,

SPATIAL AND SPATIO-TEMPORAL MODELS FOR USE IN THE MARINE ENVIRONMENT WITH APPLICATIONS TO THE SCOTIAN SHELF

by

Stuart Carson

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
December 2018

To Pam and Evie

Table of Contents

List of Tables . . . . . **vi**

List of Figures . . . . . **viii**

Abstract . . . . . **xii**

List of Abbreviations and Symbols Used . . . . . **xiii**

Acknowledgements . . . . . **xvii**

Chapter 1 Introduction **1**

1.1 Motivation: OTN 1

1.2 The grey seal on the Scotian Shelf 5

1.3 Data Sources 8

1.4 Categorizing the Spatial Problem 14

1.5 Objectives 16

1.6 Structure and statement of co-authorship 18

Chapter 2 Statistical Approaches **23**

2.1 Traditional approaches to analyzing acoustic tag tracking data . . . 23

2.2 Spatial vs. Movement Modelling 30

Chapter 3 Modelling Random Fields with R-INLA **35**

3.1 Latent Gaussian Models 35

3.2 INLA: Integrated Nested Laplace Approximation 39

3.3 SPDE: Stochastic Partial Differential Equations 43

Chapter 4 Seal encounters at sea: A contemporary spatial approach using R-INLA **47**

4.1 Abstract 47

4.2 Introduction 47

4.3 Data 50
4.4 Model Formulation 53
4.5 Results 58
4.6 Discussion 61
4.7 Acknowledgements 64
4.8 Addendum 65

Chapter 5 Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the Northwest Atlantic Ocean 66

5.1 Abstract 66
5.2 Introduction 67
5.3 Methods 71
5.4 Results 76
5.5 Discussion 79
5.6 Conclusion 82
5.7 Addendum 83

Chapter 6 Local overfishing may be avoided by examining parameters of a spatio-temporal model 86

6.1 Abstract 86
6.2 Introduction 86
6.3 Data 89
6.4 Methods 92
6.5 Results 98
6.6 Discussion 101
6.7 Acknowledgements 112

Chapter 7 An integrated modelling approach for examining spatio-temporal behaviour of a marine generalist

predator **113**

7.1 Abstract 113

7.2 Introduction 114

7.3 Materials and Methods 117

7.4 Model Framework 122

7.A Supplementary Model Details: Diagnostics 140

7.B Supplementary Model Details: Female Seal Models 142

7.C Supplementary Model Details: Male Seal Models 151

Chapter 8 Conclusions **159**

8.1 Bayesian Hierarchical Spatio-Temporal Models for Marine Data . . 159

8.2 Future work 162

Bibliography **163**

Appendix A A Derivation of the Conditional Independence in a Gauss Markov Random Field **180**

List of Tables

Table 1.1 **Seal location and dive data**. Summary statistics for instrumented grey seals. Reported are the year of data collection (Year), number of males (M) and females (F) and corresponding total (Total). The total number of locations (n Loc) obtained from the respective tracking method (ARGOS for 1995-1999, GPS for 2009-2011 and 2013-2015) and the total number of dives (n Dive) are also shown. 12

Table 3.1 **INLA vs. MCMC** Elapsed computation time (in seconds) for equivalent models. 42

Table 4.1 **Summary.** Table of seal-to-seal encounter events observed on the Eastern Scotian Shelf, 2009-2012. 51

Table 4.2 **DIC values**. DIC values for various candidate models. 59

Table 4.3 **Fixed Effects.** Summary of Fixed Effects in *model*.2, our chosen model. 60

Table 4.4 **Hyper-parameters.** Summary of hyper-parameter values for *model*.2, our chosen model. 60

Table 5.1 **Data Summary.** Region (ordered alphabetically), total sets, sets with halibut, percent with halibut, first year, last year of time series. The numbers in brackets represent the totals after removing 61 observations in the Gulf of St. Lawrence estuary. 72

Table 5.2 **Covariates, DIC and CPO Values** Selection criteria for various candidate models with an AR(1) spatio-temporal covariance structure and a Poisson likelihood. Models are listed in order of lowest to highest DIC. Bolded values indicate the lowest DIC and highest CPO an are the better fitting models. 78

Table 5.3 **Parameter Posterior Estimates** Mean, standard deviation and 95% credible interval for the penultimate model, with AR(1) spatio-temporal structure, Poisson likelihood, a latent field and a bottom temperature covariate. 78

Table 6.1 **DIC values.** DIC values for the various candidate models with an AR(1) spatio-temporal covariance structure. 99

Table 7.1 **Seal location and dive data**. Summary statistics for instrumented grey seals. Reported are the year of data collection (Year), number of males (M) and females (F) and corresponding total (Total). The total number of locations (n Loc) obtained from the respective tracking method (ARGOS for 1995-1999, GPS for 2009-2011 and 2013-2015) and the total number of dives (n Dive) as calculated by WCDAP are also shown. 120

Table 7.2 **Prey distribution and abundance data**. Summary statistics for potential grey seal prey species. Trawl Total represents the total number of OBIS trawls in which the species was present. Total Abundance is the sum of individuals present in each of those trawls. 121

Table 7.3 **Parameters of the Integrated Model**. List of the parameters required for the integrated model. 125

Table 7.4 **Diagnostic Criteria**. Comparison of WAIC values. Other diagnostics may be found in supplements 128

Table 7.5 **Random Field Parameters**. Values characterizing the spatial effects for predator and prey. (95% Credible intervals) 128

List of Figures

Figure 1.1 **A remotely piloted glider off Nova Scotia.** Used by OTN to sample the ocean environment. 2

Figure 1.2 **Sable Island and vicinity.** A map of the Eastern Scotian Shelf study area. 8

Figure 1.3 **An acoustically tagged grey seal off Nova Scotia.** A tagged animal, whose movements and diving behaviour will be recorded and analyzed as part of the OTN research project. 10

Figure 1.4 **Seal tracks.** The GPS location data mapped to the ESS. This figure based on the 2011 GPS tracks. 10

Figure 1.5 **Physical covariates.** The Bathymetry and Bottom Substrate composition of the Eastern Scotian Shelf. 14

Figure 3.1 **A simple MRF.** A very simple Markov field on a graph of 5 nodes. 37

Figure 3.2 **Posterior Parameter Distribuions.** A comparison of the estimated distributions. The approximated posteriors from R-INLA (blue) compare closely to the iterated approximated posteriors from MCMC found using R-JAGS for each of the Intercept, Slope and Precision parameters. 42

Figure 4.1 **Seal encounters off Sable Island, Nova Scotia.** Sable Island is a large (34 km^2) sand bar, the entirety of which constitutes a Canadian National Park. Seal to seal encounters (2009-2012) in the area of Sable Island are shown in red. The black line represents a bounding box that is used later for calculation of the spatial domain. Analysis of the encounters was performed after projection onto the UTM coordinate system, the right panel shows the same encounter pattern with UTM coordinates. . . . 52

Figure 4.2 **Spatial covariates**. These plots indicate there is some relationship between the response and the covariates but it is not definitive. The R^2 value for the log response was .3636 for Depth and .4399 for Dist. 53

Figure 4.3 **Triangulation.** The triangulation utilized for the SPDE approach with the encounter events superimposed in blue. 56

Figure 4.4 **Examining Results** Plot of the posterior mean (top left), random field (top right), standard deviation (bottom left) and mean response (bottom right) for 2012. 61

Figure 5.1 **Juvenile halibut presence** Raw data from ecosystem trawl surveys indicated in blue, absence in gray and the exclusive economic zone (EEZ) in red. St. Pierre and Miquelon French territory is denoted by inset EEZ. Regions of analysis are marked, Newfoundland and Labrador (NL), Nova Scotia (the Scotian Shelf and Gulf of Maine; NS), Northern Gulf of St. Lawrence (NGSL), Southern Gulf of St. Lawrence (SGSL), and the United States of America (the survey set from NMFS includes tows into Canadian waters; USA). Areas of interest for juvenile halibut abundance are also marked: the Gully, Southwest Nova Scotia (SWNS), and are denoted generally with a box. The St. Lawrence River estuary is also identified, and while there is halibut present, ultimately it was excluded from the formal analysis. 69

Figure 5.2 **Triangulated mesh** Constrained refined Delaunay triangulation mesh. The smaller triangles of the grid are where there are more data points. It is formed of triangles, the vertices are called nodes (here there are 2120). 73

Figure 5.3 **Time series trends** Stratified mean trends in juvenile Atlantic halibut from the trawl surveys by year using all data, presence $\geq$ 0 (closed circles), and positive sets only (open triangles). A: the stratified mean number of halibut from all sets on the left y-axis, and calculated for the positive sets only on the right-hand y-axis, B: stratified mean bottom temperature (^{o}C), and, C: stratified mean depth (m). Time periods are: (1) 1978-1989, (2) 1990-2003, and (3) 2004-2013. 77

Figure 5.4 **Random fields of juvenile halibut abundance** the penultimate model by fisheries management time period, the model parameter connectivity, ρ (rho) is drawn and printed on the bottom map, units are degrees Latitude. The colour scale illustrates the random latent field (log abundance) and all three panels are the same scale. The log abundance for panel A: time period (1) 1978-1989 ranged from -0.30 to 1.07 with a mean of 0.16, for B: (2) 1990 to 2003; ranged from -0.37 to 1.27 with a mean of 0.15, and C: (3) 2004-2013; ranged from -0.25 to 1.55 with a mean of 0.21. Bathymetric contours illustrate the 100 m and 1000 m depths representing the continental shelf and larger banks. . . . 85

Figure 6.1 **NAFO Areas and Named Features** Designations of the NAFO zones and sub-zones on and around the Scotian shelf; 4X, 4W and 4V. The area within the contour lines marks the approximates the Scotian Shelf. 90

Figure 6.2 **Triangulation.** The triangulation utilized for the SPDE approach with n=548 vertices. Red symbols mark the data locations. 96

Figure 6.3 **Results** Posterior distributions for the model parameters. . . . 100

Figure 6.4 **Results** Plot of the posterior mean for the years 1986 and 1992 (pre and post collapse). The lower panels show the corresponding plots of uncertainty (the response SD). The scale is log(predicted count). 101

Figure 6.5 **Relative Exploitation and Temperature Effects.** The covariate plots show the value of the covariate on the X axis, and the impact of the covariate on the response on the Y axis and their 95% credible intervals (the dashed lines). Viewed as functions of the covariates, lower Relative exploitation and colder temperatures produce higher predicted abundance. 102

Figure 6.6 **Core Range** Plot of the posterior mean for the years 1986 and 1992, showing only those areas where the predicted mean is > 75th percentile. Viewed as 'core range'. The species range has contracted with the reduced abundance but maximum density in the aggregations has not changed, scale is log(abundance). . . . 103

Figure 6.7 **Hyperstability** The maximum *observed* count (LogY), the maximum *predicted* count (η) are nearly constant, even though the stock is collapsing. 104

Figure 6.8 **Failure to recover** Plot of the posterior mean for the years 1994 through 2000, showing only those areas where the predicted mean is greater than the 75th percentile. Viewed as 'core range'. The species range fails to recover abundance and the habitat remains unchanging. 107

Figure 6.9 **Observed Parameter Movements** What the parameters were observed to do. The coloured arrows correspond to the years discussed in Figures 10,11. 108

Figure 6.10 **The 'first collapse'** Plot of the posterior mean for the years 1973 through 1975, showing a dramatic flattening of the RF. This is indicative of circumstance in which heaving fishing was *eroding* the structure of the cod distribution leading to a partial collapse. In this period, $\rho = 3.1 \rightarrow \rho = 2.61 \rightarrow \rho = 7.50\, \sigma = 3.7 \rightarrow \sigma = 2.8 \rightarrow \sigma = 2.2$. A simultaneous large increase in ρ and decrease in σ. 109

Figure 6.11 **The 'second collapse'** Plot of the posterior mean for the years 1988 through 1990, showing a dramatic flattening of the RF. This is indicative of circumstance in which heavy fishing was *effectively destroying* the structure of the cod distribution leading to near total collapse. In this period the flattening is more severe than what was seen in the 1970s, $\rho = 4.40 \rightarrow \rho = 26.43 \rightarrow \rho = 28.6$, $\sigma = 4.2 \rightarrow \sigma = 2.1 \rightarrow \sigma = 1.6$. A simultaneous *very* large increase in ρ and decrease in σ. the collapse of the early 1990s is much more pronounced. 110

Figure 6.12 **Schematic of parameter behaviour** A schematic diagram of expected parameter values under different conditions in the fishery. Note that large ρ values are at the bottom of the schematic diagram. 111

Figure 7.1 **Named Features on the Scotian Shelf** Designations of the North American Fisheries Organization zones and sub-zones on and around the Scotian shelf; 4X, 4W and 4V. The area within the contour lines marks the approximates the Scotian Shelf. . . 117

Figure 7.2 **Random Fields: Females and Sandlance** Random field for Sandlance Abundance (left) and Bottom Time (Right). 129

Figure 7.3 **Random Fields: Males and Cod and/or Haddock** Random field for Cod and/or Haddock Abundance (Left) and Bottom Time (Right). 132

Abstract

The marine environment is a particularly challenging place for anyone interested in the animals which inhabit it. Unlike on land, where animals go and what they do is usually unobservable. Efforts to learn often rely upon a *tag*, a device attached to an individual animal that records or transmits information about the animal, where it goes, and perhaps what it does (often via some ancillary information that is also recorded). Alternatively, locations may be sampled and animals captured, and counted, at these locations, so as to learn about numbers and distribution. Such studies are usually expensive, and the number of tagged or captured animals small, such that a large gap exists between knowing where that small sample of animals went or was found, and knowing where animals of that population go, and what they do, in their habitat. One means of bridging this gap is to develop models that use the small set of discrete locations generated by the tagged or captured animals to model, or predict, at all locations in the ecosystem, the number, or the behaviour, expected at those unobserved locations.

This thesis explores the use of Bayesian hierarchical spatio-temporal random field models in the marine environment, using several different forms of available marine data. Models discussed include: methods appropriate to novel data forms being produced by deployed acoustic tags; applications in population distribution and stock structure modelling based upon research trawl data; and integrated models, which combine several different data types (physical, environmental, biological and acoustic tracking), into a single modelling framework needed to examine interdependencies between species, habitat wide.

The results of these developments clearly demonstrate that random field models are a useful and practical modelling approach. They are reasonably easy to fit, are able to capture spatio-temporal trends when present, have a parameter set that is interpretable in ways that are biologically meaningful, and, provide new means of looking at interspecies relationships. The ready interpretability of the parameters also lends them to direct practical use when applied to populations or species under commercial exploitation and/or protective management. This same easy interpretability of parameters, and the flexibility of the modelling format, allow simultaneous modelling of predator and prey to view interspecies spatial relationships otherwise hidden beneath the sea.

List of Abbreviations and Symbols Used

AIC ..... Akaike Information Criterion

AR(1) ..... Auto-regressive(first order)

CPO ..... Conditional Predictive Ordinate

CPUE ..... Catch Per Unit Effort

DCRW ..... Discrete Correlated Random Walk

DCRWS ..... Discrete Correlated Random Walk Switching

DDHS ..... Density Dependent Habitat Selection

DFO ..... Department of Fisheries and Oceans (Canada)

DIC ..... Deviance Information Criterion

E ..... East

EEZ ..... Exclusive Economic Zone

ENE ..... East North East

ESE ..... East South East

ESS ..... Eastern Scotian Shelf

GAM ..... Generalized Additive Model

GF ..... Gaussian Field

GMRF ..... Gauss Markov Random Field

GSL ..... Gulf of St. Lawrence

INLA ..... Integrated Nested Laplace Approximation

km ..... kilometer

LPML ..... Logarithm of the Pseudo-Marginal Likelihood

M ..... Mortality

MCMC ..... Markov Chain Monte Carlo

MRF ..... Markov Random Field

MVN ..... Multivariate Normal

N ..... North

NAFO ..... North American Fisheries Organization

NE ..... North East

NGSL ..... Northern Gulf of St. Lawrence

NL ..... Newfoundland and Labrador

NNE ..... North North East

NNW ..... North North West

NW ..... North West

NMFS ..... National Marine Fish Service (USA)

NOAA ... National Oceanographic and Atmospheric Administration (USA)

NS ..... Nova Scotia

NSERC .... Natural Sciences and Engineering Research Council of Canada

OBIS ..... Ocean Biogeographic Information System

OTN ..... Ocean Tracking Network

ppm ..... parts per million

QFASA ..... Quantitative Fatty Acid Signature Analysis

RF ..... Random Field

RMSEE ..... Root Mean Squared Estimation Error

RV ..... Research Vessel

S ..... South

SE ..... South East

SGSL ..... Southern Gulf of St. Lawrence

SPDE ..... Stochastic Partial Differential Equation

SSE ..... South South East

SSM ..... State Space Model

SSW ..... South South West

SW ..... South West

SWNS ..... South West Nova Scotia

TDR ..... Time Depth Recorder

TMB ..... Template Model Builder

US ..... United States

USA ..... United States of America

VHF ..... Very High Frequency

VMT ..... VEMCO Mobile Transceiver

W ..... West

WAIC ..... Watanabe-Akaike Information Criterion

WCDAP ..... Wildlife Computers Dive Analysis Package

WNW .... West North West

WSW .... West South West

Acknowledgements

The creation of any work such as the one you have in your hands is no solo effort. The are many people, both at Dalhousie University and in the wider world, to whom my thanks are due, and without whom finishing this project would have been impossible.

To my thesis supervisor, Dr. Joanna Mills Flemming, my heartiest thanks. Joanna, ever ready to raise a cheer, extend a hand, or crack a whip, (as required), you were, ultimately, the source of will and inspiration that brought me to the finish line. I could not have done it without you. Thank you for sharing the experience and making this thesis so much better.

To the other members of my committee, and the readers of this thesis, Dr. W. Don Bowen and Dr. Michael Dowd, thank you ever so much for the guidance, encouragement, and patience that such a task requires. Don, I am privileged to have been able to benefit from your ability and experience producing scientific papers; I'm sure that my writing is much, much, improved after your many, many, patient readings. Mike, thanks for reading the thing! You door was always open to me and without the benefit of your sound advice and focus on practicalities I think I would have lost all perspective, and all hope, many times over. It is your kindness, sense of humour, and interest in my ultimate success that I will remember most.

To the co-authors of the constituent papers of this paper, Dr. Shelley Lang, Dr. Nancy Shackell, Dr. Cornelia den Heyer and Dr. Stephanie Boudreau, and of course, Don and Joanna, thank you for allowing me to be part of your labs, your research, and your passions. You all tirelessly read draft after draft of the papers as they came together; your generosity with your time and insight has made them what they are. It is astonishing to me that only a few years ago I didn't know seals didn't eat plants; "They aren't manatees, Stu" - Shelley Lang, and in equal measure the amazing growth of my knowledge of the Scotian Shelf ecosystem is entirely due to the generosity of Nancy, Stephanie and Nell, as they

patiently explained to me the dynamics, the behaviours, and the politics, of the fish populations for which they speak so passionately: I'm just happy I could contribute.

To the ladies of the Department of Mathematics and Statistics, Gretchen Smith, Queena Crooker-Smith, Tanya Timmins, Ellen Lynch, Paula Flemming and Maria Fe Elder, who helped me navigate all the paperwork, administration and forms needed to register, get paid, and eventually graduate from this University that this dedicated bureaucra-phobe tried so hard to avoid, thank you. Thank you too for listening when I needed to escape my office and keyboard and talk to someone who had seen it all before. On more than one occasion it was one of you who kept me sane and in the game.

To the Ocean Tracking Network, my sponsors in this degree, thank you for the support and the funding which made it possible for me to return to school full time. No mention of the OTN is complete without acknowledging Dr. Sara Iverson, who, to me, embodies the entire grand idea that is the OTN research project. Thank you for your dedication to the OTN, and your continued interest in my tiny piece of it. Thanks also to the Patrick Lett Student Assistance Bursary, who generously gave me the wherewithal to continue when funds grew short.

To those of you who have become my friends and colleagues over the last few years, Dr. Bruce Smith, Dr. Karl Dilcher, Dr. Ed Susko, Dr. Ammar Sarhan, thank you all for answering my questions and giving me the benefit of your tremendous wealth of knowledge and experience. Your advice, encouragement and humour throughout were truly priceless.

