STATISTICA, anno LXXIII, n. 1, 2013

SPATIO-TEMPORAL MODELLING OF ZERO-INFLATED DEEP-SEA SHRIMP DATA BY TWEEDIE GENERALIZED ADDITIVE MODELS

Simona Arcuti, Crescenza Calculli, Alessio Pollice
Department of Economics and Mathematics, University of Bari Aldo Moro, Bari, Italy
Gianfranco D'Onghia, Porzia Maiorano, Angelo Tursi
Department of Biology, University of Bari Aldo Moro, Bari, Italy

1. INTRODUCTION

In the Mediterrean Sea population parameters of demersal resources fluctuate over spatial and temporal scales due to the variability of abiotic and biotic factors as well as to human activities. Many changes in the environmental conditions occurred in the Ionian Sea at the end of 1980, due to the Eastern Mediterranean Transient (EMT) and to an internal process occurring between Adriatic and Ionian water masses on a decadal scale called Bimodal Adriatic-Ionian Oscillations (BiOS) (Gačić *et al.*, 2010). The EMT-BiOS related regime shift recorded in the early 1990s involved sea surface temperature, sea level pressure, surface and deep circulation of water masses, salinity and nutrient changes, affecting several marine ecosystem components encompassing more trophic levels and different Mediterranean basins (Civitarese *et al.*, 2010; Conversi *et al.*, 2010).

The changes in abundance observed in some demersal resources in the North-Western Ionian Sea were identified to be linked to both changes in physical environment and those due to fishing (D'Onghia *et al.*, 2012). These changes were only detected during time when no spatial analyses were carried out. Indeed, the North-Western Ionian Sea, along the Italian coasts, covers a geographic area of about 3° 50' in latitude and 3° in longitude. In this area, the deep-water shrimps *Parapenaeus longirostris* (deep-water rose shrimp, Lucas, 1846) and *Aristaeomorpha foliacea* (giant red shrimp, Risso, 1816) represent two demersal resources of primary importance. *P. longirostris* is widespread at depths between 18 and 711 m (Maiorano *et al.*, 2010) and mostly on continental shelf and shelf edge (D'Onghia *et al.*, 1998b), preferring waters with high temperature and salinity (14-15 °C and 38 psu, respectively) (Ghidalia and Bourgois, 1961; Ungaro and Gramolini, 2006). Though *A. foliacea* is found at depths between 127 and 1145 m (Maiorano *et al.*, 2010) it is preferentially distributed on the upper slope (350-700 m), showing a shallower distribution in the winter-spring and a deeper one during summer-autumn (Tobar and Sardá, 2010; D'Onghia *et al.*, 1998a). The giant red shrimp seems to prefer high values of the water temperature (approximately 13.5°C) and salinity (38.5 psu) that are typical characteristics of Levantine Intermediate Water (LIW) (Ghidalia and Bourgois, 1961).

The aim of the present study is to evaluate the relationships between the spatio-temporal distribution of these two shrimps and environmental variables in the North-Western Ionian Sea. Data from trawl surveys have been analyzed in a spatio-temporal domain from 10 down to 800 m and between years 1995 and 2006. Here the main modelling issue refers to accounting for spatio-temporal dependencies of zero-inflated continuous responses. Previous proposals include the use of Zero Inflated Generalized Additive Models (ZIGAM's) proposed by Liu and Chan (2008). This approach consists in assuming the existence of two data generating processes: one process regarding the zeroes and another one for non-zero values. In many cases, the two processes are influenced by common factors and we can presume that they are bound by a proportionality constraint. In this case we talk about COnstrained Zero-Inflated Generalized Additive Models (COZIGAM's). As an alternative we propose to use Generalised Additive Models with a particular distributional assumption allowing for zero-inflation. While subsection 2.1 contains a detailed description of the sampling method and data set, in 2.2 we introduce spatio-temporal Generalized Additive Models and a specific class of distributions suitable to model zero-inflated abundance data. In Section 3 we summarize the main results achieved and discuss their relevance in connection with factors affecting the fauna distribution in the marine ecosystem. Finally Section 4 is dedicated to some concluding remarks and the illustration of possible further developments of the research.

2. Materials and Methods

2.1. Data collection

Fishery data were collected during experimental trawl surveys conducted from 1995 to 2006 in the North-Western Ionian Sea as part of the international project MEDITS (MEDiterranean International Trawl Surveys). The study area runs from Capo d'Otranto (LE) (40° 06' N - 18° 31' E) to Capo Passero (SR) (36° 41' N - 15° 10' E) for a total surface of 16,350 km^2 at depths between 10 and 800 m. This Geographical region was divided into three sub-areas: Apulia; North-Calabria and South-Calabria (Figure 1).

The sampling design adopted was random-stratified by depth. With regard to each depth range, the allocation of hauls was defined proportionally to the extension of the corresponding stratum. During each experimental haul, a Scanmar Sonar System (Fiorentini *et al.*, 1994) was placed on the trawl net in order to provide information about its proper functioning and horizontal opening, the latter being used to standardize abundance data to the swept surface unit. This standardization provides density (N/km^2) and biomass (kg/km^2) indices for each species and survey at every haul location. Finally the median carapace length to the nearest mm was obtained for individuals of the two species sampled at every haul. Data for *P. longirostris* regards 247 trawl hauls carried out between 100 and 500 m while those for *A. foliacea* includes 369 trawl hauls from 500 down to 800 m.

