De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

2011
2012

FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef

spatial modeling of fish distributions

Promotor :
Prof.dr. Christel FAES

Promotor :
Dr. JANINE ILLIAN
Dr. SOPHIE SMOUT

Maxwell Tawanda Chirehwa
Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

spatial modeling of fish distributions

Promotor :
Prof.dr. Christel FAES

Promotor :
Dr. JANINE ILLIAN
Dr. SOPHIE SMOUT

Maxwell Tawanda Chirehwa

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

ACKNOWLEDGEMENTS

It would not have been possible to write this master thesis without the help and support of people around me, to some of whom it is possible to give particular mention here.

This thesis would not have been possible without the help, support and encouragement of my internal supervisor, Prof. Dr. Christel Faes. Your invaluable support is appreciated and it has been a pleasure working with you. The advice and support of my external supervisors, Dr. Janine Illian and Dr. Sophie Smout, has been priceless and I am very grateful.

I would like to thank my family and friends at Hasselt University for their support throughout my studies. Lastly, I deeply thank the Vlaamse Interuniversitaire Raad-University Development Cooperation (VLIR-OUS) for funding my studies. Without funding, I would not have been able to pursue my studies. Above all I thank the Lord for extra strength.

ABSTRACT

Fish abundance is thought to be related to environmental factors and interactions between species. Catches at locations close to each other may also be similar resulting in spatial autocorrelation. The aim of this report is to assess the relationship between fish abundance and environmental correlates, the relationship between distributions of different fish species and temporal trends in fish abundance. Data for whiting, haddock and cod fish species were collected under the North Sea-International Bottom Trawl Survey (NS-IBTS) from 1994 to 2005. Approximate full Bayesian inference was performed based on the Integrated Nested Laplace Approximation (INLA) approach. Predictions of fish catches per quarter and age class for 1994 and 2005 were done for each of the three species. Results show that there is a relationship between fish abundance and depth of the sea. Different species are more abundant at locations of different depth which varied with quarter and age class. Further to this, there is an association between fish abundance and age class as well as fish abundance and quarter in which the survey was conducted. There is a reduction in fish abundance from 1994 to 2005. Low catches of haddock are associated with higher expected catches of whiting, and the same association between cod and whiting was established. Low catches of cod are negatively associated with catches of haddock

Keywords: Bayesian analysis, Cod, Haddock, Integrated Nested Laplace Approximation (INLA), Fish abundance, Spatial distribution, Whiting

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The North Sea is a relatively small basin, with a surface area of about 575 300 km^2 and a volume of 42 300 km^3. Although the North Sea occupies a rather small area, it is by no means homogeneous as regards, for example, depth, temperature, water type, and substratum. Waters are relatively shallow over much of the area, with average depths ranging from about 30 metres in the southeast to 200 metres in the northwest (ICES, 2012). The area is important for marine shipping, fishing, military purposes, extraction of minerals, oil and gas as well as tourism. Lately, it has also become important for renewable energy installations such as wind farms (Paramor et al., 2009).

In nature, resources are often patchily distributed and such patchiness can affect both population numbers and the coexistence of species (Chesson, 2000). Fish populations, like other animal and plant populations are not distributed randomly in space but show some spatial patterns. Understanding what the underlying and pattern generating processes are, is a fundamental ecologic question and a requirement for proper management when fish species are of conservation interest (Planque et al., 2011).

Factors that influence the spatial distribution of fish can be classified into external and internal. Internally, the population size, age structure, fish condition, diversity and behaviour can modulate the spatial distribution of fish population through mechanisms such as density dependence, age or stage-dependent habitat preference and differential migration capacities (Planque et al., 2011). In addition, individual memory which determines conservatism at a population level might be responsible for the maintenance of a spatial population pattern across years and generations (Corten, 2001). Fish distribution is also affected by attracting or repulsing interactions such as during the spawning period, when males and females tend to concentrate at relatively small spatial scales to minimize gamete loss and to maximize reproductive success. Such conspecific attraction may also lead to the distribution of fish being auto-correlated in space (Loots et al., 2010).

External factors affect populations differently depending on the life history stages (egg, larva, juvenile, and adult). At each stage fish must experience suitable abiotic conditions, find food for growth, and find shelter to escape from predation or disease. Changes in fish populations may result from physiological response to changes in environmental parameters (e.g. temperature), behavioural response such as avoiding unfavourable conditions and moving into new suitable areas, changes in fecundity and/or trophic interactions. The commercial exploitation of certain species of fish greatly affects their abundance and may interact with other external factors (Rijnsdorp et al., 2009). In the North Sea, there are a number of other human activities which have the potential to impact on fish populations including the input of trace organic contaminants and nutrients from land, the input of oil and Polycyclic Aromatic Hydrocarbons (PAHs) from land and by the offshore oil and gas industry, and the input of oil, PAHs and antifoulants by shipping (Rogers and Stocks, 2001).

The dominant fish species in the North Sea include demersal species such as Whiting (*Merlangius merlangus*) and Haddock (*Melanogrammus aeglefinus*), and pelagic species including Mackerel (*Scomber scombrus*) and Horse mackerel (*Trachurus trachurus*). In shallower waters (50–100 metres depth), populations are dominated by Haddock, Whiting, Herring (*Clupea harengus*), Dab (*Limanda limanda*) and Plaice (*Pleuronectes latessa*), while at greater depth (100–200 metres), Norway pout (*Trisopetrus esmarki*) dominate.

Whiting is commonly found from 30 to 100 metres, mainly on mud and gravel bottoms, but also on sand and rock. It feeds on shrimps, crabs, mollusks, small fish, polychaetes and cephalopods. In the North Sea, time of spawning is between January and September. They mature at between three and four years of age with females growing faster than males, and can live to about ten years of age. The biggest threat to this species is over-harvesting by the fishing fleets of many nations (Arkley and Caslake, 2004).

Haddock usually lives at depths between 40 and 300 metres, but can also be seen at 15-20 metres. It feeds primarily on small invertebrates, although larger members of the species may occasionally consume fish. It thrives in temperatures of 2 to 10 °C (36 to 50 °F). Spawning occurs in the months of February to April. Juveniles prefer shallower waters and larger adults deeper water. Generally, adult haddock do not engage in long migratory behaviour as do the younger fish, but seasonal movements have been known to occur across all ages. Haddock is a popular food fish and is widely fished commercially (Cargnelli et al., 1999).

Cod (*Gadus morhua*) can be found on any depth, from the subtidal zone and down to more than 600 metres and is widespread around the northern hemisphere (Telnes, 2012). It can grow to 2 metres in length and weigh up to 96 kilograms (210 lb). It can live for 25 years and sexual maturity is generally attained between ages two and four (O’Brien et al., 1993). Most cod spawn between the months of January and April and females, if large enough, can release up to five million eggs. Depending on the temperature, the eggs hatch in two to four weeks and the young cod drift in the open ocean, feeding on small crustaceans. Atlantic cod will eat a wide variety of prey, ranging from other fish to worms; they also take swimming crabs, shrimps and prawns (Wildscreen U.K. Charity, 2012).

1.1 Objectives

The main objective of the study is to model the spatial distribution of fish using Integrated Nested Laplace Approximation (INLA). Furthermore, to assess the

i. Relationship between fish abundance and environmental correlates,
ii. Relationship between the distribution of the three fish species,
iii. Temporal trends in the distribution of the three species.

2 METHODOLOGY

2.1 Description of Data

Data on whiting, haddock and cod species for the period from 1994 to 2005 were extracted from the North Sea - International Bottom Trawl Survey (NS-IBTS). This survey monitors the abundance of both commercial and non-commercial fish species and has been conducted annually since 1965. In addition, it is now only conducted in quarter one and quarter three each year. Eight countries bordering the North Sea carry out approximately 600-700 hauls over the yearly survey period. The International Council for the Exploration of the Sea (ICES) statistical rectangles (1° longitude x 0.5° latitude = ca. 30 nm x 30 nm = ca. 56 km x 56 km) define the survey grid. In each rectangle, fishing is conducted by ships of two different countries so that at least two hauls are made per rectangle. The standard haul duration is 30 minutes with a fishing speed of 4 knots. Data recorded in the surveys include quarter, ship, gear, haul number, geographic location coordinates, date and time, area (region of the north sea), subarea (ICES statistical rectangle), species, length/age class and Catch Per Unit of Effort (CPUE). CPUE is an index for fish abundance at a location for a particular age class.

2.2 Exploratory Data Analysis

Exploratory spatial data analysis was carried out using bubble plots for catch counts at particular locations and age class. Histograms of catch counts for each haul and species were also plotted. A scatter plot for the catches by depth is presented, in order to give an indication of depth preference for species. Bar charts of total catches per year disaggregated by age class and quarter are also provided.

2.3 Spatial statistics

Spatial analysis is defined as quantitative procedures employed in the study of spatial arrangement of features (Benhart, 2006). In spatial statistics, it is often the case that some or all outcome measures exhibit spatial autocorrelation. Spatial autocorrelation arises as a result of observations at locations close to each other having more similar values than those far from each other (Dormann et al., 2007). Ignoring spatial autocorrelation when analyzing data with spatial characteristics can bias parameter estimates and yield incorrect standard error estimates (Mohebbi et al., 2011).

Spatial models are often formulated within the hierarchical Bayesian framework. INLA may be used to fit (spatial and non-spatial) latent Gaussian models, a large subset of Bayesian additive models with a structured additive predictor, in which Gaussian priors are assigned to all components of the vector of latent variables (Rue et al., 2009). More detail is provided in section 2.4.

2.4 INLA Methodology

2.4.1 Motivation

Inference for latent Gaussian models is not straightforward since in general, the posterior distribution is not analytically available. To obtain parameter estimates, Markov Chain Monte Carlo (MCMC) techniques are widely used and regarded as the standard answer to the problem (Martino and Rue, 2010). However, fundamental challenges in applying MCMC remains: computational time is long, parameter samples can be highly correlated and estimates may have a large Monte Carlo (MC) error (Schrödle and Held, 2011). Further to this, to determine convergence, users face the task of choosing between several different measures of convergence, which might give contradictory answers to the question of convergence (Wilhelmsen et al., 2009).

Integrated Nested Laplace Approximation (INLA) is a recently proposed method (Rue et al., 2009) for approximate Bayesian inference in structured additive regression models with latent Gaussian fields. It is an alternative to MCMC in which it substitutes stochastic sampling with fast and accurate deterministic approximations based on the use of Laplace approximation and numerical integration. The methodology is particularly attractive if the latent Gaussian model is a Gaussian Markov random field (GMRF) with precision matrix $\mathbf{Q}$ controlled by a few hyperparameters $\boldsymbol{\theta}$ say ≤ 6 (Beguin et al., 2012). A GMRF is a Gaussian random variable $\mathbf{x} = (x_1, \ldots, x_n)$ with Markov properties such that for $i \neq j$, x_i and x_j are independent conditional on $\mathbf{x}_{-ij}$. The Markov properties are encoded in the precision matrix $\mathbf{Q}$ as $\mathbf{Q}_{ij} = 0$ if and only if x_i and x_j are independent conditional on $\mathbf{x}_{-ij}$ (Rue et al., 2009).

2.4.2 Model formulation

The response variable $\boldsymbol{y}_i$ (i=1,…, n) is the CPUE at a location and age class, and is assumed to follow an exponential family distribution. The mean μ_i is related to a structured additive predictor $\boldsymbol{\eta}_i$ through a link function $g(\cdot)$ so that $g(\mu_i) = \boldsymbol{\eta}_i$. The effects of various covariates are accounted for in an additive way as follows:

$$\boldsymbol{\eta}_i = \alpha + \boldsymbol{z}_i^{'}\boldsymbol{\beta} + \sum_{j=1}^{m} f_j(u_{ji}) + f_s(s_i) + f_{su}(s_i), \quad i = 1, \dots, n \tag{1}$$

In model (1), α denotes an intercept, the vector **β** represents linear effects of covariates **z**, $\{f_j(\cdot)\}$'s are unknown functions of covariates **u** with m being the number covariates assumed to have non-linear smooth effects, $f_s(\cdot)$ are spatially structured effects that account for spatial autocorrelation and $f_{su}(\cdot)$ is a spatially unstructured component (Rue et al., 2009). All parameters in model (1) are considered random and can be represented as a latent field

$$\mathbf{x} = \left\{\alpha, \boldsymbol{\beta}, \left\{f_j(\cdot)\right\}, f_s(\cdot), f_{su}(\cdot)\right\}$$

Gaussian priors are assigned to all components of the above latent field **x**. Hyperparameters of these priors have hyper-prior distributions which might often be non-Gaussian (Gomez-Rubio, 2011).

The INLA approach for approximating posterior marginals is computed in three steps. The first step approximates the posterior marginal distribution of the hyperparameters **θ**, that is, the parameters of the prior distribution on the latent field **x** using the relationship

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\mathbf{x},\boldsymbol{\theta},\boldsymbol{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{y})},$$

where **θ** represents the hyperparameters, $\boldsymbol{y}$ the data and **x** the latent field described above. $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ can be approximated by

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\mathbf{x},\boldsymbol{\theta},\boldsymbol{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})},$$

where the denominator $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{y})$ is the Gaussian approximation to the full conditional distribution of **x** evaluated at the mode $\mathbf{x}^*(\boldsymbol{\theta})$ for a given value of the hyperparameters **θ**.

In the second step, INLA computes an approximation of the posterior distribution of the latent field **x**, given the observed response variables **y** and the hyperparameters $\boldsymbol{\theta}$. This can be estimated by

$$\tilde{\pi}(\mathrm{x}_i|\boldsymbol{\theta},\boldsymbol{y}) \propto \left.\frac{\pi(\mathbf{x},\boldsymbol{\theta}|\boldsymbol{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|\mathrm{x}_i,\boldsymbol{\theta},\boldsymbol{y})}\right|_{\mathbf{x}_{-i}=x^*_{-i}(\mathrm{x}_i,\boldsymbol{\theta})}.$$

Here $\tilde{\pi}_{GG}(\mathbf{x}_{-i}|\mathrm{x}_i,\boldsymbol{\theta},\boldsymbol{y})$ is a Gaussian approximation to the distribution of $\mathbf{x}_{-i}|\mathrm{x}_i,\boldsymbol{\theta},\boldsymbol{y}$ in which the vector $\mathbf{x}_{-i}$ denotes all elements of the latent field $\mathbf{x}$ except the i^{th} one and is centered around the mode $x^*_{-i}(\mathrm{x}_i,\boldsymbol{\theta})$.

