

Additive Mixed Models applied to the study of
red shrimp landings: comparison between
frequentist and Bayesian perspectives.

Valeria Mamouridis

January 14, 2011

Abstract

Relationships between Red shrimp landings in the Catalanian port of Barcelona and some explanatory variables are studied by means of regression analysis. Generalized additive mixed models (GAMMs) are proposed here as flexible alternatives to the parametric modeling approaches actually applied in fishery research (mainly GLMs). Two different approaches are considered and compared: (a) the frequentist approach, using restricted likelihood (REML), and implemented in the R-package `mgcv`, and (b) Bayesian approach based on two different methods (i) the Markov Chain Monte Carlo simulation (full Bayesian) and (ii) the empirical Bayesian REML, both implemented in the BayesX software. The main purpose of the study is to compare the three methods (frequentist REML, empirical Bayesian REML and full Bayesian) for fitting variance components, fixed and mixed effects in selected models.

Contents

1	Introduction	11
1.1	Fixed effects vs Random effects	11
1.2	The historical controversy between frequentists and Bayesians	16
1.3	Background of the application study	17
1.4	Outline of the study	19
2	Generalized additive mixed model formulation and inference	21
2.1	GAMM formulation	21
2.1.1	From LM to GLMM	21
2.1.2	From AM to GAMM	24
2.1.3	The response distribution	25
2.2	Spline representation	26
2.2.1	Polynomial spline	28
2.2.2	B-spline	30
2.2.3	P-spline	31
2.3	Inference	33
2.3.1	Frequentist perspective	33
2.3.2	Full bayesian perspective	38
2.3.3	The empirical Bayesian perspective	41
3	The application study	43
3.1	Data source	43
3.2	Data description and exploration	43
3.3	Materials and methods	52
3.3.1	Model construction	52
3.3.2	Diagnostic tools for the analysis of residuals	54
3.3.3	Selection criterion	56
3.3.4	Comparison between estimations	57
3.4	Results and discussion	58
3.4.1	Diagnostics of model assumptions	58
3.4.2	Deviance and F test	60

3.4.3	Frequentist, empirical Bayesian and full Bayesian estimations	68
3.5	Interpretation of results	70
4	Software	77
4.1	The R package mgcv	77
4.1.1	R syntax	77
4.1.2	Syntax implemented for the example in Section 1.1	82
4.2	BayesX	83
4.2.1	BayesX syntax	83
5	Conclusions	87

List of Figures

1.1	Boxplot of the CPUE expressed in kg/trip for 21 vessels. . . .	12
1.2	boxplot of residuals vs vessel code. From the top to the down: residuals of the single mean model, of the fixed effects model and the random effects model.	13
1.3	The red shrimp <i>Aristeus antennatus</i>	17
2.1	Spline representation of the smooth effect of time on the nat- ural logarithm of the CPUE.	27
2.2	The set of bases used to estimate the curve in figure 2.1. Bases: B-spline, number of knots: 13, polynomial degree: 3, knots equally spaced.	28
2.3	B-spline basis functions of degrees $d = 1$ (upper row), $d = 2$ (middle-upper row), $d = 3$ (middle-lower row) and $d = 10$ (lower row) for equally spaced knots (left panel) and unequally spaced knots (right panel).	29
3.1	From left to right. In the upper panels: Histogram (on the left) and kernel density estimation (on the right); in the mid- dle panels: boxplot (on the left) and empirical distribution function of <i>cpue</i> (on the right); in the lower panels: QQ-plot for the Gamma (on the left) and for the Normal distribution (on the right). Red lines: the theoretical Gamma distribution.	45
3.2	The temporal series of CPUE for each vessel.	47
3.3	The temporal series of CPUE monthly average across all vessels.	47
3.4	Autocorrelation plots of the vessel with <i>code</i> = 2. From above to below: time series of CPUE (on the left) and its first dif- ference (on the right); autocorrelation (ACF) and partial au- tocorrelation (partial ACF).	48
3.5	Autocorrelation plots of the vessel with <i>code</i> = 10. From above to below: time series of CPUE (on the left) and its first difference (on the right); autocorrelation (ACF) and partial autocorrelation (partial ACF).	49

3.6	Relationship between <i>cpue</i> and <i>idmonth</i>	49
3.7	Relationship between <i>cpue</i> and <i>trips</i>	50
3.8	Relationship between <i>hp</i> and <i>grt</i> , <i>cpue</i> and <i>grt</i> and $\log(cpue)$ and <i>grt</i>	50
3.9	Relationship between <i>cpue</i> and <i>nao</i>	51
3.10	Relationship between <i>cpue</i> and <i>month</i>	51
3.11	Diagnostic for Model AMM. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	62
3.12	Scatterplots of standardized residuals vs. fitted values of the AMM for each level of code.	62
3.13	QQ-plots for normality of the residuals of the AMM for each level of code.	63
3.14	Diagnostic for Model GLMM. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	63
3.15	Scatterplots of standardized residuals vs. fitted values of the GLMM for each level of code.	64
3.16	Diagnostic for Model GLM. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	64
3.17	Diagnostic for Model GAM. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	65
3.18	Diagnostic for Model GAMM. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	65
3.19	Diagnostic for Model GAMM-AR1. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	66
3.20	Diagnostic for Model GAMM-MA1. Clockwise from top left: 1) QQ plot of residuals, 2)plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.	66

3.21	Scatter-plots of standardized residuals vs. fitted values of the GAMM for each level of code.	67
3.22	QQ-plots for normality of the residuals of the GAMM for all levels of code.	67
3.23	QQ-plot for normality of the random effects of all mixed effects models.	68
3.24	Comparison of fixed effects estimations. Confidence intervals of fixed parameters estimated by the three methods (from the left to the right panels: <i>constant</i> , <i>trb</i> and <i>month</i>). Labels indicates: “fr”=frequentist, “fB”=full, Bayesian, “eB”=empirical Bayesian.	71
3.25	Partial effects of the covariates using frequentist REML inference. 95% Bayesian credible interval.	73
3.26	Partial effects of the covariates using empirical Bayesian REML inference. 80% and 95% Bayesian credible interval.	74
3.27	Partial effects of the covariates using the full Bayesian inference. 80% and 95% Bayesian credible interval.	75
4.1	Output of the code <code>summary(fit\$lme)</code> for the frequentist model in R.	80

List of Tables

3.1	List of variables	44
3.2	Basic descriptive statistics.	44
3.3	Lilliefors test for normality of residuals and random effects. For each model the statistic L and the corresponding p -value are given for residuals (resid) and for random effects (random) respectively. The last one only for mixed models.	60
3.4	Deviance explained and F tests. mod_1 and mod_2 as in the statistic formula 3.24. Null is the null deviance, with $y \in$ Γ . p =degree of freedom of mod_2 , D_R =residual de- viance, DE =deviance explained, $DE\%$ =proportion of DE , F =F statistic, p - value =p-value of the statistic.	61
3.5	Estimations for the GAMM using the frequentist inference. . .	69
3.6	Estimations for the GAMM using the empirical Bayesian in- ference.	70
3.7	Estimations for the GAMM using the full Bayesian inference. .	70

Chapter 1

Introduction

In the present study the incorporation of random effects is proposed to model CPUE (Catch Per Unit Effort) fluctuations of a fishery resource over a quite long time period. The CPUE is a standardized indicator of the amount of an exploited marine resource. Usually this index is modelled considering only a fixed effects design [2, 43], while here the possibility of the incorporation of mixed effects is implemented and discussed. Different covariates and factors were employed as distinct sources of variability in this regression analysis.

1.1 Fixed effects vs Random effects

Commonly, in regression studies, the coefficients are considered as fixed. In fact it is somewhat useful, mainly because the inference is relatively easy. However, there are cases in which it makes sense to assume some random coefficients. These cases typically occur in two situations:

- when the main interest is to make inference on the entire population, which some levels are randomly sampled from,
- and/or when the observations are correlated.

For example, in biological and medical studies, observations are often collected from the same units (e.g. individuals) over time. It may be reasonable to assume that correlations exist among the observations from the same individual.

Fixed effects are parameters associated with an entire population or with certain repeatable levels of experimental factors, while

Random effects are associated with “unrepeatable” individual experimental units, drawn at random from a population.

A model with both fixed and random effects is called **mixed effects** model.

By associating common random effects to observations sharing the same unit or level of a classification factor, mixed effects models flexibly represent the covariance structure induced by grouping data.

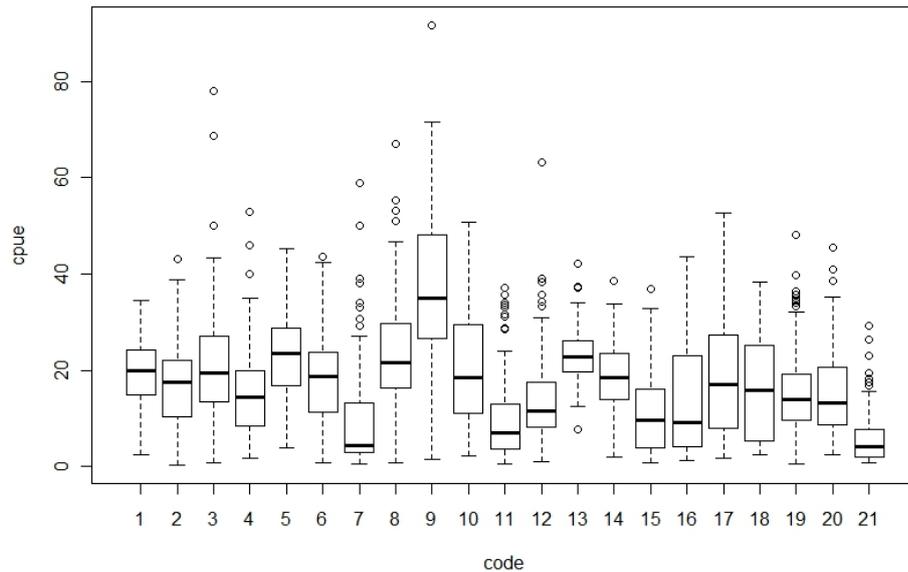


Figure 1.1: Boxplot of the CPUE expressed in kg/trip for 21 vessels.

The following example, proceeding from the application study presented in Chapter 3, could clarify the importance of incorporate fixed or random effects in a model. Data of red shrimp (Figure 1.3) landings were used to build a one-way classification model (that means observations classified according to a single characteristic), where the response CPUE is classified according to vessels. That relationship can be analyzed either with a fixed or a random effects model. The distinction between the two models is according to whether someone wish to make inference about: particular levels of the classification factor (fixed effects) or the population from which these levels come from (random effects). The quantities someone would be interested in estimating from this data set are the average CPUE for a “typical” vessel (the expected CPUE), the variation in average of CPUE among vessels (that is the **between-vessel variability**), and the variation in the observed CPUE for a single vessel (that is the **within-vessel variability**). Box-plots of the CPUE with respect to each level of factor vessel (named *code*) are presented in Figure 1.1. It is evident that there is some variability in the mean CPUE

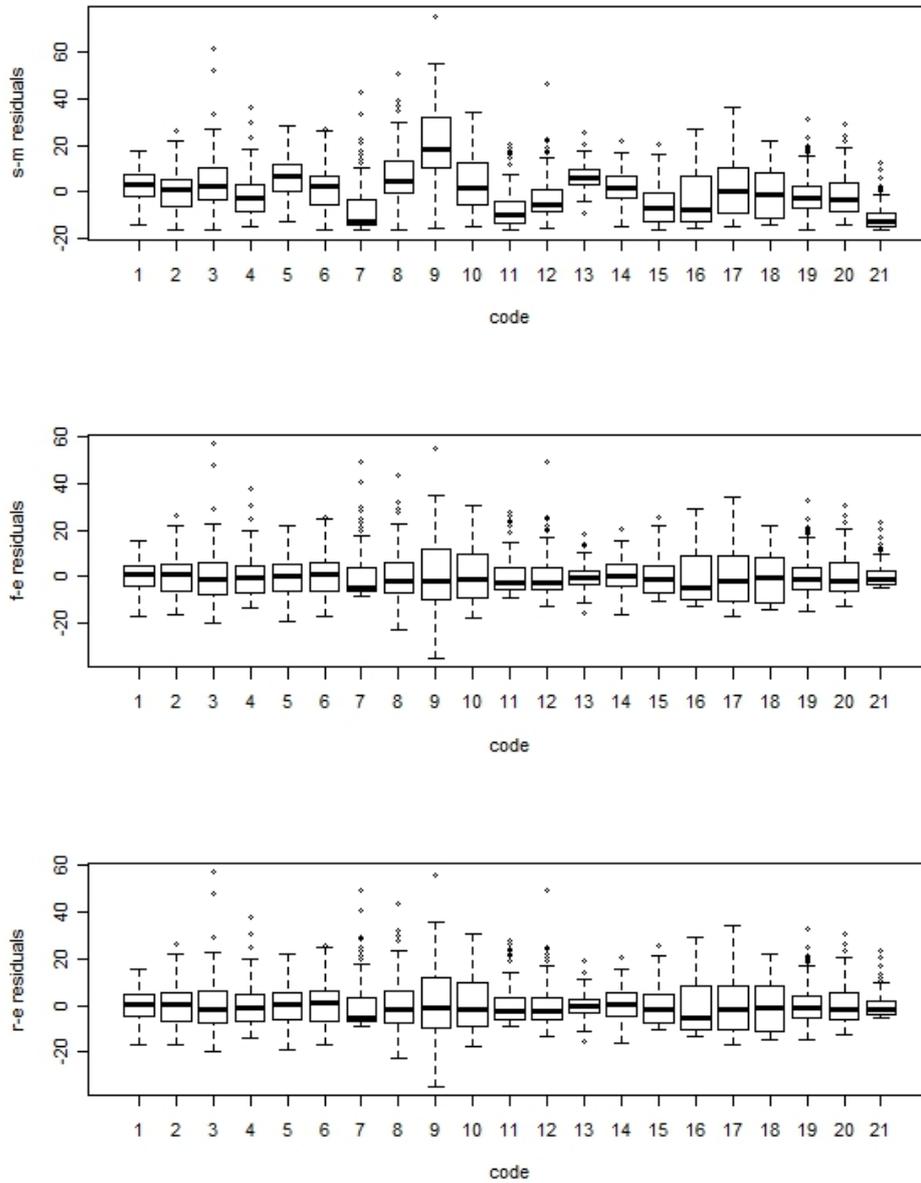


Figure 1.2: boxplot of residuals vs vessel code. From the top to the down: residuals of the single mean model, of the fixed effects model and the random effects model.

for the different vessels.
Consider first the simple model

$$y_{ij} = \alpha + \epsilon_{ij}, i = 1, \dots, M, j = 1, \dots, n_i, \quad (1.1)$$

where y_{ij} is the observed CPUE for observation j on vessel i , α is the mean CPUE across sampled vessels, and the ϵ_{ij} are independent errors belonging to the normal, $N(0, \sigma^2)$. The number of vessels is $M = 21$, while the total number of observations is $N = \sum_{i=1}^M n_i = 2354$, where n_i is the number of observations on vessel i . The estimated parameters of model 1.1 are: $\hat{\alpha} = 16.68$ and $\hat{\sigma} = 11.27$, where $\hat{\sigma}$ is the residual standard error or, in this case, within vessel variability. In this model the grouping structure by vessel is completely ignored. The box-plots of the residuals from the fit of equation 1.1 against vessels, displayed in the upper panel of Figure 1.2, illustrate the problem in which someone incures if the classification factor is ignored: the “group effects” are incorporated into the residuals, leading to an inflated estimate of the variability in the response. The “vessel effects” indicated in Figure 1.1 and in the top of Figure 1.2 may be incorporated into the model by allowing the mean of each vessel to be represented by a separate parameter. This parameter can be considered as fixed or random. The fixed effects model is given by

$$y_{ij} = \alpha_i + \epsilon_{ij}, i = 1, \dots, M, j = 1, \dots, n_i, \quad (1.2)$$

where the α_i represents the mean CPUE of vessel i and, as in model 1.1, the errors ϵ_{ij} are assumed to be independently distributed as $N(0, \sigma^2)$.

As expected, there is a difference in estimations. The residual standard error obtained, $\hat{\sigma} = 9.80$, is lower than the corresponding estimate obtained for the single-mean model 1.1, indicating that it has accounted for the vessel effects. This is illustrated by the box plot of the residuals versus vessels (middle panel of Figure 1.2).

Even though the fixed effects model 1.2 accounts for the vessel effects, it only models the specific sample of vessel, while the main interest can be in the total population of vessels. In this case the fixed effect model does not provide an estimate of the between-vessel variability. Another drawback of this fixed effects model is that the number of parameters in the model increases linearly with the number of vessels. In other words, it depends on the number of vessel sampled.

A random effects model circumvents these problems by treating the vessel effects as random variations around a population mean. The following re-parametrization of model 1.2 helps to motivate the random effects model. Hence,

$$y_{ij} = \bar{\alpha} + (\alpha_i - \bar{\alpha}) + \epsilon_{ij}, i = 1, \dots, M, j = 1, \dots, n_i, \quad (1.3)$$

where $\bar{\alpha} = \sum_{i=1}^M \alpha_i / M$ represents the average CPUE for the vessels.

The random effects model replaces $\bar{\alpha}$ by the mean CPUE over the population of vessels and replaces the deviations $\alpha_i - \bar{\alpha}$ by random variables whose distribution is to be estimated. Assume that each vessel is associated with a random effect whose value is unobservable. A random effects model for the one way classification is

$$y_{ij} = \alpha + b_i + \epsilon_{ij} \quad (1.4)$$

where α is the mean CPUE across the population of vessels being sampled, b_i is a random variable representing the deviation from the population mean of the mean CPUE for the i -th vessel, and ϵ_{ij} is a random variable representing the deviation in CPUE for observation j on vessel i from the mean CPUE. Furthermore is assumed that the random variables b_i , $i = 1, \dots, M$ and ϵ_{ij} , $i = 1, \dots, M$; $j = 1, \dots, n_i$ are independent and normally distributed random variables with mean zero and have constant variances, σ_b^2 for the b_i , or **between-vessel variability**, and σ^2 for the ϵ_{ij} , or **within-vessel variability**. That is,

$$b_i \in N(0, \sigma_b^2), \quad \epsilon_{ij} \in N(0, \sigma^2) \quad (1.5)$$

The b_i are called random effects because they are associated with the particular units (vessels in this case). Now, observations made on the same vessel share the same random effect b_i , thus, they are correlated. The covariance between observations of the same vessel is σ_b^2 corresponding to a correlation of $\sigma_b^2 / (\sigma_b^2 + \sigma^2)$. The parameters of the model 1.4 are α , σ_b^2 and σ^2 . And their estimations are: $\hat{\alpha} = 16.68$, $\hat{\sigma}_b^2 = 6.40$ and $\hat{\sigma}^2 = 9.80$. However the between-vessel variability, σ_b^2 , does not exceed the within-vessel variability, σ^2 , it is considerably high (more than $\frac{1}{3}$ of σ^2). Note that the number of parameters will always be three, irrespective of the number of vessels in the data. In the lower panel of Figure 1.2, the residuals are plotted versus the factor levels. Residuals have almost the same behavior of those in the fixed effects model. In fact the within-vessel variability is almost the same and the imperceptible difference between the residual vectors of both models is probably due to the different estimation procedure used.

1.2 The historical controversy between frequentists and Bayesians

Some backgrounds on frequentist and Bayesian “philosophies” are briefly commented in this section to introduce the estimation techniques in a conceptual point of view. The frequentist, figuring prominently in 20-th Century, is actually considered the classical approach, while the Bayesian, which dominated the 19-th Century, recently emerged as an alternative to the frequentist perspective. Which approach will predominate in the 21-st Century is an open debate [21].

Frequentist and Bayesian points of view reflects two different attitudes to the process of doing science. Frequentist statisticians are cautious with their modeling assumptions, while Bayesians tend to be optimistic. In pursuing their aims, Bayesians try to use all the information at disposal. The frequentist base his procedures on the likelihood which defines the probability distribution of observed data. Properties of the procedure are evaluated in this repeated sampling framework for fixed values of unknown parameters.

The Bayesian requires in addition to a sampling model, a prior distribution of all unknown quantities in the model. The notion of probability is related with uncertainty of knowledge rather than with variability of outcome. The prior and likelihood are used to compute the conditional distribution of the unknowns given the observed data (the posterior distribution), from which all statistical inferences arise.

Bayesian methods are older than frequentist, dating to the 1763 original paper by the Rev. Thomas Bayes, a minister and amateur mathematician. The area generated some interest by Laplace, Gauss, and others in the 19-th Century, but the Bayesian approach was ignored (or actively opposed) by the statisticians of the early 20-th Century.

Pioneering papers on the formulation of statistical inference by R. A. Fisher (e.g. 1922) laid the foundations for a frequentist theory. Despite it, during this period several prominent non-statisticians, most notably Harold Jeffreys (a physicist) and Arthur Bowley (an econometrician), continued to lobby on behalf of Bayesian ideas (which they referred to as ‘inverse probability’). Then, beginning around 1950, statisticians such as L.J. Savage, B. de Finetti, D. Lindley and many others began advocating Bayesian methods as remedies for certain deficiencies in the classical approach.

1.3 Background of the application study

The red shrimp, *Aristeus antennatus* (Risso, 1816) (Figure 1.3), is present in the entire Mediterranean Sea (except the Adriatic) and in the Atlantic from the north Iberian Peninsula to Angola [2].

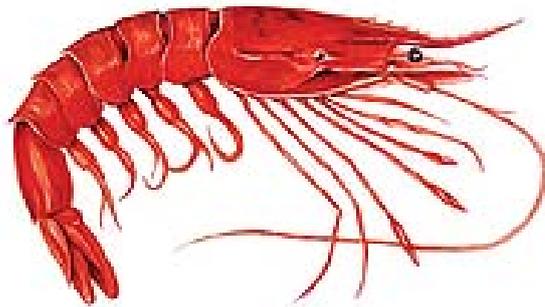


Figure 1.3: The red shrimp *Aristeus antennatus*.

