Spatio-temporal distribution patterns of North Sea herring: Analysing environmental drivers, global warming effects and density-dependent mechanisms with "Generalised Additive Models".

Dissertation

with the aim of achieving a doctoral degree at the Faculty of Mathematics,
Informatics and Natural Sciences

Department of Biology of the University of Hamburg

submitted by

Dominik Gloe

2018 in Hamburg

Institute for Hydrobiology and Fisheries Science

University of Hamburg

Olbersweg 24

22767 Hamburg

Germany

Evaluators:

Prof. Christian Möllmann

Prof. Axel Temming

Examination Committee:

Prof. Kathrin Dausmann

Prof. Christian Möllmann

Prof. Axel Temming

Prof. Myron Peck

Disputation: 21.09.2018

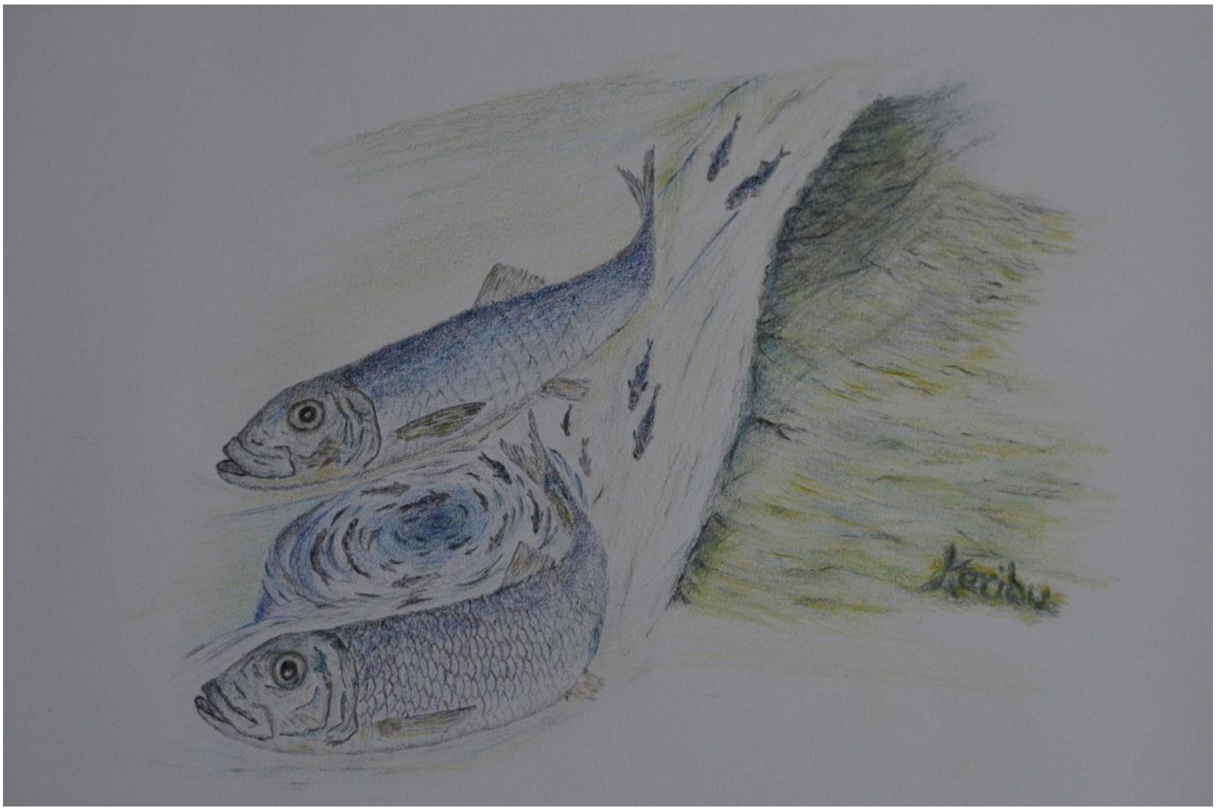
Keribu

Table of content

1. **Summary**

„Species Distribution Models“ (SDM’s) are powerful numerical tools to describe and explain distribution changes of species or populations. Analysing such changes is of great importance for understanding consequences of ecosystem changes from regional to global scales, i.e. ecosystem dynamics, varying species interactions or shifts in spatial distribution ranges. I used “Generalised Additive Models” (GAM’s) to analyse a spatio-temporal dataset of autumn spawning herring abundances for the entire North Sea for the period 1989 to 2009. The main purpose of my study was to find density-independent and density-dependent distribution changes of North Sea herring and additionally to describe environmental factors influencing the herring distribution.

In chapter 1, I showed how to test for and to deal with common statistical issues of complex spatio-temporal datasets, i.e. spatial autocorrelation (SAC) and zero-inflation (ZI). SAC was a serious problem in my herring dataset and was significantly reduced by implementing a 2-dimensional space smoother of the geographic coordinates (Latitude and Longitude) in the GAM’s. Additionally, I showed how to implement autocorrelation structures by adapting the GAM’s to mixed models. For dealing with ZI, I demonstrated the use of several distributions from the exponential family known to be able to deal high numbers of zero observations. In most cases, the simplest GAM’s with Gaussian distribution outperformed the complex models in my analysis. In conclusion, GAM’s are suitable for modelling spatio-temporal datasets to get ecological understanding of species distributions. The simplest GAM’s showed a very good performance and further adaptations can easily be performed, if necessary.

Chapter 2 consists of 3 different parts: (i) the description of temperature preferences of North Sea herring at age, (ii) the use of the developed GAM’s from chapter 1 to perform SDM’s with several environmental predictors and (iii) the analysis of potential herring distribution shifts in relation to the mean summer temperature and the “North Atlantic Oscillation” (NAO) winter index. I was able to demonstrate clear differences between immature (ages 0 and 1) and mature herring (older than age 2) concerning the temperature preferences. Mature herring showed clear preferences for cold water. These results reflected very well herring biology, as immature herring are located in coastal areas while adult herring start migrating to summer feeding areas in the northwestern North Sea after reaching maturity between age 2 and 4. Furthermore, these results approved the quality of the used oceanographic dataset. The SDM’s confirmed the clear differences between immature and mature herring, whereby herring of age 2 and 3 depicted an intermediate stage, attributable to my assumption of full maturity. Normally between 70 and 80 % of age 2 herring is mature and this introduced bias was obvious in my models. The most important environmental predictors of herring older than age 3 were sea surface temperature (SST) and bottom depth. Finally, I analysed shifts in herring distribution in relation to mean summer temperature and the NAO winter index. I related both predictors to the “Centre of Gravity” (COG), the weighted mean coordinates (Latitude and Longitude) of all mature herring age groups older than 3, based on the “Spawning Stock Biomass” (SSB). Mature North Sea herring showed no environmentally induced distribution shifts.

In chapter 3, I investigated density-dependent distribution patterns in North Sea herring. A set of 4 spatial indicators were used: (i) the amount of aggregation (Gini-Index), (ii) the occupied area (number of ICES statistical rectangles with positive abundances) and (iii) the geographical spread (Centre of Gravity and Inertia). I related the 4 spatial indicators again to the SSB of herring. Mature North Sea herring showed no density-dependent distribution patterns in the period 1989 to 2009.

To sum up, GAM’s were very suitable to model spatio-temporal datasets. Temperature turned out to be the most important environmental predictor, at least for mature herring. SDM’s for immature herring gave no satisfying results, probably due to the influence of tidal currents and river runoff in coastal waters. I could not find any shifts in the distribution of North Sea herring, neither density-independent nor density-dependent. None of the spatial indicators showed a relationship to the fluctuating SSB in the study period 1989 to 2009. Here, the overall stock status at relatively low levels since the collapse in the 1970s could be an important explanation. Obviously, no density dependent habitat selection occurred and no resource limitations existed in the study period. Relating these findings to a distribution theory like the “basin model” the stock size seemed to be below the carrying capacity of the north-western North Sea. North Sea herring showed no distribution shift as response to increasing temperatures. Here, the “conservatism” theory from Corten (2001) could be an explanation. Even when environmental parameters are slightly changing, herring tend to adhere to the existing migration pattern. Additionally, this effect could have masked an existing environmental effect in my SDM’s. Relating these findings to potential future impacts in respect to increasing temperatures, i.e. for the management of the North Sea stock, autumn spawning herring seems to be resilient to global warming.

2. Zusammenfassung

„Species Distribution Models“ sind numerische und oft eingesetzte Werkzeuge um die Verteilung von Arten und Populationen zu erklären und zu beschreiben. Solche Veränderungen zu analysieren und zu beschreiben ist von großer Wichtigkeit um die Auswirkungen von z.B. regionalen Veränderungen in Ökosystemen oder von globalen Klimaveränderungen zu verstehen. Hierzu gehören Dynamiken in Ökosystemen, Interaktionen diverser Arten oder Veränderungen in der räumlichen Verteilung von Arten. Ich habe „Generalised Additive Models“ (GAM's) verwendet, um räumliche Veränderungen in der Verteilung der herbstlaichenden Nordseeheringe in dem Zeitraum von 1989 bis 2009 zu untersuchen. Der Fokus bei den Untersuchungen waren hierbei dichteunabhängige und dichteabhängige Veränderungen der Verteilungsmuster, sowie die Beschreibung von Umweltfaktoren, die dieses Verteilungsmuster beeinflussen.

Im ersten Kapitel der vorliegenden Arbeit habe ich aufgezeigt, wie häufig vorkommende statistische Aspekte von räumlich-zeitlichen Datensätzen behandelt und aufgezeigt werden können. Der Fokus lag hierbei auf räumlicher Autokorrelation (SAC) und einer höher als erwarteten Anzahl von Nullobservationen (zero-inflation: ZI). SAC war in dem verwendeten Datensatz ein großes Problem und konnte durch die Verwendung eines 2-dimensionalen „Smoothers“ der räumlichen Koordinaten (Breite und Länge) in den GAM‘s signifikant reduziert werden. Zudem habe ich eine weitere Methode präsentiert, um SAC durch die Erweiterung zu „gemischten“ Modellen (GAMM) zu reduzieren. Hierbei wurden Autokorrelationsstrukturen in die GAM's implemetiert. Um ZI zu behandeln, wurden verschiedene statistische Verteilungen der exponentiellen Familie vorgestellt, die bekannt dafür sind, größere Anzahlen von Nullen händeln zu können. In den meisten Fällen waren die einfachsten GAM's mit Normalverteilung im Vergleich zu den komplexeren Modellen besser geeignet. Insgesamt können GAM's für das Analysieren von räumlich-zeitlichen Datensätzen überzeugen und geben sehr gute Einblicke in die ökologischen Zusammenhänge der Verteilung von Arten. Die einfachsten Modelle haben in den meisten Fällen die beste Performance gezeigt und spezielle Erweiterungen können sehr einfach eingebaut werden, wenn dies erforderlich sein sollte.

Im zweiten Kapitel wurden 3 verschiedene Themen behandelt: (i) die Temperaturpräferenzen der verschiedenen Altersklassen vom Nordseehering, (ii) die Verwendung der entwickelten GAM's aus dem ersten Kapitels um SDM's mit unterschiedlichen Umweltprädikatoren zu entwickeln und (iii) um räumliche Veränderungen der Verteilungsmuster des Nordseeherings in Relation zu den mittleren Sommertemperaturen und des „Nordatlantischen Oszillation“ (NAO) Winterindexes zu beschreiben. Es konnte demonstriert werden, dass klare Unterschiede in den Temperaturpräferenzen zwischen nicht-laichreifen (Altersklassen 0 und 1) und laichreifen (Altersklassen 2 bis 9) Heringen vorlagen.

Laichreife Heringe zeigten klare Präferenzen für kaltes Wasser. Diese Ergebnisse bestätigten das vorhandene Wissen über die Heringsbiologie. Nicht-laichreife Heringe kommen hauptsächlich in Küstengewässern vor, während laichreife Heringe migrieren und hauptsächlich in der nordwestlichen Nordsee vorkommen. Die Laichreife tritt in dem Alter zwischen 2 und 4 ein. Diese Ergebnisse zeigten zudem die sehr gute Qualität der verwendeten Daten um die Verteilung des Nordseeherings zu beschreiben. Die SDM's bestätigten die klaren Unterschiede zwischen nicht-laichreifen und laichreifen Heringen, wobei die Altersklassen 2 und 3 intermediäre Stadien beschrieben. Das kann durch die Annahme der totalen Laichreife dieser beiden Altersklassen erklärt werden, die normalerweise nicht vollständig erreicht werden kann. Für die Altersklasse 2 werden im Mittel Werte von 70 bis 80 % für die Laichreife angegeben. Die wichtigsten beeinflussenden Umweltvariablen für Heringe älter als 3 Jahre waren die die Oberflächentemperatur (SST) und die Bodentiefe. Als letztes habe ich Veränderungen in den Verteilungsmustern des Nordseehering in Relation zu der mittleren Wassertemperatur und des NAO Winterindexes untersucht. Hierbei wurden beide erklärenden Variablen in Relation zu dem „Centre of Gravity" (COG) gesetzt. Beim COG handelt es sich um die gewichtete mittlere räumliche Verteilung (mittlere Breite und Länge) aller laichreifen Heringe mit einem Alter größer als 3. Die laichreifen Heringe zeigten keine veränderten Verteilungsmuster in bezug auf die untersuchten Prädiktoren.

In Kapitel 3 habe ich dichteabhängige Veränderungen in den Verteilungsmustern des Nordseeherings untersucht. Hierfür habe ich 4 räumliche Indikatoren verwendet: (i) die Menge der Aggregationen (Gini-Index), (ii) das besetzte Gesamtgebiet (Anzahl der Gebiete mit positiven Abundanzen) und (iii) die geographische Ausbreitung (Centre of Gravity, Inertia). Die Indikatoren wurden auch hier in Relation zu der Gesamtabundanz der laichreifen Heringe betrachtet. Ich konnte keine dichteabhängigen Verteilungsmuster für den betrachteten Zeitraum von 1989 bis 2009 finden.

Abschließend kann ich sagen, dass GAM's eine hervorragende Methode zum analysieren räumlich-zeitlicher Datensätze sind. Die Wassertemperatur war die wichtigste beeinflussende Umweltvariable für laichreifen Hering. Die Ergebnisse der jungen Heringe der Altersklassen 0 und 1 waren nicht aussagekräftig. Dies kann mit der küstennahen Lokalisierung und den damit einhergehenden dynamischen Verhältnissen durch Tidenströme und Flusseinträge in diesen Gebieten erklärt werden. Ich konnte weder dichteunabhängige noch dichteabhängige Veränderungen der Verteilungsmuster des Nordseeherings beschreiben. Keiner der räumlichen Indikatoren konnte Relationen zu der fluktuierenden Gesamtabundanz der laichreifen Heringe für den Zeitraum 1989 bis 2009 nachweisen. Der Status des Nordseeheringbestands könnte hierfür eine Erklärung sein. Dieser hat sich seit dem Kollaps in den 1970er Jahren nicht erholt und ist insgesamt auf einem niedrigen Level. Offensichtlich tritt keine dichteabhängige Habitatauswahl durch limitierte Ressourcen auf. Bezieht man diese

Ergebnisse auf Verteilungstheorien wie beispielsweise das „basin model“, scheint die Größe des Heringsbestands unter der „carrying capacity“ der nordwestlichen Nordsee zu liegen. In meiner Analyse konnten auch keine Veränderungen der Verteilungsmuster als Reaktion zu steigenden Temperaturen belegt werden. Hierfür kann die „conservatism“ Theorie von Corten als Erklärung dienen. Corten beschrieb, dass selbst wenn Umweltparameter sich (leicht) verändern, der Nordseehering keine Anpassungen des Migrationsverhaltens zeigt. Zusätzlich kann angenommen werden, dass dieser Effekt maskierend in den SDM’s gewirkt haben könnte. Hierdurch könnte ein Nachweis möglicher Umwelteffekte auf die Verteilung des Herings erschwert worden sein. Bringt man diese Ergebnisse in Relation zu möglichen Effekten von ansteigenden Wassertemperaturen in der Zukunft, z.B. für das Management des Nordseeheringbestandes, scheinen die Herbstlaicher in der Nordsee sehr resilient gegenüber der Erderwärmung zu sein.

3. Outline of publications

D. Gloe, Otto, S. and C. Möllmann. Modelling species distributions in time and space – accounting for autocorrelation and zero-inflation in Generalised Additive Models. Manuscript.

D. Gloe, Floeter, J., Otto, S. and C. Möllmann. Influence of environmental and climatic changes on the spatial distribution of North Sea herring (*Clupea harengus*). Manuscript.

D. Gloe, Otto, S. and C. Möllmann. Density-dependent mechanisms in North Sea herring (*Clupea harengus*) distribution. Manuscript.

4. General Introduction

4.1 **Overview**

This thesis addresses the spatio-temporal distribution of autumn spawning herring in the North Sea. It consists of 3 chapters. In the first chapter, I will present how to apply a particular type of "Species Distribution Model" (SDM), more precisely "Generalised Additive Models" (GAM's), on comprehensive ecological spatio-temporal datasets. In the second chapter, the presented GAM's will be applied to the distribution data of North Sea herring with a focus on influencing environmental and climatic factors. The third chapter will address density-dependent mechanisms and how these impact the spatial distribution of herring. The high resolved spatio-temporal dataset is derived from acoustic surveys in the entire North Sea and covers the period from 1989 to 2009. Next, I will give a short general overview of the study area North Sea and describe threats to and recent changes of the North Sea fish community.

4.2 **Study area: The North Sea**

The North Sea is a north-western European shelf sea opened to the Atlantic Ocean and the Baltic Sea with a surface area of 575000 km² (Sherman and Hempel 2009). The North Sea is divided into a shallow southern part, the central and northern North Sea, the Norwegian Trench and the Skagerrak. The depth ranges from about 30 m in the south-eastern parts to 200 m in the north-western parts, with a maximum of 700 m in the Norwegian Deep and a mean depth of about 94 m (Sherman and Hempel 2009). The North Sea is a moderate productive ecosystem with highest production in the coastal areas due to riverine inputs. Additionally, it provides a diverse variety of habitats, e.g. fjords, estuaries, sandbanks, marches and deltas (Sherman and Hempel 2009). A high economic pressure impacts this marine ecosystem with its bordering countries Norway, Denmark, Germany, Netherlands, Belgium, France and the United Kingdom. The main pressures are shipping, wind energy, pipelines and cables, exploitation of non-living resources, mariculture and fishing (ICES 2016, OSPAR 2010). In the following, I will show the main threats for the North Sea fish community: increasing temperatures (global warming) and high fishing pressures.

4.3 **Threats to the North Sea fish community**

Climate change is known to have impacts on the North Sea fish community. The North Sea has warmed 1 – 2 °C since the 1980's and appears to be a "hot spot" for global warming (OSPAR 2010, Holt 2012). This water temperature rise is expected to lead to severe impacts on the marine ecosystems and their species, like phenological, biogeographic and community shifts (Beaugrand et al. 2014).The latter was already shown for the North Sea fish community. Diversity increased due to the introduction of more southern fish species and a distribution shift of endemic fish species to the

North occurred (Perry 2005). Dulvy (2008) described a deepening of demersal fish species as response to global warming in the North Sea. In addition, non-thermal effects like changes in the direction or location of ocean currents are expected to have impacts on the fish community (MarBef 2009). Fishing pressure also altered the community structure and foodweb in the North Sea, e.g. through depletion of large predators or species in general (ICES 2016). Furthermore, changes in size structures, trophic levels and size at maturation were observed (Teal 2011). All these examples show, how sensitive the North Sea ecosystem is to changes. All target species have specific preferences, are in (predator-prey) interactions with other species and therefore react differently to a changing ecosystem. Next, I will give an overview of the North Sea fish community and recent changes.

4.4 The North Sea fish community and recent changes

Key target species in North Sea fisheries are flatfishes such as plaice (*Pleuronectes platessa*), sole (*Solea solea*) and dab (*Limanda limanda*) in coastal areas and several gadoid species, like cod (*Gadus morhua*), Pollock (*Pollachius pollachius*), whiting (*Merlangius merlangus*) or haddock (*Melanogrammus aeglefinus*). Small pelagics such as herring (*Clupea harengus*), sprat (*Sprattus sprattus*), Norway pout (*Trisopterus esmarkii*) and sandeels (several species) have also importance for commercial fisheries (foreage fish) and furthermore as prey for larger fish species, seabirds and marine mammals (ICES 2017a, Muus and Nielsen 1999). Some major changes occurred in North Sea fish stocks in the last decades (Clark and Frid 2011). Landings of industrial (foreage) fish increased while there is a general decline in the biomass of these species. The abundances of most non-target species have increased since the 1970's. The "gadoid outburst" was characterised by a strong increase in gadoid species biomass and landings in the 1960's, peaking in the 1970's. This period was followed by a biomass decline until 1993. At the same time gadoids increased, the mackerel (*Scomber scombrus*) stock collapsed and has not recovered until now. Mackerel and horse mackerel (*Trachurus trachurus*) enter the North Sea in the meantime only seasonally during summer feeding. In the following I will start focusing on one of the most important key species in the North Sea: The autumn spawning North Sea herring.

4.5 Biology and characteristics of autumn spawning North Sea herring

Atlantic herring (*Clupea harengus*) is a key species in the North Sea ecosystem as it is commercially important for fisheries and important as prey species for large fish, seabirds and marine mammals (ICES 2014). The stock has undergone distinct changes since World War 2, with very high biomasses and landings followed by a complete depletion in the 1970's. Despite a closure of the herring fisheries and stock recovery, it never reached similar biomasses again (Simmonds 2007, Dickey-Collas et al. 2010). SSB fluctuated between 1.1 and 2.3 million tonnes from 1997 to 2016 (ICES 2017b)

Recently, the stock has full reproductive capacity and is in safe management limits, but produced until 2014 remarkable weak recruitment since 2002 (ICES 2017b, ICES 2014). The autumn spawning North Sea herring stock is distinguished into four main spawning components with distinct spawning sites and times. These are Orkney/Shetland, Buchan, Banks and Downs components (ICES 2014). The spawning takes place at limited and defined habitats as herring lay benthic eggs on gravel beds in depths of about 20 to 40 m (Cushing and Burd 1957; Parrish et al. 1959). Spawning starts in the northern North Sea with the Orkney/Shetland and Buchan components in August and September and the Banks population from August until October (Corten 2001, ICES 2014). After spawning, these 3 components migrate to the overwintering areas in the region of the Norwegian Trench. The southern Downs component spawns from November/December to January (Corten 2001, ICES 2014) and overwinters in the southern North Sea and the Norwegian Trench. Feeding starts in April and May in the north-eastern North Sea and continues in June and July in the north-western North Sea. During spring and summer feeding the components mix completely (Corten 2001). And this phase of complete mixing during summer feeding is exactly the period I was interested in. In the following I will give an overview of fisheries acoustics in general.

4.6 **Fisheries acoustics to estimate abundances**

Acoustic surveys are well established methods to assess pelagic fish stocks and are carried out for more than 20 stocks in European waters. This simple method to estimate fish abundances is very suitable since it covers the entire water column (but has a small dead zone at surface and bottom), provides continuous sampling with high resolution along the transects and has a large area coverage. The scientific echosounders should be calibrated with applied settings right before the survey to ensure comparability and quality of the outcomes. Echosounder settings are given for specific surveys and depend on, e.g. species, water depth or the frequency. Different frequencies are available, dependent on the research vessel equipment, and can be used in parallel if present. This has the great advantage of providing acoustic signatures of the target that can be used for preliminary species identifications (groups, not exact species). But this often only works in theory. The standard method for species identification during acoustic surveys is fishing. Identification hauls are done to determine the species composition and length frequencies in specific areas. In addition, subsamples are frozen to determine ages from otolith reading in the lab (if not already analysed on board if weather allowed). The species information are used to calculate abundances together with the main output of acoustic surveys, the “Nautical Area Scattering Coefficient” (NASC in $m^2/n.mi^2$). The NASC is a measure of fish density and is calculated (normally) for every nautical mile (n.mi) or other horizontal distances, the so called “Elementary Distance Sampling Unit” (EDSU). In the end, mean NASC-values are calculated for the investigated area and are partitioned to every target species (based on the species composition from the identification hauls). In a last step, abundances are calculated based on target strength (TS) equations for the species. The TS is the expected backscatter of a specific species, depending on the (quadratic) mean length (based on the length distributions from the identification hauls). TS equations are known for all target species, derived from tank experiments. This information together with the total covered area is used for abundance calculations. I tried to keep the descriptions very simple without formulas just to provide an overview of the method. To get more in touch with acoustics, I would like to recommend the acoustic bible “Fisheries Acoustics” from John Simmonds and David MacLennan (Blackwell Science 2005). In the next I will present the “Herring Acoustic Survey in the North Sea” (HERAS) in more detail.

4.7 International Herring Acoustic Surveys in the North Sea (HERAS)

To show the issues to deal with in the analysis of spatio-temporal data, I used a novel dataset from the “International Herring Acoustic Survey in the North Sea” (HERAS), coordinated by the International Council for the Exploration of the Sea (ICES). The “Working Group of International Pelagic Surveys” (WGIPS) are planning and coordinating HERAS. WGIPS supports the “Herring Assessment Working Group” (HAWG) with acoustic surveys abundance results for assessment. My study uses for the first time now the HERAS data from the period 1989 to 2009. The surveys are carried out annually in June and July to determine distribution and abundances at age of autumn spawning herring in the entire North Sea and adjacent waters. The survey design gives large area coverage and a high resolution of herring distribution. The transect spacing is 15 n.mi in most areas of the North Sea, but is reduced to 7.5 n.mi in areas with high densities and increased to 30 n.mi in areas where generally low abundances occur. As described above, NASC values, species compositions, appertaining length frequencies and ages are determined to calculate herring abundances at age. The entire North Sea is divided in ICES statistical rectangles with sizes of 1° Longitude by 0.5° Latitude. Herring abundances at age are calculated for each rectangle. All information that were given here about HERAS can be found in the “Manual for International Pelagic Surveys” (ICES 2015). In the following, I will give an overview of ecological data types for SDM’s and will describe potential statistical data issues that should be addressed when working with large spatio-temporal datasets, i.e. the HERAS data.

4.8 Ecological data to describe species distributions

Quality and characteristics of ecological datasets strongly depends on the possibilities available in the field and the applied sampling techniques. Three main types of data are common when considering spatial distribution of species (Elith and Leathwick 2009). Presence-only count data is often available for trees, birds or marine mammals. Here the number of counts is summed up on a regional scale but absence data is not available due to rarity of the species or limited resources to gather large datasets. Presence/absence data is more precisely to describe species distributions or occurrences and to link those to other factors. Presence/absence data consists of 0 and 1 (detection or non-detection in specific areas) and are collected during surveys were it is difficult to count the species numbers or the time or budget for each site is limited. This data type is very suitable for SDM’s even if some information is lost when no precise abundance data is available for the model, e.g. species aggregations. At best, abundance data is available to describe species distributions. While limitations of the applied sampling techniques must be known, abundance data gives the best requirement for SDM’s. Depending on the ecological questions behind the SDM’s, suitable datasets of explaining variables should exist.

4.9 **Ecological data issues of importance for SDM’s**

When working with ecological datasets in general, several data issues must be considered to avoid violating the statistical assumptions and to get reliable SDM results. Here, the most important are normality, heterogeneity, and independency of the data (Zuur 2010). Despite these basic statistical assumptions, I focused on 2 difficulties arising when working with complex, spatio-temporal datasets, i.e. autocorrelation (AC) and zero-inflation (ZI). Similarities of measurements that are temporally or spatially close together can lead to autocorrelation and must be handled in the SDM. Otherwise, the statistical assumption of independency can be violated (Legendre 1993). Spatial autocorrelation (SAC) can have 2 origins, (i) from “true” gradients (spatial dependency) or (ii) from “false” gradients (spatial correlation). I will treat both types as SAC in the following analysis as both have to be accounted for if present (Dormann 2007). Another commonly occurring issue in ecological datasets are large numbers of zero observations, i.e. zero-inflation. The zero observations can either origin from “real” ecological processes (“true zeros”) or from observer errors or sampling technique issue (“false zeros”). Nevertheless, ignoring ZI may lead to biased parameter estimates and standard errors (Zuur *et al.* 2010) and may cause overdispersion (Potts and Elith 2006). I will present ways to deal with autocorrelation and zero-inflation in “Generalised Additive Models” (GAM’s) in chapter 1 and will apply these methods on the HERAS data in chapter 2. Next, I’m going to introduce a special type of SDM, the “Generalised Additive Models”.

