



Fauna characterization of a cold-water coral community network along the Apulian coasts by Bayesian mixed models

C. Calculli^{1,*}, G. D'Onghia², N. Ribecco³, P. Maiorano², L. Sion²,
A. Tursi²

¹ CoNISMa Local Research Unit Bari - Italy, calculli.enza@gmail.com

² Department of Biology, University of Bari Aldo Moro, via E. Orabona 4, 70125 - Italy, gianfranco.donghia@uniba.it, porzia.maiorano@uniba.it, letizia.sion@uniba.it, angelo.tursi@uniba.it

³ Department of Economics and Mathematical Methods, University of Bari Aldo Moro, Largo Abbazia Santa Scolastica 53, 70124 Bari, Italy, nunziata.ribecca@uniba.it

*Corresponding author

Abstract. *The exploration of a cold-water coral (CWC) community network connecting the Southern Adriatic fish populations with those of the Northern Ionian Sea has many challenging implications involving biodiversity conservation and fisheries management. To characterize the species assemblages of the CWC community network, we analyze the size of six mostly abundant species to highlight the main differences among five CWC areas along the Apulian coasts. Data are surveyed by experimental longlines casted in all CWC areas between 2013 and 2014. Bayesian Generalized Additive Mixed Models (GAMMs) are applied to analyze the variation of the fishes length in the five CWC areas according to species and covariates such as depth and abundance. GAMMs allow to account for various effects of variability components on the overall length of fishes. Appropriate smooth functions are available to describe the physical effect of each covariate and specific random area effects are allowed to induce correlation among individuals of all species captured in the same area. Parameter estimation is carried out by maximization of the joint posterior probability distribution, where marginal posterior probabilities associated to specific model terms are obtained by a spike-and-slab prior structure, that can be viewed as a scale mixture of Gaussians. This approach is implemented in the R package *spikeSlabGAM*, able to deal with most common distributional assumptions and allowing efficient variable selection and model choice. Results show that the size of fishes is collectively affected by random effects of the CWC areas and by smooth effects of their depth and abundance.*

Keywords. *Bayesian GAMMs; Random effects; Cold-water coral community network; Fish length*

1 Introduction

The Apulian continental margin (Central Mediterranean) is characterized by the presence of a network of cold-water coral (CWC) communities connecting the Adriatic with the Northern Ionian Sea fishes populations [1]. The water masses flowing from the Adriatic into the Ionian Sea allow to connect areas in which cold-water corals can thrive [5]. Understanding the functioning of this CWC community network is a crucial issue because of its importance, both in terms of the habitat conservation and for

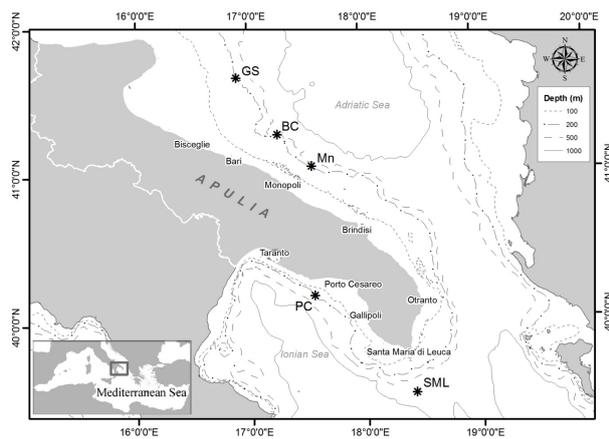


Figure 1: CWC community network along the Apulian coasts

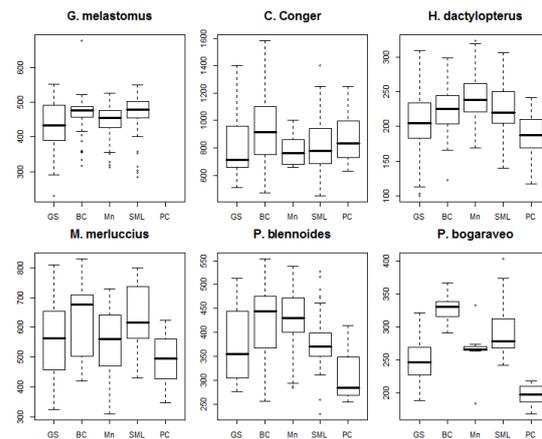


Figure 2: Size species distributions for five CWC areas

the management of the living resources impacted by commercial fishing [2, 3]. As a consequence, the protection and conservation of CWC areas have become a recognised priority whereby effective management and biodiversity objectives measures might be developed. The aim of this paper is to characterize the species assemblages of the CWC community network existing in the Southern Adriatic-Northern Ionian. We consider the fishes length as an indicator for investigating the main differences among the community network areas. An approach based on Generalized Additive Mixed Models (GAMMs) is proposed. These represent a wide class of models in the context of structured additive regression [4] and can be considered an additive extension of Generalized Linear Mixed Models using additive non parametric functions to model covariate effects and accounting for overdispersion and correlation by random terms. GAMMs are suitable for nested and crossed designs and are applicable to clustered, hierarchical and spatial data [8]. For this reason, GAMMs are particularly suited for analyzing the case study data grouped by areas. The estimation of parameters associated to specific model terms is obtained in a Bayesian perspective, following the approach suggested in [9] which is based on a flexible priors structure, named *spike-and-slab*, and implemented in the related **spikeSlabGAM** R package for fully Bayesian variable selection and model choice.