Density and biomass indices, and carapace length (mm) of the two species were considered as population features (or population variables in what follows) influenced by environmental factors. Temperature might be one of the primary factors producing changes in the mentioned population variables. Sea surface temperature (SST) data for the three areas and 12 years were obtained using the AVHRR satellite data (Advanced

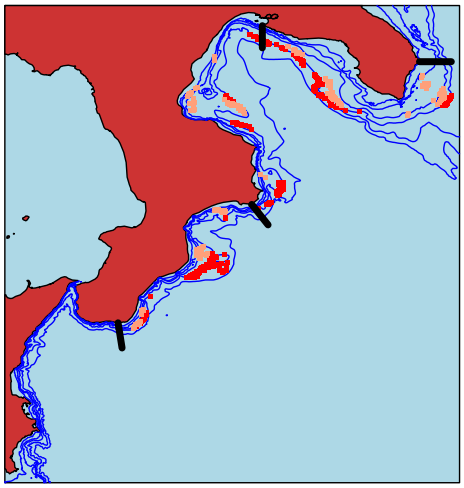

Figure 1 – Map of the study region divided into three sub-areas with bathyal curves and haul locations (*P. longirostris* in pink, *A. foliacea* in red).

Very High Resolution Radiometer, `http://podaac.jpl.nasa.gov/`). Precipitation values (expressed in mm) for the three areas and 12 years were recorded from meteorological archives on the web (`www.ilmeteo.it`). Fluctuations in the difference of atmospheric pressure between Iceland and the Azores contribute to the so called North Atlantic Oscillation (NAO). Variations of the NAO influence the weather on a large scale over the North Atlantic and Europe and have a strong impact on the oceanic conditions, including the Mediterranean region. The normalized difference in pressure between Algiers and Cairo is known as the Mediterranean Oscillation Index (MO). Values of the two barometric indices for the selected time period were obtained on the web (`http://www.cgd.ucar.edu/`) and considered for their influence on demersal resources. The depth of the experimental trawling hauls was also considered as an environmental factor influencing the species distribution. Finally, as an anthropogenic influence on the fishing resources, a proxy measure of the fishing effort was obtained as the number of potential working days according to weather and sea conditions, public holidays, required temporary closures of fishing activities and technical stops. This proxy was evaluated for every geographical area and every year involved in the analysis.

2.2. *Spatio-temporal modelling*

A high proportion of zero counts occurs in the two species (30% of 369 catches for the giant red shrimp, and 15% of 247 catches for the deep-water rose shrimp), causing zero-inflated distributions for density and biomass data (zero counts correspond to NA's in the median carapace length, which has thus a non-zero-inflated distribution for both species): such zeroes are structural depending on the adaptation of the species to the variable environmental conditions in the Mediterranean basin. Fishery trawl survey

data often contain a large number of zero catches, due to the fact that fishes swim in schools influenced by food availability and irregular current patterns (Ciannelli *et al.*, 2007). Zero-inflated data abound in ecological studies as well as in other scientific and quantitative fields, where the data contain an excess of zero responses. Statistical methods for zero-inflated data find application in a wide range of studies including the count of rare species (Cunningham and Lindenmayer, 2005), the distribution of fish eggs and larvae (Fox *et al.*, 2000) and of adult fishes (Bi *et al.*, 2007), as well as pests in agricultural systems (Ruiz Cárdenas *et al.*, 2009). The term "inflated" is used to stress that the probability of getting zero values is higher than the one expected under some classic distributional assumption (e.g. log-normal, Poisson, etc.). Such data cannot be analyzed by standard statistical distributions (e.g., log-normal, Poisson, etc.) because of the high proportion of zeroes (Zuur *et al.*, 2009).

With the purpose of accounting for zero-inflation in the data distribution, an exponential dispersion model (EDM) was assumed for the density and biomass indices. The exponential dispersion model is a two-parameter family of distributions consisting of a linear exponential family with an additional dispersion parameter (Dunn *et al.*, 2009). Members of EDM have density functions or probability mass functions of the form:

$$f(y|\theta;\phi) = \alpha(y,\phi)\exp\left[\frac{y\theta - k(\theta)}{\phi}\right] \tag{1}$$

