Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 27 (2015) 2 – 5

Procedia Environmental Sciences

Spatial Statistics 2015: Emerging Patterns - Part 2

A copula-based habitat preference index in fish spatial population modelling

Craig Marsh[a], Nokuthaba Sibanda[a,*], Matt Dunn[a], Alistair Dunn[b]

[a]Victoria University of Wellington, P O Box 600, Wellington 6140, New Zealand
[b]National Institute of Water and Atmospheric Research, P Bag 14901, Wellington 6021, New Zealand

Abstract

The habitat preference and spatial distribution of fish populations may depend on a number of environmental factors. For example, the habitat preference for a given species may depend on sea surface height, sea surface temperature and net primary productivity, among others. These variables are however not independent. We investigate an approach for determining a habitat preference index for spatial modelling of a fish population in a non-homogeneous environment. We use copulas to combine univariate habit preference variable distributions into a multivariate habit index. Such an index may be used to estimate a non-homogeneous spatial distribution of a population, with concentrations of fish in 'favourable' areas and avoidance of other less 'favourable' areas. The efficiency of copulas lies in the ability to study marginal behaviors separately from global dependence.

© 2015 The Authors. Published by Elsevier B.V This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee

Keywords: copula; habitat preference; spatial population model; fishery management; species distribution

1. Introduction

Environmental and biological factors such as sea surface temperature or amount of dissolved oxygen influence fish migration. Understanding how these factors affect the spatial distribution of fish populations will lead to better understanding of the drivers of fish population dynamics and fishery performance. In this study we investigate use of copulas to construct a single habitat preference index from multiple environmental factors for Albacore tuna in the New Zealand commercial fishery. The habitat index is then used to predict the spatial population distribution. Data used were on Albacore tuna abundance. These data are recorded by observers on surface longline vessels and stored in a centralised observer database (COD) administered by the National Institute of Water and Atmospheric Research (NIWA) for the New Zealand Ministry of Primary Industries (MPI). The observers identify and count all the catch for all species caught during the time they are observing. The data available from COD comprised information from 6515 surface longline sets collected between 25 March 1994 and 28 August 2012. This was subsequently reduced to 3642 observations between 2003 and 2012 to allow all environmental variables to be included in the analysis.

*Corresponding author. Tel.: +64-4-4636779; fax: +64-4-463-5045.
Email address: nokuthaba.sibanda@vuw.ac.nz.

1878-0296 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee
doi:10.1016/j.proenv.2015.07.107

2. Methods

2.1. Spatial population model

The aim of this study is to build a model for predicting the spatial distribution of albacore tuna in New Zealand's Exclusive Economic Zone. Since there was no appropriate data available, we assumed a homogenous fixed size population with equal movement preference across all ages and zero growth rate. Assuming a fixed abundance (population size), N, we used a deterministic spatially explicit population model over K location cells in T time-points given by

$$N_{tk} = f(N, \mathbf{X}_{tk}; \mathbf{\Phi}), \tag{1}$$

where $k = 1, 2, \ldots, K$, $t = 1, 2, \ldots, T$, N_{tk} denotes the size of the population at time point t in cell k. $f(N, \mathbf{X}_{tk}; \mathbf{\Phi})$ is a movement function given by

$$f(N, \mathbf{X}_{tk}; \mathbf{\Phi}) = N \cdot g(\mathbf{X}_{tk}; \mathbf{\Phi}), \tag{2}$$

where $g(\mathbf{X}_{tk}; \mathbf{\Phi})$ is a *habit preference index* determined from the joint density function of d environmental attributes $\mathbf{X}_{tk} = (X_1, X_2, \ldots, X_d)_{tk}$.

Our aims were to:

- identify the attributes $X_1, \ldots, X_d$ that best predict N_{tk}
- construct a habit index g that takes into account correlations between attributes
- estimate the corresponding parameter vector $\mathbf{\Phi}$

We used a temporal resolution of seasons (Summer, Autumn, Winter, Spring). This was a tradeoff between the resolutions seen in the environmental attributes and the abundance data. The spatial resolution was set at 1° latitude and longitude bins based on the most coarse resolution of the environmental variables. The area studied extended from $26°S$ to $49°S$ and from $164°E$ to $175°W$, chosen to encompass the locations of recorded fishing events.