Finally, to my girls, Pam and Evie, who have never wavered in their belief that I could, and should, do this; and upon whom, I know, a disproportionate share of sacrifice all too often fell: thanks for being there for me, I love you.

Chapter 1

Introduction

1.1 Motivation: OTN

The Ocean Tracking Network (OTN) is a global aquatic animal tracking, technology development, and partnership platform headquartered at Dalhousie University in Halifax, Nova Scotia, Canada. The OTN is a major, multi-year, effort to better understand the world's oceans and their wildlife. Despite their importance, little is known about the survival, movements, and habitat use of many aquatic species. Much more knowledge and understanding of marine ecosystems is required and OTN's mandate is to provide scientific foundations for sustainable oceans management. Since its inception, OTN has been creating a global research and infrastructure platform with the goal of integrating biological, oceanographic and social sciences, technological innovation and the fostering of the partnerships necessary to international science. Recently, (2016), OTN was recognized as a Major Science Initiative by the Canadian Foundation for Innovation, which will carry the OTN project through 2022, continuing to generate knowledge and provide training for hundreds of students and researchers.

The OTN poses the following questions: How do oceanographic and environmental features (both physical and biological) affect animal habitat use, and movement? How do aquatic species interactions and areas of ecological significance relate to habitat use, movement patterns, and oceanographic features? How do anthropogenic activities and development influence aquatic animal behaviour and ecology?

To achieve these ends, OTN has identified some large scale goals for research. The OTN exploits technology in the marine environment, taking advantage of animal tracking technology and tagging techniques to generate tracking data on subject species. But simply generating data is not enough, OTN desires to place these tracking data into wider context by also prioritizing the assimilation of these data into other, larger scale models and prioritizing visualization and statistical modelling techniques for complex aquatic and marine observations, with an eventual goal of influencing policy and developing the mechanisms for feeding results into oceans management.

The OTN deploys acoustic receivers and oceanographic monitoring equipment to better understand marine life and evaluate impacts, positive and negative, of anthropogenic actions such as fishing, fisheries management, and the consequences for other marine life. Data generators for the OTN include fixed lines; lines of acoustic detectors laid across the ocean floor which detect the presence of certain acoustically tagged animals when they pass within detection range of fixed receivers, glider-based observations of physical, biological and chemical properties, a glider being a remotely piloted vehicle which cruises through the ocean and records these various physical, biological and chemical measurements, (temperature, salinity), and telemetry based studies, which use a tag attached to an individual animal of a studied species which records its movements, locations and perhaps other ancillary information like diving depth, temperature etc.

Figure 1.1: **A remotely piloted glider off Nova Scotia.** Used by OTN to sample the ocean environment.

The data recorded by a tagged animal, usually in the form of longitude and latitude locations, along with whatever other information being recorded, are referred to as telemetry data. Typically the data are not continuous, they constitute a series of point observations for the tagged animal, a series of snapshots taken of the animal over time. Each snap shot records a location, and the ancillaries, for example, a tag may record the location via GPS every 20 minutes, and also archive the depth, the light level, the temperature, or any other desired measurement at those locations or over the interval between locations. Thus a point by point series of snap shots. While telemetry data is valuable in itself, OTN recognizes that in the marine environment there are unique difficulties since the animals themselves cannot often be observed. One may want to know where all the animals go, but have information only on where that one animal went, or one may want

to know what factors influence the animal's chosen path, knowing only the locations recorded.

Thus, the OTN has identified the need for statistical modelling techniques to extend the limited view obtained by the tracked animal to the wider population and environment, to project and predict what the untagged animals are doing, filling in the blank areas. These models should be integrated, meaning that they can assimilate varied data from varied sources, so that the factors (in addition to those recorded by the tag) which may have influenced the movement may be incorporated and examined, if additional data are available. Among the many projects overseen by the OTN pursuing these objectives, is Project IV: "*Integrating Research and Themes across Arenas and Implications for Oceans Governance (Tracking and Protecting Marine Species at Risk: Scientific Advances, Sea of Governance Challenges)*", (**www.oceantrackingnework.org**) (OTN, 2013). The umbrella of Project IV contains several sub-projects which lead directly to the research themes contained in this thesis work. The first, §4.1. "*Ocean Modelling and Assimilation*", emphasizes the critical role of modelling in understanding the movement and distribution of marine animals, and, given the difficulty of actually observing the marine environment, underscores the need for such models to "fill in the gaps" between sparse ocean observations, so to extrapolate to locations and times for which observations are not available. Models can transform the OTN's telemetry based point observations into *products* that can be used for practical applications such as management and the setting of marine policy. This OTN subproject proposed to develop new methodologies of general interest to the research community. Four species specific applications; the Atlantic sturgeon, Atlantic salmon, American eel and, of particular import here, grey seals, were to be examined to determine how these species respond to changes in their physical and biological environment.

The second pertinent subproject, §4.7. "*grey seals (Halichoreus grypus) as Bioprobes: Predicting Impacts on their Ecosystems*", involves the deployment of Vemco Mobile Transceivers (VMTs) and satellite transmitters on grey seals at Sable Island. These instruments not only record location and behavioural data for the tagged animals, they are also capable of detecting other tagged animals at sea. These encounter detections, tags recording the proximity of another tag, provide evidence of co-location and perhaps association among the grey seals and their prey. Data from these instruments were intended to be used to investigate seal-seal associations during foraging trips - and the spatial and temporal patterns of encounters with acoustically tagged Atlantic cod, salmon and

bluefin tuna [1] - and to develop technology to use VMTs in other species. Additionally, objectives included identification of foraging hotspots in the ocean and attempting to better understand grey seal habitat use, movements and migrations in relation to ocean conditions and species interactions.

A third subproject, §4.8 "*Visualization and Modelling of Complex Marine Observations*" is recognition that OTN is currently making use of new and innovative technologies to better understand continental shelf and open ocean ecosystems. These new technologies deliver new types of data, and require the development of new statistical methods to both visualize and make quantitative inferences. In order to answer scientific questions of relevance to OTN, we must have both effective visualization tools and accurate models. Modelling is crucial for designing effective tracking studies, necessary for inference about species interactions, and critical for informing conservation and management.

Understanding how marine animals use the ocean is crucial for the development of conservation and management strategies both for individual species and ocean ecosystems. In some cases, the oceanographic influence may directly affect the animal, whereas in other cases the oceanographic features may serve to concentrate prey, making foraging more profitable for the predator. Pinnipeds like grey seals are large energy requiring homeo-therms and as such their movements and habitat use ought to be strongly influenced by prey availability. OTN seeks to investigate the way grey seals respond to prey availability at scales that are relevant to the response of prey to the physical and biological environment.

The overarching goal of this thesis is to explore methods for analyzing the novel data forms being delivered to researchers by these VMT acoustic tags, (the encounters), and to find means of modelling the tracking data from individual tagged seals to examine how grey seals use their oceanic habitat structure when foraging and locating prey assemblages. That is, at the scale not of the individual animal and it's immediate surroundings, but at the scale of the population, and how the population explores and exploits the available habitat.

[1] Analysis of encounters between seals and cod, salmon, and tuna, never became practicable due to the small numbers of encounters observed. Seal to seal encounters were analyzed, see Chapter 3.

1.2 The grey seal on the Scotian Shelf

Since there is focus upon grey seals, and their prey, throughout this thesis it is necessary to provide some background on the species. Grey seals, *(Halichoerus grypus)*, are a species of large marine pinniped. They are native to the North Atlantic, both on the European and North American coasts. North American grey seals range from the shores of New Jersey north to Newfoundland. They are seen on the coasts of Quebec and throughout the Gulf of St. Lawrence. The largest single breeding colony in the world is located on Sable Island, off the coast of Nova Scotia. It is this major colony that is of most interest here, as the tracking study examined deploys tracking tags on seals from the Sable Island colony. Sable Island lies on the continental shelf off eastern Canada, adjacent to the province of Nova Scotia. The area is commonly referred to as the Eastern Scotian Shelf (ESS), and Fisheries and Oceans Canada (DFO) has the lead federal role in managing Canada's fisheries and safeguarding these waters. DFO is the regulatory authority for managing commercial activities on the ESS and also conducts many research and scientific projects concerning the wildlife of the ESS.

Presently, the grey seal population on Sable Island is at an all time high (Bowen, 2011, 2014, 2017). The reasons for increases in population levels are not well understood. Various theories have been advanced, including, decreased human hunting of seals, decreased predation upon the seals, reduced competition, and increased availability of food (Bowen, 2011, 2014, 2017). Overall, it is felt that the degree of predation on seals (mostly from sharks) is low (Baum et al., 2003), and that there is little support for a release from predation explanation for the population increase.

An alternative argument is that the increase in the grey seal population is related to an increase in the resources (prey) available to them. The collapse and failure to recover of the Atlantic cod *(Gadus morhua)* population off Canada's east coast is widely known, and has occurred over (broadly speaking) the same time period, and in the same place as, the dramatic increase in the grey seal population. As posited by DFO (2011) the seal's preferred diet may be not cod but other, higher fat fishes such as herring, mackerel and sand lance. The argument runs that collapse of the cod (which preys in part upon these smaller fish) constitutes but part of a Scotian Shelf wide ecosystem shift and that these species, preferred by the seals, have become much more abundant while filling the gap left vacant by the absence of cod and other large bodied fishes(DFO, 2011). Increased abundance of their preferred prey, in the absence of any other limitation, and with reduced competition, has driven the increase in seal numbers.

On reviewing the literature on grey seals it seems that there exists a consensus that the seals are generalist predators, able to eat many different prey. It is possible to list six or eight species which are consistently found in the diet and that constitute most of the energy intake, and the preponderance of the diet (Tucker et al., 2008). It is also agreed that the relative proportion of the diet constituted by these species must vary widely over both space and time, but how it varies is poorly understood (Tucker et al., 2008). To paint it broadly and bluntly, where the particular seals go is known from the tracking studies, but not what drives their movement - what is it about prey distribution(s) that results in variations in grey seal spatial use.

Since the seals are at sea to primarily to eat [2] , the diet, what they are eating, drives their movement and behaviour, but these diets are difficult to determine. Analyses of grey seal diets are important to us because they influence seal movements across the habitat. What grey seals eat has been derived mainly by two methods, quantitative fatty acid signature analysis (QFASA) and scat analysis. QFASA is a method wherein a tissue sample is obtained from a subject, here a grey seal, and the sample is analyzed for fatty acid composition, the presence and quantity of which are indicative of certain prey species being present and proportional in the subject animal's diet (Iverson et al., 2004; Tucker et al., 2008). Scat analysis involves analyzing fecal samples for structures that are resistant to digestion and that are diagnostic of individual prey species. Identifying bones (especially otoliths), beaks and other hard parts of consumed prey allows for the identification of species and size of prey consumed (Bowen et al., 2011; Bowen and Harrison, 1994). Conclusions reached in these studies (Arim and Naya, 2003; Bowen et al., 2011; Bowen and Harrison, 1994; Bowen and Lawson, 1993; DFO, 2011; Trzcinski et al., 2009; Tucker et al., 2008) have high degrees of uncertainty but do agree that Northern sandlance (family *Ammodytidae*) are a significant component of the diet, perhaps as much as 77% (Bowen et al., 2011), and also support the contention that the seals are generalists, with wide variety of prey species in their diet. So while there is some place to start when seeking the important prey species that influence seal movement and behaviour, we will need to cast a wide net when identifying prey species for consideration.

Each of these methods for inferring diet, QFASA and scat, have some significant limitations (Bowen and Iverson, 2013). Grey seals are a wide ranging species with a seasonal influence on where in their range they are to be found. Diets vary depending on *where*

[2]Seals may rest, or even sleep, on the bottom. Without a means to distinguish exactly what the seal is doing on the bottom, this must remain a caveat. Bottom time is still a proxy for foraging, even if imperfect.

they are in their range. The problem is obvious when looking at the dietary analysis techniques. The QFASA approach measures the fatty acids that build up in the animal's fat tissues over time as it eats certain things. What it is eating determines which fatty acids build up, so the fatty acid profile reflects a sort of long term average of what the seal has been eating, the bandwidth of this averaging depends on exactly which tissue is sampled. While it is possible to estimate the proportion of the diet made up of certain prey it is difficult to tell when or where it was consumed (Tucker et al., 2008). The scat analyses suffer from an almost opposite problem. If food takes, say, 24-72 hours to pass through a seal's gastro-intestinal tract, then scat samples will only indicate what has been consumed within 24-72 hours. Fecal samples can only be practically recovered on land, meaning that they will only reveal what has been consumed within 24-72 of haul-out, which is significant in a species that may be away from land for much longer than that. So a scat analysis is a single point in time "snap-shot" of diet (near haul-out) while QFASA provides a long term average, smoothing over both unknown locations and time. Neither method gives insight into the spatial characteristics of the seals diet, that is, the spatial structure of seal predation across the Shelf.

The spatial characteristics of the diet matter. There is strong evidence that spatial habitat use by grey seals is sex specific - that is, that the males and females have divergent ranges (Tucker et al., 2008). The evidence suggests that the greatest divergence in range male/female is during the period of heavy feeding post breeding. It has been proposed that the larger males can exploit larger prey and range farther, while the females seek out higher *quality* prey, that is prey of higher energy density, specifically the sand lance, these being found on the "banks" (Breed et al., 2006) that I have previously found to be areas favoured by the grey seal - without segregating by sex (Carson and Flemming, 2014). It is also suggested that female grey seals are able to *out-compete* the males for this higher quality prey, *relegating* the males to less profitable areas and lower quality prey; males' larger size (and proportionally larger gastro-intestinal tract) enables them to extract sufficient caloric intake from a higher volume of lower quality prey (Breed et al., 2006). Characterizing this spatial structure of seal predation across the Shelf will allow us to see the overall effect of seal predation, the 'ecological footprint' of this predator species.

1.3 Data Sources

Our study area is the Eastern Scotian Shelf (ESS). This region of the continental shelf, off eastern Canada (Figure 1.2), covers some 108,000 km^2 encompassing Sable Island ($43^o55'N$, $60^o00'W$) as well as several offshore shallow banks and basins separated by deep gullies and canyons . This continental shelf is the main foraging area for grey seals breeding on Sable Island (Breed et al., 2006, 2009). This island, home to a very large grey seal colony, is a crescent-shaped sandbar with a surface area of about 34 km^2, and is centrally located in the ESS. There exists a collection of diverse data sources relevant to this study area. Each is discussed in detail below.

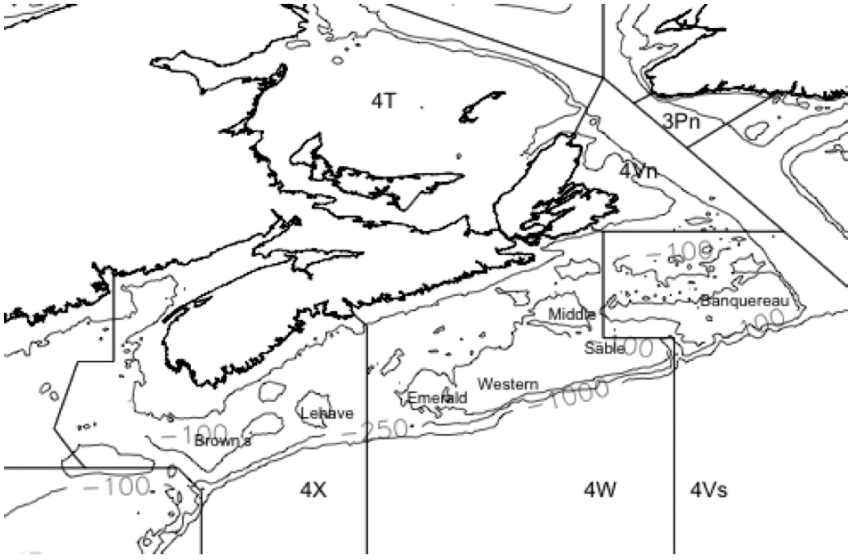

Figure 1.2: **Sable Island and vicinity.** A map of the Eastern Scotian Shelf study area.

Seal Location Data

As briefly mentioned previously, as part of the OTN's ongoing research program grey seals were fitted with electronic tags on Sable Island. Available are data from 1995 to 1999 and again from 2009 to 2015, collected to study their ranging and movement. The 2009 to 2015 electronic tags used GPS to record locations, resulting in fine resolution data in both space and time. The 1995 to 1999 electronic tags used Service Argos, an older system, which yields data which are less precise. Although the accuracy of locations provided by these tags differ, and the GPS records locations more frequently than Argos,

both see use here to provide the longest time series of observations possible. In both time periods, adult seals were captured using a hand-held net and anaesthetized to permit the animals to be measured (body length and mass) and fitted with electronic tags (Beck et al., 2003b; Breed et al., 2006; Lidgard et al., 2012). Seals were released at the site of capture (on Sable Island) after recovery from the anesthesia (30-45 min) and observed until they entered the sea. All field procedures were conducted in accordance with the requirements of the Canadian Council on Animal Care and were approved by Dalhousie University's Committee on Laboratory Animals and by Fisheries and Oceans Canada's animal care committee.

GPS Tracking Data: 2009-2015

In the years 2009 to 2015, seals were fitted with Wildlife Computers Mk10-AF Fast-loc GPS tags, (**www.wildlifecomputers.com**) and Advanced Telemetry Systems VHF transmitters (164-165 MHz, **www.atstrack.com**). The VHF transmitter was attached to the MK10-AF unit using a stainless-steel hose clamp and the whole unit was attached to the fur on the top of the head using a five-min epoxy (Breed et al., 2006). A seal so outfitted is pictured in Figure 1.3. The MK10-AF tag was programmed to archive GPS data that were downloaded on recovery of the tag following the animals return to Sable Island. Tags were programmed to record a GPS location every 15 minutes and depth every 10 sec when the seal was at sea and to suspend location attempts when the unit was dry for more than 20 min and a location had been obtained (Lidgard et al., 2012). The tag mass burden was 0.25% for adult males and 0.28% for females. The VHF tag was used to locate instrumented animals on land once they returned to Sable Island.

To obtain GPS locations, archival GPS data from each MK10-AF were analyzed using propriety software from the manufacturer (**www.wildlifecomputers.com**). This resulted in GPS locations and associated dive data for 62 females and 18 males over the 2009-2015 period. Figure 1.4 gives an idea of what the resulting location data set looks like, mapped to the ESS.

After initial data exploration it was noted that there were numerous anomalies in the 2012 data. After contacting the researcher who initially collected these data, the 2012 data were excluded from the analysis as incompatible. The parameters which control the tags recording features: when to record, when to shut off, how frequently to log data all had been given different and incomparable setting for the year 2012. This affected only the female seals tracked in 2012. This results in our having distinct blocks of data for the female seals; 2009, 2010, 2011 and 2013, 2014, 2015. The 21 male seals were not so

subdivided, results and models presented have the male seals in a single group 2009-2015.

Figure 1.3: **An acoustically tagged grey seal off Nova Scotia.** A tagged animal, whose movements and diving behaviour will be recorded and analyzed as part of the OTN research project.

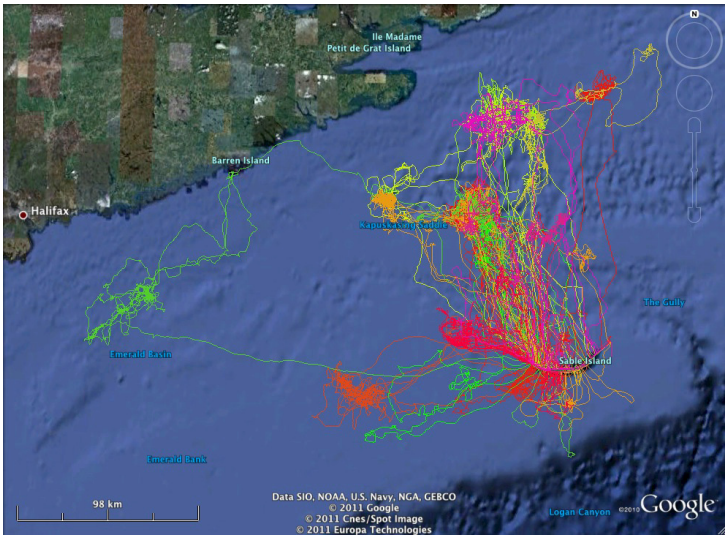

Figure 1.4: **Seal tracks.** The GPS location data mapped to the ESS. This figure based on the 2011 GPS tracks.