where: α and k are known functions, θ is the canonical parameter and $\phi > 0$ is the dispersion parameter. The domain of the canonical parameter θ is an open interval satisfying $k(\theta) < \infty$. The function $k()$ is called the cumulant function of the EDM because, if $\phi = 1$, the derivatives of k give the successive cumulants of the distribution. In particular, the mean of the distribution is $\mu = k'(\theta)$ and the variance is $\phi k''(\theta)$. The mapping from θ to μ is invertible, so we may write $k''(\theta) = V(\mu)$ for a suitable function $V(\mu)$, called the variance function of the EDM, which describes the mean-variance relationship of the distribution when the dispersion is held constant. If Y is distributed as a member of the EDM with mean μ, variance function $V()$ and dispersion ϕ, then the variance of Y can be written as $\mathrm{Var}(Y) = \phi V(\mu)$. Of special interest is the class of EDMs with power mean-variance relationships for which $V(\mu) = \mu^p$ for some p, called index parameter. Following Jørgensen (1997), we call these Tweedie distribution models. Tweedie models are natural candidates for modelling continuous positive data (Dunn and Smyth, 2008), as they include most of the important distributions commonly used with generalized linear models such as the normal ($p = 0$), Poisson ($p = 1$), gamma ($p = 2$) and the inverse Gaussian ($p = 3$) distributions. In this study a value of p between 1 and 2 was considered as in this case the Tweedie model can be represented as a mixed distributions with a positive mass at zero and with support on the non-negative reals. In this special case Tweedie distributions are also known as "compound Poisson" (Jørgensen and Smyth, 2002) as they can be obtained as Poisson mixtures of gamma distributions. Tweedie distributions with $1 < p < 2$ are thus especially appealing to model zero-inflated data. Tweedie distributions are used in a diverse range of fields including fisheries (Candy, 2004; Shono, 2008; Tascheri *et al.*, 2010) and rainfall prediction (Dunn, 2004). Dunn *et al.* (2009) give a survey of published applications. Maximum likelihood estimates of p are obtained by univariate optimization of the profile likelihood evaluated on a grid of possible values of p, implemented in the `R` package `Tweedie` (Dunn,

2013).

A general approach to data modelling for the three biological variables and for both species, consists in assuming that the distribution of the response is affected by a number of predictors. In this work Generalized Additive Models (GAMs) are considered (Hastie and Tibshirani, 1990), in which the mean of the response depends on an additive predictor through a link function and the response distribution can be any member of the exponential family. GAMs allow for unknown smooth functions of observed explanatory variables in the linear predictor and have already been widely used in the spatio-temporal modelling of marine species (Borchers *et al.*, 2009; Augustin *et al.*, 1998; Daskalov, 1999). In these studies the authors choose to use GAMs to model the spatio-temporal distribution of species abundance as a function of geographical and environmental variables. According to empirical evidence, space-time *nonseparable* models were estimated for each variable and for both species. For zero-inflated count data, an alternative nonseparable spatio-temporal model specification was adopted in Musio *et al.* (2009). Models for our zero-inflated continuous responses are specified as follows:

$$g(\mu_i) = \beta_0 + \mathrm{te}(\mathrm{lon}_i, \mathrm{lat}_i, t_i) + \sum_{j=1}^{m} s_j(x_{ij}) = \eta(x_i)$$

where g is the function linking the mean response μ_i to the additive predictor $\eta(x_i)$. Semi-parametric smooth functions of observed covariates s_j's are specified as *thin plate regression splines* by linear combinations of basis functions. The term te is a *tensor product spline*, a non-isotropic smooth function useful for representing the simultaneous effects of multiple predictors measured in different units or where very different degrees of smoothness are appropriate (such as spatial coordinates and time points). Tensor product smoothers are invariant to linear rescaling of covariates, and can be quite computationally efficient. Also tensor product splines can be expressed as linear combinations of multivariate basis functions.

The use of spline functions allows to specify different features of algorithmic and statistical efficiency by the automatic selection of the smoothing parameters, i.e. of the smoothing degree applicable to every term. In the spline case this is translated in the specification of the number of effective degrees of freedom to assign to every smooth function. To avoid overfitting estimates, possibly due to a large number of highly nonlinear smooth functions, GAMs are usually estimated by penalized likelihood maximization, solved by Penalized Iteratively Reweighted Least Squares, P-IRLS (Wood, 2006), in which the model likelihood is modified by the addition of a penalty for each smooth function, penalizing its "wiggliness". Basic GAM fitting implies the minimization of:

$$||y - X\beta||^2 + \sum_{j=1}^{m} \lambda_j \int_0^1 s_j''(x)dx \tag{2}$$

where the model matrix X contains a set of basis functions for each smoother s_j and covariate values for every linear effect, β is a parameter vector and y is the response variable. The second part of this equation is a penalty and allows to controll the trade-off between goodness of fit and model smoothness. The second-order derivatives of the smoothing functions s_j's, summarize the smoothness of the nonlinear terms. If λ_j

is very large, the corresponding smoother will be a straight line. On the other hand, if λ_j is small we are likely to end up with a considerably less smooth curve. Because s_j's are linear in the parameters, the penalty can always be written as a quadratic form, and GAM fitting can be resumed addressing a least squares problem given λ_j's. The estimation of the smoothing parameters λ_j is then addressed minimizing predictor error criteria like Generalized Cross Validation (GCV) or Un-Biased Risk Estimator (UBRE). When the scale parameter is known the expected mean square error minimization leads to estimation by UBRE, otherwise GCV is called for. In this study GAM's fit was performed by `R` function `mgcv`, (Wood, 2006) implementing P-IRLS estimation with GCV and UBRE. A Tweedie model was adopted for the zero-inflated distributions of the density and biomass of both species.

Terms significance and overall GCV scores were considered for model comparison. A given model was compared to a model with one or more smooth terms replaced by parametric terms or removed. Smooth terms candidate for replacement by parametric terms have estimated degrees of freedom close to their minimum. Smooth terms candidate for removal are identified by reference to approximate p-values. As a general criterion we choose to preserve the significance of spatio-temporal effects, eventually avoiding the inclusion of overlapping effects.

