application

An application of the theory was carried out using data for 2001 from Japanese yellowfin tuna (*T. albacores*) longline fishing in the Indian Ocean. The data set consists of the variates fishing effort (expressed in the number of hooks) and catch (in number of individuals) per year for a 5° × 5° quadrat. There were 118 pairs of fishing effort and catch data available and for each pair there was a reference point representing the quadrat where the fishery took place. The sub-areas adopted by the Indian Ocean Tuna Commission Working Party on Tropical Tunas (WPTT) (IOTC, 2002) were also incorporated in the data. For ecological reasons (habitat), sub-areas 1, 2, 3, 4 and 5 indicated in Fig. 5 were used as covariates (cofactor, i.e., categorical factor) because they were needed to help to explain the effort and catch variability. A similar example using a cofactor in a geostatistical model may be seen in Oliveira et al. (2006), where they utilize three different regions when modeling soil calcium content.

Fig. 5 also shows the 5° × 5° quadrats where the fishery took place. A system of coordinates was set for the studied area, such that the point 20°E, 40°S was adopted as the origin, and 5-degrees of latitude at the Equator represented a unit of distance. In this way, coordinates (x,y) were obtained for the central points of each quadrat (more details can be found in Nishida and Chen, 2004).

Fig. 6 shows the spatial distribution of catch and effort data where a descriptive analysis of the data showed that the statistical distributions of the effort and catch variates were quite asymmetrical, and so data were converted to a log scale.

The ln(catch) × ln(effort) presented a linear correlation of $r = 0.7$ ($n = 118$, $P < 0.001$). Fig. 7 shows the scatterplot of ln(catch) × ln(effort).

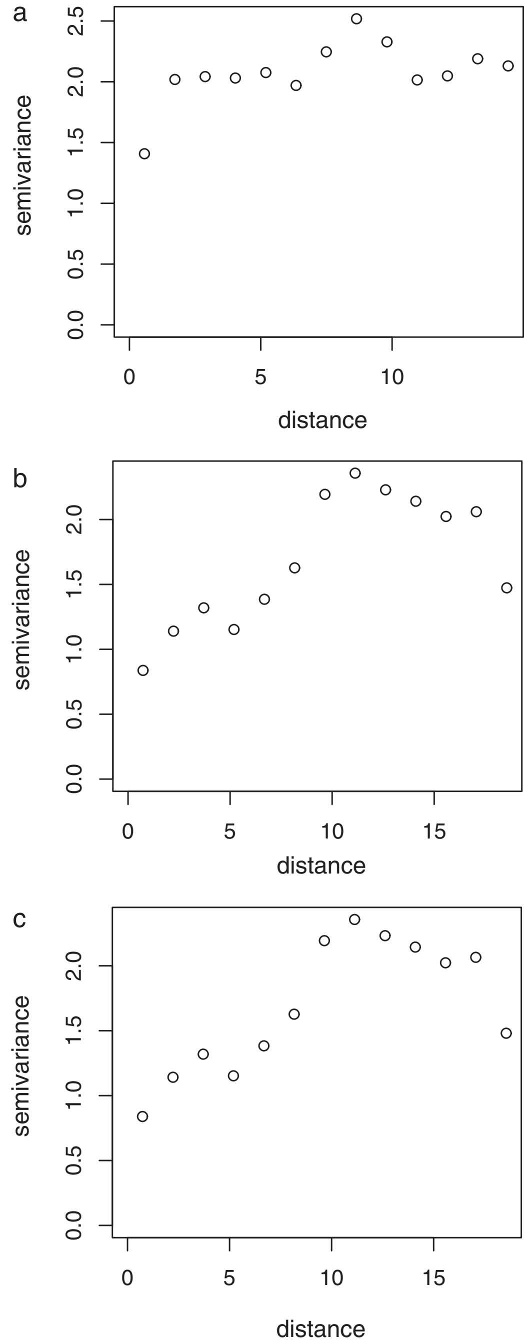

Fig. 8. Empirical semivariograms: (a) of the residual of a linear model for ln(effort) as a function of the cofactor habitat; (b) of the residual of a linear model for ln(catch) as a function of ln(effort) and habitat; and (c) of the residual of a linear model for ln(catch/effort) as a function of the habitat.

