Multivariate analysis of marine litter abundance through Bayesian space-time models

Analisi multivariata dell'abbondanza di rifiuti marini attraverso modelli bayesiani spazio-temporali

C. Calculli, A. Pollice, L. Sion, and P. Maiorano

Abstract This work focuses on the analysis of abundance data for marine litter categories, collected during trawl surveys regularly conducted at local scale, in the Central Mediterranean. Here marine litter data are modeled in order to estimate the effects affecting the dynamics of litter assemblages at different spatio/temporal scales. A correlated response model with latent variables is proposed. This modeling approach is particularly suitable to infer potential environmental covariates while controlling for correlation between litter categories and providing a method for residual ordination. MCMC estimation is implemented within the Bayesian hierachical framework that allows to integrate environmental and anthropogenic processes into a single model.

Abstract *Questo lavoro riguarda l'analisi delle abbondanze di diverse categorie di rifiuti marini, censiti durante campagne di pesca a trascico condotte su scala locale nel Mediterraneo centrale. In questo ambito, i dati sui rifiuti sono modellati al fine di investigare i fattori che influiscono sulle dinamiche di aggregazione dei rifiuti su scala spazio/temporale. Viene presentato un modello con risposte correlate che rappresenta un approccio particolarmente idoneo per investigare l'influenza di covariate ambientali, controllare la correlazione tra categorie di rifiuti e fornire un metodo per l'ordinamento residuale. Stime MCMC sono implementate nel contesto bayesiano gerarchico che permette di integrare processi ambientali e antropogenici in un unico modello.*

Key words: Bayesian hierarchical modeling approach; Correlated response model; Latent variables; Marine Litter

Crescenza Calculli, Alessio Pollice
Department of Economics and Finance, University of Bari Aldo Moro, Largo Abbazia S. Scolastica, Bari, Italy, e-mail: crescenza.calculli@uniba.it, alessio.pollice@uniba.it

Letizia Sion, Porzia Maiorano
Department of Biology, University of Bari Aldo Moro, Via E. Orabona 4, Bari, Italy, e-mail: letizia.sion@uniba.it, porzia.maiorano@uniba.it

1 Introduction

Marine litter has recently become a recognized global and local ecological concern that might jeopardize the status of marine ecosystems. The debris quantity and its distribution on the Mediterranean seafloor are still not well-known although the effects on the sea bed and on the marine living communities are pretty clear [1]. Experimental bottom trawl surveys carried out in the Mediterranean basin in the last years represent a valuable source of information about wastes; litter typologies can be seen as special items caught by the trawl net together with marine species. While single-species distribution models have been commonly used to explain and predict the response of different taxa to environmental variation, the analysis at the community-level is still lacking. Some innovative approaches that explicitly acknowledge the multivariate nature of species assemblages were recently proposed [2, 4]. These approaches model the actual processes that determine the assemblage of community samples, taking into account for the various sources of correlation across species. In this study we analyze multivariate litter abundance data using a correlated response model in the spirit of [2]. This model merges univariate Generalized Linear Models with latent variables to account for the residual correlation across litter categories, *e.g.* due to environmental interactions or unaccounted covariates. Latent variables provide a method for "residual ordination". The whole implementation is performed in a hierarchical Bayesian framework.

2 Materials and Methods

2.1 Study area and data

Litter data are collected during experimental trawl surveys conducted from 2013 to 2017 in the North-Western Ionian Sea as a complementary (voluntary) activity of the international project MEDITS (MEDiterranean International Trawl Surveys). The North-Western Ionian is the deepest sea in the Mediterranean basin characterized by a complex geomorphology and the presence of important fisheries as well as main harbours. An increasing touristic activity is developing along the Ionian coasts, thus the sea bottoms are here exposed to a strong increase in anthropogenic impact.The same 70 depth-stratified hauls are carried out between 10 and 800 m in depth every year, summing to 350 hauls in 5 years. Wastes caught during the trawl surveys are classified in 8 categories: plastic, rubber, metal, glass/ceramic, cloth/natual fibres, processed wood, paper/cardboard, other/unspecified. The number of collected items for each litter category was scaled to the swept surface unit (1 km^2), thus obtaining density indices (N/km^2) for each litter category and survey at every haul location. Litter density is a semi-continuous zero-inflated non-negative variable. Preliminarily, to investigate factors influencing the density of litter categories, we consider only the seafloor depth as environmental covariate.