3. RESULTS

In this section we report the results of applying GAMs to the data described in Section 2.1 assuming Tweedie distributions for the zero-inflated population variables density and biomass. For these two response variables, in order to get a better fit to the assumed Tweedie family, we considered the logarithmic link function. For the same reason the squared root transformation was adopted for the density of the rose shrimp and the biomass of both species. As expected, estimates of the Tweedie index parameter p obtained as mentioned in Section 2.2 and reported in Table 1 and 2 imply zero-inflation of the relative distributions. The non-zero-inflated median carapace length was assumed normally distributed with identity link function for both species.

Table 1 and 2 show the estimates of linear and smooth effects involved with the spatio-temporal distribution of the features of the two shrimp populations. Significant spatio-temporal effects correspond to the distributions of both species being related to some areas along the Ionian arc. During 1995 and 2006, *P. longirostris* was found to be more abundant along the South Calabria and *A. foliacea* along Apulia. In these years the greatest sizes were mostly detected along this latter region in both species (Figure 3). *P. longirostris* shows a nonlinear relation with depth (Figure 3) with higher values of the density for depths between 200 and 300 m and a decrease beyond 300 m. Density and biomass indices of *A. foliacea* decrease with depth, while the relationship of the median carapace length with depth is linear, with positive coefficient, for both species (Table 1 and 2). On the one hand, the geographic and depth distribution of both species along the Ionian arc reflects their life strategies (D'Onghia *et al.*, 1998b, 1998a), on the other hand, the spatio-temporal fluctuations observed throughout the study period could be related to the modification of the water mass properties and thermohaline circulation occurred in the Ionian Sea (e.g. Conversi *et al.*, 2010; Gačić *et al.*, 2010). Winter NAO index significantly influenced the three population variables of *P. longirostris* and the

TABLE 1
Model estimates for Parapenaeus longirostris [a]

		Density		Biomass		Length	
Index parameter	*p*	1.26	(1.195, 1.356)	1.26	(1.197, 1.364)		
Linear terms coefficients	Int.	3.33	(< 0.001)	0.80	(< 0.001)	25.30	(< 0.001)
	depth					0.04	(< 0.001)
	SST					-0.58	(0.0198)
Smooth effects degrees of freedom	long, lat, year	14.34	(< 0.001)	14.70	(< 0.001)	22.82	(< 0.001)
	depth	4.55	(< 0.001)	4.55	(< 0.001)		
	WNAO	4.49	(0.020)	4.60	(0.019)	2.73	(0.008)

[a] Index parameter estimates with 99% confidence intervals (in brackets) and GAMs estimates with p-values (in brackets) for *P. Longirostris* density, biomass and median length. Linear coefficients and smooth effects degrees of freedom: intercept, longitude, latitude, year, depth, Sea surface temperature, Winter NAO global barometric index.

TABLE 2
Model estimates for Aristaemorpha foliacea [b]

		Density		Biomass		Length	
Index parameter	*p*	1.74	(1.712, 1.888)	1.33	(1.249, 1.409)		
Linear terms coefficients	Int.			-2.51	(0.0621)	56.27	(< 0.001)
	depth	-0.01	(< 0.001)	-0.01	(< 0.001)	0.03	(0.002)
	SST	0.46	(< 0.001)	0.26	(< 0.001)		
	f.effort					-0.17	(< 0.001)
Smooth effects degrees of freedom	long, lat, year	21.60	(< 0.001)	11.99	(< 0.001)	16.69	(0.002)
	SST					5.67	(< 0.001)
	MO					3.01	(0.006)

[b] Index parameter estimates with 99% confidence intervals (in brackets) and GAMs estimates with p-values (in brackets) for *A. foliacea* density, biomass and median length. Linear coefficients and smooth effects degrees of freedom: intercept, longitude, latitude, year, depth, Sea surface temperature, fishing effort, MO global barometric index.