In order to determine the spatial dependence in the data, Fig. 8 presents empirical semivariogram residuals obtained from fitting a linear model to the variates: ln(effort), removing the habitat effect; ln(catch) removing the effects of ln(effort) and habitat effect; and ln(catch/effort) removing the habitat effect.

It was observed that the catch and effort data exhibit spatial dependence and the data set is in agreement with the fourth (d) of the scenarios of the simulation study. The variates are strongly correlated, there is spatial correlation, and spatial correlation for the variate ln(catch/effort) is also observed.

The data were subjected to the models used in the simulation study (SBM and SUM), incorporating sub areas A_i, $i = 1, 2, 3, 4$ and 5

Table 4
Deviance Information Criterion for the spatial bivariate model (SBM), spatial bivariate model without habitat (SBMSH), the spatial univariate model (SUM) and the spatial univariate model without habitat (SUMSH) for sample sizes of 90 and 100.

Sample size	Model			
	SBM	SBMSH	SUM	SUMSH
$n = 100$	577.568	583.3	222.876	225.7
$n = 90$	498.490	503.9	179.792	183.8

Table 5
Mean square of the predicted values of catch rate using the spatial bivariate model (SBM) and spatial univariate model (SUM), for sample sizes of 90 and 100.

Sample size	Model	
	SBM	SUM
$n = 100$	1.49e−05	1.59e−05
$n = 90$	2.54e−05	2.56e−05

(shown in Fig. 5), as a covariate. The modeling adopted a Bayesian approach as in the simulation study, and the Deviance Information Criterion (DIC) (Spiegelhater et al., 2002) was used to compare the fit of the models—smaller values of DIC indicate a more parsimonious model.

When running the models with the data set of 118 effort and catch pairs, the data utilized were adjusted as shown in the simulation runs. Sample sizes of 90 and 100 data pairs (representing omissions of 15% and 24% of the data set) were adopted for the inference process, enabling an evaluation of the predictive capacity of the models.

Table 4 presents the DIC values for the SBM and SUM models, run using the 90 and 100 sample sizes. It shows that for both sample sizes the SUM model provided lower DIC values, indicating a better fit.

In order to verify the contribution of the habitat the SBM and SUM models were also fitted without this cofactor (called SBMSH and SUMSH). The obtained DIC values are shown in Table 4. This allowed us to verify that the factor ‘habitat’ contributed to explain the variates C, f and C/f; higher DIC values were obtained from the respective models without this cofactor than from the models with it.

After fitting the models for predicting distribution, the C/f values were estimated for those locations which were excluded from the inference process. To identify the model with the highest predictive capacity, the MSE of each model was calculated for each sample size.

Table 5 shows the MSE values for the predictions using the two adjusted models and the two sample sizes. Smaller values indicate better predictions. The table indicates that even though the SBM model presents lower MSE values, the values are quite close for both models.

For this data set the SUM model presented the best fit. However, in terms of predictive capacity both models presented very close results, indicating that there is no advantage in adopting the SBM model in preference to the SUM model when predicting C/f and, consequently, estimating $cpue_1$.

4. Discussion

The fact that, under the fourth (i.e. d) of the correlation-intensity scenarios and with a sample size of 85, the fit using the univariate and bivariate models gives better estimates than the sampling-based estimates, may be due to the strong spatial correlation in the simulated data in this scenario, which was taken into consideration when calculating the adjusted estimates. Moreover, these results indicate that for this scenario it is an advantage to use the adjusted estimates; and besides, when using adjusted estimates, the credible intervals for $cpue_1$ may be directly obtained from the posterior samples from the $cpue_1$ estimates. Thus, a description of the uncertainty is obtained.