4.10 **A flexible option for SDM’s – Generalised Additive Models**

SDM’s are numerical tools to describe and explain factors that influence the distribution of species or populations. Hence, environmental estimates are very often related to species observations (Guisan and Zimmermann 2000, Elith and Leathwick 2009). One can chose between different types of SDM’s, e.g. BIOCLIM, Boosted Regression Trees, Support Vector Machines, Random Forest, Maxent, Generalised Linear Models (GLM's) or Bayesian Methods (Hijmans and Elith 2017). All those types have different advantages and disadvantages which I don't want to discuss here. Instead, I want to introduce "Generalised Additive Models" (GAM's; Hastie and Tibshirani 1990). GAM's are a very suitable type of SDM to describe and analyse species distribution because they offer a very high flexibility for modelling (Austin 2007, Elith and Leathwick 2009). Here, 2 characteristics of GAM's are very important: (i) the assumption of non-linear relationships between response and explaining variables (GAM's fit smoothers) and (ii) the implementation of any exponential family (*Poisson*, *Negative Binomial*, *Tweedie*) as error distribution. All those characteristics will be discussed in detail in chapter 1 of this thesis. An additional advantage of GAM's is its prevalence and hence a high order of support (from users and package developers). Furthermore, a high number of extensions make GAM's very suitable to model complex spatio-temporal datasets. All analysis in this thesis were done in the free software environment "*R*" (R Development Core Team 2013). One has the choice between many libraries (or packages) in R to fit GAM's to the data. But in the end, 2 packages outperform the rest. These are the packages "*gam*" and "*mgcv*". "*Gam*" is the package of the GAM-originators Hastie and Tibshirani (1990), the package "*mgcv*" is written by Simon Wood (2011). While "*gam*" was kept very simple, Simon Wood's package "*mgcv*" offers a higher amount of flexibility, e.g. it allows mixed models to implement autocorrelation structures. For this reason, all analysis were performed in the "*mgcv*" package. Now I will give an overview of the theory behind "Species Distribution Models".

4.11 **What influences distributions of species?**

"*Ecology is the scientific study of the distribution and abundance of organisms and the interactions that determine distribution and abundance*" (Begon 2006). There are 3 main levels in ecology. The level of organisms (and their interactions with the environment), the population level (presence or absence of species and abundance influencing factors) and the community level (organisation of ecological communities). This thesis addresses the population level, analysing the spatio-temporal distribution of the autumn spawning North Sea herring stock. The distribution of species is driven by individual requirements and tolerances, described as "niches". The modern concept of ecological niches was proposed by Hutchinson in the 1950's. He distinguished between the "fundamental niche" (range of conditions a species can survive) and the "realised niche" (range of conditions a species is found). The "fundamental niche" is larger, while the narrower "realised niche" is formed by various interactions and pressures of the ecosystem. These can be environmental or climatic (chapter 2) and competition on resources (intra- and interspecific; chapter 3 focuses on intraspecific mechanisms) and the presence of predators. Begon (2006) described the realised niche as follows: " … *the latter (realised niche) describes the more limited spectrum of conditions and resources that allow it to persist, even in the presence of competitors and predators*." In the end, definitions of niches are not consistent and appertaining types of models are neither. I will treat all chapters of my thesis as SDM's as proposed by Elith and Leathwick (2009). Overall, the distribution of species, or populations in the case of my studies of North Sea herring, is a complex interaction of diverse factors and SDM's are suitable tools to describe and explain these factors and relations.

4.12 **Aim of my studies**

The aim of my thesis is to present a comprehensive analysis of important drivers influencing the distribution of North Sea herring using a high resolution spatio-temporal dataset, regarding the age structure. In the first chapter I will present a methodological approach on how to analyse the spatio-temporal dataset with GAM's. Here the focus will be on treating the statistical issues of autocorrelation (temporal and spatial) and zero-inflation and the options GAM's provide. Chapter 1 is structured as a step-by-step guide making it possible for anyone to apply the methods. The developed GAM methods will be applied in chapter 2 to analyse environmental and climatic parameters that influence the distribution of North Sea herring. Several parameters, i.e. temperature, thermocline depth, salinity, zooplankton abundance and distribution, water depth, topography or climatic indices like the North Atlantic Oscillation (NAO) are known to influence the spatial distribution (Maravelias and Haralabous 1995, Maravelias and Reid 1995, Maravelias 1997, Bailey et al. 1998, Corten 2001, Volkenandt et al. 2015). Here we present for the first time an analysis of a time series of 21 years for the entire North Sea. In the last chapter, a comprehensive set of spatial indicators will be used to investigate density-dependent mechanisms for temporal changes in spatial distribution of autumn spawning North Sea herring. Knowledge is lacking about the importance of density-dependency. Therefore, I will provide insight in the intra-specific population dynamics for the first time. In summary, I will present:

I. A statistical method to analyse comprehensive spatio-temporal datasets with "Generalised Additive Models".
II. How to apply the GAM's (from chapter 1) to investigate environmental and climatic parameters that influence the spatial distribution of autumn spawning North Sea herring.
III. A study that investigates the importance of density-dependent mechanisms for the spatial distribution of North Sea herring.

5. Chapter 1

Modelling species distributions in time and space – accounting for autocorrelation and zero-inflation in Generalised Additive Models

5.1 Abstract

"Species Distribution Models" (SDM's) are powerful tools for understanding species interactions with the environment. Large spatio-temporal ecological datasets can give important insight in species behaviour but need advanced modelling techniques for the analysis. We here use a comprehensive spatio-temporal North Sea herring dataset for introducing adapted "Generalised Additive Models" (GAM's). We will show how to detect and deal with the statistical issues of autocorrelation (AC) and zero-inflation (ZI) when working with large ecological datasets. This will include the implementation of autocorrelation structures and multi-dimensional smoothers for AC and the usage of specific distributions for ZI. In our analysis, GAM's showed a high suitability for modelling our herring distribution data and we present a step-by-step guide for analysing ecological data with this powerful modelling technique. We recommend to keep the models as simple as possible. The simplest GAM's with Gaussian distribution and multi-dimensional space smoothers outperformed the complex models mostly in our study. If not, the simple GAM's can be used as a good starting point for gaining ecological insights and individual model adaptations can easily be performed.

5.2 Introduction

"Species Distribution Models" (SDM's) are widely used to provide understanding of species distributions in relation to spatial characteristics or environmental factors. The gained insights can be used to predict future distributions under a range of scenarios. Applications in terrestrial, freshwater and marine ecology are, among others, population management (Bailey *et al.* 1998, Augustin *et al.* 2013), climate change impacts (Cheung 2008, Gregory *et al.* 2012) or species interactions (Wisz *et al.* 2013).

The wide application of SDM's in various fields has been summarised by Elith and Leathwick (2009). A great number of modelling techniques are available, including (i) regression-based methods like Generalised Linear Models (Nelder and Wedderburn 1972), Generalised Additive Models (Hastie and Tibshirani 1990, Wood 2006) and Classification or Regression Trees (Breiman *et al.* 1984), (ii) Bayesian statistics or (iii) machine learning methods like Maximum Entropy Models (Phillips *et al.* 2004) or Boosted Regression Trees (Schapire 2003, Elith *et al.* 2008). Out of these, Generalised Additive Models (GAM's) are very popular in ecological modelling. In contrast to ordinary linear regression, generalised models allow error distributions other than Gaussian for the response variable and hence are more flexible. The model residuals can have the form of any exponential family distributions, e.g. Poisson, Negative Binomial or Tweedie. Additionally, GAM's assume a non-linear relationship between response and explaining variables. A smoother replaces the regression curve while the statistical assumptions are the same as for regression models. These characteristics offer great flexibility in describing the relation between response and predictor variables and makes GAM's more suitable for modelling species distributions instead of assuming linear or quadratic functions (Austin 2007, Elith and Leathwick 2009).

Modern sampling techniques and increasing availability of monitoring data lead to complex ecological datasets that need special handling when it comes to analysis. Spatio-temporal ecological data often inherit two major issues that need to be accounted for, i.e. zero-inflation (ZI) and autocorrelation (AC). A high number of zero observations can have different origins. In ZI "true zeros" and "false zeros" can be distinguished. True zeros occur where ecological processes lead to species absences, e.g. the habitat is simply unsuitable or competition is high. False zero counts are mainly due to sampling technique issues or observer errors (Martin *et al.* 2005). Consequently, ecological count data can comprise of high numbers of zero observations. ZI potentially exists if the number of zeros exceeds the expectations. The expectation depends strongly on the underlying statistical distribution. Ignoring ZI in the models can lead to bias in parameter estimates and standard errors (Zuur *et al.* 2010) and can cause overdispersion (Potts and Elith 2006). Therefore, it is important to account for zero-inflation if present.

Also two types of AC can be distinguished in spatio-temporal datasets. Temporal autocorrelation (TAC) is the association between observations over time while spatial autocorrelation (SAC) is the similaritiy of measurements that are geographically close together. Disregarding AC in a statistical analysis violates the assumption of independency and hence biases any significance test (Legendre 1993). Here it is necessary to underline the importance of the independent residual assumption. For an interesting discussion concerning SAC in the response variables and model residuals see Hawkins (2012) and Kühn and Dormann (2012).

A lot of extensions for GAM's exist to deal with the issues of ZI and AC. The number of options can be overwhelming for ecologists, especially if spatio-temporal datasets have to be analysed. Identifying ZI, SAC and TAC is not always intuitive during data exploration of such complex datasets. Concerning ZI, there is no clear definition how many zero observations makes a dataset zero-inflated and one has to evaluate this topic visually, e.g. with histograms. When it comes to AC, various visual methods and tests exist. These are autocorrelation function plots (acf, TAC), variograms (TAC, SAC), correlograms

(SAC), Moran’s I test or Mantel test (both for SAC). All of these tests have advantages, disadvantages and limitations depending on the dataset to use. Once ZI and AC are identified in the dataset, many options are available to deal with both issues. These are (i) data transformations, (ii) conversion to presence-absence data, (iii) the use of other distributions of the exponential family, (iv) particular zero-inflated distributions, (v) multi-dimensional space and time smoothers, (vi) extension of GAM’s to mixed models and implement autocorrelation structures or (vii) random effects. The combination of these methods is not possible for all options and all of these methods have limitations that could potentially influence the quality of the analysis. Here again, it depends strongly on the data and the hypothesis of the analysis which methods to use. A further challenge is to find the right computer program that offers the needed flexibility for a spatio-temporal dataset. One of the most and increasingly applied computing tools is presently the free software environment for statistical computing and graphics “R” (R Development Core Team 2013).

The aim of this study is to provide a practical guide for ecologists working with spatio-temporal data and GAM’s. We will present a step-by-step analysis from an initial ordinary GAM with abundances of North Sea herring (*Clupea harengus*) as response variable and one explanatory variable, in our case temperature. Successively we derive a suitable final model that deals with ZI and AC and can best explain spatio-temporal patterns of North Sea herring. The paper is divided into two parts. In the first part we present different solutions to test for and deal with TAC and SAC. The second part additionally deals with ZI. Based on such real datasets, we evaluate the performance of the different competing approaches. We will present criteria to evaluate model fit and performance, e.g. model diagnostics, 10-fold cross validation and prediction performance of the different models.

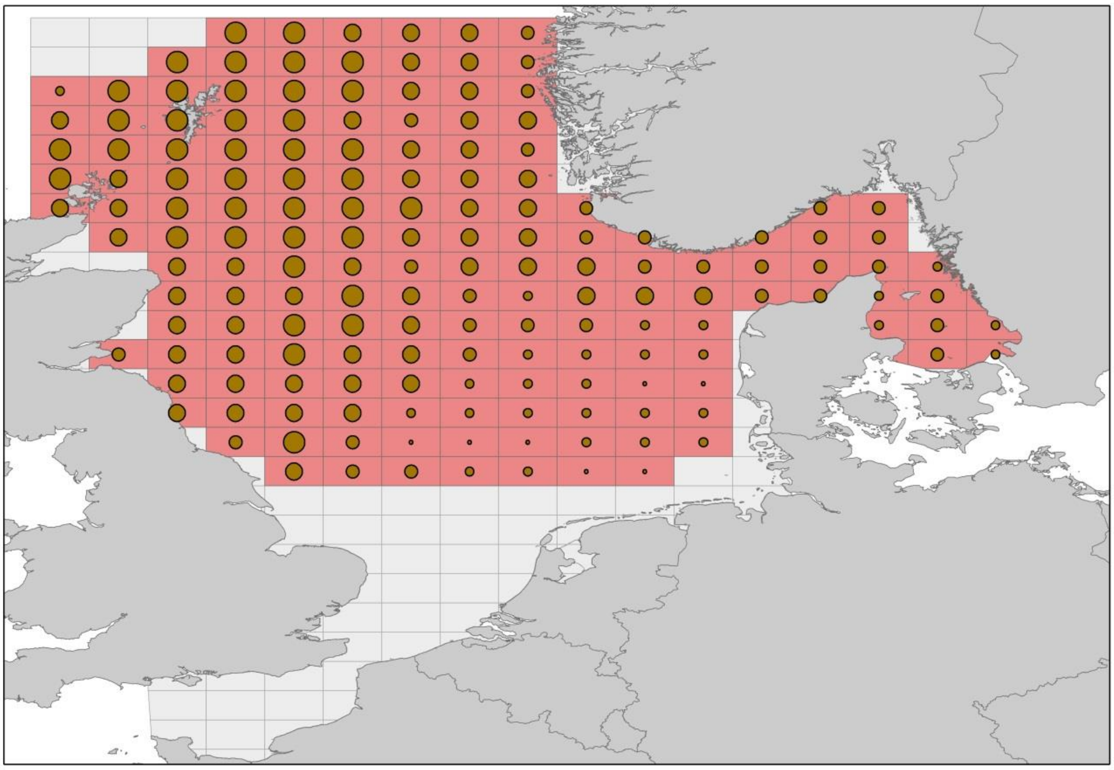

Figure 1: Map of the North Sea showing the sampling units, i.e. the 164 ICES rectangles (red). Bubbles represent the mean log-transformed abundances of all mature age-classes of autumn-spawning herring for the period 1989 to 2009.

5.3 **Material and Methods**

5.3.1. Data

We here used a novel dataset from the "International Herring Acoustic Survey in the North Sea" (HERAS), coordinated by the International Council for the Exploration of the Sea (ICES) to demonstrate the issues connected to zero-inflation (ZI) and autocorrelation (AC) in the analysis of spatio-temporal data. Our study uses for the first time the full HERAS data from the period 1989 to 2009. The survey is carried out annually in June and July to determine distribution and abundance of autumn spawning herring in the entire North Sea and adjacent waters. The acoustic survey design aims for a large area coverage and a high resolution of the herring distribution. In areas with echosounder detections, identification hauls are conducted to determine species composition and length frequencies. Outcomes from the surveys are numbers at age for immature and mature herring per ICES statistical rectangles.

Figure 1 shows an overview of the rectangles and the mean abundance of mature herring from the analysed period 1989 to 2009. The area coverage in the HERAS data was not consistent over the whole period due to bad weather conditions, technical problems or other causes in single years.

Additionally, the southern survey boundary changed from year to year due to uncertainty in the distribution limits of sprat (*Sprattus sprattus*; another target species of the survey) in the southern North Sea. Therefore, we analysed a selected number of 164 ICES rectangles to ensure spatial and temporal comparability of the data. Nevertheless, the number of analysed rectangles per year varied slightly. We used log-transformed numbers of mature herring as response for the analysis to reduce the variance of the abundances and the effect of outliers.

We used sea bottom temperature (SBT) derived from the HAMSOM (HAMburg Shelf Ocean Model) (Backhaus 1985, Pohlmann 1996), a three-dimensional baroclinic shelf sea model, as environmental forcing variable. The model was developed at the Institute of Oceanography, University of Hamburg. SBT was provided annually on the spatial resolution of the ICES statistical rectangles for the whole period. SBT represents means for the whole survey period (24.06. – 31.07.) each year.

In Part I of our analysis where we dealt with AC while ignoring ZI, we used the sums of the annual abundances of all mature age classes per rectangle (ages 2 – 9). After deleting the rows without HAMSOM data (NA) we got a dataset with 2590 total samples and 21.5% zero observations. In the analysis accounting for ZI in Part II, we only used the data from the oldest age class (age 9) with the lowest abundances and distribution range in general. This resulted in a dataset with 2590 observations in total, whereby 60% were zero values. The total dataset consisted of 20720 observations in total and 8779 zeros (42%).

5.3.2. Generalised Additive Modelling

We used the free software environment "R" 3.0.2 (R Development Core Team 2013) for statistical analysis. GAM's and GAMMs were run in the "mgcv" package (Wood 2011). GAMMs in "mgcv" implement autocorrelation structures and random effects from the package "nlme" (Pinheiro et al. 2014). Applying GAMMs is generally very stable but has some limitations that the user has to be aware of. These are for example the use of binomial data or the theta estimation for the negative binomial family in GAMM. For presence-absence data the use of the "gamm4" package (Wood and Scheipl 2013) is recommended, because penalized quasi-likelihood (PQL) iterations are used (Wood 2012). The dispersion parameter theta in the negative binomial family should be estimated in a comparable GAM and used as an approximation in the mixed model. Zero-inflated GAM's were performed in the "mgcv" package and in addition in the packages "gamlss" (Rigby and Stasinopoulos 2005) and „VGAM" (Yee and Wild 1996). The use of the latter is also associated with model limitations, e.g. no implementation of 2-dimensional interaction smooth terms or autocorrelation structures.

Simple GAM and choice of error distribution

The simple GAM approach uses the following equation:

$$y_{ij} = alpha + s(x_{ij}) + \varepsilon_{ij}, \tag{1}$$

where “alpha” is the model intercept, “s” is a smoothing functions of the explaining variables and “ε” is the error term or noise of the model. The error term depends on the underlying statistical distribution. All distributions from the exponential family have different characteristics that make them suitable for different applications and thus the model approach very flexible. Besides the Gaussian distribution that is well known from linear models, a very popular distribution for count data is Poisson. The Poisson distribution can have only positive integer values and allows for heterogeneity in the data and can handle more zeros than expected. Ecological count data is very often overdispersed. Overdispersion means that the variance is larger than the mean. For such cases, Quasi-Poisson is a better choice because an additional dispersion parameter is used to calculate the variance. Another distribution that is very common to model overdispersed ecological count data is the Negative Binomial distribution. This distribution has also a dispersion parameter that accounts for high variances in the data. At least, the Tweedie distribution can handle overdispersed count data very well. The use of Tweedie in “mgcv” is restricted to the so called “compound Poisson” distribution with $1 < p < 2$. A loop was used to find the best fitting p-value in 0.1 steps. For more details on Tweedie see Dunn and Smyth (2005). Overall, the choice of the error distribution depends strongly on the dataset to be analysed.

5.3.3. Zero-Inflation

An excessive number of zeros needs a special handling in statistical analysis. In general, zero-inflation potentially exists if the number of zeros exceeds the expected, depending on the statistical distribution. Ignoring can lead to bias in parameter estimates and standard errors (Zuur *et al.* 2010) and can cause overdispersion (Potts and Elith 2006). The already mentioned distributions from the exponential family that are suitable for overdispersed count data are a good starting point for analysis of data with a high number of zero observations. Opinions differ how to handle such datasets. Zuur et al. (2010) recommend the use of Poisson distribution prior to zero-inflated models under certain circumstances. Warton (2005) concluded that a Negative Binomial model fits the data well in most cases. Also a log-transformed Gaussian model gave a good model fit if many zeros are present. Augustin *et al.* (2013) used a Tweedie distribution with a p-value of 1.5 to account for a high percentage of zeros in their model.

Specific forms to deal with zero-inflation are zero-inflated (mixture, 2-step) or zero-altered (hurdle) models with adapted Poisson (ZIP, ZAP) or Negative Binomial (ZINB, ZANB) distributions. Both approaches use a binomial (logistic) model to describe the probability to get zero or positive response values and a second part that generates counts with the adapted distribution. The main difference between both approaches is that the count process in zero-altered models is zero-truncated and therefore cannot produce zeros. The zero-inflated model discriminates between true and false zeros. These types of models can be applied in the packages “gamlss” (Rigby and Stasinopoulos 2005) and „VGAM” (Yee and Wild 1996). The utilization is associated with model limitations, e.g. no implementation of 2-dimensional interaction smooth terms or autocorrelation structures. The “mgcv” package offers the option to use a zero-inflated Poisson distribution. This has the advantage that limitations for modelling are smaller compared to the alternative packages.

An alternative to the use of the mentioned distributions is logistic regression with binary data (Binomial distribution). Here, the abundances are converted to 1. This approach is very stable towards overdispersion but goes along with a loss of information about species quantities. As an alternative, it is not recommended to use presence-only data if absences are known. Only very few approaches use real presence-only data. For all other methods, including GAM’s, the generation of pseudo-absences should be performed. An evaluation of the full model with the real presence-only model could be done. For further reading on presence-only models Phillips *et al.* (2009) or Barbet-Massin *et al.* (2012) are recommended. However, presence-only models are easier to interpret in relation to more complex zero-inflated models.

5.3.4. Spatial and Temporal Autocorrelation

If spatial or temporal autocorrelation (SAC or TAC, respectively) is present in the model residuals, generally 2 methods can be distinguished, i.e. accounting for AC in the covariates or in the error term of the model. In GAM’s, the first method can be applied by using multidimensional smoothers for space and/or time. In GAMMs, both methods can be combined. The use of multidimensional smoothers can be extended by implementing autocorrelation structures (or random effects). In the next chapters, these methods will be presented in more detail. Furthermore, suitable methods to test for autocorrelation will be presented.

We checked for TAC in the model residuals with the autocorrelation function (acf). Acf-plots are a visual tool to detect temporal patterns in timeseries data. It is recommended to use timeseries data without missing values, but methods exist to deal with this problem (see Zuur *et al.* 2009). Another limitation is the use with spatio-temporal data. After extracting the model residuals for the whole model, every single rectangle has to be checked separately. We checked for TAC only for the rectangles with abundances for the whole period. Overall, this has been 47. For visual inspection of SAC in the model residuals, we used variograms. In variograms, semivariance is plotted as a function of distance (Bivand *et al.* 2008). Additionally, Moran’s I test statistic (Moran 1950) was calculated. Moran’s I is constructed similar to Pearson’s correlation coefficient (Borcard *et al.* 2011). Both methods have, similar to acf plots, limitations. The extracted model residuals have to be checked for every single year. We checked SAC for all 21 years.

Several smoothing types are available in “mgcv”. The default settings are thin plate regression splines, which were used to model the covariates. For these 1-dimensional smoothers we kept with the general recommendation from Keele (2008) to use a maximum number of 5 knots. The number of knots is the basis dimension to represent the smooth term and defines the amount and sensitivity of smoothing. For the multi-dimensional interaction smoother, we used tensor product smoother. As described by Wood (2016), this kind of smoother is especially useful for representing functions of covariates measured in different units. The 2-dimensional space smoother for Longitude and Latitude, were adopted from Beale *et al.* (2010). No limits for the number of knots were used. The 3-dimensional space-time smoother was adopted with the same settings for the number of knots from Augustin *et al.* (2013).

The extension of GAM’s to mixed models gives additional options to account for autocorrelation. GAMMs in “mgcv” implement autocorrelation structures (and random effects) from the package “nlme” (Pinheiro *et* al. 2014). The equation for the mixed model follows equation 1, with changes in the error term or noise that varies with autocorrelation structure. To find out the best fitting model, we tested several structure classes (corClasses in “nlme”) to account for the autocorrelation in the data. For TAC these were (i) a compound symmetry structure (corCompSymm), a simple temporal correlation structure suitable for short timeseries, (ii) an autoregressive process of order 1 (corAR1) and (iii) an autoregressive moving average process (corARMA). To account for SAC, all available correlation structures were tested. These were Gaussian (corGaus), linear (corLin), exponential (corExp), rational quadratic (corRatio) and a spherical (corSpher) correlation structure. The best performance was identified by comparing model AIC-values. It should be possible to combine corClasses in the model or to build an own one for a specific analysis.

Using mixed models with random effects may have an influence on the degree of autocorrelation. This depends strongly on the assumptions made for the analysis. If the implementation of random effects makes sense from an analytical viewpoint, it could have positive influence on autocorrelation in the model residuals. Generally, we recommend the use of specific autocorrelation structures instead of random effects.

5.3.5. Model comparison

We used the Akaike Information Criterion (AIC, Akaike 1974) to compare the various GAM's. It is not possible to compare GAM and GAMM models using AIC, because different methods are applied for calculation in the packages "mgcv" and "nlme". A solution to this problem could be to force "mgcv" to implement "nlme" also for simple GAM's by using always the "gamm" function. Instead, we used the model diagnostics (QQ-Plot and residual plots) and the explained deviance to compare model performances. However, the explained deviance is not given in the standard summary output of a GAMM. Hence, we additionally performed k-fold cross-validation to compare model performances. For every model, 80% of full data set was used as a training dataset, 10 times randomly shuffled. These training datasets were used to predict values for the remaining 20% of the data. Differences between the real and predicted values (rooted mean squared prediction error – RMSPE) from the 10 runs were then used to evaluate the performance of the model and to compare prediction performance between models. Predictions were done for the years 2008 and 2009, while the rest of the dataset was used for training. The prediction performance was evaluated visually.

5.4 **Results**

5.4.1. Part I: Aggregated herring data with autocorrelation

We first used the base model with temperature as explaining variable to test for temporal and spatial patterns according to the following equation:

$$\log(Abundance)_{ij} = alpha + s\left(Temperature_{ij}\right) + \varepsilon_{ij}, \tag{2}$$

We decided to use an "Additive Model" with Gaussian distribution and log-transformed abundance data. Compared to GAM's with Poisson, Negative Binomial and Tweedie distributions and with or without log-transformed response, the Gaussian model performed best. Except for the Tweedie model with log-transformed abundances, AIC values from the Additive Model were lowest. Additionally, the explained deviance by the model was highest and the model diagnostics from the Gaussian model appeared to be best. We will call the "Additive Model" with Gaussian distribution also GAM in the following. Only 10 of the analysed rectangles showed a slight TAC in the first lag. Figure 2 shows the frequency distribution of significant time lags, considering the 95% confidence limits. Hence, the model assumption of independency was valid. In contrast to temporal patterns, clear spatial patterns existed in the model residuals of the base model. Both Moran's I test statistic and variograms showed the same results. SAC was present in all years with varying degrees. We tried to reduce the temporal and particularly spatial patterns in the model residuals by implementing additional fixed terms (mainly smoothers), random effects or autocorrelation structures. Additionally, we combined these different approaches to evaluate whether that leads to a decrease of the residual pattern indicating autocorrelation. In total, we compared 10 alternative models (Table 1).

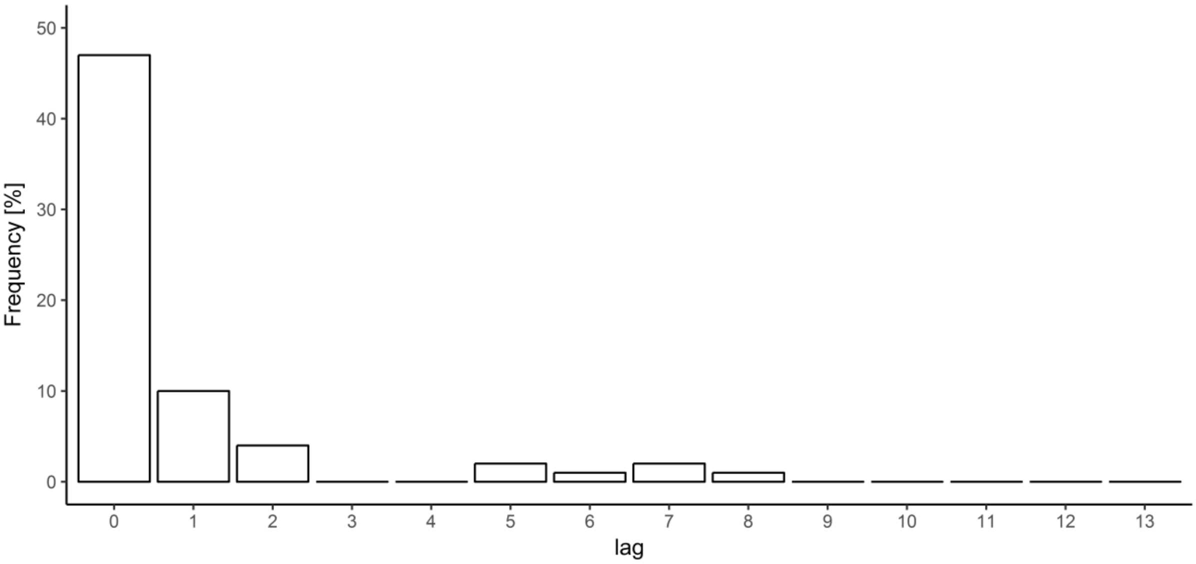

Figure 2: Frequency distribution of significant temporally autocorrelated lags in the aggregated data.

Table 1: Overview of models used to account for TAC and SAC in Part I of the analysis, including explained deviance in %, R-squared and prediction errors (RMSPE) from 10-fold cross-validation.