Within INLA, $\pi(\mathrm{x}_i|\boldsymbol{\theta},\boldsymbol{y})$ can be approximated using three methods namely Gaussian, simplified Laplace approximation and Laplace approximation in order of accuracy and reverse order of computational expensiveness. The analysis performed for this study was based on simplified Laplace approximation since it has been found to be highly accurate (Rue et al., 2009). The third step is to perform numerical integration with respect to $\boldsymbol{\theta}$, that is to find the posterior distributions of the components of the latent field $\mathbf{x}$ given observed response variables $\mathbf{y}$ as the sum as follows:

$$\tilde{\pi}(\mathrm{x}_i|\boldsymbol{y}) = \sum_k \tilde{\pi}(\mathrm{x}_i|\boldsymbol{\theta}_k,\boldsymbol{y})\,\tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y})\,\Delta_k, \qquad (2)$$

In equation (2), Δ_k is the area weight corresponding to the integration point θ_k, $\tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y})$ is the approximate posterior marginal distribution of the hyperparameters $\boldsymbol{\theta}$ at integration point k and $\tilde{\pi}(\mathrm{x}_i|\boldsymbol{\theta}_k,\boldsymbol{y})$ is the approximate posterior marginal distribution of the latent field at integration point k. (Wilhelmsen et al., 2009)

In this paragraph, a description of how model (1) was applied to the fish data is given. In all the models presented in this report, the effects of depth, adult, quarter 3 were taken as linear whilst the effect of time was taken as non-linear. The linear predictor for the model without interactions was defined as

$$\boldsymbol{\eta}_i = \alpha + \beta_1 Adult + \beta_2 Quarter3 + \beta_3 Depth + f_1(time) + f_s(s_i) + f_{su}(s_i) \qquad (2)$$

In model (2), the effect of time was modelled as random walk of order 2 (rw2). The idea is to estimate a parameter for each of the distinct values of time so that in model (2) we have a parameter for each of the 12 time points. The procedure works as follows; unique ordered values of time are represented as a vector $\boldsymbol{t} = (t_1, \ldots, t_{12})$ and a random walk model of order 2 is then constructed assuming independent second-order increments defined as

$$\Delta^2 t_l = t_l - 2t_{l+1} + t_{l+2} \sim N(0, {\tau_t}^{-1}).$$

The density for $\boldsymbol{t}$ is then derived from the (12-2=10) second order increments as:

$$\pi(\boldsymbol{t}|\tau_t) \propto \tau_t{}^{10/2}\exp\left\{-\frac{\tau_t}{2}\sum_{l=3}^{12}(\Delta^2 t_i)^2\right\}$$

A vague Gamma prior is assigned to the hyperparameter $\tau_t \sim Gamma(1, 0.001)$. The spatially structured component $f_s(s_i)$ was assumed to vary smoothly from one spatial grid to another. To account for such smoothness, $f_s(s_i)$ is modelled as a Conditional Autoregressive distribution (CAR) and specified as

$$s_i|s_j, i \neq j, \tau_s \sim N\left(\frac{1}{n_i}\sum_{i\sim j} s_j, \frac{1}{n_i\tau_s}\right),$$

where n_i is the number of neighbours of spatial grid i, $i\sim j$ indicates that the two grid are neighbours. The precision τ_s is assigned a vague Gamma prior as above. The spatially unstructured component $f_{su}(\cdot)$ is assumed to be independent and identically distributed (i.i.d) Gaussian with mean zero and unknown precision τ_{su} which was assigned a vague Gamma prior as well.

2.5 Model Building

Four distributions for count data, namely the Poisson, the Negative Binomial (NB), the Zero-Inflated Poisson (ZIP) and the Zero-Inflated Negative Binomial (ZINB) distribution as well as Gaussian distribution on log transformed CPUE (Lognormal) were applied. The log link function was used for the Poisson, the negative binomial (NB), the Zero-Inflated Poisson (ZIP) and the Zero-Inflated Negative Binomial (ZINB) distributions whereas the identity link function was used for the Lognormal distribution for the response. The Deviance Information Criterion (DIC), introduced by (Spiegelhalter et al., 2002) was used to compare models. It is defined as:

$$DIC = \overline{D} + p_D,$$

where $\overline{D}$ is the posterior mean of the deviance of the model and p_D is the effective number of parameters. A better model was defined as the one with the smaller DIC value. Model comparison based on DIC is meaningful if the response variable is the same for all models fitted. To make these values comparable, the DIC values for models assuming Gaussian distribution on log transformed CPUE had to be adjusted for the log transformation. A constant term based on the first derivative of the logarithmic function was added to the DIC value obtained from INLA. The constant term is defined as

$$C = -2\sum_{i=1}^{n} log|J_i|,$$

where J_i is the first derivative of the logarithm transformation (Ntzoufras, 2009).

In addition to the DIC values, predictive measures can be used to both validate and to compare models (Rue et al., 2009). Within INLA, Conditional Predictive Ordinate (CPO) measures are obtained from predictive density for the observed y_i based on all other observations ($CPO_i = \pi(y_i|\boldsymbol{y}_{-i})$). One way of assessing the model quality is to use the cross-validated logarithmic score (Czado et al., 2009) defined as

$$Logscore = -mean(\log(CPO_i))$$

A smaller value of the logarithmic score indicates a better prediction quality of the model. This technique is useful if the response variable for the models under comparison is the same which is not the case with the models in this report. An alternative is to compute the Probability Integral Transform (PIT) values defined as $PIT_i = Prob(y_i^{new} \leq y_i|\boldsymbol{y}_{-i})$. For one to correctly use the PIT values, a distinction between continuous and count responses has to be made for comparison of models. For a count response, to obtain the correct PIT values an adjustment has to be made to the values from INLA as follows:

$$PIT_i^{adj} = PIT_i - 0.5 * CPO_i$$

As an informal diagnostic to assess, a histogram of PIT values is used and deviations from uniformity hint at reasons for prediction failures and model deficiencies (Czado et al., 2009).

3 RESULTS

3.1 Exploratory Analysis

There were a total of 8350 hauls over the period from 1994 to 2005. The distribution of the number of hauls per year is shown in figure A1, appendix. The number of hauls per year ranged from 590 to 770. The distribution of the number of catches in 1994 and 2005 for whiting is shown in figure 1. There are many hauls where fewer catches were observed and the distribution is skewed. Similar distributions were also observed for other species in the two years (figure A2). For whiting species, there were higher catches in quarter three compared to quarter one. More juvenile fish were caught compared to the whiting adults. There is a decrease in abundance of haddock from 1995 to 1998 and a sharp increase in 1999 before it starts to decrease again. Haddock is more abundant in quarter three compared to quarter one on average except for the year 1999. With regards to catches per age class, more adults were caught compared to juveniles. There is an overall decline in the abundance of cod species. The distribution by quarter is not the same for all the years. With regards to age class, juveniles are more abundant compared to adults (figure A3, appendix).

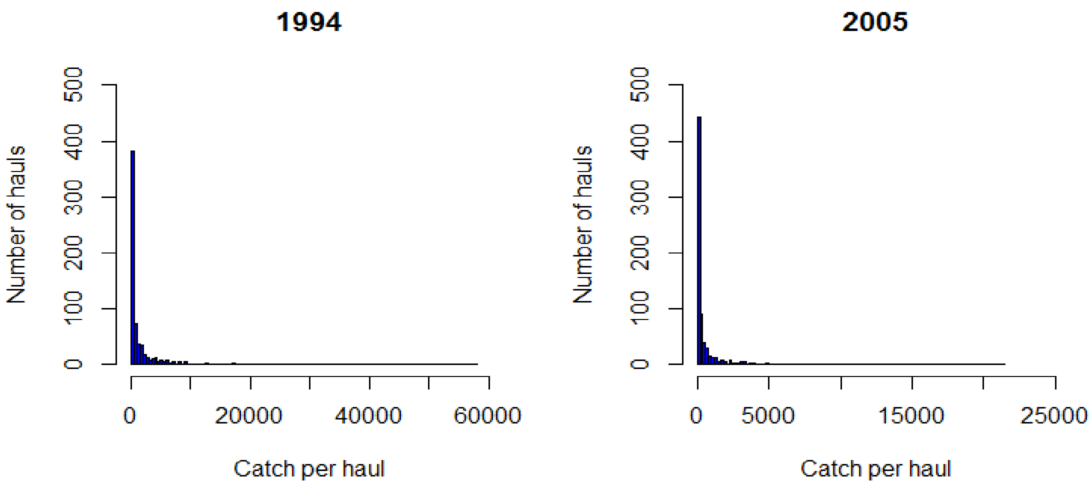
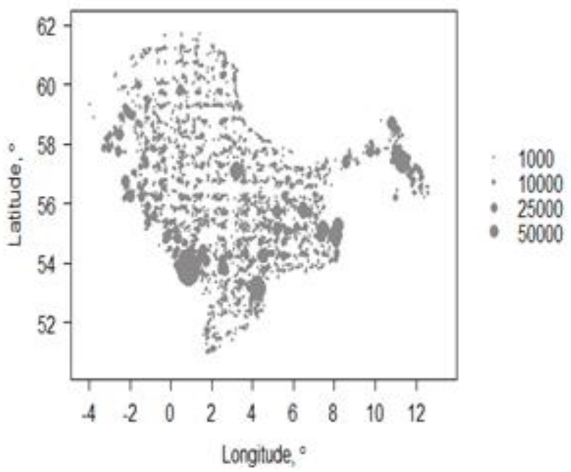
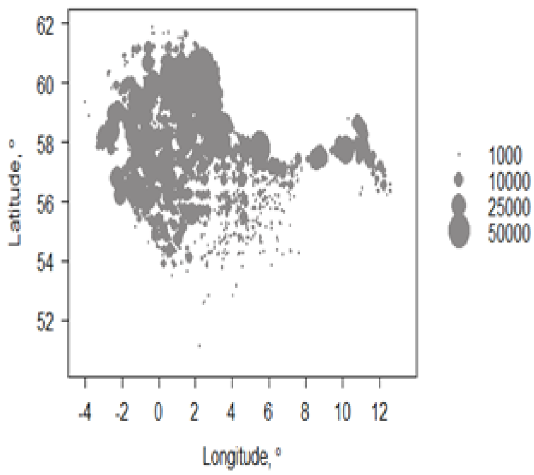


Figure 1: Distribution of catch for whiting species

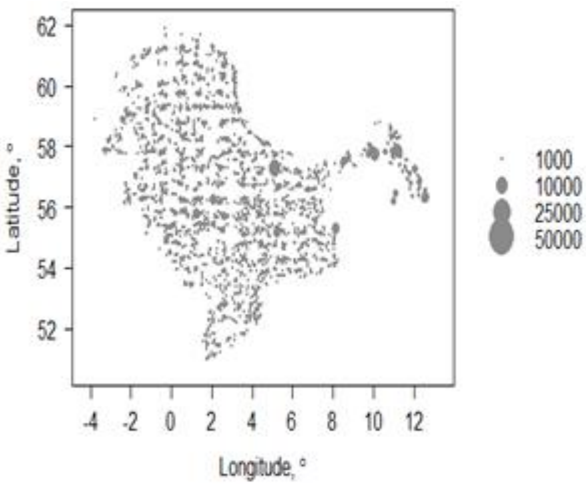
Figure 2 shows the spatial distribution of the three species under consideration for the period 1994 to 2005. The size of the circle corresponds to catch at a location. Whiting seems to be found in all the areas of the sea though there are locations where they are more abundant. Plots for spatial distribution in 1994 and 2005 are given in the appendix (Figure A4). For each of the two years, the distribution shows preference of whiting towards regions to the west of the sea as well as to the east. More catches were observed between a depth of 30 to 80 metres below sea level (figure A5, appendix). Haddock species prefers locations to the North West of the sea, between 54 degrees and 62 degrees latitude (figure A4, appendix). Higher catches were observed at location with depth of between 50 and 150 metres below sea level. Cod species is found in most locations of the Sea though in low numbers compared to other species fish abundance. Higher catches were observed at locations with a depth of 20 to 80 metres (figure A5, appendix).



(a) Whiting



(b) Haddock



(c) Cod

Figure 2: Bubble plots for catches per location for the three species, size of circle reflects CPUE

3.2 Statistical Analysis

3.2.1 Model for whiting abundance

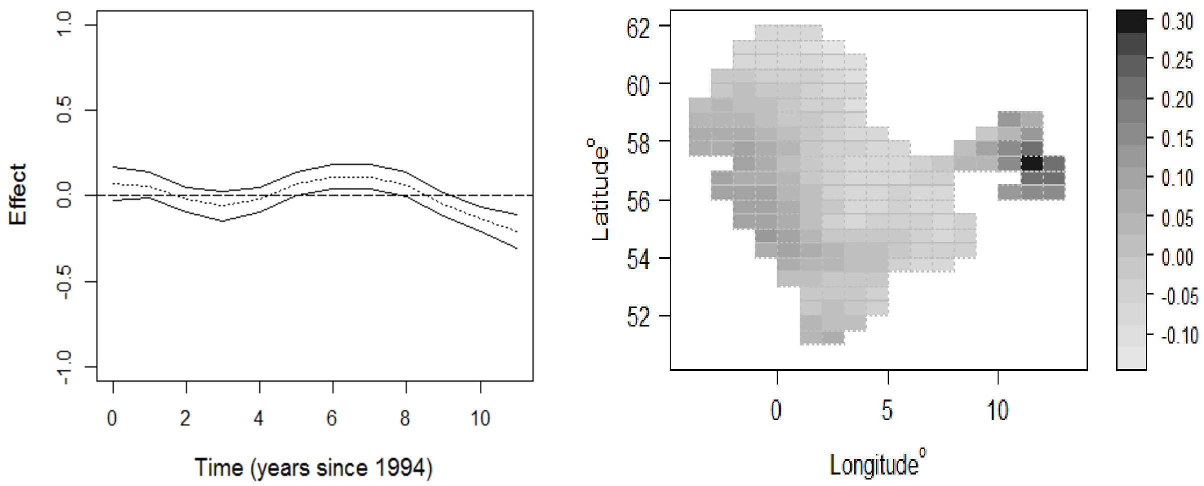
Table 1 shows results of the model to assess relationship between whiting abundance and environmental correlates. The model that assumed a normal distribution for log (CPUE + 1) had the lowest DIC value of 204 503.2. The structured additive predictor $\boldsymbol{\eta}_i$ for this model is defined as:

$$\boldsymbol{\eta}_i = \alpha + \mathbf{z}_\mathrm{i}^\mathrm{T}\boldsymbol{\beta} + f_1(T_i) + f_2(A_iT_i) + \ f_3(Q3_iT_i) + \ f_s(s_i) + \ f_{su}(s_i) \qquad (3)$$

The vector $\boldsymbol{\beta}$ represents the parameter estimates for the five linear effects namely depth, quarter 3, adult, depth*adult, depth*quarter 3. Letters used for non-linear effects in model (3) represent the variables as follows; adult (A), quarter 3 (Q3) and time (T). A combination of any two of these represents the interaction between the two variables. The variables adult and quarter 3 are binary with whilst depth and time are continuous. The effects of time $f_1(\cdot)$, interaction between time and age class $f_2(\cdot)$ as well as the interaction between time and quarter3 $f_3(\cdot)$ follows second-order random walk model with precision parameters τ_1, τ_2 and τ_3 respectively. The smooth spatial effects $f_s(\cdot)$ were modelled as a CAR distribution whilst the spatially unstructured effects were modelled as independent and .identically .distributed Gaussian with mean zero and precision defined as Gamma distributed. There is an effect of depth on fish abundance on the log scale which differs by quarter as well as age class. As depth increases, there is significantly more catches in quarter one compared to quarter three. At greater depth, adult whiting catch is 0.005 times more than juveniles. The effect of age class did not differ depending on the quarter in which the survey was conducted. There is a decline in fish abundance over time which differs by age class as well as quarter (figure 3a and A6). The spatial effects accounts for spatial autocorrelation not explained by covariates in the model (figure 3b). The pattern in the figure suggest that there could be other variables not measured that explains the spatial distribution of whiting species. Low variance of measurement error shows that adding spatial effects in the model helps to explain fish abundance.

