



**UNIVERSIDADE DE
SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE
ESTATÍSTICA E INVESTIGACIÓN OPERATIVA**

**Extended Additive Regression for Analysing LPUE Indices in
Fishery Research**

V. Mamouridis, N. Klein, T. Kneib, C. Cadarso Suárez, F. Maynou

Report 13-01

Reports in Statistics and Operations Research

Extended Additive Regression for Analysing LPUE Indices in Fishery Research

Valeria Mamouridis^{*1}, Nadja Klein², Thomas Kneib², Carmen M. Cadarso Suárez³ and Francesc Maynou¹

¹ Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain

²Georg-August University, Wilhelmsplatz 1, 37073 Göttingen, Germany

³Unidade de Bioestadística, Faculdade de Medicina e Odontologia, USC, Rúa de San Francisco, s/n, 15782 Santiago de Compostela, Spain

29/10/2013

Abstract

We analysed the landings per unit effort (LPUE) from the Barcelona trawl fleet (NW Mediterranean) of the red shrimp (*Aristeus antennatus*) using the novel bayesian approach of additive extended regression or distributional regression, that comprises a generic framework providing various response distributions, such as the log-normal and the gamma and allows estimations for location and scale or shape (as the frequentist counterpart GAMLSS). The dataset covers a span of 17 years (1992-2008) during which the whole fleet has been monitored and consists of a broad spectrum of predictors: fleet-dependent (e.g. number of trips performed by vessels and their technical characteristics, such as the gross registered tonnage), temporal (inter- and intra-annual variability) and environmental (NAO index) variables. This dataset offers a unique opportunity to compare different assumptions and model specifications. So that we compared 1) log-normal versus gamma distribution assumption, 2) modelling only the expected LPUE versus modelling both expectation and scale (or shape) of LPUE, and finally 3) fixed versus random specifications for the catching unit effects (boats). Our preliminary results favour the gamma over the log-normal and modelling of both location and shape (in the case of the gamma) rather than only the location, while not noticeable differences occur in estimation when considering catching units as fixed or random.

Citation: Mamouridis, V., Klein, N., Kneib T., Cadarso Suárez, C.M., Maynou, F. 2013. Extended Additive Regression for Analysing LPUE Indices in Fishery Research. Reports in Statistics and Operations Research. Universidade de Santiago de Compostela. Departamento de Estatística e Investigación. Vol. 29/10/2013 (2013-01). available at: <http://eio.usc.es/index.php/es/reports->

^{*}mamouridis@icm.csic.es

1 Introduction

In fishery research, the LPUE (Landings Per Unit Effort) is an index widely used in stock assessment to estimate the relative abundance of an exploited species (Mendelssohn and Cury, 1989; Marchal et al., 2002). It constitutes one of the most common pieces of information used in assessing the status of fish stocks and is reckoned in different ways depending on data availability. The “landings” portion of the measure is the quantity of the stock brought to the port by each vessel and is usually expressed as number of individuals or weight of the whole stock, while the “unit effort” portion refers to the unit of time spent by a unit of the gear used to catch (e.g. vessel or square meters of a net). Therefore, LPUE is a relative index, which use is based on the assumption that it is proportional to the natural quantity of the species, despite their proportionality has been debated in the past (e.g. Gulland, 1964; Bannerot and Austin, 1983).

The most commonly applied analyses on LPUE is its standardisation to remove the bias induced by influential factors that do not reflect the natural variability (Maunder et al., 2006). In fact, many factors can affect the index (e.g. time, seasonality, fishing area and fleet characteristics, among others), some of which (i.e. fishery related variables) if not considered can lead to biased interpretations of stock states. Here we model the LPUE using all available explanatory variables. For some of the influential variables, a simple linear impact on LPUE as often assumed in standard statistical models may not be flexible enough and alternative, semiparametric modelling approaches may be required. Moreover, in most cases LPUE data are collected repeatedly for the same catching units over time leading to the necessity to account for unobserved characteristics of these units to avoid correlations in the repeated measurements.

The most common class of regression models to determine the impact of covariates x_1, \dots, x_p on the expectation of the LPUE are generalized linear models (GLM, McCullagh and Nelder, 1989) and generalized additive models (GAM, Hastie and Tibshirani, 1986). In GLM the expectation of LPUE is related to a linear combination of the covariate effects, i.e.

$$E(\text{LPUE}_i) = h(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}), \quad i = 1, \dots, n$$

with a suitable transformation function h that ensures positivity of the expected LPUE and regression coefficients $\beta_0, \beta_1, \dots, \beta_p$ (for GLM applications in LPUE analyses see (see Goñi et al., 1999; Maynou et al., 2003). Popular special cases for modelling LPUE include the gamma distribution or the normal distribution applied to log-transformed LPUE values (which is equivalent to assume a log-normal distribution). To overcome the limitation of GLMs to purely linear effects of covariates, generalized additive models, GAM have been introduced (see Damalas et al., 2007; Denis, 2002) where now the expectation is specified as

$$E(\text{LPUE}_i) = h(\beta_0 + f_1(x_{i1}) + \dots + f_p(x_{ip})).$$

The nonlinear functions f_1, \dots, f_p remain unspecified and should be estimated flexibly from the data, for example using penalized spline approaches (see Wood, 2006; Ruppert et al., 2003).

Another direction for extending the GLM approach is the inclusion of catching unit-specific effects to acknowledge the effect that usually multiple observations are

collected and that unobserved heterogeneity remains even when accounting for some covariate effects. If the individual catching units are indexed as $i = 1, \dots, n$ and the repeated measurement for one catching unit are indexed as $j = 1, \dots, n_i$, the resulting model can be written as

$$E(\text{LPUE}_{ij}) = h(\beta_0 + \beta_1 x_{ij1} + \dots + \beta_p x_{ijp} + \alpha_i), \quad i = 1, \dots, n, j = 1, \dots, n_i.$$

The additional parameter α_i is introduced to stand for any effect specific to the catching unit that is not represented in the effects of covariates x_1, \dots, x_p . Of course, similar extensions can be defined for generalized additive models. In the statistical community, the most common assumption for α_i would be the specification as a random effect, i.e. α_i i.i.d. $N(0, \tau^2)$, to acknowledge the fact that the catching units represent a sample from the population of catching units. This leads to the class of generalized linear mixed models (GLMMs, Pinheiro and Bates, 2000) or generalized additive mixed models (GAMMs, Wood, 2006). An alternative specification is to treat the α_i as usual, fixed parameters that result from dummy coding of the catching units. This may be considered more appropriated if, for example, the complete fleet of catching units for a specific area has been observed. We will return to this debate later when discussing the methods in more detail, see also Bishop et al. (2004), Cooper et al. (2004) or Helser et al. (2004) for the use of mixed models in this field.

Another important aspect when modelling LPUE is the choice of the response distribution. In most cases, skewed distributions have been considered, including in particular the gamma distribution (Maynou et al., 2003; Stefánsson, 1996), the log-normal distribution (Brynjarsdóttir and Stefánsson, 2004; Myers and Pepin, 1990) and the delta distribution (Gavaris, 1980; Pennington, 1983). The latter provides a form of zero-adjustment where zero catches are modelled separately from the nonnegative catches via a Bernoulli distribution. As a possibility to differentiate between gamma and log-normal distribution, the Kolmogorov-Smirnov test has been applied to fitted values from corresponding GLMs. Then, the distribution leading to a lower value for the test statistic can be considered to be closer to the distribution of the data (Brynjarsdóttir and Stefánsson, 2004; Stefánsson and Palsson, 1998).

In this paper, structured additive distributional regression models (Klein et al., 2013b) has been considered as a comprehensive, flexible class of models that encompasses all special cases discussed so far and a number of further extensions. More specifically, this class of models allows to deal with the following problems:

- Selection of the response distribution: Additive distributional regression comprises a generic framework providing various response distributions and in particular the log-normal and the gamma distributions. Extensions including zero-adjustment to account for an inflation of observations with zero catch would also be possible but not required in the data set considered later. Tools for effectively deciding between competing modelling alternatives will also be considered.
- Linear versus nonlinear effects: Effects of continuous covariates can be estimated nonparametrically based on penalized splines approximations that allow for a data-driven amount of smoothness and therefore deviation from the linearity assumption.