Model type	Components	Explained deviance [%]	R-sq. (adj.)	sign. Temp.	SAC	TAC	MSE	RMSPE	Model name R	Res R
GAM (Base)	$\log(Abund_{ij}) = \alpha + s(Temp_{ij}, k=5) + \varepsilon_{ij}$	19	0.189	***	yes	slightly	42.27	6.5	M1	Res_gam
GAM	Base + $te(Lon_j, Lat_j)$	39.7	0.391	***	no	very slightly	32.11	5.66	M2	Res2D_gam
GAM	Base + te(Lon, Lat) + s(Year)	42.2	0.405	***	no	very slightly	30.67	5.53	M6	Res2d_gam_1DYear
GAM	Base + te(Lon, Lat, Year)	48	0.46	***	no	slightly	29.27	5.4	M4	Res3D_gam
GAMM	Base + corARMA(Year)	NA	0.175	***		slightly	42.84	6.54	M12	Res_gamm_corARMA
GAMM	Base + RE(Year)	NA	0.188	***	no	slightly	42.21	6.49	M15	Res_gamm_RE
GAMM	Base + corExp(Lon, Lat)	NA	0.169	***	no	slightly	43.16	6.56	M3	Res_gamm
GAMM	Base + te(Lon, Lat) + corExp(Lon, Lat)	NA	0.381	***	no	very slightly	32.39	5.68	M8	Res2D_gamm_corexp
GAMM	Base + te(Lon, Lat) + corARMA(Year)	NA	0.386	***		very slightly	32.22	5.67	M13	Res2D_gamm_corARMA
GAMM	Base + te(Lon, Lat, Year) + corExp(Lon, Lat)	NA	0.424	***	no	slightly	30.3	5.5	M9	Res3D_gamm_corexp

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We tried to reduce the slight temporal and strong spatial patterns in the model residuals by implementing additional smoother and autocorrelation structures. Besides the simple GAM, we tested a 2D-smoother for Longitude and Latitude with and without year as additional smoother, a 3D-smoother of Longitude, Latitude and Year. Additionally, 2 different autocorrelation structures (spatial: “corExp” and temporal: “corARMA”). Concerning TAC, the ARMA structure with parameter value 1 for p and q gave the best results. The comparison of the diverse spatial autocorrelation structures showed best results for “corExp”. The base model and 2D- and 3D-smoother and autocorrelation structures have been combined to test for single and additive effects of the various approaches.

Figure 3 presents the ACF-plots for 6 different models for ICES rectangle 50F0. Rectangle 50F0 showed temporal patterns in lag 1. The effect of the different model types on TAC is very small. The 2D-smoother for Longitude and Latitude is the only model extension that had an effect on TAC. All other models, also the specific ARMA-model, had no impact on reducing temporal patterns. Overall, the model residuals showed no strong problems with TAC and can therefore be neglected in our following analysis.

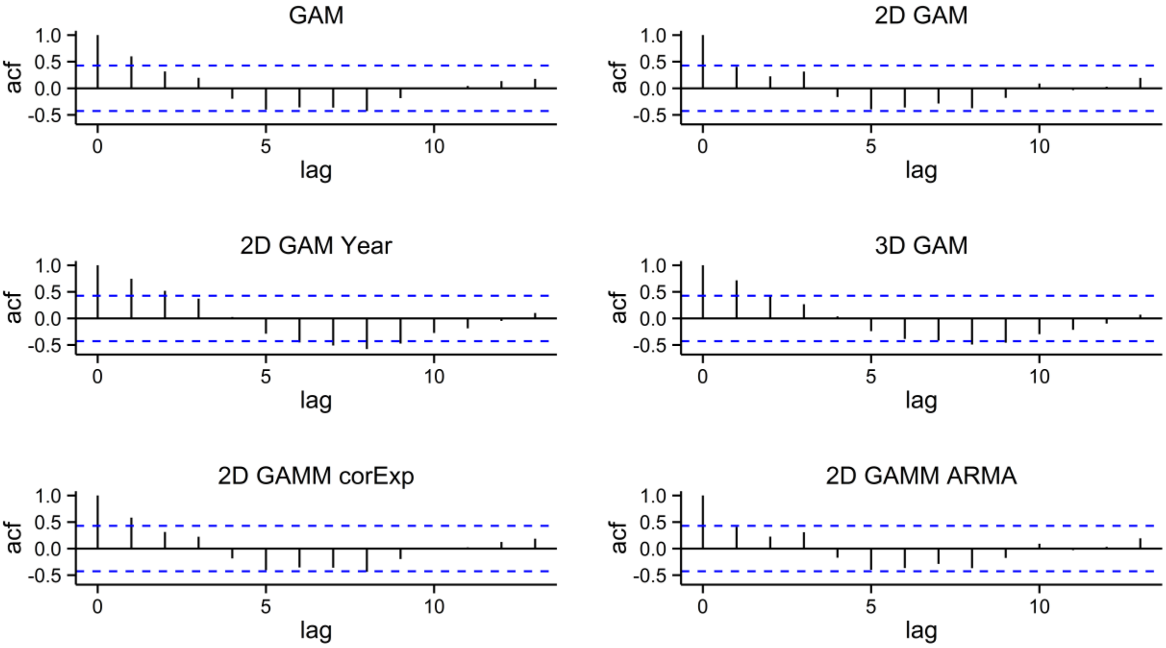

Figure 3: Autocorrelation function (acf) plots for temporal autocorrelation in ICES rectangle 50F0. Models with different complexity have only small effects on overall temporal autocorrelation (TAC).

Several models performed well to reduce the spatial patterns. The results for the Moran's I test statistic of all 10 models is presented in Figure 4. The simple GAM and the GAMM with spatial autocorrelation structure performed worst. The 2D- and 3D-smoother GAM's performed both very good. Figure 5 shows the variograms for the 4 best models. We found a clear reduction of SAC for most years. The years 1990, 1993 and 2007 showed only small residual patterns. These results were also indicated by the Moran's I test statistic.

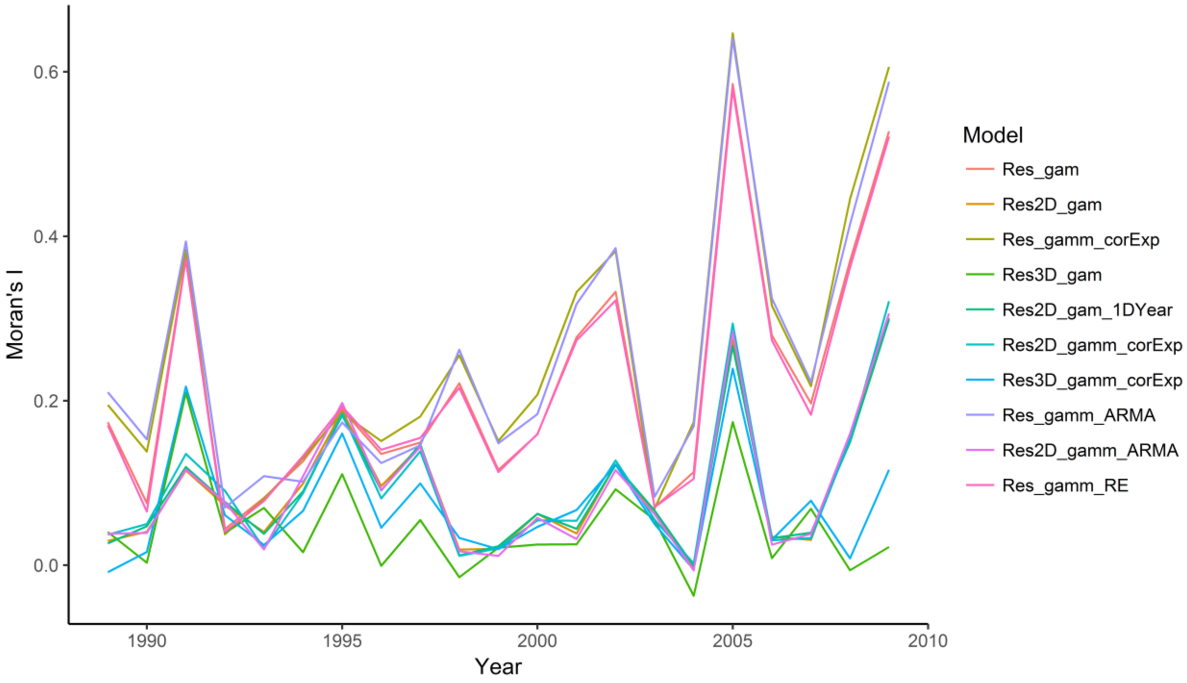

Figure 4: Moran's I for aggregated data and 10 different models. The 2D and 3D GAM's perform the best. No difference to the more complex GAMMs. The model components are listed in Table 1.

The results of the 10-fold cross validation are summarized in Table 1 (MSE, RMSPE). The 10 models had clearly different performances. As expected from the previous results, the base model and the GAMMs with single autocorrelation structure or random effect performed worse, indicated by the highest RMSPE values. All other models performed similarly, GAM's and GAMMs with 2D- or 3D-smoother had the lowest RMSPE values. We gained no better results by additionally using autocorrelation structures or random effects in the more complex GAMMs. The best model was the 3D GAM, followed by the 2D GAM with an additional smoother for Year.

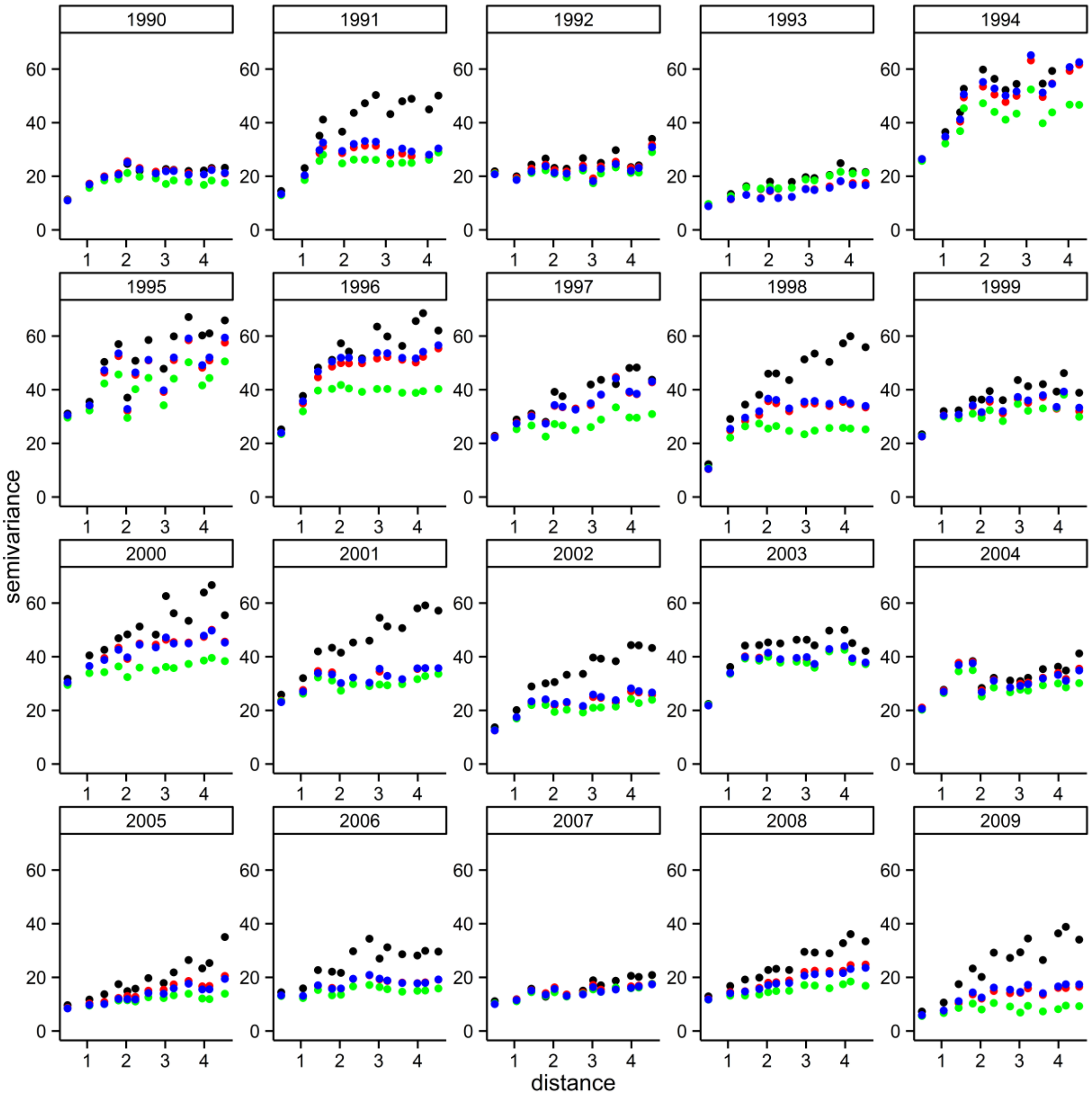

Figure 5: Variograms (semivariance) of model residuals for years 1990 to 2009. Coloured dots show changes in SAC of different model. Black – simple GAM, red – 2D GAM, green – 3D GAM and blue – 2D GAMM with autocorrelation structure "corExp".

When looking at the model predictions for the years 2008 and 2009 we obtained different results. The predicted abundances are given in Figure 6. In contrast to the cross-validation results, the 2D GAM without the Year smoother and the 2D GAMM gave the best predictions compared to the observations from 2008 and 2009. While the 2D GAM with Year smoother overestimated the abundances like the other models, the smallest deviations were given by the 2D GAM and 2D GAMM. This visual evaluation is validated by the single RMSPE estimates. The smallest values are produced by both models. Counter-intuitively, the 2D GAM with year smooth produced the worst predictions for both years.

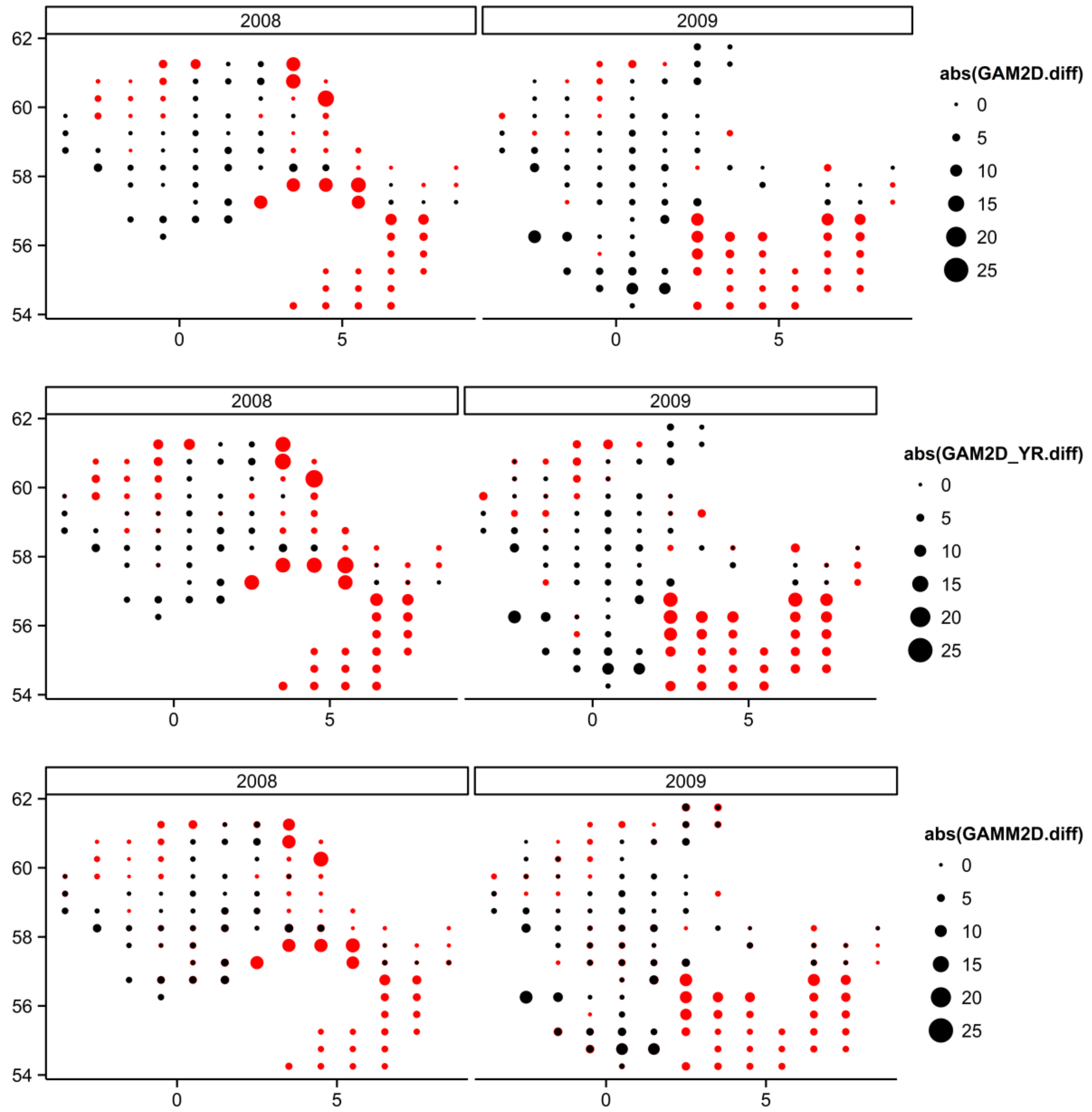

Figure 6: Differences between observations and predictions from the best models (GAM with 2D smoother, GAM with 2D smoother and year smoother, GAMM with 2D smoother). Black values represent underestimates while red dots show overestimations of the models.

The smoothers for temperature obtained from the best models were very similar and give the same results. All age classes together show temperature optima between 6 and 8 °C. All other temperature ranges show negative effects on herring abundance, with lowest abundances at temperature higher than 10 °C. The simple GAM demonstrates clearly this optimum curve. The models with multi-dimensional smoother to account for autocorrelation did not reveal such a clear drop with increasing temperatures compared to the simple GAM. The multi-dimensional smoothers show the same results. The highest herring abundances are found in the north-western North Sea, lowest abundances in the south-eastern parts of the North Sea.

5.4.2. Part II: Zero-inflated and autocorrelated data of age 9 herring

Figure 7 shows the histogram of the log-transformed abundances for the age 9 herring for the whole period. About 60% of the data points are zero values, while positive log-transformed abundances follow roughly a normal distribution. Several distributions from the exponential family are suitable to model zero-inflated count data. We again started with the base model with temperature as explaining variable. 5 distributions were compared, i.e. Gaussian, Poisson, Negative Binomial, Tweedie and zero-inflated Poisson. The lowest AIC was given by the zero-inflated model followed by Tweedie. These distributions would be our choice for analysing the dataset.

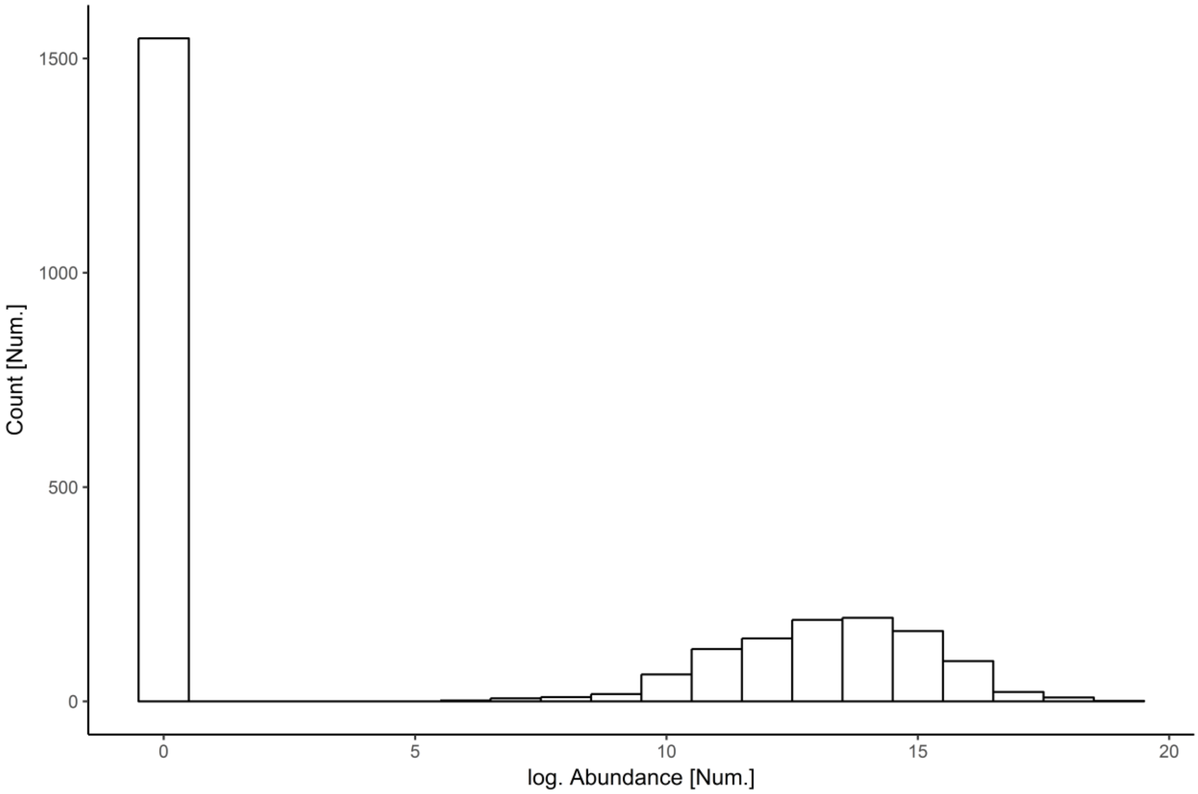

Figure 7: Histogram of the log-transformed abundances of age 9 herring. 60% of the data are zeros. Despite this, the log-transformed abundances follow roughly the normal distribution.

In a next step, we tested the base model residuals for autocorrelation. The results were similar to Part I. All models showed slight TAC and serious problems with SAC. We decided to use a 2D-smoother for Longitude and Latitude and an additional smoother for Year to account for the autocorrelation. This combination of fixed terms gave the best results in Part I of our analysis. Surprisingly, the Poisson, Negative Binomial and Tweedie distributions increased the temporal patterns in the model residuals compared to the base model. The Gaussian model increased TAC very slightly, while the zero-inflated Poisson model kept the level of TAC low. The 2D-smoother performed as expected from Part I. The amount of SAC was reduced efficiently.

For demonstration purposes, we compared all 5 models again after autocorrelation was dealt with. The results are summarised in Table 2. The zero-inflated Poisson and Tweedie distributions gave by far the lowest AIC. The least performing model included a Poisson distribution and resulted in an AIC about 2.5 times higher than the zero-inflated Poisson model. The explained deviance of Poisson, Tweedie and zero-inflated Poisson was between 43 and 44%, Negative Binomial only 33% and the highest explained deviance was given by the Gaussian model with 49%. In conclusion, zero-inflated Poisson and Tweedie appeared to be the most suitable distributions for modelling our zero-inflated dataset.

Overall, differences between models were relatively small. Due to the many zeros the model fit from all distributions was not optimal. The zero-inflated and Gaussian model showed the tendency to fit rather small values compared to the observed abundances. The Negative Binomial distribution showed the opposite effect and fitted rather high values.

Table 2: Overview of the models used to account for zero-inflation and autocorrelation in Part II of the analysis, including results of the 10-fold cross-validation. All models were based on the 2D GAM with an additional year smoother and log-transformed abundances.

Distribution	AIC	Explained deviance [%]	R-sq. (adj.)	sign. Temp.	MSE	RMSPE	presence-only AIC	presence-only RMSPE
Gaussian	15446	49	0.483	***	22.7	4.8	4167	1.77
Poisson	18838	44.1	0.479	***	23.1	4.8	4893	1.8
Negative Binomial	11148	33.3	0.392	***	26.8	5.2	4900	1.8
Tweedie	9281	43	0.478	***	23.1	4.8	4202	1.77
ziP	7043	42.9	---	***	22.3	4.7	---	---
Binomial	2152	40.1	0.447	***	---	---	---	---

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Considering the results of the 10-fold cross validation (Table 2), the differences between the distributions were very small with the exception of the Negative Binomial distribution. The prediction errors were clearly higher compared to all other models. The lowest RMSPEs were given by the zero-inflated model followed by the Gaussian, Poisson and Tweedie models. These results were approved by the visual inspection of the predictions for the years 2008 and 2009. No big differences could be seen between the distributions. The tendency of the Negative Binomial distribution to fit rather high values was validated. The prediction errors for year 2008 were obviously smaller than for 2009. Here, Gaussian, Tweedie and zero-inflated Poisson did the closest predictions (Figure 8).

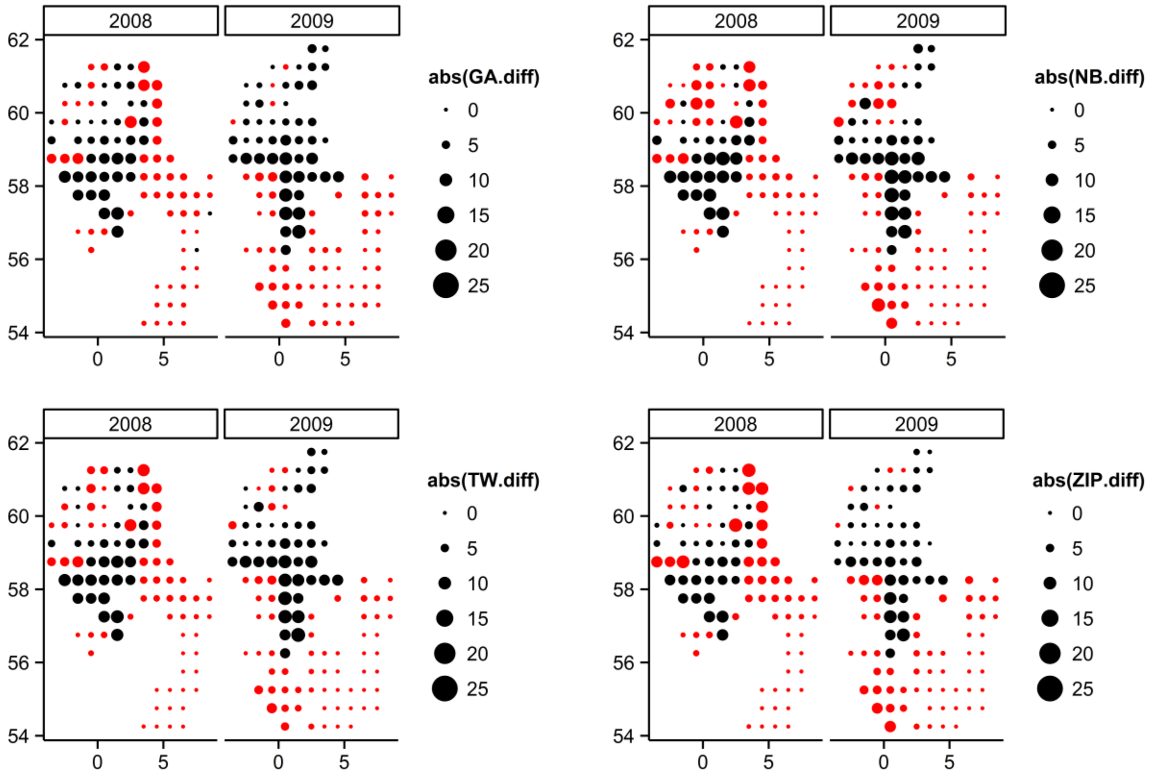

Figure 8: Differences between observations and predictions from the Gaussian, Negative Binomial, Tweedie and zero-inflated Poisson models. Black dots represent underestimates while red dots show overestimations of the models.

When comparing the binomial model and presence-only data, the binomial model gave similar results. The AIC was lowest, but due to the use of presence-absence data it cannot be compared directly to the other models. The explained deviance (40%) was in the range of the other models. We did no predictions and thus no RMSPE was calculated for this model. The presence-only models are also not directly comparable. The Gaussian presence-only model performed best. The AIC (4167) and RMSPE (1.77) were the lowest of all presence-only models. The Tweedie model was the second best with the same prediction performance but worse AIC (4202). Model diagnostics of all 4 presence-only models were comparable without big differences. Compared to the full dataset with all zero values, the model diagnostics looked much better for the presence-only models, which was unsurprising. The results from all presence-only models and the binomial model can also be found in Table 2.

The binomial model is a good starting point to compare the models related to the effect of the explaining variables on the herring abundance. As it was found for all age classes in Part I of our analysis, age 9 herring showed also a temperature optimum between 6 and 8 °C (positive effect on abundance) and a lower abundance with increasing temperature. The other models showed almost the same tendency but with different intensities. A very similar optimum curve was obtained by the Gaussian model. The zero-inflated Poisson model showed the same smoother curve but on another and smaller scale. The Tweedie and Poisson models showed also an optimum in the same temperature range, but overall the temperature effect was not as clear as for the Binomial, Gaussian and zero-inflated Poisson models. A totally different smoother curve was produced by the Negative Binomial Model. Here, the temperature optimum is in the very low and very high temperature ranges, that would not be expected for age 9 herring. The multi-dimensional smoothers for Longitude and Latitude showed different patterns. While the Gaussian model showed the expected pattern with high abundances in the north-western part and low abundances in the south-eastern part of the North Sea, this was only shown with lower intensity by the Negative Binomial and zero-inflated Poisson model. All other models showed only low abundances in the south-eastern part, but no clear positive abundance effect in the north-western part of the North Sea. The smoothers from the presence-only models gave the comparable results to the Binomial and zero-inflated models. The shape of the smoother curve changed a bit because in the presence-only dataset the temperature maximum was more than 2 °C lower compared with the full dataset. The higher abundance patterns in the north-western part and lower in the south-eastern part was approved by the presence-only models, too.