ARGOS Tracking Data: 1995-2001

In addition to these 2009-2015 GPS data there also exists several years of older seal tracking data. ARGOS satellite tracking data are similar to GPS tracking data in that location data can later be aligned to dive data and other environmental information collected by the seal. ARGOS is however, less precise. The resulting data are of a coarser nature than the GPS equivalents; instead of locations every 20 minutes or so ARGOS

locations were obtained every 6-8 hours such that there are far fewer observations and a concomitant reduction in resolution.

During the period of Argos location tracking, depth when the seal was at sea was also recorded. The earlier equipment had more rigorous memory constraints and recorded depth every 20s versus the every 10s of the later instruments. Between 1995 and 1999, 20 seals (6 males, 14 females) were instrumented with either a Satellite Data Recorder (SDR; Wildlife Computers, Redman, WA, USA) or an ST-18 (Telonics, Mesa, AZ, USA) satellite transmitter. To record diving behaviour, each animal was also instrumented with a Time-Depth Recorder (TDR; Mk3e, Mk5, Mk6 or Mk7, Wildlife Computers, Redmond, WA, USA) which was secured to the seal's lower back using 5 min epoxy. TDRs were programmed to record depth every 20s. A conductivity sensor was used to determine when the animal was at sea or hauled out on land.

Dive Data

Proprietary software, Wildlife Computers Dive Analysis Package (WCDAP; Wildlife Computers, Redman, WA, USA), extracts depth data from the tags and uses sea surface values to separate depths into individual dives. This software also generates metrics to describe each dive, including dive depth, total dive duration, and time spent at the bottom (of the dive). The operational definition of bottom time was used: time spent within 85% of the maximum depth for each dive and totalled it for all dives occurring between sequential GPS (or ARGOS) locations. While the sampling rate of the tags differs in the two periods, (every 10s vs. every 20s), the same software was used to discretize the depth reading into separate dives. The dive data from the TDRs are not dramatically different from the later dive records. The number of dives per seal are similar, as is Bottom time per dive. A handful of dives with a total duration of more than 30 minutes were excluded as they represented the concatenation of several dives. There were 127 records so excluded, resulting in a total of 194,631 dives observed in the 1995-1999 period and 1,991,211 observed dives in the 2009-2015 period.

Time spent at the bottom has previously been linked to seal foraging (Beck et al., 2003c) as grey seals are thought to forage at or near the seafloor. That is, a seal will travel to and from foraging areas spending relatively little time on the bottom, but when it arrives at its foraging area it spends relatively more time on the bottom where it is more likely to encounter prey (Beck et al., 2003c). We summed the bottom dives of the seal's dives, using our operational definition of bottom time, for all dives occurring between

sequential GPS (or ARGOS) locations. This leads to a single number summary which is thought to be a useful proxy for time during which seals are most likely to encounter prey (Beck et al., 2003c), that is, time spent foraging. The proxy is not perfect, we cannot see what the seal is doing at the bottom, foraging, resting, or sleeping; more sensors and better data from future study may resolve this question.

Year	M	F	Total	n Loc	n Dive
1995	1	3	4	985	55894
1996	1	3	4	1627	46632
1997	1	2	3	576	13929
1998	1	3	4	2732	30595
1999	2	3	5	1624	47581
Subtotal	6	14	20	7544	194631
2009	5	8	13	87111	228243
2010	5	14	19	115253	300111
2011	0	13	13	118736	318460
Subtotal	l0	35	45	321100	846814
2013	2	6	8	72769	211236
2014	2	4	6	72417	160752
2015	1	9	10	113268	277551
Subtotal	5	19	24	258454	649539
Total	21	68	89	587098	1690984

Table 1.1: **Seal location and dive data**. Summary statistics for instrumented grey seals. Reported are the year of data collection (Year), number of males (M) and females (F) and corresponding total (Total). The total number of locations (n Loc) obtained from the respective tracking method (ARGOS for 1995-1999, GPS for 2009-2011 and 2013-2015) and the total number of dives (n Dive) are also shown.

Encounter Data

Innovative tagging technologies are granting researchers access to previously unavailable descriptors of animal movement. One example, the VEMCO Mobile Transceiver (VMT) (**www.vemco.com**), acoustically records instances of proximity of instrumented animals to other, similarly tagged animals. Supplementary to the location and depth recording tags referred to above, OTN researchers have deployed VMTs on Sable Island grey seals (*Halichoerus Grypus*). In 2009, 14 grey seals on Sable Island were randomly selected and fitted with VMTs. The VMTs are anticipated to record an acoustical contact whenever one comes within approximately 500 meters (manufacturer's stated data) of another VMT instrumented seal. Thus they deliver a data set comprising the times and locations

of the pair-wise proximities of these grey seals, hereafter referred to as *encounters*. The experiment continued through 2010, 2011 and 2012. In 2010 the sample size was increased to 20 seals. The 2011 and 2012 experiments retained the sample size of 20 seals and resulted in encounters occurring in roughly, the same area of the Scotian shelf.

Data of the type delivered by the VMT, that is, times and locations of encounters between tagged marine animals, had simply not been available because its collection was infeasible prior to the advent of the VMT (O'Driscoll et al., 2000). We note that these data are at sea seal to seal encounters as distinct from *tracks* of seals (which are commonly associated with studies that use acoustic telemetry) (Jonsen et al., 2005; Patterson et al., 2008).

Prey Distribution and Abundance Data

The Department of Fisheries and Oceans Canada (DFO hereafter) Maritimes Region has conducted an annual summer ground fish research trawl study each year since 1970. Originally designed to measure distribution and abundance of commercial species these data also incorporate information on non-commercial species. Focussed upon the Scotian Shelf the DFO survey utilizes a stratified sampling plan using the three relevant North American Fisheries Organization (NAFO) zones, 4V, 4W and 4X demarcating the Scotian Shelf. Figure 1.2 presents the general geographical location and shows the boundaries of these NAFO areas and their associated sub-divisions, referred to as sub-zones. Each of the three NAFO zones comprises a stratum, with sampling effort (the number of sample trawls) proportional to it's area. The catch is sorted by species, weighed and measured for individual weight, maturity status and age. The data have been summarized in various reports (Horsman and Shackell, 2009; Ricard and Shackell, 2013; Smith et al., 2013), stored, and are publicly available in the Ocean Biogeographic Information System (OBIS) (OBIS, 2014).

OBIS is the DFO - Maritimes Region database for ground fish research trawl surveys and includes information on some 263 distinct species found on the Scotian shelf, including species we identify as likely to be prey for grey seals [3]. The database includes descriptive data for each cruise or mission resulting from about 200 trawling sets per year. The trawls are conducted primarily in July (there is some variation over the 40+ years of data, trawls may be delayed due to weather or mechanical issues, so there are sometimes observations

[3]Atlantic butterfish, capelin, Atlantic herring, Atlantic mackerel, Northern sandlance, Atlantic cod, haddock, pollock, white hake, American plaice, winter flounder, witch flounder, yellowtail flounder, Atlantic halibut, smooth skate, thorny skate, winter skate, redfish sp., longhorn sculpin, sea raven and Northern shortfin squid

present through August and September). For each set there exists trawl information: date, latitude, longitude, distance towed(km), as well as physical/water characteristics at the location and depth of the trawl; temperature(C), salinity(ppm), nitrate(ppm), phosphate(ppm) and silicate(ppm). For all species captured: genus, species, common name, total weight(kg) and count, (and total weight and count standardized by distance towed) are recorded.

Physical and Environmental Data

To integrate the physical environment, or seascape, encountered by the seals into the analysis we identified some physical covariates which may be useful for determining how seals use space. These are the bathymetry of the Scotian Shelf, bottom substrate composition (particulate size). The bathymetry data are sourced from the US National Oceanographic and Atmospheric Administration (NOAA), ETOPO1 global relief resource. Resolution is 1 arc minute in X and Y (approximately 2 km), and $\sim$ 10 meters in Z (Amante and Eakins, 2009). Bottom composition data (sediment size) was sourced from DFO; spatial resolution is 5 km x 5 km (Dr. Ken Frank, BIO). Figure 1.5 shows maps of the Bathymetry and the bottom substrate size for the ESS, the area of study.

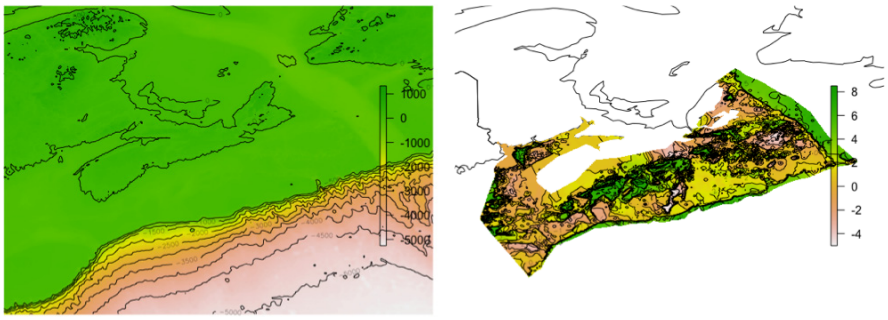

Figure 1.5: **Physical covariates.** The Bathymetry and Bottom Substrate composition of the Eastern Scotian Shelf.

1.4 Categorizing the Spatial Problem

When first confronted with the encounter data it was clear that we had spatio-temporally explicit data, that is, we had information that included location and time of the encounters, as well as information about the encounters length, duration etc. Naturally enough

we sought to classify the data type and found Cressie (Cressie, Rev. ed. 1993) classifies Spatial Data as being of 3 main types: geo-statistical, lattice and point pattern.

- Geo-statistical data: The distinguishing feature of this data type, also referred to as "spatial data with continuous variation", is that the domain (D) is fixed and the response at location x, $R(x)$, is continuous. The domain is fixed by the geographical boundary, and the response is continuous and could be measured anywhere. *Where* it is measured does not change this property of the response.

- Lattice or regional data: Typically, data reported by political boundary, postal code or census boundary is aggregated within that regional boundary and a summary measure is reported. Thus (D) is the assembly of all the reporting regions and $R(x_i)$, the response at the ith site, is an average for the ith area as opposed to a measure at x_i.

- Point pattern data: When the data does not share the fixed and non random domain (D) of geo-statistical or regional data then each realization of the process will have a set of locations, (D^*), at which the process is observed. So the number of points in the pattern, and their locations, are the outcome of a random process. The collection of points is a point pattern. If the data set also includes attributes for each point then the pattern is termed a *marked* pattern. Without the attribute(s) the set is termed *unmarked*.

Classical Methods

We first attempted to look at these data using the tools of so-called classical spatial statistics for pattern analysis. Several comprehensive works were consulted. Cressie (Rev. ed. 1993) provides an overview of the subject and in no particular order the books by Schabenberger and Gotway (2005), Daley and Vere-Jones (2003) and Møller and Waagepetersen (2004) were consulted. The field of *stochastic geometry*, as surveyed by Stoyan, Kendall and Mecke (Stoyan et al., 1983), shares theoretical aspects with the study of point processes and proved useful. The theory of *Random fields* are to be found in Vanmarcke (1988).

These fundamental writings on the topic of point patterns and point pattern data analysis led to consideration of the ideas of Cox (1955) and Matérn (1960) as likely candidates for our particular data. Examples of modern applications of these ideas include the identification and tracking of spatial and temporal lightning clusters (Woolford and Braun,

2007), using spatial point processes to model forest fire outbreaks (Møller, 2010).

Pattern data in the marine ecology context has been largely unavailable, especially at the scale we are considering, mainly due to the difficulty in obtaining point data sets of locations for mobile and wide ranging animals. This difficulty has largely limited the application of point pattern data analysis in ecology to the study of plants or sedentary invertebrates at small spatial scales (O'Driscoll et al., 2000). Indeed O'Driscoll et al. (2000) discusses analogous methods for use in the absence of the pattern information. Exceptions exist; an interesting application of spatial point pattern techniques to large animals in the marine environment is the modelling of sea turtle bycatch (Gardner et al., 2008) where locations were available and were examined in relation to the spatial distribution of the fishery activities as a whole. One attempt to link the spatial pattern of a seagoing animal to the oceanic environment itself is found in Royer et al. (2004). Here Bluefin Tuna, *Thunnus thynnus*, schools were spotted from an aircraft in the Gulf of Lions. The techniques of pattern analysis were then used to relate the locations to the hydrography of the gulf, surface temperature fronts and chlorophyll content fronts generated by the outflow of the Rhône. We found these pattern based methods to be generally inadequate for describing the encounter data.

Regional analysis would arise if we were to attempt to base our analysis or prediction on regions or zones. For example, in Chapter 3, you will find some discussion of the NAFO management zones and sub-zones. If one were to conduct an analysis discretized by zone or sub-zone, and make inference by zone or sub-zone, then one would be considering a lattice or regional model. Regionalizing the data would essentially be a form of binning the data. The discrete locations (and the associated spatial information), would be lost.

1.5 Objectives

Given this large and diverse collection of data about the marine wildlife and environment of the Scotian Shelf, what is it that we want to do? Initially, we have the problem of new data types; the encounter data. Our motivation in modelling these encounters is twofold: (1) We wish to determine what can be learned directly from encounters in space through time and, (2) Statistical methods should indeed be available with which to interpret them. Dealing with a new data stream, for which there is no body of analysis for like data establishing tried methods, requires us to look at the characteristics of the data and answer, *What kind of problem does encounter data pose*? Having characterized the data by type we need an analytical tool(s), bringing us to our next hurdle, *What sort of*

statistical tool is appropriate for our encounter data?

One answer to the above questions was that we had spatial point pattern data; random locations creating a pattern of points, and we would need spatial pattern methods to treat it. We set out to incorporate covariates, in the form of oceanography features such as temperature and bathymetry and to relate them to the encounter data, in a spatial model. The point pattern model describes the spatial distribution of the points and attempts to estimate where the locations for the next pattern would be. The ancillary variable was modelled as a *mark*. This marked point pattern approach as not a success. In a second attempt, we chose to geo-statistically model the mean of the underlying process that generated the marks. The geo-statistical spatial model for the encounter data worked well; and like most scientific success, it led to more questions.

At the conclusion of this first analysis we speculated on the direction of future work and noted it would be very useful to supplement the analysis by using prey information directly in the model, rather than relying on purely physical covariates. We noted that one could estimate the spatial distributions of candidate prey species using these spatio-temporal methods and then incorporate these prey species distribution estimates directly into models for the predators behaviour. We further noted the immediate fisheries and management interest for such an analysis for Sable Island grey seals and Atlantic cod (*Gadus morhua*).

It was in this fashion that I began to consider the next set of questions; *Would such spatial models be able to model a species' distribution of behaviour*, and, if so, *What is gained modelling a species' behaviour in a spatial model* over the more traditional modelling approaches? *Could such a model reveal new, or alternative, insight*? I took our spatial modelling format and applied it to the available abundance data for Atlantic cod (*Gadus morhua*), and Atlantic halibut, (*Hippoglossus hippoglossus*), in the form of the survey trawl data. What was revealed was not only an informative models for cod and halibut abundance, but also significant insight into the spatial structure inherent to fish stocks on the ESS. I was able to see that not only was there significant structure present, dependent on the ocean environment, but that the structures changed in time in response to outside influences, (Fishing).

Again, I found success lead to a new round of questions. There was clearly structure present in the fish stocks, and this structure changed in response to outside influence. *Could I capture interplay between species by modelling structure explicitly in a spatial model*? That is, *Could the spatial structure of separate species be linked in a single*

model?

The desire to capture interplay between species, as it applied directly to the data available, lead to questions of data integration. How do I best approach a data set (or sets) that includes the locations of these tagged predator animals from geo-locator tags, information about their dive behaviour from a depth recorder, spatio-temporal information about potential prey from a survey trawl, time varying environmental covariates such as temperature, *and* fixed covariates like bathymetry and bottom composition? These data were obtained in different places (spatial misalignment) and different times (temporal misalignment) and on different scales. *Were the spatial models able to provide a means of integrating these data into a common framework*?

This thesis concludes by showing that a model for the spatial interplay of predator and prey, integrating many diverse data forms is possible, informative, and makes intuitive biological sense. Such a model allows for spatial and temporal misalignment in predator and prey observations while jointly estimating uncertainty in each. There is a step by step sequence through the thesis in the form of individual chapters corresponding to separate articles which delineate each of the steps towards this objective. First, by demonstrating the method to be practical in that it delivers sensible results and is not inordinately difficult to use. Then, by examining carefully the results from spatial models of abundance data, showing that these models are able to capture the structure present in marine animal stocks, that these insights into structure are meaningful and that there is advantage in these models over other methods. Finally, moving to a truly integrated modelling format; integrated both in the statistical sense of bringing together diverse data sources and in the biological sense of providing a single model that incorporates both predator and prey, what contributes to their distributions, and how they interrelate, in a single model.

1.6 Structure and statement of co-authorship

This thesis is divided into a number of chapters. In addition to an abstract, introductory and concluding chapters, Chapters 4, 5, 6 and 7 correspond to individual manuscripts written for submission to various scientific journals. They each pursue aspects of our overall objectives: statistical methods for novel features of these data, constructing suitable single species distribution models, these later to be incorporated into the integrated model exploring the grey seal's habitat use and behaviour on the Scotian Shelf. Each of these chapters follows the normal format and structure of a scientific paper: introduction,

data, methods, results, discussion. Each of the co-authors of these papers contributed valuable commentary and critical review of these papers, and assisted with revisions and advice. Chapter 6 is based upon a manuscript of which I was a co-author, but not the primary author. While the submission of the manuscript was led by Dr. Stephanie Boudreau, I was very involved in the methodology and analysis and drafted the pertinent sections of the manuscript; what is presented herein is a synopsis of the eventual manuscript which conveys the key results, but which is not the manuscript as submitted. Abstracts for each of these papers are included here in Chapter 1 to assist the reader.

Seal encounters at sea: A contemporary spatial approach using R-INLA.

*Acoustic telemetry is an active field of research integral to the study of marine life. The latest generation of acoustic tags is making available new types of data. As part of the Ocean Tracking Network (OTN) (***www.oceantrackingnework.org***) acoustic tags known as VEMCO Mobile Transceivers (VMTs) (***www.vemco.com***) are being deployed on Sable Island grey seals (*Halichoerus grypus*) in order to record instances of proximity to each other (as well as any other acoustically tagged animals). The seals essentially become bioprobes yielding data exhibiting both spatial and temporal variation. Fortunately, recent developments in the field of spatial statistics have greatly facilitated the fitting of complex spatial and spatio-temporal models. Here we specifically propose a hierarchical spatio-temporal model framework and fit it to these data using both Stochastic Partial Differential Equations (SPDE) and Integrated Nested Laplace Approximations (INLA) through R-INLA (***www.r-inla.org***). In so doing we demonstrate the effectiveness and advantages of these techniques. These methods readily extend to spatially explicit data collected by any sort of mobile receiving platform (e.g. wave gliders, remotely operated underwater vehicles).*

Local overfishing may be avoided by examining parameters of a spatio-temporal model.

Spatial erosion of stock structure through local overfishing can lead to stock collapse because fish often prefer certain locations, and fisheries tend to focus on those locations. Fishery managers are challenged to maintain the integrity of the entire stock and require scientific approaches that provide them with sound advice. Here we propose a Bayesian hierarchical spatio-temporal modelling framework for fish abundance data to estimate key parameters that define spatial stock structure: persistence (similarity of spatial structure over time), connectivity (coherence of temporal pattern over space), and spatial variance (variation across the seascape). The consideration of these spatial parameters in the stock assessment process can help identify the erosion of structure and assist in preventing local overfishing. We use Atlantic cod (Gadus morhua) in eastern Canada as a case study an examine the behaviour of these parameters from the height of the fishery through its collapse. We identify clear signals in parameter behaviour under circumstances of destructive

stock erosion as well as for recovery of spatial structure even when combined with a non-recovery in abundance. Further, our model reveals the spatial pattern of areas of high and low density persists over the 41 years of available data and identifies the remnant patches. Models of this sort are crucial to recovery plans if we are to identify and protect remaining sources of recolonization for Atlantic cod. Our method is immediately applicable to other exploited species.

Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the Northwest Atlantic Ocean.