Estimates obtained after applying SBM and SUM models are quite close as both are obtained in the first instance through a common set of 85 observed effort and catch pairs; only 15 different values are predicted by both models. However, the 15 values predicted by the SBM model do not make the $cpue_1$ estimates calculated after fitting the SBM model better than the estimates obtained after fitting the SUM model. Perhaps the estimates calculated after adjustment of the SBM model are not better because with this model two variates (effort and catch) must be predicted, while with the SUM model just one variate (C/f) is predicted. Thus, uncertainties are associated with effort and catch prediction in the SBM model while with the SUM model the uncertainty is just associated with the ratio between catch and effort. Besides, Bognola et al. (2008) and Fonseca (2008) detected a better performance of these models when applying bivariate geostatistical models (when compared to univariate models) for making predictions in circumstances when the observations of the variates are co-located. Naturally in fisheries, catch and effort data are observed in the same locations, as effort causes catch; so, they are co-located. Due to the very nature of this type of observation it is not surprising that the bivariate model is not better than the univariate one.

In the yellowfin tuna application, when the number of predicated quadrats increased from 15% to 24% the MSEs of the adjusted estimate also increased. This is expected because more ‘unobserved’ quadrats need to be predicted, and hence the uncertainty will be greater.

In the simulation, when comparing the sampling-based estimates with the adjusted estimates that were calculated after fitting the SUM model, one notes that for the sample size of 85, as previously mentioned the adjusted estimates are better for the fourth (i.e. d) of the correlation-intensity scenarios, being advantageous in this case to adjust the model to estimate $cpue_1$. However, with a sample size of 76, Table 3 shows that no advantage in using the adjusted estimates for $cpue_1$. In other words, the MSEs of the adjusted estimates are higher than the MSEs of the sampling-based estimates. Using a sample size of 76 for the SUM model fitting, perhaps the number of points left for prediction (24) is so high that predictions are worse and the adjusted estimates are not better. Nonetheless, if the characteristics of the sample allows the spatial model to fit well—even if the number of observations is less than 85% of the total—it is possible that even better estimates are obtained using the spatial model compared to estimates based on sampling. It may be useful to use adjusted estimates in order to obtain more precise estimates for $cpue_1$. However, caution is needed when predicting in non-observed quadrats, especially when there are few observed and/or available sampled points for fitting the model. A small sample may not provide good estimates of model parameters—especially those associated with the spatial component, which are more difficult to estimate. A poorly determined covariance structure can lead to poor predictions; and bad predictions for the unobserved locations can lead to worse estimates.

The estimates obtained after fitting the univariate and bivariate models present values which are very close for a sample size of 85. Despite the higher scatter in estimates resulting from fitting the two models with a sample size of 76, the estimates remain close to each other. It suggests that there is no advantage in using the bivariate model in preference to the univariate one when estimating $cpue_1$. In addition to both models providing similar results, the number of parameters to be estimated in the SBM model is greater than in the SUM model, i.e. the SUM model is less complex.