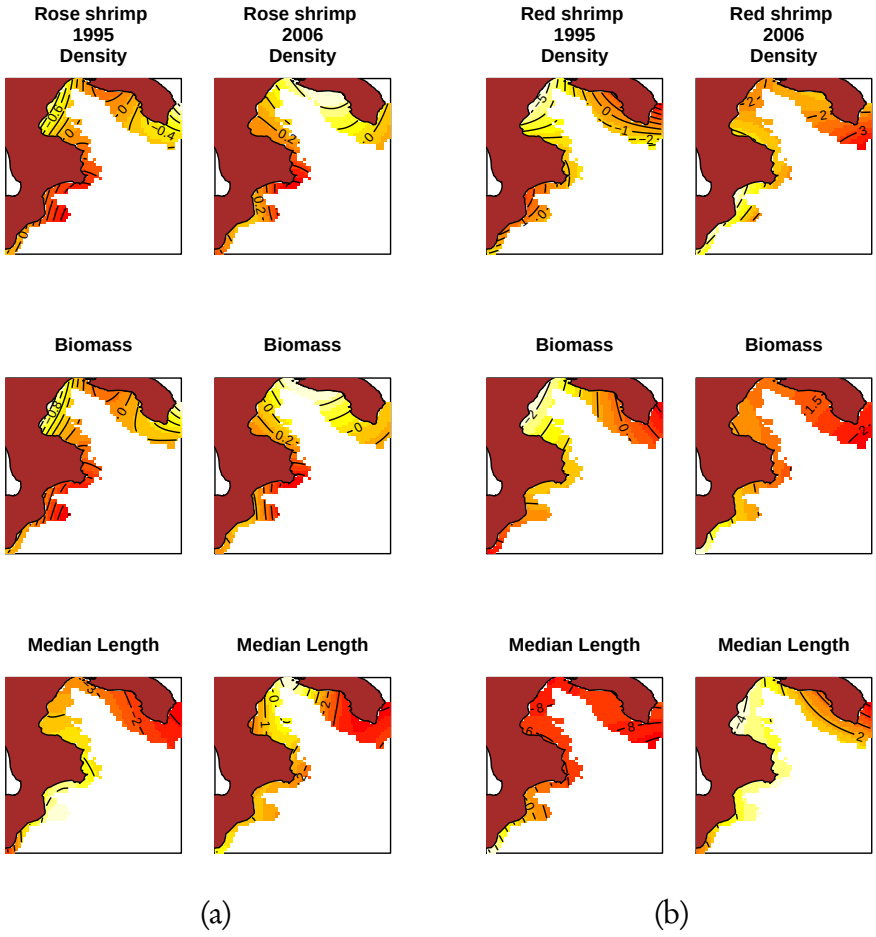

Figure 2 – Spatio-temporal smooth components of the fitted GAMs for years 1995 and 2006. (a) *P. Longirostris*, (b) *A. Foliacea*. The yellow coloured areas correspond to lower values of species density, biomass and median carapace length. In reverse, the warm colours indicate areas characterized by higher values of the population variables for the two species.

fitted smooth effects are shown in Figure 4. It is evident an increasing behavior of the Winter NAO effect for the median length, whereas it is more fluctuating for the density and biomass. Also a nonlinear effect of MO index on the median carapace length has been detected for *A. foliacea* (Figure 5). The Sea surface temperature shows a significant inverse linear relation with the median carapace length of *P. longirostris*, it has a positive linear influence on the density and biomass of *A. foliacea* (Table 2) and a nonlinear effect on its median carapace length (Figure 5). In particular, both species seem to be influenced by the variations in the thermal regime (WNAO, MO, SST) which could reflect the increased presence of the saline and warm Levantine Intermediate Waters (LIW) in the study area. This could have influenced the recruitment and, thus, the abundance level of the two shrimps. Indeed, the negative linear relation detected between SST and size in *P. longirostris* could be due to the fact that warm waters have a positive influence on recruitments with a consequent reduction of the size in the whole stock. Ungaro and Gramolini (2006) found a possible influence of the water temperature at sea bottoms on the spatial distribution of P. longirostris in the southern Adriatic. Recently, Ligas *et al.* (2010) reported an increasing trend of *P. longirostris* abundance correlated to a rise in SST, a corresponding decrease of wind circulation and to the reduction of fishing effort. Capezzuto *et al.* (2010) revealed that the increase in the abundance of *A. foliacea* verified mostly in the period 2000-2004 correlated significantly with the increase in temperature and salinity detected from 1995 to 2005 between 200 and 800 m in the Ionian Sea. This fact and present observations can be explained according to the hydrological hypothesis by Ghidalia and Bourgois (1961): *A. foliacea* is preferentially distributed in warmer and high salinity waters. In the western Mediterranean the stock abundance of another deep-water shrimp, *Aristeus antennatus*, was detected to be enhanced by high MO index periods reflecting the increased presence of LIW (Massutí *et al.*, 2008). Moreover, Maynou (2008) found a positive correlation between the average annual NAO index and the annual catches of the shrimp *Aristeus antennatus* in six ports along the Catalonia coast. This author hypothesized that the NAO-induced environmental variability could favour secondary production enhancing the food supply for *A. antennatus*. The greatest sizes observed along the Apulian coast in both species could be a consequence of a lower fishing pressure in this geographic sector. However, as expetcetd, a significant inverse relationship between fishing effort and size was only detected in *A. foliacea* (Table 2).

Due to the high vulnerability to trawling of both shrimp species, a reduction in the working days of this type of fishing could reliably favour stock recovery and, thus, an increase of the size. The fact that the same inverse relationship has not been observed in *P. longirostris* seems to be related to an increase of density, biomass and sizes in this shrimp during the last years of the esamine period. The faster recovery of this shrimp could due to its wider distribution and shorter life span (D'Onghia *et al.*, 1998b). For *P. longirostris*, it seems that the observed positive effects due to changes in environmental conditions dominated on those negative ones due to fishing pressure.

4. Concluding remarks

The two shrimps *Parapenaeus longirostris* and *Aristaeomorpha foliacea* are among the most important deep-sea demersal resources in the North-Western Ionian Sea. Their