2 Material and Methods

The CWC areas investigated are distributed along the Apulian coasts as shown in Figure 1. A total amount of five areas are located between the Southern Adriatic (Gondola Slide - GS, Bari Canyon - BC, off Monopoli - Mn) and the Northern Ionian sea (the CWC province Santa Maria di Leuca - SML, off Porto Cesareo - PC). We consider data from experimental longlines surveys collected among 2013 and 2014, minimizing the impacts on the sea-floor and benthic fauna. In particular, to study the species assemblages, the sizes of six mostly abundant fish species in each area are considered: *Galeus melastomus*, *Conger conger*, *Helicolenus dactylopterus*, *Merluccius merluccius*, *Pagellus bogaraveo*, *Phycis blennoides*. Boxplots in Figure 2 report the distributions of the lengths for each species and area, showing for some species (e.g. *P. bogaraveo* or *P. blennoides*) differences in the sizes between areas. To account for various effects of variability components on the overall length of fishes in the five CWC areas, we propose a GAMM approach modelling dependence of the size response on the fish species, depth and abundance covariates. Generally, given a set of covariates $x_j (j = 1, \dots, p)$, the distribution

	$P(\gamma = 1)$	pi	
sm(abund)	0.467	-0.005	*
lin(depth)	0.999	0.069	***
fct(species)	1.000	0.940	***
rnd(area)	0.274	-0.004	*

Table 1: Marginal posterior inclusion probabilities and terms of importance pi for each model component. *: $P(\gamma = 1) > 0.25$, **: $P(\gamma = 1) > 0.5$, ***: $P(\gamma = 1) > 0.9$.

of the response y in a GAMM belongs to the exponential family and the conditional expected value $E(y|x_1, \dots, x_p) = h(\eta)$ is determined by the additive predictor η including a wide variety of model terms as linear terms, nominal or ordinal covariates, smooth functions of continuous covariates (spline, tensor product or varying coefficient term), random effects (subject-specific intercepts and slope) and interactions between the different terms. For the case study, let y_{it} denote the standardized length of the t -th individual observed in the i -th area, with $t = 1, \dots, 1605$ and $i = 1, \dots, 5$. Here we assume that the standardized response has a Gaussian distribution and that the additive predictor is related to the expected value of the data through the identity link function (h):

$$\eta_{it} = \beta_1 f_1(abund_{it}) + \beta_2 f_2(depth_i) + \beta_3 species_t + \beta_4 area_i \quad (1)$$

where functions $f_{1,2}$ are unknown smooth functions respectively associated to the abundance and depth covariates, while the covariate species is a six levels factor forced to be a fixed component. Coefficients $\beta_j, j = 1 \dots, 4$ related to each model term are to be estimated and the area is introduced as a random effect. As a consequence the model intercept is allowed to vary for each CWC area accounting for correlation between observations belonging to the same area. In a Bayesian perspective, inference on GAMM parameters can be carried out considering all model components as random variables supplemented by appropriate prior assumptions. The estimation of marginal posterior probabilities of terms in Eq. (1) can be achieved using *spike-and-slab* priors (peNMIG) which are bimodal priors for the regression coefficients that are decomposed in a two component mixture of a narrow spike around zero and a slab with wide support for the marginal prior of the coefficients themselves [7]. The generic peNMIG prior for β_j can be viewed as a scale mixture of Gaussians, while the variance prior common to all β coefficients is Inverse Gamma. The posterior mixture weight for the spike component of a specific coefficient can be interpreted as the posterior probability of its exclusion from the model. The proposed approach is implemented by means of **SpikeSlabGAM** R package which is suitable for Gaussian, Binomial and Poisson responses and features efficient model selection as well as model choice.

3 Results

Table 1 shows the MCMC results of 3 chains each running 10000 iterations after a burn-in phase of 500 and thinning by 5. The graphical inspection of trace plots highlights strong evidence of good mixing and convergence for the 3 chains and all parameters involved with the proposed model, thus reaching the target posterior distribution. The shape of each effect is chosen according to the nature of the corresponding covariate, thus the abundance is associated with a univariate penalized spline (B-spline), the depth with a linear trend, the species with a factor effect and the area with a random intercept. The posterior probabilities of inclusion for this terms are all significant and the shape of fitted effects are reported in Figure 3.