- Fixed versus random effects: Both fixed and random specifications for the catching unit-specific effects are supported and can be compared in terms of their ability to fit the data.
- Models for location, scale and shape: instead of restricting attention to only modelling the expected LPUE, distributional regression allow to specify a further parameter of the distribution that correspond to scale or shape. This both enables for additional flexibility and a better understanding of how different covariates affect the distribution of LPUE.
- Mode of inference: Distributional regression can be formulated both from a frequentist and a Bayesian perspective and corresponding estimation schemes either rely on penalized maximum likelihood or Markov chain Monte Carlo simulations. This paper is focused on the Bayesian inference.

Structured additive distributional regression is in fact an extension of structured additive regression (STAR, Brezger and Lang, 2006; Fahrmeir et al., 2004) in the framework of generalized additive models for location, scale and shape (GAMLSS, Rigby and Stasinopoulos, 2005). The parameter specifications rely very much on STAR, a comprehensive class of regression models for the expectation of the response that comprises geoadditive models (Kammann and Wand, 2003), generalized additive models (Hastie and Tibshirani, 1986) and generalized additive mixed models (Lin and Zhang, 1999) as special cases.

Our analysis deals with the red shrimp (*Aristeus antennatus*) LPUE. The red shrimp is the target species for the deep-water trawl fishery in the Western Mediterranean (Bas et al., 2003), where catches have reached more than 1000 t/yr (FAO/FISHSTAT, 2011). This fishery is developed in deep-waters - 450 – 900 m - on the continental slope and near submarine canyons (Sardà et al., 1997; Tudela et al., 1998). *A. antennatus* LPUE has already been studied: its fluctuations have been related to changes in oceanographic conditions, e.g. at least partially triggered by changes in the North Atlantic Oscillation (Maynou, 2008) or explained by changes in the seasonal availability of the resource, linked to its life-cycle (Carbonell et al., 1999) and source demand (Sardà et al., 1997).

The main objective of this study is to demonstrate the usefulness of structured additive distributional regression in modelling and predict shrimp LPUE and to provide guidance for model choice and variable selection, questions that arise in the process of developing an appropriate model for a given data set. Therefore, we will discuss tools such as quantile residuals, information criteria and predictive mean squared errors to evaluate the ability of models to describe and predict LPUE adequately.

The rest of the paper is structured as follows: In Section 2 a description of the data set is given to illustrate the application of structured additive distributional regression. Section 3 provides an overview of the methods dealing with the real data set, including the choice of an appropriate response distribution and variable selection. In Section 4, we perform an extensive analysis of the red shrimp data, comparing different response distributions, regression specifications and random versus fixed effects for the repeated observations. The final Section 5 concludes and comments on main results and directions of future research.

2 Data description

Data proceed from the daily sale slips of the Barcelona trawling fleet, granted by the Barcelona Fishers' Association. This data set comprises the information for 21 trawlers, with their total monthly landings (*landings*, kg), their monthly number of trips performed (*trips*) and the Gross Registered Tonnage (*grt*, GRT). Furthermore, the monthly average value of the North Atlantic Oscillation index (NAO) was obtained from the web site of the Climatic Research Unit of the University of East Anglia (Norwich, UK): <http://www.met.rdg.ac.uk/cag/NAO/slpdata.html>. Then we computed the year average of NAO, whose relationship with landings of three years later has been detected through cross-correlation analysis in previous studies by Maynou (2008).

The total number of observations, N , amounts to $N = 2314$ using the whole fleet (21 trawlers) having practised deep-fishing in the period from January 1992 to December 2008 (17 complete years). Landings of the whole fleet were monitored over time, so, data depict the entire population of the fleet in the studied area and period. Red shrimp fishery is a specific fishery, thus, all the product caught on board appears in landings, due to its high commercial value, while, when landings are not reported for a given boat is due to its inactivity, rather than to zero catches of the source.

The landings and number of trips were used to estimate the “nominal” LPUE index,

$$lpue_{ij} = \frac{landings_{ij}}{trips_{ij}}, \quad (1)$$

where i and j refer to the observation i of the vessel j . The adjective “nominal” refers to the variable not standardized. Table 1 and the introductory part of this Section provide information on the variables in Equation 1. Trips are always performed in one day, hence, the *lpue* represents the daily biomass average caught by a boat during one day (kg d^{-1}) with a monthly resolution.

As in previous regression analysis (see [mamouridis2014](#)) months associated to not significant parameters of the categorical variable *months* with categories $month_k$, $k = 1, \dots, 12$ were backward assembled into two categories of a new variable *period* (period of the year): *period1* defines all months excluding June and November and *period2* refers to June and November. All variables are summarised in Table 1.

3 Methodology

It has become quite popular to model the expected LPUE as a function of linear covariate effects. The results obtained from linear mean regression are easy to interpret but depreciated by possible misspecification due to a more complex underlying covariate structure, violation of homoscedastic errors or correlations caused by clustered data. To deal with these problems we apply Bayesian distributional structured additive regression models (Klein et al., 2013b) a model class originally proposed by Rigby and Stasinopoulos (2005) in a frequentist setting. The idea is to assume a parametric distribution for the conditional behaviour of the response and to describe each parameter

Table 1: List of variables.

Variable	Description
<i>landings</i>	the total catches landed at port by each boat in one month
<i>lpue</i>	the daily LPUE index for each boat calculated as in Eq. 1
<i>code</i>	a categorical variable assigned to each boat, $c = 1, \dots, 21$
<i>time</i>	a total of 204 months from January 1992 to December 2008 coded with a letter and two digits, e.g. J92 is January 1992
<i>trips</i>	the number of trips performed by each vessel during one month
<i>grt</i>	Gross Registered Tonnage of each boat
<i>nao₃</i>	mean annual NAO index of 3 years before the year of estimated <i>lpue</i>
<i>month</i>	categorical variable with $m = 1, \dots, 12$ from January to December
<i>period</i>	binary variable with grouped months holding the same effect <i>period1</i> = all month excluding June and November <i>period2</i> = June and November

of this distribution as a function of explanatory variables. In the following, only log-normal and gamma distribution are considered for LPUE, although also an extension to mixture distributions with point masses in zero would be possible as in Heller et al. (2006) or Klein et al. (2013a). Here they are not necessary since the response in this study is always greater than zero.

We consider the log-normal distribution with parameters μ_{ij} and σ_{ij}^2 such that

$$E(lpue_{ij}) = \exp\left(\mu_{ij} + \frac{\sigma_{ij}^2}{2}\right)$$

$$Var(lpue_{ij}) = (\exp(\sigma_{ij}^2) - 1) \exp(2\mu_{ij} + \sigma_{ij}^2).$$

Accordingly, $\log(lpue_{ij})$ is normal distributed with $E(\log(lpue_{ij})) = \mu_{ij}$ and $Var(\log(lpue_{ij})) = \sigma_{ij}^2$. As an alternative, we assume a gamma distribution with parameters $\mu_{ij} > 0$, $\sigma_{ij} > 0$ and density

$$f(lpue_{ij}|\mu_{ij}, \sigma_{ij}) = \left(\frac{\sigma_{ij}}{\mu_{ij}}\right)^{\sigma_{ij}} \frac{lpue_{ij}^{\sigma_{ij}-1}}{\Gamma(\sigma_{ij})} \exp\left(-\frac{\sigma_{ij}}{\mu_{ij}} lpue_{ij}\right).$$

Then, the expectation and variance are given by

$$E(lpue_{ij}) = \mu_{ij}$$

and

$$Var(lpue_{ij}) = \frac{\mu_{ij}^2}{\sigma_{ij}}$$

such that μ_{ij} is the location parameter and the parameter σ_{ij} is inverse proportional to the variance.