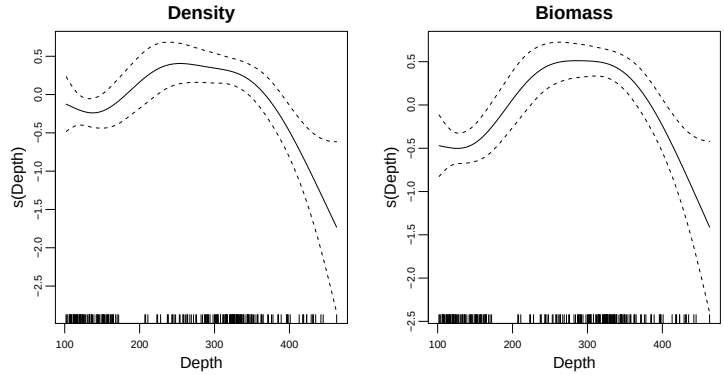

Figure 3 – Fitted depth smooth effects for *P. longirostris*

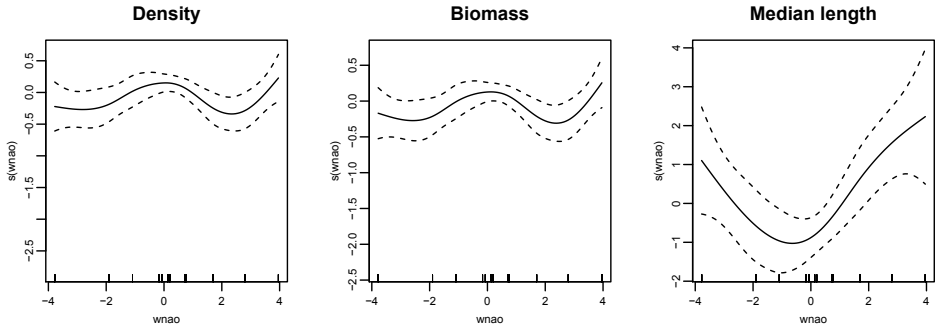

Figure 4 – Fitted WNAO smooth effects for *P. longirostris*

changes in terms of density, biomass and median length induced by anthropogenic and environmental variables are here investigated considering a spatio-temporal non-separable model. Figure 3 shows that the effect of longitude and latitude with respect to time is rather complex and supports well this choice. On the other hand depth is included as a separate additive term in the model (Figure 3 and Table 1 and 2 display the impact of depth), assuming the effect of depth to be constant with respect to time and horizontal positioning. This simplifying assumption is reasonable as, although both *P. longirostris* and *A. foliacea* can be found on a wide depth range, their constant preferential depth distribution is well documented in many studies throughout the Mediterranean. In particular, *P. longirostris* is mostly found between 150 and 400 m with adults deeper than juveniles (D'Onghia *et al.*, 1998b; Abelló *et al.*, 2002) while *A. foliacea* has preferential distribution at depths of 400-700 m without clear separation between adults and juveniles (D'Onghia *et al.*, 1998a; Cau *et al.*, 2002). With the previous assumptions, Generalized Additive Models are used to evaluate the spatio-temporal variation of the population features in both species, together with the possible nonlinear effects of biotic and abiotic

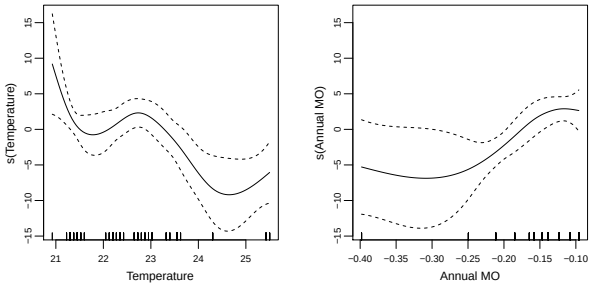

Figure 5 – Fitted Temperature and Annual MO smooth effects of the median length for *A. foliacea*

factors. Mechanisms producing changes in population variables act in high dimensional complex natural systems and often remain unexplained. For the same reason, it is often difficult to disentangle environmental effects on demersal communities from those due to fishing. The significance of the various predictors here considered reveals effects due to the geographic area, depth and year on the distribution of the two species, confirming their sensitivity to the changes in environmental conditions and human pressure (D'Onghia *et al.*, 2012). Tweedie distributions assumed for zero-inflated population features have specific advantages as they can be considered in standard GAM implementations such as the `R` library `mgcv`. Their adoption allows to analyze the interaction between space and time through a non-separable model with a tensor product splines component. Unlike the alternative ZIGAM approach, the proposed method does not provide information about the zero-generating process. This aspect is a major concern in view of possible future studies allowing to adequately estimate the proportion of zero records within spatio-temporal models. ZIGAM's and COZIGAM's have had a limited use in spatio-temporal modelling as their current implementation in R does not allow for smoothers of more than two dimensions. Separable spatio-temporal COZIGAM's were fit to *P. Longirostris* density and biomass data by the R library `cozigam`, leading to some numerical difficulties. Possible further developments include the consideration of spatial smoothers accounting for complex boundaries and the use of Bayesian hierarchical models implementing efficient computational algorithms and Bayesian methods for model choice.

REFERENCES

P. ABELLÓ, A. ABELLA, A. ADAMIDOU, S. JUKIC-PELADIC, P. MAIORANO, M. T. SPEDICATO (2002). *Geographical patterns in abundance and population structure of Nephrops norvegicus and Parapenaeus longirostris (Crustacea: Decapoda) along the European Mediterranean coast*. Scientia Marina, 66, suppl. 2, pp. 125 – 141.

N. H. AUGUSTIN, E. D. BORCHERS, S. T. CLARKE, S. T. BUCKLAND, M. WALSH (1998). *Spatiotemporal modelling for the annual egg production method of stock assessment using generalized additive models*. Canadian Journal of Aquatic and Fisheries Sciences, 55, pp. 2608 – 2621.