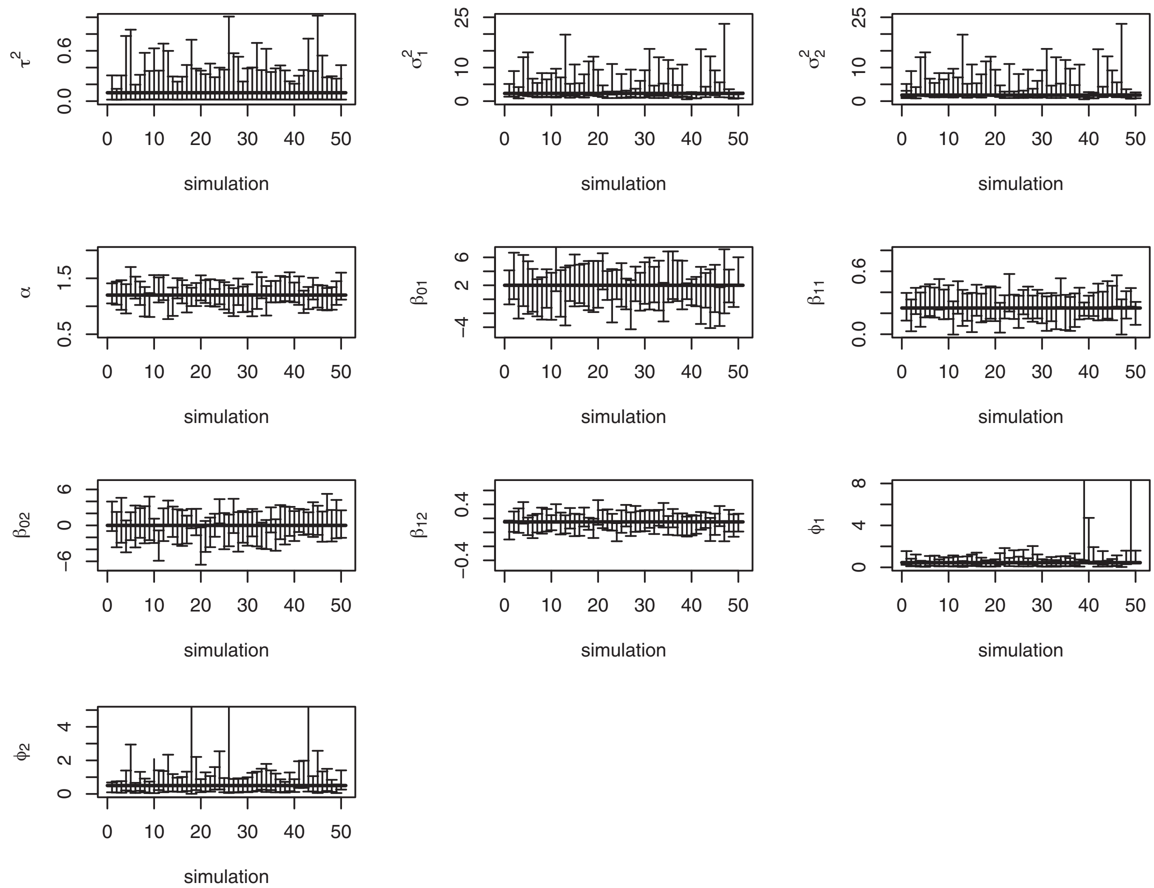

Fig. A.1. The 95% credible interval of parameters and values used in the simulation of data (horizontal line); under the scenario (d) (strong correlation between catch and effort and strong spatial correlation); and for the bivariate model applied for a sample size of 76 'observed' quadrats.

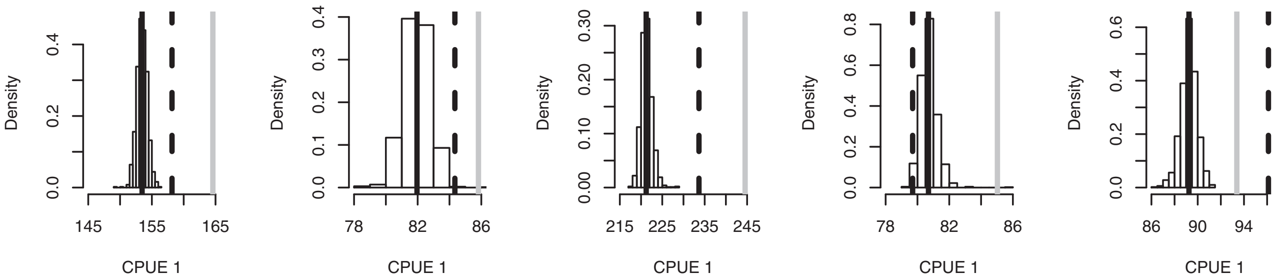

Fig. A.2. Histograms of the adjusted estimates obtained using the univariate model, for a sample size of 85 'observed' quadrats, and referring to 5 of the 50 cpue_1 samples. Each median is shown in black, each sampling-based estimate in gray, and the true value is dashed. The scenario for the model assumed strong correlation between catch and effort, and strong spatial correlation.