All parameters involved are linked to structured additive predictors, yielding

$$\eta_{ij,\mu} = \mu_{ij} \quad \eta_{ij,\sigma^2} = \log(\sigma_{ij}^2)$$

for the log-normal distribution where the log-link is used to ensure positivity of the σ_{ij}^2 . For the gamma distribution, both parameters are restricted to be positive so that we obtain

$$\eta_{ij,\mu} = \log(\mu_{ij}) \quad \eta_{ij,\sigma} = \log(\sigma_{ij}).$$

Dropping the parameter index, a generic structured additive predictor is of the form

$$\eta_{ij} = z'_{ij}\gamma + \sum_{p=1}^P f_p(x_{ijp}) + \alpha_i.$$

Here, z_{ij} is a vector containing binary, categorical or continuous linearly related variables and f_1, \dots, f_P are smooth functions of continuous variables x_{ij1}, \dots, x_{ijP} modelled by Bayesian P(enalized) splines (Lang and Brezger, 2004). The basic assumption is that the unknown functions f_p can be approximated by a linear combination of B-spline basis functions (Eilers and Marx, 1996). Hence, f_p can in matrix notation be written as $Z_p\beta_p$, where Z_p is the design matrix with B-spline basis functions evaluated at the observations and β_p is the vector of regression coefficients to be estimated. To enforce smoothness of the function estimates we use second order random walk priors for the regression coefficients such that

$$p(\beta_p | \tau_p^2) \propto (\tau_p^2)^{-0.5\text{rank}(K)} \exp\left(-\frac{1}{\tau_p^2} \beta_p' K \beta_p\right)$$

where $K = D'D$ for a second order difference matrix D and τ_p^2 are the smoothing variances with inverse gamma hyperpriors.

The additional, boat-specific effect α_i is introduced to represent any effect specific to the catching unit that is not represented in the covariate effects of $z_{ij}, x_{ij1}, \dots, x_{ijP}$. A standard assumption for this effect would be α_i i.i.d. $N(0, \tau^2)$, to acknowledge the fact that the catching units represent a sample from the population of catching units. Alternatively, α_i can be treated as a fixed effect resulting from dummy coding of the different catching units. There has been considerable debate in the past (Bishop et al., 2004; Cooper et al., 2004; Helser et al., 2004; Venables and Dichmont, 2004) about whether it is more appropriate to specify α_i as random or fixed effects from a methodological perspective, but then random effects have been rarely considered (e.g. Marchal et al., 2007). One differentiation goes along the lines discussed above, i.e. differentiating between situations where the catching units in the data set define a (random) subsample of the population of the catching unit (which would favour the specification as random effects) and situations where (almost) the complete fleet has been observed (which would favour the specification as fixed effects). From the Bayesian perspective, this differentiation provides an incomplete picture since the differentiation between random and fixed parameters only corresponds to a difference in prior specifications. From a practical point of view the random effects assumption can also be seen as a possibility to regularise estimation in case of large numbers of catching units and/or

small individual time series where estimation of fixed effects may easily become unstable. Note also that in case of a fixed effects specification, no other time-constant covariates z_i characterising the catching units can be included since they can not be separated from the fixed effects. In our data set, this applies for the gross register tonnage which may be expected to provide important information on LPUE but which can not be included in a fixed effects analysis. This problem can be avoided for example by clustering catching units but with a probable loss of information (as the solution used in Mamouridis et al., 2014). In the next section, we compare the performance of random and fixed effects specifications based on model fit criteria to decide which model has a better explanatory ability.

Our inferences is based on efficient Markov chain Monte Carlo (MCMC) simulation techniques (for more details on distributional regression see (Klein et al., 2013b)). In principle, the approach in all models could also be performed in a frequentist setting (Stasinopoulos and Rigby, 2007) via direct optimization of the resulting penalized likelihood which is often achieved by Newton-type iterations with numerical differentiation. However, many models turned out to be numerically unstable leading to no estimation results or warnings concerning convergence. Therefore the study is restricted to the Bayesian analysis. The Bayesian approach with MCMC also reveals several additional advantages, e.g. simultaneous selection of the smoothing parameters due to the modularity of the algorithm, credibility intervals which are directly obtained as quantiles from the samples and the possibility to extend the model for instances with spatial variations. All models have been estimated in the free open source software BayesX (Belitz et al., 2012).

The performance of models is compared in terms of the Deviance Information Criterion, DIC (Spiegelhalter et al., 2002). The DIC is similar to the frequentist Akaike Information Criterion, compromising between the fit to the data and the complexity of the model. Furthermore it can easily be computed from a sample $\theta^1, \dots, \theta^M$ of the posterior distribution $p(y|\theta)$,

$$\text{DIC} = 2\overline{D(\theta)} - D(\bar{\theta}),$$

with deviance $D(\theta) = -2\log(p(y|\theta))$ and $\overline{D(\theta)} = \frac{1}{M} \sum_{m=1}^M D(\theta^m)$, $\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \theta^m$ respectively. We also use the DIC to determine important variables and optimal predictors $\eta_{i,\mu}$ and η_{i,σ^2} or $\eta_{i,\sigma}$.

To validate the distribution assumption we used normalized quantile residuals. That allowed to decide between equivalent models under different response assumptions. Normalized quantile residuals are defined as $r_i = \Phi^{-1}(u_i)$. Here, Φ^{-1} is the inverse cumulative distribution function of a standard normal distribution and u_i is the cumulative distribution function of the estimated model and with plugged in estimated parameters. For consistent estimates, the residuals r_i , $i = 1, \dots, n$ follow approximately a standard normal distribution if the estimated distribution is the true distribution. Therefore, models can be compared graphically in terms of quantile-quantile-plots.

Finally, to assess the predictive accuracy of the models we performed a k-fold Cross Validation using the mean squared error of prediction (MSEP)

$$\text{MSEP}_k = \frac{\sum_{i=1}^N [lpue_{i,k} - \tilde{E}(lpue_{i,k})]^2}{N}.$$

Here $lpue_{i,k}$ is the observation i of subset k , $\tilde{E}(lpue_{i,k})$ is the expectation of the prediction in the validation set, given parameters estimated on the k -th training set and N refers to the number of observations of the corresponding test set. We performed a 10 fold stratified (within each catching unit) random partition of the whole dataset to ensure a minimum number of observations for each boat in both the training and the validation sets. Taking the 10% of the data to build the latter, we ensure at least 10 observations per unit in the validation set. If at least one catching unit is not represented in one of the partitions, the prediction for the missing catching unit in fixed effects models could not be computed. This clarifies the usefulness in using mixed effects specification when interested on predictions for unobserved catching units.

4 Data analysis

4.1 Model diagnostics and comparison

During model building variables were selected using a stepwise forward procedure according to the DIC scores and the significance of their effect. Single models have been built first for each variable, to assess its explanatory potential. For each distribution the predictor for location has been modelled, adding one variable at a time till finding the best predictor. Then, using this “best” predictor for location, also the predictor for the second parameter, the scale or shape for log-normal or gamma respectively, has been modelled using the same procedure.

For both distributions, all models with single explanatory variable returned significant effects, except some categories of *code* and *month*. So that, we decide to model *month* effect as binary (*period*), after grouping categories (see Methodology Section and Mamouridis et al. (2014)). Conversely, *code* variable has not been merged, to allow the comparison between fixed and mixed effects models.

Assuming log-normal distribution and according to the ascending DIC, variables are ordered from *code* (DIC=17059.4 as random and DIC=17060.0 as fixed effect), then *trips*, *time*, *grt*, *nao3*, to *period* (DIC=17735.7). The same ordination has been found assuming the gamma distribution with DIC scores ranged between DIC=16763.0 for *code* and DIC=17290.0 for *period*. Variables have been added in this order till the saturated model. Variables *nao3* and *grt* do not strongly improve model in terms of DIC scores (less than 20 units for each variable). However parameters are significantly different from zero and their incorporation effective to be discussed. We used the same procedure for the second predictor. According to the DIC, variables are ordered as follows: *code*, *trips*, *time*, *grt*, *nao3* and *period*. The effects of variables *nao3* *period* and *grt* for the second parameter were not significant.