H. Bi, R. E. Ruppel, W. T. Peterson (2007). *Modeling the pelagic habitat of salmon off the Pacific Northwest coast using logistic regression*. Marine Ecology Progress Series, 336, pp. 249 – 265.

D. L. Borchers, S. T. Buckland, I. G. Priede, S. Ahmadi (2009). *Improving the precision of the daily egg production method using generalized additive models*. Canadian Journal of Aquatic and Fisheries Sciences, 54, pp. 2727 – 2742.

S. G. Candy (2004). *Modelling catch and effort data using generalized linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects*. CCAMLR Sci., 11, pp. 59 – 80.

F. Capezzuto, R. Carlucci, P. Maiorano, L. Sion, D. Battista, A. Indennidate, G. D'Onghia, A. Tursi (2010). *The bathyal benthopelagic fauna in the NW Ionian Sea: structure, patterns and interactions*. Chemistry and Ecology, suppl. 26, pp. 199 – 217.

A. Cau, A. Carbonell, M. C. Follesa, A. Mannini, G. Norrito, L. Orsi Relini, C. Y. Politou, S. Ragonese, P. Rinelli (2002). *MEDITS-based information on the deep-water red shrimps Aristaeomorpha foliacea and Aristeus antennatus (Crustacea: Decapoda: Aristaeidae)*. Scientia Marina, 66, suppl. 2, pp. 103 – 124.

L. Ciannelli, K. S. Chan, N. C. Stenseth (2007). *Phenological and geographical patterns of Walleye Pollock spawning in the Gulf of Alaska*. Canadian Journal of Aquatic and Fisheries Sciences, 64, pp. 713 – 722.

G. Civitarese, M. Gačić, M. Lipizer, G. L. Eusebi Borzelli (2010). *On the impact of the Bimodal Oscillating System (BiOS) on the biogeochemistry and biology of the Adriatic and Ionian Seas (Eastern Mediterranean)*. Biogeosciences, 7, pp. 3987 – 3997.

A. Conversi, S. Fonda Umani, T. Peluso, J. C. Molinero, M. Edwards (2010). *The Mediterranean Sea regime shift at the end of the 1980s, and intriguing parallelisms with other European basins*. Plos one, 5, no. 5, pp. 1 – 15.

R. B. Cunningham, D. B. Lindenmayer (2005). *Modeling count data of rare species: some statistical issues*. Ecology, 85, pp. 1135 – 1142.

G. Daskalov (1999). *Relating fish recruitment to stock biomass and physical environment in the black sea using generalized additive models*. Fish. Res., 41, no. 1, pp. 1 – 23.

G. D'Onghia, A. Giove, P. Maiorano, R. Carlucci, M. Minerva, F. Capezzuto, L. Sion, A. Tursi (2012). *Exploring relationships between demersal resources and environmental factors in the Ionian Sea (Central Mediterranean)*. Journal of Marine Biology, 2012. Article ID 279406, 12 pages.

G. D'Onghia, P. Maiorano, A. Matarrese, F. Perri (1998a). *Distribution, biology, and population dynamics of Aristaeomorpha foliacea (Risso, 1827) (Decapoda, Natantia, Aristeidae) in the North-Western Ionian Sea (Mediterranean Sea)*. Crustaceana, 71, no. 5, pp. 518 – 544.

G. D'ONGHIA, A. MATARRESE, P. MAIORANO, F. PERRI (1998b). *Valutazione di Parapenaeus longirostris (Lucas, 1846) (Crustacea, Decapoda) nel Mar Ionio*. Biol. Mar. Medit., 5, pp. 273 – 283.

P. K. DUNN (2004). *Occurrence and quantity of precipitation can be modelled simultaneously*. International Journal of Climatology, 24, pp. 1231 – 1239.

P. K. DUNN (2013). *tweedie: Tweedie exponential family models*. R package version 2.1.7.

P. K. DUNN, G. K. SMYTH (2008). *Evaluation of Tweedie exponential dispersion model densities by Fourier inversion*. Statistics and Computing, 18, pp. 73 – 86.

P. K. DUNN, G. K. SMYTH (2009). *Series evaluation of Tweedie exponential dispersion model*. Statistics and Computing, 15, pp. 267 – 280.

L. FIORENTINI, G. COSIMI, A. SALA, V. PALUMBO (1994). *Caratteristiche e prestazioni delle attrezzature a strascico impiegate per la valutazione delle risorse demersali in italia*. Biol. Mar. Medit., 1, suppl. 2, pp. 115 – 134.

C. J. FOX, C. M. O'BRIEN, M. DICKEY-COLLAS, R. D. M. NASH (2000). *Patterns in the spawning of cod (Gadus morhua l.), sole (Solea solea l.) and plaice (Pleuronectes platessa l.) in the Irish Sea as determined by generalized additive modeling*. Fisheries Oceanography, 9, pp. 33 – 49.

M. GAČIĆ, G. L. E. BORZELL, G. CIVITARESE, V. CARDIN, S. YAIRI (2010). *Can internal processes sustain reversals of the ocean upper circulation? The Ionian Sea example*. Geophysical Research Letters, 37, no. 9, pp. 1 – 5.

W. GHIDALIA, F. BOURGOIS (1961). *Influence de la temperature et de l'éclairement sur la distribution des crevettes des moyennes et grandes profondeurs*. In: Stud. Rev. Gen. Fish. Couns. Medit., FAO, 16, pp. 1 – 53.