Table 1: Posterior mean and standard deviation (SD) and quantiles of the linear effects and variance estimates of the model for whiting

Variable	**Mean**	**SD**	**2.5% quantile**	**97.5% quantile**
Intercept	3.200	0.041	3.119	3.281
Depth	0.012	0.001	0.011	0.014
Quarter 3	1.365	0.048	1.270	1.459
Adults	0.104	0.048	0.009	0.198
Depth*Quarter3	-0.014	0.001	-0.015	-0.013
Depth*Adults	0.005	0.001	0.004	0.006
Variance estimates				
Response	4.553	0.048	4.451	4.638
Time	0.006	0.005	0.001	0.018
Adults*Time	0.005	0.005	0.0004	0.017
Quarter3*Time	0.006	0.005	0.0001	0.019
Spatial Structured	0.011	0.003	0.007	0.017
Spatial unstructured	0.006	0.002	0.004	0.010

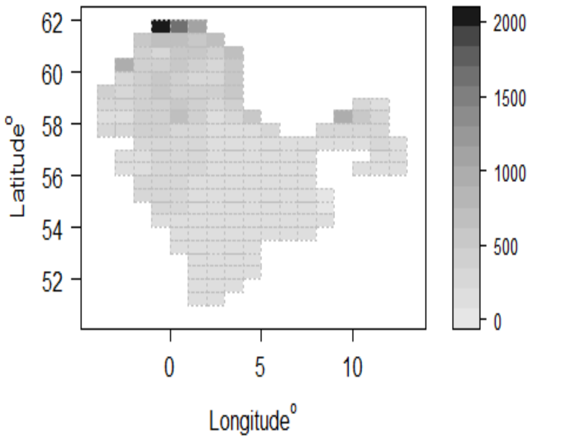


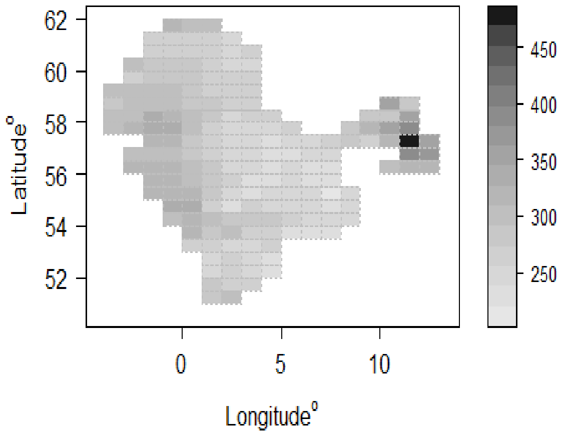
(a) Posterior mean of smooth effect of time

(b) Posterior mean of the smooth spatial effects

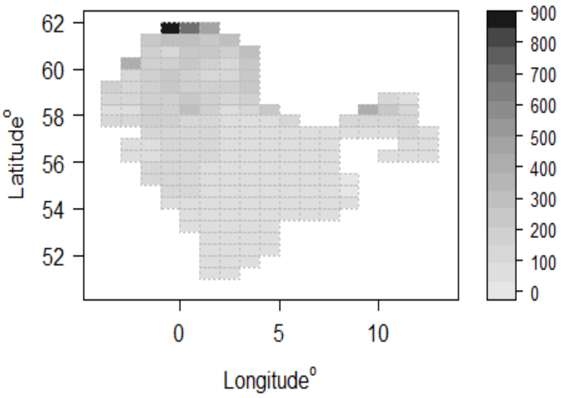
Figure 3: Estimated effects of smooth effects and spatial effects of the model for whiting

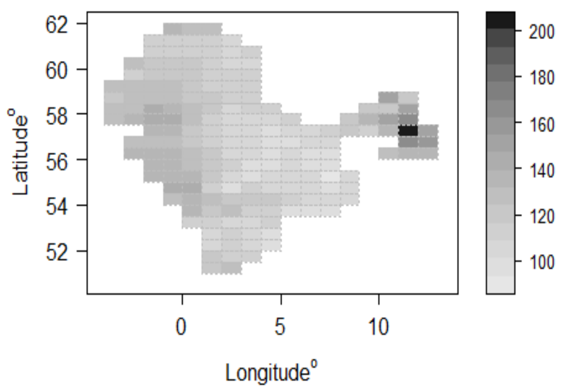
Figure 4 shows the predicted catches of whiting per statistical rectangle in quarter one and three of 1994 and 2005. With regards to the depth of the rectangle, the average depth was taken using all the locations in that rectangle. The plots show certain preferred locations where the predicted number of fish is higher. There is a shift in locations of high fish abundance depending on the quarter. In addition, there is low predicted CPUE in 2005 compared to 1994 for both quarters. Predicted catches by year and age class are shown in figure A7, appendix. There seems to be a balance between the predicted catches for juveniles and adults though in some rectangles, there are more adults than juveniles. The pattern from the plot suggests that juveniles and adults tend to be found in the same locations and almost equal numbers.





(a) Predicted catches in quarter one of 1994 (b) Predicted catches in quarter 3 of 1994





(c) Predicted catches in quarter one of 2005 (d) Predicted catches in quarter 3 of 2005

Figure 4: Predicted catches of whiting by quarter and year (1994 and 2005)

3.3.2 Model for haddock abundance

Results of the model describing the relationship between log (CPUE +1) for haddock and age, quarter, depth and year as covariates are presented in table 2. The model had the lowest DIC value of 161 637.8 and the structured additive predictor $\boldsymbol{\eta}_i$ is defined as:

$$\boldsymbol{\eta}_i = \alpha + \mathbf{z}_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{\beta} + f_1(T_i) + f_2(A_i T_i) + f_3(Q3_i T_i) + f_s(s_i) + f_{su}(s_i) \quad (4)$$

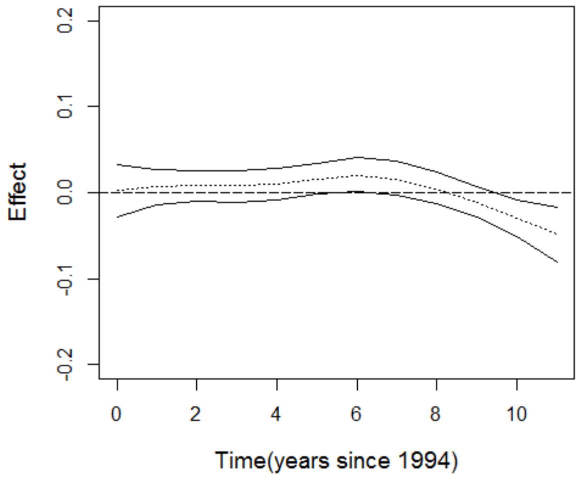
The vector $\boldsymbol{\beta}$ represents the linear effects for depth, quarter 3, adult, depth*adult, depth*quarter 3 and adult*quarter 3. Letters used in the definition of non-linear effects are represent the same variables as in model 3. There is a significant effect of depth on log (CPUE +1) for haddock. However, this effect depends on quarter as well as age class. The effect of depth is greater in quarter one compared to quarter three. A one metre increase in depth increases log (CPUE +1) in quarter one by 0.002 more than it does in quarter three. In the same way, a one metre increase in depth increases log (CPUE +1) for juveniles by 0.012 more than adults. There is higher log (CPUE +1) in quarter 3 compared to quarter one which is even higher by 0.087 for adults compared to juveniles.

There is an effect of time on fish abundance and this effect depends on quarter and age class. The effect of time on fish abundance is not different between quarter one and quarter three from 1994 (time 0) to 2003. Thereafter, time has a negative effect on fish abundance in quarter three compared to quarter one. Similarly, after 2003, time has a negative effect of fish abundance in adults than juveniles. There is low variability in the spatially structured effect as well as spatially unstructured effects. This suggests that unexplained variability in fish abundance is almost similar in all statistical rectangles of the sea.

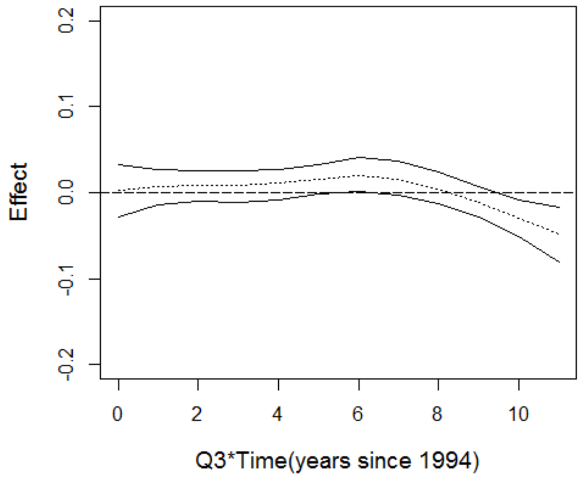
Predicted catches of haddock species are shown in figure A8 and A9, appendix. In both 1994 and 2005, predicted catches are higher in quarter three compared to quarter one. Pattern of predicted number of catches of juveniles and adults suggests that they co-exist though in different numbers.

Table 2: Posterior mean and standard deviation (SD) and quantiles for the linear effects and variance estimates of the model for haddock

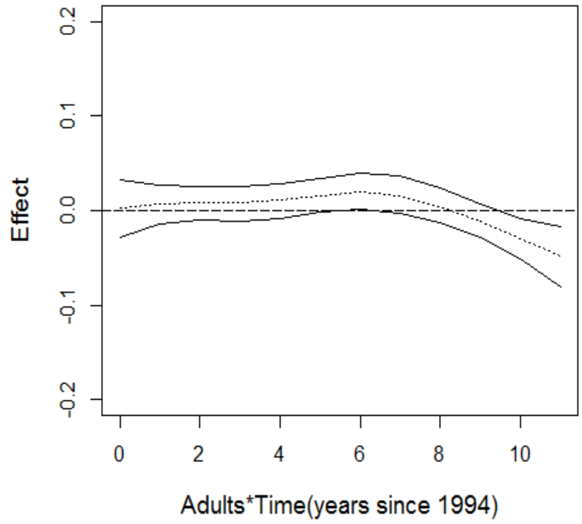
Variable	**Mean**	**SD**	**2.5% quantile**	**97.5% quantile**
Intercept	0.227	0.020	0.187	0.266
Depth	0.046	0.001	0.046	0.048
Quarter 3	0.204	0.021	0.162	0.246
Adults	0.157	0.022	0.114	0.200
Quarter3*Adults	0.087	0.022	0.044	0.129
Quarter3*Depth	-0.002	0.001	-0.003	-0.001
Adults*Depth	-0.012	0.001	-0.013	-0.011
Variance Estimates				
Response	5.704	0.072	5.567	5.851
Time	1.1E-04	1.2E-04	0.1E-04	1.3E-04
Quarter3*Time	0.9E-04	1.0E-04	0.1E-04	3.7E-04
Adults*Time	0.9E-04	0.9E-04	0.1E-04	3.7E-04
Spatial Structured	0.6E-04	0.3E-04	0.1E-04	1.3E-04
Spatial unstructured	1.7E05	0.8E-05	0.7E-05	3.7E05



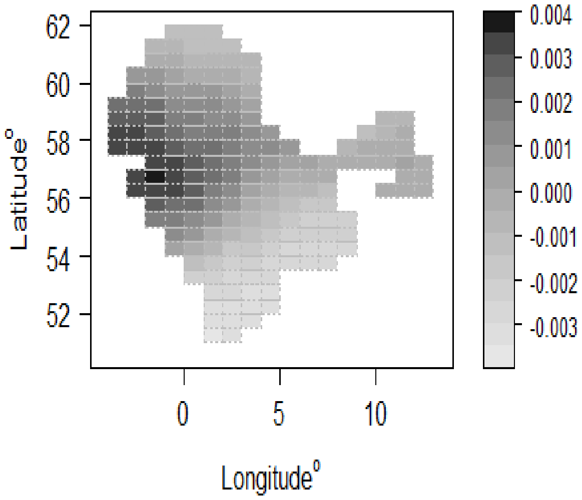
(a) Posterior mean of smooth effect of time



(b) Posterior mean of smooth effect of Quarter3* time



(c) Posterior mean of smooth effect of Adults*time



(d) Posterior mean of the smooth spatial effects

Figure 5: Estimated smooth effects and spatial effects of the model for haddock

3.2.3 Model for cod abundance

Table 3 shows results of the model describing the relationship between log (CPUE + 1) for cod and depth, time (years since 1994), age and quarter. The model selected had the lowest DIC value of 76 260.4 and the structured additive predictor $\boldsymbol{\eta}_i$ is defined as:

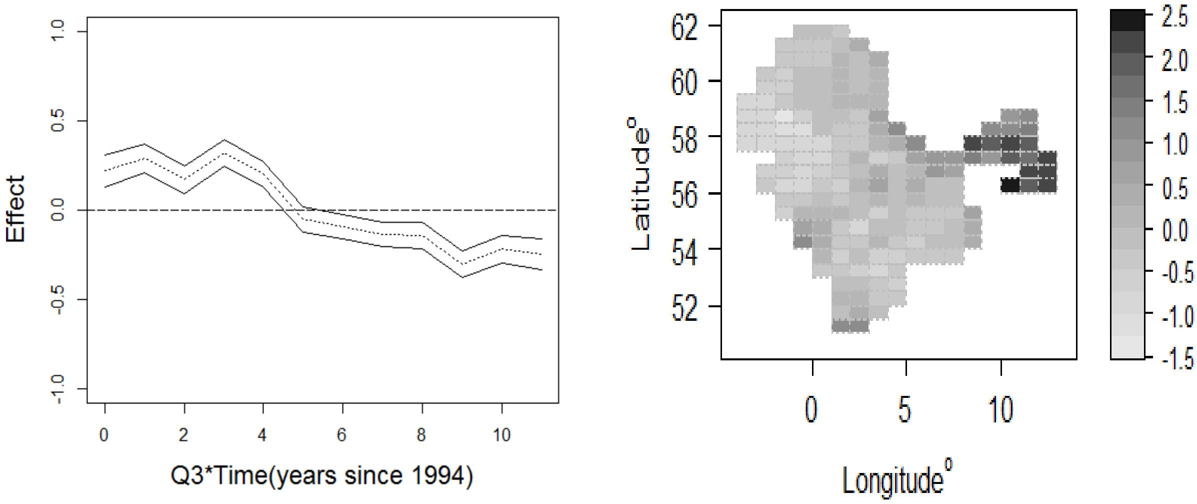
$$\boldsymbol{\eta}_\mathrm{i} = \alpha + \mathbf{z}_\mathrm{i}^\mathrm{T}\boldsymbol{\beta} + f_1(Q3_iT_i) + f_s(s_i) + f_{su}(s_i). \quad (5)$$

In model (5) α is the intercept, $\boldsymbol{\beta}$, represents the parameter estimates for the eight linear effects namely depth, quarter 3, adult, depth*adult, depth*quarter 3, adult*quarter 3, time, time*adult. In addition, $f_1(\cdot)$ is the non-linear smooth effect for interaction between time and quarter 3 which follows a second-order random-walk model with unknown precision τ_1 assigned a Gamma prior distribution. Similar to the previous models, $f_s(\cdot)$ and $f_{su}(\cdot)$ were modelled as a CAR distribution, and independent and identically distributed Gaussian respectively.