This crustacean is found within 100 – 3300 m depth [54], but in the NW Mediterranean it is abundant only from 600 to 800 m. Its shoals undergo important movements over a short time scale (weekly to seasonal [16, 55]) resulting in some uncertainty the exact location of the high-density shoals. Movements are linked to feeding and reproductive behavior [15, 56].

The red shrimp represents an important target species of deep-water trawl fisheries in the Western Mediterranean [2]. Its fishery is very lucrative, due to the high commercial value of the product (from 20 € kg⁻¹ in 1992 to 50 € kg⁻¹ in 2004), but involves high costs and risks because of conducted in deep-waters of the continental slope (450 – 900 m) and near submarine canyons [61].

In the Catalan Sea, *A. antennatus* is fished by the largest vessels of the trawl fleet [44] along the continental slope between 400 and 800 m depth. As for many Spanish Mediterranean harbors, the trawl fleet of Barcelona operates the red shrimp deep-water fishery on a daily trip basis.

Fishery showed important fluctuations over the mid- and short-term (cycles of minimum catches every 8 – 10 years [8]) and seasonal fluctuations (Autumn-Winter fishing occurs usually around 400 m, while during Summer in deeper waters).

In the period between years 1950 – 2000 a general decrease in catches was observed [2], due, among other things, to an important increase in the engine power of the trawlers. Nowadays the catches increased, but at a very high energetic cost: the engine power required to produce one unit of red shrimp is ten times what it was more than 50 years ago. The high prices fetched by red shrimps (typically above 30 € kg⁻¹ on first sale) explain the increased effort towards this deep-water living resource, raising concern about its long-term sustainability [44].

Maynou et al. [44] suggest that the high temporal variability of the CPUE of red shrimp is due to both inter- and intra-annual variations of the resource and the technical characteristics of boats (e.g. HP and length). As pointed out elsewhere [57], the characteristics of the Spanish Mediterranean trawl fishery give a premium to larger boats with powerful engines, able to reach far-off fishing grounds in shorter periods (as the fishery is limited to 1-day trips). More powerful engines also allow for the trawling of larger nets and for large powerful winches suitable for towing at great depth.

Also the contribution of the individual vessel must be considered, because of fishing skills of the individual fishermen or other factors not captured by the technical characteristics, such as investment in technology, are important in this fishery. As an example, short-time migrations of *A. antennatus* make difficult to point to the exact location of high-density shoals and imply a certain mastering by the fishermen of red shrimp life history [43].

More recent studies have related red shrimp catches by the North Atlantic Oscillation (NAO) as a simplified environmental indicator [45].

The NAO refers to the most prominent atmospheric phenomenon over the middle and high latitudes of the Northern Hemisphere [34]. It has profound implications on the weather and also affects human economic activities, such as agricultural harvests and fisheries yields [34]. NAO [62] has a demonstrated influence (both direct and indirect) on the biology of marine organisms, including fish stocks [17]. It may act on biological organisms at different levels (individual, population) through physiology (metabolic and reproductive processes) or through trophic relationships, including ecological cascade effects.

Relationship between NAO index and the evolution of catches of *A. antennatus* were observed in many representative ports of the Catalan Sea [45]: the mean annual NAO index is positively correlated with the annual catches at different time lags in each port (from 1 to 4 years).

Maynou [45] proposed that NAO-induced environmental variability may enhance food supply to *A. antennatus* and hence strengthen the reproductive potential of particular year classes, which result in increased catches 1 to 3

years later, although other possible effects of environmental variability on the population dynamics of this species are worth investigating.

1.4 Outline of the study

The first purpose of the present study was to validate the use of mixed models in modelling fisheries data sets as an effective alternative to the widely used fixed effects models, often unsuitable to data. This aim was carried out building and comparing a wide range of linear and additive models, from the simple linear model (LM) up to the generalized additive mixed model (GAMM) (see Chapter 2). The analysis of residuals was used to check the validity of models and the proportion of deviance explained ($DE\%$) to select the best model.

The second purpose was to compare the fitting procedures at date available for generalized additive mixed models. In particular, the following inference approaches were compared:

1. the frequentist REML procedure;
2. the full Bayesian perspective, that benefits of MCMC techniques;
3. and the empirical Bayesian REML procedure.

The mean squared error, MSE [66] was used as evaluation criterion to compare the three fits.

The sources of variability taken into account during the regression analysis were:

- technical characteristics of vessels, such as HP and/or GRT;
- intensity of fishing activity, estimated by the monthly number of trips performed by each vessel;
- the environmental variable NAO, using mean annual values;
- all source of variability not captured by the above cited variables, such as the time and the vessels.

This study highlights the importance to take into account the between-group variability of vessels as random effect. Technical characteristics of vessels, such as HP or GRT, give a limited information regarding to the effect that vessels can generate on the response: boats strongly differ each

other due to non strictly technical reasons, e.g. fishermen ability. Moreover, since boats change along years, a fixed number of levels of this factor is not available. That also comprises that the variability in the entire population of boats becomes of main interest. Considering vessel as a random effect also permits predictions of biomass indices and fishery landings.

Another problem related to fishery data is that time series are not available for all boats in the same time. Mixed models allow to combine data sets in a single continuous time-series.

To date, the modelling of catches [43] has been carried out mainly by GLMs, that do not allow to point out possible non parametric relationships nor the incorporation of vessel variability as a random effect. Only three studies are available in the literature regarding to linear mixed models ([12, 4, 32]). Besides in the past many continuous explanatory variables have been used as categorical (e.g. the time, grouped in categories of years and months). The possible non parametric influence of some covariates and the replacement of categories by continuous, justify the use of additive models.

The work is structured as follows: in Chapter 2 a theoretical introduction to the Generalized Additive Models (GAMMs) and spline regression is given as well as some notion about inference methodologies; in Chapter 3 GAMMs are applied to the case of red shrimp fishery; in Chapter 4 the software used to fit models are briefly introduced and some example of the syntax implemented; and in Chapter 5 a conclusion of the study is discussed.

Chapter 2

Generalized additive mixed model formulation and inference

In this chapter some theory on generalized additive mixed models (GAMM) [39] is given. First the linear model (LM) is introduced as the simplest regression model. Then, moving on through intermediate models and adding some complexity, the GAMM formulation is achieved.

There are three ways to add complexity and flexibility in these regression models:

1. incorporate random effects
2. admit responses not belonging only to the Gaussian distribution
3. admit smooth functions.

Thus, to accomplish an ordered presentation of all models, they are split into two main groups: linear and additive models. Linear models, LM, LMM, GLM and GLMM will be presented in 2.1.1 first, while moving into the additive context, AM, AMM, GAM, and finally GAMM will be exposed in 2.1.2.

2.1 GAMM formulation

2.1.1 From LM to GLMM

Consider n pairs of observations (x_i, y_i) , where y_i is an observation of random variable, Y_i , with expectation, $\mu_i \equiv \mathbf{E}(Y_i)$. Y is called response variable or

dependent variable, while x is the predictor or independent variable. In the case of a fixed design, the simplest model which describes the relationship between x and y is:

$$y_i = \mu_i + \epsilon_i \tag{2.1}$$

where $\mu_i = x_i\alpha$ and α is an unknown parameter while, ϵ_i 's, called random errors, are mutually independent random variables, supposed to be $\epsilon_i \in N(0, \sigma^2)$.

If there are more than one predictor, x_j , where $j = 1, \dots, p$ is the number of different predictors, the equation 2.1, using matrix notation, becomes

$$\mathbf{y} = \mathbf{X}\alpha + \epsilon \tag{2.2}$$

where \mathbf{y} is the $n \times 1$ vector of the response, \mathbf{X} is a $n \times p$ matrix of predictor variables, usually called the design matrix of the model, α is a $p \times 1$ vector of unknown parameters and ϵ is $n \times 1$ vector of random errors, with $\epsilon \in N(\mathbf{0}, \mathbf{I}\sigma^2)$. The vector $\mathbf{0}$ denotes a vector with n zero's and \mathbf{I} is the identity matrix of order $n \times n$.

Note that by considering $x_j = x^j$, $j = 1, \dots, p$, the model 2.2 becomes a polynomial regression model. Thus, it is linear in the coefficients even though it is a nonlinear function of the predictor variable, hence also the polynomial is a particular case of linear regression.

The linear model in 2.2 is based on many limiting assumptions, which are:

- Linearity: the dependence between variables can be described only by a straight line and it implies the estimation of parameters (the intercept and the slope parameters for each one of the independent variables);
- Homoscedasticity: the error variance is the same whatever is the value of the explanatory variable, $Var(\epsilon | X = x) = \sigma^2 \forall x$;
- Normality: the error is normally distributed, $\epsilon \in N(\mathbf{0}, \mathbf{I}\sigma^2)$;
- Independence: the errors are uncorrelated.

All those assumptions are useful simplifications to carry out inference procedures, but in real cases if data do not comply with them, the model loses validity.

The incorporation of random effects generalizes in some way the model 2.2. Let consider q vectors of predictor variables, \mathbf{z} of length n . A LMM can be easily built as an extension of the LM and has the form

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} + \epsilon \quad (2.3)$$

where \mathbf{b} is a $q \times 1$ vector containing random effects, $\mathbf{b} \in N(\mathbf{0}, \mathbf{G}_\theta)$, while the vector of random errors has order $n \times 1$ and $\epsilon \in N(\mathbf{0}, \mathbf{R})$. Both \mathbf{b} and ϵ are unobservable. The matrix \mathbf{Z} is the design matrix for the random effects and has order $n \times q$. The covariance matrix \mathbf{G}_θ is positive definite and depends on unknown parameters θ , usually called variance components. Finally \mathbf{R} is a positive definite matrix, sometimes used to model residual correlation. Usually it is equal to $\mathbf{I}\sigma^2$ matrix.

The basic assumptions for 2.3 are that the random effects and errors have mean zero and finite variances. Typically, the covariance matrices $G_\theta = cov(\mathbf{b})$ and $R = cov(\epsilon)$ involve some unknown dispersion parameters, or variance components. It is also assumed that \mathbf{b} and ϵ are uncorrelated.

These models have the ability to model the mean structure (fixed effects) and the covariance structure (random effects and random errors) simultaneously.

LM and LMM permit only gaussian responses. GLM [46](with only fixed effects) and GLMM [3] (with both fixed and random effects) allow response to follow also some other distribution. That is what the letter G (=generalized) refers to.

Thus, a GLM has the form

$$G(\mathbf{y}) = \mathbf{X}\alpha + \epsilon \quad (2.4)$$

and a GLMM is represented as

$$G(\mathbf{y}) = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} + \epsilon \quad (2.5)$$

where $G(\cdot)$ is a monotonic **link function**. If $\mu^b \equiv E(\mathbf{y} | \mathbf{b})$, is the conditional mean of the response, model 2.5 can also be written

$$G(\mu^b) = \eta = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} \quad (2.6)$$

and η is usually called **linear predictor** of the model, while, in the case of more than one covariate, $\eta_j = \mathbf{X}_j\alpha_j$, represents the **partial effect** of covariate x_j .

The assumptions for generalized models are:

- response belongs to one exponential family distribution (that includes Gaussian and categorical responses)
- the mean of the observation is associated with a linear function of some covariates through a link function $G(\cdot)$
- the variance of the response is a function of the mean, that is $\text{var}(\mathbf{y}) = u(\phi)v(\mu)$, where ϕ is called dispersion parameter.

2.1.2 From AM to GAMM

In the same way is possible to build a GAMM adding complexity to a simple AM. The AM is the nonparametric version of the LM.

Consider the response vector, \mathbf{y} , and the corresponding p-vectors of predictors, \mathbf{x} . The AM has a structure something like

$$\mathbf{y} = \mathbf{X}^* \alpha + \sum_{j=1}^p f_j(x_j) + \epsilon \quad (2.7)$$

where \mathbf{X}^* is a model matrix for all strictly parametric model components, α is the corresponding parameter vector and the $f_j(\cdot)$'s are smooth functions of covariates x_j 's, ϵ is the vector of random errors, belonging to a Gaussian $N(\mathbf{0}, \mathbf{I}\sigma^2)$. Note that assumptions of AM are the same of the LM except the linearity.

If random effects are incorporated into the AM, thus the AM becomes an additive mixed model (AMM):

$$\mathbf{y} = \mathbf{X}^* \alpha + \sum_{j=1}^p f_j(x_j) + \mathbf{Z}\mathbf{b} + \epsilon. \quad (2.8)$$

Finally, if we let the response to be some other distribution function and not only the Gaussian, thus, the AMM is converted to a GAMM.

A GAMM, [39, 22], represents the model with higher flexibility and complexity, where mixed effects, smooth terms and a not normal response are admitted. A GAMM has the following structure

$$G(\mathbf{y}) = \mathbf{X}^* \alpha + \sum_{j=1}^p f_j(x_j) + \mathbf{Z}\mathbf{b} + \epsilon \quad (2.9)$$

where $G(\cdot)$ is a monotonic differentiable link function, α is the vector of fixed parameters; \mathbf{X}^* is the fixed effects model matrix, the f_j is the smooth function of covariate x_j (and it is centered), \mathbf{Z} is the random effects model matrix, $\mathbf{b} \in N(0, \mathbf{G}_\theta)$ is the vector of random effects coefficients with unknown positive definite covariance matrix \mathbf{G}_θ , $\epsilon \in N(\mathbf{0}, \mathbf{R})$ is the residual error vector with covariance positive definite matrix \mathbf{R} . The structure of those models allows element of the response vector, \mathbf{y} , to be not longer independent [69].

In analogy to 2.5, the conditional mean of the response, μ^b , is linked to the linear predictor, η , and model 2.9 can be written

$$G(\mu^b) = \eta = \mathbf{X}^* \alpha + \sum_{j=1}^p f_j(x_j) + \mathbf{Z} \mathbf{b} \quad (2.10)$$

and $\eta_j = f(x_j)$ is the partial effect of covariate x_j .

Model 2.10 encompasses various study designs, such as clustered, hierarchical and spatial designs. This is because it is possible to specify a flexible covariance structure of the random effects \mathbf{b} .

Note that in the generalized additive model framework, linear and polynomial models are specific cases of the more general additive model, when smooth effects reduce to linear.

2.1.3 The response distribution

Generalized regression models assume that, given a vector of covariates, \mathbf{x} , the distribution of the response variable y belongs to an exponential family, i.e.

$$p(y|\mathbf{x}) = \exp\left(\frac{y\theta - b(\theta)}{\phi}\right) c(y, \phi) \quad (2.11)$$

where $b(\cdot)$, $c(\cdot)$, θ and ϕ determine the specific response distribution. Many distributions belong to that family: Gaussian, Exponential, Gamma, Chi-square, Beta, Dirichlet, Bernoulli, Binomial, Multinomial, Poisson, and many others. In the following Gamma distribution is briefly discussed, as it is used in the application study.

The Gamma distribution

In the literature, the density function of the Gamma distribution is parameterized in various ways. In the context of regression analysis, the density

is usually parameterized in terms of the mean μ and the scale parameter s . Then, the density of a Gamma distributed random variable y is given by

$$p(y) \propto y^{s-1} \exp\left(-\frac{s}{\mu}y\right) \quad (2.12)$$

for $y > 0$. For the mean and the variance we obtain $E(y) = \mu$ and $Var(y) = \mu^2/s$ and we write $y \in G(\mu, s)$. A second parametrization is typically employed (e. g. for hyperparameters a and b of priors for variance parameters in the Bayesian approach). In this case, the density is given by

$$p(y) \propto y^{a-1} \exp(-by) \quad (2.13)$$

for $y > 0$. In this parametrization we obtain

$$E(y) = a/b, \quad Var(y) = a/b^2 \quad (2.14)$$

for the mean and the variance, respectively and we write $y \in G(a, b)$.

2.2 Spline representation

Smooth effects of continuous variables in model 2.10 must be represented in such a way to be estimated.

Many techniques were developed in last decades, such as the running mean, the Nadaraya-Watson Kernel Smoothing [27], the locally weighted regression [10, 11] and the smoothing spline [50, 67, 64] among others.

In this section, the spline regression is discussed, being the method chosen for the application study. The basic idea is that the shape of a smooth function, $f(\cdot)$, can be approximatively represented with a linear combination of some basis functions:

$$f(\cdot) = \sum_{k=1}^r S_k(\cdot)\beta_k \quad (2.15)$$

where β_k 's are unknown parameters to be estimated and $S(\cdot)$ is the $n \times r$ design matrix consists of the basis functions evaluated at specified observations. Using 2.15 to represent smooth functions in 2.10 yields to a linear model and that allows to make inference in the framework of the linear model [69, 6] (see section 2.3).

A **spline** is a function defined piecewise by polynomials, i.e. dividing the codomain of the covariate into not overlapped intervals and fitting separate

polynomials, called bases (see de Boor [14] and Wahba [64]). The break-points between two adjacent intervals consist of a vector of ordered points, called **knots**. Knots at the borders of the curve are called boundary knots, while the inner are called interior or inner knots. At these points adjacent basis are joined to allow the function to appear continuous.

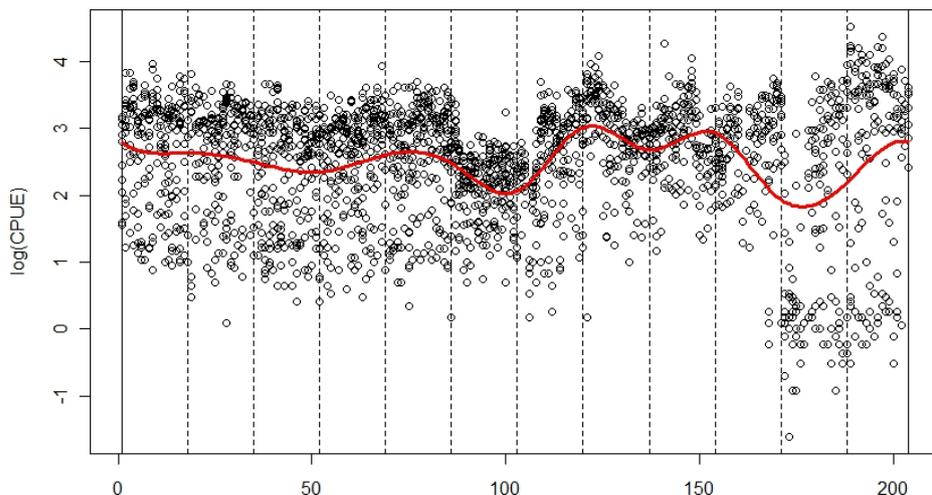


Figure 2.1: Spline representation of the smooth effect of time on the natural logarithm of the CPUE.

In Figure 2.1 a spline estimation is represented. It is an example taken from red shrimp data: the natural logarithm of the CPUE, is plotted with respect to the time. More detailed information about the variables, will be given in chapter 3. Vertical dashed lines represent the knots. Note that the curve is continuous at the knots. This curve is the result of a linear combination of the bases represented in Figure 2.2. Knots can be equally spaced or not. The latter allows the model to be noticeably more flexible. Figure 2.3 shows the behaviour of a basis changing both the degree of polynomial and the position of the knots.

Spline bases perform well in such circumstances, because of they have good approximation theoretic properties. They are simple to construct, easy to estimate and capable to approximate complex shapes. They can be ap-

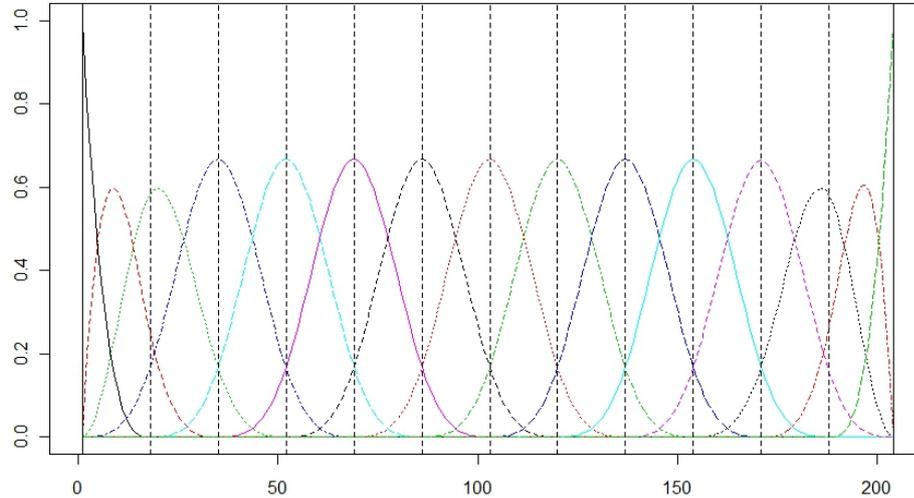


Figure 2.2: The set of bases used to estimate the curve in figure 2.1. Bases: B-spline, number of knots: 13, polynomial degree: 3, knots equally spaced.

plied on either one-dimensional or multi-dimensional data.

The term “spline” usually refers to a wide class of functions, e.g. polynomial splines [14], thin plate regression splines [68], P-splines [18] and Bayesian P-splines [38] among others.

2.2.1 Polynomial spline

A function $f : [a, b] \rightarrow \mathbf{R}$ of the covariate x , is called **polynomial spline** of degree d , with $d \in \mathbf{N}_0$, based on knots $a = \xi_1 < \dots < \xi_m = b$, if it satisfies the following conditions.

- $f(x)$ is a polynomial of degree d for $x \in [\xi_t, \xi_{t+1})$, with $t = 1, \dots, m-1$
- $f(x)$ is $(d-1)$ times continuously differentiable

and represented by 2.15. The space of polynomial spline is a $(m+d-1)$ -dimensional space and a subspace of the space of $(d-1)$ times continuously differentiable functions.

