Ecological Modelling 265 (2013) 74–84

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

Review

# Modeling the habitat associations and spatial distribution of benthic macroinvertebrates: A hierarchical Bayesian model for zero-inflated biomass data

J.B. Lecomte[a,b,*], H.P. Benoît[c], M.P. Etienne[a,b], L. Bel[a,b], E. Parent[a,b]

[a] INRA, UMR 518 Math. Info. Appli., F-75005 Paris, France
[b] AgroParisTech, UMR 518 Math. Info. Appli., F-75005 Paris, France
[c] Gulf Fisheries Centre, Fisheries and Oceans Canada, Moncton, NB E1C 9B6, Canada

ARTICLE INFO

*Article history:*
Received 10 April 2013
Received in revised form 7 June 2013
Accepted 8 June 2013

*Keywords:*
Zero-inflated data
Bayesian hierarchical modeling
Habitat associations
Spatial dependencies
Macro-invertebrates

ABSTRACT

Biomass samples from marine scientific surveys are commonly used to investigate spatial and temporal variations in stock abundances. Biomass records are often characterized by a high proportion of zeros on the one hand, and occasional large catches on the other. These features induce a modeling challenge when trying to understand the state of populations and their ecological associations with one another and with habitat. We develop a hierarchical Bayesian model to represent the spatial structure of biomass and analyze the spatial distribution and habitat associations of three species of macro-invertebrates sampled in the southern Gulf of St. Lawrence (Canada). A zero-inflated distribution based on a compound Poisson with Gamma marks is used for the observation layer, and a linear model with spatial correlated errors accounts for the role of habitat variables (temperature, depth and sediment type) in the process layer. Maps of quantities of interest (e.g. probability of presence, quantity of biomass) are produced, taking into account the uncertainty of the estimated parameters and observation errors. This hierarchical Bayesian modeling approach provides a useful tool for spatial management of human activities that may affect living resources that may affect living resources, such as marine protected areas.

Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.

**Contents**

1. Introduction ........ 75
2. Methods ........ 75
  2.1. Data description ........ 75
  2.2. The statistical model for zero-inflated continuous positive data ........ 75
    2.2.1. Observation layer ........ 76
    2.2.2. Spatial distribution layer ........ 76
  2.3. Bayesian inference ........ 78
  2.4. Validation and model selection ........ 78
    2.4.1. Posterior predictive checking ........ 78
    2.4.2. Model comparison ........ 78
  2.5. Predictions ........ 78
3. Results ........ 79
  3.1. Green sea urchin ........ 79
  3.2. Starfish ........ 81
  3.3. Sea cucumber ........ 81
4. Discussion ........ 82
  Appendix A. Model code ........ 83
  References ........ 83

* Corresponding author at: INRA, UMR 518, 16 rue Claude Bernard, 75005 Paris, France. Tel.: +33 144087292.
*E-mail address:* jean-baptiste.lecomte@agroparistech.fr (J.B. Lecomte).

0304-3800/$ – see front matter. Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.06.017

## 1. Introduction

Understanding species spatial distribution and habitat associations are key challenges when managing harvested, endangered or invasive species (Welsh et al., 1996; Engler et al., 2004; Cook et al., 2007). Marine spatial management measures, in which the spatial and temporal distribution of human activities is restricted to achieve ecological, social and economic objectives (e.g., marine protected areas), have been the focus of many studies in the recent decades (Shea, 1998; Hilborn et al., 2004; Hobday and Hartmann, 2006; Hartog et al., 2011). In many applications, these management approaches require knowledge of habitat use by the targeted species to be effective (Perry and Smith, 1994; Williams and Bax, 2001).

Linear or additive models are often developed to infer distributions and habitat use and preferences using survey or other ecological data (as reviewed by Guisan and Thuiller, 2005). Efficient models must be able to address two common characteristics of ecological data: observations can be dominated by a large number of null values combined with skewed positive values, and abundance can be strongly spatially correlated. Failure to address both of these characteristics is well known to impact model parameter estimates and their uncertainty, leading to incorrect statistical inference and therefore, in turn, potentially inappropriate management actions (Zuur et al., 2009; Sileshi et al., 2009). Ideally, the models should also be able to address possible spatial misalignment between the available data for abundances and for habitat characteristics.