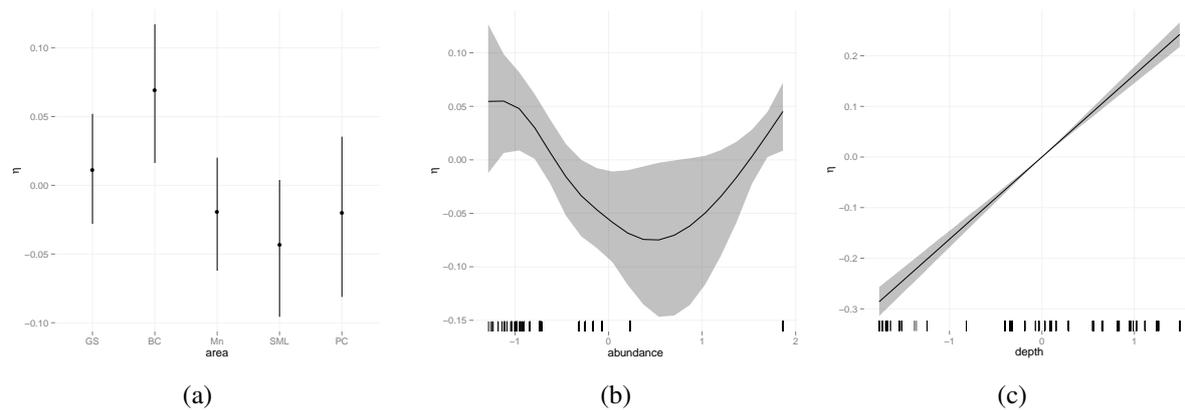


Figure 3: Posterior means and pointwise 80% credible intervals for some fitted model components.

The abundance has a non linear decreasing effect on the length of fishes (smaller individuals are more aggregated), while the depth has a linear increasing trend according to the bigger-deeper phenomenon. The fitted random effect area shows as the lengths differ in the five CWC areas as a consequence of bathymetric ranges investigated. The species factor has a natural expected effect on the length of all fishes (not reported in Figure 3).

In this paper an efficient alternative to deal with GAMMs is proposed. The approach based on the Bayesian estimation of model parameters is very appealing allowing to specified complex models concisely. The handling of many kind of effects makes this method particularly recommended for the analysis of nested or hierarchical structured ecological data.

References

- [1] Angeletti, L., Taviani, M., Canese, S., Fogliani, F., Mastrototaro, F., Argnani, A., Trincardi, F., Bakran-Petricioli, T., Ceregato, A., Chimienti, G., Mačić, V., Poliseo, A. (2014). New Deep-water Cnidarian Sites in the Southern Adriatic Sea. *Mediterranean Marine Science*, **15**(2), doi: <http://dx.doi.org/10.12681/mms.558>.
- [2] Armstrong, C. W., Grehan, A. J., Kahui, V., Mikkelsen, E., Reithe, S., Van den Hove, S. (2009). Bioeconomic Modeling and the Management of Cold-Water Coral Resources. *Oceanography*, **22**(1),86–91.
- [3] Grehan, A.J., Unnithan, V., Olu-Le Roy, K., Opderbecke, J. (2005). Fishing impacts on deepwater coral reefs: making a case for coral conservation. In: Barnes, P.W., Thomas, J.P. (Eds.), *Benthic Habitats and the Effects of Fishing*. American Fisheries Society, Bethesda, MD, 819–832.
- [4] Fahrmeir, L., Kneib, T., Lang, S. (2004). Bayesian Regularization in Structured Additive Regression: a Unifying Perspective on Shrinkage, Smoothing and Predictor Selection. *Statistics and Computing*, **20**(2), 203–219.
- [5] Freiwald, A., Beuck, L., Rüggeberg, A., Taviani, M., Hebbeln, D. (2009). The White Coral Community in the Central Mediterranean Sea Revealed by ROV Surveys. *Oceanography*, **22**(1),36–52.
- [6] Hastie, T.J., Tibshirani, R.J. (1990). *Generalized Additive Models*. London: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.
- [7] Ishwaran, H., Rao, J.S. (2005). Spike and Slab Variable Selection: Frequentist and Bayesian Strategies. *Annals of Statistics*, **33**(2), 730–773.
- [8] Lin, X., Zhang, D. (1999). Inference in generalized additive mixed model using smoothing splines. *Journal of the Royal Statistical Society, Series B*, **61**, 381–400.
- [9] Scheipl, F. (2011). spikeSlabGAM: Bayesian Variable Selection, Model Choice and Regularization for Generalized Additive Mixed Models in R. *Journal of Statistical Software*, **43**(14), 1–24.