5.5 **Discussion**

In our analysis we showed ways to analyse spatio-temporal ecological datasets with GAM’s and GAMMs. In a first part we demonstrated the analysis of autocorrelated data, while the second part shows an analysis of zero-inflated and autocorrelated data. Using North Sea herring we provide a guide on how to deal with these properties of complex datasets. Generally, the performances of the models with different complexities were relatively small. Concerning autocorrelation, the simple solution to model the autocorrelation in the fixed term of the model in form of multi-dimensional smoothers appeared to be the best option to analyse our dataset. The use of Gaussian, Tweedie or zero-inflated Poisson models showed the best results in the context of zero-inflation.

We first used simple additive models with a Gaussian distribution as the base model for Part I of our analysis. The use of other distributions from the exponential families, e.g. Poisson, Negative Binomial or Tweedie brought no improvements in model performance. In the next step, we tested for autocorrelation in the model residuals. It is important to note that no method exist to test for TAC in a complete spatio-temporal dataset. We hence tested residuals of the time-series for each rectangle separately. Our dataset showed only slight temporal patterns in the acf plots for the first lag which were neglectable for the analysis. Despite that, we tried to reduce the temporal residual patterns with different methods. The simplest option was the implementation of a smoother for time, e.g. day of the year, month or year. In our case, the implementation of year as a smoother in the model gave no model improvements. Also the extension to GAMMs with specific temporal autocorrelation structures did not improve the model. Besides the “corAR1” structure, “corARMA” with different combinations of p and q-values were tested. The performance of the autocorrelation structures was probably due to the weak temporal patterns in our dataset and should not be seen as generally valid.

We do not recommended to use factors on a time-scale or random effects as alternatives to account for TAC. In long time-series, the use of factors goes along with a loss of many degrees of freedom and one may not be interested in the year effect. Random effects on a time-scale do not account for TAC. Instead, a correlation structure within the years would be implemented which was not the purpose of our analysis.

In contrast to temporal patterns, spatial patterns were strong in our aggregated dataset. Different methods to test SAC exist, but as for temporal patterns, none of these methods can cope with spatio-temporal data. The tests have to be performed separately for extracted residuals of each year from the full model. Variograms turned out to be a very suitable method and a lot of options exist in the statistical environment “R” to produce these. The Moran’s I statistic gives expectations and test values which makes the evaluation of the results very easy. Some authors, however, do not recommend the use of Moran’s I for model residuals of GAM and GAMMs. In our analysis, both tests showed the same results and appeared to be suitable methods. Two methods to deal with SAC are available for GAM’s and GAMMs: Implementation in the covariates or in the error term of the model (see Beale 2010 for more information). The last option can be realised only in mixed models (GAMMs). We found that the use of multi-dimensional tensor product smoothers worked most effective for our dataset. We tested a 2D-smoother for space (Longitude and Latitude) against a 3D-smoother of space and time (Longitude, Latitude, Year). The 3D-smoother model gave the best results, shown in the highest explained deviance and R-squared value. Additionally, the RMSPE was slightly lower compared to the 2D-smoother model. Despite that, we decided to use the model with a 2D-space-smoother and a 1D-time-smoother to keep model complexity low. The use of specific spatial autocorrelation structures gave no further model improvement. Both methods to reduce SAC can be combined in GAMMs. It is well known, that multi-dimensional smoothers are effective in reducing large-scale spatial patterns. If SAC on smaller spatial scales is accounted better with autocorrelation structures has to be tested individually.

It is important to highlight the evaluation of model performance when comparing GAM's and GAMMs. We compared the model diagnostics plots, the explained deviance as percentage of the model and the prediction error from a 10-fold cross-validation of the models. Furthermore, we compared the prediction performance for 2 selected years. The AIC, which is a common tool for comparing model performances, should not be used to compare GAM's and GAMMs. Different algorithms are used for calculation of AIC-values in "mgcv" if using the function "gam" (mgcv) or "gamm" (nlme, MASS). One should be careful to compare these kinds of models. Also the comparison of different distributions in GAMMs should not be performed with AIC.

The shape of the smoother-curve for Temperature changed due to the use of the spatial 2D-moother in the model. While the base model gave an "optimum-curve" for temperature, the effect of temperature seemed to be more positive in higher temperature-ranges in our best model. This effect was probably due to the temperature-effect with changing Latitudes and Longitudes. Here, the explaining variables Latitude and Longitude could be seen as proxies for temperature and probably also for other effects (unknown covariates). This could be the explanation for the changing smoother shape for temperature in the chosen model.

In the second analysis, we again started with the base model and temperature as only explaining variable. For our zero-inflated dataset we compared the performance of the different distributions of the exponential family. Different recommendations on how to cope with zero-inflated data are given in the literature. Among those are simple log-transformed Gaussian models, Poisson or Negative Binomial models or specific zero-inflated models. The latter include zero-inflated Poisson (ZIP), zero-inflated Negative Binomial (ZINB), zero-altered Poisson (ZAP) or zero-altered Negative Binomial (ZANB) distributions. The use of these specific distributions in GAM's is restricted to other packages than "mgcv" and goes along with limitations in the model flexibility. If, like in our case, autocorrelation is present in the data, these limitations are serious for the modelling approach, because multi-dimensional smoothers or autocorrelation structures are not available in these particular packages. For these reasons we restricted our analysis to the use of the "mgcv" package, including a zero-inflated Poisson modelling approach. The tests for autocorrelation showed the same result as for the aggregated dataset and we decided to use again the 2D-space-smoother and a 1D-time-smoother to account for the autocorrelation. Here again, the use of specific autocorrelation structures gave no model improvements as indicated by diagnostic plots and prediction error (RMSPE) from 10-fold cross validation. Therefore, we limited our analysis to GAM's which had the additional advantage to use AIC values to compare models more comfortable.

We cannot approve the frequent recommendations from literature to use Poisson or Negative Binomial distributions for zero-inflated data with our herring data. The best approach for our data was the simple Gaussian model, the Tweedie and zero-inflated Poisson models within the “mgcv” package, indicated by the already named methods. One limitation of our model approach is the lack of additional explaining variables. The use of temperature and a space-smoother is probably not sufficient to describe the distribution of age 9 herring. On the other side, suitable explaining variables are often sparse and the use of proxies is common. But the limitation of only 1 covariate is noticeable in the model diagnostics. The diagnostics plots are not optimal as it would be expected for a final model.

Nevertheless, our best models showed ecologically meaningful results. The temperature preference range and the core distribution area of herring obtained from the 2D-space-smoothers are biologically sensible. This result was validated by comparing the results from the “full” model with the presence-absence and presence-only models. Here, only the model output plots for the smoothers and the summary output significances are compared. Overall, all different model approaches showed the same results.

5.6 **Conclusion and general recommendations**

Probably it is not possible to transfer all our findings and recommendations to ecological count datasets in general. But we presented options and alternatives on very important topics when analysing ecological datasets: autocorrelation and zero-inflation. Not accounting for these important issues violates important statistical assumptions and increases the probability of Type I or Type II errors. Our analysis importantly showed that in most cases, it is the best choice to use the simplest modelling approach instead of using fancy distributions or extending to a very complex model. The use of simple Additive Models with Gaussian distribution and log-transformed response variables gave often similar results than more adapted models. The differences in model diagnostics, prediction performance and significance levels of the explaining variables were neglectable. One should exploit the great flexibility GAM’s offers in the “mgcv” package concerning distribution to fit the data, the use of multi-dimensional smoothers or specific autocorrelation structures. But one should restrain to the simplest model possible to describe the dataset that have to be analysed. Overall, we recommend log-transformed Gaussian models for most purposes, at least as a starting point for modelling. The model results can be validated, depending on the dataset and the linked statistical issues, with specific distributions, binomial data or presence-only data to make the analysis a multi-model approach that can justify the use of a particular model and maybe the violation of statistical assumptions.

6. Chapter 2

Influence of environmental and climatic changes on the spatial distribution of North Sea herring

6.1 **Abstract**

The effect of environmental and climatic predictors on the distribution of North Sea herring (*Clupea harengus*) at age was analysed using a complex spatio-temporal dataset for the period 1989 to 2009. Adapted "Generalised Additive Models" (GAM's) were used to investigate the influence of environmental predictors on the spatial distribution of herring, i.e. temperature, amount of stratification, thermocline depth, bottom depth and zooplankton distribution. Furthermore, we investigated northward distribution shifts of North Sea herring in relation to increasing mean summer temperatures of the North Sea and the North Atlantic Oscillation (NAO) winter index. Such changes were already shown for other marine fish species as response to global warming on a global scale and for the North Sea. Our results showed the importance of temperature as environmental predictor for spatial North Sea herring distribution. But in contrast to this and the findings in other studies, North Sea herring showed no northward shift triggered by global warming. The core distribution area, indicated by the "Centre of Gravity" (COG) remained stable, even in an increasing temperature scenario based on "International Panel for Climate Change" (IPCC) climate projections.

6.2 **Introduction**

Global warming already has multitude effects on ecosystems and will likely continue to do so (IPCC 2014). Consequences are changes in ecosystem dynamics, species phenology, species interactions and shifts in species distribution ranges due to increasing temperatures (Walther *et al.* 2002, Parmesan 2006). Marine ecosystems exhibit several additional changes with increasing water temperatures, like altered stratification and circulation processes, a rising sea level and changes in precipitation levels and freshwater input (Doney *et al.* 2012). In this context, spatially diverse regional changes are important drivers for changes in marine ecosystems.

Ecological responses of species are evident on a global scale. Increasing temperatures and associated changes can affect phyto- and zooplankton growth and timing (Beaugrand *et al.* 2002, Hays *et al.* 2005). Additionally, geographical extensions are changed by warming oceans. Responses of marine zooplankton and fish species are in general distribution shifts to higher latitudes (Beaugrand *et al.* 2002, Richardson 2008, Cheung *et al.* 2013, Gamito *et al.* 2015). This phenomenon was found on a global scale for exploited marine fishes and invertebrates (Jones and Cheung 2014).

Rapid warming was observed in the European Seas (Belkin 2009). The North Sea appears to be a “hot spot” for global warming due to the combination of regional scale processes and global warming (Holt *et al.* 2012) and has warmed 1 to 2 °C since the mid 1980s (OSPAR 2010). Two-thirds of the exploited and unexploited North Sea fishes show the same distribution shifts to higher latitudes (Perry *et al.* 2005). Thus, temperature is an important direct or indirect driver for the distribution of demersal and pelagic fish species. Changes in distribution shifts can have consequences for the whole ecosystem which in turn impacts fisheries (Cochrane *et al.* 2009).

Herring (*Clupea harengus*) plays an important role in many temperate marine ecosystems as prey for fish species, marine mammals and seabirds. Furthermore, herring are important targets for commercial fisheries. However, due to high fishing pressures with new fishing techniques and changing environmental conditions, herring suffered and worldwide most herring populations declined dramatically since the 1950s (Simmonds 2007, Schweigert *et al.* 2010, Lorentzen and Hannesson 2010, Dickey-Collas *et al.* 2010).

The North Sea herring stock collapsed in the mid 1970’s as a consequence of recruitment overfishing (Saville and Bailey 1980). After a fishery closure, the stock has recovered until the mid 1990s but never reached similar spawning stock biomass (SSB) levels again. Since the year 2000, SSB levels are fluctuating between 1 and 3 mio. t. The 4 major spawning components (i.e. Orkney/Shetland, Buchan, Banks and Downs) re-colonised their traditional spawning grounds (Schmidt et al. 2009) as soon as the stock recovered. After spawning season that lasts from August to January, herring migrate to the specific overwintering areas in the Norwegian Trench and the southern North Sea and mix completely during spring and summer feeding (Cushing 1981, ICES 2014, Corten 2001). Feeding starts in April and May in the north-eastern North Sea and continues in June and July in the north-western North Sea (Corten 2001).

We here analysed changes in spatial distribution of autumn spawning North Sea herring in relation to environmental and climatic parameters. Earlier studies have shown the importance of the environment for fluctuations in herring distribution. Temperature seemed to be the most important driver, but also other parameters influenced herring distribution, i.e. thermocline depth, salinity, zooplankton abundance and distribution, water depth, topography or climatic indices like the North Atlantic Oscillation (NAO) (Maravelias and Haralabous 1995, Maravelias and Reid 1995, Maravelias 1997, Bailey et al. 1998, Corten 2001, Volkenandt et al. 2015). We here used for the first time a time series of 21 years (1989 – 2009) of spatio-temporal herring abundance data derived from the

"International Herring Acoustic Survey" (HERAS). As opposed to the other studies we here used no landing data or density indices for specific areas of the North Sea. Instead we used numbers (abundances) of herring at age for the entire North Sea, divided in ICES statistical rectangles with sizes of 1° Longitude by 0.5° Latitude. Furthermore, we treated statistical issues arising from the analysis of complex spatio-temporal datasets in fitted "Generalised Additive Models" (GAM's), i.e. temporal autocorrelation, spatial autocorrelation and zero-inflation. From here, we focused on analysing (i) temperature preferences at age of North Sea herring, (ii) environmental age specific drivers for herring distribution, (iii) retrospective herring at age distribution changes (Centre of Gravities) in the context of mean summer temperatures of the North Sea and (iv) possible future distribution changes for herring age groups from climate change projections until the year 2099. Our age specific GAM's showed clear distinctions between immature (ages 0 and 1) and mature (age > 3) North Sea herring. Immature herring showed preferences for warmer temperatures and almost no significant environmental drivers influencing the spatial distribution. Mature herring distribution was mainly influenced by surface temperature and bottom depth, but was restricted to a core area in the north-western North Sea.

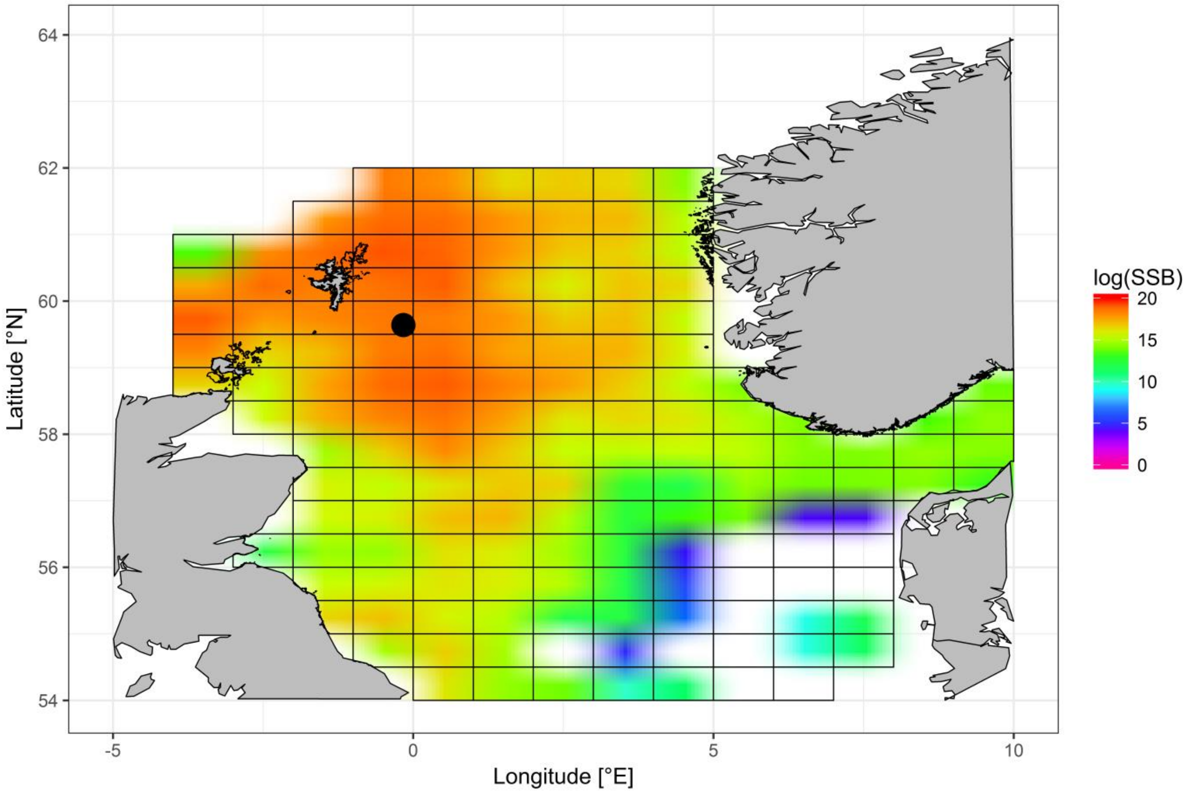

Figure 1. Overview of the study area and the analysed 164 ICES statistical rectangles. The black dot represents the mean distribution centre of North Sea herring in the considered period (1989 – 2009). The spawning stock biomass (log-transformed SSB; average of the entire period) is interpolated in the background.

6.3 **Material and Methods**

6.3.1. Data

We used data from the “International Herring Acoustic Survey” (HERAS) to explore the distribution of autumn spawning herring in the North Sea (International Council for the Exploration of the Sea – ICES – management area 47d3). Data are available from the “Regional DataBase FishFrame” (http://www.ices.dk/marine-data/data-portals/Pages/RDB-FishFrame.aspx). HERAS covers the whole North Sea annually during June/July on predefined transects. Scientific echosounders are used to record so-called “Nautical Area Scattering Coefficient” (NASC in $m^2/n.mi^2$) values which represent the fish density along the transects. During the surveys, pelagic trawling is conducted to determine the species composition and length distribution of the acoustically determined fish abundances. Subsamples of the trawl catches are used for age-reading in the lab. Eventually herring abundances (numbers) at age of immature and mature herring are available per ICES statistical rectangle (Figure 1). We limited our analysis to 164 ICES rectangles located between 54° and 62° N as well as -4° and 9° E that had a good coverage of herring data for all ages (Figure 1). Most of the rectangles we used have been investigated more than 15 times over our 21 years study period (1989 – 2009). Occasional smaller area coverage was due to bad weather or technical problems during the surveys. We log-transformed herring abundances prior to analysis to reduce the variance between records in the dataset.

We used climate, hydrography and zooplankton data as predictors for the herring distribution. As large scale climate variable we used the North Atlantic Oscillation (NAO) winter index (December to March). The NAO describes the differences of normalized sea level pressure between Portugal (Lisbon) and Iceland (Reykjavik) and is a suitable descriptor for climate fluctuations in the North Atlantic and adjacent areas like the North Sea (Hurrell 1995). We used the NAO to represent atmospheric changes that influence thermal conditions but also affect stratification and current patterns in the North Sea (Hurrell and Deser 2010). Data are derived from the National Center for Atmospheric Research (NCAR – https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-pc-based).

High resolution physical oceanographic data for the North Sea were derived from HAMSOM (HAMburg Shelf Ocean Model; Backhaus 1985, Pohlmann 1996, Schrum and Backhaus 1999), a three-dimensional baroclinic shelf sea model. The HAMSOM data was scaled on the basis of the ICES statistical squares with a resolution of 0.5° in latitudinal and 1° in longitudinal direction. The vertical resolution was 5 m in the upper 50 m and 10 m for higher depths. We here used surface temperature (SST), bottom temperature (SBT) and the depth of the thermocline at the midpoint of the HERAS survey period (Julian day 193).

Zooplankton is a major food source of herring and hence may be a significant bottom-up indicator of herring distribution in the North Sea (Maravelias and Reid 1997, Maravelias *et al.* 2000, Corten 2001). We derived spatially-explicit fields of zooplankton for every year we have herring distribution data from ECOHAM4 (ECOsystem Model HAMburg Version 4; Pätsch and Kühn 2008). ECOHAM is a NPZD ecosystem model which is coupled to HAMSOM to describe the dynamics of the phytoplankton, zooplankton and microbial loop dynamics. The spatial resolution of the model output was identical to HAMSOM, based on the ICES statistical squares. Only the meso-zooplankton output was used as driver for herring distribution in the models.

6.3.2. Statistical analysis

We used "Generalised Additive Models" (GAM's, Hastie and Tibshirani 1990) in the free software environment "R" 3.0.2 (R Development Core Team 2013) within the package "mgcv" (Wood 2011) for modelling the relationship between herring spatial distribution and environmental variables. While the statistical assumptions of GAM's are the same as for linear regression models, GAM's allow for non-linear relationships between response and explaining variables. This makes GAM's suitable for modelling species distributions (Austin 2007, Elith and Leathwick 2009).

The basic GAM formulation

$$y_{ij} = alpha + s(x_{ij}) + \varepsilon_{ij}, \tag{1}$$

contains the model intercept *alpha*, smoothing functions *s* of the explaining variables *x* and the error term ε, where *i* represent the years and *j* the sampled areas. All analysis were done with Gaussian distribution error.

First, we investigated age-dependent temperature preferences of North Sea herring applying simple univariate GAM's with herring abundance at age as response variable and SST and SBT as predictor variables in separate models.

Next, we modelled the impact of environmental factors on the herring distribution based on a GAM model selection approach for complex spatio-temporal datasets developed by Gloe et al. (in prep.). In the approach spatial autocorrelation was reduced by implementing 2-dimensional space smoothers of mean Longitudes and Latitudes of the ICES statistical rectangles. We also tested for temporal autocorrelation affecting our results, but found this to be unproblematic for our dataset. The high number of zero observations, especially for older age classes (60% for age 9), furthermore produced no bias in the GAM results and was hence neglected (Gloe et al. in prep). Forward and backward model selection was conducted based on the Akaike Information Criterion (AIC; Akaike 1974), resulting in the same final models. We used either surface or bottom temperature, a temperature stratification index (difference from surface and bottom temperature), depth of the thermocline, bottom depth and (meso-)zooplankton abundances as explaining variables in our initial statistical models. The presence of multicollinearity between the predictor variables was tested by calculating the “variance inflation factor” (vif) in the “car” package (Fox and Weisberg 2011). A vif value smaller 3 indicates no collinearity between the predictors (Zuur et al. 2010)

We furthermore explored changes in the age-specific distribution of North Sea herring using a Centre of Gravity (COG) approach. The COG represents the mean Longitudes and Latitudes of the ICES rectangles weighted by herring abundance. We tested for a northward shift of the herring population by using the Latitude of the COG for all years as response variable in GAM’s. Here we used temperature data and the winter NAO index as explaining variables. Temperature data from the HAMSOM model was again used for the Julian day 193 and additionally.

Finally, we explored how the age-dependent herring distribution in the North Sea will likely change in response to expected global climate change. Towards this goal we predicted future changes (2010-2099) in herring distribution applying above derived models using SST together with a 2-dimensional space smoother as well as an additional year smoother as explaining variables. Based on these predictions, the COG for all predicted years and ages was calculated. Future North Sea SST data was provided by the Max Planck Institute (MPI) for Meteorology in Hamburg (Mathis and Pohlmann 2014) using on climate projections based on the International Panel for Climate Change (IPCC) future emission scenario A1B from the Special Report of Emission Scenarios (SRES). The scenario A1B is basically based on the future assumptions of very rapid economic growth, global population that peaks in mid-century and declines thereafter, and the rapid introduction of new and more efficient technologies. Energy supply is balanced across all sources. For further information about this and other developed scenarios see the IPCC report or homepage (http://www.ipcc.ch/ipccreports/sres/emission/index.php?idp=93).

6.4 **Results**

6.4.1. Temperature preferences of North Sea herring

We first investigated the thermal preferences of North Sea herring and found a clear age-dependent pattern in the response of abundance to sea bottom (SBT) and sea surface temperature (SST) (Figure 2). Ages 0 and 1 showed a clear preference for warmer waters compared to older herring. Age group 0 abundances increased rapidly above 8 °C SBT and 14 °C SST. Age 1 herring showed a slightly different and unique relationship to temperature. Here, the curve increased towards plateaus at about 8 °C and 15 °C for SBT and SST, respectively. In contrast, age 3 and 4 herring showed SBT optimum curves, with an optimum between 5 °C and 8 °C. These optimum curves were less pronounced for SST. Here the abundance decreased clearly at temperatures higher 13 °C. Age groups 5+ showed no optima but a clear relationship to lower temperatures. The highest abundances were found at SBTs below 10 °C and SSTs below 15 °C. Eventually the older part of the herring population (ages > 2) clearly displayed preferences for lower temperatures since abundances continuously decreased with both SBT and SST.

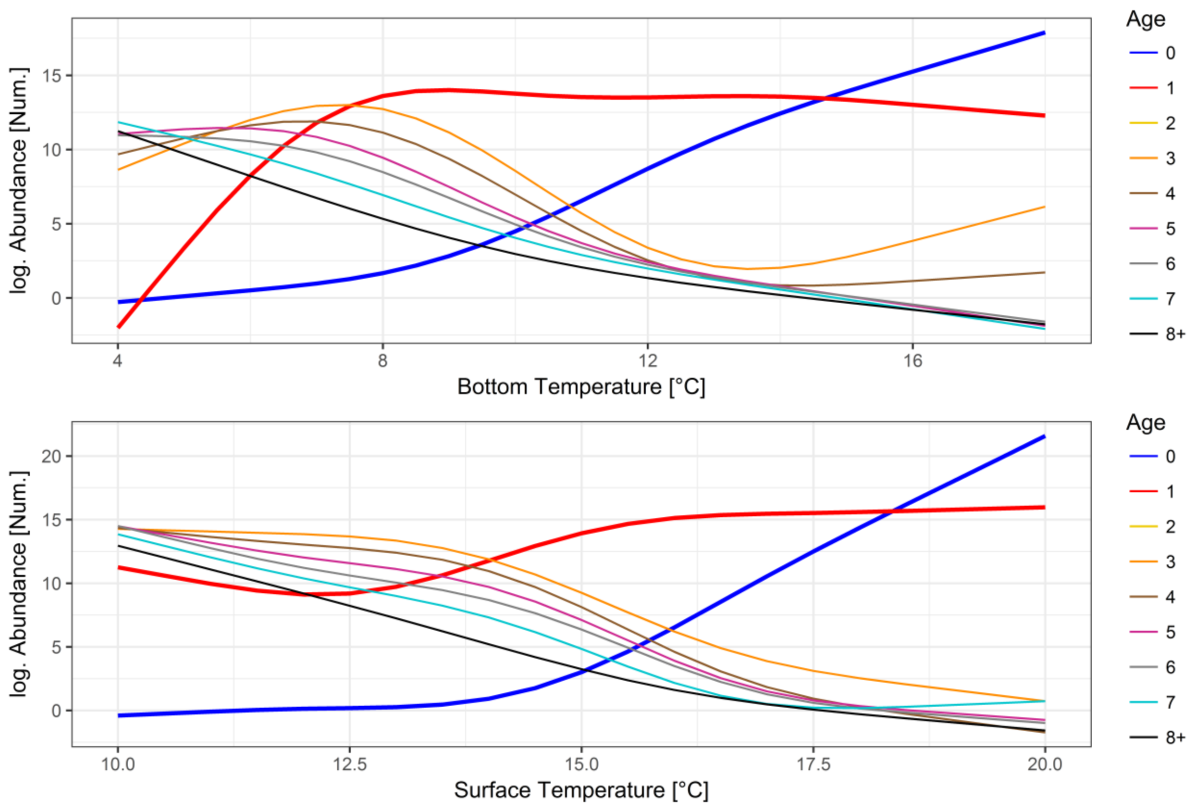

Figure 2. Bottom (upper) and surface (lower) temperature preferences for all age groups derived from the basic GAM's.

6.4.2. Herring spatial distribution models

In the next step of our study we explored a wider range of environmental drivers for the spatial distribution of herring in the North Sea (Table 1). Initial GAM's included the predictor variables SBT and SST, a water column stratification index, the depth of thermocline and the bottom as well as zooplankton abundance. This predictor set showed no multicollinearity. The "variance inflation factor" (vif) was below a threshold of 3, which indicates no issues in this context (Zuur 2010). The finally selected models explained between 31 and 60 % of the variation in herring spatial distribution. The 2D-space smoother (Latitude and Longitude) reduced spatial autocorrelation (SAC) efficiently in the model residuals as proposed by Gloe et al. (in prep.) and was highly significant in all models. The final models for all ages are given in Table 1, appertaining smoothers are shown in Figure 3.

Table 1: Final GAM's of the environmental models for each age group. Asterisks indicate the significance levels of the variables (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The final models included a 2D-space smoother of Latitude and Longitude to account for spatial autocorrelation and the predictor variables sea surface (*SST*) and sea bottom temperature (*SBT*), a stratification index (*Strat.*), the bottom depth (*Bot. Depth*), Zooplankton abundance (*Zoopl.*) and the depth of thermocline (*Therm. Depth*).