2.2. Habitat index

The traditional approach for determining a habitat index evaluates the preference function

$$p(\mathbf{x}_{tk}; \mathbf{\Phi}, \alpha) = \prod_{j=1}^{d} f(x_{j,tk}; \boldsymbol{\phi}_j)^{\alpha_j} \tag{3}$$

and then determines the habit index at time point t as the relative preference across all locations using

$$g(\mathbf{x}_{tk}; \mathbf{\Phi}) = \frac{p(\mathbf{x}_{tk}; \mathbf{\Phi}, \alpha)}{\sum_{l=1}^{K} p(\mathbf{x}_{tl}; \mathbf{\Phi}, \alpha)}.$$

In equation 3, $\mathbf{x}_{tk}$ is the vector of observed values for the environmental attributes at location k and time-point t; ϕ_j and α_j are the parameters and weights for attribute j, respectively; and $f(x_{j,tk}; \phi_j)$ is a function that reflects the population's preference for $x_{j,tk}$ relative to other possible values.

The traditional approach assumes that environmental attributes act independently to create preferred habitats. We argue that incorporation of information about the dependence structure among attributes would improve estimation and prediction of the spatial structure of the population. We propose use of copulas to create a single habitat index that takes into account the dependence structure among attributes.

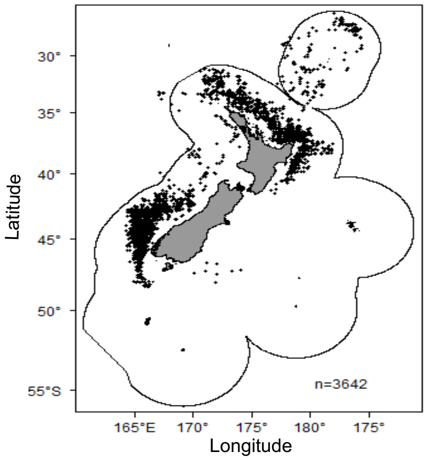
2.2.1. Copula-based habitat index

A d−dimensional copula, C, is a multivariate distribution on $[0, 1]^d$ with uniformly distributed marginals [1]. Based on Sklar's theorem [2], a joint density function h, based on a copula density, c, and marginal densities $h_1, \ldots, h_d$, is given by

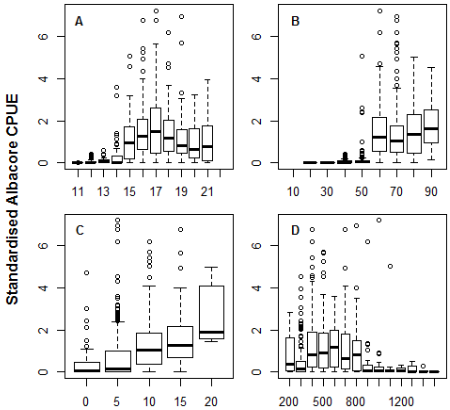
$$h(x_1, \ldots, x_d; \mathbf{\Phi}) = c(u_1, \ldots, u_d) h_1(x_1; \boldsymbol{\phi}_1) \ldots h_d(x_d; \boldsymbol{\phi}_d),$$

where $u_1 = H_1(x_1), \ldots, u_d = H_d(x_d)$ and

$$c(u_1, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C(u_1, \ldots, u_d).$$



(a) Spatial distribution of CPUE for the NZ Albacore tuna fishery, 2003-2012.



(b) Graphs of standardised CPUE against: (A) Sea Surface Temperature, (B) Sea Surface Height, (C) Sea Surface Gradient, (D) Net Primary Productivity