According to the DIC scores, for log-normal assumption, the appropriate predictor structures for location η_μ and scale η_{σ^2} , are

$$\begin{aligned}\eta_\mu &= \beta_{0,\mu} + \beta_{1,\mu} \text{period}_2 + f_{1,\mu}(\text{trips}) + f_{2,\mu}(\text{time}) + f_{3,\mu}(\text{nao3}) + \sum_i \alpha_{i,\mu} \\ \eta_{\sigma^2} &= \beta_{0,\sigma^2} + f_{1,\sigma^2}(\text{trips}) + f_{2,\sigma^2}(\text{time}) + \sum_i \alpha_{i,\sigma^2}\end{aligned}\tag{2}$$

for both fixed and mixed effects specification. Here the $\beta_{k,\cdot}$ are parameters associated to the intercept and linear fixed effects of the variable *period*. The $\alpha_{i,\cdot}$ are parameters associated to the effects of *code*, specified as fixed in fixed effects models and as random in mixed effects models. Instead, $f_{l,\cdot}$ are smooth functions associated to nonlinear effects of the variables *trips*, *time* and *nao3*. The second sub-index in all parameters, μ or σ^2 , identifies which predictor the parameter or function belongs, η_μ or η_{σ^2} respectively.

Using fixed effects specification for *code* the inclusion of variable *grt* leads to instability, due to the reasons mentioned in Section 3, so then, the model with *grt* cannot be estimated in these cases. On the contrary, in mixed effects specification *grt* could be estimated but it leads to equal or lightly higher values of DIC score. Thus it has been backward eliminated, however the associated parameter was significantly different from zero and positive.

Under the gamma distribution assumption, the log-link function has been chosen, since the support of both parameters is the positive real domain and the final predictor structures for location and shape are

$$\begin{aligned}\eta_\mu &= \beta_{0,\mu} + \beta_{1,\mu} \text{period}_2 + f_{1,\mu}(\text{trips}) + f_{2,\mu}(\text{time}) + f_{3,\mu}(\text{nao}_3) + \sum_i \alpha_{i,\mu} \\ \eta_\sigma &= \beta_{0,\sigma} + f_{1,\mu}(\text{trips}) + f_{2,\sigma}(\text{time}) + \sum_i \alpha_{i,\sigma}\end{aligned}\quad (3)$$

for both fixed and mixed effects specification. The notation here is the same specified for the log-normal models however here the second parameter, the shape, is denoted σ .

Table 2: Global scores of selected models. Columns indicate: M, refers to model coding (see specifications in the text); DEV, the residual deviance; EP: Effective total number of Parameters, DIC: Deviance Information Criterion, MSEP, mean and sd of the mean square error of predictions calculated through 10-fold validation.

M	DEV	EP	DIC	MSEP
M1	16163.9	47.1	16258.0	98.3 ± 12.3
M2	16164.2	46.8	16257.7	96.9 ± 12.4
M3	15138.4	91.2	15320.8	436.2 ± 434.9
M4	15142.7	87.2	15317.2	315.0 ± 307.4
M5	16095.7	45.6	16187.0	77.8 ± 11.5
M6	16096.8	43.9	16184.7	77.3 ± 11.7
M7	15027.1	90.9	15208.9	71.5 ± 9.2
M8	15032.4	86.9	15206.1	71.7 ± 9.4

The Table 2 provides a selection of models that we used for comparison purposes and their corresponding global parameters: the deviance, the effective number of parameters and the DIC, estimated on the whole dataset, and the MSEP calculated by predictions on the validation subsets as described in the methodology.

The eight models in table 2 present a combination between alternatives of the following assumptions:

- (A) The log-normal (LN) or gamma (GA) as the underlying distribution assumption;
- (B) Only location (LO) or both location and scale/shape (LS) parameters explicitly modelled using one or more explanatory variables;
- (C) Effects of *code* as fixed or random leading to a fixed effects model (FI) or mixed effects model (MI) respectively;

So, model M1 is specified by A=LN, B=LO and C=FI; M2 by A=LN, B=LO and C=MI; M3 by A=LN, B=LS and C=FI; M4 by A=LN, B=LS and C=MI; M5 by A=GA, B=LO and C=FI; M6 by A=GA, B=LO and C=MI; M7 by A=GA, B=LS and C=FI; and M8 by A=GA, B=LS and C=MI. The predictors for location in models denoted as M1, M2, M5 and M6 (with B=LO) have the same structure of η_{mu} in Equations 2 and 3, while the corresponding predictor for scale/shape is simply the constant. Both predictors in models M3, M4, M7 and M8 correspond to Equations 2 and 3.

Concerning to (A), model specifications widely favour the gamma over the log-normal distribution. In fact DIC scores are lower under the gamma assumption, with approximately 100 scores of difference between analogous models (i.e. same variables specified in the predictors). The benefit in assuming the gamma distribution is also evident comparing MSEP scores (lower scores for better predictions). Nevertheless, results of log-normal models' MSEP is not entirely satisfactory, because when accounting for both predictors MSEP should behave as in the gamma models, i.e. lower scores than accounting only for the predictor η_{μ} . Regarding to the estimation of both predictors for first and second parameters (B), models under the gamma assumption show an improvement when explicitly modelling the dependence of second parameter σ from explanatory variables in both DIC and MSEP scores (both decrease). Contrariwise, under the log-normal assumption, however the DIC decreases, the MSEP increases and presents higher variance, suggesting worse predictions, when the second parameter is explicitly modelled. But as discussed few lines above, that is not realistic and follows a strange behaviour of η_{σ^2} , thus, we still working on this point. Finally no notable leaps have been observed between fixed and random effects models (C). Whereby the best DIC and MSEP scores and QQplots lead to the final model with predictors given in (3).

According to the DIC (Table 2) the model that better fits the data is the gamma model M8 whose predictors are given in Equation 3 specifying catching units as random effects.

The boxplots of MSEPs also favour the gamma assumption (see Figure 1 and values in Table 2). The minor MSEP better the prediction, MSEP results divide models into three distinguishable groups from highest to lowest mean MSEP: 1) log-normal models considering heteroscedasticity, 2) log-normal models considering constant variance (compare M1-M2 with M3-M4), and 3) all gamma models (compare M1-M4 with M5-M8). Within this group, modelling the shape in dependence to some variables, consistently decreases MSEP estimates, in terms of average and variance (compare M5-M6 with M7-M8).

In order to validate the distribution assumptions, QQplots for residuals are reported in Figure 2, from which it follows that:

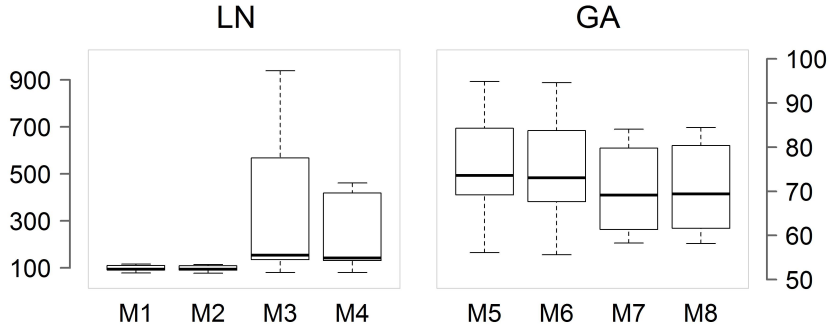


Figure 1: Boxplots for MSEP calculated for all models. See Table 2 and Equations in the text for model specifications.

- the residuals in gamma models almost follow the straight line (M5-M8 in the Figure), while in the log-normal they show upward-humped curves (M1-M4), suggesting a definitively “better approximation” of models to the gamma distribution.
- Modelling the second parameter in gamma models improves QQplot outputs, while the opposite happens for log-normal models. Focusing only in gamma models, the outlier is “absorbed” into the straight line in the right part validating the improvement in estimations (compare M5-M6 and M7-M8 in Figure 2).

We finally assess the normality for random effects for the model M8 corresponding to Equation 3. Figure 4.1 provides the QQplots of the random effects for both predictors. The majority of sample quantiles approximatively follow the normal quantiles, however they depart from it at the extremes, especially evident in the lower tails and for α_μ (on the left).