2.2 Statistical framework

To investigate assemblage of litter in terms of density and composition, a mixture model where latent variables are included alongside the measured covariates is proposed. In particular, densities of litter categories are jointly modeled as semi-continuous zero-inflated multivariate responses assuming the Tweedie distribution model. The mean density μ_{ij} of the j-th litter category at the i-th haul is specified by the following mixture model:

$$g(\mu_{ij}) = \alpha_1(t_i) + \alpha_2(s_i) + \beta_{0j} + \sum_{k=1}^{p} \beta_{jk} X_{ik} + z_i' \theta_j \quad i = 1,\ldots,350; \quad j = 1,\ldots,8 \quad (1)$$

where $g(\cdot)$ is the link function, $\alpha_{1,2}(\cdot)$ are effects adjusting for differences in site and time (year) on the overall litter density, β_{0j} is the litter type-specific intercept and β_{jk} is the type-specific regression coefficient of the k-th covariate (preliminarily, only the seafloor depth). Finally, $z_i = (z_{i1},\ldots,z_{iq})'$ is a q-dimensional vector of latent variables, while $\theta_j = (\theta_{j1},\ldots,\theta_{jq})$ are the corresponding litter type-specific loadings. Latent variables can be considered here as missing informative predictors for the multivariate response inducing residual correlation between litter categories. Independent weakly informative $N \sim (0,10)$ priors were assumed for all site and time effects, type-specific intercepts, type-specific regression coefficients, latent variables and loadings. Uniform priors $U \sim (0,30)$ are adopted for all dispersion and variance parameters in the model. Inferences for model in Eq. 1 were implemented by the `boral` package [2] that provides an interface between `R` and `JAGS` [5] for multi-species models with latent variables.

3 Results

The model in Eq.1 was fitted with 1 to 3 latent variables and with fixed or random site and time effects. All model estimates were obtained using 20,000 iterations, discarding the first 5,000. The Geweke diagnostic and the graphical inspection of trace plots provided clear evidence of the convergence of MCMC chains for all model parameters. As reported in Table 1, the best models in terms of lowest BIC consider random site/time effects. Results for the model with two latent variables, that also enable to draw a scatterplot of the ordinations as for distance-based techniques, suggest a positive correlation between plastic and glass litter due to depth (Figures 1A-B). Strong, positive residual correlations are observed: plastic is correlated with all other materials except for metal and other/unspecified wastes. Estimated spatial effects represented in Figure 1C, allow to identify some *hot-spots* assemblages for all litter categories. This work represents a starting point for the analysis of space/time structured multivariate litter data. Further developments include the selection among

Table 1: Values of the BIC for models with 1-3 latent variables (LVs) and fixed/random site and time effects

	site/time effects	
	fixed	random
1 LV	8731.59	8387.27
2 LVs	8777.11	8434.66
3 LVs	8833.30	8486.78

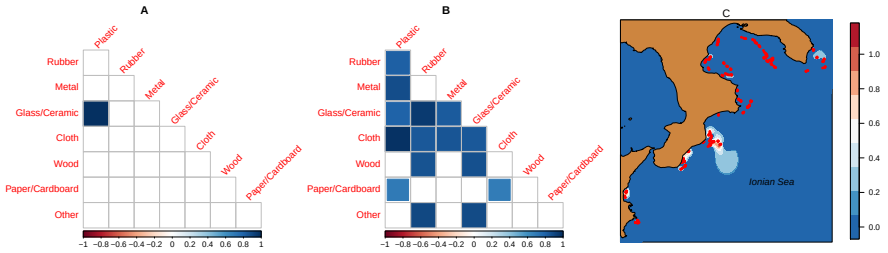

Fig. 1: (A) Correlations between litter categories due to seafloor depth; (B) Residual correlations (we report correlations with 95% credible intervals excluding zero); (C) Predictions of spatial effects of sites.

a larger number of environmental predictors by Boosted Regression Trees (BRT) as already suggested in [3] for ecological data.

Acknowledgements C. Calculli and A. Pollice were supported by the PRIN2015 project EphaStat - *Environmental processes and human activities: capturing their interactions via Statistical methods*, funded by the Italian Ministry of University and Research (MIUR).

References

1. Gall, S.C., Thompson,R.C.: The impact of debris on marine life. *Marine Pollution Bulletin* **92**, 170–179 (2015)
2. Hui, F. K. C.: boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in R. *Methods in Ecology and Evolution* **7**, 744–750 (2016)
3. Leathwick, J. R., Elith, J., Francis, M., Hastie T., Taylor, P.: Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* **321**, 267–281 (2006)
4. Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. and Abrego, N.: How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letter* **20**, 561–576 (2017)
5. Plummer, M. et al.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)* Vienna, Austria (2003)