Depth has an effect on log (CPUE +1) which differs depending on quarter and age class. Compared to juveniles, a one metre increase in depth is associated with less catches of adults For a one metre increase in depth, the difference between log (CPUE +1) for quarter three and quarter one is 0.004 with higher catches in quarter one. Expected log (CPUE +1) of adult fish is 0.061 more than that of juveniles for a one year increase in time. There are significantly more catches in quarter three from 1994 to 1998 and then more in quarter one thereafter (figure 6a). Variability of the spatially structured effects shows that the differences in spatial effects. This is evident in the plot of spatial effects where the effect is higher in certain grids (figure 6b). Predicted catch of cod are shown in figure A10 and A11, appendix. Predicted catches are higher in the small part of the Sea between 9 degrees and 13 degrees longitude. The pattern is the same for the two age classes and two quarters.

Table 3: Posterior mean, standard deviation (SD) and quantiles for the linear effects and variance estimates of the model for cod

Variable	**Mean**	**SD**	**2.5% quantile**	**97.5% quantile**
Intercept	1.489	0.044	1.403	1.574
Depth	0.010	0.001	0.009	0.011
Time	-0.029	0.007	-0.042	-0.016
Quarter 3	0.352	0.044	0.265	0.438
Adults	-0.286	0.056	-0.395	-0.177
Adults*Time	0.061	0.009	0.043	0.080
Adults*Depth	-0.009	0.001	-0.011	-0.008
Quarter3*Depth	-0.004	0.001	-0.005	-0.003
Variance Estimates				
Response	1.327	0.020	1.285	1.362
Quarter3*Time	1.4E-04	0.7E-04	0.4E-04	3.0E-04
Spatial Structured	0.841	0.189	0.600	1.321
Spatial unstructured	1.5E-04	2.0E-04	0.2E-04	8.1E-04



(a) Posterior mean of smooth effect of the difference between quarter three and one over time

(b) Posterior mean of the smooth spatial effects

Figure 6: Estimated smooth effects and spatial effects of the model for cod

3.3 Relationship between fish species

To assess the relationship between fish species, a model with whiting as response and the other two species as covariates together with age, quarter, depth and time was developed. Another model with haddock as response variable was developed with covariates being whiting, cod and the above mentioned covariates.

3.3.1 Relationship between whiting and other species

Table 4 shows results of the model for the relationship between whiting and other species controlling for other covariates. The model has a DIC value of 219 879.2 and the response is the log (CPUE) for whiting. The structured additive predictor $\boldsymbol{\eta}_\mathrm{i}$ for model is defined as:

$$\boldsymbol{\eta}_\mathrm{i} = \alpha + \mathbf{z}_i^\mathrm{T}\boldsymbol{\beta} + f_1(T_i) + f_2(H_i) + f_3(C_i) + f_4(Q3_i * H_i) + f_5(Q3_i * C_i)$$
$$+ f_6(A_i * H_i) + f_7(A_i * C_i) + f_s(s_i) + f_{su}(s_i). \qquad (6)$$

The vector $\boldsymbol{\beta}$ represents parameter estimates for linear effects $\mathbf{z}$ which are depth, adult, quarter 3, depth*quarter3 and depth*adults. Letters in model (6) are explained as follows: H represents log (CPUE+1) for haddock, C represents log (CPUE+1) for cod, Q3 represents Quarter3 and A represents Adult. The non-linear smooth effects $\{f_1(\cdot), \dots, f_7(\cdot)\}$ follows a second-order random-walk model with precision parameters $\{\tau_1, \dots, \tau_7\}$. The smooth spatially structured effects $f_s(\cdot)$ and spatially unstructured effects $f_{su}(\cdot)$ are defined in the same way as in models (3) to (5). There is a relationship between the distribution of fish species and the relationship changes with quarter as well as age class. Low catches of haddock has a positive effect in quarter three compared to quarter one. However, as log (CPUE +1) for haddock increase, there is no difference in between quarter one and quarter three. The relationship between cod and whiting is similar to the one for haddock and whiting. Low catches of haddock are associated with higher catches of whiting with more catches of the whiting being from the adult age class. The same result was also observed for the relationship between cod and whiting.

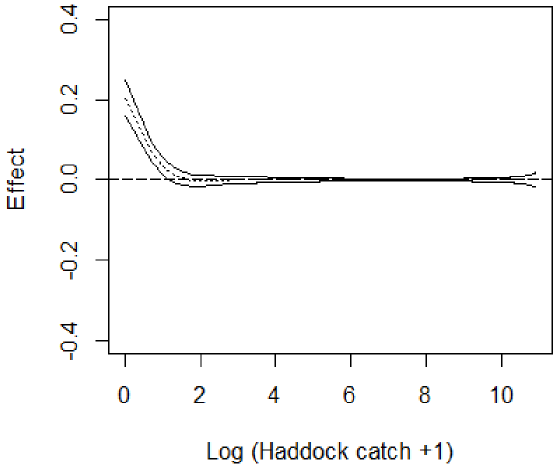
The relationship between depth of location and catches is such that in quarter three, there are less catches for a one metre increases in depth compared to quarter one. In addition, as depth increases, log (CPUE) for adult whiting is higher by 0.005 times compared to juveniles. There is an overall reduction in catches over time which become significant in 2003 (time 9)

Spatial effects suggest that there is some pattern not explained by the covariates in the model that explain high abundance of fish in parts of the sea between 9 degrees to 13 degrees longitude. There is a reduction in the variance of spatially unstructured effects compared to the one model presented in table 1. This shows that the model which include other species as covariates better explain variation in abundance of whiting species.

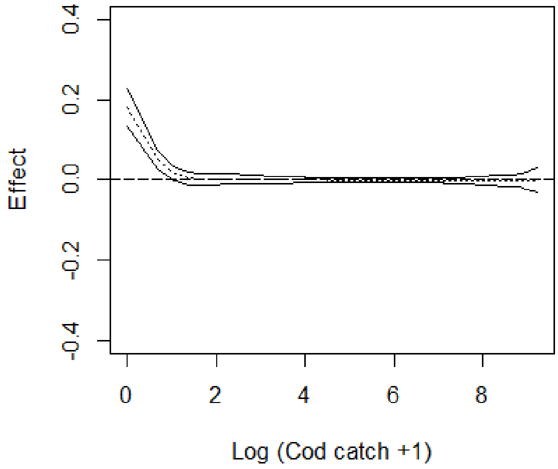
Table 4: Posterior mean, standard deviation (SD) and quantiles for the relationship between whiting and other species

Variable	**Mean**	**SD**	**2.5% quantile**	**97.5% quantile**
Intercept	0.700	0.022	0.658	0.743
Depth	0.029	0.001	0.028	0.030
Quarter 3	0.374	0.022	0.332	0.417
Adults	0.205	0.022	0.162	0.247
Depth*Quarter 3	-0.006	0.001	-0.007	-0.005
Depth*Adults	0.005	0.001	0.004	0.006
Variance Estimates				
Response	8.352	0.099	8.156	8.547
Time	2.1E-04	2.2E-04	0.2E-04	8.9E-04
Log(Haddock + 1)	0.002	0.001	0.001	0.004
Log(Cod + 1)	0.003	0.002	0.001	0.007
Q3*log(Haddock + 1)[1]	0.002	0.001	0.001	0.004
Q3*log(Cod+ 1)[1]	0.003	0.002	0.001	0.007
Adult* log(Haddock + 1)	0.002	0.001	0.001	0.004
Adult* log(Cod+ 1)	0.003	0.002	0.001	0.007
Spatial Structured	4.9E-05	3.0E-05	1.4E-05	1.2E-04
Spatial unstructured	1.8E-05	1.7E-05	0.7E-05	3.7E-05

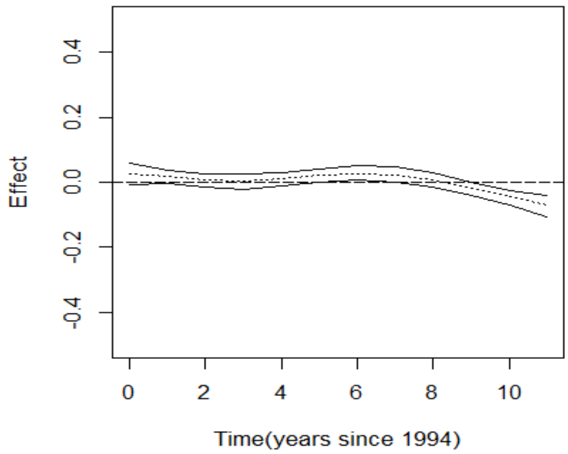
[1] Q3 represent Quarter 3



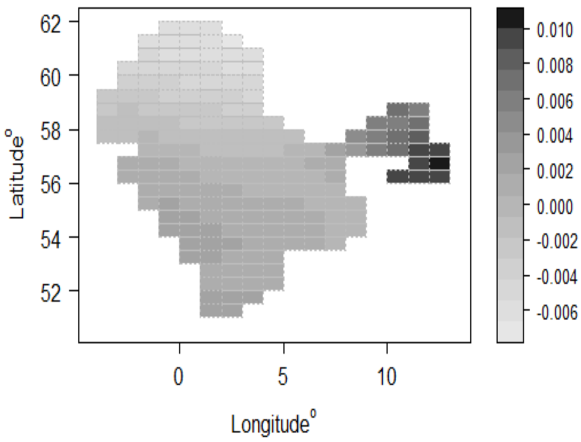
(a) Posterior mean of smooth effect of Haddock



(b) Posterior mean of smooth effect of Cod



(c) Posterior mean of smooth effect of time



(d) Posterior mean of the smooth spatial effects

Figure 7: Estimated smooth effects for relationship between whiting and other species

3.3.2 Relationship between haddock and other species

Posterior means together with other statistics for linear effects and variance estimates for the model describing the relationship between haddock and other species are shown in table 5. This model has a DIC value of 172 193.7 with the response being the log transformed catch counts plus one. The structured additive predictor $\boldsymbol{\eta}_{\mathrm{i}}$ of the model is defined

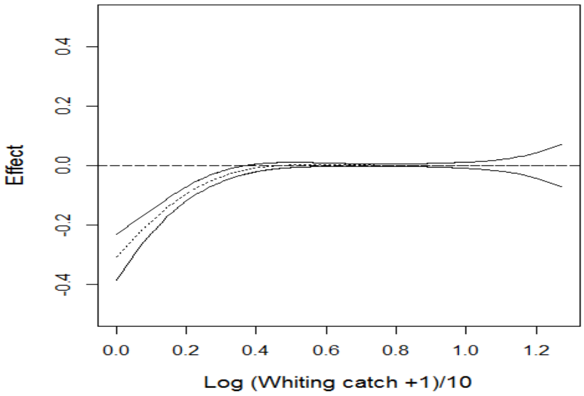
$$\boldsymbol{\eta}_{\mathrm{i}} = \alpha + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta} + f_1(T_i) + f_2(W_i) + f_3(C_i) + f_4(Q3_i * W_i) + f_5(Q3_i * C_i)$$
$$+ f_6(A_i * W_i) + f_7(A_i * C_i) + f_s(s_i) + f_{su}(s_i). \quad (7)$$

The vector of parameters $\boldsymbol{\beta}$ as well as linear effects $\mathbf{z}$ is defined as in model (6). The letter W in model (7) represents (log (CPUE + 1))/10 for whiting whilst the rest of the letters remain the same as they were described in model (6). The smooth effects also follow a second-order random-walk model. The spatially structured $f_s(\cdot)$ and unstructured effects $f_{su}(\cdot)$ are modelled as a CAR distribution and independent and identically distributed Gaussian respectively.

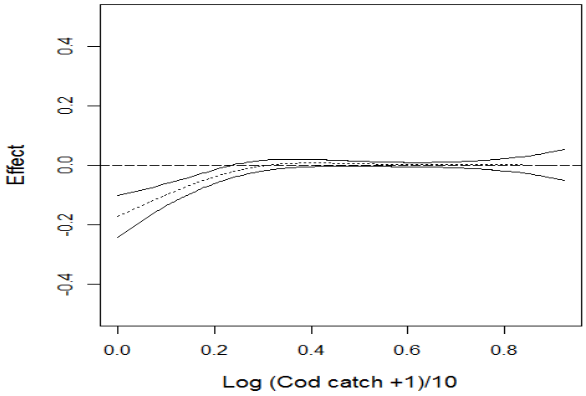
There is an effect of depth on catches of haddock controlling for the presence of other species. This effect differs by age class as well as quarter. For a one metre increase in depth, there are more catches of compared to adults. Similarly, as depth increases, there are more catches in quarter one compared to quarter three and the difference in catches is 0.01 on the logarithmic scale. Low catch of whiting below (log (catch +1))/10 of 0.3 has negative effect on haddock abundance that differ by age class and quarter. For values of (log (catch +1))/10 above 0.3, there is no relationship between catches of whiting and catches of haddock. Similarly, low catch of cod below (log (catch + 1))/10 of 0.2 has negative effect on haddock catch that differ by age class and quarter (figure 8 and A12). There are fluctuations in fish abundance over time between year 1994 and year 2000. After year 2000, there is a decline in haddock abundance till 2005. Spatially structured effect show that there are regions of the sea associated with high catches of haddock.