*The Northwest Atlantic Ocean has long been a productive area for commercial fishing, and long overtaxing of the resources eventually led to a well documented, fisheries driven, ecosystem shift in the early 1990s. The ecosystem shift has seen the Atlantic cod (*Gadus morhua*), once very abundant, be reduced to historically low levels, and fail to recover after the implementation of measures to protect it. This failure of an overexploited species to recover has been accompanied by an increased abundance in other fishes. One species which has been increasing in abundance, at least since the 2000s, is the Atlantic halibut (*Hippoglossus hippoglossus*). This species is of commercial interest, but knowledge of halibut ecology is limited and the failure to recover of other groundfish species stocks collapsed by local overexploitation serves warning as to the dangers. In this Chapter we examine the application of a Bayesian hierarchical spatio-temporal model to juvenile halibut over a time span of 36 years; during which time there were 3 distinct fisheries management regimes. We characterize the structure of juvenile halibut using three parameters: persistence (similarity of spatial structure in time), connectivity (coherence of structure over space) and spatial variance (variation across the seascape). We find that areas of high juvenile abundance persist throughout the 36 years, but find that some of these areas are much reduced while others retain their high levels; in spite of increased abundance and landings throughout the study area. These persistent areas of high abundance overlap substantially with full and seasonal area fishery closures, which may create refuges from fishing. We estimate the connectivity to be 250km, far less than the distance assumed by Canadian fishery management units (2000km). This smaller scale of spatial coherence suggests a more complex structure than previously assumed, with attendant consequences for management of the fishery.*

An integrated modelling approach for examining spatio-temporal behaviour of a marine generalist predator.

Understanding the ecological footprint of a predator begins with an understanding of how that species or population uses space through time. Space use by predators is driven by many factors including the availability of prey and the nature of the physical landscape. Understanding spatial behaviour is particularly challenging in marine environments where animals and their interactions with other species are rarely observable. Recently the wide-spread use of location telemetry

*and data loggers for the study of marine animal movement has made large datasets (big data) available that can be used to model how marine animals use space and the factors underlying these patterns. Here we develop an integrated model framework to describe diving behaviour of grey seals (*Halichoerus grypus*) on the Scotian Shelf of eastern Canada. Geo-referenced metrics of diving behaviour (time spent at the bottom of dives, a proxy for foraging effort) for 89 adult seals are integrated with environmental factors (e.g. bottom temperature, bathymetry, and bottom sediment) and prey information derived from stratified-random, bottom trawl ecological surveys. We model the predator and prey simultaneously using a Bayesian hierarchical model incorporating latent spatial random effects represented using random fields. The resulting models reveal that the amount of bottom time exhibited by seals is relatively stable in both time and space, but is spatially differentiated by sex. Similarly, overall patterns seem not to be tied to the presence of particular prey species for the grey seal population as a whole, but males and females do appear to respond differently to certain species and prey assemblages. The pattern of habitat use for the seals showed little local spatial pattern, unlike that of the prey themselves, which tend to be more highly structured with respect to their physical environment.*

The journal and publication status of these manuscripts at the time of submission of this thesis is as follows:

S. Carson and J. Mills Flemming. Seal encounters at sea: A contemporary spatial approach using R-INLA. *Ecological Modelling*, Vol. 291, pp 175-181, 2014
https://doi.org/10.1016/j.ecolmodel.2014.07.022

S. Carson, N. Shackell and J. Mills Flemming. Local overfishing may be avoided by examining parameters of a spatio-temporal model. *PLoS ONE*, Vol. 12(9): e0184427, 2017
https://doi.org/10.1371/journal.pone.0184427

S.Boudreau, N. Shackell, **S. Carson**, C. den Heyer. Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the Northwest Atlantic Ocean *Ecology and Evolution*, Volume 7, Issue 22, pp 9193–9844, 2017
https://doi.org/10.1002/ece3.3495

S. Carson, W. D. Bowen and S. L. C. Lang and J. Mills Flemming An integrated modelling approach for examining spatio-temporal behaviour of a marine generalist predator. *Ecosphere*: Submitted

Chapter 2

Statistical Approaches

Here we provide an overview of traditional approaches to analyzing acoustic tracking data. These traditional approaches include techniques such as the state-space models (SSMs) and their outgrowth, switching models. Here distinctions are drawn between these established traditional modelling treatments and the spatio-temporal modelling treatments of behaviour proposed in this thesis. I then review ecological applications of geo-statistical modelling as well as the literature on Bayesian hierarchical spatial and spatio-temporal modelling approaches for geo-statistical data, with particular emphasis on those pertaining to animal modelling.

2.1 Traditional approaches to analyzing acoustic tag tracking data

Traditionally, acoustic tags have delivered tracking data, which detail the movement (and perhaps associated behaviour) of an *individual animal* over time. Such data are usually analyzed at the individual track (seal) level. In these studies of movement, the tracking data are not approached in the habitat context, (i.e. not spatially) but in the movement context. Covariates are usually recorded at the animals observed location, often by the same tag. These data are then treated in a manner akin to a time series (Breed et al., 2009; Jonsen et al., 2005, 2003). In many such studies the quantification of measurement error is of importance. The animal's observed locations may be subject to substantial measurement error (i.e. error in location), depending on the exact method of measurement. Older methods of geo-location, such as Argos satellite transmitters (SDR-Wildlife Computers, Redman, WA, USA, ST-18, Telonics, Mesa AZ, USA or SRDL 7000, Sea Mammal Research Unit, St Andrews, UK), were uncertain, and before any inference for the animal's behaviour could be made it was necessary to estimate the animals *true* locations given the Argos observations. The observed locations were filtered, per these previous references, or in (Austin et al., 2003), to estimate actual positions

and then the behaviour of the animal could be examined. Methods such as state-space modelling (Jonsen et al., 2003), including extensions to switching versions (Jonsen et al., 2005), which categorize behaviour along the trajectory based on the characteristics of the inferred true movement; have as their goal a description of how the animal moves; accomplished by modelling a most probable trajectory given uncertain observations and then examining how that trajectory may respond to outside factors (environmental covariates) and/or be a function of the actions of the animal (inferred behaviour modes).

Technologies advance, and recent generations of acoustic tags have increased their data collection and storage capacities, such that more variables, each at higher resolution, can now be recorded than could be recorded by earlier tags. Geo-location via Global Positioning Satellite (GPS) has greatly increased the precision with which animal locations may be known. However, when modelling or examining movement, via the animal's trajectory through space, covariates are not often considered for locations other than those occupied by the tagged animal, that is, not observed where the track does not go. It is this spatial difference, the distinction between inferring (only) at the animal's observed or inferred locations via a movement model, and examining the whole of the habitat, and the footprint of the whole of the species on the whole of the habitat, that marks a key distinction between these traditional approaches and the spatial models that are presented in this thesis. Before I move on to detail these spatial models however, let us examine the traditional modelling frameworks in some detail.

State-Space Models

One approach to modelling the movement of the tagged animal(s) from acoustic tag data is the state-space model (SSM) as found in (Jonsen et al., 2003). In the SSM the animal's observed location is a function of the true location and a location error. This function is expressed in a measurement equation. The location error may be significant for Argos location data, but is often assumed negligible for GPS location data. The SSM seeks to estimate the true (unobserved) locations for the animal and to model the movements of the animal between true locations from the

observed locations using a transition equation. The transition equation parameterizes the movement between subsequent true (unobserved) locations subject to some uncertainty (process noise). The measurement and transition equations are given below. The equations that follow in this subsection are in the notation and style of (Jonsen et al., 2003).

1. The measurement equation:

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \boldsymbol{\epsilon}_t), \tag{2.1}$$

   - $t = 1, 2, \ldots, T$,
   - $\mathbf{y}_t$: Vector of ***observed*** location of animal at time t,
   - $\mathbf{x}_t$: Vector of ***true*** **(unobserved)** location of animal at time t,
   - $\boldsymbol{\epsilon}_t$: Vector of location **errors** for animal at time t, i.e. $\epsilon_t \sim N(0, \sigma_\epsilon)$.

2. The transition equation:

$$\mathbf{x_t} = f_t(\mathbf{x}_{t-1}, \boldsymbol{\eta}_t; \boldsymbol{\gamma}), \tag{2.2}$$

   - $\boldsymbol{\eta}_t$: Process error at time t, mutually independent of $\boldsymbol{\epsilon}_t$, i.e. $\eta_t \sim N(0, \sigma_\eta)$
   - $\boldsymbol{\gamma}$: Parameter vector describing the movement process.

So in this SSM format the true position of the animal at time t is a function of the animal's true position at time $t-1$, process noise η, (noise in the movement itself), and some movement parameters γ (which describe the properties of the movement: direction, distance, speed, etc.) and the effects of any explanatory variables.

Each of these equations, the measurement and the transition, include uncertainty and so imply a density. From the measurement equation we get $p_y(y_t|x_t)$, and from the transition equation we get $p_x(x_t|x_{t-1}; \gamma)$, these being the probability of the observed location given the true one, and the probability of the true location, given the previous true location and the properties of movement. By combining these densities and integrating over the (unobserved) true locations preceding time

t, we get the prediction equation; a density for predicting the next true location given the previous observed (known) locations and the properties of movement. This is called the prediction equation.

3. The prediction equation:

$$p(\boldsymbol{x}_t|\mathbf{Y}_{t-1};\boldsymbol{\gamma}) = \int p_{\boldsymbol{x}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\boldsymbol{\gamma})p(\boldsymbol{x}_{t-1}|\mathbf{Y}_{t-1};\boldsymbol{\gamma})d\boldsymbol{x}_{t-1}, \tag{2.3}$$

- $\mathbf{Y}_{t-1}$: Vector of all observed locations prior to time t.

The prediction equation predicts the true location at time t using all information previously observed. State space modelling consists of 2 steps; 1, take all information from previously observed data ($\mathbf{Y}_{t-1}$), and, using the prediction equation, predict the new location. Then, 2, use an update equation (which is Bayes) to update the prediction, using new information provided by the latest observed location, y_t, along with the previous information given by $\mathbf{Y}_{t-1}$.

4. The update equation is:

$$p(\boldsymbol{x}_t|\mathbf{Y}_t;\boldsymbol{\gamma}) = \frac{p_y(\mathbf{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\mathbf{Y}_{t-1};\boldsymbol{\gamma})}{\int p_y(\mathbf{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\mathbf{Y}_{t-1};\boldsymbol{\gamma})d\boldsymbol{x}_t}. \tag{2.4}$$

This updating process is done for times $t = 1, 2, \ldots, T$ from some initial known position (e.g. the point of release of tagged animal). So $p(\boldsymbol{x}_1|\mathbf{Y}_0;\boldsymbol{\gamma}) = p(\boldsymbol{x}_1|\boldsymbol{x}_0)$. The likelihood for $\boldsymbol{\gamma}$, which is termed the innovation, is the probability of observing each of the $\mathbf{y}_t$s, given the previously observed positions and explanatory variables.

5. The innovation is:

$$\begin{aligned} L(\boldsymbol{\gamma};\mathbf{Y}_T) = p(\mathbf{Y}_T;\boldsymbol{\gamma}) &= \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{Y}_{t-1};\boldsymbol{\gamma}) \\ &= \prod_{t=1}^{T}\left[\int p_y(\mathbf{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\mathbf{Y}_{t-1};\boldsymbol{\gamma})d\boldsymbol{x}_t\right]. \end{aligned} \tag{2.5}$$

These SSMs may be fitted either using a Bayesian approach or a frequentist approach. The Bayesian is used in (Jonsen et al., 2003), using MCMC. Frequentist approaches use linearization and Kalman-filter and/or numerical integration.

The (Jonsen et al., 2003) paper goes on to provide a simple example where the transition and measurement equations (Equations 2.1 and 2.2) are:

$$\boldsymbol{x}_t = f_t(\boldsymbol{x}_{t-1}, \boldsymbol{\eta}_t; \boldsymbol{\gamma})$$

$$=> \boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \boldsymbol{\eta}_t$$

$$\mathbf{y}_t = \mathbf{h}_t(\boldsymbol{x}_t, \boldsymbol{\epsilon}_t) = \boldsymbol{x}_t + \boldsymbol{\epsilon}_t$$

- $t = 1, 2, \ldots, T$
- Measurement error is included as $\boldsymbol{\epsilon}_t \sim N(0, \tau^2)$,
- An explanatory covariate, s_t, is introduced,
- Movement north-south and east-west are independent draws from the same normal distribution with variance $\boldsymbol{\sigma}_\eta = \sigma exp(-\beta_{s_t})$. Thus the covariate enters into the model via the process noise. In the specifics of the example in (Jonsen et al., 2003), β determines how quickly the distance moved at time t declines with increasing temperature.
- The model is fitted by estimating the parameter vector $\Theta = (\beta, \sigma, \tau)$.

These SSMs above all model the movement location by location, that is, they attempt to the *true* locations (the αs, given the observed locations (y). Other models would focus more upon the movements of the animal, as opposed to the locations. Movements at each time step are examined by differencing the locations; movement at time t is the difference between the present location x_t and the previous, x_{t-1}.

1. One example of such a model is the Differences Correlated Random Walk (DCRW).

$$\mathbf{d}_t \sim \gamma \mathbf{T} \mathbf{d}_{t-1} + \mathbf{N}_2(0, \boldsymbol{\Sigma}). \tag{2.6}$$

   - γ: is an autocorrelation. If $\gamma = 0$ it is a simple random walk, $0 \leq \gamma \leq 1$ produces a correlated random walk, γ determines the degree of autocorrelation.

- $\mathbf{d}_t$: is the difference between locations $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$
- $\mathbf{d}_{t-1}$: is the difference between locations $\mathbf{x}_{t-1}$ and $\mathbf{x}_{t-2}$
- $\mathbf{T}$: is a transition matrix, describes the rotation in the DCRW

$$\mathbf{T}(\theta) = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

- θ: is the mean turning angle,
- $\mathbf{N}_2$: is a bivariate Normal with mean 0 and covariance matrix Σ, where:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2_{lon} & \rho\sigma_{lon}\sigma_{lat} \\ \rho\sigma_{lon}\sigma_{lat} & \sigma^2_{lat} \end{bmatrix}$$

- σ^2_{lon}: is the process variance in longitude,
- σ^2_{lat}: is the process variance in latitude,
- ρ: is the correlation coefficient.
- The model is fitted by estimating the parameter vector $\Theta = (\theta, \gamma, \rho, \sigma_{lon}, \sigma_{lat})$.
- In (Jonsen et al., 2005) it is assumed $\rho = 0$ (uncorrelated).

State-Space models have been extended from attempting to model the trajectory of the animal as it interacts with the environment. Efforts to elucidate not just how the animal's movement reacts to the environment, but to try to infer what it is doing along the trajectory have been made. So called switching models model the trajectory, but assign the animal one of 2 or more behaviour categories; e.g. feeding or transiting. Switching models in the state-space framework, such as found in (Jonsen et al., 2005), follow a similar logic to state-space approaches. I will examine the switching model in a similar way to what I have just done with (Jonsen et al., 2003) above.

2. The DCRWS: Differences Correlated Random Walk: Switching

1. The transition equation is:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \boldsymbol{\eta}_t. \tag{2.7}$$

- $\mathbf{x}_t$: 2D vector of unobserved states at time t, ***true*** **(unobserved)** locations (lat, long), regularly spaced in time.
- $\boldsymbol{\eta}_t$: The process variability

The transition equation describes a Markov process where unobserved states evolve over regular time intervals, given the previous state, process variability, and biological parameters. Describes the dynamics of the movement process being modelled. The DCRWS essentially allows the parameters to be different for each behaviour; this is equivalent to having a transition equation for each behaviour. Probabilities are defined, α_1, the probability of being in behaviour 1 at time t, given the animal was in behaviour 1 at time $t-1$, and, α_2, the probability of being in behaviour 1 at time t, given the animal was in behaviour 2 at time $t-1$.

- The model is fit by estimating the parameter $\Theta = (\theta_1, \theta_2, \gamma_1, \gamma_2, \alpha_1, \alpha_2, \sigma_{lon}, \sigma_{lat})$.
- Again it is assumed $\rho = 0$ (uncorrelated).
- Thus a model for each behaviour and some transition probabilities.

Such a switching SSM was applied to grey seals in (Breed et al., 2009). In this instance the SSM with a correlated random walk able to switch between 2 behaviour modes, foraging and travelling. In this paper, (Breed et al., 2009), having categorized the likely behaviour of the seal at each regularized time step as either foraging or travelling, extended the inference to environmental factors. Noting that the locations categorized as foraging tended to be somewhat clustered, bounding boxes were drawn about these clusters of inferred foraging locations. The physical and oceanographic characteristics present at the clusters was examined and similarities noted.

Other models for tracking or telemetry data

Direct observation of the seal's behaviour while at sea is usually impossible and therefore behaviour must be inferred from what is observed. Other formats of observed data are also seen in animal telemetry study. For example, along with locations the tag may record ancillary variables such as ambient light or diving

depth. Attempts to discern the animals at sea behaviour or link it to potential prey, either via tracking data or ancillary telemetry data, are limited in the literature. Beck et al. (2003b,c), in two related papers at different temporal scales, have examined the characteristics of the seal's dive behaviour - as evidenced by such metrics as bottom time, dive depth, dive shape or dive duration - as an indicator of behaviour, attempting to classify the observed dives by shape and thereby infer behaviour by associating "types" of dive to travel or to forage behaviour on the seal's part. Analysis of the impact of the grey seal as a predator in relation to a prey specie(s) in the context of attempting to explain the high observed mortality (**M**) in the Atlantic Cod of the Gulf of St. Lawrence (Benoit et al., 2011; Mohn and Bowen, 1996; Trzcinski et al., 2006, 2009), or in the context of linking seal foraging to prey density (biomass) (Harvey et al., 2012) have been attempted. These studies use the idea of overlap, either in areal data as in Benoit et al. (2011) or by inferring behaviour from the tracking data (i.e. using the tracking data to infer foraging areas of seals) as in Harvey et al. (2012) and comparing these inferred areas of forage to available prey data for those areas. In Sterling et al. (2014) a state space model with two states, resident and transit, is used to attempt to understand the circumstances and conditions the animal is likely to be in, in each state.

2.2 Spatial vs. Movement Modelling

In the SSMs previously described the emphasis is on modelling the most probable actual trajectory of the animal through the environment. This is done by modelling how the animal moves, often as a function of environmental covariates. For example, does the animal move faster of slower if it is cold? The behaviour of the animal, what it is doing at any given location and time, is extrapolated from the nature of the movement; rapid and straight line movement versus slow and circuitous movement may be interpreted as transiting between locations and foraging at a feeding area respectively. This means the animal's behaviour at time t is inferred from how it moves at time t, not from any direct measurement or observation of behaviour itself. In this thesis I ultimately seek to examine the

behaviour of a predatory species with respect to its use of habitat via a spatio-temporal model of behaviour, and, simultaneously attempt to link that observed pattern of spatial behaviour to possible explanatory variables. That is, I wish to consider not where a single animal goes in it's potential range, or how it moves in response to environmental factors it sees there, but, by incorporating other information available, to directly model the behaviour and then infer *why* animals of that species go where they do and *what* they will do there. I propose a spatio-temporal modelling framework that integrates data sources potentially observed across the entire range, to use the tracking data from many animals to infer spatial use by the species (as represented by the sample tracked) so as to infer effect of the species on the habitat as a whole; the *ecological footprint* of the species, if you will.

In particular, I consider the structure and distribution of behaviour of grey seals, how this structure may depend upon the locations and distributions of prey, and, to explore the relationship of these two together with the physical and/or oceanographic environment in which they exist. The goal is to derive *population*, spatial distributions of behaviour and interdependencies for the grey seals from tags attached to *individual* seals. In other words what is required is statistical methodology to incorporate individual seal tracking data into population level geo-statistical distribution or behaviour models so that I can estimate the impact of grey seals, and the nature of their interdependencies with other (prey) species present, at all locations in their Scotian Shelf habitat, which is required in order to understand these impacts and interdependencies at locations where they are not observed.