4.2 Description of partial effects

Estimations of linear fixed effects for A) μ and B) σ predictors of final model (3), to which we referred in the text as M8, are reported in table 3.

Table 3: Estimations of linear fixed effects for the final model, Eq. (3) associated to A) μ and B) σ respectively.

	A)		B)	
	mean	sd	mean	sd
<i>const</i>	2.891	0.058	-1.551	0.099
<i>period₂</i>	-0.153	0.024		

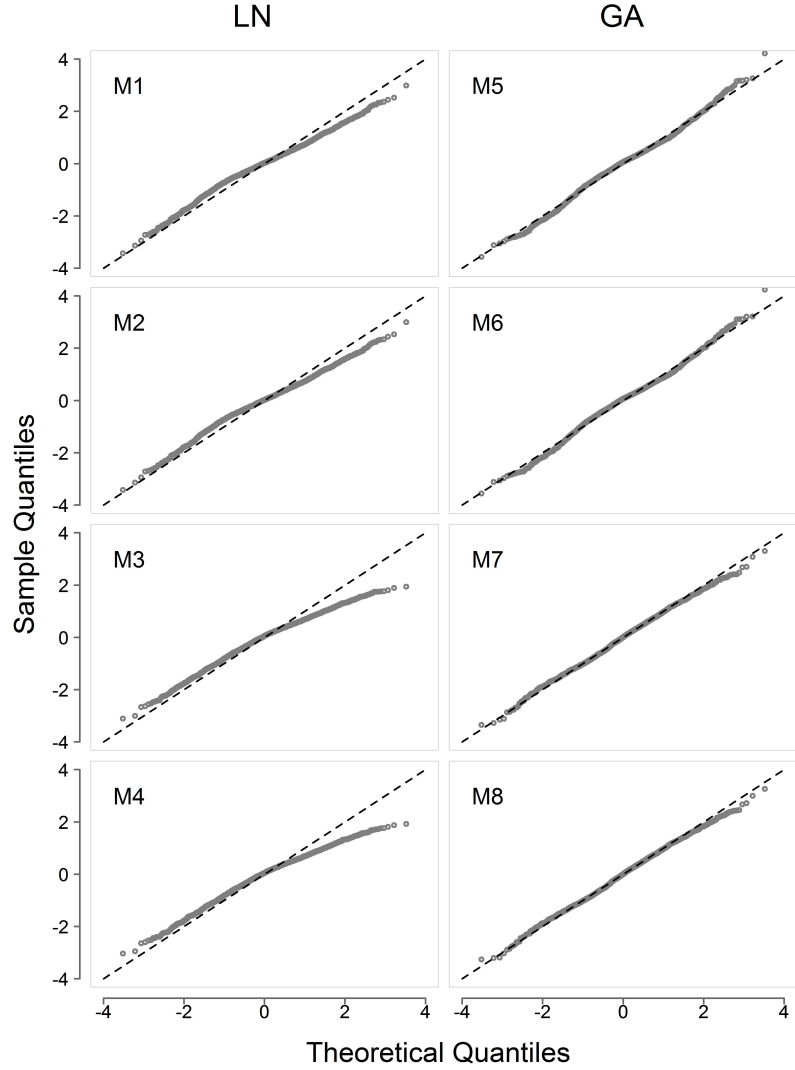


Figure 2: QQplots of residuals for selected log-normal and gamma models.

The categorical variable *period* describes the intra-annual variability and shows a negative effect during *period*₂, corresponding to June and November in comparison to the rest of the year. That should be related to a lower demand of this source during these months, as suggested by Sardà et al. (1997).

The variable *grt*, referring to the gross registered tonnage of boats, (not incorporated into the model because did not improve the DIC score) bears a significantly positive slope parameter (0.009 ± 0.003 when included), causes a linear increment on the predictor for location. We consider important to quantify its effect in order to compare

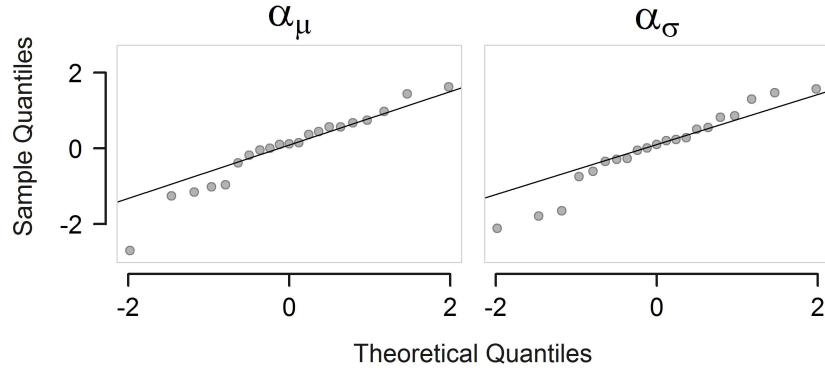


Figure 3: QQplots for normality of catching units as random effects in the mixed model M8. α_μ refer to random effects in the predictor for location, while α_σ refers to the predictor for the shape.

it with other fisheries, where, contrariwise, it could have a major effect. *grt* is not the only variable characterising a fleet, nevertheless is the only reliable for this fishery.

The second variable we can consider as fishery-related variable is represented by the catching units. Variable *code* captures all abilities of fishermen and technical characteristics of the fleet, appropriate technologies and strategies, e.g. the power and type of the engine, the net shape and the skipper's expertise and ability. Results (Figure ??) show that many trawlers have similar effects, while few of them hold or positive either negative effects. The former are very specialised and powerful boats that capture large amounts of the resource so then that leads to higher values of LPUE while the latter are not specialised trawlers that accordingly capture lower amounts, leading to lower LPUE (see the partial effects on μ). At the same time catching units associated to higher effects on the predictor of μ also present higher effects on the predictor for σ , while boats associated to lower effects on μ also hold lower effects on σ . In other words, more specialised catching units are able to capture more quantities of the resource, and they also present less variability. Contrariwise landings of not specialised trawlers present more variability. This fraction of the fleet more likely is represented by boats that fish usually on the continental shelf and occasionally displace towards deeper waters going in search of the red shrimp, representing one of the most lucrative resources for the NW Mediterranean fisheries. It is likely to think that these boats have less knowledge of red shrimp fishing grounds (Maynou et al., 2003) and catch less.

Concerning to nonparametric effects (Figure 4.2), *trips* and *time* influence both η_μ and η_σ , while *nao3* slightly affects only η_μ .

trips shows a negative effect on the predictor for μ when $trips \leq 8$, while positive otherwise. The rate of the effect decreases moving through the covariate interval till rising a plateau beginning around $trips = 17$. For extreme high values this covariate has

an uncertain effect. It is plausible that increasing the number of trips per month, more likely increase the ability to find high-concentration shoals inside the fishing grounds in a process of trial and error (as suggested by Sardà and Maynou 1998).

The effect of *time* on the predictor for μ is the most difficult to interpret, showing high inter-annual variability, certainly caused from unobservable multiple factors. Between 1992-1996, the function decreases while increases in next three years. We could not find a reasonable explanation to this trend. Afterwards, between 1999-2000, it drops till a minimum low followed by a rapid increase up to a pick in 2004. We believe that the minimum is related to both negative NAO observed in previous years and to the rising of fuel price started in 2000, that in turn is related to a lower number of trips performed by trawlers (see comments below and the discussion in Mamouridis et al., 2014). Then, for five years it presents a slightly oscillatory trend till the last year characterized by another positive pick probably related to the rise of the economic value of the resource, that offsets the increase in the fuel price.