Age	Final Model	Explained Deviance [%] (Adjusted R^2)
0	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij}) + f_4(Strat._{ij})* + f_5(Bot.\ Depth_{ij})** + \varepsilon_{ij}$	59.9
1	$f_1(Lat_{ij}, Lon_{ij})*** + f_3(SBT_{ij})* + f_4(Strat._{ij})** + f_5(Bot.\ Depth_{ij})*** + f_6(Zoopl._{ij})** + \varepsilon_{ij}$	31.1
2	$f_1(Lat_{ij}, Lon_{ij})*** + f_3(SBT_{ij})*** + f_5(Bot.\ Depth_{ij})*** + f_6(Zoopl._{ij})*** + \varepsilon_{ij}$	34.5
3	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij})*** + f_4(Strat._{ij})*** + f_5(Bot.\ Depth_{ij})*** + f_6(Zoopl._{ij}) + \varepsilon_{ij}$	43.4
4	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij})*** + f_4(Strat._{ij})*** + f_5(Bot.\ Depth_{ij})*** + f_7(Therm.\ Depth_{ij}) + \varepsilon_{ij}$	48.4
5	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij})*** + f_4(Strat._{ij}) + f_5(Bot.\ Depth_{ij})*** + f_7(Therm.\ Depth_{ij}) + \varepsilon_{ij}$	46.2
6	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij})*** + f_4(Strat._{ij})* + f_5(Bot.\ Depth_{ij})*** + f_7(Therm.\ Depth_{ij}) + \varepsilon_{ij}$	46.6
7	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij})*** + f_4(Strat._{ij}) + f_5(Bot.\ Depth_{ij})*** + f_7(Therm.\ Depth_{ij}) + \varepsilon_{ij}$	46.7
8+	$f_1(Lat_{ij}, Lon_{ij})*** + f_2(SST_{ij})*** + f_4(Strat._{ij})** + f_5(Bot.\ Depth_{ij})*** + f_6(Zoopl._{ij}) + f_7(Therm.\ Depth_{ij}) + \varepsilon_{ij}$	45.7

The final model of age 0 herring showed, with an exception for the space smoother, only significance levels smaller than 0.01. SST stayed in the final model, but was in contrast to all other age groups not significant ($p = 0.125$). Here temperature was the main driver for herring distribution, as it was expected from the temperature preferences results. Age 1 and 2 herring differed from older herring insofar that SBT instead of SST was a significant predictor of distribution, with a smaller p-value for age 1 herring (p = 0.013). Instead, stratification of the water column seemed to be an important

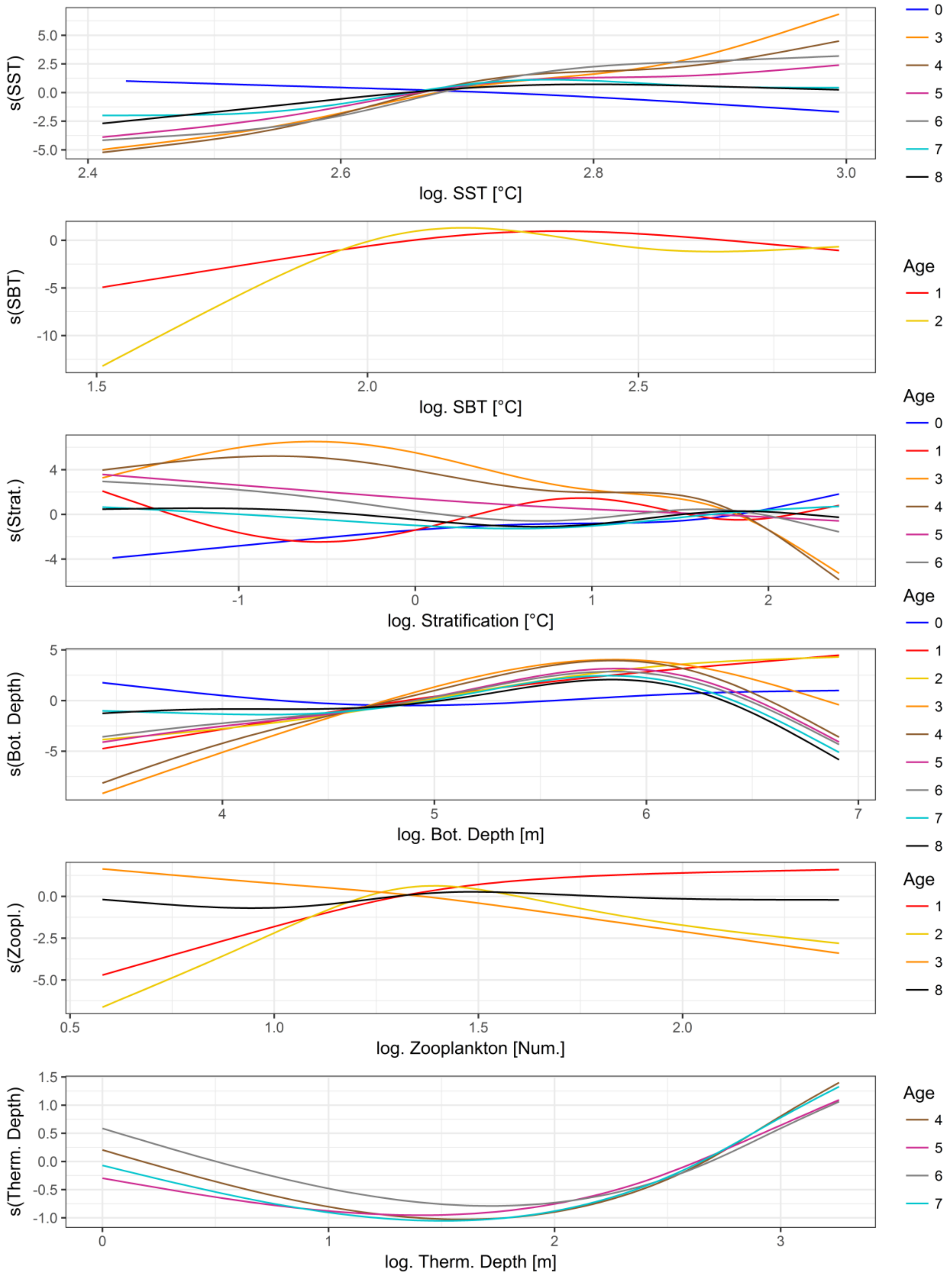

Figure 3. Smoothers from the final models with the exception of the 2D-space smoothers (Latitude and Longitude) are shown for all age groups. The smoothers were extracted from the final model outputs. Standard errors are not shown to provide a better overview of the results.

driver for age 1 herring distribution. Bottom depth was an important driver for both ages 1 and 2. There were only minor differences in the final models of age 3+ herring. SST and bottom depth were highly significant for all ages. A difference was made up by the stratification index which was only significant on levels smaller 0.01 for ages 3, 4 and 8+. Overall, differences of the environmental drivers existed between the youngest herring (ages 0, 1 and 2). For 3+ herring no major differences between the final models were obvious and the smoother shapes were similar (Figure 3).

6.4.3. Changes in mean population distribution

We investigated changes in the distribution of North Sea herring by analysing the Centre of Gravities (COGs) at age. The COGs represent the weighted mean Longitudes and Latitudes of herring distributions at age. Figure 4 shows the mean COGs per age while the distributional range shift of the survey period is indicated by the lines in longitudinal and latitudinal direction. We found distinct differences between the youngest and oldest herring. Age 0 and 1 herring showed the most eastern COGs with a slightly overlapping distributional range. Age 0 herring COG was closest to the German and Danish coast. The ranges of age 2 herring were located more or less in between the youngest herring (ages 0 and 1) and 3+ herring, distinctly located more in the Central North Sea compared to the young herring. Both ages showed a high variability in distribution over the investigated period. The latitudinal range was 365 km for age 0 and 267 km for age 1 herring, the longitudinal 259 km and 234 km, respectively. We found only small differences in the COGs of 3+ herring. The mean distribution shifted slightly to the North with increasing age. The mean COG-Latitudes for all 3+ herring were close together within a range of 85 km. Furthermore, the variability over the whole period was comparable for all age 3+ herring in both directions. It ranged on a scale of 291 km in latitudinal and 171 km in longitudinal direction. The variability of distribution for older herring was overall smaller compared to younger age groups 0 and 1.

Next we analysed if the shift in the northern distribution of approximately 300 km for 3+ herring was dependent on mean North Sea SBTs. We found an almost linear relationship between SBTs and an increasing northern distribution, reaching the most northern distribution at 9 °C (Figure 5). Years with high SBTs were characterized by a more northern distribution compared to years with lower temperatures. SBT was significant with a p-value smaller 0.01 and the model explained 9 % of the total variation of northern herring distribution.

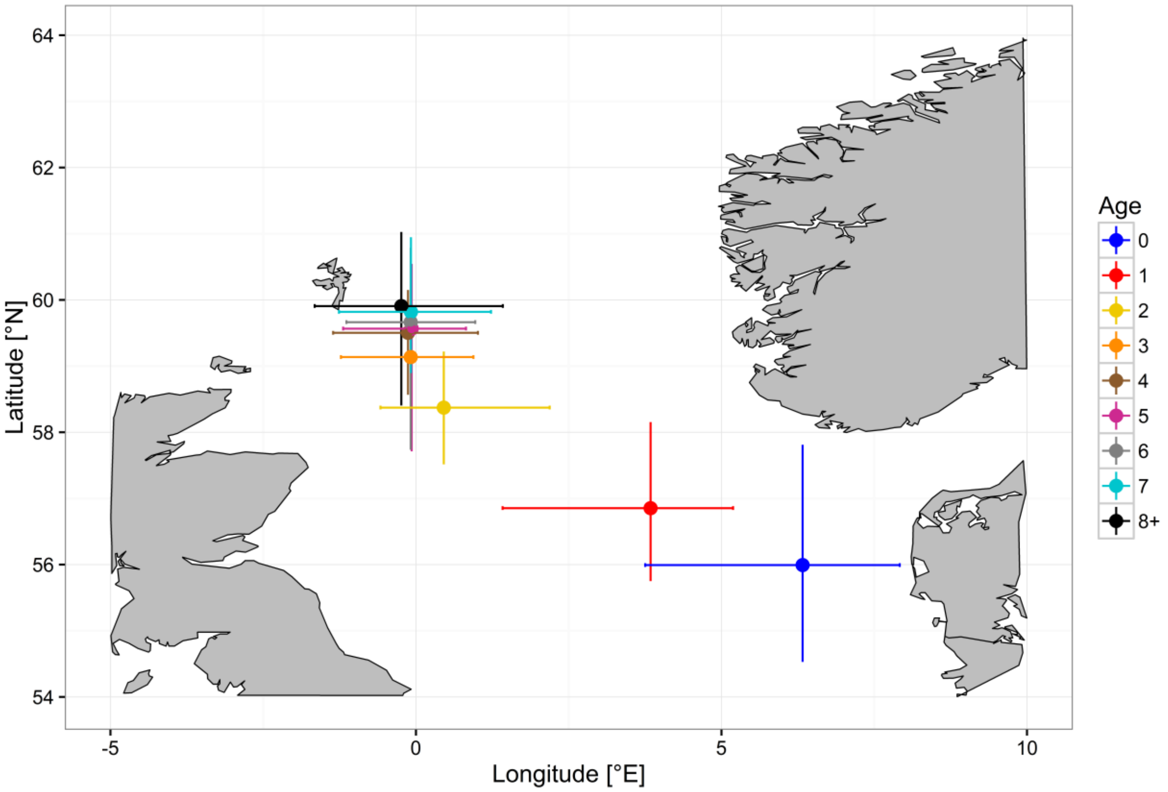

Figure 4. Mean Centre of Gravities (COGs) for all herring age groups for the entire study period. The lines in latitudinal and longitudinal directions show the distribution variability around the COGs.

Then we analysed possible relationships between the northern distribution of the mature herring stock and the North Atlantic Oscillation (NAO). Here we found no clear relations. The most northern distribution occurred at the lowest (-4) and the highest (4.5) NAO values (Figure 5). Furthermore, a more northern distribution occurred when the NAO Index had values around 0.5. The NAO Index was significant (p-value < 0.001) and explained 15 % of the herring distribution. Unfortunately, the NAO showed no clear relationship and was therefore inappropriate to explain northward shifts of North Sea herring.

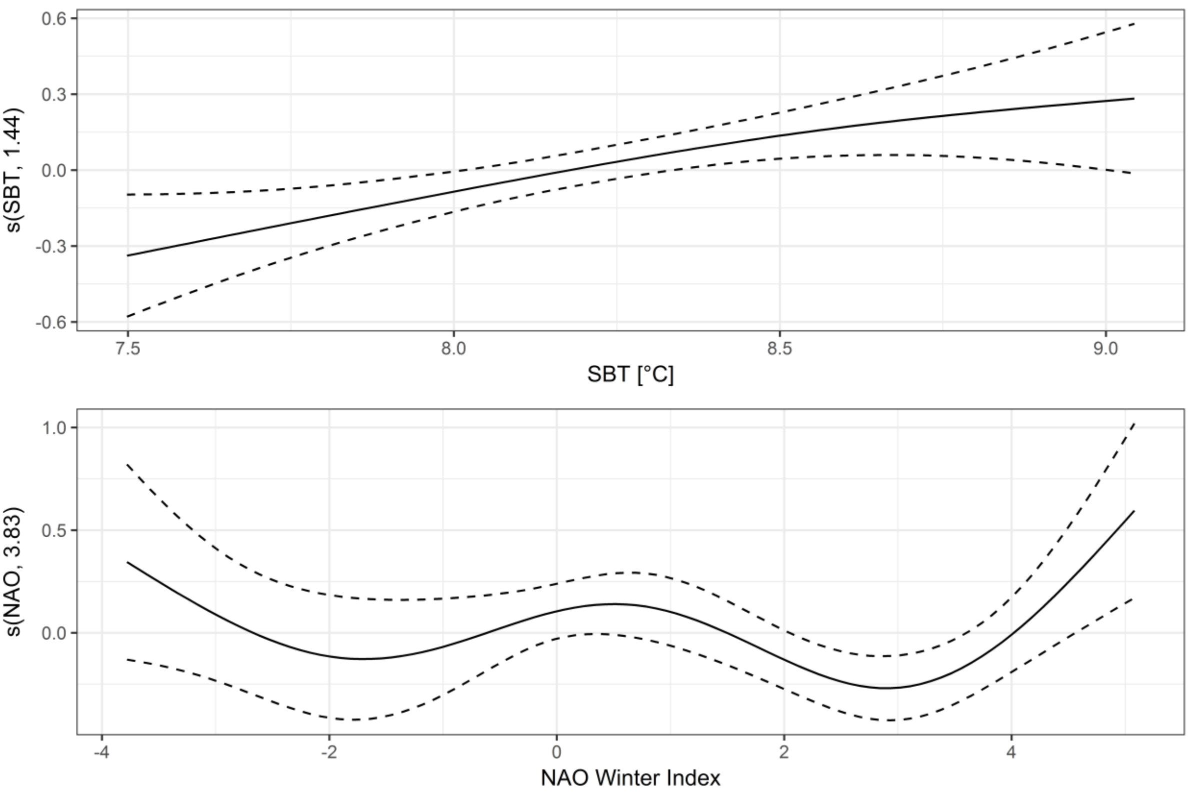

Figure 5. Relationships between the northern distribution of North Sea herring and the sea bottom temperatures (SBT, upper graph) and the North Atlantic Oscillation (NAO, lower graph) winter index. Smoothers from the GAM's are shown.

6.4.4. Projections of the future herring distributions

The last step of our study consisted of predicting the future mean population distribution, i.e. the COG, based on projected SSTs for the North Sea. The mean SST increase per ICES square over the projected period (2010 to 2099) was 4.14 °C, based on the IPCC future emission scenario A1B. These SST differences are independent from the area, e.g. German Bight, central North Sea or northern North. The COG predictions showed clear differences between young and old herring ages (Figure 6). In the projection of the year 2099, age 0 herring was distributed close to the German and Danish Bights with a low variability in the model predictions. The longitudinal and latitudinal range shifts of distribution were 74 km and 69 km. The COG of age 1 herring was located relatively wide in the north-western part of the North Sea. But in contrast to age 0 herring, the distributional range over the projected period was high. In longitudinal and latitudinal direction the range was 380 km and 292 km, respectively. A similar behaviour was found for age 2 herring. The COG was a bit more in the north-west compared to age 1. The variability of the distribution derived from the model predictions was relatively high, with 437 km in Longitude and 150 km in Latitude. Opposed to this, older herring (ages 3 – 7) showed stable COGs in the projections of the year 2099, located in the north-western part of the North Sea. The ranges in distribution in both directions were low and showed only small variability over the projected period. The longitudinal and latitudinal ranges were less than 74 km and 66 km, respectively. An exception for old herring was built by the oldest 8+ age group. The COG showed the most distinct north-western distribution and additionally a high variability in distribution. The longitudinal and latitudinal ranges were 300 km and 100 km and therefore multiple higher compared to older herring of the ages 3 – 7.

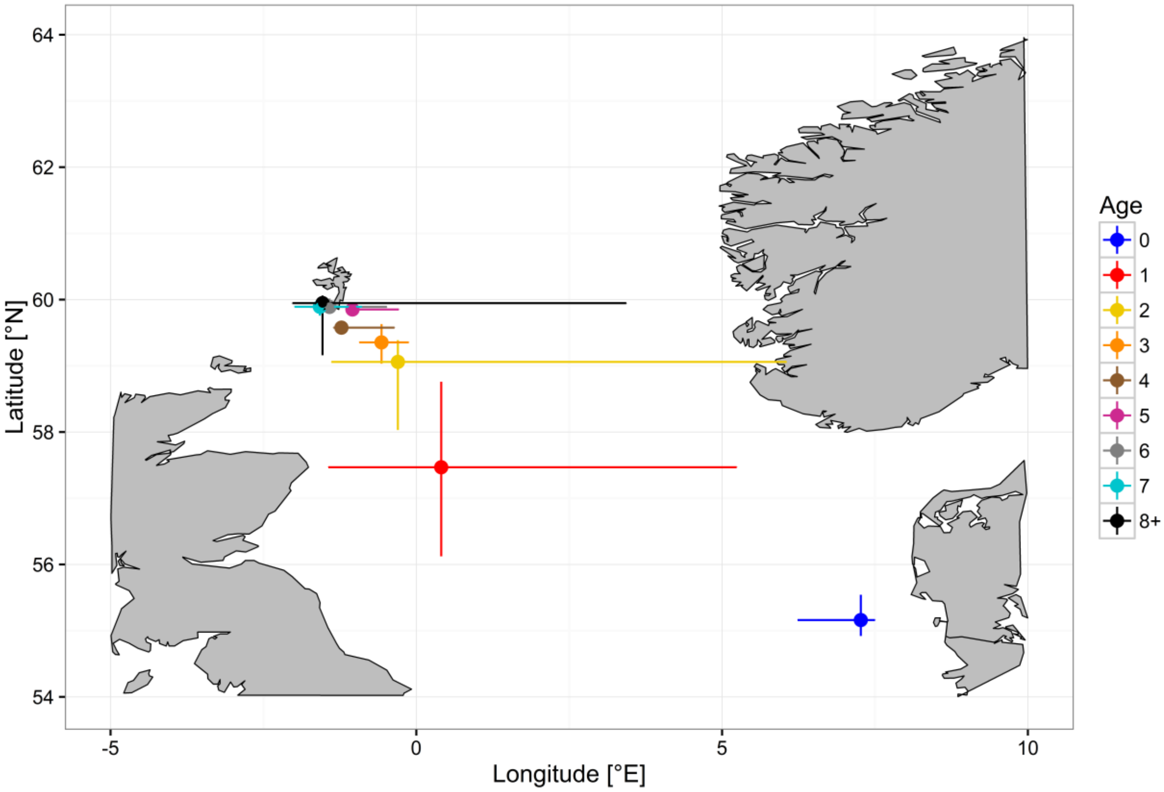

Figure 6. Mean Centre of Gravities (COGs) derived from the projections for all herring age groups for the period 2010 – 2099. The lines in latitudinal and longitudinal directions show the distribution variability around the COGs.

6.5 **Discussion**

6.5.1. Temperature preferences

First, we investigated the thermal preferences of autumn spawning North Sea herring age groups by doing simple GAM's with the HERAS data and modeled temperature (SST and SBT) data. Atlantic herring (*Clupea harengus*) is known as temperate species with preferences for colder waters in general (http://www.fishbase.org/summary/SpeciesSummary.php?ID=24&AT=herring). Brunel and Dickey-Collas (2010) give mean SSTs for Atlantic herring between 11.2 – 13.1 °C and 4.7 – 6.2 °C for the southern and northern distribution borders, respectively. For the North Sea, Dietrich et al. (1959) described herring occurrence in cold bottom waters. Maravelias *et al.* (2000) found highest abundances of mature herring between SSTs of 11 - 12 °C (period 1992 to 1995 around Shetlands and Orkneys).

Our analysis was based on annual herring abundances at age per ICES statistical rectangles and mean SSTs and SBTs for Julian day 193 (midpoint of the HERAS surveys). Working with Julian day 193 was preferred over calculating temperature means for a period of 6 weeks in terms of losing information. Performing the analysis with 164 ICES squares gave a good resolution, although the most southern part of the North Sea (< 54 °N) was not covered. The boundary of the German HERAS area shifted to the south during the study period to find the distribution boundary of sprat. In this context, a regularly coverage of the ICES squares was more important for our statistical analysis. This could have introduced bias in presented temperature ranges especially for young herring (ages 0 and 1), as young herrings are widespread in coastal areas that were not covered at Latitudes smaller 54 °N. We did not distinguish between immature and mature age 2 and 3 herring. About 70 to 80 % of age 2 herring is assumed to be mature (HAWG 2015), but variability is high. Due to the high percentage of mature herring in these age groups and general uncertainties in dividing immature and mature herring (sample size, staging in the lab and associated assumptions), we treated all age 2 and 3 herring as mature.

Our results confirmed the current understanding of the herring life cycle in the North Sea. Young herring (ages 0 and 1) are located close to the German and Danish Bight (Dutch Bight not covered in our dataset), where temperatures are generally higher compared to the central and northern North Sea. These habitat differences are displayed by the temperature preferences very well. Age 0 herring with a unique preference curve (increasing with higher temperatures for both SST and SBT) and age 1 herring with a completely different preference curve compared to older herring. All age groups older than 2 showed very similar behaviour with small differences in ages 2 and 3. Here, temperature preferences look more like optimum curves for SBT compared to herring of ages older than 4. This effect could be triggered by our assumption that all age 2 and 3 herring are mature. A small percentage of age 2 and 3 herring was immature (at least 20 to 30 %) and therefore could have introduced bias in our temperature preferences calculations for mature herring. Furthermore it could have been an effect of decreasing age group abundances with increasing age (abundances in numbers per age groups decline by half) and therefore probably smaller distributional ranges. The SST optima of mature North Sea herring fitted very well with the ranges given by Brunel and Dickey-Collas (2010) and Maravelias et al. 2000 (SST preferences between 11 – 13 °C). Overall, our analysis and data setup seemed very suitable to show temperature preferences of different age groups (life-stages) of North Sea herring. Young herring showed preferences for warmer waters, while migrating (mature) herring showed clear preferences for cold water.

6.5.2. Impact of environmental predictors on herring distribution

The next part of our study consisted of analyzing environmental drivers affecting the distribution of herring at age. Temperature (SST and SBT), stratification of the water column, thermocline depth, bottom depth and zooplankton abundance were used as predictors. These predictors are known to have an impact on herring distribution. Several studies addressed this topic, using abundance indices from HERAS, or landing data with distinct predictors for varying periods and areas. Temperature was an important predictor, but also other parameters like thermocline depth, salinity, zooplankton, water depth or topography influenced the spatial distribution herring (Maravelias and Haralabous 1995, Maravelias and Reid 1995, Maravelias 1997, Bailey et al. 1998, Corten 2001, Volkenandt et al. 2015). Most of the listed studies analysed a shorter period, a smaller area or did not consider a potential age effect.

Several statistical methods were used in the above named studies (Spearman Rank Correlation, Geostatistical Mapping, GAM) to analyse environmental predictors influencing North Sea herring distribution. We went a step further with adapted GAM's to address potential statistical issues of spatio-temporal datasets, i.e. autocorrelation and zero-inflation. We used a 2-dimensional space smoother (Longitude and Latitude) to reduce spatial autocorrelation in the models. Temporal autocorrelation and zero-inflation were no issues in our dataset. Data limitations that were described above were also valid for this part of our study (coverage, maturity). Here again, modeled data from the Julian day 193 was used, also for the zooplankton abundance from the ECOHAM models, to avoid averaging data for the period of 6 weeks. Frontal systems are known to have important impacts on ecosystems, by enhancing productivity (Le Fèvre 1986) and are also described to have an effect on herring (Maravelias and Reid 1997). We tried to analyse the effect of frontal systems with frontal indices on the basis of ICES squares from satellite data (kindly provided by Peter Miller from the Plymouth Marine Laboratory). It turned out that our grid of ICES squares was scaled too large for analysing the effects of frontal systems in the North Sea on herring distribution. Hence, we could not use the provided frontal indices from PML in our analysis.

Clear distinctions between environmental predictors influencing the spatial distribution of young and old herring were found. Compared to old herring, the distribution of young herring (ages 0 and 1) was explained by similar drivers, i.e. temperature, stratification and bottom depth, but with lower significance levels. Young herring occur preferably in coastal waters before getting mature and start migrating with the adult stock. Coastal waters are highly dynamic, influenced by tidal currents and river runoff. This could be a possible explanation for the lower significance levels of the environmental predictors, opposed to the migrating herring older than age 2. After recruiting, herring start migrations and are located in their summer feeding areas during HERAS surveys. Probably for this reason, all mature herring age groups showed similar model results. Significance levels of the stratification index differed among ages and zooplankton was not included in every final model (only ages 2, 3 and 8+). Besides bottom depth, SST seemed to be the most important predictor. The only exception being age 2 herring, were SBT was more important, equally to age 1 herring. Ages 1 and 2 final models have in common that both have SBT (instead of SST) and additionally zooplankton abundance occurred in the final model, as opposed to herring older than age 4. The treatment of all age 2 and 3 herring as mature probably introduced bias in the analysis. This could lead to the conclusion that especially ages 2 and 3 depicted intermediate stages. Overall, mature North Sea herring older than age 4 occurred in a core area in the north-western North Sea during summer feeding. In this area, variations in spatial distribution were mainly driven by the environmental predictors temperature (SST) and bottom depth.

6.5.3. Changes in spatial population distribution in relation to summer temperature and NAO

Changes in the spatial distribution of North Sea herring at age were analysed by investigating the Center of Gravities (COGs). The COG describes the weighted mean herring locations (Longitudes and Latitudes), weighted by the abundance at age of each ICES square. Summer temperatures and the NAO winter index were used to analyse shifts in herring distribution. Increasing temperatures in marine ecosystems are known to cause distribution shifts. This phenomenon was found on a global scale (Jones and Cheung 2014) and for the North Sea (Perry *et al.* 2005). Corten (2001) described an additional indirect temperature effect for North Sea herring. He found a relation between summer distribution of herring and low abundances of the important food item *Calanus finmarchicus*, caused by warm winter temperatures (high NAO winter index). In this context, Fromentin and Planque (1996) described a negative correlation between *Calanus finmarchicus* distribution and abundance and the winter NAO index. Comparable relationships were described by Broms *et al.* (2012) for Norwegian Spring Spawning herring. Here, temperature had a secondary effect on herring spatial distribution, on the scale of water masses. Primary driver was the *C. finmarchicus* abundance (overwintering generation). An additional secondary temperature effect for herring distribution could be controlling metabolisms by choosing optimal temperatures for gonad ripening (Corten 2001, Broms *et al.* 2012).

We tested several different summer and winter temperatures for the whole North Sea to check for an effect on distribution. Significance levels and the degree of explained variance were relatively low for the most temperature models. Among these different approaches were mean temperatures for the entire North Sea for June and July, provided by the “Bundesamt für Seeschifffahrt und Hydrographie” (BSH) and mean winter temperatures for the months January and February as mean values from the HAMSOM data. Best performance was given by our SBT values for the Julian day 193, representing the midpoint of the HERAS survey. As climatic predictor we used the winter NAO index (NCAR 2014), which represents the mean NAO index of the months December to March. The winter NAO index could not be linked to North Sea herring distribution shifts.

The COGs of herring at age corresponded well with those of the first 2 parts of our study. Age 0 and 1 herring COGs were located close to the German and Danish coast. The spatial ranges of both in longitudinal and latitudinal direction were relatively high. This could be an indicator for the high abundances of both age groups. Another reason for the wider distribution range could be the higher dynamics and environmental variabilities associated with coastal systems in the North Sea. Herring older than age 2 showed a very consistent picture and were located in a core area in the north-western North Sea. Ages 2 and 3 were located a bit more in the south, which again could be triggered by our assumption of full maturity in both age groups. All other ages (4+) showed identical COGs. There was almost no effect of the mean summer temperatures in the entire North Sea. In warmer years, herring older than age 3 showed only a small northern distribution shift in relation with temperature. The summer temperature difference was about 1.5 °C in our period of 21 years. This represented the documented rate of warming of 1 – 2 °C over the last 25 years in the North Sea very good (IPCC 2014). Nevertheless, the temperature increase (SBT) explained only 9 % of the model variance. The maximal range shift in latitudinal and longitudinal direction was about 291 km and 171 km, respectively. Increases in temperature were not distinctly related to northward shifts in herring distribution in our analysis as it was shown for other fish species in the North Sea (Perry 2005). The winter NAO index gave inconsistent results, with northern herring distribution at low and high indices and additionally at indices around 0. Thus, the NAO index was not suitable to explain distributional shifts. This was probably due to the relatively short time series for linking large scale climatic effects like the NAO. Additionally, the NAO showed high variations during the study period while the herring SSB fluctuated only at small levels. The findings from the environmental predictors and the distribution shifts in this part did not allow conclusions about primary or secondary temperature effects, i.e. zooplankton as indirect driver. Overall, we conclude here, that our results are indicative for a conservative behaviour of North Sea herring, where distribution of mature North Sea herring was very similar with COG Latitudes between ages in a range of 85 km, regardless of temperature.