Table 5: Posterior mean, standard deviation (SD) and quantiles for the relationship between haddock and other species

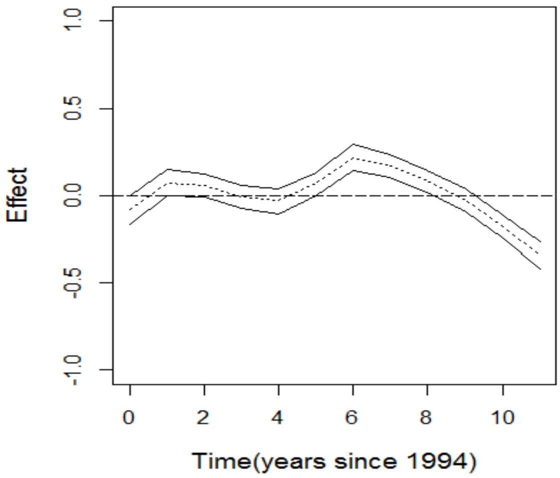
Variable	**Mean**	**SD**	**2.5% quantile**	**97.5% quantile**
Intercept	0.459	0.042	0.376	0.459
Depth	0.047	0.001	0.046	0.047
Quarter 3	0.882	0.047	0.789	0.882
Adults	-0.100	0.048	-0.195	-0.005
Depth*Adults	-0.011	0.001	-0.012	-0.011
Depth*Quarter3	-0.009	0.001	-0.010	-0.009
Variance Estimates				
Response	4.625	0.045	4.528	4.701
Time	0.014	0.011	0.003	0.043
Log (cod+1)/10	0.534	0.545	0.057	2.233
Log (whiting+1)/10	0.510	0.396	0.102	1.578
Q3*log (cod+1)/10	0.561	0.591	0.055	2.366
Q3*log (whiting+1)/10	0.511	0.389	0.100	1.547
Adult* log(cod+1)/10	0.479	0.470	0.047	1.817
Adult* log(whiting+1)/10	0.504	0.386	0.105	1.546
Spatial structured	0.012	0.003	0.013	0.026
Spatial unstructured	0.011	0.001	0.008	0.014



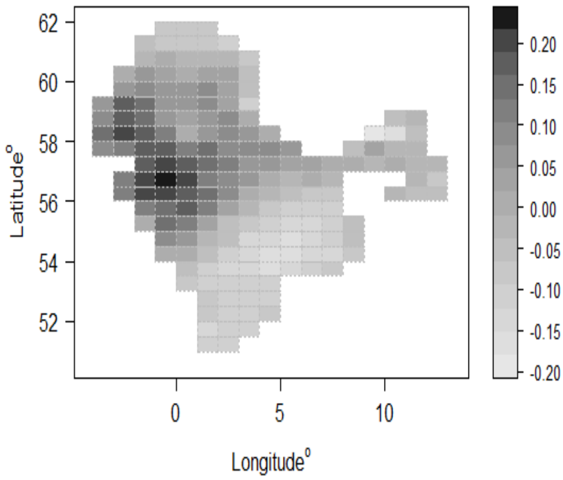
(a) Posterior mean of smooth effect of Whiting



(b) Posterior mean of smooth effect of Cod



(c) Posterior mean of smooth effect of time



(d) Posterior mean of the spatial effects

Figure 8: Estimated smooth effects for relationship between haddock and other species

4 DISCUSSION AND CONCLUSION

This study aims at establishing the relationship between fish abundance and environmental correlates, the relationship between distribution of fish species as well as temporal effects in fish abundance. In this report, three species were considered namely whiting, haddock and cod. Catch counts from the North Sea for the period from 1994 to 2005 were extracted from the NS-IBTS fish database.

The surveys were carefully planned so as to cover the entire North Sea. In each statistical rectangle, there were supposed to be at least two locations where data on catch count were collected and this was independent of previous measurements of fish CPUE at a location. In such cases, there is no interest in modelling locations hence these would be treated as fixed. However, this might not be the case all the time. In commercial fishing where fishermen tend to visit locations were they think they will have more catches, it will be important to model the locations where fishermen visit. One strategy is to perform joint modelling of fishing location as well as catches using the marked point pattern methodology (Illian et al., 2011). In this case, the locations are modelled as a point process and the catch at the locations as marks Fish abundance at locations where fishermen visit might not correspond to the amount the fishermen expected hence this need to be incorporated into the model by adding a fishermen error component (Pennino et al., 2012).

When using INLA, there is no distinction between spatial data type since all data are converted to lattice data for analysis. It has the flexibility of modelling data on both a regular grid as well as on non-regular grids. Data on regular grids are easily handled in the current version of INLA package and users have the ability to change the size of the grids to make them finer thus making approximations very accurate but at the expense of computational time (Simpson et al, 2011). For the North Sea surveys, locations are already defined within ICES rectangles and these were used as the grids for the analysis. This could have been improved if the locations were not systematically visited by changing the size of the grid for to accurately predict fish abundance. However, in this case the area of the grid has to be considered when modelling hence INLA users would run into boundary problems where the area of grids on shore is not the same as those off shore.

Recent developments in spatial analysis using INLA focus on the use of Stochastic Partial Differential Equations (SPDE) as a way to deal with grids on the boundary of the survey region. This involves discretization of the survey region and creating Delaunay triangulation using the observation locations as vertices. Another advantage is that other vertices which are not observation locations can be used for prediction purposes (Lindgren et al., 2011).

There are other environmental variables not measured in the survey that could have had an effect of the fish abundance in certain location. Only depth was recorded and results show that there is an association between fish abundance and depth which differs depending on quarter as well as age class. Temperature levels could also have had an effect on fish abundance since other studies have established the association (Järvalt et al., 2005). Further to this, different eutrophication levels at different locations within the North Sea might explain some of the spatial patterns in fish abundance.

There is a relationship between the three species studied taking into account the depth as well as the quarter of survey. Low catches of haddock are associated with high expected catches of whiting and low catches of cod are also associated with high expected catches of whiting controlling for the spatial distribution of whiting. For the model with haddock as response, low catches of whiting are associated with low expected catches of haddock and low catches of cod are associated with low catches of haddock as well. This relationship could further be improved by looking at the relationship between the effects of adults one species on the juvenile of the other species. There is a reduction in fish abundance from the period 1994 to 2005 for all the species. There is an increase in fish abundance for whiting and haddock in 1998 and 2000 which is then followed by a further decline.

In conclusion, INLA is a powerful inferential tool for latent Gaussian models. However, for large datasets with many non-linear smooth effects, more computer memory required. Overall, there is an overall reduction in fish abundance in the North sea for the species considered here. In addition, these species are more abundant in certain parts of the sea. Results from this study could be used by fishermen to visit locations where fish are more abundant and modelling of such data will have a reduced fishermen error. If there is need to conserve any species reported here, the results will be useful in identification of areas and habitats to protect

5 REFERENCES

Arkley, K., and Caslake, G. (2004). Off-bottom trawling techniques for the sustainable exploitation of non-pressure stocks in Cornish inshore waters. Project Reference No. FEP 592. September 2004. 2089 pp.: Defra Cornish Objective 1 Programme.

Beguin, J., Martino, S., Rue, H., and Cumming, S. G. (2012). Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution*.

Benhart J E. (2006). Spatial Analysis Techniques Notes. Pennsylvania: Indiana University of Pennsylvania.

Cargnelli, L. M., Griesbach, S. J., Berrien, P. L., Morse, W. W., and Johnson, D. L. (1999). *Essential fish habitat source document. Haddock, Melanogrammus aeglefinus, life history and habitat characteristics*. Woods Hole, Mass. : [Springfield, Va. :: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Region, Northeast Fisheries Science Center.

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* **31**, 343-366.

Corten, A. (2001). The role of ``conservatism" in herring migrations. *Reviews in Fish Biology and Fisheries* **11**, 339-361.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive Model Assessment for Count Data. *Biometrics* **65**, 1254-1261.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., and Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609-628.

ICES. (2012) *Factors affecting the distribution of North Sea fish* Available at: http://www.ices.dk/marineworld/fishmap/pdfs/factors.pdf. [Accessed July 2 2012]

Illian, J., Borchers, D., and Pennino, M. G. (2011). Spatial modelling in a marine environment when sampling with unknown inclusion probabilities. In *58th World Statistics Congress*. Dublin.

Järvalt, A., Krause, T., and Palm, A. (2005). Diel Migration and Spatial Distribution of Fish in a Small Stratified Lake. *Hydrobiologia* **547**, 197-203.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 423-498.

Loots, C., Vaz, S., Planque, B., and Koubbi, P. (2010). What controls the spatial distribution of the North Sea plaice spawning population? Confronting ecological hypotheses through a model selection framework. *ICES Journal of Marine Science: Journal du Conseil* **67**, 244-257.

Martino, S., and Rue, H. (2010). Case studies in Bayesian computation using INLA. In *COMPLEX DATA MODELING AND COMPUTATIONALLY INTENSIVE STATISTICAL METHODS*, *Contributions to Statistics*, P. Mantovan, and P. Secchi (eds), 99-114: Springer.

Mohebbi, M., Wolfe, R., and Jolley, D. (2011). A poisson regression approach for modelling spatial autocorrelation between geographically referenced observations. *BMC Medical Research Methodology* **11**, 133.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken, NJ Wiley.

O'Brien L, Burnett J, and Mayo R. K (1993). Maturation of Nineteen Species of Finfish off the Northeast Coast of the United States, 1985-1990 NOAA Tech. Report. NMFS 113, 66 p.

Paramor O.A.L. , Allen K.A., Aanesen M., Armstrong C., Hegland T., Le Quesne W., Piet G.J., Raakær J., Rogers S., van Hal R., van Hoof L.J.W., van Overzee H.M.J., and Frid C.L.J. (2009). *MEFEPO North Sea Atlas*: University of Liverpool.

Pennino, M. G., Illian, J., Muñoz, F., Simpson, D., Conesa, D., López-Quílez, A., and Bellido, J. M. (2012). Accounting for preferential sampling in the fishery context. In *International Statistical Ecology Conference (ISEC)*. Norway.

Planque, B., Loots, C., Petitgas, P., LindstrøM, U. L. F., and Vaz, S. (2011). Understanding what controls the spatial distribution of fish populations using a multi-model approach. *Fisheries Oceanography* **20**, 1-17.

Rijnsdorp, A. D., Peck, M. A., Engelhard, G. H., Möllmann, C., and Pinnegar, J. K. (2009). Resolving the effect of climate change on fish populations. *ICES Journal of Marine Science: Journal du Conseil* **66**, 1570-1583.

Rogers, S., and Stocks, R. (2001). North sea fish and fisheries. In *Strategic Environmental Assessment - SEA2*, Technical Report TR_003: CEFAS.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 319-392.

Schrödle, B., and Held, L. (2011). Spatio-temporal disease mapping using INLA. *Environmetrics* **22**, 725-734.

Simpson D, Illian JB, Lindgren F, Sørbye S, Rue H, Going off grid: Computationally efficient inference for log-Gaussian Cox processes**.** NTNU Technical report 10/2011

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583-639.

Telnes, K. (2012) *Cod - Gadus morhua*, Available at:
http://www.seawater.no/fauna/chordata/morhua.html. [Accessed 19 July 2012]

Wildscreen U.K. Charity. (2012) *ArKive: Gadus Morhua*, Available at:
http://www.arkive.org/species/ARK/fish/Gadus_morhua/more_info.html. [Accessed 19 July 2012]

Wilhelmsen, M., Dimakos, X. K., Husebø, T., and Fiskaaen, M. (2009). Bayesian Modelling of Credit Risk Using Integrated Nested Laplace Approximations. Oslo: Norwegian Computing Center.

6. APPENDIX

Additional Figures

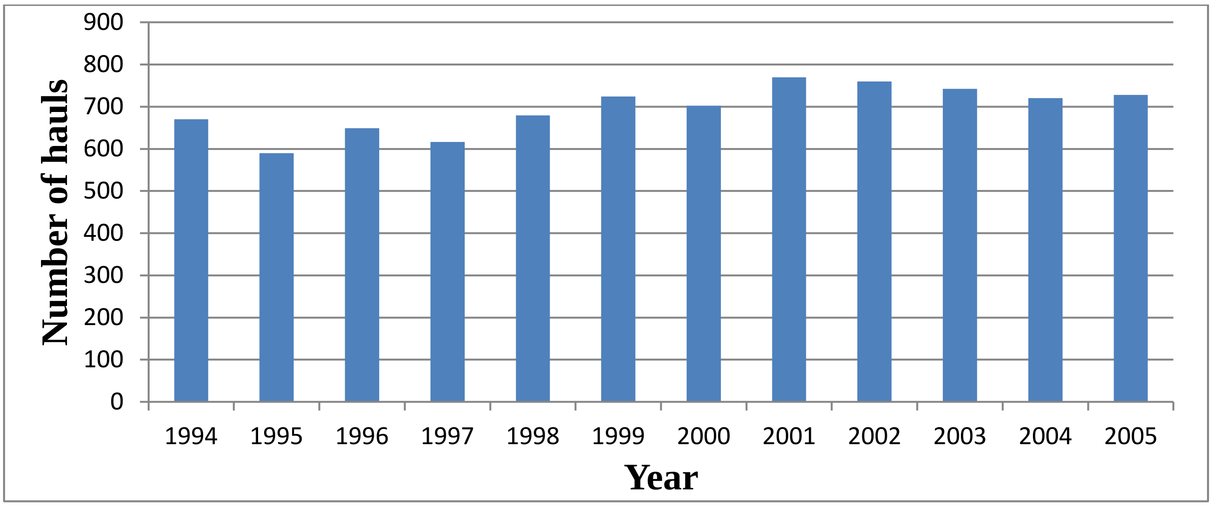
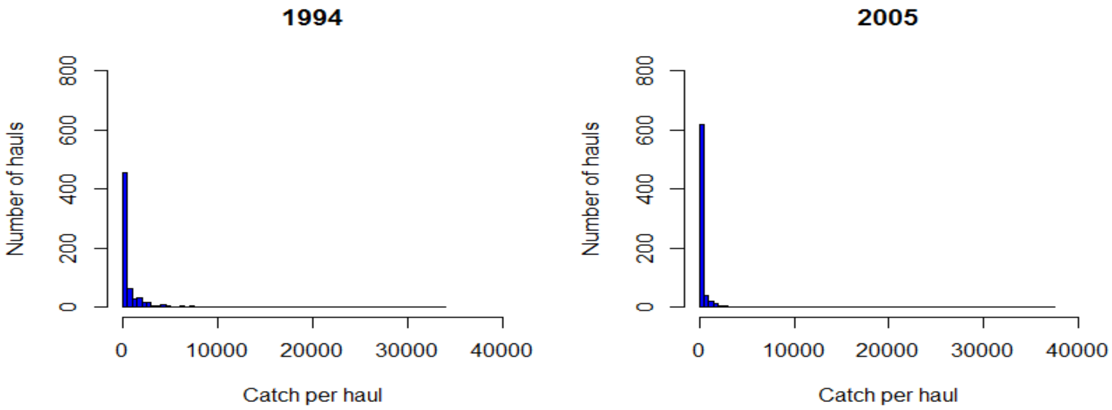
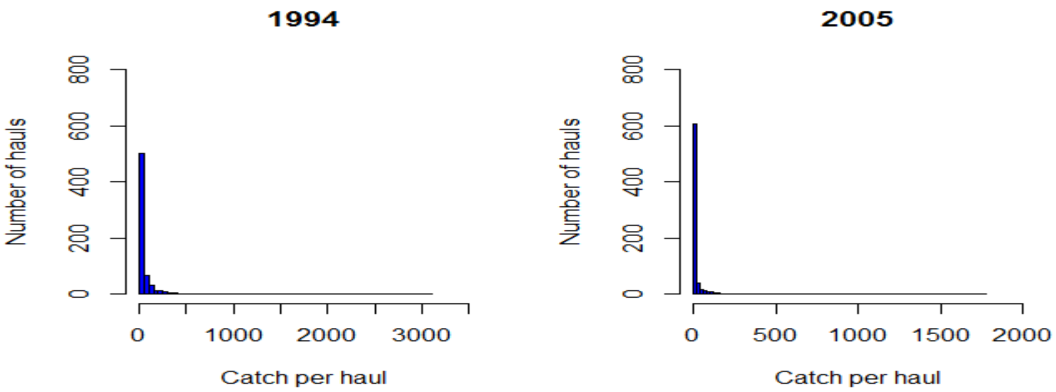