Hence every polynomial spline can be represented by a set of $r = m+d-1$ basis functions.

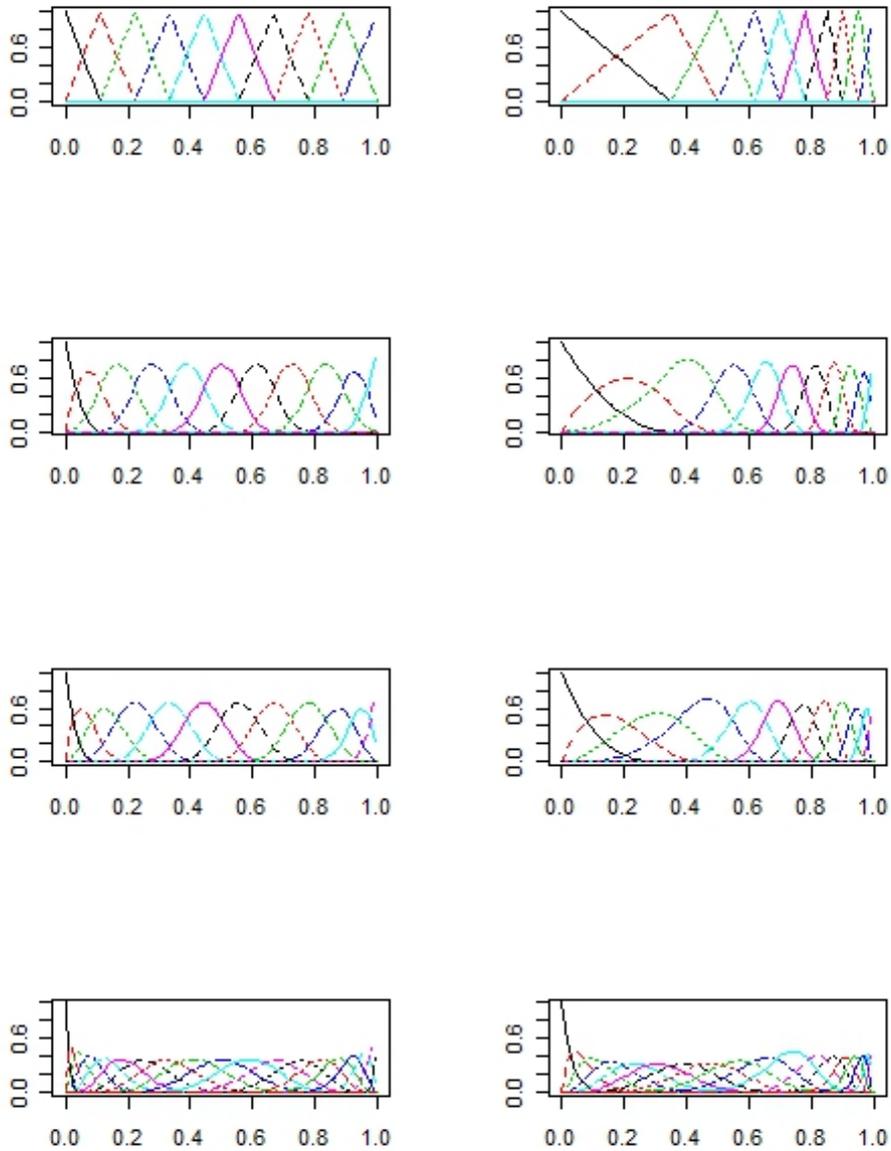


Figure 2.3: B-spline basis functions of degrees $d = 1$ (upper row), $d = 2$ (middle-upper row), $d = 3$ (middle-lower row) and $d = 10$ (lower row) for equally spaced knots (left panel) and unequally spaced knots (right panel).

One commonly used polynomial spline is the cubic spline with $d = 3$.

The polynomial spline $f(x)$ is a **cubic spline** if

1. $f(x)$ is a cubic polynomial over $x \in [\xi_t, \xi_{t+1})$, with $t = 1, \dots, m - 1$
2. $f(x)$ has first and second continuous derivatives at inner knots.

Hence, the cubic polynomial spline of order $M = 4$, has $d = 3$ degree of freedom and $d - 1$ derivatives at the knots. More generally an M -th order spline is a piecewise $d = M - 1$ degree polynomial with $M - 2$ continuous derivatives at the knots.

The **natural spline** is a spline that is linear beyond the boundary knots, and the **natural cubic spline** is a cubic spline that is linear on the boundaries.

2.2.2 B-spline

The Basic spline basis (or simply B-spline) [18] is a particular polynomial spline with excellent computational properties.

B-splines of degree d are obtained by fusing $d + 1$ polynomials of degree d with $d - 1$ derivatives at inner knots. The choice of extra knot is arbitrary. The basis functions are defined recursively as follows.

B-spline of degree $d = 0$ and order $M = 1$ is given by [14]:

$$B_k^0 = \mathbf{1}_{[\xi_t, \xi_{t+1})}(x) = \begin{cases} 1 & \text{if } \xi_t \leq x < \xi_{t+1} \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

Next, higher order B-spline are defined recursively as

$$B_k^d = \frac{x - \xi_t}{\xi_{t+d} - \xi_t} B_t^{d-1} + \frac{\xi_{t+d+1} - x}{\xi_{t+d+1} - \xi_{t+1}} B_{t+1}^{d-1}(x) \quad (2.17)$$

Some properties of the B-splines are:

- they form a local basis
- all basis functions have the same shape for equidistant knots
- at a given position $x \in [a, b]$ exactly $d + 1$ basis functions are positive
- at a given position $x \in [a, b]$ the basis functions sum up to one, that is,

$$\sum_{k=1-d}^{m+l} B_k(x) = 1 \quad (2.18)$$

- each basis function overlaps with $2d$ neighbouring basis functions (except for the boundaries)
- the basis functions are bounded from above and this make them numerically stable
- B-splines determine a large linear model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with

$$\mathbf{X} = \begin{pmatrix} B_{1-d}^d(x_1) & \cdots & B_{m-d}^d(x_1) \\ \vdots & \ddots & \vdots \\ B_{1-d}^d(x_n) & \cdots & B_{m-d}^d(x_n) \end{pmatrix} \quad (2.19)$$

2.2.3 P-spline

One great problem in using splines is to avoid subjectivity proceeds from too many “conditions” to be selected. In fact the functional form of splines is determined by:

- the degree of the basis
- the number of knots
- the position of knots

Obviously that subjectivity also leads to problems of over- or under-fitting data, thus, some regularization is needed.

Two approaches are actually used in estimation procedures to avoid this problem:

1. **Adaptive knot selection procedure**
2. **Penalization approach**

however the latter is usually preferred.

The first approach controls the dimensionality of the basis looking for a set of potential knots which mark the major trend as well as the detailed features of the curve. The idea is to choose the combination of knots using a stepwise forward or backward selection scheme to minimize a model selection criterion. For example the wide used GCV [13]

$$GCV = \frac{n \|\sqrt{\mathbf{W}}(\mathbf{y} - \mathbf{X}\beta)\|^2}{n - \gamma p} \quad (2.20)$$

where \mathbf{W} is a weight matrix, p are the effective degree of freedom of the model and γ an additional smooth parameter; or other model selectors such as the AIC [1], BIC [59] and the C_p statistic [41]. During the last decade procedures which optimize both number and location of knots were proposed [47].

However, before contemplating such an approach it is worth considering the number of possible models. If there are R candidate knots then there are 2^R possible models, assuming the overall intercept and linear term are always present. Thus, often this approach becomes either highly computationally intensive or impossible [52].

The second approach works on the coefficients, β_k of equation 2.15, for which an appropriate penalty with an associated penalty (or smoothing) term is specified. This penalty can be built, e.g., through the second derivative of $f(x)$ [48] and depends on one smoothing term, λ . This parameter controls the trade-off between fidelity to data (λ small) and smoothness of the function estimate (λ large). In that context each kind of spline basis is called **penalized spline** or simply P-spline. Anyhow note that P-spline refers indeed to an estimation procedure rather than to a kind of basis.

P-splines were introduced by Eilers and Marx [18] for (generalized linear) smoothing. They proposed B-splines choosing a sufficiently large number of basis, so that the penalty gives additional, continuous, control over smoothness. Because the penalty is discrete, it can be implemented extremely easily, in contrast to penalties that use the integral of squared higher-order derivatives of the fitted function [48]. They also extended their field of application in several directions, exclusively choosing a B-spline basis with an equispaced grid of knots and (higher-order) differences in the penalty (see e.g. [42, 19]).

The P-spline approach proposed by Eilers and Marx [18] is following presented. Consider for simplicity, a B-splines, \mathbf{B} , and the model $E(y) = \mu = \mathbf{B}\beta$ and minimize the following objective function:

$$Q_B = |\mathbf{y} - \mathbf{B}\beta|^2 + \lambda|\mathbf{D}_d\beta|^2 \quad (2.21)$$

where \mathbf{D}_d is a matrix such that $\mathbf{D}_d = \Delta^d\beta$, the d -th differences of β . Note that $\Delta\beta_j = \beta_j - \beta_{j-1}$, $\Delta^2\beta_j = \Delta(\Delta\beta_j) = \beta_j - \beta_{j-1} - (\beta_{j-1} - \beta_{j-2}) = \beta_j - 2\beta_{j-1} + \beta_{j-2}$, and so on for higher d . Mostly $d = 2$ (such as in the present study) or $d = 3$ is used. In practice the difference penalty on a B-spline, forces the coefficients to follow a smooth pattern. Hence, minimization of Q_B leads to the system of equations

$$(\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_d^T \mathbf{D}_d) \hat{\beta} = \mathbf{B}^T y \quad (2.22)$$

from which the estimation of response variable, \hat{y} , can be obtained. Notice that for $\lambda = 0$ this reduces to the normal equations for linear regression of y on \mathbf{B} . The number of basis functions in \mathbf{B} is chosen “too large”, which means that for $\lambda = 0$ the fitted curve is over-fitting the data, giving a result with too many fluctuations. Depending on the application, the size of the basis can be anywhere from 10 to over 1000. By increasing λ the smoothness can be tuned [18].

2.3 Inference

Statistical inference in GAMMs involves estimation of the nonparametric functions $f_j(\cdot)$, the smoothing parameters, λ , and on all the variance components θ (including σ^2).

In previous sections the implicit linearity behind spline representation 2.10 was emphasized, being it crucial in inference procedures. In fact, estimation of and inference on parameters can be carried out within the framework of GLM with some devices.

2.3.1 Frequentist perspective

The estimation of nonparametric functions, smoothing and variance parameters in the context of generalized additive mixed models is achieved by using REML (Restricted Maximum Likelihood) in the case of Gaussian responses and identity link function and PQL (Penalized Quasi Likelihood) or DPQL (Double Penalized Quasi Likelihood) otherwise [39]. All such methods are likelihood-based techniques and take origin from the ML technique, which has a straightforward application only in fixed models with Gaussian response. ML was used for LMM as well [28], but maximum likelihood estimators (MLE) of variance are, in general, biased.

Could be useful discuss briefly ML and REML estimation of LMM parameters as introductory. Afterwards, the PQL methodology, used to estimate GAMM parameters, is presented, overtaking the GLMM inference, which is a particular case of GAMM.

ML estimation

Under Gaussian mixed model, such as 2.3, the distribution of \mathbf{y} is

$$\mathbf{y} \in \mathbf{N}(\mathbf{X}\alpha, \mathbf{V}) \text{ with } \mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}_\theta\mathbf{Z}^T \quad (2.23)$$

and the log-likelihood is given by

$$l\{\mathbf{y}; \alpha, \theta\} = c - \frac{1}{2} \log(|\mathbf{V}|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\alpha)^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha) \quad (2.24)$$

where c is a constant and θ is the vector of variance components involved in \mathbf{V} . Then, partial derivatives of $l\{\mathbf{y}; \alpha, \theta\}$ w.r.t. the parameters, θ and α can be obtained

$$\frac{\partial l}{\partial \alpha} = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} - \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \alpha \quad (2.25)$$

$$\frac{\partial l}{\partial \theta_r} = \frac{1}{2} \{ (\mathbf{y} - \mathbf{X}\alpha)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_r} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha) - \text{tr}(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_r}) \} \quad (2.26)$$

where θ_r is the r -th component of θ of dimension q . Assuming that α has dimension p and $\text{rank}(\mathbf{X}) = p$, then the MLE is obtained by solving 2.25 and 2.26. With some assumptions, the MLE for α is:

$$\hat{\alpha} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y} \quad (2.27)$$

that requires the estimation of \mathbf{V} and of its components θ . Such estimators are obtained by solving

$$\mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_r} \mathbf{P} \mathbf{y} = \text{tr}(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_r}) \quad (2.28)$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \quad (2.29)$$

Then $\hat{\alpha}$ is obtained by plugging $\hat{\mathbf{V}}$ into 2.27.

REML estimation

The MLE of the variance components are, in general, biased. The “trick” in the REML estimation is to apply a transformation to the data to eliminate the fixed effects, then use the transformed data to estimate the variance component. As before, assume, w.l.o.g., that $\text{rank}(\mathbf{X}) = p$. Let A be an $n \times (n - p)$ matrix such that

$$\text{rank}(\mathbf{A}) = n - p, \mathbf{A}^T \mathbf{X} = 0 \quad (2.30)$$

Then, define $z = \mathbf{A}^T y$, where $z \in N(0, \mathbf{A}^T \mathbf{V} \mathbf{A})$. It follows that the log-likelihood based on z , that is the restricted log-likelihood, is given by

$$l_R\{\mathbf{z}; \theta\} = c - \frac{1}{2} \log(|\mathbf{A}^T \mathbf{V} \mathbf{A}|) - \frac{1}{2} \mathbf{z}^T (\mathbf{A}^T \mathbf{V} \mathbf{A})^{-1} \mathbf{z} \quad (2.31)$$

By differentiating the $l_R\{\mathbf{z}; \theta\}$, one obtains in terms of y

$$\frac{\partial l_R}{\partial \theta_i} = \frac{1}{2} \left\{ y^T \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{P} y - \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_i} \right) \right\} \quad (2.32)$$

where $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{V} \mathbf{A})^{-1} \mathbf{A}^T$ and $i = 1, \dots, q$. Although the REML estimator is defined through a transforming matrix \mathbf{A} , it does not depend on \mathbf{A} (note that, by 2.32, the REML equations do not depend on \mathbf{A}). This fact is important because, obviously, the choice of \mathbf{A} is not unique, and one does not want the estimator to depend on the choice of the transformation.

Note that the restricted log-likelihood is a function of θ only. In other words, the REML method is a method of estimating θ (not α , because the latter is eliminated before the estimation). However, once the REML estimator of θ is obtained, α is usually estimated in the same way as the ML, that is, by 2.27, where $V = V(\hat{\theta})$ with $\hat{\theta}$ being the REML estimator. ML and REML are based on the normality assumption, that is violated in many cases in the practice and likelihood-based inference is difficult, or even impossible. First if the distributions of the random effects and errors are not specified, the likelihood function is simply not available. Furthermore, even if the (non-normal) distributions of the random effects and errors are specified (up to some unknown parameters), the likelihood function is usually complicated and not have an analytic expression. Finally, like normality, any other specific distributional assumptions may not hold in practice. These difficulties have led to consideration of methods other than maximum likelihood. One such method is Gaussian-likelihood, or, as we call it, quasi-likelihood. approach, the REML estimators can be derived from a quasi-likelihood [33]. Furthermore, it has been shown (Richardson and Welsh 1994; Jiang 1996, 1997a) that the REML estimator is consistent and asymptotically normal even if normality does not hold. Therefore, the quasi-likelihood approach is, at least, well-justified from an asymptotic point of view. data are not really normal. Jiang (1996) has pointed out that exactly the same equations will arise if one starts with a multivariate t-distribution. More generally, Heyde (1994, 1997) showed that the REML equations can be derived from a quasi-likelihood. For such a reason, the (Gaussian) REML estimation may be re-

garded as a method of quasi-likelihood. For simplicity, the corresponding estimators are still called REML estimators.

Laplace approximation

When the exact likelihood function is difficult to compute, approximation becomes one of the natural alternatives. A well-known method of approximate integrals is the Laplace approximation. Suppose to need to approximate an integral of the form,

$$\int \exp\{-q(x)\}dx \tag{2.33}$$

where $q(\cdot)$ achieves its minimum value at $x = \tilde{x}$ with $q'(\tilde{x}) = 0$ and $q''(\tilde{x}) > 0$. where q' and q'' denote the gradient (i.e., the vector of first derivatives) and Hessian (i.e., the matrix of second derivatives) of q , respectively. Then, we have

$$\int \exp\{-q(x)\}dx \approx c|q''(\tilde{x})|^{-1/2}\exp\{-q(\tilde{x})\} \tag{2.34}$$

where c is a constant depending only on the dimension of the integral and $|q''(\tilde{x})|$ denotes the determinant of the Hessian.

PQL estimation

With Laplace approximation, one may proceed as in maximum likelihood, treating the approximated likelihood function as the true likelihood function.

Using Laplace approximation, an approximated likelihood can be calculated instead of the exact likelihood. Such approximated likelihood is called Penalized Quasi-Likelihood. PQL is required in the case of non-Gaussian models.

The discussion of GLMM inference is skipped here, because of it can be seen as a particular case of GAMM. Following the estimation procedure proposed by Lin and Zhang [39] is discussed. They propose to estimate jointly smoothing parameters and variance components by using a marginal Quasi-Likelihood, that is an extensions of the REML approaches used by Wahba [63] and Kohn [37]. Smoothing parameters are treated as extra variance components, an idea proceed from inference in Gaussian nonparametric mixed models (see e.g. Wang [65] among others).

According to Lin and Zhang [39], for given values of λ and θ , the spline estimator of $f_j(\cdot)$ maximizes the following penalized log-quasi-likelihood

$$\begin{aligned}
l_{PQ}\{y; \alpha, f_1, \dots, f_p, \theta\} - \frac{1}{2} \sum_{j=1}^p \lambda_j \int_{t_j^0}^{t_j^n} f_j''(x)^2 dx &= \\
l_{PQ}\{y; \alpha, f_1, \dots, f_p, \theta\} - \frac{1}{2} \sum_{j=1}^p \lambda_j f_j^T K_j f_j & \quad (2.35)
\end{aligned}$$

where (t_j^0, t_j^n) defines the range of the j -th covariate and K_j is the nonnegative definite penalty matrix of f_j evaluations (see Green and Silverman [25]). Differentiating equation w.r.t α , smoothing functions and b respectively, yields to a system equations that can be solved by Fisher scoring algorithm with working vectors of response and estimated (centered) smooth functions. Lin and Zhang proposed an alternative to 2.35, since it still usually requires numerical integration, the DPQL approximation (see [39] for further details).

These estimators can be obtained by iteratively fitting a working GLMM to an updated response. The basic idea of this approach is to re-parameterize a GAMM as a GLMM. In fact, the GAMM in equation 2.9 can be reformulated as a GLMM as follows

$$G(\mu^b) = \mathbf{X}\beta + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b} \quad (2.36)$$

by assuming that smooth function estimations can be sundered into a fixed component and a random component. That derives from each $f_j = \mathbf{X}_j\beta_j + \mathbf{U}_j\mathbf{a}_j$, where β_j represents the fixed effects while \mathbf{a}_j the random effects. In particular if \mathbf{B}_k is a set of spline bases with $k = 1, \dots, r$, then the model is specified by $\mathbf{X} = (\mathbf{B}_k)_{k=1,2}$ and \mathbf{U} , such a transformed matrix of remaining bases matrix $\mathbf{B} = (\mathbf{B}_k)_{3 < k < r}$ (see [6]).

But the estimation of smoothing functions, $f(\cdot)$, needs previously the estimation of λ and θ .

The smoothing parameters, λ , and the variance components, θ , can be jointly estimated by using the marginal quasi-likelihood by extending the REML approach of Wahba [63]. They can be obtained by fitting a working GLMM via iteratively fitting a working LMM using REML, with $\tau = (1/\lambda_1, \dots, 1/\lambda_p)$ treated as extra-variance components in addition to θ . Then the GLMM can be fitted iteratively. Hence a marginal quasi-likelihood of (τ, θ) , $l_m PQ\{y; \tau, \theta\}$, can be constructed (eq. 21 in Lin and Zhang [39]). The l_m reduces to REML under AMM.

Equation 21 in [39] is often uncalculable for intractable numerical integration and is necessary an approximation of it, e.g. using Laplace's method (equation 22 in [39]). This approximated $l_m PQ\{y; \tau, \theta\}$ corresponds to the REML log-likelihood under the LMM

$$\mu^b = \mathbf{X}\beta + \mathbf{U}a + \mathbf{Z}b \quad (2.37)$$

with a and b random effects. It follows that τ and θ can be easily estimated by iteratively fitting model 2.37 using REML. After estimating τ and θ , it is possible to use the BLUP estimators of β_j and a_j to construct approximate spline estimators \hat{f}_j by PQL (or DPQL).

2.3.2 Full bayesian perspective

Priors

The unknown functions $f_j(\cdot)$, fixed parameters and variance components for the model 2.10 are considered as random variables and must be supplemented by appropriate prior assumptions. In absence of any knowledge, the appropriate choice for fixed effects parameters is the diffuse prior, i. e.,

$$p(\alpha_i) \propto \text{const} \quad (2.38)$$

while for random effects a Gaussian prior is specified.

Priors for the unknown functions, $f_j(\cdot)$, depend on the type of the covariate and on prior beliefs about the smoothness of $f_j(\cdot)$. Using spline representation of $f_j(\cdot)$ is possible to rewrite the predictor in 2.10 as a long linear model, i.e.

$$\eta = \mathbf{X}\alpha + \sum_{j=1}^p \mathbf{S}_j\beta_j + \mathbf{Z}b \quad (2.39)$$

where $f_j = \mathbf{S}_j\beta_j$.