To do this I will use the location data provided by the tag, corrected for measurement error if necessary, and integrate that with the ancillary data. In fact, what I do is to choose an ancillary data source which is a proxy for the behaviour, and then, from the observed locations and observed ancillary variable, estimate the (expected) value of the ancillary variable fo all points in the habitat (observed and unobserved) in order to have an indicator of the ecological footprint of the population. That is, I would seek some metric of the seals impact in the ecosystem, and then *map* the expected value of that metric of behaviour to the environment, producing an estimate of what any seal may be expected to do at any point in the

usable spatial domain. It is this idea of estimating the mean at all points in the habitat which brings us into the geo-statistical modelling arena. Our animals are a sample, which represent all animals in our population. The behaviours exhibited by our sampled animals are to be extrapolated to the population. So rather than simply model the sampled animals as a realization of a pattern at specific locations, I looked instead to use them to model the behaviour of all animals at all locations, a spatially continuous response which could be, at least theoretically, measured anywhere in the domain.

The encounter data, the survey trawl data for prey species, and the tracking data for the predators from Argos/GPS are all data sets exhibiting both spatial and temporal variation. Treating them geo-statistically means instead taking the pattern of points, each with the associated ancillary variable value and attempting to model the mean of the underlying process which generated those values at those points. That is, not guessing where the next sample of locations will be located, but, I wish to infer what the characteristics of the ancillary would be, at any point. At any arbitrary point on the ESS, what would they do? We will utilize the spatio-temporal encounters to fixed physical oceanography and a latent field as a motivating example to demonstrate the effectiveness and advantages of these techniques. What follows this analysis is a new statistical approach to Bayesian hierarchical modelling for abundance data, based on research trawls, for two of our prey species. The approach allows the rapid analysis of these data, and in a way which yields insight and answers to practical questions. Ultimately, I will use the seals tracking data not to model the individual trajectories, i.e. the next series of locations for that seal, but to instead model the expected characteristics of what is underlying, If a seal were observed at any arbitrary point on the ESS, what would I expect it would be doing? Before I proceed to these analyses though, I set out what exists in the literature for applications of these methods, both to animal data and otherwise.

Recall the classification of spatial data into geo-statistical, regional and pattern categories. This division into data types centres on how the space is regularized for the analysis. In the geo-statistical case there is a single spatial domain D and the response varies continuously over it. Imagine a map of Canada (D) and air

temperature across the nation. The temperature could be measured anywhere and exists continuously across D. Even though the temperature is only measured at relatively few locations, it is useful to know, or at least predict, what it might be in other places. A model which predicts temperature anywhere from a limited number of observation points would be a geo-statistical model for temperature. Naturally enough, this type of geo-statistical model has found favour in atmospheric applications like our example, see Cameletti et al. (2013) for a thorough explanation and example of the technique applied to geo-statistical data on airborne particulates in the Po river valley. See also (Musenge et al., 2012) for a disease mapping application and (Moller et al., 2016) for an example predicting weather patterns.

More recently, the methods have been finding wider application. Interest in complex problems continues to grow and the approach is being extended and new models, likelihoods and extensions have appeared (Bivand et al., 2015). Several spatial studies have appeared in widely separated fields. In epidemiology, Alegana et al. (2016b) use the fine spatial resolution maps produced to target resources to specific local populations, and to specific months of the season, whilst combatting malaria. In a second paper on similar issues the spatial effects of covariates in the study area (Alegana et al., 2016a) are examined over time. Another geo-statistical paper with focus upon a time-varying covariate in a veterinary study of bovine Coxiellosis can be seen in (Schrodle and Held, 2011). Common issues such as zero-inflation may be provided for within the geo-statistical framework; see (Arab, 2015) for statistical solutions such as zero inflated and hurdle models applied to disease mapping and epidemiology and (Hu et al., 2016) for a zero-inflated, negative binomial model describing spatio-temporal schistosomiasis risk in China's Yangtze Basin.

Animal based publications include efforts to estimate total population size for terrestrial animals in finland from encounters occurring during transects of a habitat Jousimo and Ovaskainen (2016), and from random observations of annually breeding bird species (Meehan et al., 2017). Examples of particular interest to this study are a study of Greenland Shark by-catch (Cosandey-Godin et al., 2015) in the Canadian Arctic, where a spatio-temporal geo-statistical model was used

for expected bycatch from commercial fishers. Finally, an exploration of predator–prey interactions of mobile marine species, including among others, grey seals, using the traditional idea of habitat overlap to view the relationships, can be seen in (Sadykova et al., 2017). A study with similar objectives, utilizing a spatial GAM is found in (Swain and Wade, 1993a). These overlap based studies look for areal correlation seeking to use predator and prey values on a grid cell by grid cell basis or on an area by area basis (a BYM model is used in Sadykova et al. (2017)).

Chapter 3

Modelling Random Fields with R-INLA

3.1 Latent Gaussian Models

INLA, for Integrated Nested Laplace Approximation is the computational approach to Bayesian modelling proposed by Rue et al. (2009a) as an alternative means to Markov Chain Monte Carlo (MCMC). The method is intended to compute posterior densities for latent gaussian models, a large sub-class of Bayesian hierarchical models. Hierarchical models are used to explore complex dependence structures in data. Uncertainty in parameters, latent variables, and/or processes can be modelled with appropriate prior distributions using a Bayesian framework. R-INLA (**?**) is the R (R Core Team, 2013) package which implements these methods. The implementation works for models of the following form;

$$(\boldsymbol{\theta}) \sim p(\boldsymbol{\theta})$$

$$(\boldsymbol{x}|\boldsymbol{\theta}) \sim N(\mathbf{0}, Q(\boldsymbol{\theta})^{-1})$$

$$\eta_i = \sum_j c_{ij} x_j$$

$$(y_i|\boldsymbol{x}, \boldsymbol{\theta}) \sim p(y_i|\eta_i, \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ are the parameters, $\boldsymbol{x}$ is a latent Gaussian field (a joint distribution of all the parameters in the linear predictor), $\boldsymbol{\eta}$ is a linear predictor based on known covariate values, c_{ij} (and which may be related to them by a link function), and $\boldsymbol{y}$ is the data vector (Lindgren and Rue, 2015). In a latent Gaussian model the response, each element of $\boldsymbol{y}$, the data vector, y_i, is assumed to belong to an exponential family with the mean, μ_i linked to an additive predictor, η_i through a link function, $g(.)$, such that $g(\mu_i) = \eta_i$. Hence we have priors, $p(\boldsymbol{\theta})$, a latent field, $(\boldsymbol{x}|\boldsymbol{\theta}) \sim N(\mathbf{0}, Q(\boldsymbol{\theta})^{-1})$, a linear predictor, $\eta_i = \sum_j c_{ij} x_j$, and a likelihood, $(y_i|\boldsymbol{x}, \boldsymbol{\theta}) \sim p(y_i|\eta_i, \boldsymbol{\theta})$.
The model, still in general form, can be written:

$$g(\mu_i) = \eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(\mu_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i. \tag{3.1}$$

Here the $\left\{f^{(j)}\right\}$ are unknown functions of the covariates $\mathbf{u}$, the α is an intercept, $\{\beta_k\}$ are the linear effects of covariates $\mathbf{z}$ (regression coefficients), and the ϵ_i are unstructured. That is to say that the model consists of the sum of various components, random effects and linear effects of covariates. (Bannerjee et al., 2008; Cressie and Johannesson, 2008; Kammann and Wand, 2003). Many popular models can be expressed in this form and modelled using INLA depending on the choices of the functions $\left\{f^{(j)}\right\}$. These functions may be non-linear, such as i.i.d. random effects, spatially and/or temporally correlated effects or smoothing splines. Examples are the Besag-York-Mollié model for disease mapping in regions (Besag et al., 1991), continuous Gaussian models, GLMMs and GAMs (Bannerjee et al., 2004; Diggle and Ribiero, 2006). Spatio-temporal dependance can be introduced via a covariate (s, t) like a kernel smoothed spatial covariate on a grid (Illian et al., 2012b) or a spatio-temporal Gaussian field as follows here. These models are all latent Gaussian models so long as x has a joint normal distribution.

Gauss Markov Random Fields

I intend to use a *random field* to model responses across the entire domain. A random field, to use an analogy taken from Vanmarcke (Vanmarcke, 1988), is like a large laboratory in which there are a large number of experiments. Each experiment has an outcome, X. Observing the outcome of all of the experiments is observing a *realization* of the random field. There are many types of random fields. A key distinction necessary for the direction I intend to go is between continuously indexed random fields and discretely indexed random fields. A continuously indexed random field has a random value at every location in the domain, a two-dimension smooth surface, for example. A discretely indexed field has a discrete domain and has a value at each coordinate in the domain. So if one imagines a grid, and the field is a random value on each cell or at each node then one has a discretely indexed random field.

Markov random fields (MRFs) are a particular type of discretely indexed random

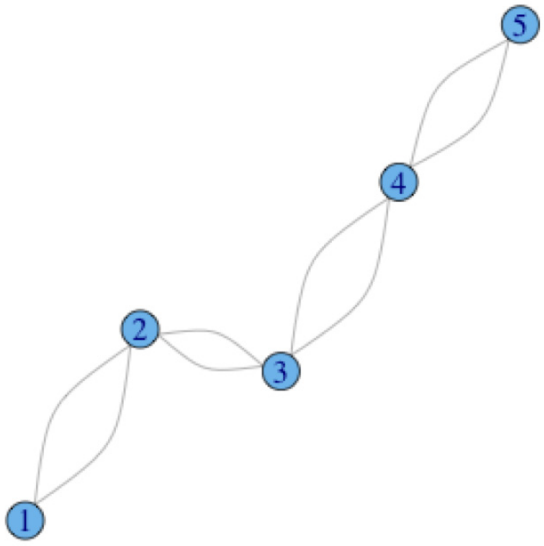

Figure 3.1: **A simple MRF.** A very simple Markov field on a graph of 5 nodes.

field wherein the random variables have the *Markov property* and are defined on, or described by, a graph (Illian et al., 2012b; Rue and Held, 2005). In one dimension a stochastic process (in time) has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present values) depends only upon the present state, not on the sequence of events that preceeded it. An MRF extends this property to two or more dimensions for random variables defined for an interconnected network of nodes as on a graph (Rue and Held, 2005). This is the concept of the *Markov Blanket*, which in the case of the MRF simply means a set of *neighbours*, the neighbours of node l defined as the set of nodes sharing an edge with node l.

Let us consider the small example in Figure 3.1. The graph has 5 nodes and 4 edges. Let $x_l, l \in (1, ..., 5)$ be the value of an MRF at locations l, so $\mathbf{x} = (x_1, ..., x_5)'$ is an MRF. What is important here and later is that an MRF has the distribution of $x_l, l \in (1, ..., 5)$ depend only on a portion of $\mathbf{x}$, (the neighbours). Calling the set of neighbours to component l, δ_l, we get that the conditional distribution of

$x_l, (l = 1, \ldots, 5)$ is,

$$p(x_l|x_{-l}) = p(x_l|x_{\delta_l}), \tag{3.2}$$

which is the same as stating that given δ_l, x_l and $x_{-\{l,\delta_l\}}$ are independent, where $x_{-\{l,\delta_l\}}$ includes neither x_l nor the neighbours of l, δ_l. So, in this instance, the value of the field at node 3, x_3, is conditionally independent of $x_{-\{3,\delta_3\}} = \{x_1, x_5\}$ given $\delta_3 = \{x_2, x_4\}$.

Gauss markov random fields (GMRFs) arise when a multivariate normal distribution forms a Markov random field with respect to a graph, that is if the values $\mathbf{x} = (x_1, ..., x_5)'$ have a multivariate normal distribution $\mathbf{x} \sim MVN(\mu, \Sigma)$ and $\Sigma^{-1}_{ij} = 0$ if node i and node j do not share an edge. Calling this matrix (the *precision* matrix) $\Sigma^{-1} = Q$, then for x_i and $x_j, i \neq j$, $i, j \in (1, ..., n)$,

$$x_i \perp x_j|x_{-\{i,\delta_i\}} \Leftrightarrow Q_{ij} = 0 \tag{3.3}$$

which means that the non-zero elements of Q are determined by the neighbours. So $Q_{ij} \neq 0$, if and only if $j \in \{i, \delta_i\}$. Looking again at our example of Figure 3.1, the matrix Σ may be dense (have many or all element non-zero) but the matrix Q will be sparse, in fact tri-diagonal. If we have this conditional independence then the latent field x is a GMRF.

All this is important because inverting a matrix is computationally demanding when n gets large, $\mathcal{O}(n^3)$. Sparse matrices, with many zeroes, are easier to invert. A covariance matrix (Σ) with lots of zeroes is easy to invert, but to get elements of this matrix to be zero requires many of the elements in x to be *marginally independent*, since $x_i \perp x_j \Leftrightarrow \Sigma_{ij} = 0$ which is a strong assumption. The conditional independence, via the Markov property, is much weaker. Having the precision matrix Q contain many zeroes aids computation greatly, (it is a large matrix) (Rue et al., 2009b).

Laplace Approximation

Taylor series expansion is a method of expanding a function around a point a into a series of sums.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... \tag{3.4}$$

The series of sums may be infinite, but the function may be approximated using a finite number of terms of the expansion. A Taylor series can be used to estimate a normal probability distribution in what is termed a Laplace approximation. Laplace approximation uses the first three terms (i.e a quadratic function) Taylor series expansion around the mode x^*, of a function ($h(x)$ to approximate $log[h(x)]$ (the log simplifies differentiation):

$$log[h(x)] \approx log[h(x^*)] + \frac{\delta log[h(x^*)]}{\delta x}(x-x^*) + \frac{\delta^2 log[h(x^*])}{2\delta x^2}(x-x^*)^2 \tag{3.5}$$

The term $\delta log[h(x^*)]$ is zero at x^*, and the expression simplifies. Using a Laplace approximation the distribution $h(x)$ is approximately normal with mean x^* and variance $\frac{-1}{h"[x^*]}$.

3.2 INLA: Integrated Nested Laplace Approximation

We want the marginal distributions of the parameters of x, that is, the regression coefficients, intercepts, and parameters of the the functions $f^{(j)}$. What Rue et al. (2009a) propose is to approximate the posterior density for $(\boldsymbol{\theta}|\boldsymbol{y})$ using a Gaussian approximation $\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ for the posterior of the latent field, calculated using the Taylor series expansion, evaluated at the posterior mode, $x^*(\theta) = argmax_x p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$,

$$p(x_j|\boldsymbol{y}) \approx \int \tilde{p}(x_j|\boldsymbol{\theta},\boldsymbol{y})\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}, \tag{3.6}$$

$$p(\theta_k|\boldsymbol{y}) \approx \int \tilde{p}(\theta|\boldsymbol{y})d\boldsymbol{\theta}_{-k}, \tag{3.7}$$

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{p(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{x=x^*} \approx \left.\frac{p(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})}{\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{x=x^*}. \tag{3.8}$$

This is called the Laplace approximation. The terms we need to approximate are: $p(\theta|y)$ and $p(x_j|\theta,y)$. The first term can be used to estimate all the marginals of

interest for θ and is also needed to estimate the marginals for the latent field. The marginal posteriors for each of the θ_k and x_j are then found by numeric integration over $\boldsymbol{\theta}$, with another Laplace approximation.

INLA can either use the normal approximation $\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ already calculated above. Very fast, but very strong assumption (often skewed or heavy tailed). Alternatively, partition x and get $p(x_{-j}|x_j, \theta, y)$, use Laplace on each element. These conditionals behave better (more normal), but are more computationally expensive. This last is INLA default. The INLA algorithm uses Newton-like methods to explore the joint posterior distribution for the hyper-parameters $\tilde{p}(\theta|y)$ to find suitable points for the numerical integration.

The reason that INLA invites comparison to MCMC methods is that they are designed to do similar things, even though they take very different approaches. Where INLA uses a deterministic *approximation* to the unknown distribution, MCMC attempts to *find* an approximation. It does this using an algorithm (Bolstad, 2010; Metropolis et al., 1953).

Let $f(x)$ be a function that is proportional to the desired probability distribution $\pi(x)$. We can do this because, as above, we can find the joint distribution $p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$ that constitutes the numerator in

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}, \tag{3.9}$$

but the denominator, which INLA approximates deterministically, is not. To follow the MCMC algorithm, first, an arbitrary point x_0 is chosen to be the first sample, and an arbitrary probability density $P(x|y)$ which is used as the density for the next sample value x, given the previous sample value y. Often this density $P(x|y)$ is chosen to be a Gaussian distribution centred at y. This distribution $P(x|y)$ is referred to as the proposal density. Next, generate a candidate x' for the next sample by picking from the distribution $P(x'|x_t)$. Third, calculate the acceptance ratio $\alpha = \frac{f(x')}{f(x_t)}$, which will be used to decide whether to accept or reject the candidate. Because $f(x)$ is proportional to the density of $P(x)$, we have that $\alpha = \frac{f(x')}{f(x_t)} = \frac{P(x')}{P(x_t)}$. If $\alpha \geq 1$, it means the candidate is more probable than x_t; so you automatically accept the candidate by setting $x_{t+1} = x'$ and so resample the same distribution again. Otherwise, accept the candidate with probability α; if

the candidate is rejected, set $x_{t+1} = x_t$, instead. The algorithm is iterated many times, sometimes accepting the moves and sometimes remaining in place. The key is the acceptance ratio α indicates how probable the new proposed sample is with respect to the current sample, according to the distribution $P(x)$. We will always accept the move to a point that is more probable than the existing point (i.e. a point in a higher-density region of $P(x)$). If we attempt to move to a less probable point, we will only move sometimes. The more the relative drop in probability, the more likely we are to reject the new point. So we tend to return large numbers of samples from high-density regions of $P(x)$, and small numbers from low-density regions. The algorithm returns samples that follow the desired distribution $P(x)$. The disadvantage in the method is the large number of iterations required to get the approximation, which is still subject to some uncertainty. For a full discussion of the relative merits of the methods, see (Taylor and Diggle, 2012).

So both methods have arrived at a means to obtain the desired $p(\boldsymbol{\theta}|\boldsymbol{y})$ despite not knowing the denominator in

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}, \tag{3.10}$$

one by a fast but deterministic approximation and the other by a slow, but non-deterministic, sampling method.

Use of INLA naturally begs the question "What about the approximation error? How confident can we be in these results?" INLA, like any other statistical model, can fail. Empirically, the MCMC error and INLA error are frequently very similar, as has been shown in many simulation studies (Taylor and Diggle, 2012).

Our model in general form is: $g(\mu_i) = \eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(\mu_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i$, which is Equation 3.1 above. Linear regression is a special case where the outcome is Gaussian, the link function is identity, and all the $f^{(j)}(\mu_{ji}) = 0$. To briefly compare I can look at a simple linear regression:

$$E(y_i|x_i) = \alpha + \beta x_i,$$

and see how INLA and MCMC compare in fitting such a Bayesian model. I use R-JAGS (Plummer, 2009), an R implementation of JAGS, to fit the MCMC model (Lunn et al., 2009). INLA claims the advantage of speed, and a look at Table 3.1

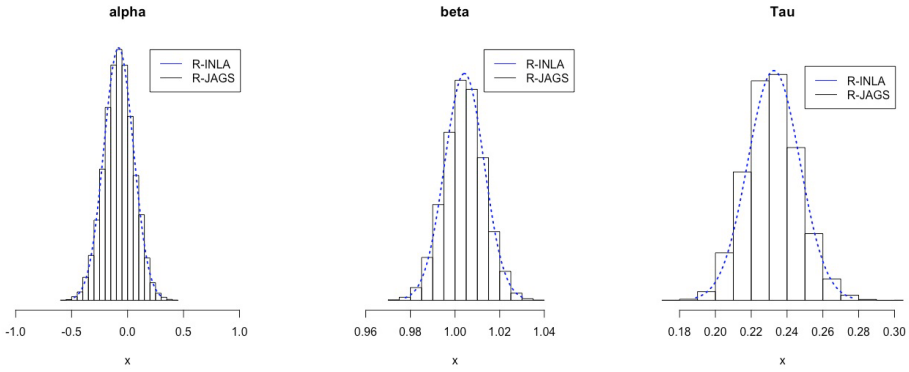

Figure 3.2: **Posterior Parameter Distribuions.** A comparison of the estimated distributions. The approximated posteriors from R-INLA (blue) compare closely to the iterated approximated posteriors from MCMC found using R-JAGS for each of the Intercept, Slope and Precision parameters.

shows that INLA is indeed much faster. INLA models run in seconds, whereas the MCMC (burn in 5000, 10000 iterations) took much longer. Examining Figure 3.2 there does not seem to have been significant difference, or much lost, when approximating the posterior deterministically using INLA. The distributions are nearly coincident. Furthermore, when one considers that the largest models which appear later in this thesis took some 40 hours to run using INLA, and that there were dozens of them, the study presented in Chapter 7 would not have been possible via MCMC, just due to the constraints of the computing power available to me, (Dalhousie University, Department of Mathematics and Statistics, 2 x Quad Core Xeon Nehalem X5550 2.66Ghz processors with 512GB of RAM).