Figure A1: Number of hauls per year

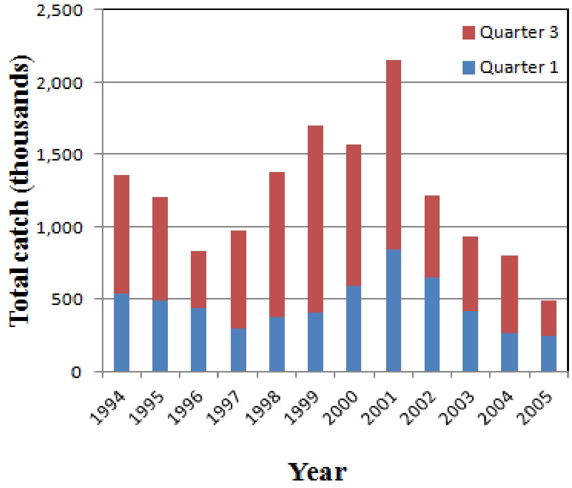


(a) Haddock

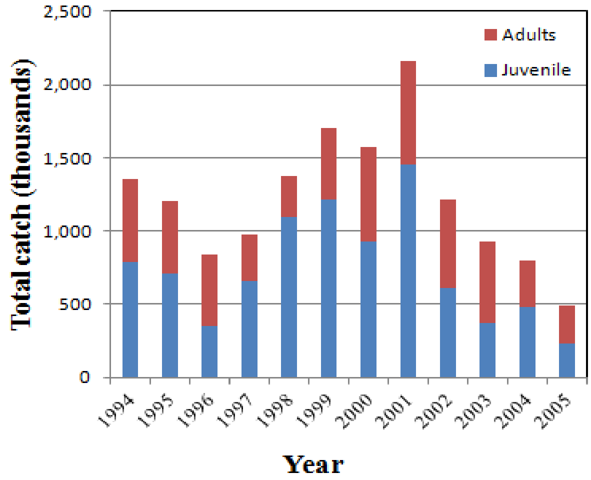


(b) Cod

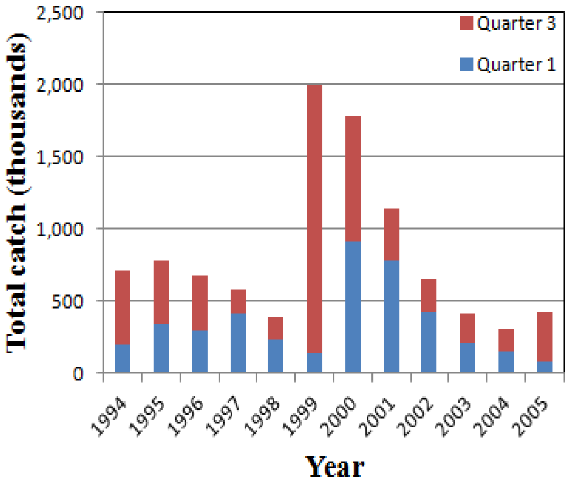
Figure A2: Distribution of catch in 1994 and 2005 by species



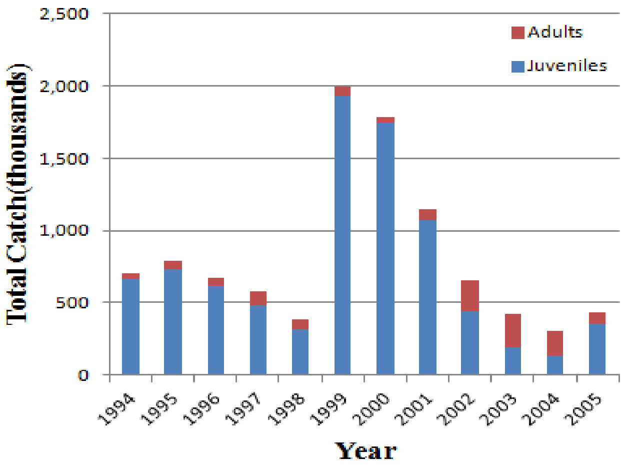
(a) Catches by quarter and year- Whiting



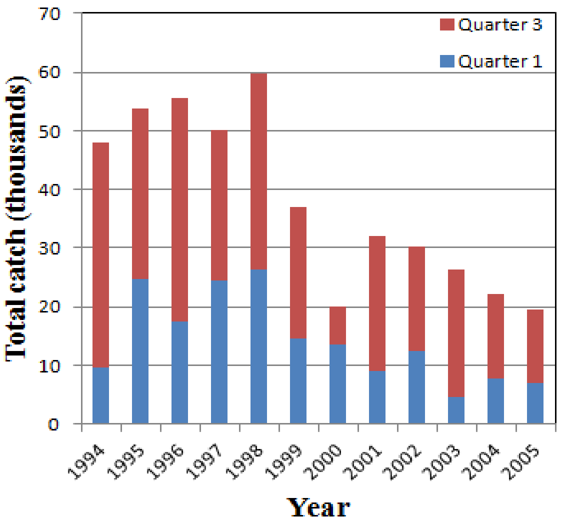
(b) Catches by age and year- Whiting



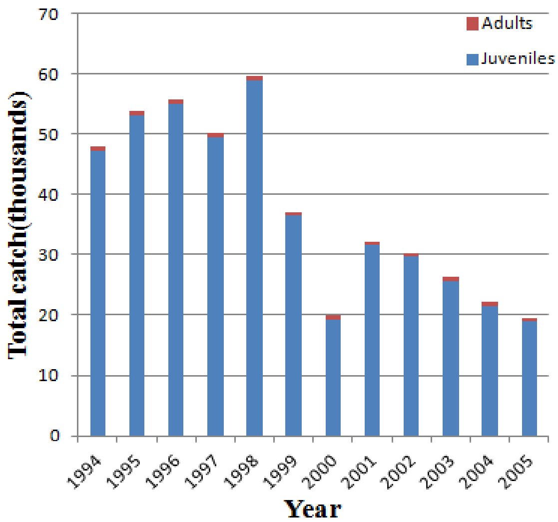
(a) Catches by quarter and year- Haddock



(b) Catches by age and year- Haddock



(c) Catches by quarter and year- Cod



(d) Catches by age and year-Cod

Figure A3: Distribution of total catch per year by species, quarter and age class

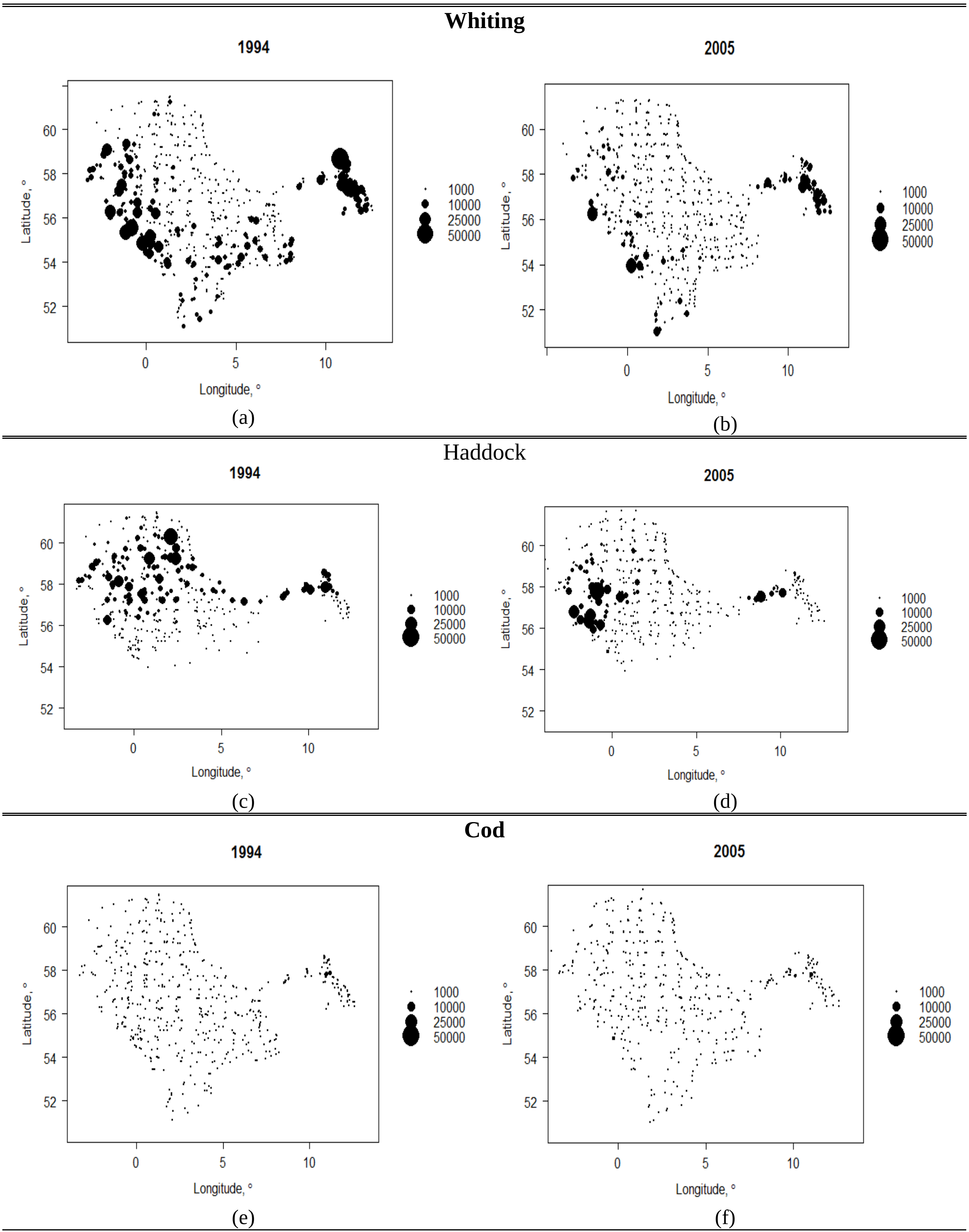
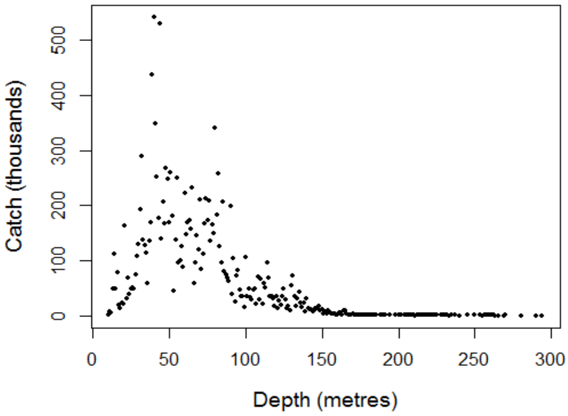
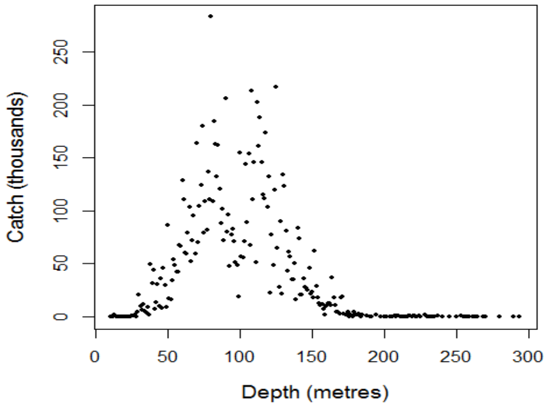


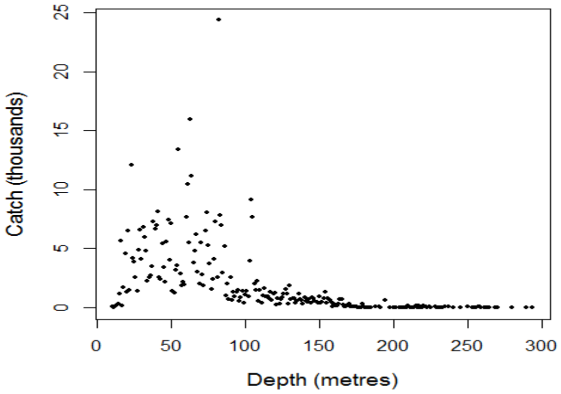
Figure A4: Plot of number of catches per unit of effort at a particular location, size of circle reflect catches