Environmental variations seemed not to have an important impact on large-scale distribution in our analysis. The time series seemed to be too short to link climatic effects as the NAO gave a very inconsistent picture. Similar conclusions were drawn by Corten (2001), were the response of herring to environmental changes was restrained to “conservatism”.

6.5.4. Projections of the future herring distributions

Last, we analysed possible future herring distribution changes. Therefore, we used projected SSTs based on the International Panel for Climate Change (IPCC) future emission scenario A1B from the Special Report of Emission Scenarios (SRES) which is based on the future assumptions of very rapid economic growth, a global population peaks in the mid-century followed by a decline, and the rapid introduction of new and more efficient technologies. Energy supply is balanced across all sources. The mean temperature increase per ICES square for the projected period (2010 to 2099) was 4.14 °C. This is far more compared to the SBT range from the COG analysis with 1.5 °C. This part of the analysis focused only on potential future temperature changes and herring responses. Other environmental predictors were not considered that could have direct or indirect effects on herring distribution at age. Furthermore, the scale of our models was relatively large with predicted abundances at age for the projected period 2010 to 2099 for 164 ICES squares. But preceding results for temperature preferences, environmental predictors and COG analysis showed considerably good results at the same scale. The projected SSTs were only available up to 60 °N, while the preceding analysis was performed with data up to 62 °N. Nevertheless, the predicted future herring distribution did not differ much from recent COGs. With the exception of age 1 herring with a far more western distribution and very large ranges in both directions, COGs of 3+ herring were comparable. Only the ranges in latitudinal direction differed, especially for age 2 and 8+. Overall, these results also underline the conservative behaviour of herring. Even increasing temperatures did not impact the herring distribution in our study. Adult herring were distributed in core areas in the north-western North Sea and young herring close to the coasts. This findings are in contrast to behaviour of other fish species in the context of global warming. While a shift to higher Latitudes was shown for marine fishes on a global scale (Cheung et al. 2013, Gamito et al. 2015) and also in the North Sea (Perry et al. 2005), we found a contrasting result in our analysis. North Sea herring showed in our analysis no northward shift to higher Latitudes in regard to increasing temperatures.

7. Chapter 3

Density-dependent mechanisms in North Sea herring distribution

7.1 Abstract

Density-dependent mechanisms in North Sea herring distribution were analysed using a spatio-temporal dataset for the period 1989 to 2009. Therefore, 3 distribution characteristics were considered: (i) the amount of aggregation within the population, (ii) the total area occupied by the population and (iii) the geographical spread of the population. Getting insight into density-dependent mechanisms of marine population can be important for management purposes, e.g. aggregating behaviour increases the vulnerability to overexploitation. Such behaviour was already shown for some herring stocks. In contrast to these findings, the North Sea herring population showed no density-dependent distribution patterns in our analysis. An important reason for this could be the currently low Spawning Stock Biomass (SSB) compared to times before the stock collapse in the 1970's. Relating these findings to a distribution theory like the basin model suggests that there is currently no habitat limitation for North Sea herring since stock size is below carrying capacity and hence no density-dependent mechanisms could be observed.

7.2 Introduction

Understanding patterns in the spatio-temporal distribution of fish populations is a major topic in marine ecology. The ability to explain changes in species distributions is especially important for the implementation of an ecosystem-based management (Casini *et al.* 2011), including e.g. marine protected area (MPA) designs (Fisher and Frank 2004). Furthermore, understanding responses of species to environmental changes can help predicting effects of global warming, such as distribution shifts (Jones and Cheung 2014). Hence, numerical tools, “Species Distribution Models“ (SDM's), have been developed and are available to analyse relationships between species distributions and the physical environment (Elith and Leathwick 2009).

Besides environmental parameters, density-dependent effects are important when considering spatial distributions of populations. Competition, either inter- or intra-specific, is a common density-dependent mechanism caused by limitations of resources or habitats and hence can have effects on the spatial distribution of populations. The spatial distribution of populations can generally be described by three characteristics (Rindorf and Lewy 2012). These are (i) the amount of aggregation within a population, (ii) the total area occupied by a population and (iii) the geographical spread of a population. For some fish populations it is known that an aggregation of the population occurred at low abundances, making the population even more vulnerable to overexploitation (Beverton 1990, Hutchings 1996). An increasing occupied area with increasing population size was shown for several marine species (Gaston *et al.* 2000, Fisher and Frank 2004), known as positive abundance-occupancy relationship. These changes in the spatial distribution of populations are associated with changes in the geographical spread, e.g. distributional shifts (Casini 2011).

Changes in spatial distribution patterns have been shown for the North Atlantic herring (*Clupea harengus*). Fisher and Frank (2004) analysed abundance-occupancy relationships for exploited fish stocks in the North Atlantic and described a positive relationship for herring, i.e. the occupied area of the population increases with increasing spawning stock biomass (SSB). A change in distribution patterns was also shown for Norwegian Spring Spawning Herring during spawning and feeding periods (Dragesund et al. 1997). High SSBs resulted here in a change in migration patterns and as a consequence in higher occupied areas. The autumn spawning herring population in the North Sea has historically experienced great fluctuations in recruitment and SSB (Dickey Collas et al. 2010). The stock has collapsed in the 1970’s due to high fishing pressure and the fishery was closed for several years. The SSB recovered slowly and since the mid 1990’s is still fluctuating around two million tonnes (HAWG 2014). Up to now, climatic and environmental parameters are known to influence the spatial distribution of North Sea herring, most importantly temperature, but also bottom depth or the amount of stratification (Maravelias 1997, Bailey et al. 1998, Corten 2001, Gloe in prep.). However, knowledge of the importance of density-dependent mechanisms for the spatial distribution of North Sea herring is lacking.

In this study, we investigated the importance of density-dependence for temporal changes in spatial distribution of autumn spawning North Sea herring. We used a comprehensive dataset derived from the “Herring Acoustic Survey in the North Sea” (HERAS) that covers the entire North Sea for more than two decades and is coordinated by the *International Council for the Exploration of the Sea* (ICES) (HAWG 2014, WGIPS 2014). This dataset includes information on herring abundance (numbers) at age (ages 0 to 9+) during the summer feeding phase. We restricted our analysis to herring older than age 3, the age at which they are fully mature (HAWG 2014). An overview of the studied area and the distribution centre of mature herring for the entire period are given in Figure 1.

In the lack of knowledge of density-dependent effects for the North Sea herring stock we followed 2 approaches for our analysing. First we analysed age-independent data by summing up the abundances derived from the HERAS dataset for all herring age-groups older than age 4. Since the abundances from the youngest to the oldest age-groups decline by half, numbers at age differ strongly in the North Sea herring stock. Hence, we investigated in a second analysis potential density-dependent age effects arising from the described stock structure. We explored distribution characteristics of North Sea herring, i.e. geographical spread, occupied area and the amount of aggregation, applying several spatial indicators (Rindorf and Lewy 2012, Woillez 2009). *Geographical spread* was analysed by calculating the *Centre of Gravity* (COG), i.e. the mean location weighted by its abundance, and *Inertia* which is the squared distance (in nautical miles) of the population to the COG. *Occupied area* was analysed using the number of spatial units with positive abundances in the full dataset. Eventually the *amount of aggregation* was described with the *Gini Index* (Gini 1912). With this set of spatial indicators we investigated (i) whether changes in population size cause density-dependent distribution patterns, and (ii) if density-dependent mechanisms are age-specific. Our analysis showed no density-dependent distribution patterns for autumn spawning North Sea herring. The present SSB is relatively low compared to the times before the stock collapse and probably no limitations occur in the north-western North Sea. Our findings provide an insight into the "conservative" behaviour of North Sea herring.

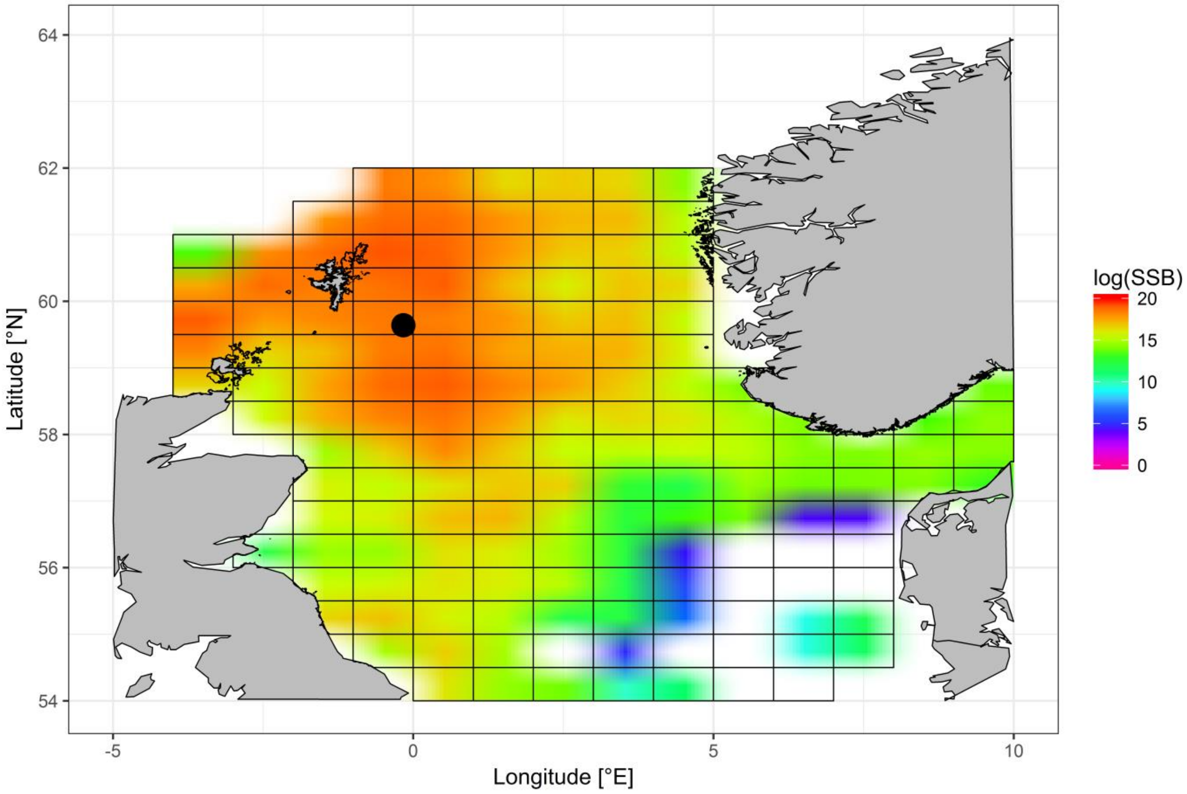

Figure 1. Overview of the studied area and the analysed ICES statistical rectangles. The black dot represents the mean distribution centre of North Sea herring in the considered period (1989 – 2009). The spawning stock biomass (log-transformed SSB; average of the entire period) is interpolated in the background.

7.3 **Material and Methods**

7.3.1. Data

Spatially resolved abundance data for North Sea herring were derived from the “International Herring Acoustic Survey in the North Sea” (HERAS) that is carried out annually in June and July in the entire North Sea. The survey is coordinated by the “ICES Working Group for International Pelagic Surveys” (HAWG 2014, WGIPS 2014). During the surveys, fish densities are measured along defined transects with scientific echosounders. Identification hauls are done to assess species composition and appertaining length frequencies. Age of the species is determined from otholith reading either directly onboard or from frozen sub samples in the lab. With this biological information and the acoustic backscatter strength (Nautical Area Scattering Coefficient – NASC [m^2/n.mi^2]) species abundances and biomasses at age are calculated for so-called statistical rectangles in the North Sea. Our analysis comprised the period 1989 to 2009 and we used a defined number of 164 regularly sampled statistical rectangles (Figure 1) to ensure comparability between years and interpretability of the spatial indicators. The size of the statistical rectangles is 0.5° in latitudinal and 1° longitudinal direction. The boundaries of our study area were 54 to 62 °N and -4 to 10 °E. The annual sum of abundances for ages 4 to 9 was used for the analysis and is assumed to reflect herring SSB since it shows the same trend as the SSB (tonnes) provided by the ICES “Herring Assessment Working Group” (HAWG). Additionally spatial indicators were calculated for herring age groups separately to study potential age effects on herring distribution. All calculations were done in the free software environment “R” 3.0.2 (R Development Core Team 2013).

7.3.2. Spatial Indicators

We used several spatial indicators to analyse density-dependent mechanisms in North Sea herring distribution suitable to describe population distribution characteristics, i.e. amount of aggregation, total occupied area and geographical spread.

Amount of aggregation – The Gini-Index

Gini (1912) developed the Gini-Index as a measure of unequality in a socio-economic context. We used the index here to describe the amount of aggregation of North Sea herring. The Gini-Index ranges from 0 to 1 with value of 0 representing perfect equality, while a value of 1 shows a completely unequal distribution. The calculations were done using the “R” package “reldist” (Handcock 2016). The following equation is used to estimate the Gini-Index:

$$G_j = \sum_{i=1}^{N-1} \frac{i(N-i)[n_{(i+1)} - n_{(i)}]}{(N-1)\sum_{i=1}^{n} n_i}$$

where N is the number of observations (ICES rectangles), sorted in ascending order ($n_{(1)}$, $n_{(2)}$, …) for every year *j*.

Total Occupied Area – Number of ICES rectangles with positive abundances

We used the number of occupied ICES rectangles to describe the occupied area for mature herring. For each year, all ICES rectangles with positive abundances were enumerated and related to SSB. Only the pre-defined 164 rectangles were considered.

Geographical Spread - Centre of Gravity (COG) and Inertia

The COG is computed as the weighted mean geographic location of the autumn spawning North Sea herring stock. The COGs were calculated by weighting the mean geographic coordinates (mean Longitudes and Latitudes) of the ICES rectangles with the appertaining herring abundances. As result, we obtained the distribution centre for each year:

$$COG_j = \frac{\sum_{i=1}^{n} x_{ij} z_{ij}}{\sum_{i=1}^{n} z_{ij}}$$

where x_{ij} is the mean Latitude or Longitude for the ICES rectangle *i* in year *j*, while z_i is the herring abundance.

Inertia is the dispersion of the population around its COG. In more detail, Inertia is defined as the mean squared distance between a local component of the population (in our case the abundance per ICES rectangle) and the COG measured in square nautical miles ($n.mi^2$). Therefore, the rooted Inertia should be preferred as this is a measure in nautical miles (Woillez *et al.* 2009). A higher dispersion value does not allow assertions of the total area covered by the population. The R-script to calculate Inertia, its definition and equation were adopted from Woillez *et al.* (2009). The Inertia has the following equation:

$$I_j = \frac{\sum_{i=1}^{n} (x_{ij} - COG_j)^2 z_i}{\sum_{i=1}^{n} z_i}$$

7.3.3. Statistical analysis

Generalized Additive Models for analysing the aggregated 4+ herring data

To test for density-dependent mechanisms in spatial distribution changes of herring we used “Generalised Additive Models” (GAM’s, Hastie and Tibshirani 1990) in the R package “mgcv” (Wood 2011).

The basic GAM formulation is:

$$y_{ij} = \alpha + s(x_{ij}) + \varepsilon_{ij}$$

with y being the response variable for the spatial unit (i.e. rectangle) *i* and year *y*, *α* is the intercept, *s* the smoothing function of the explaining variables *x* and *ε* the error term. All analysis were done with a Gaussian distribution error that turned out to be best suited for this kind of model in a previous analysis (Gloe et. al in prep). Density dependent mechanisms in herring distribution were analysed by using the spatial indicators as response variables (i.e. y) and log-transformed SSBs as explaining variable in the GAM’s. We used a maximum number of 4 knots (k-value) to avoid overfitting of the models (Keele 2008).

Analysing age effects in HERAS data

Testing for the effect of age was performed by fitting varying coefficient GAM’s to the seperated data for 4+ herring with the following formula:

$$y_{ij} = \alpha + s(x_{ij}, by = f(Age)) + f(Age) + \varepsilon_{ij}$$

The *by* variable allows different slopes for the herring ages in the model. The ages were implemented in the models as a factor (*f*(Age)). We performed a model for each spatial indicator (y_i) with the total abundance of the appertaining age group. Varying coefficient models with significant age-SSB interactions ($p < 0.01$) were further investigated by checking if the age group differences were statistically significant. This was tested with an additional varying coefficient model with 2 groups. The significant age group (ID=1) was modelled against all other age groups (ID=2).

Change Point Analysis

To analyse if change points occurred in our time series, we conducted a Bayesian Change Point analysis using the R package bcp (Wang et al. 2016). Bcp returns the posterior probability for a change point to occur in each year of the study period. Posterior probabilities of ≥ 0.9 indicate significant change points (Lee and Kim 2014). We applied the test to herring SSB and all spatial indicators used in our analysis of the aggregated mature herring stock. To test whether the potential change points are related to herring stock size and therefore to density-dependent mechanisms, we applied additional varying coefficient GAM's. As opposed to the above described models, we here used the periods derived from the change point analysis as grouping factors.

7.4 **Results**

7.4.1. Trends of the spatial indicators and SSB during 1989 - 2009

The time series of the spatial indicators and the SSB is shown in Figure 2. The COG-Longitude decreased strongly from 1990 to 1997, with the most western distribution in 1997 (< -1 °E). This phase was followed by increasing COG Longitudes until 2003, a more western distribution in 2004, and again an increasing trend with easterly distributions (around 0 °E). The COG-Longitude variability of the time series was 2°E (~ 63 n.mi). A similar trend was depicted by the COG-Latitude. The COG-Latitude decreased at the beginning of the time series with the most southern distribution in 1995 (58.5 °N). From 1996 to 2004 the distribution of 4+ North Sea herring was relatively stable at about 60 °N, followed by a slightly more southern distribution. The variability of the COG-Latitude was about 1.8 °N (106 n.mi). The Inertia showed overall a decreasing trend, but was varying strongly over the study period. The highest dispersion of North Sea herring occurred in 1995 (133 n.mi). In the following period the Inertia fluctuated around 90 n.mi. The number of occupied rectangles was relatively high at the beginning of the time series (> 90) and decreased strongly in 1994 (< 70). After a permanent increase of the occupied area until 2004, the number of rectangles decreased again slightly in the following years and fluctuated around a value of 80. The Gini-Index was relatively high in the entire studied period. Values below 0.8 occurred only in the beginning of the time series with a minimum of 0.63 in 1992. From 1993 onwards, values exceeded 0.8 with an exception in 2007 (0.77). The SSB of autumn spawning North Sea herring decreased strongly from 1990 onwards with an abundance minimum of 1521170833 (log.: 21.14) in 1995. This minimum was followed by a constant increase up to the maximum of 8181302400 (log.: 22.83) in 2005. A small decrease of the SSB occurred at the end of the time series.

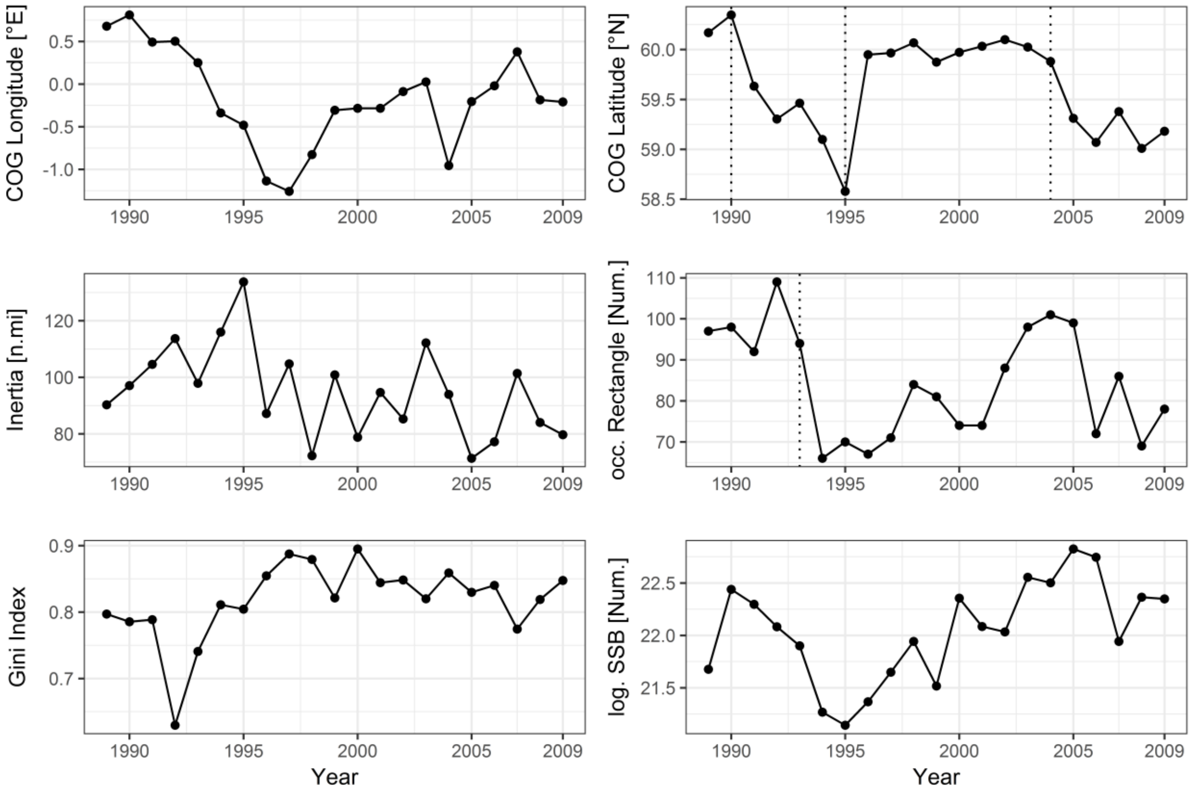

Figure 2. Time series of all spatial indicators and SSB of North Sea herring for the study period 1989 to 2009. The vertical dashed lines represent the significant change points.

7.4.2. Total occupied area

The total occupied area in relation to the log-transformed herring SSB showed an almost linear positive relationship (Figure 3). However, the effect of the SSB in the GAM was not significant ($p > 0.05$) indicating no positive abundance-occupancy relationship for the entire mature herring population. We found evidence for a change point in total occupied area in 1993 indicating differing abundance-occupancy relationships for the periods 1989 – 1993 and 1994 – 2009. However, the varying coefficient model showed no significant differences between both periods.

Visual inspection indicated positive relationships between population size and occupied area for all age groups separately, except for age 9 (Figure 3). However, the varying coefficient model indicated no significant differences between age groups. Ages 4 and 6 seemed to be slightly different from all other age groups. But the p-values of the model indicated no clear statistically differences ($p > 0.01$).

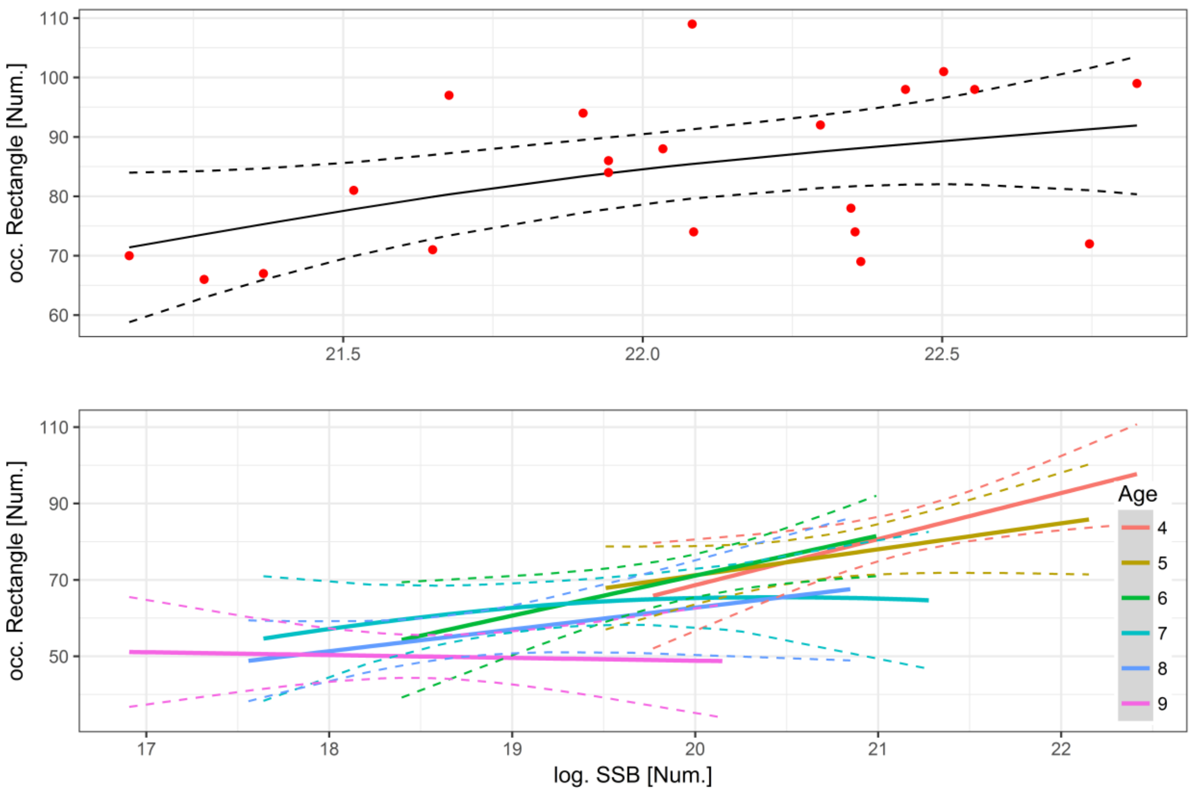

Figure 3. Relationship between total occupied area (number of ICES rectangles with positive abundances) and SSB (log-transformed) of the aggregated (above) and seperated (below) North Sea herring stock. Smoothers from the models are shown with standard deviations (dashed lines). Red dots in the upper graph represent the original aggregated data.

Amount of aggregation

Gini-Index values were never below 0.6 pointing towards a fair amount of aggregation in distribution of the entire herring population. The relationship between the Gini-Index and the SSB of the herring population is positive linear indicating a higher amount of aggregation with increasing stock size (Figure 4). The resulting GAM was however not significant ($p > 0.05$; see Table 1 in Appendix 1). No change points were found in the time-series of the Gini-Index.

Age-specific relationships between the Gini-Index and population sizes were not statistically significant between age groups ($p = 0.166$). Visual inspection however indicated negative relationships between the Gini-Index and population size for age groups 6 and 7, which is in contrast to the remaining age groups. A separate varying coefficient model could not find significant differences in the slopes between age groups 6 and 7 as well as the other age groups.

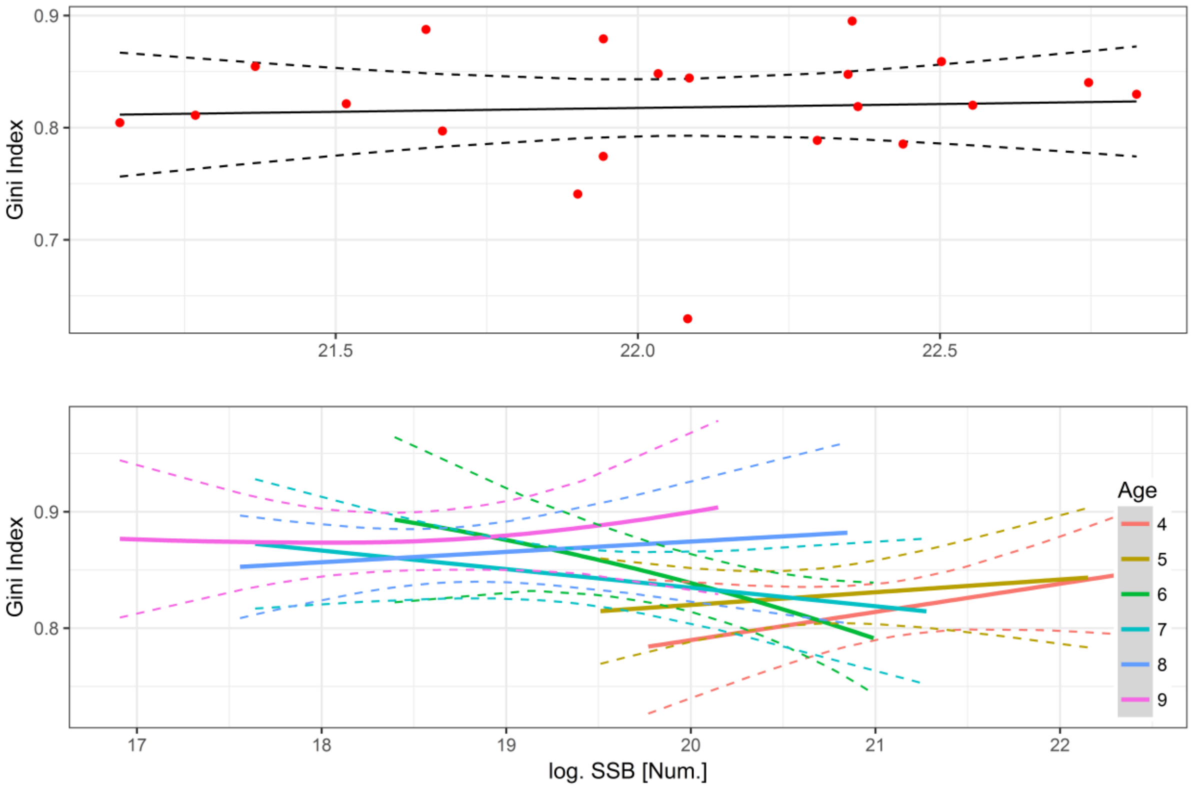

Figure 4. Gini-Index is shown for the aggregated (above) and the separated (below) dataset. Smoothers are presented with standard deviations (dashed lines). Red points in the upper graph shows the original aggregated data.