N	R-INLA	R-JAGS
100	0.24	4.15
500	1.26	9.64
5000	3.95	114.61
25000	18.51	606.77
50000	31.66	3311.88
100000	71.24	7922.44

Table 3.1: **INLA vs. MCMC** Elapsed computation time (in seconds) for equivalent models.

3.3 SPDE: Stochastic Partial Differential Equations

The mean of our response, at some arbitrary location s and time t, $E[Y(s,t)] = \mu(s,t)$, can be mapped by a link function g(.) (Zuur et al., 2009) to a linear predictor $\eta(s,t)$ as in the Generalized Additive Model framework (Hastie and Tibshirani, 1990). That is,

$$g[\mu(s,t)] = \eta(s,t) = z(s,t)\beta + \xi(s,t), \tag{3.11}$$

where the linear predictor is of two parts, a linear combination of fixed covariates $z(s,t)$ and a spatial temporal random effect $\xi(s,t)$. The first, $z(s,t)$ is the vector of covariates for location s and β is the vector of coefficients (the t in $z(s,t)$ is redundant here since neither of our covariates vary in time but retain it for generality). The second piece, the latent process, $\xi(s,t)$, is of particular interest. It represents the effect of unknown factors impacting the response such that any spatial and/or temporal variation in $\xi(s,t)$ may be thought of as variation due to any number of factors influencing the response that are not explicitly included in the model. For illustration, suppose the response of interest in our marine environment varies both spatially and temporally due to latent factors. Reasonable conjecture is that we may get larger responses when and where the latent response is large. As an example, one factor sure to influence this is the distribution of the specie(s) being preyed upon. If this prey abundance is not included explicitly in our model, it will influence the latent variable. By considering $\xi(s,t)$ as a Gaussian random field (GF) this random field can be viewed as a *proxy* of the effect of these latent factors. Let $X(s) \equiv \{x(s), s \in D \subseteq R^2\}$ denote the field that to model. The SPDE approach (Lindgren and Rue, 2015) represents this GF with a Gauss Markov Random Field (GMRF), Lindgren and Rue (2015) showed that there is a mapping of the covariance function of the Gaussian field to the GMRF such that the spatio-temporal model can be rewritten in terms of a GMRF. A GMRF is a spatial process that models spatial dependence on a grid or lattice (Illian et al., 2012b), (or on a map of geographical regions if a set of neighbours is defined (Gardner et al., 2008; Ross et al., 2012). Thus, this approach is estimating not the *continuously* indexed GF with attendant computational expense but instead

an *discretely* indexed GMRF. The SPDE approach uses linear combinations of basis functions defined on the locations of the set of vertices in the triangulation to represent the field. The basis function representation of the original field $X(s)$ is;

$$X(s) = \sum_{l=1}^{n} \psi_l(s)\epsilon_l, \tag{3.12}$$

where n is the number of vertices in the triangulation, $\psi_l(s)$ are the basis functions and ϵ_l are gaussian weights. The basis function $\psi_l(s)$ is equal to 1 at vertex l and 0 at all other vertices. The value of the field at any vertex is given by ϵ_l and values for the interior of the triangles are determined by linear interpolation. The objective is to find a GMRF and temporal correlation structure that best represents $X(s)$.
A spatio-temporal hierarchical model framework is a natural choice where data are *observed* encounters and of interest is the *process* generating them. A hierarchical model allows one to account for spatial and temporal correlation while also being able to separate the observations from the process as sources of uncertainty. This is important when attempting to interpret the resulting model.
The speed of the INLA approach allows a variety of candidate models to be fitted to the encounter event data and compared. We commenced with models with no spatial or temporal effect at all which amounts to simply modelling the response as a function of the fixed covariates. We also consider models with a single spatial effect constant in time as well as those with temporally varying fields. In the latter case the models considered have spatial effects replicated each year, that is, there is a single spatial effect for each year (without temporal correlation), spatial effects correlated between years using an ar(1) model and a spatial effect with exchangeable correlation between years. Note that if the fixed covariates completely capture the response variation $\xi(s,t)$ would not be needed in the model (it's spatial effect would be insignificant). Conversely, if $\xi(s,t)$ remains a significant predictor, then there remains variation in the response unexplained by the covariates.
The values in the field may be related, for example, values close to one another spatially may differ less than values that are far apart. When this exists it is termed *covariance* and the field may have a covariance structure (Vanmarcke, 1988). This covariance may have a functional form, and there are many valid functional forms

of which the Matérn (Matérn, 1960) which we focus on throughout the remainder of this document is but one.

Lindgren and Rue (2015) show that one can express a large class of random field models as solutions to continuous domain Stochastic Partial Differential Equations (SPDEs), and further, that there is an explicit link between the parameters of the SPDE and the elements of the precision matrix if one chooses a discrete basis function representation of the continuous random field. Take a SPDE (this is just a requirement-the field must have this property-like an assumption, *why* this assumption, tractable);

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}}(\tau x(s)) = W(s), \qquad s \in \Omega, \tag{3.13}$$

where $x(s)$ is the spatial field, κ is the spatial scale parameter, Δ is the Laplacian, α controls the smoothness, τ controls the variance, Ω is the spatial domain and $W(s)$ is Gaussian spatial white noise. Stationary solutions to this SPDE on R^d have Matérn covariance functions (Whittle, 1953, 1963), meaning,

$$COV(x(\mathbf{0}), x(\mathbf{s})) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa||s||)^{\nu}K_{\nu}(\kappa||s||). \tag{3.14}$$

This implies that to estimate the spatial field one needs to estimate the parameter (κ, σ^2) where

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{\frac{d}{2}}\kappa^{2\nu}\tau^2}. \tag{3.15}$$

These then are the salient features of the models to be demonstrated in this thesis. Latent Gaussian models using a GMRF to model the spatial structure present in our spatio-temporally explicit data which are able to give insight into the underlying biological process being observed. These LGMs, fit via INLA and the SPDE, have relatively few parameters and these parameters will be seen to be interpretable and understandable in biological terms when applied to our marine data. The latent GMRF models are flexible enough to allow for the integration of multiple data sources, with multiple resolutions, within the same model and, all this is available in a package with large computational advantage over the alternative. In what follows in this paper each of these features is explored in detail.

Chapter 3 explores the idea of modelling the mean of a biological process in a geo-statistical way, and shows the resulting model to be sensible and meaningful. Chapter 4 explores an abundance model for a species and lays out the idea that the parameters of our GMRF models correspond well to the biological criteria by which a population and its abundance are characterized. Chapter 5 takes this basic correspondence much further, and explores the dynamic responses of the population structure to external forces (overfishing) and reveals the developing crisis in the fish stock as one of the erosion and eventual near destruction of the population structure, all as seen in the patterns of movement of the parameters of our GMRF based abundance model. Finally, we examine the ability of these models to integrate varied and diverse data in Chapter 7, in which data from multiple sources and with various resolutions are combined onto a single model which examines, and models jointly, predator behaviour and prey abundance. These integrated models in Chapter 7 are the outcome possible from the speed of INLA: with more than one million observations of the response INLA enabled the fitting of not just an integrated model of both predator and prey as described, but enabled the exploration of multiple different prey and multiple different model formulations where even a single one might have been impossible without such speed and flexibility.

Chapter 4

Seal encounters at sea: A contemporary spatial approach using R-INLA

4.1 Abstract

Acoustic telemetry is an active field of research integral to the study of marine life. The latest generation of acoustic tags is making available new types of data. As part of the Ocean Tracking Network (OTN) (**www.oceantrackingnework.org**) acoustic tags known as VEMCO Mobile Transceivers (VMTs) (**www.vemco.com**) are being deployed on Sable Island grey seals (*Halichoerus grypus*) in order to record instances of proximity to each other (as well as any other acoustically tagged animals). The seals essentially become *bioprobes* yielding data exhibiting both spatial and temporal variation. Fortunately, recent developments in the field of spatial statistics have greatly facilitated the fitting of complex spatial and spatio-temporal models. Here we specifically propose a hierarchical spatio-temporal model framework and fit it to these data using both Stochastic Partial Differential Equations (SPDE) and Integrated Nested Laplace Approximations (INLA) through R-INLA (**www.r-inla.org**). In so doing we demonstrate the effectiveness and advantages of these techniques. These methods readily extend to spatially explicit data collected by any sort of mobile receiving platform (e.g. wave gliders, remotely operated underwater vehicles).

4.2 Introduction

Innovative tagging technologies are granting researchers access to previously unavailable descriptors of animal movement. One example, the VEMCO Mobile Transceiver (VMT) (**www.vemco.com**), acoustically records instances of proximity of instrumented animals to other, similarly tagged animals. The Ocean Tracking Network (OTN) (**www.oceantrackingnework.org**), has deployed VMTs on

Sable Island grey seals (*Halichoerus grypus*). This island lies about 300 km South-East of Halifax, Nova Scotia, Canada. In 2009, 14 grey seals on Sable Island were randomly [1] selected and fitted with VMTs. Seals so instrumented are referred to as *bioprobes*, reflecting the fact that the organism itself is being used as the means of interrogating it's environment. These bioprobes are anticipated to record a contact whenever one comes within approximately 500 meters of another seal (or any other tagged animal for that matter). Thus they deliver a data set comprising the times and locations of the pair-wise proximities of these grey seals, hereafter referred to as *encounters*. The experiment has continued through 2010, 2011 and 2012. In 2010 the sample size was increased to 20 seals. The 2011 and 2012 experiments retained the sample size of 20 seals and resulted in encounters occurring in roughly, the same area of the Scotian shelf.

Data of the type considered here, that is, times and locations of encounters between marine animals, has simply not been available until recently because its collection was infeasible prior to the advent of the VMT (O'Driscoll et al., 2000). However, other emergent means of interrogating the oceanic biosphere, such as wave gliders and remotely operated underwater vehicles, will yield similar data. It is natural to propose spatial approaches for this type of data since they allow us to formally describe the data generating processes, hence a thorough demonstration of their utility is both timely and warranted.

We note that the data of interest here are at sea seal to seal encounters as distinct from *tracks* of seals (which are commonly associated with studies that use acoustic telemetry) (Jonsen et al., 2005; Patterson et al., 2008). Although these data sources, when available simultaneously, could be considered together, here we choose to model only the encounters. Our motivation is twofold: (1) We wish to determine what can be learned directly from encounters in space through time and, (2) Statistical methods should indeed be available with which to interpret them. Encounters can be, and are, being recorded between bioprobes associated

[1] Seals are selected from amongst animals already branded by the Department of Fisheries and Oceans Canada. These animals number approximately 1000 and are known to return to Sable Island, an important factor when tags must be physically recovered in order to obtain the data.

with this study and compatible tags deployed in the ocean environment *not associated with this study*. Given this fact, analysts will see cases where tracking data will be available in conjunction with encounter data, and cases where it will not.

Scientists who study seal behaviour have noted that the Sable Island grey seals display a preference for the banks off of Sable Island; areas where the seabed rises to form underwater bluffs or banks. Use of the banks by the seals should reveal itself by showing a relationship between the encounters and depth of water. It has also been suggested that grey seal foraging behaviour is governed by the need to maximize their (caloric) benefit whilst minimizing their energy expended to achieve the benefit (Lidgard et al., 2012) so we might therefore also expect there to be some relationship between the nature or duration of the encounters and the distance the seals must travel from Sable Island. Our approach allows us to gain valuable insight into the nature and extent of grey seal encounters at sea and test hypotheses concerning their behaviour. However, our intent here is to demonstrate the utility of these approaches by showing that they lead to credible results, not to make definitive statements about the results themselves. The data sets available are based on small samples, necessarily so due to the high costs of the technology [2]. These small samples are of the very large seal population using Sable Island to haul out (i.e. come ashore) which is in the order of 350,000 animals (Lidgard et al., 2012).

A spatio-temporal hierarchical model framework is a natural choice here as we have *observed* encounters and are interested in the *process* generating them. A hierarchical model allows us to account for spatial and temporal correlation while also being able to separate the observations from the process as sources of uncertainty. This is important when attempting to interpret the resulting model. Despite their attractive features, hierarchical models present difficulties to the analyst. Markov Chain Monte Carlo (MCMC) methods can be used for their fitting but are computationally intensive as well as demanding a fair degree of statistical sophistication in order to both fit and interpret models as well as interpret convergence diagnostics. Recently, a viable method for obtaining posterior distributions for a Bayesian hierarchical model has been proposed that avoids MCMC methods.

[2] VMTs are $3300 USD each.

Called INLA, for Integrated Nested Laplace Approximation (Rue et al., 2009b) and available as an R (R Core Team, 2013) package (**www.r-inla.org**),(R Core Team, 2013), R-INLA offers the analyst the opportunity to perform Bayesian hierarchical modelling relatively easily and efficiently (Illian et al., 2012a). This computational advantage also allows various candidate models to be expeditiously fitted and compared using the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). In addition to the flexibility available with which to specify the data generating process and the spatio-temporal correlation, R-INLA can straightforwardly incorporate any covariates of interest (Illian et al., 2012b), here we utilize this feature to link the distribution of the encounters to characteristics of the physical ocean where they arise.

In Section 4.3 we fully describe the data and present some summary statistics. In Section 4.4 we present our hierarchical model formulation, first outlining the spatio-temporal problem and the rationale for the Gauss Markov random field (GMRF) representation of the Gaussian Field. We then discuss the means of the SPDE approach used by R-INLA to fit such models and close Section 4.4 by listing what it is we want to extract from the model. In Section 4.5 we present and interpret the model and describe how the model parameters are presented to give interpretable results. We discuss these results in Section 4.6, relating the model to the scientific hypotheses and placing them in context. We finish by outlining possibilities for future work.

4.3 Data

Each grey seal bioprobe carries a VMT and a SPLASH tag (**www.wildlifecomputers.com**). For each encounter the SPLASH tag records the GPS location[3] while the VMT records the identity of the other transmitter involved along with the date and time. Both tags are archival such that they must be physically retrieved in order to recover the data. Once recovered the data are aligned to produce a complete encounter dataset where an individual encounter record consists of the date, time, latitude and longitude of the encounter and the identities of the seals involved.

[3]The SPLASH (Mk 10AF) tag can also record diving depth, light level and temperature.

In 2009 there were 1,444 encounters. Those in which the transmitting seal identification number and the receiving seal identification number were one and the same were removed on the advice of the tag manufacturer; as spurious self detections. After cleaning in this way there remained 892 encounters in the 2009 data set. In 2010 and later the technology was improved such that cleaning in this way was no longer required.

Analysis of the seal encounters is complicated somewhat by the functioning of the VMTs. Each has a signal pulse rate of approximately two minutes. That is, it sends out a signal, a pulse, every two minutes on average. Since, once in proximity a pair of seals are likely to be re-observed at the sample rate of the tag there will likely be subsequent encounters over the ensuing time at the 2 minute refresh rate of the tag. Eventually the seals separate and the series of encounters ends. Each of these series of encounters can be viewed together as an *encounter event*. This has the effect of introducing spatial dependence and temporal dependence since, given an initial encounter it is very likely that there will be further encounters proximate in both space (the seals don't go far in 2 minutes) and time (2 minutes later). Indeed, for each encounter event there is all but zero spatial displacement, the initial encounter and all of the subsequent encounters are co-located, at least at the resolution of the sensor which is 0.01 degrees latitude/longitude.

Grouping the data into encounter events yields 129 encounter events for 2009, 195 for 2010, 207 for 2011 and 412 for 2012. Note that there are no instances in the dataset where more than two seals encountered each other simultaneously. This is primarily due to the very large area of the Scotian shelf relative to the detection range of the tags combined with the sample size, the approach we take could accomodate any multi seal encounters that did occur since what we are interested in is a measure of how long any given seal remains at a given location. A summary is provided in Table 4.1 while Figure 4.1 serves to locate these data geographically, in the context of the coast of Nova Scotia and Sable Island.

Year	2009	2010	2011	2012	ALL
Encounter events	129	195	207	412	943

Table 4.1: **Summary.** Table of seal-to-seal encounter events observed on the Eastern Scotian Shelf, 2009-2012.

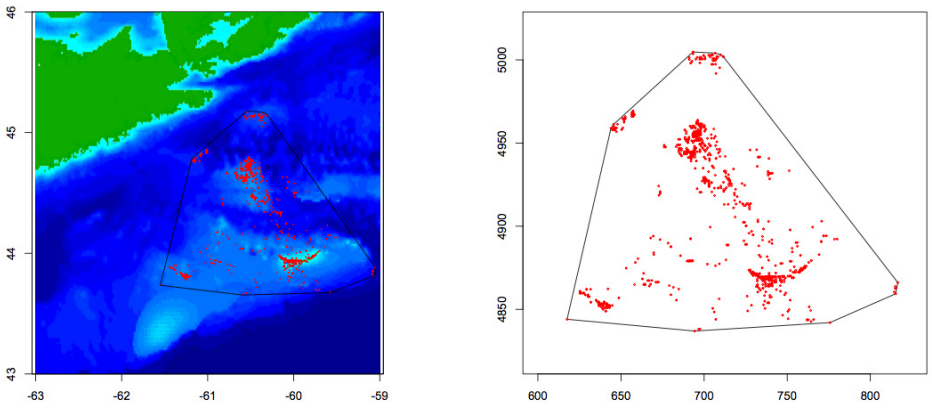

Figure 4.1: **Seal encounters off Sable Island, Nova Scotia.** Sable Island is a large (34 km^2) sand bar, the entirety of which constitutes a Canadian National Park. Seal to seal encounters (2009-2012) in the area of Sable Island are shown in red. The black line represents a bounding box that is used later for calculation of the spatial domain. Analysis of the encounters was performed after projection onto the UTM coordinate system, the right panel shows the same encounter pattern with UTM coordinates.

The encounter data set is temporally discontinuous as the tags were deployed in the late summer and recovered around the new year for each of 4 years (and continue to be collected as at the time of writing). The dates of deployment and recovery of the tags vary such that there are several month and year combinations that are not observed. In addition, in 2012 there were a very small number of encounters observed well to the Northwest in the area of the Madeleine Islands in the Gulf of St. Lawrence. Since all of the 2009, 2010 and 2011 data lie in the vicinity of Sable Island we excluded this small subset. This keeps the size of our domain reasonable (see Figure 4.1) and common to all years.

In order to examine behavioural hypotheses about preference for the banks and conservation of energy expended in foraging, we created two spatial covariates, DEPTH and DIST. The bathymetry of the ocean area where the seals interact (DEPTH) was obtained from the US National Oceanic and Atmospheric Administration (NOAA) (**www.ngdc.noaa.gov/mgg/global/global.html**) with a resolution of 1 minute of longitude and latitude (Amante and Eakins, 2009). Distance

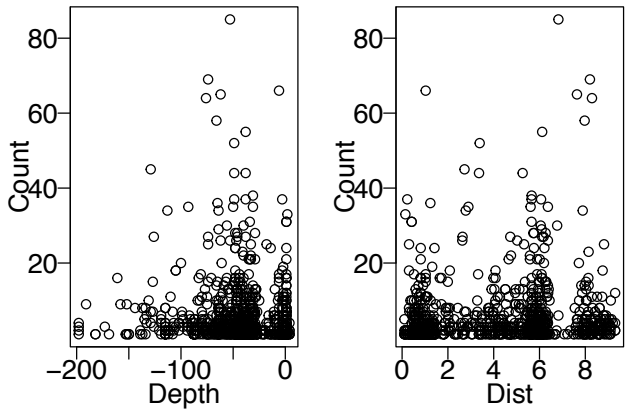

Figure 4.2: **Spatial covariates**. These plots indicate there is some relationship between the response and the covariates but it is not definitive. The R^2 value for the log response was .3636 for Depth and .4399 for Dist.

from Sable Island (DIST) was calculated as the great circle distance from an (arbitrary) point at the centre of Sable Island. A plot of these spatial covariates is provided below (see Figure 4.2).