(a) Whiting



(b) Haddock



(c) Cod

Figure A5: Scatter plots of depth and total catch at a location

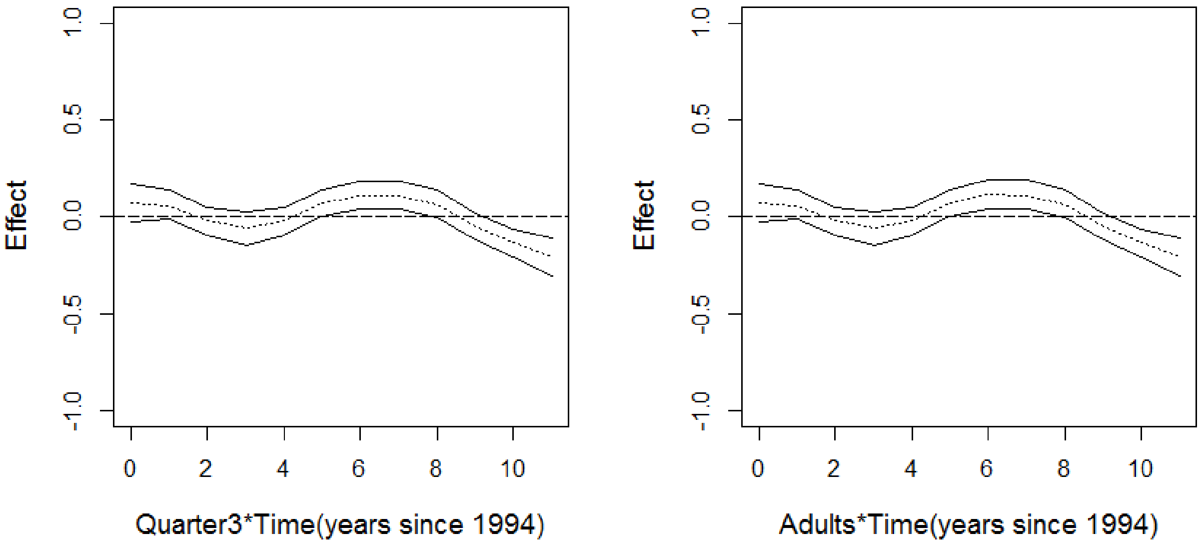
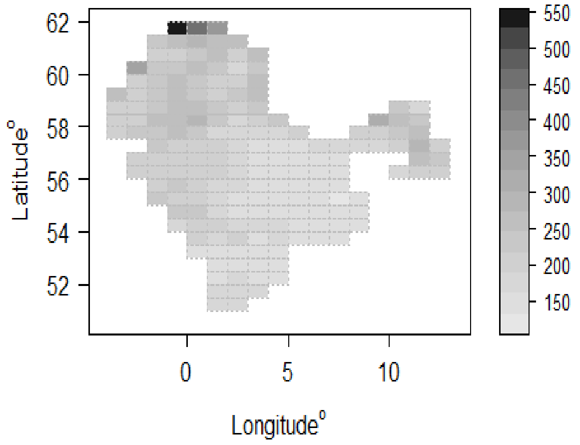
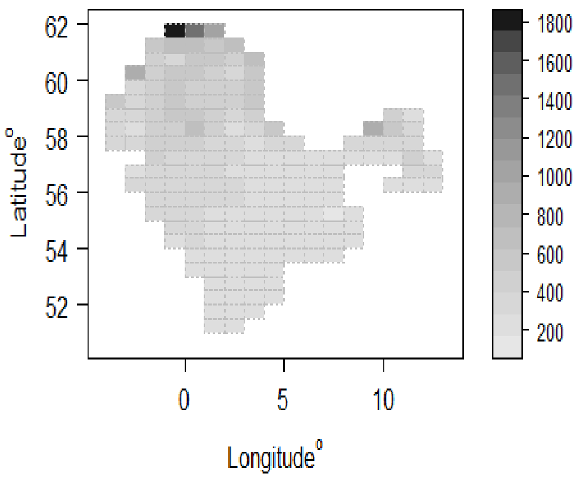


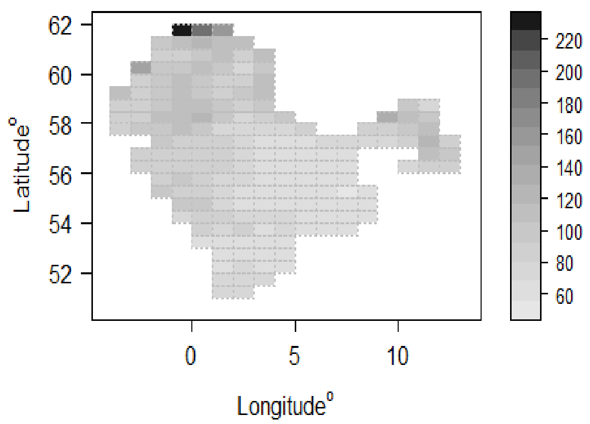
Figure A6: Smooth effect of interaction between time and age class, and time and quarter



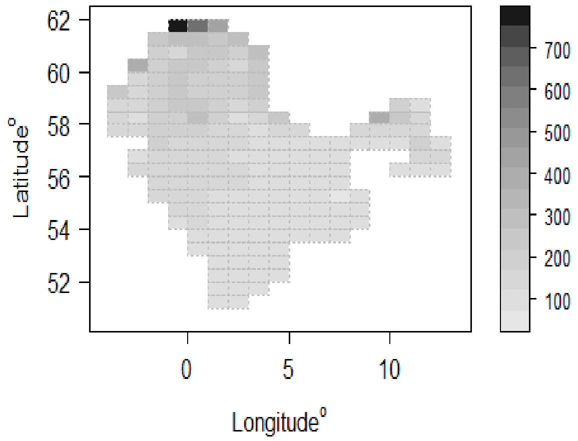
(a) Predicted juvenile catches 1994



(b) Predicted adult catches 1994

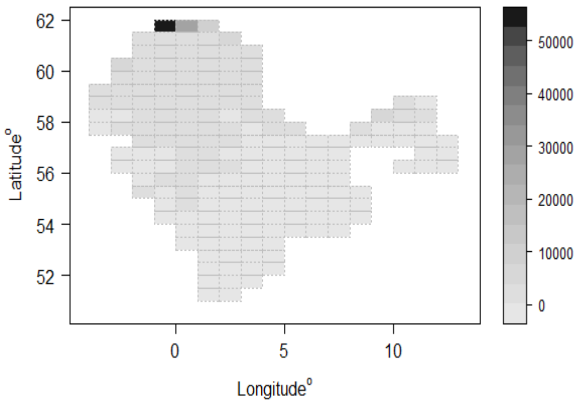


(c) Predicted juvenile catches 2005

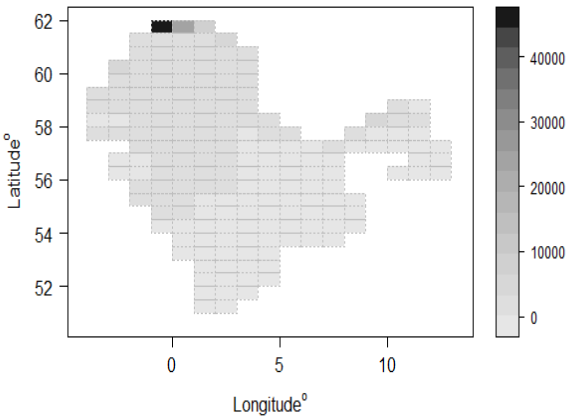


(d) Predicted adult catches 2005

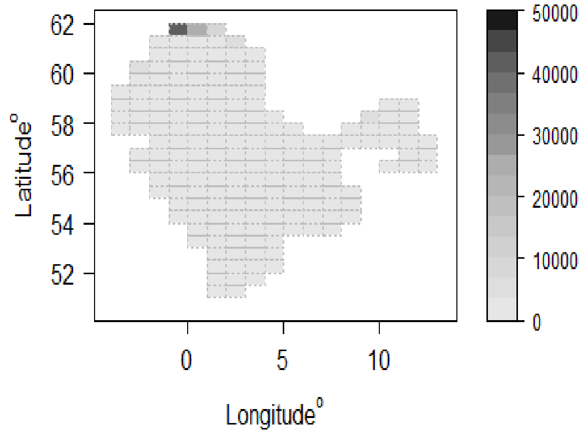
Figure 7: Predicted catches of Whiting by year and age class



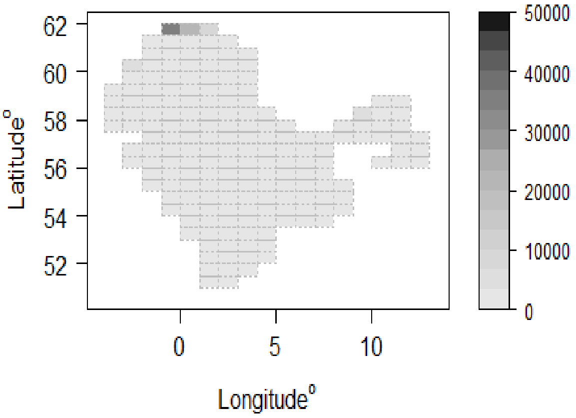
(a) Predicted catches in quarter 1, 1994



(b) Predicted catches in quarter 3, 1994

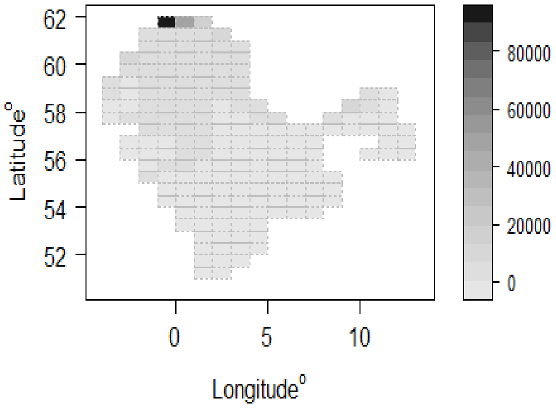


(c) Predicted catches in quarter 1, 2005

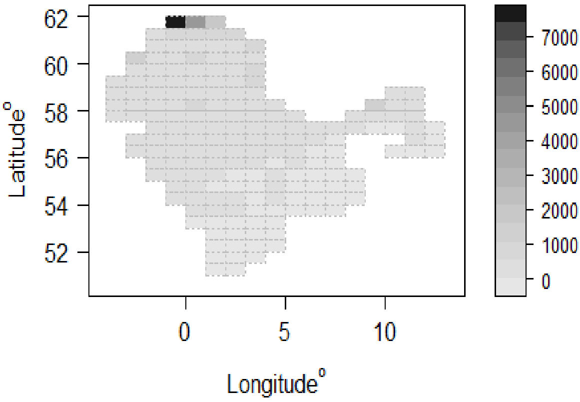


(d) Predicted catches in quarter 3, 2005

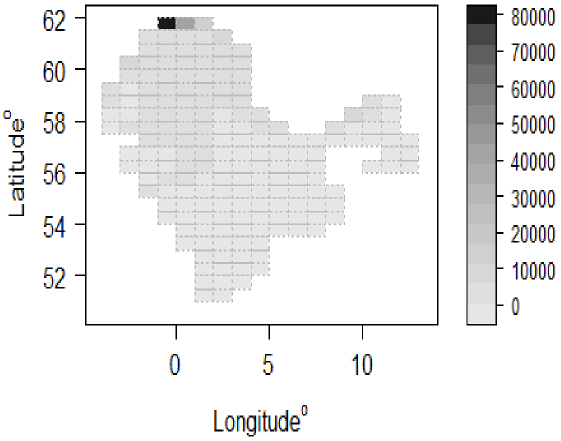
Figure A8: Predicted catches of Haddock by year and quarter



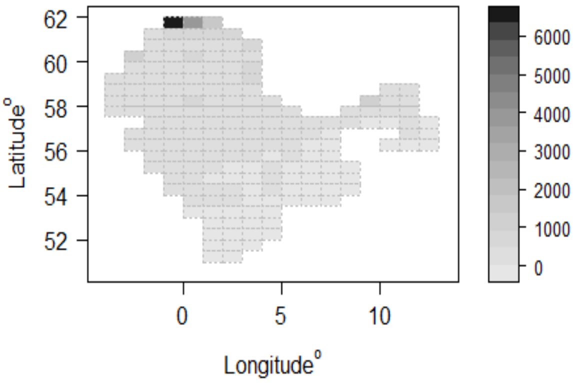
(a) Predicted juvenile catches 1994



(b) Predicted adult catches 1994

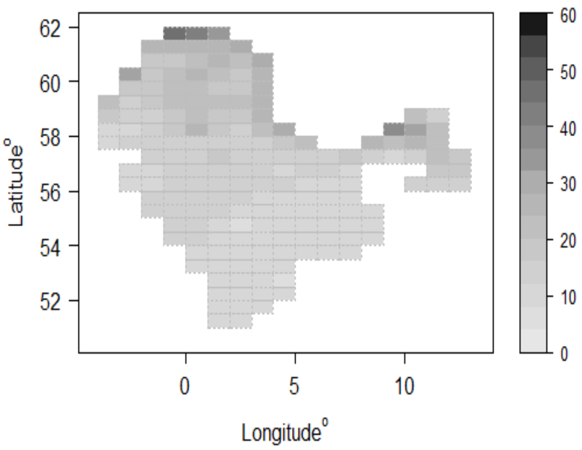


(c) Predicted juvenile catches 2005

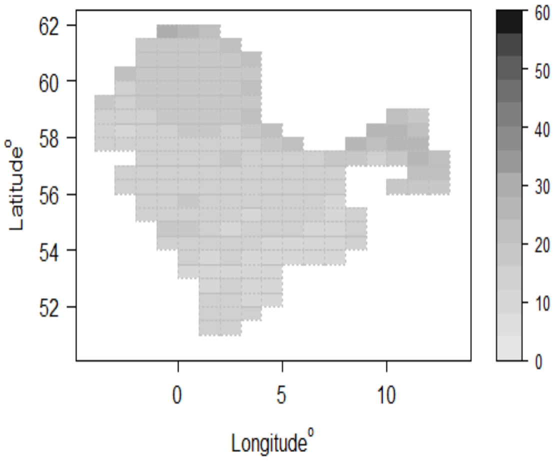


(d) Predicted adult catches 2005

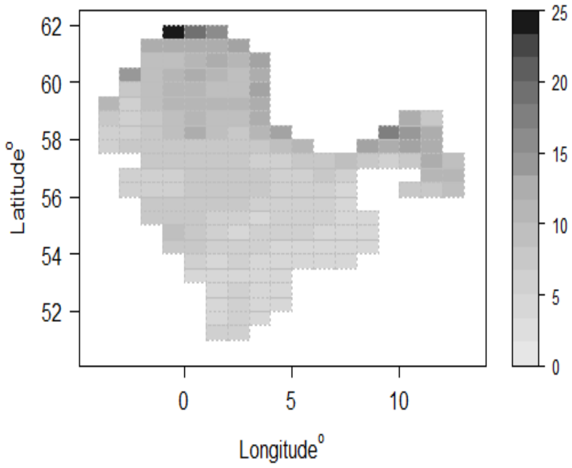
Figure A9: Predicted catches of Haddock by year and age class



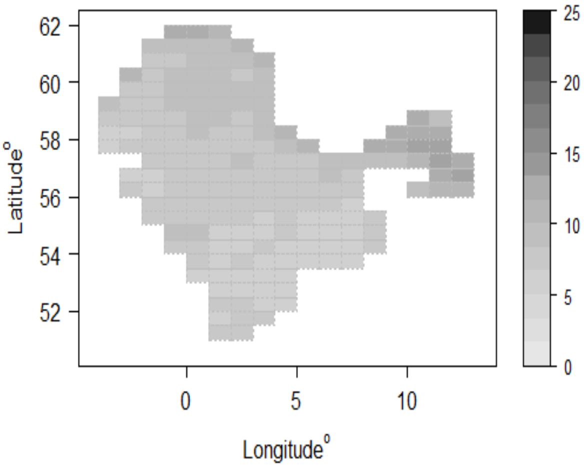
(a) Predicted catches in quarter 1, 1994