Table 1. P-values and explained deviances [in %] from the GAM's (aggregated data) and the varying coefficient models (seperated data) are presented.

Model **(aggregated and unaggregated data)**	**p-value**	**explained deviance [%]**
COG-Lat ~ $f_1(SSB) + \varepsilon_{ij}$	0.213	23.5
COG-Lat ~ $f_1(SSB, by=factor(Age)) + factor(Age) + \varepsilon_{ij}$	not sign.	25
COG Lon ~ $f_1(SSB) + \varepsilon_{ij}$	0.233	8.7
COG-Lon ~ $f_1(SSB, by=factor(Age)) + factor(Age) + \varepsilon_{ij}$	Age 9: 0.0002	18.2
Inertia ~ $f_1(SSB) + \varepsilon_{ij}$	0.011	28.9
Inertia ~ $f_1(SSB, by=factor(Age)) + factor(Age) + \varepsilon_{ij}$	not sign.	21.8
Num ~ $f_1(SSB) + \varepsilon_{ij}$	0.086	22.2
Num ~ $f_1(SSB, by=factor(Age)) + factor(Age) + \varepsilon_{ij}$	Age 4: 0.012 Age 6: 0.0219	49.4
Gini ~ $f_1(SSB) + \varepsilon_{ij}$	0.804	0.3
Gini ~ $f_1(SSB, by=factor(Age)) + factor(Age) + \varepsilon_{ij}$	not sign.	19.3

7.4.3. Geographical spread

The geographical range of North Sea herring, described by the Centres of Gravity (COGs) with respect to Latitude and Longitude, as well as Inertia. We modelled these indicators as a function of SSB using GAM's. The smoother for COG Latitude is found to be a bell-shaped curve, with a peak northern distribution at intermediate population sizes (Fig. 5). COG Longitude is positively linear related to population size, hence the population has a tendency toward an easterly distribution with increasing SSB. Both smoothers for COGs in the GAM's were insignificant ($p > 0.05$). We found the relationship for Inertia to be negatively linear and significantly related to SSB ($p = 0.02$). The latter indicates that the spread of the population is decreasing with population size.

The dispersion of 4+ herring around the COG was decreasing with increasing SSBs. Figure 5 shows the geographical ranges of North Sea herring for the years 1990 to 2009. Overall, the mean spatial distribution of North Sea herring shows no strong variability. The COG of 4+ herring is relatively stable within a range of 106 n.mi in north-south and 63 n.mi in east-west directions. The ranges were calculated from the minimum and maximum values of the COG Latitudes and Longitudes. The Inertia in Figure 5 showed slightly more variation in some years and ranged between 80 and 110 n.mi.

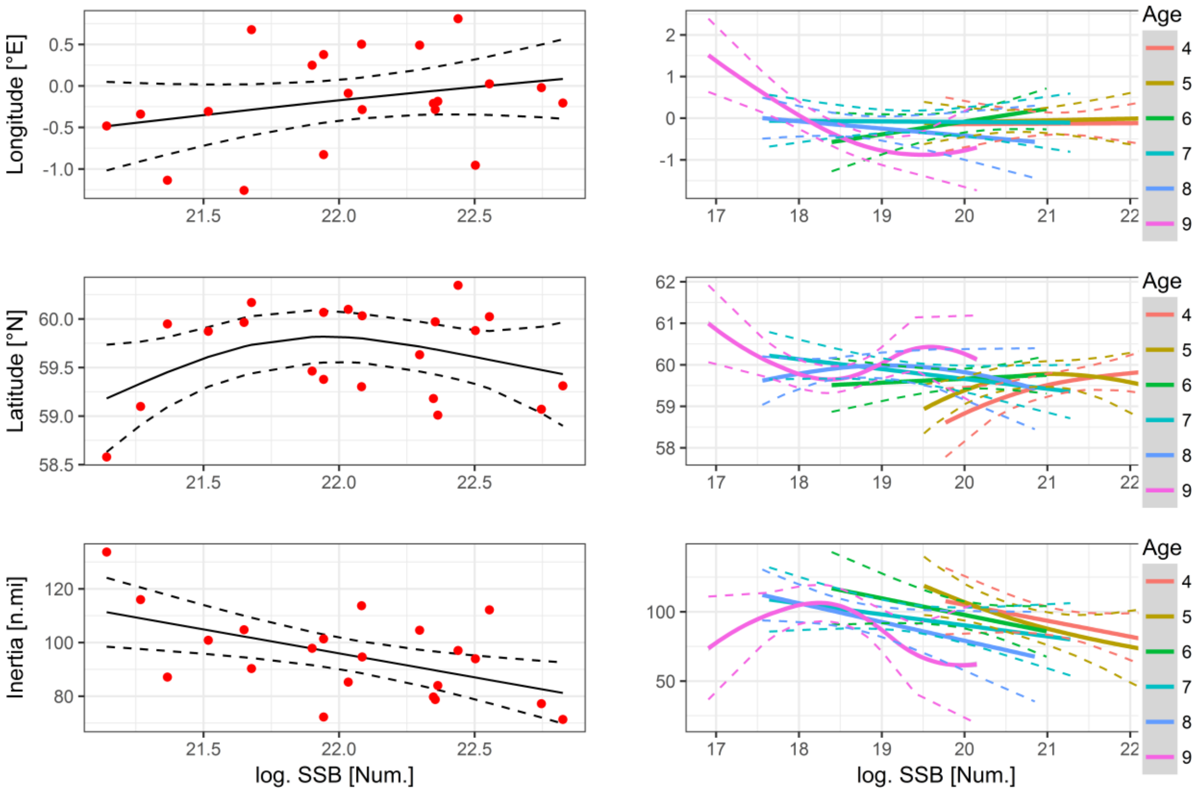

Figure 5. COGs (Latitude and Longitude) and the Inertia are shown for the aggregated (left) and the separated dataset (right). Smoothers and appertaining standard deviations are presented. Red dots show the original aggregated data.

Except for Inertia we found the age-specific relationships of the two COG indicators to population size differ. COG-Latitude increased for the younger age groups 4 to 6, but decreased for older ages 7 - 9. Especially age group 7 showed a strong tendency of a more southern spatial distribution with increasing abundances. Overall, the changes were relatively small considering the predicted range of about 1 °N (~60 n.mi) in total. The relationships for COG-Longitude were quite variable and negative for the oldest age groups 8 and 9, while positive for age group 6. The remaining age groups displayed no trend in COG-Longitude with population size. Considering the Inertia, no differences between the age groups were obvious. The differences between the age-specific relationships were tested statistically by varying coefficient models. Statistical significant differences existed only between age group 9 and the other age groups for the COG-Longitude (p-value 0.0002).

We found evidence for change points in our time series only for COG-Latitude. Significant change points (posterior probability ≥ 0.9) occurred 1990, 1995 and 2004. Additionally, varying coefficient models for the periods 1989 – 1990, 1991 – 1995, 1996 – 2004 and 2005 – 2009 revealed significant differences in the individual relationships between COG-Latitude and overall mature stock size. Here, the period 1991 – 1995 differed from the others with a more northern distribution in abundant years and a clear relationship between COG-Latitude and SSB existed.

7.5 **Discussion**

In the following, we will discuss the 3 analysed distribution characteristics of North Sea herring and explain our results of density-independent distribution patterns in the North Sea herring stock. Possible sampling errors and pro and cons of our analysis will be outlined.

7.5.1. Total occupied area

We described the occupied area of mature North Sea herring by the number of ICES squares with positive abundances in each year of the analysed period. Clupeids are known to show stock expansion in relation to stock size (Mac Call 1990). This effect is also known as positive “abundance-occupancy relationship” and has been demonstrated for other North Atlantic herring stocks (Dragesund et al. 1997, Fisher and Frank 2004). We expected similar abundance-occupancy relationships in our HERAS dataset which describes the distribution of North Sea autumn spawning herring during their feeding period in summer. Acoustic surveys give high resolution data of species distributions and are therefore very suitable for this kind of analysis. The HERAS outcome are numbers of herring per statistical ICES square at age and we used all squares with positive abundances for our analysis, either aggregated (summed herring 4+ abundances) or separated for each age group. To produce comparable results, we used a defined number of 164 ICES squares which was necessary due to sampling irregularities during the acoustic surveys caused by bad weather or other complications like technical problems. As a consequence, the most southern North Sea (< 54 °N) was not considered in our analysis. It is assumed that no bias was introduced as a result of this because mature North Sea herring is located in the summer feeding areas in the north-western North Sea (Cushing 1981, ICES 2014, Corten 2001). No abundance threshold (minimum considered abundance) was used because the calculated herring numbers per ICES square were generally high and representative.

Our results showed no clear abundance-occupancy relationship for North Sea herring. The occupied area did not increase with increasing SSBs. There were slight differences between the beginning of our time series and the end, indicated by a high probability of a change point in the year 1993. But no significantly different abundance-occupancy relationships were observed between both periods (1989 – 1993 and 1994 – 2009). A candidate reason for the lacking relationship would be the general state of the autumn spawning North Sea herring stock. The stock was back in 2010 harvested sustainably but at risk of having a reduced reproductive capacity (HAWG 2010), i.e. SSB was at a low level compared to periods before the stock collapse in the 1970’s. Since then, the SSB and recruitment exhibit great fluctuation. In the period 1997 to 2016 the SSB fluctuated between 1.1 to 2.3 million tons (ICES 2017b). Relating these findings to a distribution theory like the basin model (Mac Call 1990) suggests that there is currently no habitat limitation for North Sea herring since stock size is below carrying capacity and hence no abundance-occupancy relationship could be observed for the study period.

We additionally analysed age- specific relationships, but again found no significant differences. Only the visual inspection of our ANCOVA results showed differences between age group 9 and all other age groups. Age group 9 showed almost no relationship between the number of ICES squares and the total age group abundance, while all other groups showed an increasing relation. These differences were statistically significant. The abundance at age decreased approximately by half each year in a cohort. Thus, age 9 herring have the lowest abundance. The total occupied area of age 9 herring was very stable in our analysis, which could indicate conservative behaviour as described by Corten (2001). Ages 7 and 8 showed also a little different relationship of the area occupied compared to the youngest age groups (in general lower numbers of ICES squares, lower slope). But we could not show significant differences between those age groups, even if younger herring (ages 4 – 6) showed the highest abundances at age and a higher total occupied area. Overall, we found no positive abundance-occupancy relationships for North Sea herring as expected. The spatial distribution seemed to be stable during our observation period and independent from the stock size. Since the

SSB of North Sea herring was relatively low in the studied period, probably no habitat limitations occur currently.

7.5.2. Amount of aggregation

The next distribution characteristic we analysed in the context of density-dependent distribution patterns in North Sea herring was the amount of aggregation. Aggregations in specific areas are known for several fish populations at low stock sizes, making them more vulnerable to overexploitation (Beverton 1990, Hutchings 1996). Therefore, aggregation is of great importance when considering density-dependent mechanisms of fish stocks. We used the Gini-Index which can take values between 0 (uniformity) and 1 (unequality) and can be applied to any kind of distributions. We were not able to find density-dependent aggregation effects in our dataset. The relation between the Gini-Index and the SSB was linear for the aggregated data, but the (predicted) range of the Index was relatively low. All calculated values were higher than 0.6, indicating an unequal spatial distribution of mature herring in the North Sea. This was expected as most mature herring were found in the north-western North Sea. The effect was found to be independent from age. We hence conclude that no aggregations occurred in our North Sea herring. Similar to the analysis of the occupied area it seems that aggregation is not influenced by stock size at currently low SSBs.

Another reason that should be considered in our analysis, are the open boundaries of the HERAS survey to the North. The survey samples the entire North Sea during summer feeding, but the northern boundary is 62 °N. It is likely that a part of the population regularly crosses the boundary affecting the calculation of our spatial indicators.

Regarding overexploitation of the herring stock in the North Sea due to aggregations, there obviously exist a risk for the mature herring. The mature herring stock is unequally distributed with a core centre in the north-western North Sea, expressed by high Gini-Indices and small occupied areas. Thus, herring is concentrated in small areas (at least during summer feeding), making mature herring vulnerable to overexploitation. On the other hand, juvenile and adult herring have different habitats and do not occur in the same areas. Therefore, a total stock depletion is unlikely.

7.5.3. Geographical spread

Finally, we analysed the geographical spread of North Sea herring in relation to stock size. The geographical spread can be used as measure for spatial shifts or to analyse changes of the spatial distribution. Changes of the geographical spread of marine populations in relation to the population size are known for example for sprat in the Baltic Sea (Casini 2011). We used the Centres of Gravity (COG, Latitude and Longitude) to describe the geographical spread of North Sea herring. The COG is the mean of the location in each year weighted by the SSB (or age group abundance). Furthermore, we calculated the Inertia, describing the dispersion in square nautical miles around the COG. These spatial indicators were successfully used to describe the geographical spread for marine populations (Woillez 2007, Rindorf and Lewy 2012).

Our results gave a very consistent picture of North Sea herring geographical spread exhibiting generally low variability both in COGs and Inertia. The relation between spread and SSB was also not significant. The change point analysis gave a high probability of 3 change points for the COG-Latitude. The 4 periods seemed to be significantly different. The period from 1991 – 1995 showed a clear increasing relationship, in detail a more northern distribution in abundant years. This period was characterised by relatively small biomasses, lowest in 1995. Especially the COG in 1995 showed a very southern distribution compared to the other years. This could have produced a strong relation for this short period. No report from the “Working Group for International Pelagic Surveys” (WGIPS, formerly PGHERS) could be found. Therefore, information is lacking if problems occurred during the HERAS. The covered area was among average for this period. Nevertheless, bias could exist in the dataset and a historically low abundance of the North Sea herring stock is known – similar to 1978 – for 1995 (Simmonds 2007). An age effect was found only for COG-Longitude. Here again, age group 9 differed strongly from the other age groups. The varying coefficient model was slightly significant with a p-value of 0.012. Overall, the variability of the geographical spread was minimal and we can conclude here that the spatial distribution of herring in the North Sea showed conservative behaviour.

8. General Discussion

In my work, I analysed spatio-temporal distribution changes of a key species in the North Sea ecosystem: the Atlantic herring (*Clupea harengus*). The North Sea autumn spawning herring is commercially important for fisheries and additionally very important as a prey species for large fish, seabirds and marine mammals (ICES 2014). The stock was depleted in the 1970's and never fully recovered again (Simmonds 2007, Dickey-Collas et al. 2010). In the study period I considered (1989 to 2009), the North Sea herring stock was in safe management limits but produced remarkable weak recruitment (ICES 2017b, ICES 2014). The Spawning Stock Biomass (SSB) fluctuated between 1 and 3 million tonnes from 1989 to 2009 (ICES 2017b). In the same period, the North Sea appeared to be a "hot spot" for global warming and has warmed 1 – 2 °C since the 1980's (OSPAR 2010, Holt 2012). Since the North Sea fish community has undergone several changes in the last decades, I was interested in the response of North Sea herring to all those threats. Hence, my thesis was divided in 3 parts. How to model distribution changes of North Sea herring using a complex spatio-temporal dataset? Which environmental or climatic predictors can describe changes or shifts in the spatial distribution of North Sea herring? And does the North Sea herring stock size has an effect on spatial distribution? In the end, I developed a comprehensive analysis of North Sea herring behaviour in respect to density-independent and density-dependent factors. Such an analysis was done for the first time now for a period of 21 years and the entire North Sea.

8.1 Fisheries Acoustics and "International Herring Acoustic Survey" (HERAS) in the North Sea

To investigate changes in spatial distribution of North Sea herring, I used data from the "International Herring Acoustic Survey" (HERAS). This survey was established in the 1980's and is coordinated by the "International Council for the Exploration of the Sea" (ICES). Scientific echosounders (Simrad EK500, EK60 or EK80) are used to determine the density of herring schools along pre-defined transects in the entire North Sea during summer feeding in June and July. The spatial resolution of the survey are ICES statistical rectangles with a size of 0.5° in latitudinal and 1° longitudinal direction.

In my analysis I considered almost the entire North Sea with a total number of 164 rectangles. It was necessary to exclude the most southern part of the North Sea (< 54 °N) because this area was not covered regularly. The southern survey boundary was shifted southwards to find the distributional boundary of sprat (*Sprattus sprattus*). For my analysis, a regularly sampled area was more important for the statistical analysis than a larger area coverage. Bias could be produced in the analysis of immature herring, since young herring age groups occur preferably in coastal areas, consequently also in the southern North Sea. This could have introduced bias in the second part of my thesis. In the third part, only mature herring were considered. Mature herring were probably not affected by this area limitation as older herring migrate to the summer feeding areas in the north-western North Sea as I confirmed in my work and which was described by several authors (Cushing 1981, ICES 2014, Corten 2001).

The spatial resolution of ICES statistical rectangles was sufficient for my analysis of herring distribution in the entire North Sea. Since I focused on the complete autumn spawning herring stock in the North Sea, I disregarded small-scale effects, e.g. originating from frontal systems (Maravelias and Reid 1995, Maravelias and Reid 1997, Maravelias 1997). For this kind of analysis, NASC-values (Nautical Area Scattering Coefficients [$m^2/n.mi^2$]) were used as response variables in the statistical analysis. The NASC-values would have been available from all surveys in resolutions of nautical miles or 15 minutes. But environmental data was not available sufficiently to analyse small-scale effects for the entire North Sea. There was a lot of oceanographic data available from the ICES oceanographic data page with better data coverage in the end of the considered period (1989 – 2009). But overall, there was not enough environmental data available. Therefore, I decided to work with modelled environmental data from HAMSOM for Julian day 193 (the mid of HERAS surveys) for the spatial resolution of ICES statistical rectangles. For this reason I did not consider effects of frontal systems on the distribution of North Sea herring. I tried to dissolve frontal systems derived from satellite data (thanks to Peter Miller from Plymouth Marine Lab), but the ICES rectangle scale appeared to be too large. For all other aspects of my analysis, the scale of ICES rectangles was sufficient.

It turned out that data from HERAS was suitable for my spatio-temporal analysis of herring distribution in the entire North Sea. Standard settings can be found in the “Manual for International Pelagic Surveys (IPS)”, provided by the ICES “Working Group for International Pelagic Surveys” (WGIPS). Prior to the survey beginning, the transducers must be calibrated with the applied survey standard settings to ensure precise fish abundance estimates. Otherwise, a significant bias could be introduced by assuming constantly wrong abundance estimates (Simmonds and MacLennan 2005). The standard-target method from Foote (1987) is recommended for the calibration procedure, where a sphere with known acoustic properties (target strength) is measured in the sonic cone of the transducer and related to the backscattered acoustic energy. The advantages of a successfully calibrated system outperform the disadvantages, i.e. blind zones in the transducer near field and the bottom, scaring of fish due to vessel noise, dependency on relatively good weather to avoid air bubbles under the transducer and a relatively small sonic cone, compared to sonar systems. The blind zones are no big deal since herring schools are found generally in the pelagial close to but not at the bottom. The small sonic cone has its origin in the 7° circular beamwidth of the standard 38 kHz transducer (Simrad). But in practice, the distribution of herring is represented very well by acoustic surveys. Bad weather issues are a general problem when working on research vessels. During summer time, loss of time due to bad weather is in general a smaller problem. Also technical problems onboard are generally rare but nevertheless can cause severe problems when occurring, i.e. a smaller area coverage or an inappropriate number of identification hauls during the survey.

Overall, I would highly recommend data from acoustic surveys for “Species Distribution Models” in general. Especially the HERAS data gave a very high resolution and accurate herring abundance estimates for the entire North Sea (> 54° N). This data made it possible to develop an appropriate statistical method using “Generalised Additive Models” (GAM’s) to get insight in North Sea herring behaviour during summer feeding in the next step. Considering small-scale effects was not possible due to a lack of environmental data available. A higher number of CTD-casts during HERAS could be the next step in analysing small-scale effects for herring in the entire North Sea. In this context, a general discussion would be needed to outline a suitable spatial resolution of CTD casts. Especially in areas with high herring densities the number of casts should be increased to dissolve smaller effects. This would not be time-killing in the relatively shallow North Sea and would give high benefit. In areas with low herring densities a total number of two identification hauls is recommended. This would also be a sufficient number for CTD casts. At least, the hydrographic HERAS data should be available from a database, similar to the acoustic data (DATRAS). Such a database does not exist until now, hydrographic data storage is the responsibility of the HERAS participating nations (ICES 2016). At least, hydrographic HERAS data availability in the ICES oceanographic database would be an advantage and a starting point for analysing small-scale environmental effects.

8.2 “Species Distribution Model’s” (SDM’s) and statistical HERAS data issues

Since the spatio-temporal HERAS dataset covers almost the entire North Sea and a period of 21 years, a suitable method for “Species Distribution Modelling” should be used. I decided to use “Generalised Additive Models” (GAM’s) due to the high flexibility, especially in the very stable package “mgcv” (Wood 2011). GAM flexibility is given by allowing non-linear relationships between response and explaining variables, the use of any form of statistical distributions from the exponential family and the availability of several options to deal with statistical issues like autocorrelation or zero-inflation. Autocorrelation are measurements in the dataset that are close together both or either in time (temporal autocorrelation – TAC) or in space (spatial autocorrelation – SAC). Because of the lack of temporal patterns in our dataset, I will focus on SAC in the following discussion. Another important statistical issue of ecological datasets is zero-inflation (ZI). ZI could be a problem when the number of zero observations exceeds the expected number based on the underlying statistical distribution. In the following I will discuss ways of detecting and dealing with those important statistical issues using GAM's.

8.2.1. Spatial Autocorrelation

For checking SAC in the model residuals several options exist. I used Moran's I test statistic (Moran 1950) and additionally variograms for the ability of the detection of small-scale patterns. A big disadvantage of both methods is that each year has to be checked separately for SAC. Both methods showed serious SAC in the model residuals. GAM's offer 2 ways to deal with SAC. The implementation of a space smoother in the fixed term of the model or alternatively the extension to mixed models (GAMMs) by implementing an autocorrelation structure. Generally, checking for reduction of SAC in the model residuals with both methods separately and additionally together is the way to choose a final approach. Overall, if SAC is present in the model residuals, it is most important to take the patterns into account instead of finding the perfect solution (Zuur 2010). In my case, the extension to GAMMs showed almost no effect in reducing the spatial patterns. Here the origin of the spatial patterns was probably of importance. Spatial patterns can origin from 'true' or 'false' gradients. 'True' gradients, e.g. temperature gradients, lead to spatial dependency in the data while 'false' gradients, e.g. internal community driven factors lead to spatial autocorrelation (Legendre 1993, Borcard et al. 2011). In my thesis, I treated both types of spatial patterns as SAC as both have to be accounted for if present in the model residuals (Dormann 2007). In my models, the implementation of a space smoother including the geographical coordinates (Latitude and Longitude) reduced the SAC significantly. This indicates that a temperature gradient in the North Sea from the North to the South was probably responsible for the spatial patterns detected in my analysis of spatial herring distribution. Nevertheless, I presented a comprehensive overview to detect and deal with all kinds of SAC using GAM's.

8.2.2. Zero-Inflation

Another important issue that is often associated with ecological datasets is zero-inflation. Here, the characteristic of GAM's to use other error distributions than *Gaussian* turned out to be a great advantage in dealing with this statistical issue. Several distributions from the exponential family are able to deal with a higher number of zeros, i.e. *Poisson*, *Negative Binomial* and *Tweedie.* Alternatively, the package "*mgcv*" offers a *zero-inflated Poisson* (*ziP*) distribution which could be another alternative if zero-inflation exists in the dataset. Another option could be the use of binomial models with presence/absence data. But here a loss of information goes along with the transformation of abundances in presence data. If all these distributions did not fit the dataset, distributions of particular kind can be used, i.e. zero-inflated (or mixture) models or zero-altered models. The latter are also known as two-step or hurdle models, whereby the main difference between zero-inflated and zero-altered models is the way the zeros are treated. Two-step models are composed of two parts, a first binomial to model the probability of a zero count and a second zero-truncated (no zeros can be produced) part. The zero-inflated models work very similar, but the binomial part distinguishes between true and false zeros (the probability for a zero measure is implemented in the model). For both types, *Poisson* and *Negative Binomial* models are available (*ZIP, ZINB, ZAP, ZANB*). Particular packages are necessary to model *ZIP, ZINB, ZAP* or *ZANB* that did not support the model adaptions, i.e. space smoother or autocorrelation structures. Therefore I did not use those packages but applied zero-altered models manually by multiplying the outcomes from the binomial part (probabilities) with the outcomes from the presence only part. The 2-step models showed good overall performance and are a very suitable option for zero-inflated datasets using GAM's. Nevertheless, only predictions are possible if applying the zero-altered model manually and no explicit information of the covariates effects are given from the combined model. In the end, many options to deal with ZI existed. It turned out in my analysis, that the simplest model with log-transformed response variable did the best model fit. Also the *Tweedie* distribution showed good model performances similar to the findings from Augustin *et* al. (2013). Furthermore, the *ziP* distribution from "mgcv" showed good performances. For count data with many zeros, the *Negative Binomial* distribution is recommended without transforming the response in general (Warton 2008). Binomial models are often recommended for zero-inflated datasets as the distribution is stable against overdispersion. But for my opinion the disadvantage of losing ecological information due to transformations if abundances were available would be too large. Here the combination of both models in form of a two-step model should be preferred.

Overall, GAM's turned out to be very suitable for modelling complex spatio-temporal datasets like the HERAS data. I would highly recommend the simplest models with Gaussian distribution at least as a starting point for the analysis. From here, model diagnostics and above discussed tests should be used to further improve and adapt the model, i.e. by changing the distribution or implementing additional smoothers or structures to reduce existing statistical issues. Especially when it comes to model adaptions, I want to highlight the strengths of GAM's. A diverse number of options exist, as already described to account for special issues of the dataset. And furthermore, the adaptation and structure implementation within "mgcv" is very easy and does not need many lines of code and usually does not decrease overall performance. But this is definitely depending on the kind of ecological data and model characteristics and complexity used. In the case problems occur in adjusting GAM's, several easy ways to find help exist. Despite very useful textbooks (Wood 2006, Dalgaard 2008, Zuur 2009, Borcard 2011) a huge community of users is working with GAM's in "mgcv". In my experience, for most occurring problems, solutions can be found easily in particular blogs. Here, a positive experience was that many questions were answered by competent people, very often by "mgcv" developer Simon Wood. So there is an active community with a lot of assistance and constantly improved software. All those experiences together let me state that "Generalised Additive Models" have many advantages in modelling ecological dataset and I recommend GAM's for "Species Distribution Modelling" due to their flexibility and a variety of adaptations. Starting with simple models, developing more suitable complex models and combine the knowledge from both should be a very good way for interpretation of the ecological mechanisms and meanings behind the analysis.

8.3 Spatio-temporal distribution of North Sea herring in the period 1989 – 2009

I did several analysis of the spatio-temporal distribution of North Sea herring in the period 1989 to 2009 based on the HERAS dataset. In the second and third part of my thesis I investigated density-independent and density-dependent mechanisms influencing the herring distribution. Here, the analysis can be distinguished into 2 main parts. Firstly, I investigated temperature preferences at age and additionally environmental predictors that potentially could have an influence on the distribution of North Sea herring. Secondly, I did an analysis of density-independent distribution shifts in relation to mean summer temperatures of the North Sea, the North Atlantic Oscillation winter index (NAO, both in manuscript 2) and furthermore an analysis of density-dependent distribution shifts of North Sea herring in relation to the total stock size (SSB, manuscript 3).