High proportions of zeros in survey data stem from three general causes. An observed zero value can be a true zero if the species is not present in the studied area, while a false zero, also called pseudo-absence, results from a low probability of detection even though the species is present. A third class of zeros results from an observer effect, whereby a species normally found in the study area is frightened away by some inappropriate data collection procedure. Numerous approaches exist for such zero-inflated data when dealing with counts, as reviewed in Martin et al. (2005). The two main approaches, Zero-inflated Poisson (ZIP) and Zero-inflated binomial (ZIB), are mixture models and the presence–absence is modeled separately from the number of counts (i.e. individuals). The development of zero-inflated models for continuous abundance data (i.e. densities or biomasses) has also received attention (Stefansson, 1996; Maunder and Punt, 2004; Fletcher et al., 2005; Shono, 2008; Ancelet et al., 2010). The simplest approach consists in adding a positive constant to all the observations, typically followed by a logarithmic transformation, as is often performed in generalized linear modeling (GLM). This approach requires choosing an arbitrary constant that could severely bias model estimates (Maunder and Punt, 2004; Shono, 2008). An alternative is to remove the zero catches from data prior to the analysis. However removing zero values often affects the results and can also bias the analysis (Martin et al., 2005), though this is not necessarily the case (Maunder and Punt, 2004). A common and slightly more complex approach for continuous data, named the delta approach (Stefansson, 1996; Shono, 2008), models separately the presence–absence using a binomial distribution and positive values using a standard probability distribution function such as the log-normal (leading to a delta-lognormal model) or the gamma (delta-gamma). The approach reduces bias since the expected biomass is the product of the probability of presence and the average positive biomass. This family of models treats all absences as true zeros. Furthermore, sampling effort, which can vary between sites for a number of logistical and operational reasons, is mostly addressed by a prior standardization of the data (Stefansson, 1996). However, performing such a standardization may obscure the relationship that exists between expected values (for a given sampling effort) and their associated variance for count probability density functions.

In this paper, we develop a hierarchical Bayesian spatial model for biomass data that overcomes these shortcomings. We apply this approach to describe the distribution and habitat associations of epibenthic invertebrates in the southern Gulf of St. Lawrence (sGSL), Canada. The biomass records come from an annual bottom trawl survey in which invertebrates and fish are collected at randomly chosen locations by sweeping the ocean floor over targeted distances which can vary between sites. We use a model based on two substructures that are linked probabilistically using a hierarchical approach. The first substructure, the observation layer, consists of a compound Poisson model with Gamma marks, which heuristically models the process of observing a Poissonian number of patches of a species, each containing a random biomass given by the Gamma mark. This approach constitutes a generalization of the one proposed by Bernier and Fandeux (1970) and applied in ecology by Ancelet et al. (2010) which used exponential marks. It also allows for explicit accounting for the duration or volume of sampling for individual sampling events. The second model substructure explicitly models habitat associations using a linear model that accounts for spatial autocorrelation using a geostatistical approach. Jointly, these model substructures result in a modeling approach that is very flexible, likely making it a useful tool for spatial analysis and planning.

## 2. Methods

### 2.1. Data description

Fisheries and Oceans Canada has conducted an annual bottom-trawl survey in the sGSL each September since 1971 (Chadwick et al., 2007; Benoît et al., 2009). Since its inception, the main objective of this survey has been to quantify the abundance and the distribution of marine fishes and certain commercially important invertebrates. Since 1988, data for epibenthic invertebrates such as urchins, starfish, whelks and anemones have been collected. The domain for the sGSL survey is split into 27 strata defined so as to be homogeneous in terms of depth and geographic location. Every year, since the mid 1980s, 140–200 sites have been chosen according to a stratified random design. The number of sites per stratum is generally proportional to stratum size, making the selection of sites at the survey level approximately randomly balanced. Sites are sampled using a straight-line tow for a target duration of 30 min. at 3.5 knots. All captured organisms are identified to the lowest taxonomic level possible and weighed in kilograms per tow. Habitat information, such as bottom temperature (°C) and depth (m), is also collected at each bottom-trawl site. Moreover the type of sediments is interpolated at each sampling site from an existing map of surficial geology for the Gulf of St. Lawrence (Loring and Nota, 1973). This study focuses on three epibenthic macroinvertebrates sampled during the 1997 survey to illustrate the modeling approach: green sea urchin (*Strongylocentrotus droebachiensis*), starfish (*Asterias sp*), and sea cucumber (*Cucumaria frondosa*). These three taxa were chosen for their differences in density distribution and habitat preferences so as to demonstrate the model's ability to confront different data situations (Figs. 1 and 2). In fact, the majority of epibenthic macroinvertebrates in the sGSL are distributed in patches of localized variable abundance, interspersed by numerous and relatively large areas where the species is absent. Consequently, the dataset contains a very large proportion of sites where the species are not observed.

### 2.2. The statistical model for zero-inflated continuous positive data

The model description is split into two parts, as is classically done in hierarchical Bayesian modeling. The first section describes

Download English Version:

https://daneshyari.com/en/article/6297069

Download Persian Version:

https://daneshyari.com/article/6297069

Daneshyari.com