4.4 Model Formulation

Spatio-temporal model structure

Here we present a Bayesian hierarchical spatio-temporal model framework for the at sea seal to seal encounter events. A hierarchical model has a number of components, the selection of which determines the model (or models) which best describe the data. A data component specifies the distribution of the data, a process component specifies the underlying mechanism which gives rise to the data and a parameter component specifies the distributions of any parameters contained in the process component. Such hierarchical formulations have the advantage that the observations and the process are in some sense modelled separately.

The response variable $y(s,t)$ is the total number of encounters (subsequent to the initial one) taking place between two seals during a particular encounter event (0 to a maximum observed count of 84) at location s at time point t, $t \in (1, ..., 4)$ corresponding to the four years in the study (2009-2012). As such we consider

candidate distributions suitable for count data including the Poisson and negative binomial distributions (each with their respective canonical link functions (Zuur et al., 2009))[4].

The mean of our response, $E[Y(s,t)] = \mu(s,t)$, is mapped by a link function to a linear predictor $\eta(s,t)$ as in the Generalized Additive Model framework (Hastie and Tibshirani, 1990). That is,

$$\log[\mu(s,t)] = \eta(s,t) = z(s,t)\beta + \xi(s,t) \tag{4.1}$$

where the linear predictor is of two parts, a linear combination of fixed covariates $z(s,t)$ and a spatial temporal random effect $\xi(s,t)$. The first, $z(s,t)$ is the vector of covariates for location s and β is the vector of coefficients (the t in $z(s,t)$ is redundant here since neither of our covariates vary in time but we retain it for generality). The second piece, the latent process, $\xi(s,t)$, is of particular interest. It represents the effect of unknown factors impacting the response such that any spatial and/or temporal variation in $\xi(s,t)$ may be thought of as variation due to any number of factors influencing the response that are not explicitly included in the model. For illustration, suppose the foraging quality in our marine environment varies both spatially and temporally due to latent factors. We conjecture that we may get larger responses when and where the seals are foraging in the best locations. As an example, one latent factor sure to influence this is the distribution of the specie(s) being preyed upon. By considering $\xi(s,t)$ as a Gaussian random field (GF) this random field can be viewed as a *proxy* of the effect of these latent factors.

GMRFs and the SPDE approach

GFs are conventionally defined by their mean and a covariance function $Cov[(s,t),(s',t')]$ defined for each $(s,t),(s',t')$ in $R^2 \times R$. Direct implementation of such a field becomes difficult with large datasets,(Simpson et al., 2012a,b; Taylor and Diggle, 2012), the so-called *big n* problem. It is in response to this problem that the SPDE approach has been proposed, which essentially amounts to representing

[4]One could consider the response as continuous time, discretized by the 2 minute sampling rate. We did examine models with continuous responses but the discrete response proved preferable.

the GF with a GMRF. That is, the continuously indexed Gaussian random field is represented as a *discretely* indexed random process, a GMRF, with attendant computational advantages. Readers interested in a thorough explanation are referred to (Lindgren et al., 2011) for proofs and theoretical details. If, for example, we assume first order auto-correlated spatio-temporal effects between years with coefficient a ($|a| < 1$) then,

$$\xi(s,t) = a\xi(s,t-1) + \omega(s,t), \tag{4.2}$$

where $\omega(s,t)$ has a zero mean gaussian distribution, is temporally independent, and has a spatial covariance function

$$Cov[\omega(s,t), \omega(s',t')] = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C(h;\nu,\kappa) & \text{if } t = t', \end{cases}$$

for $s \neq s'$ where,

$$C(h;\nu,\kappa) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^\nu K_\nu(\kappa h). \tag{4.3}$$

$C(h;\nu,\kappa)$ is parameterized by ν and κ, $\nu > 0, \kappa > 0$, and K_ν is the modified Bessel function of the second kind. The parameter ν determines smoothness and κ determines spatial scale and the covariance function depends only upon the distance separating the locations $h = \|s - s'\| \in R$. In practice, the parameter ν is usually fixed (we take $\nu = 1$) and $\rho = \frac{\sqrt{8\nu}}{\kappa}$ is reported empirically with ρ being the distance at which the spatial correlation is reduced to approximately 0.1 (Cameletti et al., 2011) (Cameletti et al., 2013).

Let $X(s) \equiv \{x(s), s \in D \subseteq R^2\}$ denote the Matérn field that we wish to model, i.e. the continuously indexed GF with Matérn covariance function, defined by parameters κ and ν on some spatial domain D. A GF with a Matérn covariance function can be represented as a GMRF. A GMRF is a spatial process that models spatial dependence on a grid or lattice or graph (Illian et al., 2012b), (a map of geographical regions is a graph if a set of neighbours is defined (Gardner et al., 2008; Ross et al., 2012). The grid or lattice or graph is used to define a neighbourhood structure. Notationally, if we let x_l, $l \in (1, ..., n)$ be the value of such a field at locations l, and, calling the set of neighbours to component l, δ_l, then

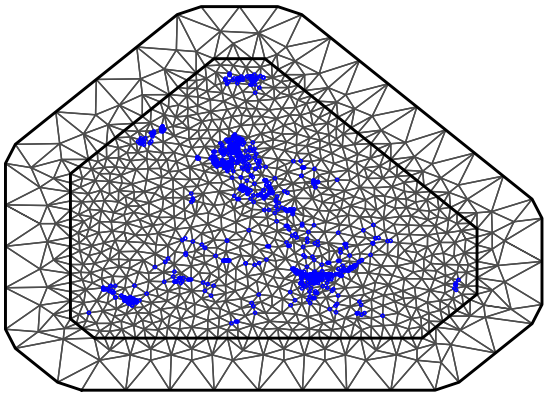

Figure 4.3: **Triangulation.** The triangulation utilized for the SPDE approach with the encounter events superimposed in blue.

x_i and $x_j, i \neq j$, $i, j \in (1, ..., n)$, are independent unless they are neighbours. The importance of this is seen in the precision matrix Q, where the non-zero elements of Q are determined by the neighbours. So $Q_{ij} \neq 0$, if and only if $j \in \{i, \delta_i\}$. The conditional dependence based on this neighbourhood structure gives a sparse matrix Q, the computational advantage of a GMRF is a consequence of this property (Lindgren and Rue, 2015; Lindgren et al., 2011).

The objective is to find a GMRF with some local neighbourhood and sparse precision matrix that best represents $X(s)$. As an alternative to using a regular grid, the SPDE approach utilizes a triangulation of the domain (Lindgren et al., 2011). The domain is subdivided into non-intersecting triangles with vertices at the data locations. Additional vertices are added sufficient to get a useful triangulation. Figure 4.3 presents a triangulation of the space creating a domain for the encounter event data.

The SPDE approach uses linear combinations of basis functions defined on the locations of the set of vertices in the triangulation to represent the field. The basis function representation of the original field $X(s)$ is;

$$X(s) = \sum_{l=1}^{n} \psi_l(s)\epsilon_l, \tag{4.4}$$

where n is the number of vertices in the triangulation, $\psi_l(s)$ are the basis functions

and ϵ_l are gaussian weights. The basis function $\psi_l(s)$ is equal to 1 at vertex l and 0 at all other vertices. The value of the field at any vertex is given by ϵ_l and values for the interior of the triangles are determined by linear interpolation. (Lindgren et al., 2011) showed that once written this way there is a mapping of the covariance function $C(h; \nu, \kappa)$ of the Gaussian field to the precision matrix Q of the GMRF such that the spatio-temporal model can be rewritten in terms of a GMRF.

The R-INLA package provides numerous options for specifying the triangulation, allowing for the choice of maximum side lengths and internal angles to determine size and shape of the triangles, specifying convex hulls or other boundaries for the points, construction of the triangulation with or without the points themselves and offering control of the degree of inflation provided to compensate for the boundary effect.

The construction of the triangulation is an important step in constructing the model and some care is needed in choosing the triangulation upon which to perform the analysis. It is important to have triangles that are of regular size and regular shape. Further, the triangulation needs to be extended spatially in order to avoid the *boundary effect*, an inflation of the variance at the borders of the space (Lindgren et al., 2011). The R-INLA package provides numerous options for specifying the triangulation. Controlling the maximum allowable triangle edge length and minimum allowable triangle internal angle allows the user to avoid very large or very small triangles, or triangles with very odd shape. Importantly, the triangulation may need to be re-evaluated after running the initial models and obtaining an estimate of the spatial range parameter ρ. Since both the degree of spatial inflation required and the largest triangle size desired are affected by ρ. We obtained a triangulation with n=3728 vertices using a maximum side length of 4 kilometers.

Fitting the hierarchical model using R-INLA

INLA is the computational approach to Bayesian modelling proposed by (Rue et al., 2009b) as an alternative means to MCMC for obtaining approximated posterior marginals for latent variables and hyper-parameters of the model. If we let the parameter to be estimated be $\theta = (\mu, \phi, a, \sigma^2_\omega, \kappa)$, where $\mu = \{\mu(s,t)\}$ is the

vector of mean responses $E[Y(s,t)] = \mu(s,t)$, ϕ is the parameter of the response distribution (the size parameter of the negative binomial distribution, for example), a is the persistence parameter in the ar(1) model for $\xi(s,t)$ as assumed in Equation 4.2, σ_ω^2 is the variance of the Matérn field and κ is as in Equation 4.3, then the posterior distribution is given by.

$$\pi(\theta, \xi|y) \propto \pi(y|\xi, \theta)\pi(\xi|\theta)\pi(\theta), \tag{4.5}$$

and $\xi = \{\xi_1, \xi_2, \xi_3, \xi_4\}$ for $t = 1, ...4$, that is, one realization of the spatial random field for each year. We assign independent prior distributions to parameters (using the default priors as specified in R-INLA), such that we can rewrite the posterior distribution as,

$$\pi(\theta, \xi|y) \propto \left(\prod_{t=1}^{4} \pi(y_t|\xi_t, \theta)\right)\left(\pi(\xi_1|\theta)\prod_{t=2}^{4}\pi(\xi_t|\xi_{t-1}, \theta)\right)\pi(\theta). \tag{4.6}$$

The posterior marginal for β_k is

$$\pi(\beta_k|y) = \int \pi(\beta_k|\theta, y)\pi(\theta|y)d\theta, \tag{4.7}$$

for $k = 1, 2$. Estimating this distribution will give us the distribution of the coefficients for the fixed covariates. The random field has the following posterior marginal distribution,

$$\pi(\xi_t|y) = \int \pi(\xi_t|\theta, y)\pi(\theta|y)d\theta, \tag{4.8}$$

for $t = 1, ..., 4$. ξ_t is the spatial effect, that is the effect of the latent variables, for year t.

4.5 Results

A variety of candidate models are fitted to the encounter event data. We commence with models with no spatial or temporal effect at all (*model*.0) which amounts to simply modelling the response as a function of the fixed covariates. We also consider models with a single spatial effect constant in time (*model*.1) and well as those with temporally varying fields. In the latter case the models considered have

spatial effects replicated each year ($model.2$), that is, there is a single spatial effect for each year (without temporal correlation), spatial effects correlated between years using an ar(1) model ($model.3$) as well as a spatial effect with exchangeable correlation between years ($model.4$). Note that if the fixed covariates completely capture the response variation $\xi(s,t)$ would not be needed in the model (it's spatial effect would be insignificant). Conversely, if $\xi(s,t)$ remains a significant predictor, then there remains variation in the response unexplained by the covariates.

The models described herein were run at Dalhousie University's department of Mathematics and Statistics using 2 x Quad Core Xeon Nehalem X5550 2.66Ghz processors with 18GB or RAM. Fitting the described models with $n = 3728$, $max.edge = 4km$ took anywhere from a few seconds for the simplest ($model.1$) to 29 minutes for the most complex ($model.3$ or $model.4$).

We select the best time structure the same way as we select the best distribution for $y(s,t)$. That is, by running models using the various alternative constructions and comparing them using DIC. Results are summarized in Table 4.2.

	poisson	nbinomial
model.0	12123.58	4971.12
model.1	6549.50	4721.76
model.2	5721.98	4650.31
model.3	5722.15	4647.83
model.4	5721.47	4650.21

Table 4.2: **DIC values**. DIC values for various candidate models.

Examining Table 4.2 reveals that for each spatio-temporal model formulation considered, the DIC for the negative binomial is lower than that of the Poisson. Hence the negative binomial distribution is preferred. Amongst only these, the lowest DIC is obtained by $model.3$ although $model.2$ and $model.4$ are very close. While the inclusion of the ar(1) temporal structure ($model.3$) reduces the DIC slightly the required parameter is not well estimated, the case is similar for $model.4$. As such, results and discussion that follow pertain to $model.2$, which is both parsimonious and well performing.

Fixed Covariates

Posterior summary statistics for fixed covariates DEPTH and DIST are provided in Table 4.3. Both of the fixed covariates contribute significantly to the model.

model.2	mean	sd	0.025quant	0.5quant	0.975quant
β_{DEPTH}	-0.0062	0.0024	-0.0108	-0.0062	-0.0015
β_{DIST}	0.1494	0.0308	0.0889	0.1499	0.2099

Table 4.3: **Fixed Effects.** Summary of Fixed Effects in *model*.2, our chosen model.

Hyper-parameters

It is the hyper-parameters that specify the spatial effect (the GMRF) in the model. A summary of the hyperparameter values for *model*.2 is provided in Table 4.4.

m.2	mean	sd	0.025quant	0.5quant	0.975quant
ρ	8.555	2.418	4.852	8.183	14.303
σ^2_ω	0.9452	0.1636	0.6641	0.9306	1.306
ϕ	0.6841	0.0523	0.5871	0.6820	0.7926

Table 4.4: **Hyper-parameters.** Summary of hyper-parameter values for *model*.2, our chosen model.

$\rho = 8.555$ is the distance (in kilometers) at which the spatial correlation is reduced to approximately 0.13. The variance of the GMRF is $\sigma^2_\omega = 0.9452$. For the negative binomial distribution assumed for the responses, $\phi = log(n) = 0.6841$. The size parameter of a negative binomial distribution reflects dispersion and is related to the variance, $\sigma^2_{nbin} = \mu(s,t)(1+\frac{\mu(s,t)}{n})$. As n becomes large the variance appraoches the mean, smaller values of n are indicative of over-dispersion. Our estimate is indicative of over-dispersion, common in ecological data (Bolker, 2008)

By assembling the above components we can nicely display the results. Combining the effects of the fixed covariates with the model output for the random field gives us the posterior mean of the linear predictor. A plot of this mean for 2012, is shown in the top left panel of Figure 4.4 and consists of the sum of the linear predictors in the model and the elements of the random field $[z(s,t)\beta + \xi(s,t)]$. It is also possible to examine the random field (the GMRF) on it's own as is shown in the top right panel of Figure 4.4. Recall that the posterior mean of the linear

predictor is the logarithm of the mean of the negative binomial distribution used to model the response. By exponentiating we can produce a spatial plot on the natural scale of the mean event lengths as in the bottom right panel of Figure 4.4.

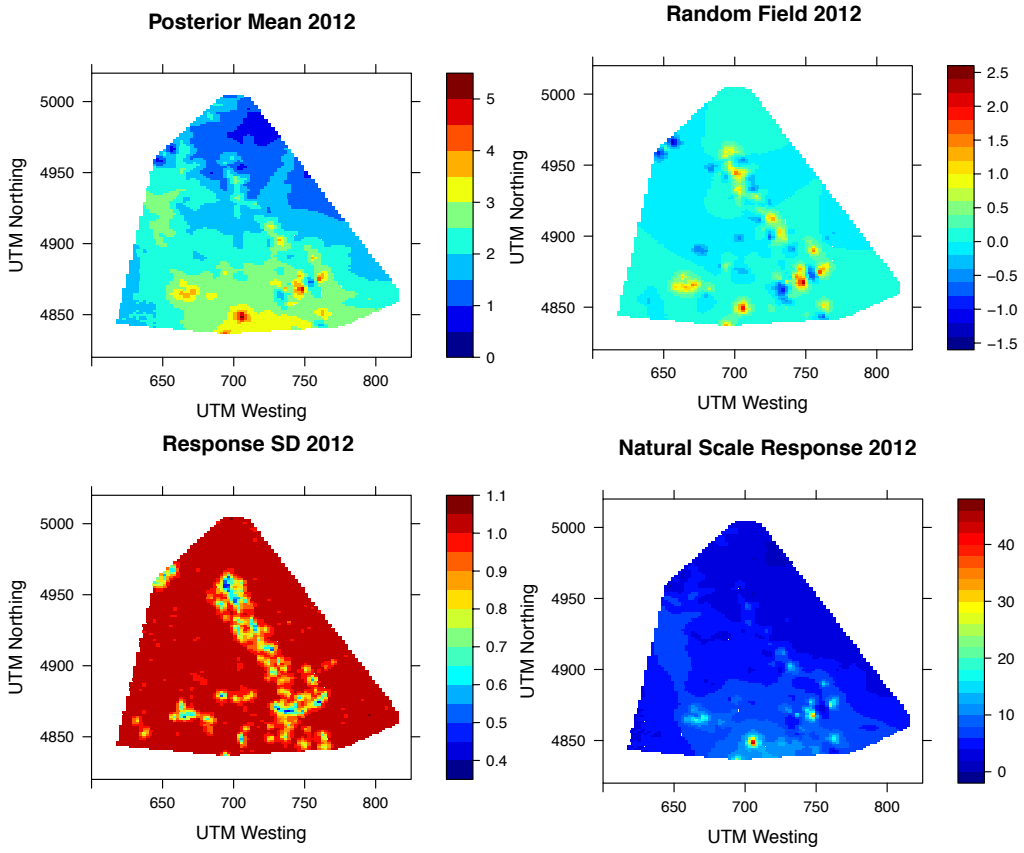

Figure 4.4: **Examining Results** Plot of the posterior mean (top left), random field (top right), standard deviation (bottom left) and mean response (bottom right) for 2012.

4.6 Discussion

We have proposed here hierarchical spatio-temporal models for novel acoustic telemetry data, in particular encounter event data currently being collected by bioprobes. These models allow for the inclusion of available covariates and for investigation of latent factors. Using SPDEs and INLA they can be fit quickly thereby allow for efficient model comparison via DIC.

We have provided statistical evidence of a negative relationship between bathymetry and response where large response values are indicative of encounter events involving many subsequent encounters (i.e. encounter events of long duration) . We find that long encounter events tend to occur in shallower water, presumably corresponding to the hypothesis that foraging behaviour is occurring on the banks. Conversely, shorter duration encounter events in deeper water can be envisioned as seals *passing* in transit between the forage area and their location of haul out (Sable Island).

Statistical evidence of a positive relationship between the response and the distance from haul out was found. Large response values occur to northwest and southwest of Sable Island, within the domain created by drawing a convex hull about the data. Since the large values occur away from Sable, the parameter estimate is positive. If one looks at the bathymetry and the domain however it is plain that the domain encompasses Sable Island and the two closest (and nearly equidistant) areas of undersea bank. We are seeing exactly what we anticipated, that seals use preferentially the *closest* area of bank water.

Examining Figure 4.4 we see that there are areas of high mean response (the red areas), that is to say that the red areas are locations where the predicted count of subsequent encounters is large. These areas of high mean response are consistent with the fixed effects (i.e. they are *on* the banks), but it is even more complex. Each red spot (or spots) on a bank is (are) small in relation to the area of bank water. Some other (local and latent) factor may be present at these locations and influencing seal behaviour. By local we mean what is it about *that* spot that leads to a high response as opposed to another area nearby with the same depth and distance from Sable Island, but with a much lower posterior mean. Whatever is causing this difference is presently unknown and hence not explicitly included in the model. One could postulate that this could be the effect of, say, simply the distribution of the prey species. Where the prey is abundant the seals linger and the measured response is large.