Despite no good grounds for fitting the SBM model when the main objective is to estimate cpue_1, this model has other advantages. For example, it allows the preparation of separate prediction maps for effort and catch, thus providing scope to adopt a different function of catch and effort as an alternative index to cpue_1.

The simulation study has shown that:

(i) Under scenario (d) of the correlation-intensity scenarios, when 15% of the observations were left for prediction, the adjusted estimates using the SUM and SBM models are better than sampling-based estimates.
(ii) Leaving more points—24%—for prediction, there is no advantage in using adjusted estimates rather than sampling-based estimates for cpue_1.
(iii) In general, estimates for cpue_1 obtained after fitting SBM and SUM models are very close, indicating little advantage in using the SBM model; the SUM model might actually be preferable because it is a simpler model.

Another possibility that could be explored in the simulation would be to study a catch univariate model and another independent for the fishing effort; then, with these models cpue_1, cpue_2 and cpue_3 could be estimated and compared with estimates obtained after adjusting the SBM model. It might be expected that in cases showing strong correlation between catch and effort the SBM model would present better results.

It would also be interesting to compare, through simulations, the behavior of the SUM model, SBM model and the sampling-based method for estimating cpue in a scenario where the sampling is biased, for example when the fishing targets areas of greater abundance. Simulations involving this scenario might show improvements over the sampling-based method in more general correlation structures.

In both models investigated in this paper it is assumed that the data follow a Gaussian process. However, when analyzing data from fisheries it is common to find inflated zero-catch data. In such cases different statistical approaches have been used, including the use of generalized linear models, where different probability distributions can be assumed for catch or for cpue data. A review of some of these approaches (ad hoc approach; using the Poisson; negative binomial; and delta-type two-step distributions) can be found in Shono (2008) where the author applied the Tweedie distribution to catch-and-effort data for silky shark and yellowfin tuna.

Acknowledgments

To Professor Alexandra Mello Schmidt of UFRJ for her insights. The senior author is indebted to the CNPq for financial support via a post-graduate scholarship. UFSCar—Sorocaba, ESALQ/USP and UNESP-Rio Claro are acknowledged for the use of their facilities.

Appendix A.

The effective range d_0 is obtained making $\rho(d_0) = \exp(-\phi d_0) = 0.05$. Taking natural logs we have $-\phi d_0 = \ln(0.05)$ or $-\phi d_0 \approx -3$, and so $\phi \approx 3/d_0$. Taking $d_0 = max.dist/2$, we get a prior $E(\phi) \approx 6/max.dist$, where *max.dist* represents the maximum distance among the observed locations.

Appendix B.

Posterior complete conditional distributions of the parameters of the univariate model for the ratio between catch and effort.

1. Posterior complete conditional distribution for $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is bivariate normal $\boldsymbol{\beta}|\sigma^2, \tau^2, \phi, y \sim N(\mathbf{Bb}, \mathbf{B})$, where $\mathbf{B} = \left(\mathbf{X}'[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]^{-1}\mathbf{X} + (\sigma^2_\beta \mathbf{I})^{-1}\right)^{-1}$ and $\mathbf{b} = \mathbf{X}'[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]\mathbf{y}$
2. $\pi(\phi|\boldsymbol{\beta}, \sigma^2, \tau^2, \mathbf{y}) = |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-1/2} \exp\left\{-\frac{1}{2}[\mathbf{y} - \mathbf{X}\beta]^T[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]^{-1}[\mathbf{y} - \mathbf{X}\beta] - b_\phi\phi\right\}\phi^{a_\phi - 1}$
3. $\pi(\sigma^2|\boldsymbol{\beta}, \tau^2, \phi, \mathbf{y}) = |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-1/2} \exp\left\{-\frac{1}{2}[\mathbf{y} - \mathbf{X}\beta]^T[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]^{-1}[\mathbf{y} - \mathbf{X}\beta] - \frac{b_\sigma}{\sigma^2}\right\}(\sigma^2)^{-a_\sigma - 1}$
4. $\pi(\tau^2|\boldsymbol{\beta}, \sigma^2, \phi, \mathbf{y}) = |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-1/2} \exp\left\{-\frac{1}{2}[\mathbf{y} - \mathbf{X}\beta]^T[\sigma^2\mathbf{R} + \tau^2\mathbf{I}]^{-1}[\mathbf{y} - \mathbf{X}\beta] - \frac{b_\tau}{\tau^2}\right\}(\tau^2)^{-a_\tau - 1}$