Thus a prior for a function, f_j , is defined by specifying a design matrix, \mathbf{S}_j , and a prior distribution for the vector β_j of unknown parameters, i.e.

$$p(\beta_j/\tau_j^2) \propto \frac{1}{(\tau_j^2)^{\text{rank}(\mathbf{K}_j)/2}} \exp\left(-\frac{1}{2\tau_j^2} \beta_j^T \mathbf{K}_j \beta_j\right) \quad (2.40)$$

where \mathbf{K}_j is a penalty matrix. The variance parameter, τ_j^2 , is equivalent to the inverse of the smoothing parameter, λ_j , in the penalized likelihood approach.

The unknown variance parameters, τ_j^2 , are supplemented with hyperprior assumptions, e.g. dispersed inverse Gamma priors $p(\tau_j^2) \sim IG(a_j, b_j)$, with the corresponding probability density function

$$\tau_j^2 \propto (\tau_j^2)^{-a_j-1} \exp(-b_j/\tau_j^2) \quad (2.41)$$

for $a_j > 0$ and $b_j > 0$. Usually small values are given to a_j and b_j . Also for the overall variance σ^2 the same hiperprior is given.

Several alternatives have been proposed to specify smoothness priors for continuous covariates. Here, random walk priors or more generally autoregressive priors [22] and Bayesian P-splines [38] are presented.

The random walk

Suppose that \mathbf{x} is a time scale or continuous covariate with equally spaced ordered observations $x^{(1)} < x^{(2)} < \dots < x^{(K)}$, where $K \leq n$ denotes the number of different observations for x . A common approach is to estimate one parameter β_k for each distinct $x^{(k)}$, i.e. $f(x^{(k)}) = \beta_k$, and penalize abrupt jumps between successive parameters using random walk priors. For example, first and second order random walk models are given by

$$\beta_k = \beta_{k-1} + \epsilon_k \text{ and } \beta_k = 2\beta_{k-1} + \beta_{k-2} + \epsilon_k \quad (2.42)$$

with errors $\epsilon_k \in N(0, \tau^2)$ and diffuse priors $p(\beta_1) \propto \text{const}$ or $p(\beta_1) \propto \text{const}$ and $p(\beta_2) \propto \text{const}$, for initial values, respectively. Both specifications act as smoothness priors that penalize too rough functions. A first order random walk penalizes abrupt jumps $\beta_k - \beta_{k-1}$ between successive states while a second order random walk penalizes deviations from the linear trend $2\beta_{k-1} - \beta_{k-2}$. The joint distribution of the regression parameters β is easily computed as the product of conditional densities defined by 2.42 and can be brought into the general form 2.40. The penalty matrix is of the form $\mathbf{K} = \mathbf{D}^T \mathbf{D}$ where \mathbf{D} is a first or second order difference matrix. For example, for a random walk of first order the penalty matrix is given by

$$\mathbf{K} = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 1 & \end{pmatrix} \quad (2.43)$$

The design matrix \mathbf{X} is a simple 0/1 matrix where the number of columns equals the number of parameters, i.e. the number of distinct covariate values. If for the i -th observation $x_i = x^{(k)}$ the element in the i -th row and k -th column of \mathbf{X} is one and zero otherwise. In case of non-equally spaced observations slight modifications of the priors defined in 2.42 are necessary, see [22] for details.

Bayesian P-spline

In alternative to random walk for continuous variables the Bayesian P-spline can be used, such as in the present study. In the Bayesian approach, the difference penalties in 2.21 are replaced by their stochastic analogue, i.e. first or second order random walks (e.i. equations in 2.42) as priors for the regression coefficients.

The amount of smoothness allowed by particular priors specifications for τ_j^2 depends (weakly) on the scale of the response. To avoid the problem, the vector \mathbf{y} is standardized before estimation and results are retransformed afterwards (this also avoid numerical difficulties with MCMC simulations).

The posterior

For the given priors, the posterior of the model is given by

$$p(\beta_1, \dots, \beta_n, \tau_1^2, \dots, \tau_p^2, \alpha \mid \mathbf{y}) \propto l\{\mathbf{y}; \beta_1, \dots, \beta_p, \alpha\} \prod_{j=1}^p (p(\beta - j \mid \tau_j^2) p(\tau_j^2)) \quad (2.44)$$

where $l(\cdot)$ denotes the likelihood which under the assumption of conditional independence is the product of individual likelihood probabilities.

MCMC simulation techniques

Full Bayesian inference is based on MCMC (Markov Chain Monte Carlo) simulation techniques. They overcome difficulties in calculation of the posterior distribution, which is usually numerically intractable. These technics allow to draw random samples from the posterior, whose characteristics (such as the mean, standard deviation and quantiles) can be easily estimated by their empirical analogues. Instead of drawing samples directly from posterior, MCMC devices a way to construct a Markov chain with the posterior as a stationary distribution. Hence the iterations of the transition kernel of this Markov chain converge to the posterior yielding a sample of independent random numbers. Usually the first part of the sample (the burn-in phase) is discarded because of the algorithm needs time to converge. The sampling scheme depends on the supposed distribution of the response variable, e.g. Gaussian or categorical (see the methodology manual of *BayesX* programm for futher details).

2.3.3 The empirical Bayesian perspective

In the empirical Bayesian inference the GAMM is re-parametrized as a GLMM, as proposed by Green [24] and used in the actual frequentist procedures. Basically its prerogatives are:

- to use priors as in the full Bayesian approach
- to use the REML procedure of estimation.

Chapter 3

The application study

3.1 Data source

Available data were obtained from two different sources. The monthly total landings of red shrimp by vessel (expressed in $\text{kg month}^{-1} \text{vessel}^{-1}$), the monthly number of trips performed by each vessel and their technical specifications (Engine Power: HP, and Gross Registered Tonnage: GRT) were obtained from the Barcelona Fishers Association. The average monthly value of the NAO index 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>.

3.2 Data description and exploration

The total number of observations amounts to 2354 for 21 trawlers having prictised deep-fishing in the period from January 1992 to December 2008 (17 complete years).

Table 3.1 contains all variables examined in this analysis, not necessarily all included in final models. A basic statistical description is reported in table 3.2.

The response variable, *cpue* (CPUE, Catch Per Unit Effort), is a ratio commonly used to eliminate temporal and regional trends in fish stocks. The “catch” portion of the measure is expressed as the weight of the entire catch (sometimes as the number). The “unit effort” portion of the rate refers to the unit of time spent by a unit of the tool used to catch (e.g. vessel in this case or peace of gear elsewhere).

Table 3.1: List of variables

Variable	Description
<i>biom</i>	the total monthly landings per vessel of red shrimp
<i>cpue</i>	monthly average catch per unit effort of red shrimp
<i>idmonth</i>	time index of T progressive months, $T = 204$
<i>month</i>	categorical variable of I categories, $I = 12$
<i>trips</i>	number of trips performed by each vessel during the t -th month
<i>hp</i>	Engine Power of vessels
<i>grt</i>	Gross Registered Tonnage of vessels
<i>code</i>	code assigned to each vessel, $C = 21$
<i>nao_j</i>	mean annual NAO index of j years before the year of observed CPUE

Table 3.2: Basic descriptive statistics.

	<i>biom</i>	<i>cpue</i>	<i>trips</i>	<i>hp</i>	<i>grt</i>	<i>nao₁</i>	<i>nao₂</i>	<i>nao₃</i>
Min	0.20	0.20	1.00	160.00	37.70	-1.01	-1.01	-1.01
1-st Qu	26.85	7.90	3.00	300.00	42.83	-0.30	-0.25	-0.25
Median	125.05	15.60	8.00	375.00	48.10	-0.04	0.04	0.05
Mean	173.06	16.68	8.99	370.60	53.85	-0.05	0.04	0.09
3-rd Qu	275.31	23.50	13.75	430.00	54.99	0.25	0.34	0.51
Max	1064.00	91.80	43.00	700.00	113.50	1.11	1.23	1.23
sd	171.57	11.27	6.48	110.95	17.54	0.47	0.57	0.59

	<i>code</i>	<i>month</i>	<i>year</i>
levels	21	12	17

Thus, CPUE was calculated dividing *biom* (the total monthly kg per boat) by the monthly number of fishing trips:

$$cpue = \frac{biom}{\#trips} \quad (3.1)$$

Hence, the CPUE is the monthly average kg per boat [43].

A brief study of density and distribution functions of *cpue* is necessary to select both response distribution and link function in generalized regression models (see Chapter 2). Two estimations of the density function of the *cpue* are reported in the upper panels of Figure 3.1: the histogram and the gaussian kernel estimation.

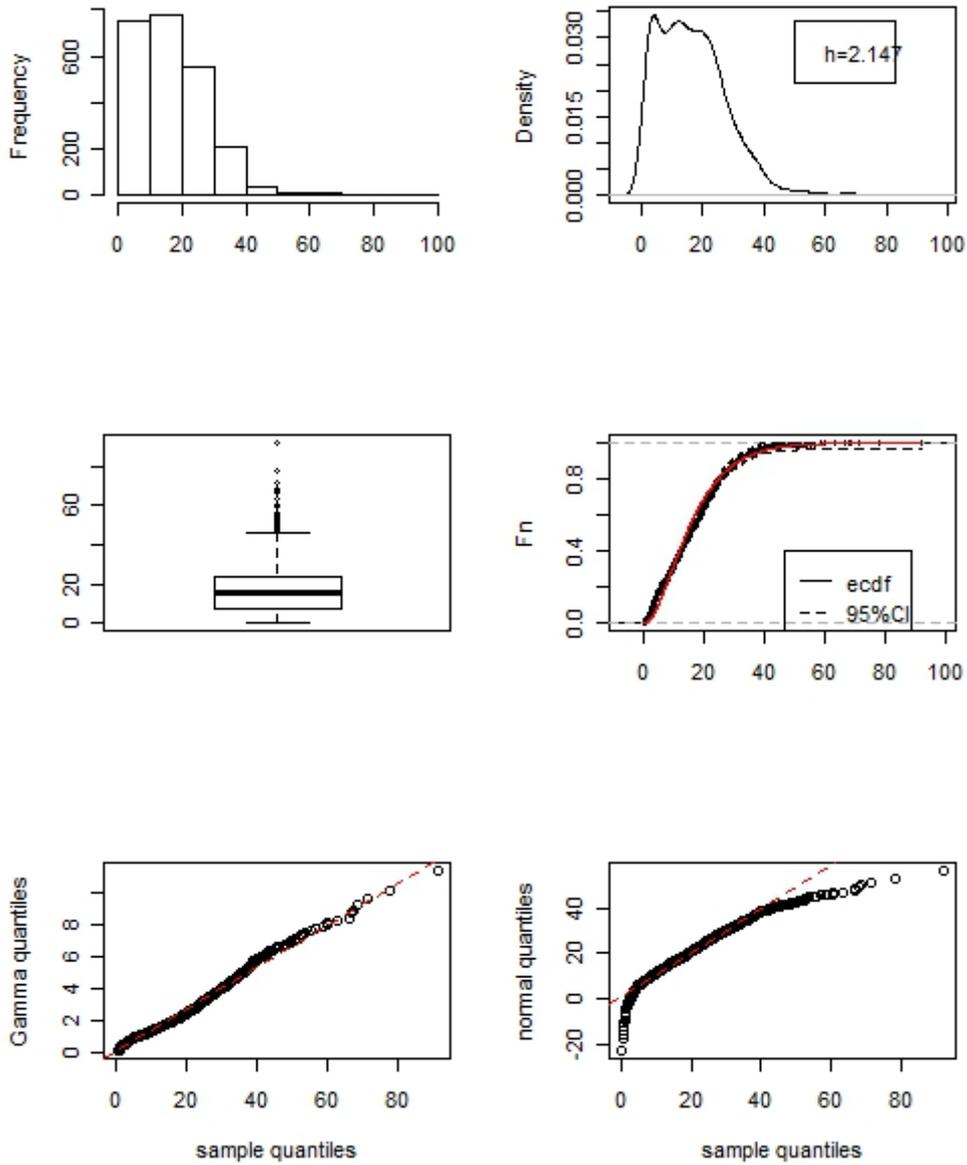


Figure 3.1: From left to right. In the upper panels: Histogram (on the left) and kernel density estimation (on the right); in the middle panels: boxplot (on the left) and empirical distribution function of *cpue* (on the right); in the lower panels: QQ-plot for the Gamma (on the left) and for the Normal distribution (on the right). Red lines: the theoretical Gamma distribution.

Both density estimations show an asymmetric shape, due to many low values and few high values of the response. Those high values could be “atypical” as can be observed in the box-plot in the middle left panel of the same figure. In the right middle panel, also the empirical cumulative distribution function (ecdf) of *cpue* and the 95% confidence intervals, jointly with the theoretical Gamma distribution (a and b estimated parameters) are reported. The theoretical Gamma distribution lies inside calculated intervals of the ecdf, hence it can be considered as the underlying theoretical distribution of the response in the regression analysis. Finally, in the lower panels of the figure also the QQ-plots for the Gamma and the Normal distributions are reported, given one more evidence of the good approximation of data to the Gamma (on the left) and bad approximation to the Normal (on the right).

Confidence intervals were calculated as follows [66]:

$$L(\cdot) = \max\{F_n(\cdot) - \epsilon_n, 0\} \quad (3.2)$$

$$U(\cdot) = \min\{F_n(\cdot) + \epsilon_n, 1\} \quad (3.3)$$

where ϵ_n is given from the DKW (Dvoretzky-Kiefer-Wolfowitz) inequality, by:

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)} \quad (3.4)$$

Parameters a and b were obtained by 2.14.

The temporal series of the CPUE for each vessel are reported jointly in Figure 3.2. In spite of non complete series is available for any of the boats, a similar trend can be observed in the plot: a quite constant trend during first 10 years (up to $idmonth = 90$), an abrupt decreasing around $idmonth = 100$ and high variability after this crash.

The monthly average CPUE within all vessels (Figure 3.3) highlights more clearly the fluctuations above commented and allows to detect another crash in *cpue* around $idmonth = 175$.

The presence of a dependence structure between observations of each vessel was also investigated. Thus, the autocorrelation functions of vessel series were calculated looking for such structure in the data and all series showed some autocorrelation. Here only the autocorrelations for two of the 21 vessels are reported, i.e. with codes 2 and 10, in Figures 3.4 and 3.5 respectively.

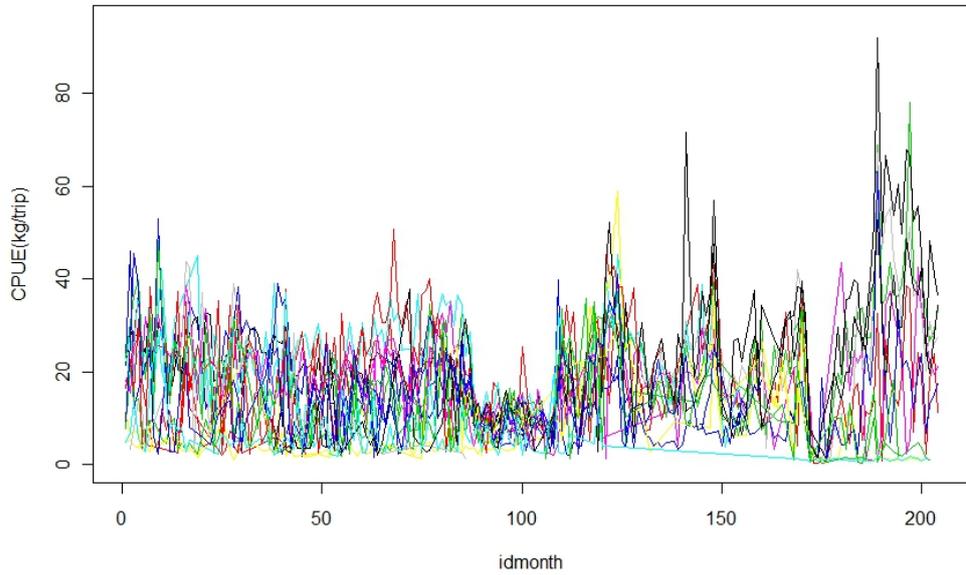


Figure 3.2: The temporal series of CPUE for each vessel.

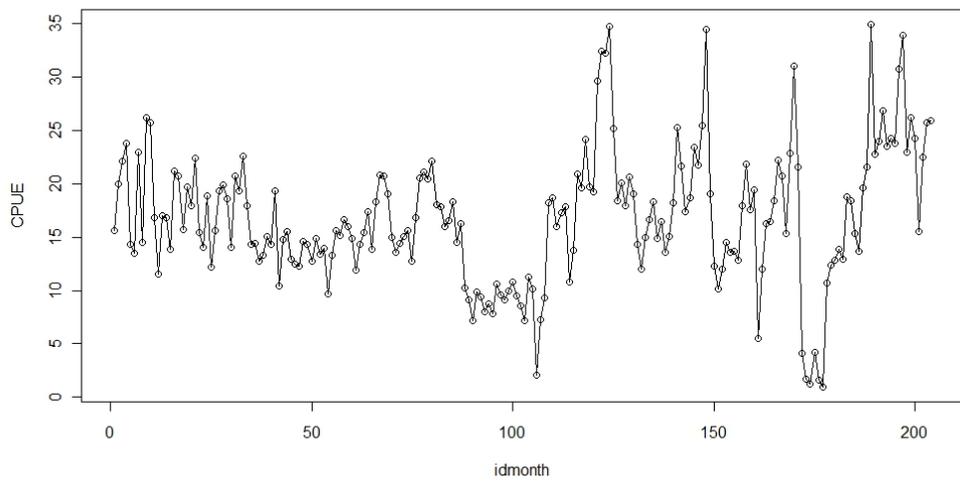


Figure 3.3: The temporal series of CPUE monthly average across all vessels.

Both series show some kind of trend (both in the series and in the ACF in left panels). Hence to avoid it, series were differentiated (1-st order dif-

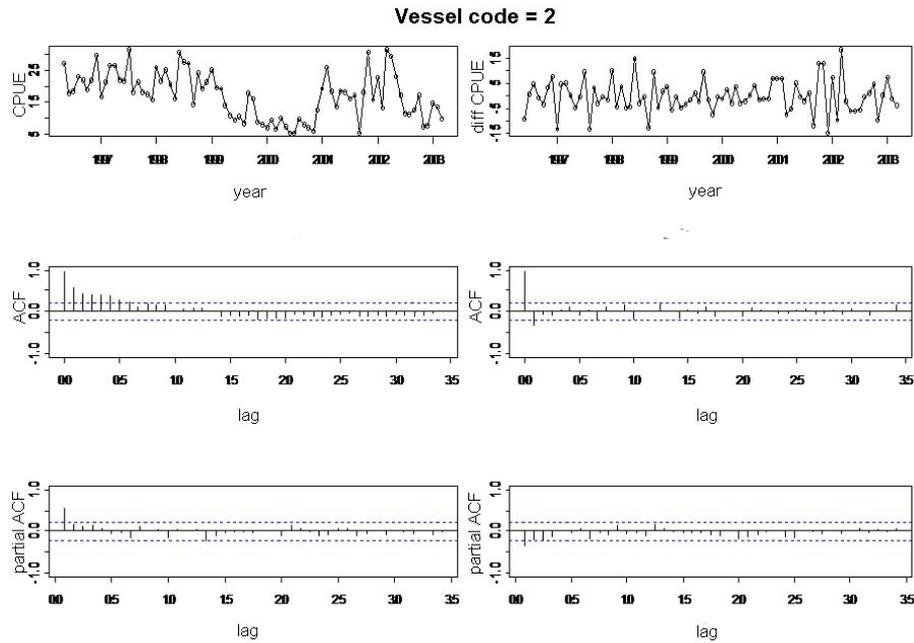


Figure 3.4: Autocorrelation plots of the vessel with *code* = 2. From above to below: time series of CPUE (on the left) and its first difference (on the right); autocorrelation (ACF) and partial autocorrelation (partial ACF).

ferentiation). The partial ACFs of differentiated series show the possibility of an autoregressive or mean average relationship of 1-st order (difficultly of more order).

Follows an exploratory analysis of relationships between variables, used as a guideline in building models.

The Figure 3.6 shows sample points of *cpue* and its logarithm versus *idmonth*. The logarithmic transformation allows to better identify which periods showed extrem values of CPUEs.

days of a month, because it can be that vessels performed more trips for day.

The *cpue* seems to show a positive relationship with the number of trips performed by each boat, if a logarithmic transformation on it is implemented (Figure 3.7). An abnormal pattern can be observed in the first and second panels of Figure 3.7, around *idmonth* = 90 – 110, corresponding to years 1999 and 2000. In those years, high values of number of *trips* correspond to the lowest *cpue* observed along all the time period. That trend will be

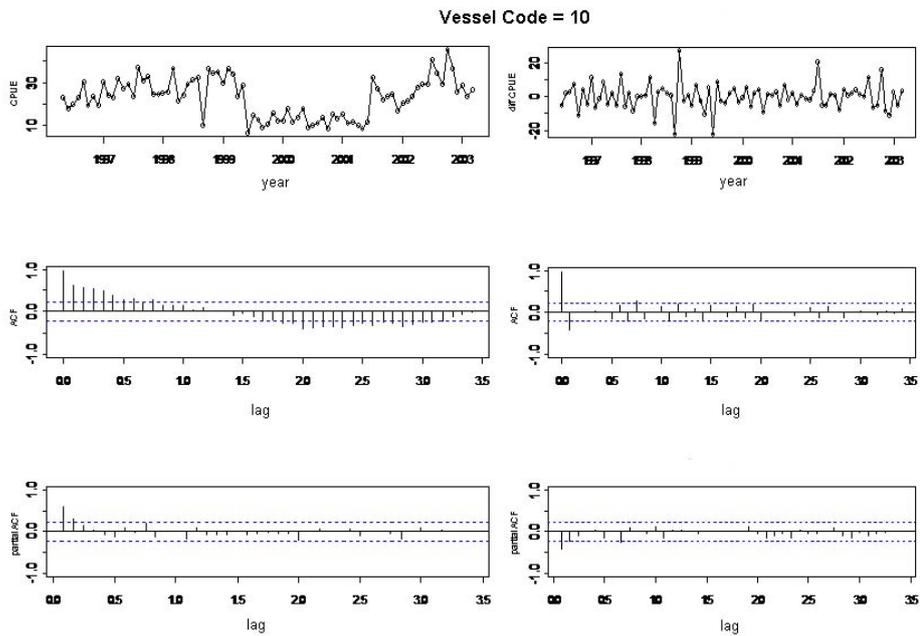


Figure 3.5: Autocorrelation plots of the vessel with *code* = 10. From above to below: time series of CPUE (on the left) and its first difference (on the right); autocorrelation (ACF) and partial autocorrelation (partial ACF).