Finally, the nao_3 has a moderate effect?? for this deep-sea species, being notoriously evident only when reaches anomalous values. Numerous studies on this and other fish stocks (e.g. Maynou, 2008; Báez et al., 2011) demonstrated that the NAO can have important effects when it reaches extreme values??, whether they are positive or negative. Our results show that nao_3 has a moderate effect?? for this deep-sea species, however they suggest that it can lead to the reduction of its biomass when reaches very negative values. Combining results of *time* and nao_3 effects and comparing data series

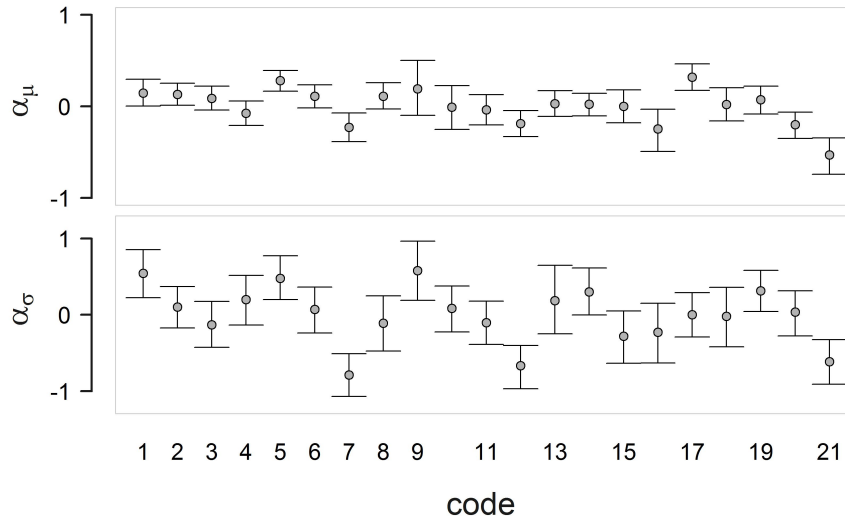


Figure 4: Interval plots of estimated random effects in the predictor for location (α_μ) and shape (α_σ) of model M8 on the upper and lower plots respectively. Bars indicates 95% CI.

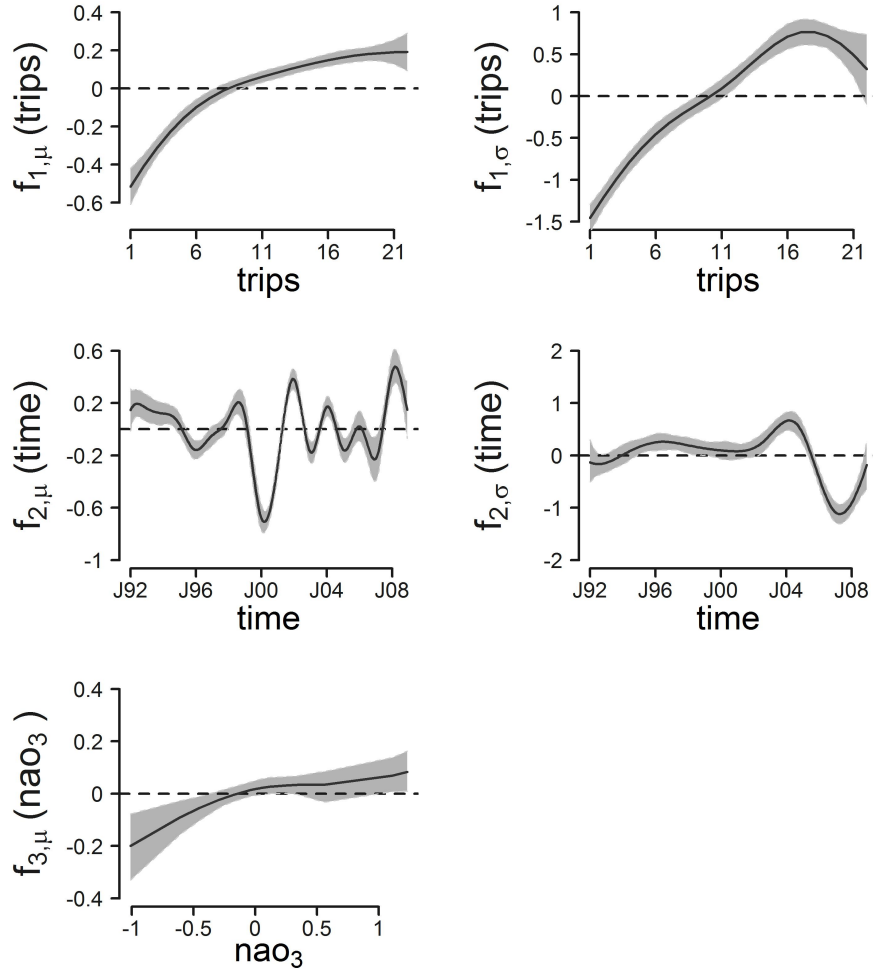


Figure 5: Nonparametric effects for the GA mixed effect of the selected gamma model (M8). Effects on predictor for μ (left side) and for σ (right side). Grey shapes represent 95% credible intervals.

of both the raw (or nominal) LPUE and the NAO, we believe that the LPUE low starting at the end of 1999 can be related to consecutive negative NAO during the previous four years, especially during 1996, corresponding to three years before the beginning of the decline of LPUE ??see[[mamouridis2014analysis. NAO leads to paucity of resources when it is low, while enhances productivity when it is high. Between $nao_3 = 0.50$ and $nao_3 = 1$ the effect increases however without significant evidences. In the middle region of the observed variable span, NAO shows any effect, that could represents a “buffering region” related to “normal” weather conditions for the stock.

All partial effects are linked to the expectation, $E(LPUE)$, through the exponential of η_μ , such that, when partial effects of *trips*, *time* and *nao3* hire positive values, $E(LPUE)$ is positive while negative otherwise. Regarding to the variance of LPUE, it is affected by both predictors being directly proportional to μ and inverse proportional to σ .

The effect of *trips* is negative for low values of the covariate ($trips \leq 10$) and positive for higher values. It also shows a high increment till a maximum corresponding to $trips = 17$, while decreasing again for higher values, although always positive. The effect of *time* is slightly negative before 1995 and slightly positive between 1995 and 2000. Then for two years has no effect and, in 2002, it switches clearly positive again till 2006, showing a high pick in 2004 and finally negative in the last years during 2006-2008, reaching an abrupt drop in late 2007.

Thus, regarding to the $Var(LPUE)$, results show that values of covariates associated to positive effects in η_μ and negative effects in η_σ , in turn affect positively (increase) to the variance. We can also deduce that *time* and *trips* are drivers in causing the heteroscedasticity in LPUE, *times* mainly in last years when its effect on η_μ is positive and its effect on η_σ is strongly negative. This high variability could be related to different factors, probably of economic origin, such as the fuel and ex-vessel shrimp prises.

5 Conclusions

In this study distributional structured additive models have been proposed for the first time to model the LPUE, index widely used in fisheries research. Data deal with the LPUE of red shrimp (*A. antennatus*) from the Barcelona's fleet during years 1992 - 2008.

Our aims were: 1) find the best distribution in relation to the response variable, comparing gamma and log-normal distributions, 2) improve estimations modelling both predictors for first and second parameter 3) compare parameter specification for catching units, fixed versus mixed effects models, and finally 4) achieve new insights in the understanding of the effect that the explanatory variables considered have on the LPUE of red shrimp.

On a methodological viewpoint, distributional structured additive models, DSTAR, as the frequentist counterpart GAMLSS, permit the estimation of both first and second order moments of the LPUE, allowing more accurate estimations and the analysis of both the expectation and the variance of the response. Results indicate that explicitly modelling the second moment in dependence to appropriate explanatory variables can lead to better estimations and predictions. We also rose a more detailed understanding of the LPUE, that bears some amount of heteroscedasticity in the time span studied. This heteroscedasticity had been observed but could not be described in previous analysis (Mamouridis et al., 2014). In fact the analysis performed on almost the same data set, a frequentist approach using GAM accounting only for the location, could not avoid the heteroscedasticity in the residuals (see Figure 4 within). Here, the modelling of the shape, in the case of the gamma assumption, permitted to infer about the variance of he

response. Here we demonstrated that both the number of trips and the time influence the second parameter leading to changes in the variance.