Appendix C.

Posterior complete conditional distributions of the parameters of the bivariate model for effort and catch.

1. $\boldsymbol{\beta}_1|\phi_1, \sigma^2_1, \mathbf{y}_1 \sim N(\mathbf{Bb}, \mathbf{B})$, where $\mathbf{B} = [\mathbf{X}_1^T(\sigma_1^2\mathbf{R}_1)^{-1}\mathbf{X}_1 + \boldsymbol{\Sigma}^{-1}_{\beta_1}]^{-1}$ and $\mathbf{b} = \mathbf{X}_1^T(\sigma_1^2\mathbf{R}_1)^{-1}\mathbf{y}_1 + \boldsymbol{\Sigma}^{-1}_{\beta_1}\boldsymbol{\mu}_{\beta_1}$
2. $\sigma^2_1|\boldsymbol{\beta}_1, \phi_1, \mathbf{y}_1 \sim IG\left(\frac{n}{2} + a_{\sigma_1}, c_{\sigma_1}\right)$, where $c_{\sigma_1} = \frac{1}{2}[\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1]^T\mathbf{R}_1^{-1}[\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1]$
3. $\pi(\phi_1|\boldsymbol{\beta}_1, \sigma^2_1, \mathbf{y}_1) = |\mathbf{R}_1|^{-1/2}\exp\left\{-\frac{1}{2}[\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1]^T[\sigma_1^2\mathbf{R}_1]^{-1}[\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1] - b_{\phi_1}\phi_1\right\}\phi_1^{a_{\phi_1} - 1}$
4. $\boldsymbol{\beta}_2|\sigma^2_2, \tau^2_2, \phi_2, \mathbf{y}_1, \mathbf{y}_2 \sim N(\mathbf{B}_2\mathbf{b}_2, \mathbf{B}_2)$, where $\mathbf{B}_2 = [\mathbf{X}_2^T(\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})^{-1}\mathbf{X}_2 + \boldsymbol{\Sigma}^{-1}_{\beta_2}]^{-1}$ and $\mathbf{b}_2 = \mathbf{X}_2^T(\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})^{-1}y_2 + \boldsymbol{\Sigma}^{-1}_{\beta_2}\boldsymbol{\mu}_{\beta_2}$
5. $\pi(\phi_2|\boldsymbol{\beta}_2, \sigma^2_2, \tau^2_2, \mathbf{y}_1, \mathbf{y}_2) \propto |\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-1/2}\exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2]^T[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2] - b_{\phi_2}\phi_2\right\}\phi_2^{a_{\phi_2} - 1}$
6. $\pi(\sigma^2_2|\boldsymbol{\beta}_2, \tau^2_2, \phi_2, \mathbf{y}_1, \mathbf{y}_2) \propto |\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-1/2}\exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2]^T[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2] - \frac{b_{\sigma_2}}{\sigma_2^2}\right\}(\sigma_2^2)^{-a_{\sigma_2} - 1}$
7. $\pi(\tau^2_2|\boldsymbol{\beta}_2, \sigma^2_2, \phi_2, \mathbf{y}_1, \mathbf{y}_2) \propto |\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-1/2}\exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2]^T[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2] - \frac{b_{\tau_2}}{\tau_2^2}\right\}(\tau_2^2)^{-a_{\tau_2} - 1}$

Appendix D. Credible intervals

Appendix E.