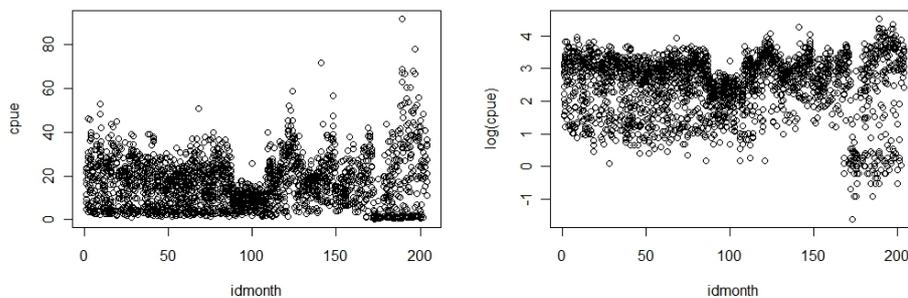


Figure 3.6: Relationship between *cpue* and *idmonth*.

discussed in the biological conclusion in section 3.5.

index, it is affected by the number of trips performed such as it differs within months and vessels.

The two technical characteristics of vessels, *hp* and *grt*, are positively correlated (first panel of Figure 3.8), but high values of *hp* are sub-estimated because of fishermen underreported the real value, thus it could not be a good

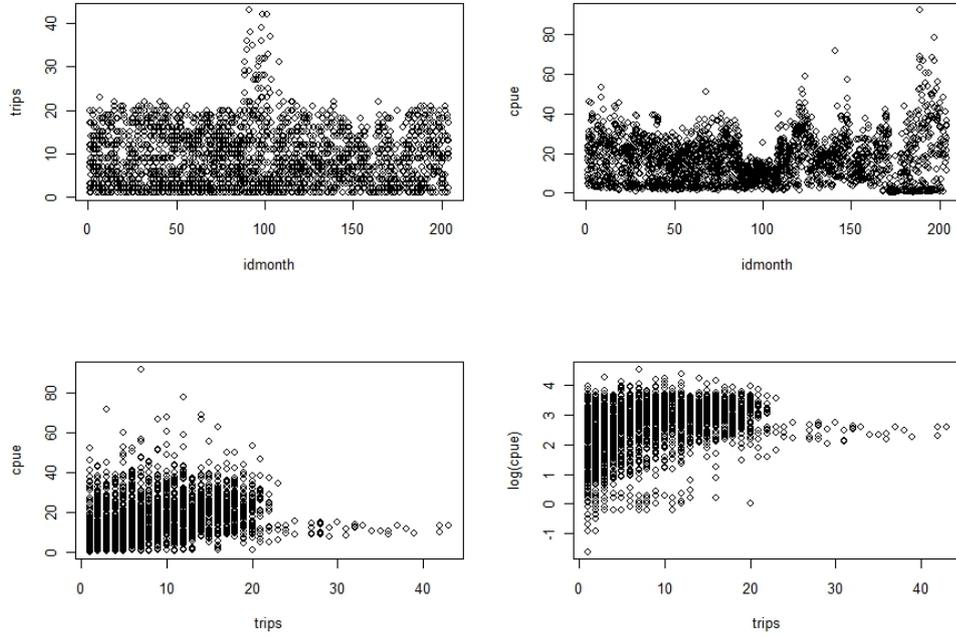


Figure 3.7: Relationship between *cpue* and *trips*.

predictor and only the *grt* were incorporate in the final regression model. The relationship between *cpue* and *grt* is shown in middle and right panels of Figure 3.8. The relationship is more evident applying a logarithmic transformation on the response variable.

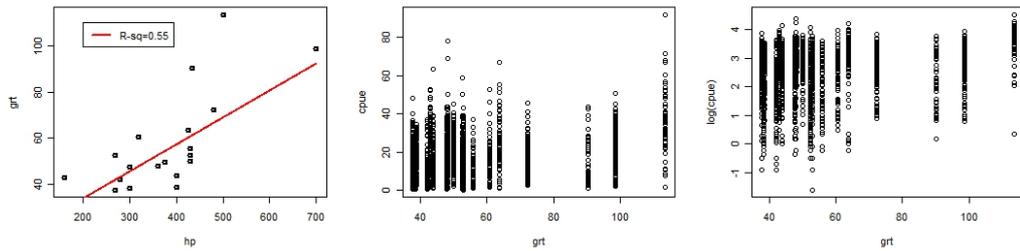


Figure 3.8: Relationship between *hp* and *grt*, *cpue* and *grt* and $\log(\text{cpue})$ and *grt*.

The relationship between *cpue* and *nao* is displayed in Figure 3.9. It can be appreciated that negative values of *nao* affect negatively to the *cpue*.

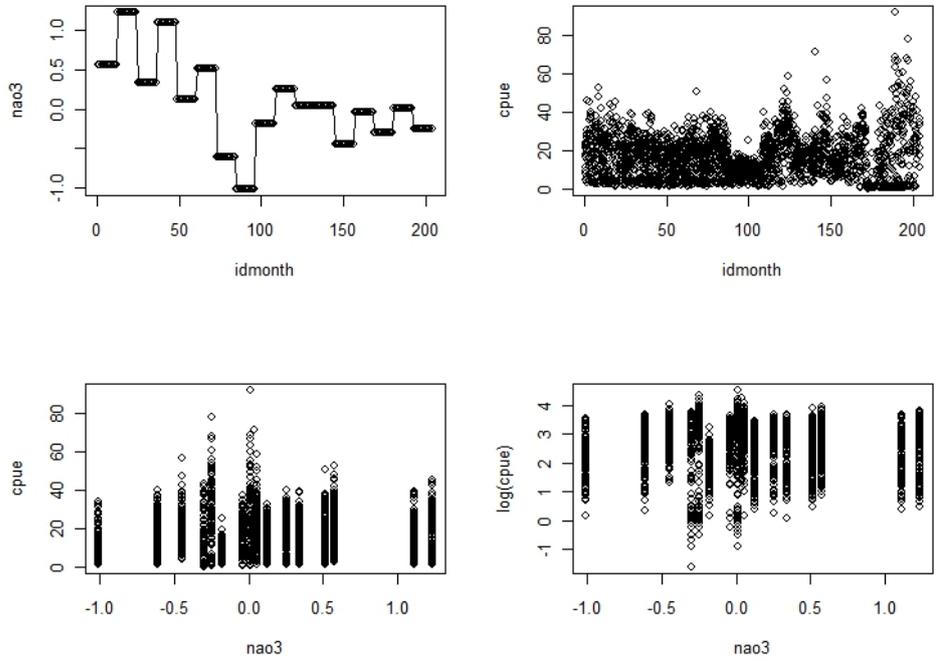


Figure 3.9: Relationship between *cpue* and *nao*.

affects to the *cpue* fluctuations, controlling inter and intra-annual variability of *cpue*: *idmonth* captures the year to year variability and *month* the seasonal variability.

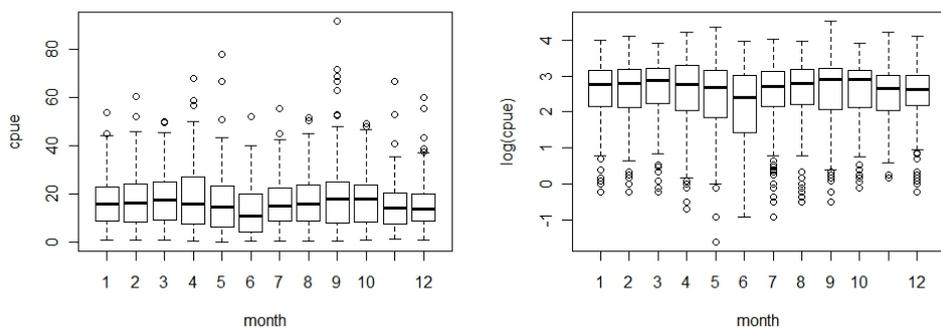


Figure 3.10: Relationship between *cpue* and *month*.

3.3 Materials and methods

3.3.1 Model construction

The relationship between *cpue* and all predictors was initially checked through the frequentist approach using R2.12.0 (packages: “mgcv”, “nlme” and “MASS” as appropriate). REML algorithm was set as the estimation procedure of all models and the P-spline as the basis for smooth functions.

Initially independent models were constructed for all covariates to check their effect on the *cpue* and, if it resulted significant, its nature (linear or nonparametric).

Regarding to the variable NAO, three subsets of the covariate were initially considered, from 1 to 3 years before the year of estimated *cpue*, due to biological reasons commented in Section 1.3. Only the subset with best explanatory potential (percentage of deviance explained, $DE\%$, see below for details about this criterion) was introduced in the final model.

To the other hand, components, i.e. parameters, that were initially significant, but not in more complicate models, were simply eliminated from the model itself. That is, if a covariate was significant as unique predictor, but not significant jointly with other predictors, it was delated. For example, it occurred in the case of the variable *naolag3*, that had a significant parameter in the model GAMM, but it was eliminated in models with autocorrelation structures (see below) because its effect resulted no more significant. In the case of categorical variables, i.e. *month* and *code*, levels with no significant coefficient were one by one eliminated and merged with the reference category. This rearrangement was made separatelly for each class of model built.

This preliminary analysis was omitted from the results for shortness. This procedure allowed to built the best model within each of the theoretical classes of models presented in Section 2.1, i.e. the best LM, the best LMM, and go on up to the GAMM.

As mixed models permits the incorporation of autocorrelation structures in residuals, e.g. autoregressive and moving everage structures (ARMA modelling), also additive mixed models with the incorporation of $ARMA(p,q)$ (as suggested by the exploratory analysis) were investigated. All GAMM- $ARMA(p,q)$ were built up to the orders p and q with no more significant coefficients. GAMM-MA(1) are presented in the results because of all

Thus, the collection of models compared can be summarized as follows:

The linear model (LM):

$$\begin{aligned}
 cpue = & \alpha_0 + grt\alpha_{grt} + idmonth\alpha_{idmonth} + \\
 & trips\alpha_{trips} + nao_3\alpha_{nao_3} + \\
 & month\alpha_{month} + code\alpha_{code} + \epsilon
 \end{aligned} \tag{3.5}$$

The additive model (AM):

$$\begin{aligned}
 cpue = & \alpha_0 + grt\alpha_{grt} + f_{idmonth}(idmonth) + \\
 & f_{trips}(trips) + f_{nao_3}(nao_3) + \\
 & month\alpha_{month} + code\alpha_{code} + \epsilon
 \end{aligned} \tag{3.6}$$

The generalized linear model (GLM):

$$\begin{aligned}
 G(cpue) = & \alpha_0 + grt\alpha_{grt} + idmonth\alpha_{idmonth} + \\
 & trips\alpha_{trips} + nao_3\alpha_{nao_3} + \\
 & month\alpha_{month} + code\alpha_{code} + \epsilon
 \end{aligned} \tag{3.7}$$

The generalized additive model (GAM):

$$\begin{aligned}
 G(cpue) = & \alpha_0 + grt\alpha_{grt} + f_{idmonth}(idmonth) + \\
 & f_{trips}(trips) + f_{nao_3}(nao_3) + \\
 & month\alpha_{month} + code\alpha_{code} + \epsilon
 \end{aligned} \tag{3.8}$$

The linear mixed model (LMM):

$$\begin{aligned}
 cpue = & \alpha_0 + grt\alpha_{grt} + idmonth\alpha_{idmonth} + \\
 & trips\alpha_{trips} + nao_3\alpha_{nao_3} + \\
 & month\alpha_{month} + code b_{code} + \epsilon
 \end{aligned} \tag{3.9}$$

The additive mixed model (AMM):

$$\begin{aligned}
 G(cpue) = & \alpha_0 + grt\alpha_{grt} + f_{idmonth}(idmonth) + \\
 & f_{trips}(trips) + f_{nao_3}(nao_3) + \\
 & month\alpha_{month} + code b_{code} + \epsilon
 \end{aligned} \tag{3.10}$$

The generalized linear mixed model (GLMM):

$$\begin{aligned}
G(cpue) = & \alpha_0 + grt\alpha_{grt} + idmonth\alpha_{idmonth} + \\
& trips\alpha_{trips} + nao_3\alpha_{nao_3} + \\
& month\alpha_{month} + code b_{code} + \epsilon
\end{aligned} \tag{3.11}$$

The generalized additive mixed model (GAMM):

$$\begin{aligned}
G(cpue) = & \alpha_0 + grt\alpha_{grt} + f_{idmonth}(idmonth) + \\
& f_{trips}(trips) + f_{nao_3}(nao_3) + \\
& month\alpha_{month} + code b_{code} + \epsilon
\end{aligned} \tag{3.12}$$

The generalized additive mixed model (GAMM-ARMA(p,q)):

$$\begin{aligned}
G(cpue) = & \alpha_0 + grt\alpha_{grt} + f_{trips}(trips) + \\
& f_{idmonth}(idmonth) + month\alpha_{month} + \\
& code b_{code} + \epsilon, \text{ with } \epsilon \sim ARMA(p, q)
\end{aligned} \tag{3.13}$$

In the cases of generalized response assumption, the Gamma distribution was used as the underlying distribution function with logarithmic link as the descriptive analysis in section 3.2 suggested.

Then, the residuals were analyzed and those models whose residuals irrefutably did not pass the diagnostics were discarded in further analyses.

Finally the deviances were calculated and an F test, based on deviances was used to compare pairs of models. The model with the highest percentage of deviance explained ($DE\%$) was considered the model that best explains the dataset.

3.3.2 Diagnostic tools for the analysis of residuals

In the analysis of residuals, several diagnostics were used to assess the following theoretical assumptions:

1. Normality of residuals. It was checked for the residuals of all vessels jointly and in the case of mixed models for each vessel separately.
2. Homocedasticity of residuals. It was checked for all vessels jointly and in the case of mixed models within each category of vessel.
3. Normality of random effects in mixed models.
4. Independence between random effects and residuals.
5. Also the relationship between response and fitted values was checked as a visual goodness of fit verification.

and the corresponding diagnostics were:

1. QQ-plots and histograms of residuals and the lilliefors test for normality.
2. Scatterplots of residuals vs linear predictor.
3. QQ-plots of the random coefficients and the lilliefors test.
4. Scatterplots of residuals for each vessel.
5. Scatterplots of response vs fitted values.

In the construction of QQ-plots, the distribution function of residuals, $F(\epsilon)$, was estimated with the empirical cumulative distribution function (ecdf) as

$$\hat{F}(\epsilon_i) = \frac{\text{rank}(\epsilon_i) - 0.5}{n} \quad (3.14)$$

Then, the QQ-plot is the scatter-plot of the collection of points

$$(\epsilon_i, F_0^{-1}(\hat{F}(\epsilon_i))), \text{ with } i = 1, \dots, n \quad (3.15)$$

where, $F_0^{-1}(\cdot)$ is the quantile function of the distribution function F_0 , that in this case is the Normal distribution.

The Lilliefors test was used to check the normality of both residuals and random coefficients, i.e. the null assumption

$$H_0 : x \in N(\mu, \sigma^2) \text{ against } H_1 : x \notin N(\mu, \sigma^2) \quad (3.16)$$

and the test statistic, that is a modified Kolmogorov-Smirnov statistic, is:

$$L = \sup_x | F_n(x) - F_{(\hat{\mu}, \hat{\sigma}^2)}(x) | \quad (3.17)$$

where F_n is the ecdf of the sample vector \mathbf{x} and $F_{(\hat{\mu}, \hat{\sigma}^2)}(x)$ is the distribution function of a normal with mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$. The null hypothesis is rejected for large values of the statistic.

Residuals used in the construction of QQ-plots, were the “deviance residuals”,

$$\hat{\epsilon}_i^d = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i} \quad (3.18)$$

where d_i is the i -th component of the deviance (3.20) contributed by the datum i -th, as suggested by Wood [69].

Also the “standardized” or “pearson” residuals were sometimes implemented in other graphics and analyses (i.e. in the scatterplots of residuals for each vessel). They are defined as

$$\hat{\epsilon}_i^p = \frac{y_i - \mu_i}{\sqrt{V(\hat{\mu}_i)}} \quad (3.19)$$

which have approximate mean 0 and variance ϕ .

3.3.3 Selection criterion

After the diagnostic, the criterion used to compare and select models was the proportion of deviance explained ($DE\%$). The deviance or residual deviance (D_R) of a model is defined as:

$$D_R = 2[l(\hat{\beta}_{max}) - l(\hat{\beta})]\phi \quad (3.20)$$

where $l(\hat{\beta}_{max})$ indicates the maximized likelihood for the saturated model, i.e. the model with one parameter per datum. It represents the highest value that the likelihood could possibly achieve for a given data set. $l(\hat{\beta})$ is the likelihood of the estimated model, while ϕ is the unknown scale parameter. In the practise D_R can be estimated independently from ϕ and that permits the construction of a F test to compare models through their deviances [69]. In the case that response $y \in N(\mu, \sigma)$, the D_R is simply the sum of squared residuals. Then, the deviance explained (DE) by a model is

$$DE = D_N - D_R \quad (3.21)$$

where D_N is the “null deviance”, given by setting the model with a constant, $y_i = \alpha_0 + \epsilon_i$. Finally, the $DE\%$ is simply

$$DE\% = DE \times 100/D_N. \quad (3.22)$$

The F test was performed between pairs of several models as follows. Consider the following test hypothesis

$$H_0 : mod_1 = mod_2 \text{ against } H_1 : mod_1 \neq mod_2 \quad (3.23)$$

under H_0 , the F statistic is:

$$F = \frac{(D_{R1} - D_{R2})/(p_2 - p_1)}{D_{R2}/(n - p_2)} \sim F_{p_2 - p_1, n - p_2} \quad (3.24)$$

where D_{R1} and D_{R2} are the residual deviances of models mod_1 and mod_2 and p_1 and p_2 are the corresponding degrees of freedom and n the total number of observations. If $F > F_{.05}$, the test is significant and H_0 is rejected. The test was built with the aim to check if the degrees in deviance of mod_2 justify the increasing number of parameters. The degree of freedom of a model, p , was setted as given by the sum of effective degrees of freedom of smooth functions plus the number of the fixed parameters. Hence, within models that passed the diagnostic of residuals, the model with highest $DE\%$ and significant F test was selected as the best model.

3.3.4 Comparison between estimations

Then the selected best model was fitted applying the empirical Bayesian REML and the full Bayesian methods using the software BayesX [5].

The same initial conditions, when possible, were set for the three procedures:

1. P-spline¹ as bases for all smooth functions;
2. the same number of initial knots for each smoother: $nk_{idmonth} = 20$, $nk_{trips} = 10$, $nk_{nao3} = 5$.

Then, the three estimations were compared using the estimate mean squared error, MSE:

$$\widehat{MSE} = \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n} \quad (3.25)$$

where $\hat{\epsilon}$ are the response residuals, i.e. $(y_i - \hat{y}_i)$, n the number of observations. The estimation with the lowest value of \widehat{MSE} was considered the best estimation.

¹cubic penalized spline with a B-spline basis and 2-nd order difference penalty

3.4 Results and discussion

3.4.1 Diagnostics of model assumptions

Basic diagnostic plots of the AMM 3.10 (Figures 3.11, 3.12 and 3.13) were selected as example of residual behaviour recorded for all models with Gaussian response assumption, i.e. the LM, the LMM, the AM and obviously the AMM.

Residuals of all these models produced almost the same patterns. These are:

- The QQ-plot of residuals shows an evident departure of residual quantiles from the theoretical normal quantiles and a pronounced curvature, concave upwards, related to a pronounced asymmetry observable in the histogram of residuals as well (upper and lower left panels in Figure 3.11).
- The scatter-plot of residuals versus the linear predictor shows an accentuated heteroscedasticity with a notable increasing of variance for higher values of the linear predictor (upper right panel in Figure 3.11).
- The same scatterplot shows another problematic behaviour of residuals, that is, some kind of trend evidenced by the grouping of negative residuals for lower values of the linear predictor.
- The response versus fitted values shows an increasing badness in the fit while accompanying the response (lower right panel of the same figure).
- The remaining two figures show residuals behaviour inside each group of variability, scatterplots allow to highlight in which vessel residuals are more dispersed or concentrated and for which vessel residuals clearly follows the same trend (see e.g. the upper and the middle plots on the left 3.12).
- The information proceeding from this graphics allows sometimes the construction of more general models where categories of random effects are grouped into more general categories.
- Moreover F_n of residuals within each vessel shows the same pattern of the residual joint sample Figure 3.13.
- Finally the QQ-plot of random coefficients is the unique valid diagnostic plot of this model (3.23).