Concerning to fixed versus mixed specification, in our study the fixed effects can be considered appropriate representing sampling units the whole population of the studied fleet, however mixed models permit more flexibility, such as the estimation of GRT effect. Consider that mixed models also allow to get prediction for unobserved catching units and in turn to generalise predictions outside the observed fishing population.

Acknowledgements

Authors would like to thank the Fisheries Directorate of Catalonia's Autonomous Government to facilitate access to the sales data of the Barcelona Fishers' Association, as well as fishers of Barcelona. This study was financed by the Spanish National Council of Research (CSIC) through the JAE-predoc grant program. Financial support from the German Research Foundation (DFG) grant KN 922/4-1 is gratefully acknowledged.

References

- J.C. Báez, J.M. Ortiz de Urbina, R. Real, and D. Macías. Cumulative effect of the north atlantic oscillation on age-class abundance of albacore (*thunnus alalunga*). *Journal of Applied Ichthyology*, 27(6):1356–1359, 2011.
- S. P. Bannerot and C. B. Austin. Using frequency distributions of catch per unit effort to measure fish-stock abundance. *Transactions of the American Fisheries Society*, 112:608–617, 1983.
- C. Bas, F. Maynou, F. Sardà, and J. Lleonart. *Variacions demogràfiques a les poblacions d'espècies demersals explotades: els darrers quaranta anys a Blanes i Barcelona*. Inst. Est. Catalans. Arxiu de la Sec. Ciències, Barcelona, 2003.
- C. Belitz, A. Brezger, T. Kneib, S. Lang, and N. Umlauf. Bayesx, 2012. - software for bayesian inference in structured additive regression models. version 2.1. available from <http://www.bayesx.org>. 2012.
- J. Bishop, W.N. Venables, and Y.-G. Wang. Analysing commercial catch and effort data from a penaeid trawl fishery: A comparison of linear models, mixed models, and generalised estimating equations approaches. *Fisheries Research*, 70(23):179 – 193, 2004. ISSN 0165-7836. doi: <http://dx.doi.org/10.1016/j.fishres.2004.08.003>. URL <http://www.sciencedirect.com/science/article/pii/S0165783604001651>. <ce:title>Models in Fisheries Research: GLMs, {GAMS} and GLMMs</ce:title>.
- A. Brezger and S. Lang. Generalized structured additive regression based on bayesian p-splines. *Computational Statistics & Data Analysis*, 50(4):967 – 991, 2006.
- J. Brynjarsdóttir and G. Stefánsson. Analysis of cod catch data from icelandic ground-fish surveys using generalized linear models. *Fisheries Research*, 70(23):195 – 208, 2004. ISSN 0165-7836. doi: <http://dx.doi.org/10.1016/j.fishres.2004.08.004>. URL <http://www.sciencedirect.com/science/article/pii/S0165783604001663>.

- A. Carbonell, M. Carbonell, M. Demestre, A. Grau, and S. Monserrat. The red shrimp *Aristeus antennatus* (risso, 1816) fishery and biology in the balearic islands, western mediterranean. *Fisheries Research*, 44(1):1–13, 1999.
- A. B. Cooper, A. Rosenberg, G. Stefansson, and M. Mangel. Examining the importance of consistency in multi-vessel trawl survey design based on the U.S. west coast groundfish bottom trawl survey. *Fisheries Research*, 70(2-3):239–250, December 2004. ISSN 01657836. doi: 10.1016/j.fishres.2004.08.006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0165783604001699>.
- D. Damalas, P. Megalofonou, and M. Apostolopoulou. Environmental, spatial, temporal and operational effects on swordfish (*Xiphias gladius*) catch rates of eastern Mediterranean Sea longline fisheries. *Fisheries Research*, 84(2):233–246, April 2007. ISSN 01657836. doi: 10.1016/j.fishres.2006.11.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S0165783606003845>.
- V. Denis. Spatio-temporal analysis of commercial trawler data using General Additive models: patterns of Loliginid squid abundance in the north-east Atlantic. *ICES Journal of Marine Science*, 59(3):633–648, June 2002. ISSN 10543139. doi: 10.1006/jmsc.2001.1178. URL <http://icesjms.oxfordjournals.org/cgi/doi/10.1006/jmsc.2001.1178>.
- P. H. C. Eilers and B. D. Marx. Flexible smoothing with b -splines and penalties. *Statistical Science*, 11(2):89–102, 1996.
- L. Fahrmeir, T. Kneib, and S. Lang. Penalized structured additive regression for space-time data: a bayesian perspective. *Statistica Sinica*, 14(3):731–762, 2004.
- FAO/FISHSTAT. Fao fisheries department, fishery information, data and statistics unit. fishstatj, a tool for fishery statistical analysis, release 2.0.0, 2011.
- S. Gavaris. Use of a multiplicative model to estimate catch rate and effort from commercial data. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(12):2272–2275, 1980. doi: 10.1139/f80-273. URL <http://www.nrcresearchpress.com/doi/abs/10.1139/f80-273>.
- R. Goñi, F. Alvarez, and S. Adlerstein. Application of generalized linear modeling to catch rate analysis of western mediterranean fisheries: the castellón trawl fleet as a case study. *Fisheries Research*, 42(3):291 – 302, 1999. ISSN 0165-7836. doi: [http://dx.doi.org/10.1016/S0165-7836\(99\)00039-9](http://dx.doi.org/10.1016/S0165-7836(99)00039-9). URL <http://www.sciencedirect.com/science/article/pii/S0165783699000399>.
- J. A. Gulland. Catch per unit effort as a measure of abundance. *Rapp. P.-V. Reun., Comm. Int. Explor. Mer Mediter*, 155:8 – 14, 1964.
- T. Hastie and R. Tibshirani. Generalized additive models. *Statistical science*, pages 297–310, 1986.

- G. Heller, Stasinopoulos D. M., and Rigby R. A. The zero-adjusted inverse Gaussian distribution as a model for insurance data. In J. Newell J. Hinde, J. Einbeck, editor, *Proceedings of the 21th International Workshop on Statistical Modelling*, 2006.
- T. E. Helser, A. E. Punt, and R. D. Methot. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. *Fisheries Research*, 70(23):251 – 264, 2004. ISSN 0165-7836. doi: <http://dx.doi.org/10.1016/j.fishres.2004.08.007>. URL <http://www.sciencedirect.com/science/article/pii/S0165783604001705>.
- E. E. Kammann and M. P. Wand. Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):1–18, 2003.
- N. Klein, M. Denuit, T. Kneib, and S. Lang. Nonlife ratemaking and risk management with bayesian additive model for location scale and shape. Technical report, 2013a. URL <http://eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2013-24.pdf>.
- N. Klein, T. Kneib, and S. Lang. Bayesian structured additive distributional regression. Technical report, 2013b. URL <http://eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2013-23.pdf>.
- S. Lang and A. Brezger. Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212, 2004.
- X. Lin and D. Zhang. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):381–400, 1999. ISSN 1467-9868. doi: 10.1111/1467-9868.00183. URL <http://dx.doi.org/10.1111/1467-9868.00183>.
- V. Mamouridis, F. Maynou, and G. Aneiros Pérez. Analysis and standardization of landings per unit effort of red shrimp *Aristeus antennatus* from the trawl fleet of barcelona (nw mediterranean). *Scientia Marina*, 78:INSERT, 2014. doi: 10.3989/scimar.03926.14A.
- P. Marchal, C. Ulrich, K. Korsbrekke, M. Pastoors, and B. Rackham. A comparison of three indices of fishing power on some demersal fisheries of the north sea. *ICES Journal of Marine Science: Journal du Conseil*, 59(3):604–623, 2002. doi: 10.1006/jmsc.2002.1215. URL <http://icesjms.oxfordjournals.org/content/59/3/604.abstract>.
- P. Marchal, Bo A., B. Caillart, O. Eigaard, O. Guyader, H. Hovgaard, A. Iriondo, F. Le Fur, J. Sacchi, and M. Santurtún. Impact of technological creep on fishing effort and fishing mortality, for a selection of european fleets. *ICES Journal of Marine Science: Journal du Conseil*, 64(1):192–209, 2007.
- M. N. Maunder, J. R. Sibert, A. Fonteneau, J. Hampton, P. Kleiber, and S. J. Harley. Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science: Journal du Conseil*, 63(8):1373–1385, 2006. doi: 10.1016/j.icesjms.2006.05.008. URL <http://icesjms.oxfordjournals.org/content/63/8/1373.abstract>.