Fig. 1: CPUE and its relationship with environmental attributes

The copula density, c, captures the dependence structure of $X_1, \ldots, X_d$. The habit index is then given by

$$g(\mathbf{x}_{tk}; \mathbf{\Phi}) = h(x_1, \ldots, x_d; \mathbf{\Phi})_{tk},$$

for assumed marginal densities $h_{X_1}(x_1; \boldsymbol{\phi}_1), \ldots, h_{X_d}(x_d; \boldsymbol{\phi}_d)$ and an assumed dependence structure. For example if the bivariate Clayton copula is used, then

$$g(\mathbf{x}_{tk}; \mathbf{\Phi}) = \frac{(1+\theta)}{(u_1 u_2)^{\theta+1}} \left[\frac{(u_1 u_2)^{\theta}}{u_1^{\theta} + u_2^{\theta} - (u_1 u_2)^{\theta}} \right]^{\frac{1+2\theta}{\theta}} h_1(x_1; \boldsymbol{\phi}_1) h_2(x_2; \boldsymbol{\phi}_2).$$

3. Data

Fishing of Albacore tuna is done using longline fishing, in which baited hooks are attached to the mainline using shorter branch lines. We defined CPUE as the number of individuals caught per hook. Each fishing event was assigned to a location cell using the latitude, longitude and season at the start position of the event. Figure 1a shows the locations of the 3642 fishing events recorded between 2003 and 2012 in the study area.

All environmental attributes considered were determined using a combination of remote sensing via satellite imagery and in situ data collected using, for example, research vessel instruments or controlled and uncontrolled buoys. Interpolation, informed by satellite imagery and in situ data, is then used to fill in any gaps. The attributes considered include sea surface temperature (SST), sea surface height (SSH), sea surface gradient (MAG), net primary productivity (NPP) and depth. Figure 1b shows the influence of the first four attributes on CPUE. To investigate possible bivariate correlation structures among the environmental attributes, we used a scatterplot matrix (Figure 2).

4. Model fitting and evaluation

The model in equation 2 was fitted using the `Spatial Population Model` software programme [3]. Maximum likelihood estimation was used to estimate the model parameters. Competing models were evaluated using the AIC.

A number of candidate bivariate copula and marginal density combinations for the environmental attributes were used to obtain the habit preference index $g(\mathbf{X}_{tk}; \mathbf{\Phi})$. The form of the habit index that gave the

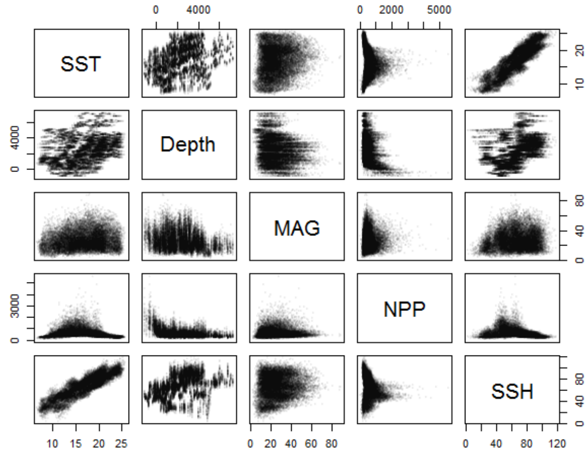


Fig. 2: Pairwise scatter plots showing the dependence structure between pairs of environmental attributes.

Table 1: Comparison of traditional (independence) and copula approach for modelling spatial population distribution. The best fitting model has the lowest $-\log L$.

Bivariate dependence structure	$-\log L$	Parameters
Independence	3088.25	6
Frank	3066.70	6
Gaussian	3015.36	6
Gumbel	3088.07	6

lowest negative log-likelihood was identified as having the best fit. Table 1 shows results for the traditional approach (Independence) and three candidate bivariate dependence structures.

5. Discussion

The results show that incorporating the dependence structure improves the model fit compared to an assumption of independence. The bivariate Gaussian copula gave the best fit for a model that included SST and MAG.

Acknowledgements

We are grateful to the New Zealand Ministry for Primary Industries and the National Institute of Water and Atmospheric Research for providing funding and commercial fishery data.

References

[1] R. B. Nelsen, An introduction to copulas, Springer, 2007.
[2] A. Sklar, Fonctions de rèpartition á n dimensions et leurs marges, Publications de l'Institut de statistique de l'Universitè de Paris, 1959.
[3] A. Dunn, S. Rasmussen, S. Mormede, Spatial population model user manual, SPM, Hobart, Australia, v1.1-2015-03-05 (rev. 1248) Edition (2015).