An INLA spatio-temporal model for zero-inflated marine plastic litter abundance

C. Calculli[1,*], A. Pollice[1], I. Paradinas[2], L. Sion[3] and P. Maiorano[3]

[1] *Department of Economics and Finance, University of Bari, Largo Abbazia S. Scolastica 53, 70124 Bari, Italy; crescenza.calculli@uniba,it, alessio.pollice@uniba.it*
[2] *Asociación Ipar Perspective, C/Karabiondo, 48600 Sopela, Spain; paradinas.iosu@gmail.com*
[3] *Department of Biology, LRU CoNISMa, University of Bari, via E. Orabona 4, 70125 Bari, Italy; letizia.sion@uniba.it, porzia.maiorano@uniba.it*
[*] *Corresponding author*

Abstract. *The marine plastic litter pollution is a worldwide growing environmental concern. Despite its negative effects on marine ecosystems, the phenomenon is still not well-known at global and local scale. This work aims at assessing the spatio-temporal distribution of plastic litter amounts found at the sea-floor in a region of the central Mediterranean (Ionian sea). Inspired by species distribution models, we propose a two-parts model to accommodate the excess of zeros and the spatio-temporal correlation characterizing abundance monitoring data. A common spatial effect that links the plastic abundances and the probabilities of occurrences is implemented with the Stochastic Partial Differential Equation approach extended to a non-stationary barrier model. The INLA methodology allows to efficiently perform Bayesian inference to fit complex spatio-temporal models including effects of environmental covariates and enables to investigate the assemblages of plastic litter over the study region.*

Keywords. *Hurdle model; Integrated nested Laplace approximation (INLA); Marine ecology*

1 Introduction

The extensive use of disposable plastic items with a cultural propensity of increasingly over-consuming, discarding and littering, has become a lethal combination for marine ecosystems. While it is true that not all marine garbage is plastic, recent literature clearly indicates that plastic is the dominant material littering seas [3]. Despite the known negative effects of plastic accumulation on habitats and communities [7], the magnitude of the marine litter pollution has yet to be deeply investigated at global and local scale. In the Mediterranean basin, the majority of plastic comes from terrestrial inputs (coastal human populations and rivers) as well as from discarded fishing gears and shipping traffic that greatly contribute to the overall litter of sea bottoms [2]. Even though devoted to the study of benthic and demersal fish stocks, experimental bottom trawl surveys regularly carried out in the Mediterranean (MEDITS program), represent a valuable source of information about wastes. Litter categories might be seen as special abiotic items or additional "species", caught by trawl nets together with real marine species. Therefore, the analysis of litter abundances is not far different from species distribution modeling (SDM). In this spirit, we propose to analyze the spatio-temporal distribution of plastic abundances at the sea-floor in a region of the central Mediterranean (Ionian sea), investigating environmental factors that might affect litter assemblage dynamics at local scale. To this end, a suitable modeling approach is proposed in order to accommodate the zero-inflation and the spatio-temporal dependence characterizing abundance data. Hurdle models assume that zero and non-zero data are generated by two independent processes, one for the probabilities of occurrences and the other for the intensities of the non-zero responses. For semi-continuous data, the commonly assumed independence between the two processes might be considered unsuitable since it neglects the relation between abundances and probabilities of occurrences. It is far more realistic to expect low abundance intensities associated with low probabilities of occurrences and vice versa. According to [5, 6], a framework with a common latent gaussian random field (GRF) that allows the two processes to be related, is considered. To model the spatio-temporal structure, we refer to the stochastic partial differential equations (SPDE) approach that consists in defining the continuously indexed Matérn Gaussian field (GF) as a discretely indexed spatial random process (GMRF) using piece-wise linear basis functions defined on a triangulation of the domain of interest. SPDE provides a representation of the whole spatial process that varies continuously in the considered domain [4]. The SPDE approximation is currently implemented using the Integrated Nested Laplace approximation (INLA) [8] via the R-INLA package (`http://www.r-inla.org`) designed to make Bayesian inference accessible for a large class of latent Gaussian models. Using this approach, accurate approximations of the posterior marginals are obtained with computationally efficient tools alternative to cumbersome MCMC simulations. The INLA method is an efficient approach for modeling spatio-temporally correlated data with excessive zeros considering the effects of environmental covariates affecting the dynamics dynamics of plastic litter assemblages in the central Mediterranean.