(b) Predicted catches in quarter 3, 1994

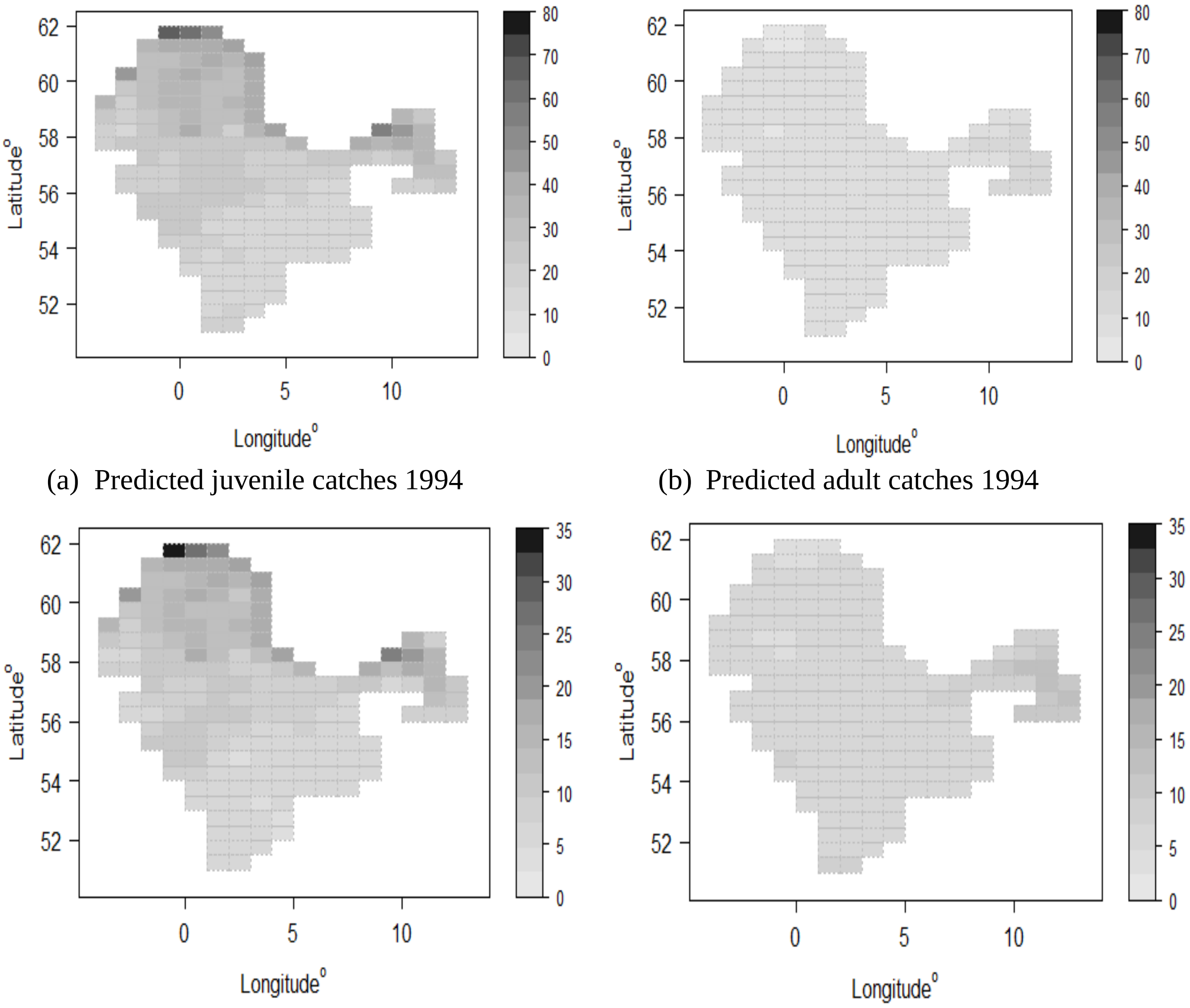


(c) Predicted catches in quarter 1, 2005



(d) Predicted catches in quarter 3, 2005

Figure A10: Predicted catches of Cod by year and quarter



(a) Predicted juvenile catches 1994
(b) Predicted adult catches 1994
(c) Predicted juvenile catches 2005
(d) Predicted adult catches 2005

Figure A11: Predicted catches of Cod by year and age class

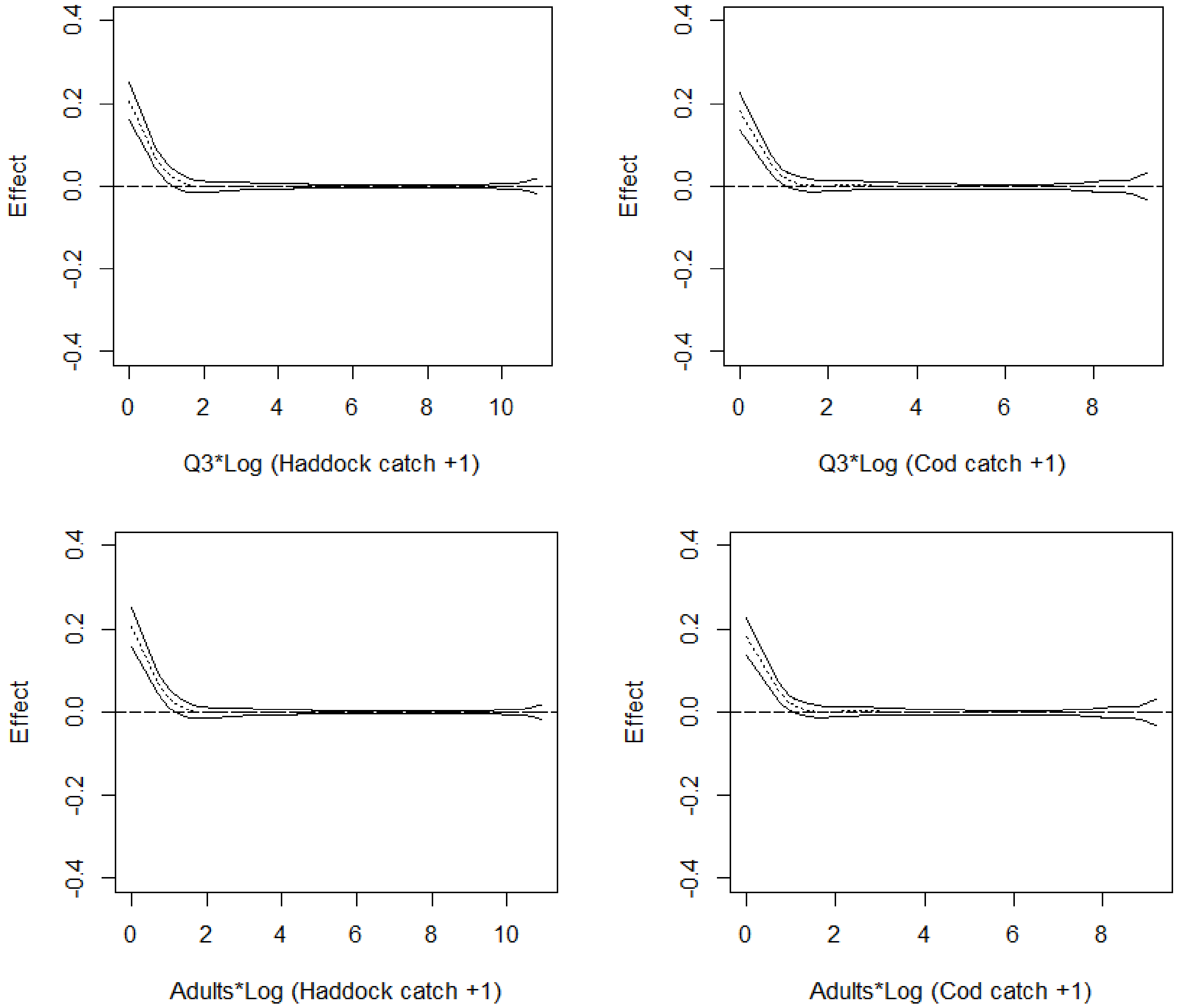


Figure A11: Smooth effects for the difference in effects of species for two quarters and age classes for relationship between whiting with other species

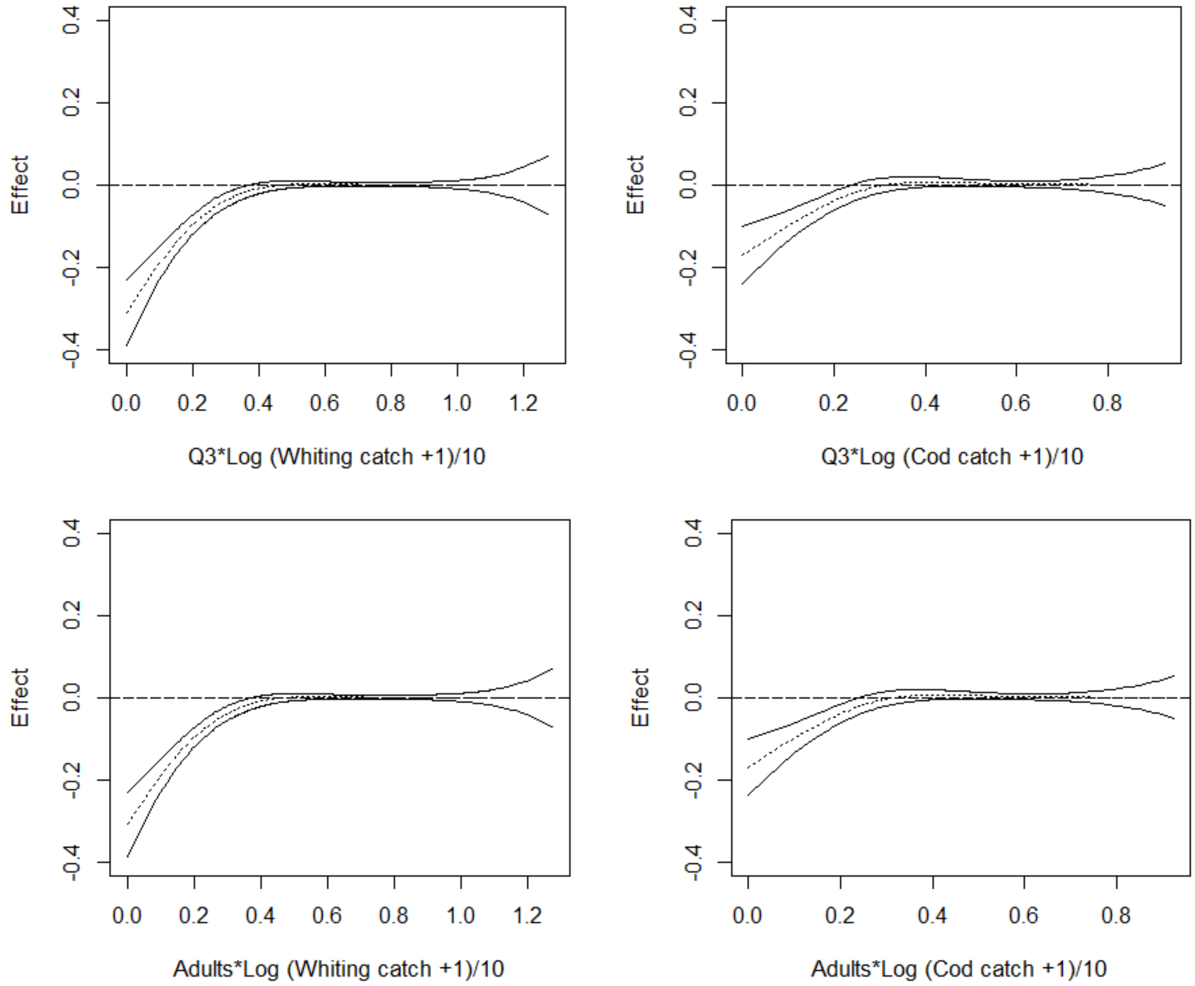


Figure A12: Smooth effects for the difference in effects of species for two quarters and age classes for relationship between haddock with other species

DIC values for selected models

Distribution	Species: DIC Whiting	Haddock	Cod
Lognormal *	204503.7	161637.8	76260.41
Negative binomial	204911.9	162810.3	77056.47
Zero Inflated NB	204914.2	162812.6	77059.59
Poisson	27440301	13249129	836478.8
Zero inflated Poisson	27088868	12797896	766136.2

*Adjusted for the log transformation

R-code for whiting model (3)

```
library(spdep)
library(foreign)
library(INLA)
setwd("C:\\Users\\Maxwell\\Documents\\HASSELT\\YEAR 2\\thesis\\CPUE")

# read shape files
nseacomb.shp<-  readShapeSpatial("nseacomb.shp", IDvar="ID")

# read whitting data
dwtemp<- read.dta("speciesint.dta")
attach(dwtemp)

# log of catch
dwtemp$log.whitcatch<- (log(dwtemp$whitcatch+1))
dic.whit.const=2*sum(dwtemp$log.whitcatch)

# add one column for the unstructured spatial effect
dwtemp$rec.unstr  =  dwtemp$recID

# rescale time
dwtemp$year1=year-1994

# bin quarter
dwtemp$Q.bin= ifelse(dwtemp$quarter>1, c(1),c(0))

# Interaction between quarter and year
dwtemp$Q.year <- dwtemp$year

# interaction between depth and year
dwtemp$D.year<- dwtemp$depth*dwtemp$year1

# generate binary age
dwtemp$A.bin= ifelse(dwtemp$age>1, c(1),c(0))

# interaction between age and quarter
dwtemp$Q.age= dwtemp$Q.bin*dwtemp$A.bin
```

```
# interaction between age and time
dwtemp$A.year= dwtemp$year1
```

Model for whiting

```
fw5 = log.catch ~ 1 +  depth*Q.bin +  f(year1, model="rw2")+depth*A.bin + f(A.year, model="rw2",
replicate=A.bin)   + f(Q.year, model="rw2", replicate=Q.bin) +
f(recID, model="besag",graph="nbcomb.adj")  + f(rec.unstr,model="iid")

m.fw5 = inla(fw5, data=dwtemp, control.fixed = list(prec.intercept = 0.001), family="gaussian",
control.inla=list(diagonal = 100), control.compute=list(dic=TRUE,cpo=TRUE), verbose=T)
```

#Variance estimates parameters: precision data stored in Stata format

```
#save precision data
write.dta(as.data.frame(m.fhs4$marginals.hyperpar$"Precision for the Gaussian observations"),
"Gprec.dta")

# convert precision into variance
gaus<- read.dta("Gprec.dta")
gaus.variance = inla.tmarginal(function(x) 1/x, gaus)
summary(gaus.variance)
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
spatial modeling of fish distributions

Richting: **Master of Statistics-Biostatistics**
Jaar: **2012**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Chirehwa, Maxwell Tawanda

Datum: **14/09/2012**