The diagnostic graphics for the GLMM show mainly the same, including somewhat accentuated, failures (see e.g. the QQ-plot of joint residuals in Figure 3.14). Comments are omitted, being figures of immediate interpretation (3.14 and 3.15). The QQ-plots per vessel were omitted having exactly the same behaviour than the AMM's residuals. On the contrary the QQ-plot of random coefficients showed a good approximation to the normal distribution (graphic was omitted as well).

The diagnostic plots of all remaining models, GLM, GAM, GMM, GAMM-AR(1) and GAMM-MA(1) show evident gains with respect to the models with Gaussian assumption of the response. That is mainly the gain given by strike with the Gamma assumption for the response. Furthermore they show increasing gain moving through the GLM, the GAM and finally to GAMMs (both with and without correlation structures). In fact, the sequential observation of Figures 3.16, 3.17, 3.18, 3.19 and 3.20, highlights a rearrangement of residuals versus a more homogeneous variability in the scatter-plots against the linear predictor (upper right panel of figures). Also the QQ-plots in the upper left panels of all figures show better trends, however in the fixed design models the residuals look like more normal distributed. The scatter-plots of residuals for each vessel (only the plots for GAMM are presented, Figure 3.21) show residuals centered and homogeneously distributed around the zero, without drastically alarming trends. However, the within-group variability changes considerably among vessels. Finally, both residual per vessel and random coefficients quantiles show a reasonable approximation to the normal, however less evident in both residuals and random coefficients of the GAMM-MA(1) (Figures 3.22 and 3.23). The Lilliefors's test (3.17) gave significant results for residuals of all models (Table 3.3). That means rejection of the normality assumption, but it is not alarming because of the test for large sample sizes detects each minimum difference between the ecdf and the theoretical distribution as significant. On the contrary, the test performed on the coefficients was not significant in all cases (Table 3.3).

In conclusion, data clearly violate the assumptions of normality and independence, as it was easy to suppose from the exploratory analysis. Thus, Gaussian models were no longer considered. The GLMM gave clearly bad diagnostics and was set aside as well. In contrast, analysis of residuals can be reasonably accepted in the case of remaining models, giving some better diagnostics regarding to trends in the residuals in the case of GAMMs and regarding to the normality assumptions in the case of fixed effects models. The following section is devoted to a deeper comparison of remaining models.

Table 3.3: Lilliefors test for normality of residuals and random effects. For each model the statistic L and the corresponding p – value are given for residuals (resid) and for random effects (random) respectively. The last one only for mixed models.

model	L (resid)	p – value	L (random)	p – value
LM	0.065	$\leq 2e-16$		
LMM	0.066	$\leq 2e-16$	0.146	0.279
GLM	0.073	$\leq 2e-16$		
GLMM	0.073	$\leq 2e-16$	0.185	0.058
AM	0.065	$\leq 2e-16$		
AMM	0.045	$\leq 2e-16$	0.109	0.738
GAM	0.048	$\leq 2e-16$		
GAMM	0.075	$\leq 2e-16$	0.162	0.159
GAMM-MA1	0.062	$\leq 2e-16$	0.144	0.307
GAMM-AR1	0.057	$\leq 2e-16$	0.166	0.137

3.4.2 Deviance and F test

The results about the deviance of the remaining models are presented in Table 3.4. All tests with p – values < 0.05 were significant. The models were ordered from lower to higher $DE\%$ for an easier interpretation. The GAM was the model with higher deviance explained, while the GLM with the lower. The GLM was significantly different from the Null model as well as the GAMM–AR(1). Hence, succeeding models, with higher p and lower DE are significantly different from the Null model, without need to calculate the F statistic. The GAMM–AR(1) does not apport improvement w.r.t. the GLM, neither the GAMM–MA(1), neither the GAMM, while the GAM was significantly different from the GLM. The GAM was also significantly better than the GAMM–AR(1) and GAMM–MA(1). The test between GAM and GAMM produced obviously NA due to that statistic F cannot be negative, but the fact to have both less df and DE , is a sufficient proof of its explanatory potential, confirmed by its $DE\%$. Finally, within mixed additive models, results show that the GAMM has the best explanatory potential ($DE\% = 28.5$), but it also presents the highest number of parameters.

The results of deviance regarding to GAMMs with higher orders p and q of ARMA, showed increasing values of D_R , thus they were discarded from the analysis, even if they presented significant added parameters. These results about autoregressive and moving average incorporation were somewhat surprising, because ACF and PACF of data presented evidense of these

Table 3.4: Deviance explained and F tests. mod_1 and mod_2 as in the statistic formula 3.24. Null is the null deviance, with $y \in Gamma$. p =degree of freedom of mod_2 , D_R =residual deviance, DE =deviance explained, $DE\%$ =proportion of DE , F =F statistic, $p-value$ =p-value of the statistic.

mod_2	p	D_R	DE	$DE\%$	mod_1	F-val	p-val
Null	2	1446					
GLM	9	1081	364	25.2	Null	18.8	$\leq 2e-16$
GAMM-AR1	24.8	1054	391	27.1	Null	6.1	$2e-15$
					GLM	0.6	0.87
GAMM-MA1	25.6	1046	400	27.7	GLM	0.7	0.79
					GAMM-AR1	3.6	$7e10-02$
GAMM	27.2	1033	413	28.5	GLM	0.9	0.51
					GAMM-AR1	2.7	$6e10-02$
					GAMM-MA1	2.9	$7e-02$
GAM	27.1	922	524	36.2	GLM	3.5	$26e-06$
					GAMM-AR1	3.4	$1e-05$
					GAMM-MA1	33.4	$2e-11$
					GAMM	<i>NA</i>	<i>NA</i>

structure. At the evidence of results, the dependence structure is negligible.

In conclusion, the GAM resulted the best model according to this criterion of selection. And in fact, if the main interest is to describe the particular situation of *cpue* only for the sampled vessel, thus, the GAM remain the best model. In other words, if there is any intention in make estrapolations, but just only describe the particular dataset, GAM is the best choice. But it is razonable to suppose that vessel effect changes in time (e.g. with improving tecnology) or in space. Thus, it must be clear that vessel levels of the entire population can not be defined a priori, can not be “predictable” by the scientist, simply because they are not available (e.g. there are not information about vessel in the future). Many times throughout the manuscript, the importance of considering the effect of vessel as random was pointed out. As a result, it is in the context of mixed models that data must be modelled. Thus, in the spectrum of possibilities used to model red shrimp’s data, the best model within mixed models, must be selected. And the best, in this case, is the GAMM, that presented the best explanatory potential ($DE\% = 28.5$) within this class of models.

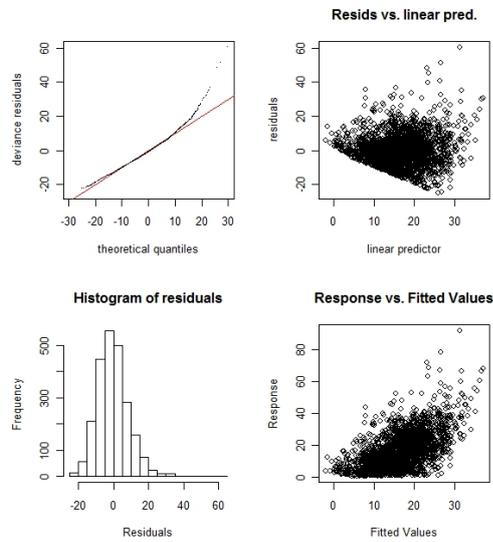


Figure 3.11: Diagnostic for Model AMM. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

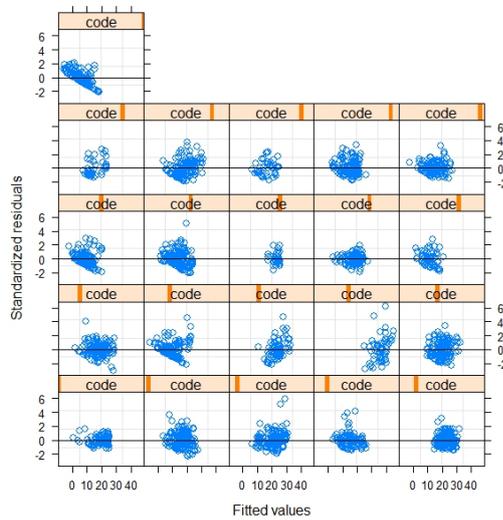


Figure 3.12: Scatterplots of standardized residuals vs. fitted values of the AMM for each level of code.

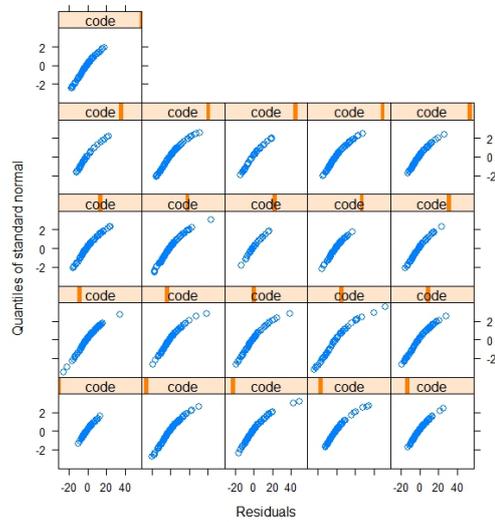


Figure 3.13: QQ-plots for normality of the residuals of the AMM for each level of code.

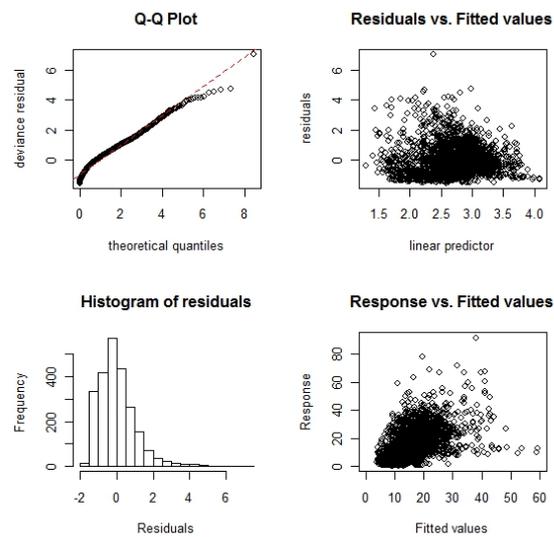


Figure 3.14: Diagnostic for Model GLMM. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

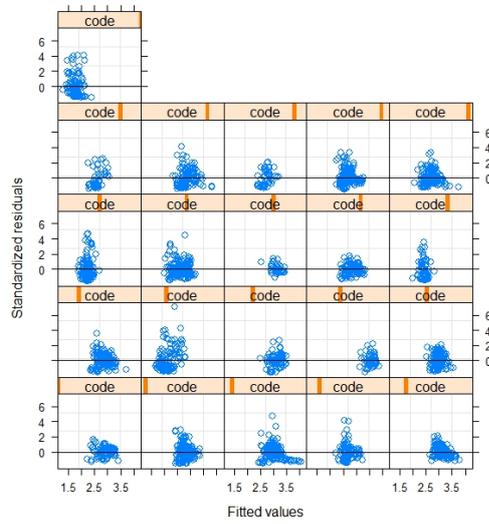


Figure 3.15: Scatterplots of standardized residuals vs. fitted values of the GLMM for each level of code.

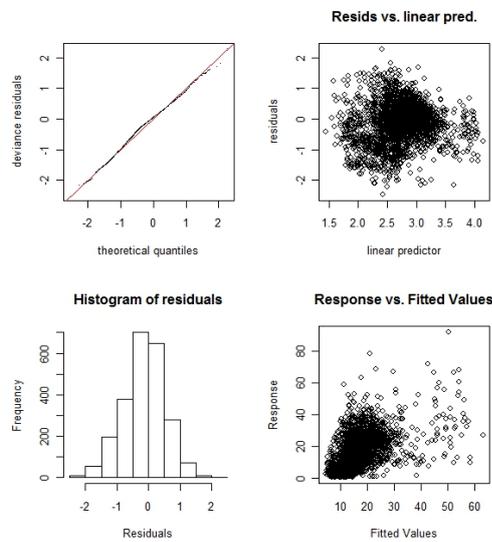


Figure 3.16: Diagnostic for Model GLM. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

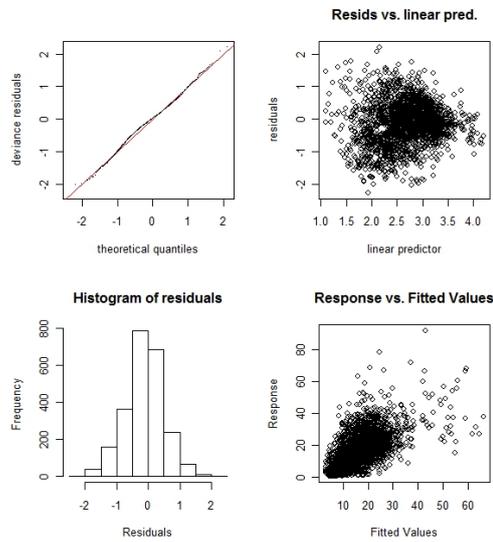


Figure 3.17: Diagnostic for Model GAM. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

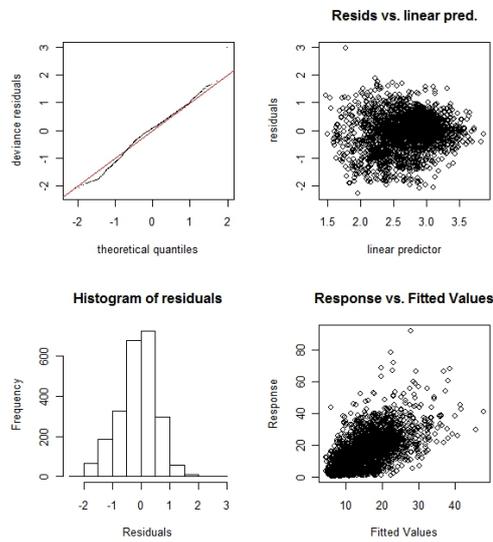


Figure 3.18: Diagnostic for Model GAMM. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

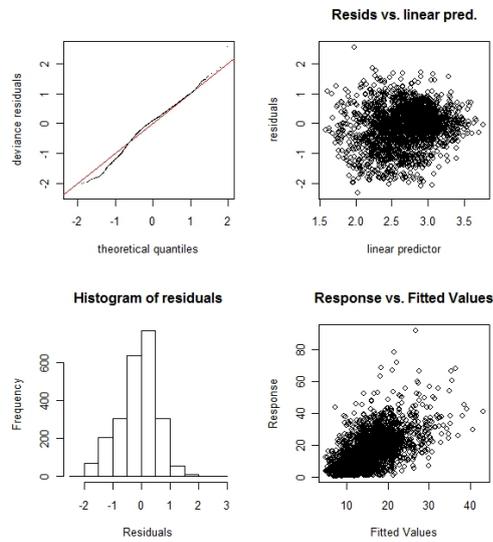


Figure 3.19: Diagnostic for Model GAMM-AR1. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

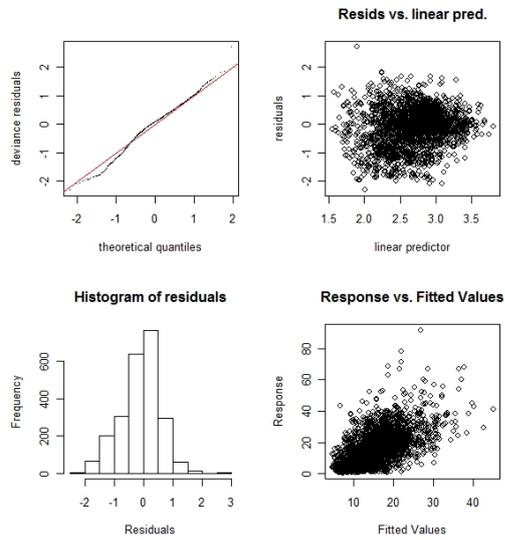


Figure 3.20: Diagnostic for Model GAMM-MA1. Clockwise from top left: 1) QQ plot of residuals, 2) plot of residuals vs. linear predictor, 3) histogram of the residuals and 4) plot of the response vs. fitted values.

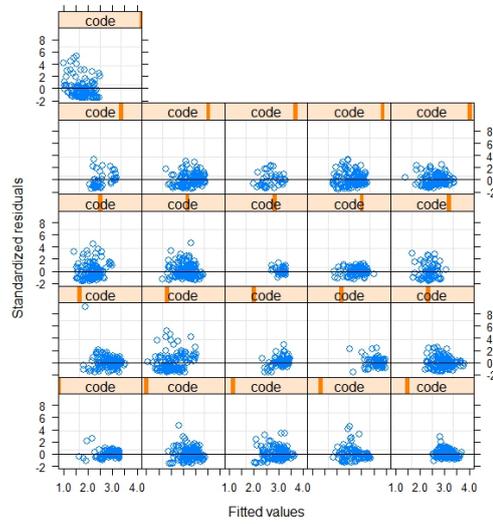


Figure 3.21: Scatter-plots of standardized residuals vs. fitted values of the GAMM for each level of code.

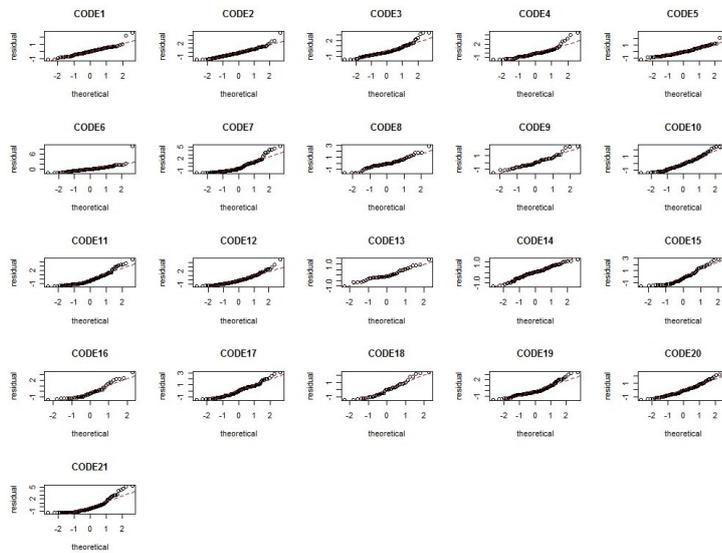


Figure 3.22: QQ-plots for normality of the residuals of the GAMM for all levels of code.

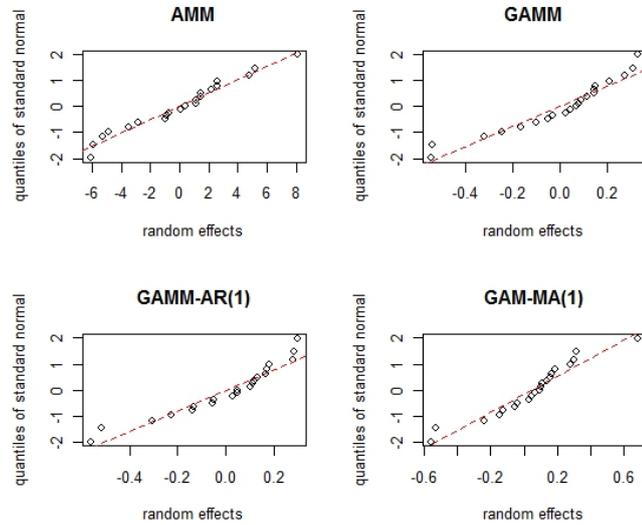


Figure 3.23: QQ-plot for normality of the random effects of all mixed effects models.

3.4.3 Frequentist, empirical Bayesian and full Bayesian estimations

This section is focused on the comparison of the three different procedures at date available for estimating generalized additive mixed models. The model estimated with the three approaches was the model selected at the end in section 3.4.2, i.e. the GAMM.

A compendium of results of each procedure is reported in Tables 3.5, 3.6, 3.7 and the partial effects of predictors are displayed in Figures 3.25, 3.26, 3.27 respectively for the frequentist, the empirical Bayesian and the full Bayesian approaches. In the following section some biological interpretation of results are given.

In the Tables are reported the estimations of: fixed effects, smoothing terms, the two variance estimation and the estimated mean squared error, \widehat{MSE} , calculated as in 3.25. The inference procedure that gave the smallest mean squared error was the empirical Bayesian ($\widehat{MSE} = 71.83$), followed in order by the full Bayesian ($\widehat{MSE} = 73.41$) and at the end by the frequentist ($\widehat{MSE} = 82.71$).

Though, paying attention on the various parameter estimated, the three approaches gave quite similar results as can be observed by comparing esti-

mations of the fixed coefficients and of scale parameters of each model. As an example, Figure 3.24 shows three different boxplots, where, in each of them, the three estimations (“fB”=full Bayesian, “eB”=empirical Bayesian, “fr”=frequentist) of the three fixed parameters (in order “coef const”, “coef month2”, “coef trb”) are represented. It is evident that the coefficient intervals overlapped.

Also comparing graphically the smooth partial effects in figures 3.25, 3.26, 3.27, no substantial differences can be appreciated.

Table 3.5: Estimations for the GAMM using the frequentist inference.

FIXED	EFFECTS			
	mean	std	t-val	p-val
<i>const</i>	2.22	0.17	13.17	$\leq 2e-16$
<i>grt</i>	0.01	0.003	3.34	8.6e-04
<i>month2</i>	-0.17	0.03	-4.98	6.64e-07
SMOOTH	TERMS			
	$\hat{\lambda}$	<i>df</i>	F-val	p-val
<i>s(trips)</i>	0.35	4.12	87.42	$\leq 2e-16$
<i>s(nao₃)</i>	0.04	2.08	5.18	5.07e-03
<i>s(idmonth)</i>	0.02	16.17	16.86	$\leq 2e-16$
GLOBAL	ESTIM			
$\widehat{MSE} = 82.71$	$\hat{\sigma}^2 = 0.35$		$\hat{\sigma}_b^2 = 0.06$	

Table 3.6: Estimations for the GAMM using the empirical Bayesian inference.