2 Data description

Monitoring data are collected during experimental trawl surveys conducted from 2013 to 2016 in the North-Western Ionian Sea as part of MEDITS project (MEDiterranean International Trawl Surveys) activity. The same 70 depth-stratified hauls are carried out between 10 and 800 m in depth every year, summing to 280 hauls in 4 years. For each survey at every haul location, plastic density indices (N/km^2) are obtained scaling the number of collected items to the swept surface unit (1 km^2). Data concerning *sea currents* and *fishing activities*, characterized by different spatio-temporal support, were first aligned and then considered to investigate environmental factors affecting the distribution of plastic litter densities over the study region. In particular, we investigate the effects of: 1) the superficial eastward and northward sea water velocities (U and V) retrieved from the Copernicus Marine Environment Monitoring Service (`http://marine.copernicus.eu/`); 2) the daily average transit (MVH) and fishing time (MFH) for 3 vessel types (Drifted Longlines, Fixed Gears and Trawlers) provided by the International Global Fishing Watch organization (`https://globalfishingwatch.org/`).

3 The Bayesian spatio-temporal hurdle model

Density-based spatio-temporal abundance processes are commonly measured in R_+, resulting in semi-continuous non-negative datasets. A convenient representation of such datasets is obtained by Hurdle models that consider two independent sub-processes: an occurrence process and a conditional-to-presence continuous process. Let y_{st} and z_{st} being the occurrence and the conditional-to-presence abundance sub-processes at time t ($t = 1,\dots,T$) and location s ($s = 1,\dots,n_t$), then for the plastic litter densities we get:

$$
\begin{aligned}
& z_{st} \sim Ber(\pi_{st}) \\
& logit(\pi_{st}) = \beta_0^{(1)} + \sum_{i=1}^{p} \beta_i^{(1)}(x_{ist}) + V_{st}(1) \quad \text{with} \quad s = 1, \ldots, 70; \quad t = 1, \ldots, 4 && (1) \\
& y_{st} \sim Gamma(a_{st}, b_{st}) \\
& log(\mu_{st}) = \beta_0^{(2)} + \sum_{i=1}^{p} \beta_i^{(2)}(x_{ist}) + V_{st}(2) \quad \text{with} \quad s = 1, \ldots, 70; \quad t = 1, \ldots, 4 && (2)
\end{aligned}
$$

where π_{st} and $\mu_{st} = a_{st}/b_{st}$ are modeled through the logit and logarithm links, respectively. In linear predictors, the $\beta_0^{(1)}$ and $\beta_0^{(2)}$ represent the intercepts, $\beta^{(1)}$ and $\beta^{(2)}$ are fixed effects of spatio-temporally varying covariates x_i (U, V, MVH and MFH). In order to account for information shared by related occurrence and abundance sub-processes, the V_{st} components in Eq.(1)-(2) are assumed common and modeled by a Gaussian field through the SPDE approach [4] extended to the non-stationarity case to ensure that the spatial correlation seeps around coastlines [1], thus $\boldsymbol{V} \sim N(0, Q(\kappa, \tau))$ and $(log(\kappa), log(\tau)) \sim MVN(\mu, \rho)$ where the covariance function of the spatial effect Q depends on a range effect (κ) and a total variance parameter (τ). Due to the short time series of 4 years available, a first order autoregressive (AR1) model is preliminary adopted to capture the temporal effect. The lack of information leads to assign vague prior distributions to all model parameters, as implemented by default in INLA.

4 Main results

Figure 1 reports estimated fixed effects for the positive plastic densities of the Hurdle model in Eq.(2). Although none of the estimated effects is relevant in affecting the intensity of plastic densities, some insights come from these results. In particular, the left panel shows estimated higher densities with respect to current towards north-western direction. The right panel highlights, instead, higher densities associated with increasing transit time and the opposite negative effect for the fishing time covariate (the more fishing activity, the lesser sampling litter items observed in the area). The shared spatio-temporal field, represented in Figure 2, shows changes concerning densities and presence of plastic hot-spots from year to year. The spatial variation clearly distinguishes hauls with higher densities of plastic in the same areas of Sicily, Calabria and Apulia over the years. On the other hand, the available time series is too short for the identification of a temporal trend.