T. HASTIE, R. J. TIBSHIRANI (1990). *Generalized Additive Models*. Chapman and Hall, London.

B. JØRGENSEN (1997). *The Theory of Dispersion Models*. Chapman and Hall, London.

B. JØRGENSEN, G. K. SMYTH (2002). *Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling*. ASTIN Bulletin, 32, pp. 143–157.

A. LIGAS, S. DE RANIERI, D. MICHELI, B. REALE, P. SARTOR, M. SBRANA, P. BELCARI (2010). *Analysis of the landings and trawl survey time series from the Tyrrhenian Sea (NW Mediterranean)*. Fisheries Research, 105, suppl. 1, pp. 46 – 56.

H. LIU, K. S. CHAN (2008). *Constrained generalized additive model with Zero-Inflated data*. Technical Report 388, The University of Iowa, Department of Statistics and Actuarial Science.

P. MAIORANO, L. SION, R. CARLUCCI, F. CAPEZZUTO, A. GIOVE, G. COSTANTINO, M. PANZA, G. D'ONGHIA, A. TURSI (2010). *The demersal faunal assemblage of the Northwestern Ionian Sea (central Mediterranean): Current knowledge and perspectives*. Chemistry and Ecology, 26, Suppl. June 2010, pp. 219 – 240.

E. MASSUTÍ, S. MONSERRAT, P. OLIVER, J. MORANTA, J. L. LOPEZ-JURADO, M. MARCOS, M. HIDALGO, B. GUIJARRO, A. CARBONELL, P. PEREDA (2008). *The influence of oceanographic scenarios on the population dynamics of demersal resources in the western Mediterranean: Hypothesis for hake and red shrimps of Balearic Islands*. Journal of Marine Systems, 71, pp. 421–438.

F. MAYNOU (2008). *Environmental causes of the fluctuation of red shrimps (Aristeus antennatus) landings in the Catalan Sea*. Journal of Marine Systems, 71, pp. 294 – 302.

M. MUSIO, E. A. SAULEAU, A. BUEMI (2009). *Bayesian semi-parametric ZIP models with space-time interactions: An application to cancer registry data*. Mathematical Medicine and Biology, 27, pp. 181 – 194.

R. RUIZ CÁRDENAS, R. M. A. AO, C. DEMÉTRIO (2009). *Temporal modelling of coffee berry borer infestation patterns accounting for inflation of zeroes and missing values*. Scientia Agricola, 66, pp. 100 – 109.

H. SHONO (2008). *Application of the Tweedie distribution to zero-catch data in CPUE analysis*. Fisheries Research, 93, pp. 154 – 162.

R. TASCHERI, J. C. SAAVEDRA-NIEVAS, R. ROA-URETA (2010). *Statistical models to standardize catch rates in the multi-species trawl fishery for Patagonian grenadier (Macruronus magellanicus) of Southern Chile*. Fisheries Research, 105, pp. 200 – 214.

R. TOBAR, F. SARDÁ (2010). *Análisis de la evolución de las capturad de gamba rosada, Aristeus antennatus (Risso, 1817) en los utimos decenios en Cataluña*. Informes Técnicos del Instituto de Ciencias del Mar de Barcelona, 142, pp. 1 – 20.

N. UNGARO, R. GRAMOLINI (2006). *Possible effect of bottom temperature on distribution of Parapenaeus longirostris (Lucas, 1846) in the Southern Adriatic (Mediterranean Sea)*. Turkish Journal of Fisheries and Aquatic Sciences, 6, pp. 109 – 116.

S. WOOD (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall CRC, Boca Raton, Florida.

F. ZUUR, E. N. IENO, N. WALKER, A. A. SAVELIEV, G. M. SMITH (2009). *Mixed effects models and extensions in ecology with R*. Springer, New York.

SUMMARY

Spatio-temporal modelling of zero-inflated deep-sea shrimp data by Tweedie generalized additive models

In the Mediterrean Sea the population features of demersal resources fluctuate over spatial and temporal scales due to the variability of abiotic and biotic factors as well as to human activities. The two shrimps *Parapenaeus longirostris* and *Aristaeomorpha foliacea* are among the most important deep-sea demersal resources in the North-Western Ionian Sea. Their changes in terms of density, biomass and median length induced by anthropogenic and environmental variables (fishing effort, sea surface temperature, precipitations, Winter North Atlantic Oscillation (NAO) and Annual Mediterranean Oscillation (MO) indices) were investigated. Biological data were collected during trawl surveys carried out from 1995 to 2006 as part of the international program MEDITS (International Bottom Trawl Survey in the Mediterranean). Generalized Additive Models were used to evaluate the spatio-temporal variation of both species, together with the possible nonlinear effects of biotic and abiotic factors. Density and biomass were assumed to be distributed according to a member of the Tweedie family in order to account for zero-inflation in the relative data. Space-time interaction was considered within a non-separable model with smooth spatio-temporal component based on tensor product splines. The results show significant spatio-temporal and depth effects in the three population parameters of these resources. Winter NAO index significantly influenced the density, biomass and length of *P. longirostris*. Sea surface temperature significantly influenced the size of this species and the three population features of *A. foliacea*. The size of this shrimp resulted also influenced negatively by fishing effort and positively by the MO index.