FIXED	EFFECTS		
	mean	std	p-val
<i>const</i>	2.37	0.30	1.83e-11
<i>grt</i>	0.01	0.003	1.7e-03
<i>month2</i>	-0.17	0.03	1.37e-05
SMOOTH	TERMS		
	$\hat{\lambda}$	<i>df</i>	$\hat{\theta}$
<i>s(trips)</i>	11.09	4.50	0.03
<i>s(nao₃)</i>	21.31	2.23	0.02
<i>s(idmonth)</i>	0.50	15.93	0.72
GLOBAL	ESTIM		
$\widehat{MSE} = 71.83$	$\hat{\sigma}^2 = 0.36$	$\hat{\sigma}_b^2 = 0.07$	

Table 3.7: Estimations for the GAMM using the full Bayesian inference.

FIXED	EFFECTS		
	mean	std	p-val
<i>const</i>	2.37	0.20	1.83e-11
<i>grt</i>	0.009	0.003	1.38e-05
<i>month2</i>	-0.17	0.04	0.002
SMOOTH	TERMS		
	$\hat{\lambda}$	$\hat{\theta}$	
<i>s(trips)</i>	14.00	0.05	
<i>s(nao₃)</i>	52.71	0.07	
<i>s(idmonth)</i>	0.54	0.83	
GLOBAL	ESTIM		
$\widehat{MSE} = 73.41$	$\hat{\sigma}^2 = 0.36$	$\hat{\sigma}_b^2 = 0.08$	

3.5 Interpretation of results

In this study for the first time vessel characteristics (*grt*), fishing intensity (*trips*) and environmental incidences (*nao₃*) were jointed in one global model with temporal variability (*idmonth*, *month*) to predict catch index fluctuations. Furthermore, for the first time generalized additive mixed model (GAMM) was used for this purpose, being data obtained from a multi-vessel

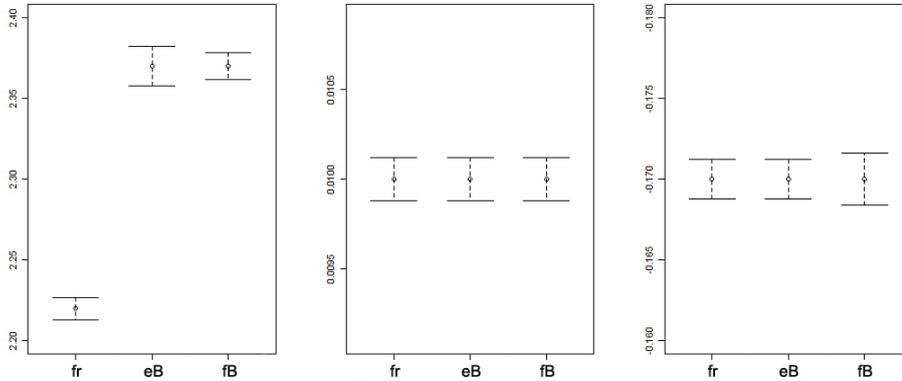


Figure 3.24: Comparison of fixed effects estimations. Confidence intervals of fixed parameters estimated by the three methods (from the left to the right panels: *constant*, *trb* and *month*). Labels indicates: “fr”=frequentist, “fB”=full, Bayesian, “eB”=empirical Bayesian.

trawl fishery context. In the past few studies proposed mixed models (in particular GLMM), which take into account vessel as a random effect [12], but, we found that linear mixed model are not flexible enough to describe complicated effects such those observed for some variables (e.g. *idmonth*).

The model here proposed confirmed that all variables, often considered separately (fishery fleet effects: [43], [60]; environmental effects: [45]) synergistically influence the CPUE fluctuations. In previous studies, models were built by means of GLM or descriptively and did not capture the complex relationships existing between each explanatory variable and the response, while the model here proposed identified many non linear relationships.

The NAO (with a lag of 3 years before the year of observed response, nao_3), has a non linear influence indicating that CPUE increases, with NAO increasing for negative values of that variable, while it reaches a maximum for $NAO \approx 0.2$ keeping constant for higher values (we can say that it reach a saturation). Thus, values higher than $NAO = 0.2$ does not seem to influence CPUE. Note that for high values of NAO, the curve of the partial effect tends to decrease, but the large credible bands of that region do not permit to consider that curvature significant.

Maynou [45] found that nao_2 influences Barcelona landings more than nao_3 , however here we found that is nao_3 the subset that most influences CPUE. It probably could be due to the different period observed in the two studies (larger in this study). Furthermore, CPUE depends on both datasets

of nao ($lag = 2$ or 3), sometimes more on the former or the latter according to the dominant length class of the species captured (2 or 3 years length class) [45].

The partial of monthly number of trips ($trips$) also shows a non linear effect on the CPUE: the two variables are positively related and the rate of the curvature decreases for increasing values of the covariate. That occurs because the CPUE, such as it is actually calculated [43] does not take into account the difference in the number of trips (the average is not a weighted average). The non linear effect observed is probably due to both the increasing fishing effort, which provides higher catches and to the experience of fishermen (higher effort is related to vessels which are specialized in the deep sea fishery, thus they reach higher values of CPUE for $trips \approx 15-25$. In other words, experienced fishermen reach higher catches than not experienced, as observed by Maynou et al. ([43]).

The partial effect of the time ($idmonth$) also shows an evident non linear relationship with different relative maxima and minima along the years. Figure 3.25 indicates that the CPUE generally increase then decrease along the time observed, reaching minima after about 8-10 years, as was commented in [45]. Note that this effect incorporates between years variability of the CPUE, while losing the within years variability (that allows the introduction of month as categorical variable, such as it was done). The absolute minimum was observed in time around of $idmonth = 90 - 110$ (which correspond to year 1999-2000). That low CPUE observed was probably due to the low nao_3 observed in preceding months (3.9). These years of low CPUE encouraged fishermen to make more trips to catch as more as possible. This condition explains the high effort observed in these years and that was apparently in contrast with the marginal of $trips$ (see Figure 3.7).

The $idmonth$ effect did not evidence intra-annual variation of red shrimp CPUE, which was better captured by the categorical variable $month$. In figure 3.25 effect of monthly variation results significantly lower for June and November grouped in level 2 of the variable $month$. This result could be related with the reproduction behaviour of *A. antennatus*. The model probably mask partially the effect of month variability due to the high number of explanatory variables Results about grt marginal effect suggest that bigger boats obtain higher CPUE in this offshore fishery.

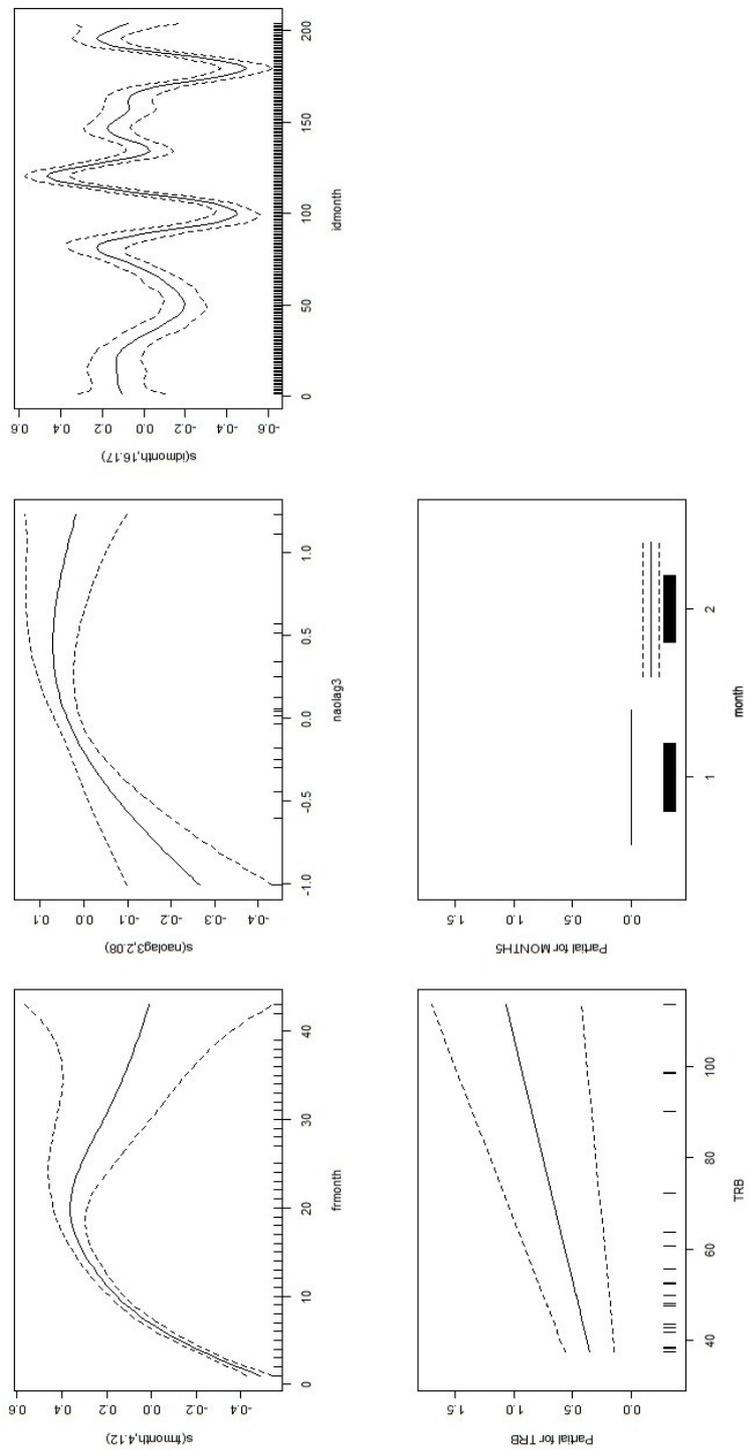


Figure 3.25: Partial effects of the covariates using frequentist REML inference. 95% Bayesian credible intervals.

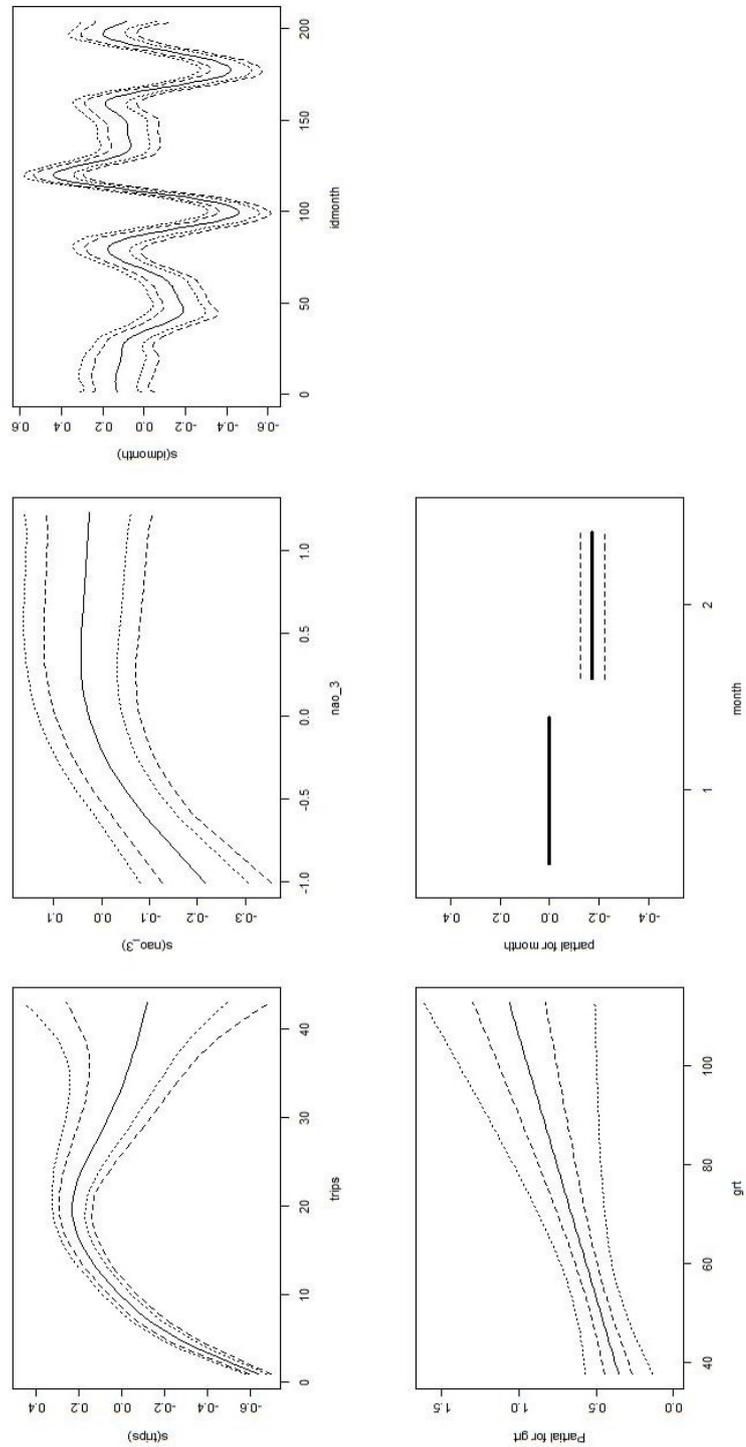


Figure 3.26: Partial effects of the covariates using empirical Bayesian REML inference. 80% and 95% Bayesian credible intervals.

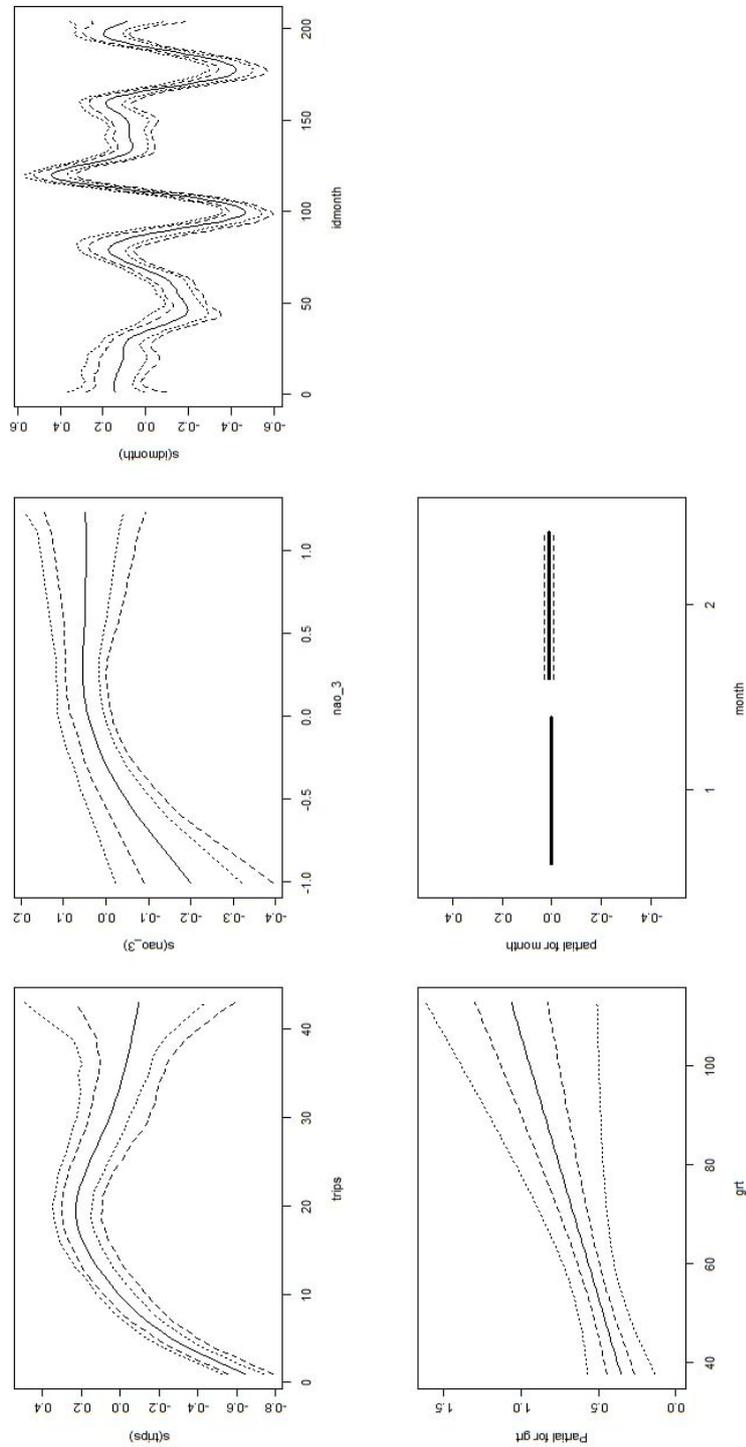


Figure 3.27: Partial effects of the covariates using the full Bayesian inference. 80% and 95% Bayesian credible intervales.

Chapter 4

Software

4.1 The R package `mgcv`

R library `mgcv` allows to fit GAMMs using the function `gamm`, by a call to the library `lme` in the case of normal response and identity link, or by a call to `gammPQL` otherwise. Theoretical motivations of that settings are given in Section 2.3. `gammPQL` is a modified version of `glmmPQL` and in that case the estimates are only approximated. All fixed effects and smooths are specified as in a call to `gam` as part of the fixed effects model formula, but the “wiggly” components of the smooth are treated as random effects (that is the representation discussed in Chapter 2, Section 2.3.1). The random effects structures available through `lme` are used to specify other random effects. It is also assumed that the prime interest is in inference about the terms in the fixed effects model formula including the smooths. For this reason the routine calculates a posterior covariance matrix for the coefficients of all the terms in the fixed effects formula, including the smooths.

Other functions sometimes useful in constuctiong mixed models are `mle` and `glmmPQL` from the libraries `nlme` and `MASS` respectively. These functions can be used in alternative to `gam` in construction of LMM and GLMM models. Although, in theory `gamm` comprises all the models presented in this study, it often crashes, probably for computational instability. Thus, also in the present study, for models LMM and GLMM, `mle` and `glmmPQL` were implemented.

4.1.1 R syntax

First load required libraries and dataframe:

```

> library(nlme)

> library(mgcv)

> data<-read.table("C://dots data.txt",header=T)

```

The construction of model 3.12 is reported below.

Before fitting model, ensure that the variable *month* is read as factor, if it is not, type:

```

> MONTH=factor(data$month)

```

Otherwise, intercept and slope parameters are estimated rather than deviations from the mean of the reference level! Note that in R the first level is chosen as reference level.

`gamm` formula is quite similar to `gam`:

```

> fit<-gamm(cpue ~ s(trips,bs="ps",k=10)+grt+
+ s(nao3,bs="ps",k=5)+s(idmonth,bs="ps",k=20)+
+ MONTH,random=list(code= 1),gamma=1.4,
+ family=Gamma(link="log"), data=data,method=REML,
+ niterPQL=100)

```

Response variable is in the left side of the symbol \sim , that means equal to in the formula, while all predictors and options are in the right part after \sim . All fixed terms can be smooth, parametric or categorical. Random effects can be put in the model using the argument `random`. In the simple case of no nested random effects the code is as in the example above.

Smoothing terms, `s` or `te` (the latter specifies the tensor product, that allows the construction of smooths of more than one covariate), are implemented as in `gam`. In present analysis, only two parameters were explicitly controlled: `bs` and `k`:

`bs` indicates the smoothing basis to use. (e.g. `tp` for thin plate regression spline, `ps` for penalized spline). Typing `?smooth.terms`, it shows which options are available. For example, regarding to `ps`: a univariate P-spline as

proposed by Eilers and Marx (1996), is used. This command allows to use B-spline bases penalized by discrete penalties applied directly to the basis coefficients. These bases can be also implemented in tensor product smooths (see `te`).

`k` is the dimension of the basis used to represent the smooth term. The default depends on the number of variables that the smooth is a function of. `k` should not be less than the dimension of the null space of the penalty for the term (see `null.space.dimension`), but will be reset if it is.

The computation of a GAMM within the `gamm` formula returns a list with two items:

`gam`: an object of class `gam`, less information relating to GCV/UBRE model selection. It allows to use `predict`, `summary` and `print` methods and `vis.gam`, but not to use e.g. the “`anova`” method function to compare models.