- F. Maynou. Environmental causes of the fluctuations of red shrimp (*Aristeus antennatus*) landings in the Catalan Sea. *Journal of Marine Systems*, 71(3-4):294–302, June 2008. ISSN 09247963. doi: 10.1016/j.jmarsys.2006.09.008. URL <http://linkinghub.elsevier.com/retrieve/pii/S0924796307001935>.
- F. Maynou, M. Demestre, and P. Sanchez. Analysis of catch per unit effort by multivariate analysis and generalised linear models for deep-water crustacean fisheries off barcelona (nw mediterranean). *Fisheries Research*, 65(1-3):257–269, December 2003. ISSN 01657836. doi: 10.1016/j.fishres.2003.09.018. URL <http://linkinghub.elsevier.com/retrieve/pii/S0165783603002479>.
- P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman & Hall, London, 1989.
- R. Mendelssohn and P. Cury. Temporal and spatial dynamics of a coastal pelagic species, sardinella maderensis off the ivory coast. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(10):1686–1697, 1989. doi: 10.1139/f89-214. URL <http://www.nrcresearchpress.com/doi/abs/10.1139/f89-214>.
- R. A. Myers and P. Pepin. The robustness of lognormal-based estimators of abundance. *Biometrics*, pages 1185–1192, 1990.
- M. Pennington. Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, 39(1):281–286, 1983.
- J. C. Pinheiro and D. M. Bates. *Mixed Effects Models in S and S-PLUS*. 2000.
- R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554, 2005. ISSN 1467-9876. doi: 10.1111/j.1467-9876.2005.00510.x. URL <http://dx.doi.org/10.1111/j.1467-9876.2005.00510.x>.
- D. Ruppert, M. P. Wand, and Ra. J. Carroll. *Semiparametric regression Vol. 12*. Cambridge University Press, 2003.
- F. Sardà and F. Maynou. Assessing perceptions: do catalan fishermen catch more shrimp on fridays? *Fisheries Research*, 36:149157, 1998.
- F. Sardà, F. Maynou, and L. Talló. Seasonal and spatial mobility patterns of rose shrimps *Aristeus antennatus* in the western mediterranean: results of a long-term study. *Marine Ecology Progress Series*, 159:133141, 1997.
- D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4):583–639, 2002.
- D. M. Stasinopoulos and R. A. Rigby. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46, 2007.

- G. Stefánsson. Analysis of groundfish survey abundance data: combining the glm and delta approaches. *ICES Journal of Marine Science: Journal du Conseil*, 53(3):577–588, 1996. doi: 10.1006/jmsc.1996.0079. URL <http://icesjms.oxfordjournals.org/content/53/3/577.abstract>.
- G. Stefánsson and O. K. Pálsson. Points of view: A framework for multispecies modelling of arcto-boreal systems. *Reviews in Fish Biology and Fisheries*, 8:101–104, 1998.
- S. Tudela, F. Maynou, and M. Demestre. Influence of submarine canyons on the distribution of the deep-water shrimp, *aristeus antennatus* (risso, 1816) in the nw mediterranean. *Crustaceana*, 7(2):217225, 1998.
- W. N. Venables and C. M. Dichmont. Glms, gams and glmms: an overview of theory for applications in fisheries research. *Fisheries research*, 70(2):319–337, 2004.
- S. Wood. *Generalized Additive Models: an introduction with R*. CRC Press, 2006.

Reports in Statistics and Operations Research

2005

- 05-01 SiZer Map for Evaluating a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models. M. D. Martínez-Miranda, R. Raya-Miranda, W. González-Manteiga and A. González-Carmona.
- 05-02 The Role of Commitment in Repeated Games. I. García Jurado, Julio González Díaz.
- 05-03 Project Games. A. Estévez Fernández, P. Borm, H. Hamers.
- 05-04 Semiparametric Inference in Generalized Mixed Effects Models. M. J. Lombardía, S. Sperlich.

2006

- 06-01 A unifying model for contests: effort-prize games. J. González Díaz.
- 06-02 The Harsanyi paradox and the "right to talk" in bargaining among coalitions. J. J. Vidal Puga.
- 06-03 A functional analysis of NO_x levels: location and scale estimation and outlier detection. M. Febrero, P. Galeano, W. González-Manteiga.
- 06-04 Comparing spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.
- 06-05 On the spectral simulation of spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal.
- 06-06 An L₂-test for comparing spatial spectral densities. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.

2007

- 07-01 Goodness-of-fit tests for the spatial spectral density. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.
- 07-02 Presmoothed estimation with left truncated and right censored data. M. A. Jácome, M. C. Iglesias-Pérez.
- 07-03 Robust nonparametric estimation with missing data. G. Boente, W. González-Manteiga, A. Pérez-González.
- 07-04 k-Sample test based on the common area of kernel density estimators. P. Martínez-Camblor, J. de Uña Álvarez, N. Corral-Blanco.

07-05 A bootstrap based model checking for selection-biased data. J. L. Ojeda, W. González-Manteiga, J. A. Cristobal.

07-06 The Gaussian mixture dynamic conditional correlation model: Bayesian estimation, value at risk calculation and portfolio selection. P. Galeano, M. C. Ausín.

2008

08-01 ROC curves in nonparametric location-scale regression models. W. González-Manteiga, J. C. Pardo Fernández, I. Van Keilegom.

08-02 On the estimation of α -convex sets. B. Pateiro-López, A. Rodríguez-Casal.

2009

09-01 Lasso Logistic Regression, GSoft and the Cyclyc Coordinate Descent Algorithm. Application to Gene Expression Data. M. García-Magariños, A. Antoniadis, R. Cao, W. González-Manteiga.

2010

10-01 Asymptotic behaviour of robust estimators in partially linear models with missing responses: The effect of estimating the missing probability on simplified marginal estimators. A. Bianco, G. Boente, W. González-Manteiga, A. Pérez-González.

10-02 First-Price Winner-Takes-All Contents. J. González-Díaz.

10-03 Goodness of Fit Test for Interest Rate Models: an approach based on Empirical Process. A. E. Monsalve-Cobis, W. González-Manteiga, M. Febrero-Bande.

2011

11-01 Exploring wind direction and SO₂ concentration by circular-linear density estimation. E. García-Portugués, R.M. Crujeiras, W. González-Manteiga.

11-02 Utilities for Statistical Computing in Functional Data Analysis: The R Package fda.usc. M. Oviedo de la Fuente, M. Febrero-Bande.

11-03 Multivariate uniformity tests: the distance to boundary method for the case of unknown support. J. R. Berrendero, A. Cuevas, B. Pateiro-López.

11-04 On the meaning, properties and computing of a coalitional Shapley value. J. M. Alonso-Mejide, B. V. Casas-Méndez, A. M. González-Rueda, S. M. Lorenzo-Freire.

11-05 Share functions for cooperative games with levels structure of cooperation. M. Álvarez-Mozos, R. van den Brink, G. van der Laan, O. Tejada.

2012

- 12-01 A goodness of fit test for the functional linear model with scalar response. E. García-Portugués, W. González-Manteiga, M. Febrero-Bande.
- 12-02 Bootstrap independence test for functional linear models. W. González-Manteiga, G. González-Rodríguez, A. Martínez-Calvo, E. García-Portugués
- 12-03 Allocation rules for games with optimistic aspirations. L. Carpenté, B. Casas-Méndez, I. García-Jurado, A. van den Nouweland
- 12-04 Kernel density estimation for directional-linear data. E. García-Portugués, R.M. Crujeiras, W. González-Manteiga

2013

- 13-01 Extended Additive Regression for Analysing LPUE Indices in Fishery Research. V. Mamouridis, N. Klein, T. Kneib, C- Cadarso Suárez, F. Maynou.

Previous issues (2001 – 2003):

<http://eio.usc.es/reports.php>