8.3.1. Environmental Predictors

Concerning the analysis of temperature preferences and environmental predictors, separating immature and mature herring was very important for my analysis. Immature herring occur preferably in coastal areas before reaching maturity and start migrating with the adult stock. This was confirmed in my analysis, since temperature preferences of immature herring of ages 0 and 1 were higher and the environmental models had lower significance levels of the models, indicating that the model cannot explain the variance of immature herring distribution clearly. I would explain this with the high dynamics of coastal waters influenced strongly by tidal currents and river runoff. Herring of the ages 2 and 3 depicted a kind of intermediate stage in my analysis. This was mainly due to my assumptions of full maturity of both ages. This is at least not true for herring of age 2. About 70 to 80 % of age 2 herring is assumed to be mature (HAWG 2015), but variability is high. Due to the high percentage of mature herring in these age groups and general uncertainties in dividing immature and mature herring (sample size, staging in the lab and associated assumptions), I treated all age 2 and 3 herring as mature, in the knowledge that bias is introduced in the analysis. This bias was relatively small considering the temperature preferences. All herring older than age 2 showed similar preferences with only marginal differences. Considering the environmental predictors, herring of age 2 showed clearly different results with zooplankton and SBT as significant environmental predictors, compared to ages older than 4. Age 3 herring model results were similar to all older herring age groups. After getting mature, herring start migrating. During the investigated period, adult herring is located in the summer feeding areas in the north-western North Sea (Cushing 1981, ICES 2014, Corten 2001). This was confirmed by the temperature preferences in the analysis. Herring older than age 4 showed clear preferences for cold water. The most important environmental predictors in the feeding area in the north-western North Sea were the sea surface temperature (SST) and the depth of the bottom. Overall, clear differences were found between immature and mature herring with ages 2 and 3 depicting intermediate stages. Temperature was highlighted as the most important environmental predictor for herring distribution in my analysis.

8.3.2. Distribution Shifts

Finally, I investigated distribution shifts of North Sea herring in relation to density-independent (manuscript 2) and density-dependent (manuscript 3) mechanisms. I used the mean summer temperature of the North Sea and the NAO winter index as density-independent factors. The spawning stock biomass (SSB) was used as density-dependent factor. All environmental and climatic factors were related to the “Centre of Gravity”, describing the weighted mean Latitudes and Longitudes of all herring age groups. The COGs of all mature herring were similar and as expected located in the north-western North Sea. For this reason, the analysis was performed for all mature age groups together (4+ herring), representing the total SSB based on the HERAS data. The mean summer temperature difference in the investigated 21 years was 1.5 °C. This represented the documented rate of warming of 1 – 2 °C over the last 25 years in the North Sea very good (IPCC 2014). Nevertheless, the temperature increase explained only 9 % of the model variance. The maximal range shift in latitudinal and longitudinal direction was about 291 km and 171 km, respectively. Increases in temperature were not distinctly related to northward shifts in herring distribution in my analysis. This was surprising, since distribution shifts of marine fish species to higher latitudes are known for many species (Richardson 2008, Cheung et al. 2013, Jones and Cheung 2014, Gamito et al. 2015) and were also shown for North Sea species (Perry 2005). The analysis of distributional shifts in relation to the NAO winter index gave inconsistent results and the NAO showed very high variations in the study period. It was not possible to relate the NAO winter index to shifts in herring distribution. This was probably due to the relatively short time series for linking large scale climatic effects like the NAO. The last step was the investigation of density-dependent distribution patterns in North Sea herring. The SSB in the study period fluctuated around 1 and 3 million tons. I used a set of 5 spatial indicators to study the distribution characteristics of North Sea herring: (i) the amount of aggregation (Gini-Index), (ii) the occupied area (number of ICES statistical rectangles with positive abundances) and (iii) the geographical spread (Centre of Gravity and Inertia). None of the spatial indicators showed a relationship to the fluctuating SSB in the study period 1989 to 2009. In addition, the age group had no effect on spatial distribution patterns of North Sea herring. When considering the overall stock status, SSB of autumn spawning North Sea herring is relatively low since the collapse in the 1970's. The poor stock status is associated with recruitment below average and only 2 strong year classes (1998 and 2000). The stock status may be the main reason for the small variations in spatial distribution. I would conclude here, that no density dependent habitat selection occurred and no resource limitations existed in the study period. Relating these findings to a distribution theory like the basin model (Mac Call 1990) the stock size seemed to be below the carrying capacity of the north-western North Sea.

Considering North Sea herring biology, expectations based on prevailing knowledge about herring temperature preferences at age and distribution ranges at age were confirmed. Young and immature herring were located close to the coastal areas with generally higher temperatures, while mature herring older than age 2 were found mainly in the north-western North Sea at clearly lower temperatures. Hence, the spatio-temporal HERAS dataset was suitable to describe the biology of North Sea herring in combination with the modeled HAMSOM temperature data. Thus, both datasets were a good starting point for SDM's. As I already described, several environmental parameters are known to influence the distribution of herring. Among those, the most important is temperature. I confirmed the importance of temperature for herring distribution in the North Sea. Since herring is a migrating species, the time of the year I considered was the summer feeding period. Herring was located in the north-western North Sea as described by Cushing (1981) or Corten (2001). This area was characterised in the study period by relatively low temperature variations. The biggest difference in temperature between the coldest and warmest years was 1.5 °C. We could explain distribution shifts of North Sea herring only to a small degree. This could have several reasons. One explanation could be the relatively stable temperature regime in the north-western North Sea. This stability did not cause herring to respond to higher temperatures on a larger scale. But on the other side, the maximum range shift in north-south direction that I found was 291 km. In this context, mature herring of all ages showed similar behaviour and the range shift is not neglectable. Nevertheless, the statistical models explained only a small part of this distribution shift variance. Here, the "conservatism" theory from Corten (2001) could be an explanation. Even when environmental parameters are slightly changing, herring tend to adhere to the existing migration pattern. This effect could have masked an existing environmental effect and could make it difficult to demonstrate environmental impacts. Another explanation could be the open northern boundary of the North Sea to the Atlantic. If herring would shift to the North as response to higher temperatures, it would not be possible to find this effect within the HERAS dataset. The northern boundary of HERAS is at 62 °N. However, the low variations of abundances at age and replicable survey results between the years in the whole period did not show evidence for herring shifts regularly exceeding the northern HERAS boundary. Comparing my findings to responses of other marine fish species, clearly showing northward distribution shifts with increasing temperatures (Perry 2005, Richardson 2008, Cheung et al. 2013, Jones and Cheung 2014, Gamito et al. 2015), I conclude here, that the impact of global warming on herring distribution seems to be small, at least for autumn spawning North Sea herring at currently low stock sizes. But there are many uncertainties for drawing reliable conclusions in this context. As already mentioned, the northward shift higher 62 °N cannot be estimated precisely based on the HERAS data, even if the effect is expected to be neglectable. Furthermore, I did not consider the effect of Atlantic inflow in my analysis. This could potentially have major effects for herring, i.e. changes in zooplankton distribution (Corten 2001). The zooplankton distribution is another important point here. Prey abundance in my analysis was reproduced based on modelled data from the “Continuous Plankton Recorder” (CPR, SAHFOS). Large areas of the North Sea are interpolated and more precise zooplankton abundance estimates could improve SDM’s of North Sea herring. Overall, my SDM’s showed good performance and gave good insight into herring biology. Temperature is an important driver for North Sea herring but conservative migration patterns seemed to mask environmental effects. Low total stock size could further contribute to this “conservatism” effect, since obviously no resource limitations occurred. This conservative behaviour was confirmed by my findings of the spatial herring distribution. Neither a distribution shift nor an increase in the distribution range was indicated by my analysis. Relating these findings to potential future impacts in respect to increasing temperatures, i.e. for the management of the North Sea stock, autumn spawning herring seems to be resilient to global warming.

9. References

Akaike, H. (1974). ”A new look at the statistical model indentification''. *IEEE Transactions on Automatic Control*, 19:716--722.

Augustin, N. H., V. M. Trenkel, S. N. Wood and P. Lorance (2013). "Space-time modelling of blue ling for fisheries stock management." Environmetrics 24(2): 109-119.

Austin, M. (2007). "Species distribution models and ecological theory: A critical assessment and some possible new approaches." Ecological Modelling 200(1-2): 1-19.

Backhaus, J. (1985). "A three-dimensional model for the simulation of shelf sea dynamics." Deutsche Hydrografische Zeitschrift 38(4): 165-187.

Bailey, M. C., C. D. Maravelias and E. J. Simmonds (1998). "Changes in the spatial distribution of autumn spawning herring (Clupea harengus L.) derived from annual acoustic surveys during the period 1984-1996." Ices Journal of Marine Science 55(3): 545-555.

Barbet-Massin, M., F. Jiguet, C. H. Albert and W. Thuiller (2012). "Selecting pseudo-absences for species distribution models: how, where and how many?" Methods in Ecology and Evolution 3(2): 327-338.

Beale, C. M., J. J. Lennon, J. M. Yearsley, M. J. Brewer and D. A. Elston (2010). "Regression analysis of spatial data." Ecology Letters 13(2): 246-264.

Beaugrand, G., P. C. Reid, F. Ibañez, J. A. Lindley and M. Edwards (2002). "Reorganization of North Atlantic Marine Copepod Biodiversity and Climate." Science 296(5573): 1692-1694.

Beaugrand, G., E. Goberville, C. Luczak and R.R. Kirby (2014). “Marine biological shifts and climate.” *Proc. R. Soc. B* **281**:20133350.

Belkin, I. M. (2009). "Rapid warming of Large Marine Ecosystems." Progress in Oceanography 81 (1–4): 207-213.

Beverton, R.J.H. (1990). Small marine pelagic fish and the threat of fishing; are they endangered? J. Fish Biol. 37(Suppl. sA): 5–16. doi:10.1111/j.1095-8649.1990.tb05015.x.

Breiman, Leo, Jerome Friedman, R. Olshen and C. Stone (1984). “Classification and Regression Trees.” Belmont, California: Wadsworth.

Broms, C., W. Melle and J. K. Horne (2012). "Navigation mechanisms of herring during feeding migration: the role of ecological gradients on an oceanic scale." Marine Biology Research 8 (5-6): 461-474.

Brunel, T. and M. Dickey-Collas (2010). "Effects of temperature and population density on von Bertalanffy growth parameters in Atlantic herring: a macro-ecological analysis." Marine Ecology Progress Series 405: 15-28.

Bivand, S. R., E. J. Pebesma and V. Gómez-Rubio (2008). “Applied Spatial Data Analysis with R”. 2008 Springer Science + Business Media, LLC.

Borcard, D., F. Gillet and P. Legendre (2011). “Numerical Ecology with R”. 2011 Springer Science + Business Media, LLC.

Casini, M., G. Kornilovs, M. Cardinale, C. Moellmann, W. Grygiel, P. Jonsson, T. Raid, J. Flinkman and V. Feldman (2011). "Spatial and temporal density dependence regulates the condition of central Baltic Sea clupeids: compelling evidence using an extensive international acoustic survey." Population Ecology 53(4): 511-523.

Cheung, W. W. L. L., Vicky W Y; Pauly, Daniel (2008). "Modelling present and climate-shifted distribution of marine fishes and invertebrates." University of British Columbia Fisheries Centre Research Reports 16(3): 1-72.

Cheung, W. W. L., R. Watson and D. Pauly (2013). "Signature of ocean warming in global fisheries catch." Nature 497(7449): 365-368.

Clark, R.A. and C.L.J. Frid (2011). “Long Term Changes in the North Sea Ecosystem.” *Environmental Reviews, 2001, 9(3): 131.187.*

Cochrane, K., C. De Young, D. Soto, T. Bahri (2009). “Climate Change implications for fisheries and aquaculture: overview of current scientific knowledge.” FAO Fisheries and Aquaculture Technical Paper No. 530.

Corten, A. (2001). "Northern distribution of North Sea herring as a response to high water temperatures and/or low food abundance." Fisheries Research 50(1-2): 189-204.

Corten, A. (2001). "The role of "conservatism" in herring migrations." Reviews in Fish Biology and Fisheries 11(4): 339-361.

Cushing, D.H. (1981). “Fisheries Biology. A Study in Population Dynamics.” The University of Wisconsin Press. Second Edition.

Daan, N., Bromley, P. J., Hislop, J. R. G., and Nielsen, N. A. (1990). “Ecology of North Sea fish.” Netherlands Journal of Sea Research 26(2-4): 343-386.

Dickey-Collas, M., R. D. M. Nash, T. Brunel, C. J. G. van Damme, C. T. Marshall, M. R. Payne, A. Corten, A. J. Geffen, M. A. Peck, E. M. C. Hatfield, N. T. Hintzen, K. Enberg, L. T. Kell and E. J. Simmonds (2010). "Lessons learned from stock collapse and recovery of North Sea herring: a review." Ices Journal of Marine Science 67(9): 1875-1886.

Doney, S. C., M. Ruckelshaus, J. E. Duffy, J. P. Barry, F. Chan, C. A. English, H. M. Galindo, J. M. Grebmeier, A. B. Hollowed, N. Knowlton, J. Polovina, N. N. Rabalais, W. J. Sydeman and L. D. Talley (2012). "Climate Change Impacts on Marine Ecosystems." Annual Review of Marine Science 4(1): 11-37.

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

Dragesund, O., A. Johannessen and 0. Ulltang (1997). Variation in migration and abundance of Norwegian spring spawning herring *(Clupea harengus* L.). Sarsia 82:97-105. Bergen. ISSN 0036-4827.

Dulvy, N.K., S.I. Rogers, S. Jennings, V. Stelzenmüller, S.R. Dye and H.R. Skjoldal (2008). “Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas.” *Journal of Applied Ecology* 2008, 45, 1029-1039.

Dunn, P. and G. Smyth (2005). "Series evaluation of Tweedie exponential dispersion model densities." Statistics and Computing 15(4): 267-280.

Elith, J. and J. R. Leathwick (2009). "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." Annual Review of Ecology, Evolution, and Systematics 40(1): 677-697.

Elith, J., J. R. Leathwick and T. Hastie (2008). "A working guide to boosted regression trees." Journal of Animal Ecology 77(4): 802-813.

Fisher, J. A. D. and K. T. Frank (2004). "Abundance-distribution relationships and conservation of exploited marine fishes." Marine Ecology Progress Series 279: 201-213.

Fox, J. and Sanford Weisberg (2011). An R Companion to Applied Regression, Second Edition. Thousands Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Fromentin, J. M. and B. Planque (1996). "Calanus and environment in the eastern North Atlantic. II. Influence of the North Atlantic Oscillation on C. finmarchicus and C. helgolandicus." Marine Ecology Progress Series 134: 111-118.

Gamito, R., M. J. Costa and H. N. Cabral (2015). "Fisheries in a warming ocean: trends in fish catches in the large marine ecosystems of the world." Regional Environmental Change 15(1): 57-65.

Gaston, K. J., T. M. Blackburn, J. J. D. Greenwood, R. D. Gregory, R. M. Quinn and J. H. Lawton (2000). "Abundance-occupancy relationships." Journal of Applied Ecology 37: 39-59.

Gregory, S. D., B. W. Brook, B. Goossens, M. Ancrenaz, R. Alfred, L. N. Ambu and D. A. Fordham (2012). "Long-Term Field Data and Climate-Habitat Models Show That Orangutan Persistence Depends on Effective Forest Management and Greenhouse Gas Mitigation." PLoS ONE 7(9): e43846.

Guisan, A. and N.E. Zimmermann (2000). „Predictive habitat distribution models in ecology.“ *Ecological Modelling 135* (2000) 147-186.

Handcock, M. S. (2016). Relative Distribution Methods. Version 1.6-6. Project home page at http://www.stat.ucla.edu/~handcock/RelDist.

Hastie, T. J. and R. J. Tibshirani (1990). “Generalized Additive Models”. Chapman and Hall/CRC.

Hawkins, B. A. (2012). "Eight (and a half) deadly sins of spatial analysis." Journal of Biogeography 39(1): 1-9.

Hays, G. C., A. J. Richardson and C. Robinson (2005). "Climate change and marine plankton." Trends in Ecology & Evolution 20(6): 337-344.

Hemelrijk, C.A. and H. Kunz (2004). “Density distribution and size sorting in fish schools: An individual-based model.” Behavioral Ecology. doi:10.1093/beheco/arh149.

Hijmans, R.J. and J. Elith (2017). “Species distribution modelling with R.” January 8, 2017. https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf

Hoare, D.J., G.D. Ruxton, J.G.J. Godin and J. Krause (2000). “The social organization of free- ranging fish schools.” Oikos 89, 546–554.

Holt, J., S. Hughes, J. Hopkins, S. L. Wakelin, N. Penny Holliday, S. Dye, C. González-Pola, S. S. Hjøllo, K. A. Mork, G. Nolan, R. Proctor, J. Read, T. Shammon, T. Sherwin, T. Smyth, G. Tattersall, B. Ward and K. H. Wiltshire (2012). "Multi-decadal variability and trends in the temperature of the northwest European continental shelf: A model-data synthesis." Progress in Oceanography 106(0): 96-117.

Hurrell, J. W.and C. Deser (2010). „North Atlantic climate variability: The role of the North Atlantic Oscillation.” J. Mar. Syst., 78, No. 1, 28-41.

Hutchings, J.A. (1996). “Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock’s collapse.” Can. J. Fish. Aquat. Sci. 53(5): 943–962. doi:10.1139/f96-097.

ICES (2014). Report of the Working Group of International Pelagic Surveys (WGIPS), 20-24 January 2014. ICES CM 2014/SSGESST:01. 360 pp.

ICES (2014). Report of the Herring Assessment Working Group for the Area South of 62ºN (HAWG), 11-20 March 2014, ICES HQ, Copenhagen, Denmark. ICES CM 2014/ACOM:06. 1257 pp.

ICES (2015). Manual for International Pelagic Surveys (IPS). Series of ICES Survey Protocols SISP 9 – IPS. 92 pp.

ICES (2016). ICES Acoustic Database Specification Document. http://www.ices.dk/marinedata/Documents/Specification%20of%20the%20acoustic%20database.pdf

ICES Advice (2016), Book 6. ICES Ecosystem Overviews. Greater North Sea Ecoregion.

ICES Advice (2017a). ICES Fisheries Overviews. Greater North Sea Ecoregion.

ICES Advice (2017b). ICES Advice on fishing opportunities, catch and effort. Greater North Sea Ecoregion.

IPCC (2014). “Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)].” IPCC, Geneva, Switzerland, 151 pp.

Jones, M. C. and W. W. L. Cheung (2014). "Multi-model ensemble projections of climate change effects on global marine biodiversity." ICES Journal of Marine Science: Journal du Conseil.

Keele, L. J. (2008). “Semiparametric Regression for the Social Sciences”. Wiley and Sons Ltd.

Kühn, I. and C. F. Dormann (2012). "Less than eight (and a half) misconceptions of spatial analysis." Journal of Biogeography 39(5): 995-998.

Le Fèvre, J. (1986). “Aspects of the biology of frontal systems.” Adv. mar. Biol. 23, 164-299.

Legendre, P. (1993). "Spatial Autocorrelation - Trouble or new Paradigm.” Ecology 74(6): 1659-1673.

Lorentzen, T. and R. Hannesson (2004). “The Collapse of the Norwegian Herring Fisheries in the Late 1950s and 60s. Crisis, Adaptation, and Recovery.” Working Paper No. 12/04. Institute for Research in Economics and Business Administration. Bergen. ISSN 1503-2140.

MacCall, A. D. (1990). “Dynamic Geography of Marine Populations.” Washington Sea Grant Press, Seattle, WA. 153 pp.

Maravelias, C. D. and J. Haralabous (1995). "Spatial distribution of herring in the Orkney Shetland area (Northern North Sea): A geostatistical analysis." Netherlands Journal of Sea Research 34(4): 319-329.

Maravelias, C. D. and D. G. Reid (1995). "Relationship between herring (*Clupea harengus*) distribution and sea surface salinity and temperature in the northern North Sea." Scientia Marina 59(3-4): 427-438.

Maravelias, C. D. (1997). "Trends in abundance and geographic distribution of North Sea herring in relation to environmental factors." Marine Ecology-Progress Series 159: 151-164.

Maravelias, C. D. and D. G. Reid (1997). "Identifying the effects of oceanographic features and zooplankton on prespawning herring abundance using generalized additive models." Marine Ecology Progress Series 147: 1-9.

Maravelias, C. D., D. G. Reid and G. Swartzman (2000). "Modelling spatio-temporal effects of environment on Atlantic herring, Clupea harengus." Environmental Biology of Fishes 58(2): 157-172.

Maravelias, C. D., D. G. Reid and G. Swartzman (2000). "Seabed substrate, water depth and zooplankton as determinants of the prespawning spatial aggregation of North Atlantic herring." Marine Ecology-Progress Series 195: 249-259.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre and H. P. Possingham (2005). "Zero tolerance ecology: improving ecological inference by modelling the source of zero observations." Ecology Letters 8(11): 1235-1246.

Mathis, M. and T. Pohlmann (2014). "Projection of physical conditions in the North Sea for the 21st century." Climate Research 61(1): 1-17.

Moran, P. A. P. (1950). “Notes on Continuous Stochastic Phenomena”. *Biometrika*. 37 (1): 17 – 23.

Muus, B.J. and J.G. Nielsen (1999). „Die Meeresfische Europas in Nordsee, Ostsee und Atlantik“. Franckh-Kosmos Verlags-GmbH & Co., Stuttgart.

National Center for Atmospheric Research Staff (Eds). Last modified 08 Aug 2017. "The Climate Data Guide: Hurrell North Atlantic Oscillation (NAO) Index (Station-based)."

Nelder, J. A. and R. W. M. Wedderburn (1972). "Generalised Linear Models." Journal of the Royal Statistical Society. Series A (General) 135(3): 370-384.

OSPAR (2010). “Quality Status Report 2010.” OSPAR Commission. London. 197 pp.

Parrish, B. B., Saville, A., Craig, R. E., Baxter, I. G., and Priestley, R. 1959. Observations on herring spawning and larval distribution in the Firth of Clyde in 1958. Journal of the Marine Biological Association of the UK, 38: 445 – 453.

Pätsch, J. and W. Kühn (2008). „Nitrogen and carbon cycling in the North Sea and exchange with the North Atlantic – a model study. Part I.” Nitrogen budget and fluxes, Cont. Shelf Res., 28, 767–787, doi:10.1016/j.csr.2007.12.013.

Parmesan, C. (2006). "Ecological and Evolutionary Responses to Recent Climate Change." Annual Review of Ecology, Evolution, and Systematics 37(1): 637-669.

Perry, A. L., P. J. Low, J. R. Ellis and J. D. Reynolds (2005). "Climate change and distribution shifts in marine fishes." Science 308(5730): 1912-1915.

Phillips, S., M. Dudik, J. Elith, C. Graham, A. Lehmann, J. Leathwick and S. Ferrier (2009). "Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data." Ecological Applications 19(1): 181-197.

Phillips, S. J., M. Dud, #237 and R. E. Schapire (2004). A maximum entropy approach to species distribution modeling. Proceedings of the twenty-first international conference on Machine learning. Banff, Alberta, Canada, ACM: 83.

Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2014). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-117, http://CRAN.R-project.org/package=nlme.

Pohlmann, T. (1996). "Predicting the thermocline in a circulation model of the North sea — part I: model description, calibration and verification." Continental Shelf Research 16(2): 131-146.

Potts, J. M. and J. Elith (2006). "Comparing species abundance models." Ecological Modelling 199(2): 153-163.

R Core Team (2014). “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.

Richardson, A. J. (2008). "In hot water: zooplankton and climate change." ICES Journal of Marine Science: Journal du Conseil 65(3): 279-295.

Rigby R.A. and Stasinopoulos D.M. (2005). Generalised additive models for location, scale and shape,(with discussion), Appl. Statist., 54, part 3, pp 507-554.

Rindorf, A. and P. Lewy (2012). "Estimating the relationship between abundance and distribution." Canadian Journal of Fisheries and Aquatic Sciences **69**(2): 382-397.

Saville, A., and R. S. Bailey (1980). “The assessment and management of the herring stocks in the North Sea and to the west of Scotland.” Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer 177: 112-142.

Schapire, R. (2003). The Boosting Approach to Machine Learning: An Overview. Nonlinear Estimation and Classification. D. Denison, M. Hansen, C. Holmes, B. Mallick and B. Yu, Springer New York. 171: 149-171.

Schmidt, J. O., C. J. G. Van Damme, C. Roeckmann and M. Dickey-Collas (2009). "Recolonisation of spawning grounds in a recovering fish stock: recent changes in North Sea herring." Scientia Marina 73: 153-157.

Schrum C**,** Backhaus J. O**.** (1999). „Sensitivity of atmosphere-ocean heat exchange and heat content in the North Sea and the Baltic Sea.” *Tellus A* 51(4): 526–549.

Schweigert, J. F., J. L. Boldt, L. Flostrand and J. S. Cleary (2010). “A review of factors limiting recovery of Pacific herring stocks in Canada.” ICES Journal of Marine Science, 67: 1903–1913.

Sherman, K and G. Hempel (Editors) (2009). The UNEP Large Marine Ecosystem Report: “A perspective on changing conditions in LMEs of the world’s Regional Seas.” UNEP Regional Seas Report and Studies No. 182. United Environment Programme. Nairobi, Kenya.

Simmonds, E. J. (2007). “Comparison of two periods of North Sea herring stock management: success, failure, and monetary value.” – ICES Journal of Marine Science, 64: 686–692.

Simmonds, J. and D. MacLennan. “Fisheries Acoustics. Theory and Practice.” Second Edition. Blackwell Science.

Teal, L.R. (2011). “The North Sea fish community: past, present and future.” Background document for the 2011 National Nature Outlook. Wageningen UR.

Verberk, W. (2011). “Explaining General Patterns in Species Abundance and Distributions.” Nature Education Knowledge 3(10): 38

Volkenandt, M., S. Berrow, I. O'Connor, J.-M. Guarini and C. O'Donnell (2015). "Prespawning herring distribution in the Irish Celtic Sea between 2005 and 2012." ICES Journal of Marine Science: Journal du Conseil 72(2): 498-507.

Walther, G.-R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J.-M. Fromentin, O. Hoegh-Guldberg and F. Bairlein (2002). "Ecological responses to recent climate change." Nature 416(6879): 389-395.

Wang, X. and J.W. Emerson (2015). “Bayesian Change Point Analysis of Linear Models on General Graphs.” Working Paper.

Warton, D. I. (2005). "Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data." Environmetrics 16(3): 275-289.

Wisz, M. S., J. Pottier, W. D. Kissling, L. Pellissier, J. Lenoir, C. F. Damgaard, C. F. Dormann, M. C. Forchhammer, J.-A. Grytnes, A. Guisan, R. K. Heikkinen, T. T. Høye, I. Kühn, M. Luoto, L. Maiorano, M.-C. Nilsson, S. Normand, E. Öckinger, N. M. Schmidt, M. Termansen, A. Timmermann, D. A. Wardle, P. Aastrup and J.-C. Svenning (2013). "The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling." Biological Reviews 88(1): 15-30.

Woillez, M., J. Rivoirard and P. Petitgas (2009). "Notes on survey-based spatial indicators for monitoring fish populations." Aquatic Living Resources 22(02): 155-164.

Wood, S. (2006). “Generalized Additive Models: An introduction with R”. Chapman & Hall/ CRC Texts in Statistical Science. 1st Edition.

Wood, S.N. (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.” Journal of the Royal Statistical Society (B) 73(1):3-36.

Wood, S. and F. Scheipl (2013). “Generalized Additive Mixed Models using lme4 and mgcv”. Package “gamm4” R documentation.

Wood, S. (2016). “Generalized additive models with integrated smoothness estimation.” Package “mgcv” R Documentation.

Yee, T. W. and C. J. Wild (1996). Vector Generalised Additive Models. Journal of Royal Statistical Society, Series B, 58(3), 481-493.

Zuur, A. F., E. N. Ieno and C. S. Elphick (2010). "A protocol for data exploration to avoid common statistical problems." Methods in Ecology and Evolution 1(1): 3-14.

Zuur, A.F., E. N. Ieno, N.J. Walker, A.A. Saveliev and G.M. Smith (2009). Mixed Effects Models and Extensions in Ecology with R. Springer Science+Business Media, LLC 2009.

10.Acknowledgements

First, I would like to thank my supervisor Prof. Christian Möllmann for helping me with the process of finishing my dissertation and for funding me for 6 years. Furthermore, I would like to thank Dr. Saskia Otto for great support in ecological and statistical questions and suggesting new approaches when my dissertation stucked. Without the input and data supply from the Thünen Institute for Sea Fisheries this thesis would not have been possible. Thanks to: Dr. Norbert Rohlf, Dr. Heino Fock, Dr. Matthias Schaber, Dr. Eckhard Bethke and Eberhard Götze. Furthermore, I have to thank all my great colleagues for the fantastic time at the IHF! Special thanks to Prof. Axel Temming for important support and input during the finalisation of my dissertation. Also special thanks to Dr. Ute Jacob for great support and feedback always within minutes and being the best office mate ever! Finally, I want to thank Dr. Jens Floeter who supported me anytime, gave me new insights in all aspects of my dissertation and funded me for 2 further years in the project “synKoFiMo”. Thank you for all the great surveys and Triaxus night watches onboard “RV Heincke”!

11.Eidesstattliche Versicherung, Declaration on oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

____

Heidekamp, the 10th April 2018