`lme`: the fitted model object returned by `lme` or `gammPQL`. This output may appear to be rather difficult to understand, because of the manner in which the GAMM is split up and the calls to `lme` and `gammPQL` are constructed. This output contains:

- information of the technique used for estimation
- AIC, BIC and Log-likelihood value
- information on the random components of smooth functions
- information on random effects
- information on fixed effects and their correlation

Thus, we can print the results of the model in the following way:

```
> summary(fit$gam)
```

```
> summary(fit$lme)
```

The output of `summary(fit$lme)` is given in Figure 4.1. This output gives the information of the random components of smoothing functions. Each smooth starts with 10, 5 and 20 dimensionality of basis, but one of these is lost to the GAM centering constraint (that ensure the additive identifiability) and one is treated as a fixed effect [69], so each smooth has

```

Linear mixed-effects model fit by maximum likelihood
Data: data
      AIC      BIC    logLik
4441.825 4562.866 -2199.912

Random effects:
Formula: ~Xr.1 - 1 | g.1
Structure: pdIdnot
      Xr.11  Xr.12  Xr.13  Xr.14  Xr.15  Xr.16  Xr.17  Xr.18
StdDev: 1.022712 1.022712 1.022712 1.022712 1.022712 1.022712 1.022712 1.022712

Formula: ~Xr.2 - 1 | g.2 %in% g.1
Structure: pdIdnot
      Xr.21  Xr.22  Xr.23
StdDev: 3.374137 3.374137 3.374137

Formula: ~Xr.3 - 1 | g.3 %in% g.2 %in% g.1
Structure: pdIdnot
      Xr.31  Xr.32  Xr.33  Xr.34  Xr.35  Xr.36  Xr.37  Xr.38  Xr.39
StdDev: 6.414079 6.414079 6.414079 6.414079 6.414079 6.414079 6.414079 6.414079 6.414079
      Xr.310  Xr.311  Xr.312  Xr.313
StdDev: 6.414079 6.414079 6.414079 6.414079

Formula: ~1 | code %in% g.3 %in% g.2 %in% g.1
      (Intercept) Residual
StdDev: 0.2463403 0.5977506

Variance function:
Structure: fixed weights
Formula: ~invwt
Fixed effects: list(fixed)
      Value Std.Error DF t-value p-value
X(Intercept) 2.2450242 0.1727060 2319 12.999106 0.0000
XTRB 0.0093838 0.0028002 19 3.351093 0.0034
XMONTH2 0.0078888 0.0577885 2319 0.136512 0.8914
XMONTH3 0.0142683 0.0580811 2319 0.245662 0.8060
XMONTH4 0.0615503 0.0579352 2319 1.062399 0.2882
XMONTH5 -0.0679558 0.0593098 2319 -1.145777 0.2520
XMONTH6 -0.2420252 0.0608291 2319 -3.978776 0.0001
XMONTH7 -0.0842715 0.0606516 2319 -1.389436 0.1648
XMONTH8 -0.0381056 0.0603671 2319 -0.631231 0.5280
XMONTH9 0.0137026 0.0629035 2319 0.217835 0.8276
XMONTH10 -0.0411533 0.0619570 2319 -0.664223 0.5066
XMONTH11 -0.1448167 0.0613953 2319 -2.358759 0.0184
XMONTH12 -0.0775308 0.0583213 2319 -1.329374 0.1839
Xs (Ermonth)Fxl -0.5186768 0.4703666 2319 -1.102707 0.2703
Xs (naolag3)Fxl 0.5945308 0.2968637 2319 2.002706 0.0453
Xs (idmonth)Fxl -0.3735845 0.5854150 2319 -0.638153 0.5234
Correlation:
      X(Int) XTRB XMONTH2 XMONTH3 XMONTH4 XMONTH5 XMONTH6 XMONTH7 XMONTH8 XMONTH9
XTRB -0.919
XMONTH2 -0.166 -0.003
XMONTH3 -0.165 -0.005 0.503
XMONTH4 -0.171 0.001 0.502 0.501
XMONTH5 -0.164 -0.002 0.492 0.494 0.494
XMONTH6 -0.166 0.004 0.478 0.478 0.483 0.474
XMONTH7 -0.164 0.001 0.481 0.484 0.484 0.480 0.465
XMONTH8 -0.163 -0.002 0.482 0.484 0.487 0.481 0.468 0.473
XMONTH9 -0.159 0.001 0.463 0.464 0.468 0.459 0.449 0.453 0.458
XMONTH10 -0.163 0.003 0.469 0.470 0.475 0.466 0.457 0.461 0.465 0.448
XMONTH11 -0.163 0.003 0.474 0.475 0.478 0.467 0.459 0.461 0.466 0.450
XMONTH12 -0.174 0.004 0.499 0.500 0.504 0.492 0.482 0.486 0.489 0.471
Xs (Ermonth)Fxl 0.000 -0.003 0.016 0.023 -0.005 0.020 0.006 0.029 0.005 -0.002
Xs (naolag3)Fxl -0.001 0.001 -0.001 0.000 -0.002 0.006 0.004 0.001 0.000 -0.002
Xs (idmonth)Fxl 0.025 -0.007 -0.009 -0.024 -0.035 -0.040 -0.051 -0.059 -0.069 -0.074
      XMONTH10 XMONTH11 XMONTH12 Xs (f)F1 X(3)F1
XTRB
XMONTH2
XMONTH3
XMONTH4
XMONTH5
XMONTH6
XMONTH7
XMONTH8

```

Figure 4.1: Output of the code `summary(fit$lme)` for the frequentist model in R.

two less number of random coefficients. Each of these coefficients is treated as having the same variance, after re-parameterizations, but this variance (which plays the role of smoothing parameter) is unknown and is estimated. From g.1 to g.3 are dummy grouping variables each having only one level. They force the smooths to apply to all data. What follows in the same output is information about fixed effects. In this case there are: intercept, effects of TRB, MONTH levels and all smoothers.

The following syntax produces some diagnostic informations about the fitting procedure and the results:

```
> gam.check(fit$gam)
```

In particular it plots four standard diagnostics plots, such as those in Figure 3.18, and some other convergence information.

Some diagnostics for the random effect part of the mixed model are available typing:

```
> plot( fit$lme, form = resid(., type = "p") fitted(.)|code,
abline = 0 )
```

stand residuals vs fitted values for heach code.

```
> qqnorm(fit$lme, resid(.)|code)
```

to check the normality assumption separatly for residuals within each code.

To check the normality of random effects, the following syntax is available, but it works well only for lme and glmmPQL object, while it does not work for gamm\$lme, for which is necessary to programm directly the qq-plots.

```
> qqnorm( fit6LMMlme, ranef(.))
```

The syntax to display partial effects is:

```
> plot(fit$gam,all.terms=T, pages=1,scale=0)
```

4.1.2 Syntax implemented for the example in Section 1.1

```
#Boxplot of cpue vs levels of vessel

> boxplot(data$cpue ~ factor(code),xlab="code",ylab="cpue")

#construction of models

> m0<-lm(cpue ~ 1 , data = data ) #simple mean model

> m1<-lm(cpue ~ factor(code)-1, data = data ) #fixed effects model

#The -1 is used in the model formula to prevent the default inclusion of
an intercept term in the model.
# More clearly: in this way the reference value in taeting significantness of
heach term of code is the average of the whole set of observations, otherwise
the reference value becomes the average within observations of the 1-st level
of the factor.

> m2<-lme(cpue ~ 1, data = data, random = ~ 1 | code) #random
effects model

#outputs

> summary()

#diagnostics

> plot()

#example of boxplot of residuals: residuals vs code in the SIMPLE-mean
model

> boxplot(residuals(m0) ~ factor(code),xlab="code",ylab="residuals")

#alternatively:

> res=data$cpue-m0$coefficients[1]
```

```
> boxplot(res ~ factor(code), xlab="code", ylab="residuals")
```

4.2 BayesX

Currently, BayesX is available in two different versions. The first one is intended for the various versions of the Windows operating system and includes a graphical user interface that enables visualization of estimated effects. While the computational kernel of BayesX has been implemented in C++, the graphical user interface has been implemented in Java. As a supplement to both versions, an additional R package, *BayesX*, is available from CRAN (<http://www.r-project.org>). While this package does not provide direct access to BayesX from within R at the moment (this is planned for the future), it provides additional visualization routines. The current releases of both BayesX versions can be downloaded from <http://www.stat.uni-muenchen.de/~bayesx>. Manuals are also available in the same website.

4.2.1 BayesX syntax

```
# Read data from an ASCII file:
```

Before uploading dataframes, create the object where it be load. Thus, e.g., create the data set object *shrimps*:

```
> dataset shrimps
```

and fill the created object of data:

```
> shrimps.infile using C:/ ...
```

It is also possible to print the data set:

```
> shrimps.describe
```

Note that in the ASCII file, variable names and observations must be separated by blanks or tabs. If missing value is present, only "." or "NA" signs are read by default. Otherwise specify the sign used in the missing option in the infile method.

The empirical bayesian regression syntax:

Create the remlreg object *fit1*:

```
> remlreg fit1
```

Create the path in which to save the results of the regression model:

```
> fit1.outfile = C:/ ...
```

Create the folder of results before typing the command and at the end of the path write which will be the name of the output files.

The formula of the model:

```
> fit1.regress cpue = time(psplinerw2,nrknots=15)+  
nao3(psplinerw2,nrknots=10)+fr(psplinerw2,nrknots=10)+  
code(random)+grt,family=gamma using shrimps
```

That command allows to fit the regression model: *fit1* is the name of the model, *regress* is the method applied, *psplinerw2* is used for smooth terms to apply p-splines bases with penalization on the second derivative (rw2), *nrknots* is the number of knots

Quite all information are printed in the output window.

Full bayesian regression syntax:

With the same logic is constructed the completely bayesian model:

```
> bayesreg fit2
```

```
> fit2.outfile = C:/...
```

```
> fit2.regress cpue=time(psplinerw2,nrknots=15)+  
nao3(psplinerw2,nrknots=10)+fr(psplinerw2,nrknots=10)+  
code(random)+grt,family=gamma using shrimps
```

here the only difference from the first code is the object type, *bayesreg*.

```
# Visualizing estimation results:
```

After estimating model plots of non parametric effects can be plotted in BayesX. Plotting parametric effects is not allowed by the program, but it can be done importing results in the R environment (see below for the implementation of the R-packege BayesX).

```
> fit1.plotnonp
```

```
> fit2.plotnonp
```

```
# Visualizing results with R:
```

```
> library("BayesX")
```

```
> plotnonp("C:/.../results_idmonth_pspline.res",ylab="s(idmonth)")
```

where many graphical settings can be changed.

Chapter 5

Conclusions

In the study brought up here, a mixed effects approach was proposed to model the CPUE of the red shrimp within the framework of the Generalized Additive Mixed Models (GAMMs). The main purposes were:

1. Propose a new form to model data proceeding from the fishery;
2. Investigate the ability of GAMMs to flexibly model real data;
3. Investigate and compare the methodologies currently in use to estimate such models, which follow in turn, the frequentist, the Bayesian and the empirical Bayesian perspectives;
4. Define a brief guideline of the syntax required to construct such models in the environments of two software to date available.

The first and second items were carried out comparing a wide class of models (from LM to GAMM with autocorrelation structures).

The third item was achieved comparing the three estimators using the mean squared error, MSE.

About the last purpose Chapter 4 explains the most important steps in the construction of generalized additive mixed models.

Mixed model was considered the proper model design to deal with the application study. In section 1.1, the importance of considering effects as random rather than fixed was explained. The mixed model considers as source of variance both error and random effects variance, that permits extrapolations of results to the entire population.

A series of different models with increasing complexity was constructed, from the LM to the GAMM with the incorporation of autoregressive and moving

average structures. Moreover GAMMs has the ability to partially eliminate undesirable trends in residuals however it did not always deal with the application study .

The comparison between bayesian versus frequentist techniques, detected better behaviours (in terms of the lowest MSE estimated) of the Bayesian in comparison to the frequentist approach. The comparison does not give significant different results in the estimation of the parameters, whose intervals overlapped (except to the constant).

Bibliography

- [1] Akaike, H., 1973. Information theory and an extension of the maximum-likelihood principle. Second International Symposium on Information Theory, Ed. B. F. Petrov and F. Csiki, Akadémiai Kiadó, Budapest.
- [2] Bas, C., Maynou, F., Sardà, F., Lleonart, J., 2003. Variacions demogràfiques a les poblacions d'espècies de Marsals explotades els darrers quaranta anys a Blanes i Barcelona. Inst. Est. Catalans. Arxiu de la Sec. Ciències. CXXXV. 202p.
- [3] Breslow, N. E., Clayton, D. G., 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88: 9 – 25.
- [4] Bishop, J., Venables, W.N., Wang, Y.G., 2004. Analysis commercial catch and effort data from a Penaeid trawl fishery. A comparison of models, mixed models, and generalised estimating equations approaches. *Fisheries Research* 70: 179 – 193.
- [5] Brezger, A., Kneib, T., Lang, S., 2005. BayesX: analyzing Bayesian structured additive regression models. *Journal of Statistical Software* 14 (11).
- [6] Brumback, B., Rice, J. A., 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of American Statistical Association* 88: 9 – 25.
- [7] Bryk, A., Raudenbush, S., 1992. *Hierarchical Linear Models for Social and Behavioral Research*, Sage, Newbury Park, CA.
- [8] Carbonell, A., Carbonell, M., Montserrat, S., Grau, A., Demestre, M., 1999. The red shrimp *Aristeus antennatus* (Risso, 1818) fishery and biology in the Balearic Islands, western Mediterranean. *Fisheries Research* 44: 1 – 13.

- [9] Cartes, J.E., 1998. Feeding strategy and partition of food resources in deep-water decapod crustaceans (400 – 2300 m). *Journal of Marine Biological Association of the United Kingdom* 78: 509 – 524.
- [10] Cleveland, W. S., 1979. Robust locally weighted regression and smoothing scatterplots, *Journal of American Statistical Association* 74 (368): 829 – 836
- [11] Cleveland, W. S., Devlin, S. J., 1988. The shape parameter of a two-variable graph. *Journal of the American Statistical Association* 83: 289–300.
- [12] Cooper A. B., Rosenberg A. A., Stefánsson G., Mangel, M., 2004. Examining the importance of consistency in multy-vessel trawl survey design based on the U.S. west coast groundfish bottom trawl survey. *Fisheries Research* 70: 239 – 250.
- [13] Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the generalized cross-validation. *Numerische Mathematik* 31: 337 – 403.
- [14] de Boor, C., 1978. *A Practical Guide to Splines*. Revised edition. *Applied Mathematical Sciences* 27. Springer–Verlag Inc, New York.
- [15] Demestre, M., 1995. Moulting activity-related spawning success in the Mediterranean deep-water shrimp *Aristeus antennatus* (Decapoda: Dendrobranchiata). *Marine Ecology Progress Series* 127: 57 – 64.
- [16] Demestre, M., Martin, P., 1993. Optimum exploitation of a demersal resource in the western Mediterranean: the fishery of the deep water shrimp *Aristeus antennatus* (Risso, 1816). *Scientia Marina* 57: 175 – 182.
- [17] Drinkwater, K.F., 2002. A review of the role of climate variability in the decline of northern cod. *American Fisheries Society Symposium* 32: 113 – 130.
- [18] Eilers, P. H. C., Marx, B. D., 1996. Flexible smoothing using B-splines and penalties (with comments and rejoinder). *Statistical Science* 11, 89 – 121.
- [19] Eilers, P.H.C., Marx, B.D., 2002. Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics* 11(4): 758 – 783.

- [20] Eilers, P. H. C., Marx, B. D., 2004. Splines, knots and penalties, unpublished manuscript. URL <http://www.stat.lsu.edu/faculty/marx/splinesknotspenalties.pdf>
- [21] Efron, B., 2005. Bayesians, Frequentists, and Scientists. To appear in Journal of American Statistical Association, 100.
- [22] Fahrmeir, L. Lang, S., 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. Journal of the Royal Statistical Society 50: 201 – 220.
- [23] Goldstein, H., 1995. Multilevel Statistical Models, Halstead Press, New York.
- [24] Green, P. J., 1987. Penalized Likelihood for General Semi-Parametric Regression Models. International Statistical Review 55 (3): 245 – 259.
- [25] Green, P. J., Silverman, B. W., 1994. Nonparametric Regression and Generalized Linear Models. London, Chapman and Hall.
- [26] Gu, C., 2002. Smoothing Spline ANOVA Models. Springer, New York.
- [27] Härdle, W., 1990. Applied Nonparametric Regression, Cambridge Univ. Press.
- [28] Hartley, H. O., Rao, J. N. K., 1967. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54: 93 – 108.
- [29] Hastie, T., Tibshirani, R., 1986. Generalized additive models (with discussion). Statistical Science 1: 297 – 318.
- [30] Hastie, T., Tibshirani, R., 1990. Generalized additive models Chapman Hall, London.
- [31] Hastie, T., Tibshirani, R., 2000. Bayesian Backfitting. Statistical Science 15: 193 – 223.
- [32] Helser, T.E., Punt, A.E., Methot, R.D., 2004. A generalized linear mixed analysis of a multy-vessel fishery resource survey. Fisheries Research 70: 251 – 264.
- [33] Heyde, C. C., 1997. Quasi-likelihood and Its Application, Springer, New York.

- [34] Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M., 2003. An overview of the North Atlantic Oscillation. In: Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M. (Eds.), *The North Atlantic Oscillation. Climatic Significance and Environmental Impact*. Geophysical Monograph, vol. 134. American Geophysical Union, Washington, DC. 279 pp.
- [35] Jang, W., Lim, J., 2006. PQL Estimation Biases in Generalized Linear Mixed Models. Scenario, p.1 – 15. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.68.6778&rep=rep1&type=pdf>.
- [36] Kammann, E. E., Wand, M. P., 2003. Geoadditive models. *Journal of the Royal Statistical Society* 52: 1–18.
- [37] Kohn, R., Ansley, C. F., Tharm, D., 1991. The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of American Statistic Association* 86: 1042 – 1050.
- [38] Lang, S., Brezger, A., 2004. Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13: 183 – 212.
- [39] Lin, X., Zhang, D., 1999. Inference in generalized additive mixed models using smoothing splines. *Journal of the Royal Statistical Society, Series B* 61: 381 – 400.
- [40] Lindley, D., Smith, A., 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Ser. B* 34: 1 – 41.
- [41] Mallows, C. L., 1973. Some comments on C_p . *Technometrics* 15: 661 – 675.
- [42] Marx, B.D., Eilers, P.H.C., 1998. Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis* 28: 193 – 209.
- [43] Maynou F., M Demestre, P Sánchez, 2003. Analysis of catch per unit effort by multivariate analysis and generalised linear models for deep-water crustacean fisheries off Barcelona (NW Mediterranean). *Fisheries Research* 64 (1 – 3): 257 – 269.
- [44] Maynou, F., Sardà, F., Tudela, S., Demestre, M., 2006. Management strategies for red shrimp (*Aristeus antennatus*) fisheries in the Catalan Sea (NW Mediterranean) based on bioeconomic simulation analysis. *Aquatic Living Resources* 19: 161 – 171.

- [45] Maynou, 2008. Environmental causes of the fluctuations of red shrimp (*Aristeus antennatus*) landings in the Catalan Sea . Journal of Marine Systems 71 (3-4): 294 – 302.
- [46] McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models, 2nd ed., Chapman and Hall, London.
- [47] Miyata, S., Shen, X., 2005. Journal of Japan Statistical Society 35 (2): 303 – 324.
- [48] O’Sullivan, F., 1986. A Statistical Perspective on Ill-Posed Inverse Problems (with Discussion). Statistical Science 1: 505 – 527.
- [49] Pinheiro, J.C., Bates, D.M., 2000. Mixed effects Models in S and S-PLUS. Springer.
- [50] Reinsch, C., 1967. Smoothing by spline functions. Numer. Math. 24: 383 – 393.
- [51] Rubin, D.B., 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics 12: 1151 – 1172.
- [52] Ruppert D., Wand M. P., Carroll R. J., 2003. Semiparametric regression. Cambridge University Press.
- [53] Sakamoto, Y., Ishiguro, M. and Kitagawa, G., 1986. Akaike Information Criterion Statistics. Reidel, Dordrecht, Holland.
- [54] Sardà F., J.B. Company, Castellón A., 2003. Intraspecific aggregation structure of a shoal of a western Mediterranean (Catalan coast) deep-sea shrimp, *Aristeus antennatus* (Risso, 1816), during the reproductive period. Journal Shellfish Research 22(2): 569 – 579.
- [55] Sardà, F., and Maynou, F. 1998. Assessing perceptions: do Catalan fishermen catch more shrimp on Fridays? Fisheries Research 36: 149 – 157.
- [56] Sardà, F., Cartes, J. E., and Norbis, W. 1994. Spatio-temporal structure of the deep-sea shrimp *Aristeus antennatus* (Decapoda: Aristeidae) population in the western Mediterranean. Fishery Bulletin 92: 599 – 607.
- [57] Sardà, F., G. D’Onghia, C.Y., Politou, J.B. Company, P. Maiorano and K. Kapiris., 2004. Deep-sea distribution, biological and ecological aspects of *Aristeus antennatus* (Risso, 1816) in the western and central Mediterranean sea. Scientia Marina 68(3): 117 – 127.

- [58] Schoenberg, I. J., 1964. Spline functions and the problem of graduation. Proc. Nat. Acad. Sci. U.S.A. 52: 947 – 950.
- [59] Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6: 461 – 464.
- [60] Stefánsson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J. Mar. Sci. 53: 577 – 588.
- [61] Tudela S. F., Sardà F. M., Maynou F., Demestre M., 2003. Influence of submarine canyons on the distribution of the deep-water shrimp (*Aristeus antennatus*, Risso 1816) in the north western Mediterranean. Crustaceana 76 (2): 217 – 225.
- [62] Visbeck, M., Chassignet, E.P., Curry, R.G., Delworth, T.L., Dickson, R.R., Krahnmann, G., 2003. The ocean's response to North Atlantic Oscillation variability. In: Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M. (Eds.), The North Atlantic Oscillation. Climatic Significance and Environmental Impact. Geophysical Monograph, vol. 134. American Geophysical Union, Washington, DC. 279 pp.
- [63] Wahba, G., 1985. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. Annals of Statistics 13: 1378 – 1402.
- [64] Wahba, G., 1990. Spline Models for Observation Data. Society for Industrial and Applied Mathematics.
- [65] Wang, Y., 1998. Mixed effects smoothing spline analysis of variance. Journal of the Royal Statistical Society 60 (1): 159 – 174.
- [66] Wasserman, L., 2005. All of Nonparametric Statistics. Springer.
- [67] Whittaker, E.T., 1923. On a new method of graduation. Proc. Edinburgh Math. Soc. 41: 63 – 75.
- [68] Wood, S.N., 2003. Thin plate regression splines. Journal of the Royal Statistical Society 65 (1): 95 – 114.
- [69] Wood, S.N., 2006. Generalized Additive Models: An Introduction with R CRC/Chapman Hall, Boca Raton, Florida.