This work represents a starting point for the analysis of spatio-temporal structured monitoring data for multiple litter categories. To combine environmental information from multiple sources and different temporal sampling frequencies, further development includes the identification of a suitable modeling framework to face the change of support problem.

Acknowledgments. 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). The MEDITS surveys have been carried out within the Data Collection Framework. For fishing activities data availability, we thank W. Zupa, COISPA Tecnologia & Ricerca, Bari, Italy.

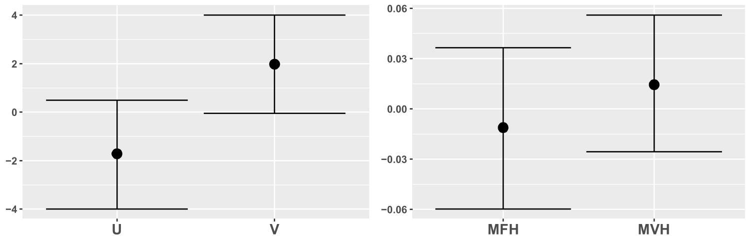

Figure 1: Estimated covariates effects for positive plastic densities (bars represent 95%CI)

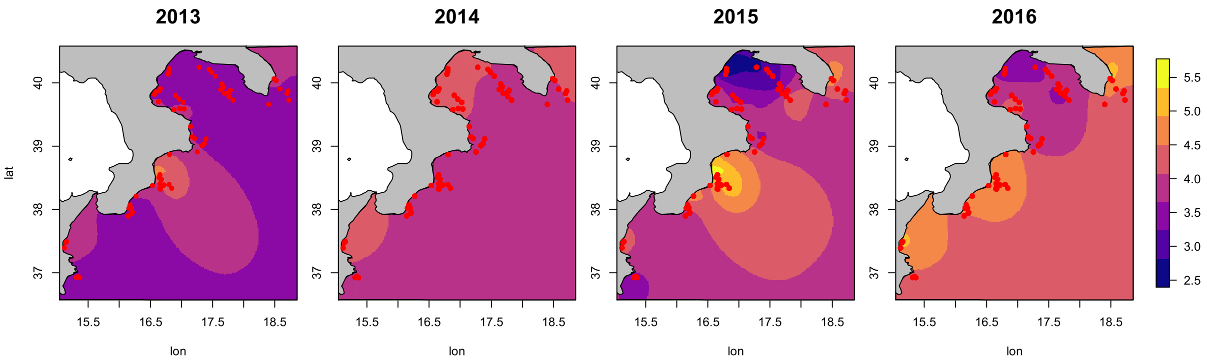

Figure 2: Yearly posterior means of spatial effect

References

[1] Bakka, H., Rue, H., Fuglstad, G-A., Riebler, A,, Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgren, F. (2018) Spatial modeling with R-INLA: A review. *WIREs Comput Stat*, 10:e1443

[2] Consoli, P., Romeo, T., Angiolillo, M., Canese, S., Esposito, V., Salvati, E., Scotti, G., Andaloro, F., Tunesi, L.(2019). Marine litter from fishery activities in the Western Mediterranean sea: The impact of entanglement on marine animal forests. *Environmental Pollution* **249**, 472–481

[3] Law, K. L.(2017). Plastics in the marine environment. *Annual review of marine science* **9**, 205–229

[4] Lindgren, F., Rue, H., Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society*, **73**, 423–498

[5] Paradinas, I., Conesa, D., López-Quílez, A., Bellido, J. M. (2017). Spatio-Temporal model structures with shared components for semi-continuous species distribution modelling. *Spatial Statistics*, **22**, 434–450

[6] Quiroz, Z. C., Prates, M. O., Rue, H. (2014) A Bayesian approach to estimate the biomass of anchovies off the coast of Perù. *Biometrics*, **71**, 208–217

[7] Rochman, C. M., Browne, M. A., Underwood, A. J., Van Franeker, J. A., Thompson, R. C., Amaral-Zettler, L. A. (2016). The ecological impacts of marine debris: unraveling the demonstrated evidence from what is perceived. *Ecology*, **97**, 302–312

[8] Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by usingintegrated nested Laplace approximations. *Journal of the Royal Statisitcal Society, Series B*, **71**, 319–392