References

Anganuzzi, A.A., 2004. An Aggregate Model of Effort Distribution for the Eastern Pacific Tuna Fishery. IATTC Bulletin. Fisheries and Aquaculture Department, FAO, Roma.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC, New York, 452 pp.

Bognola, I., Ribeiro Jr., P., Silva, E., Lingnau, C., Higa, A., 2008. Modelagem uni e bivariada da variabilidade espatial de rendimento de *Pinus taeda* L. Floresta 38, 373–385.

Fonseca, B.H.F., 2008. Um estudo sobre estimação e predição em modelos geoestatísticos bivariados. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, Brasil.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman & Hall/CRC, New York, 323 pp.

Gelfand, A.E., Schmidt, A.M., Banerjee, S., Sirmans, C.F., 2004. Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). Test 13, 263–312.

Gelman, A., 1996. Inference and monitoring convergence. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 131–143.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall, New York, 668 pp.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 4, 457–511.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman & Hall, New York, 570 pp.

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

Matsunaga, H., Hosono, T., Shono, H., 2006. Analysis of longline CPUE of major pelagic shark species collected by Japanese research and training vessels in the Pacific Ocean. WCPFC Report. National Research Institute of Far Seas Fisheries WP-10, Manila, Philippines.

Nadal-Egea, J.A., 1996. Esfuerzo y Captura: Tecnología y Sobreexplotación de Recursos Marinos Vivos. El Colegio de México, Cidade del México, 476 pp.

Nishida, T., Chen, D.G., 2004. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. Fish. Res. 70, 265–274.

Nóbrega, M.F., 2008. Modelagem temporal e espacial para a avaliação de estoques de *Ocyurus chrysurus* e *Coryphaena hippurus* da região nordeste do Brasil. Tese (Doutorado em Oceanografia Biológica) – Universidade Federal do Rio Grande, Rio Grande, Brasil.

Oliveira, M.C.N., Barbin, D., Silva, E.F., Macedo, J.R., Capeche, C.L., Hissa, H.R., 2006. Estimação de parâmetros de uma função de covariância no estudo da variabilidade espacial do teor de cálcio no solo. Rev. Mat. Esta. 24 (4), 117–132.

Paez, M.S., Gamerman, D., De Oliveira, V., 2005. Interpolation performance of a spatio-temporal model with spatially varying coefficients: application to PM10 concentration in Rio de Janeiro. Environ. Ecol. Stat. 12, 169–193.

Pereira, J.C., 2009. Estimação do índice de abundância de um estoque pesqueiro com estrutura de correlação espacial: uma abordagem bayesiana. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, Brasil.

Pereira, J.C., Leandro, R.A., Petrere Jr., M., Nishida, T., 2009. Comparing three indices of catch per unit effort using Bayesian geostatistics. Fish. Res. 100, 200–209.
Petrere Jr., P., Giacomini, M., De Marco Jr., H.C., 2010. Catch-per-unit effort: which estimator is best? Braz. J. Biol. 2, 631–637.
Quinn II, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, Oxford, 542 pp.
Reis, I.A., 1998. Testando a aleatoriedade espatial de uma configuração de pontos. Dissertação de Mestrado, Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
Schmidt, A.M., Gelfand, A.E., 2003. A Bayesian coregionalization approach for multivariate pollutant data. J. Geophys. Res. Atmos. 108, 24–87.
Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fish. Res. 93, 154–162.
Spiegelhater, D.J., Thomas, A., Best, N.G., 2002. WinBugs version 1.4 user manual. Technical Report. Medical Research Council Biostatistics, Cambridge.
Vignaux, M., 1996. Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. Can. J. Fish. Aquat. Sci. 53, 963–973.
Walters, C., 2003. Folly and fantasy in the analysis of spatial catch rate data. Can. J. Fish. Aquat. Sci. 60, 1433–1436.