Figure 4.4 (top right) shows several red areas, where there are large values of the random field. These areas are again quite small (local). Recall that the data is characterized by spatial correlation that decreases fairly quickly over distance,

the value of ρ corresponding to a decrease in spatial correlation to approximately 0.13 over a distance of about 8.5 kilometers. These locales may be thought of as *hot spots* where the response is large even after allowing for the fixed covariates. The response is quite variable, because it is presumably driven by the need to forage, and foraging, like most animal behaviours, is patchy and variable, highly dependent upon local conditions. The short range of the spatial effect combined with the small numbers of observations in parts of the domain imply that caution should be used when evaluating the predicted mean in these sparsely observed areas. The bottom left panel of Figure 4.4 shows the variance of the response. We see that this variance has similar magnitude to the variance of the GMRF, $\sigma^2_\omega = 0.9452$, over most of the domain indicating the two are about equivalent as sources of uncertainty. The spatial effect, which in the model is a predictor of spatial variation, is, crudely, the spatial driver of the seal encounter distribution after the fixed covariates impact has been considered. Examining it gives a a glimpse into what the patchy and variable local conditions driving behaviour looks like, quantifying the strength of the driver and giving an idea of just how variable these local conditions actually are.

Our results conform closely with what was expected based on previous knowledge of seal behaviour with the caveat that such encounter events have not previously been observed. It has long been known that the bank areas of the offshore are highly productive; for human fishing activity as for marine fauna. Those who have studied such marine mammals postulate that their foraging behaviour is directed by the imperative of maximizing feeding benefit while simultaneously minimizing energy expenditure (Lidgard et al., 2012). In this context a model which uses the seafloor topography, i.e. the banks, and penalizes distance from Sable Island, i.e. the seal's base, makes much ecological sense.

While these results are reasonable it is important to bear in mind that they arise from small samples. Our intent here is not to make definitive statements about seal behaviour but rather to demonstrate the suitability of these methods for the data being generated by the latest tagging technologies, data which will certainly be forthcoming. We place more weight upon the results in terms of what types conclusions are possible as opposed to evaluating the conclusions themselves.

Future work

Methods for fitting complex spatio-temporal models using INLA (Illian et al., 2012a) have been relatively recently developed. They offer a practical method with which to fit models to complex data sets. Information such as fisheries data on purported prey species, is one likely candidate for study. Recently, methods similar to those used here have been turned to the estimation of the density of fish species directly (Munoz et al., 2013). By replacing or augmenting the fixed covariates with a density estimate of prey species we may be able to directly model seal response to prey distribution. This would be a very nice refinement to our model. It would also be very useful to compare spatial distribution estimates for candidate prey species generated using the methods in (Munoz et al., 2013) to the latent variables thought to be proxies or to incorporate these prey species estimates directly. Highly topical, and of immediate fisheries and management interest would be such an analysis for Sable Island grey seals and atlantic cod (*Gadus morhua*). This article elected to consider encounter events in isolation but having the tracking data for the animals in conjunction with the encounters, offers yet another prospective avenue of exploration.

There are other examples of encounter-type data both within and outside of the OTN, yet still in the marine realm that we may consider in future. These include data that will be derived by wave gliders (they will have VMT type technology on board) that are currently being considered for deployment by OTN as well as whale-ship collision data.

4.7 Acknowledgements

We are grateful to Dr. Daniel Simpson for visiting Dalhousie University to offer an INLA workshop as well as much advice and insight. This work was supported by a research grant (NETGP 375118 08) from the Natural Sciences and Engineering Research Council of Canada (NSERC) for OTN.

4.8 Addendum

A note on priors: While in the original article it is stated that the default priors provided in the R-INLA package were used in the preparation of the analysis, the article does not explicitly state what the priors were. For clarity, and to assist the reader, they are listed here.

1. The size parameter for the Negative Binomial $\phi = log(n)$ was given a Gaussian prior with mean 0 and precision 0.01

2. The range parameter ρ, reparameterized in INLA as $log(\kappa)$ was given a Log Gamma prior (1, .01)

3. The variance parameter σ^2, reparameterized in INLA as $log(\tau)$ was given a Log Gamma prior (1, .00005)

4. β_{DIST} and β_{DEPTH}, the regression coefficients for the covariates, were given Gaussian priors each with mean 0 and precision 0.01

Chapter 5

Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the Northwest Atlantic Ocean

Chapter 5 is based upon a manuscript of which I was a co-author. While the submission of the manuscript was lead by Dr. Stephanie Boudreau, I was responsible for the proposed methodology, analysis and interpretation. What is presented herein are the details of these sections along with a synopsis of the eventual manuscript which conveys the results.

5.1 Abstract

The Northwest Atlantic Ocean has long been a productive area for commercial fishing, and overtaxing of its resources eventually led to a well documented, fisheries driven, ecosystem shift in the early 1990s. The ecosystem shift has seen Atlantic cod (*Gadus morhua*), once very abundant, be reduced to historically low levels, and fail to recover even after the implementation of protective measures. This failure of an overexploited species to recover has been accompanied by an increased abundance of other fishes. One species which has been increasing in abundance, at least since the 2000s, is the Atlantic halibut (*Hippoglossus hippoglossus*). This species is of commercial interest, but knowledge of halibut ecology is limited. The failure to recover of other groundfish species stocks collapsed by local overexploitation serves warning as to the dangers of exploitation without adequate knowledge of the effects of exploitation. In this Chapter we examine the application of a Bayesian hierarchical spatio-temporal model to juvenile halibut abundance over a time span of 36 years; during which time there were 3 distinct fisheries management regimes. We characterize the structure of juvenile halibut using three

parameters: persistence (similarity of spatial structure in time), connectivity (coherence of spatial structure over space) and spatial variance (variation across the seascape). We find that areas of high juvenile abundance persist throughout the 36 years, but that some of these areas are much reduced while others retain their high levels; in spite of increased abundance and landings throughout the study area. These persistent areas of high abundance overlap substantially with full and seasonal area fishery closures, which may create refuges from fishing. We estimate the connectivity to be 250km, far less than the distance assumed by Canadian fishery management units (2000km). This smaller scale of spatial coherence suggests a more complex structure than previously assumed, with attendant consequences for management of the fishery.

5.2 Introduction

The subject species for this study is the Atlantic halibut (*Hippoglossus hippoglossus*; halibut). The Atlantic halibut is a large flatfish with a wide range; found from the Canadian Arctic south to the United States' Atlantic region and encompasses French (St. Pierre and Miquelon), Icelandic and Danish (Greenland) territory (R. Froese and D. Pauly, eds., 2015; Trumble et al., 1993). Presently, Canadian halibut is managed as two stocks, a smaller stock in the the Gulf of St Lawrence, (DFO, 2015a), and a large single stock running from the Gulf of Maine southwest of Nova Scotia to the southeastern Newfoundland, (DFO, 2015b), referred to as the Scotian Shelf and Southern Grand Banks management unit.

Unregulated until 1988, management units and minimum legal catch lengths were then established (Neilson and Bowering, 1989; Trumble et al., 1993); the "one large stock" for the scotian Shelf/Grand Banks based primarily on tagging results which seemed to show intermixing of the Scotian Shelf with the Grand Banks, but not with the Gulf of St. Lawrence (Neilson et al., 1987), see Figure 5.1. More recently, tagging studies, (Bris et al., 2017), and age at length comparisons, (Bowering, 1986), support this distinction of the Gulf of St. Lawrence stock.

Long exploited, halibut was once abundant in US waters (Grasso, 2008), but became overfished there and the active fishery moved northwards to maintain catch rates; eventually reaching as far north as the Davis strait (Grasso, 2008; Trumble

et al., 1993). Decades of halibut fishing pressure resulted in large declines of halibut and the fishery through the 1990s (Trzcinski and Bowen, 2016). In American waters the halibut fishery has been under moratorium since 1999 (DofC, 1999), and the species remains a "Species of Concern" under the US *Endangered Species Act*, (DofC, 2012), a designation which includes halibut in Canadian waters under the provisions of that act ("a single stock"). Since the early 2000s however, halibut landings have been steadily increasing, (DFO, 2015a,b; Trzcinski and Bowen, 2016), so much so that by 2013 the Scotian Shelf/Grand Banks fishery was able to achieve certification by the Marine Stewardship Council (Martell et al., 2013). Indeed, juvenile halibut abundance in Canadian waters has been estimated to be 5 times higher than US despite ample suitable waters in US territory, (Shackell et al., 2016). These circumstances suggest that halibut population structure exists at smaller, or more local, scale than previously assumed, and that US halibut stocks have never fully recovered from historic overfishing (Shackell et al., 2016). Other recent studies support the hypothesis that halibut exists as a series of local sub-populations (den Heyer et al., 2012; Kanwit, 2007; Seitz et al., 2016, 2017).

The historical pattern of the halibut fishery, depleting the southernmost stock and then moving progressively further northward so as to maintain catch rates is suggestive of a pattern of sequential local overfishing of adjacent sub-populations (Grasso, 2008; Maury and Gascuel, 2001). Local overfishing of sub-populations can lead to localized extinction by collapsing small, weakly connected sub-populations (Frank and Brickman, 2000; Kerr et al., 2010b; Reich and Dealteris, 2009; Sterner, 2007). Examples of such local overfishing, (reviewed in (Cianelli et al., 2013; Safina et al., 2005)), include the North Sea herring (*Clupea harrengus*), (Payne, 2010; Ruzzante et al., 2006), and the Northwest Atlantic's Atlantic cod, (*Gadus morhua*), (Hutchings, 1996; Rose et al., 2000). The Atlantic cod collapse is well studied and is a singular example; high density areas of cod, ultimately discrete spawning grounds, were targeted aggressively by the fishery, leading to the spatial erosion of the Atlantic cod meta-population and disappearance of sub-populations (Hu and Wroblewski, 2009; Smedbol and Wroblewski, 2002). Atlantic cod has not recovered from the basin wide collapse of the early 1990s caused by this local overfishing (Shelton et al., 2006), and it is now understood that spatial scale of the

fishery management was much larger than that of the exploited subpopulations, allowing the localized over fishing to occur (Carson et al., 2017; Roney et al., 2016).

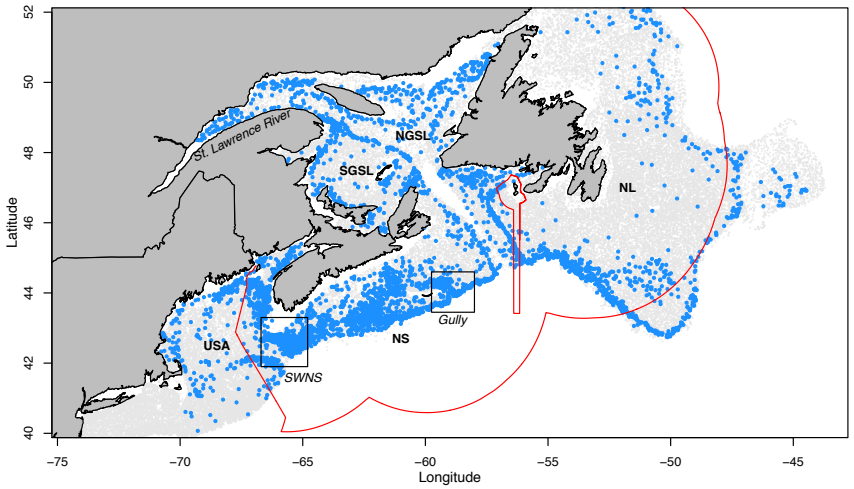

Figure 5.1: **Juvenile halibut presence** Raw data from ecosystem trawl surveys indicated in blue, absence in gray and the exclusive economic zone (EEZ) in red. St. Pierre and Miquelon French territory is denoted by inset EEZ. Regions of analysis are marked, Newfoundland and Labrador (NL), Nova Scotia (the Scotian Shelf and Gulf of Maine; NS), Northern Gulf of St. Lawrence (NGSL), Southern Gulf of St. Lawrence (SGSL), and the United States of America (the survey set from NMFS includes tows into Canadian waters; USA). Areas of interest for juvenile halibut abundance are also marked: the Gully, Southwest Nova Scotia (SWNS), and are denoted generally with a box. The St. Lawrence River estuary is also identified, and while there is halibut present, ultimately it was excluded from the formal analysis.

Spatial Analyses

The need to understand the spatial structure of an exploited population was recognized even in the 1800s, but since the 1960s fisheries science has focussed upon

quantitative methods of assessing the fishable biomass (Stephenson, 2002), with the underlying assumption that fish removed from any local area will be replaced by in-migration from surrounding areas (Svedäng et al., 2001). While it is possible for this sort of recolonization to occur (Corten, 2013), local depletion does happen, suggesting that local recolonization is not always a simple density dependence process. Possible explanations for the lack of recolonization of suitable habitat are philopatric feeding behaviour (Svedäng et al., 2001; Svedäng and Svenson, 2006), changes in population structure or demographic rates (Payne, 2010), or simply that the rate of recolonization cannot counter excessive fishing pressure (Shackell et al., 2005). The collapse and non-recovery of the Atlantic cod has spurred renewed interest in spatial stock structure (Cadrin et al., 2014; Carson et al., 2017) as it is now recognized that a commercial fishery can erode local concentrations if overall stock structure is ignored.

Research Questions

The Northwest Atlantic Continental Shelf has undergone a well documented ecosystem shift in recent decades; the shift has seen a change from an ecosystem dominated by large bodied groundfish to one abundant in invertebrates and forage fish (Frank et al., 2005; Shackell and Frank, 2007; Shackell et al., 2010; Worm and Myers, 2003). In marked contrast to the other ground fishes, Atlantic halibut has been experiencing population growth, supported by high recruitment in Canadian waters since the early 2000s (DFO, 2015a,b; Trzcinski and Bowen, 2016). We build upon evidence which suggests that halibut are not habitat limited (Shackell et al., 2016); most tagging studies suggest local residency for halibut, (citations above), and postulate that the lack of halibut recovery in US waters is interpretable as a change in local structure brought about by historical, local, overfishing (Grasso, 2008).

In this paper we use a Bayesian hierarchical spatio-temporal modelling approach (Carson and Flemming, 2014; Cosandey-Godin et al., 2015) to model the spatial structure of juvenile halibut abundance, as an index of halibut fisheries recruitment. Fisheries independent groundfish research vessel trawl survey data from

Canada and the USA are used to explore the spatial and temporal patterns of juvenile halibut abundance in the Northwest Atlantic over the last 36 years (1978-2013), a period which encompasses 3 distinct Canadian fisheries management regimes: (1), 1978-1989, post implementation of the Exclusive Economic Zone (EEZ), (2), 1990-2003, a period of moratoria and low groundfish abundance, and, (3), 2004-2013, a period of high halibut recruitment. The model, which contains no *a priori* assumptions about management units or international borders, identifies areas of persistently high relative abundance, which varied among regimes. We argue that the protection of areas of persistent high abundance contributed to the recovery of this stock, or stocks, and that sustainable management must consider stock structure.

5.3 Methods

Data

Data were combined from 27 fisheries independent research trawl surveys conducted by the National Marine Fish Service (NMFS, USA), and the Department of Fisheries and Oceans (DFO, Canada). These surveys are stratified random sampling designs, and sample on the on the continental shelf off eastern North America, see Figure 5.1. The resulting time series contained trawls using seven different types of trawl gear, capturing primarily juvenile halibut (that is, body size $<$ 80cm). Various surveys covered different seasons, different years and different geographic regions; Newfoundland and Labrador (NL), Nova Scotia (the Scotian Shelf and Gulf of maine, NS), the Northern and Southern Gulf of St. Lawrence (NGSL, SGSL), and the American and Canadian waters in the Gulf of Maine and Bay of Fundy sampled by the NMFS. In total there were 75149 survey trawls taken between 40^o N and 52.25^o N, during the research time period 1978-2013. Of the 75149 trawls, 4509 (6%), caught juvenile Atlantic halibut, see Table 5.1. Date, location (Latitude, Longitude), bottom temperature, depth, abundance (count) and biomass (weight) were recorded for each trawl. Abundance was standardized within each survey to account for differences in set duration and distance trawled (OBIS, 2014). Annual estimates of stratified mean abundance were used to show

regional abundance trends in time, see Figure 5.3.

Region	Total Sets	Sets with halibut	Percent with halibut	Start Year	End Year
Newfoundland(NL)	34170	1142	3.34	1978	2013
Northern Gulf of St. Lawrence(NGSL)	3487	385	11.04	1990	2013
Nova Scotia(NS)	12965	2126	16.4	1978	2013
Southern Gulf of St. Lawrence(SGSL)	6802(6741)	494(433)	7.26(6.42)	1978	2013
United States of America	17725	362	2.04	1978	2013
Summary	75149(75088)	4509(4448)	6.00(5.92)		

Table 5.1: **Data Summary.** Region (ordered alphabetically), total sets, sets with halibut, percent with halibut, first year, last year of time series. The numbers in brackets represent the totals after removing 61 observations in the Gulf of St. Lawrence estuary.

Approach: The Spatio-Temporal Model

To explore abundance and distribution of juvenile Atlantic halibut we employ a Bayesian hierarchical spatio-temporal model extending the methods in Carson et al. (2017). The model encompasses the 1978-2013 time series, which we divide into three time steps defined by changes in fisheries management regime:

1. 1978-1989 - After the establishment of the EEZ in 1977/8 the 'foreign fleet' was absent from Canadian waters and the halibut, and most groundfish stocks, rebounded (Horsman and Shackell, 2009). The halibut fishery was regulated in 1988, after which time management units and minimum catchable size were defined by DFO. Research survey trawls covered all areas except NGSL, and SGSL catches were very low (DFO, 2015b), n= 12 years, 1550 captures of juvenile halibut in 15 separate surveys.

2. 1990-2003 - The NGSL survey commenced in 1990. Moratoria were declared on the collapsed Newfoundland Atlantic cod fishery in1992, and on Gulf of St. Lawrence cod in 1993, and on Eastern Scotian shelf groundfish in 1994, n= 14 years, 1298 captures of juvenile halibut in 16 separate surveys.

3. 2004-2013 - This decade marked a period of high recruitment and Atlantic halibut population recovery, n= 10 years, 1600 captures of juvenile halibut in 15 separate surveys.

Models are fitted using INLA, for Integrated Nested Laplace Approximation, (Rue et al., 2009b), which is available as an R package (R Core Team, 2013) called R-INLA (**?**). R-INLA has been used previously in the marine environment to model

fisheries bycatch, animal abundance and animal movement, (Carson and Flemming, 2014; Cosandey-Godin et al., 2015; Quiroz et al., 2014). R-INLA uses a Stochastic Partial Differential Equations (SPDE) approach to model spatial dependence on a triangulated mesh. The mesh is built from the data and the domain, see Figure 5.2, and ultimately regularizes the space to visualize spatio-temporal structure.

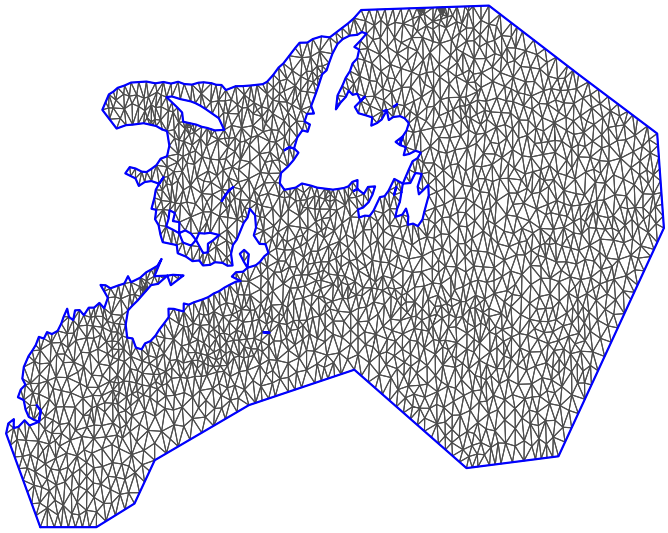

Figure 5.2: **Triangulated mesh** Constrained refined Delaunay triangulation mesh. The smaller triangles of the grid are where there are more data points. It is formed of triangles, the vertices are called nodes (here there are 2120).

For spatial analysis of large data sets SPDE models are efficiently fit; they represent the continuous response (a Gaussian Random Field, GRF) using a discretely indexed spatial process, (a Gauss Markov Random Field, GMRF). Spatial dependence is modelled on the mesh constructed as a above, and parameters for the covariance function (which is Matérn) are estimated (Lindgren et al., 2011; Rue

and Held, 2005). Following Carson et al. (2017), we estimated connectivity (coherence across space), spatial variance (variation across the seascape), and persistence (coherence in time) for juvenile halibut in the modelled domain.

The response variable of interest is juvenile atlantic halibut abundance (number) at each location in each time period. All true zero values in the original data were removed from the data set, but were represented in the spatio-temporal model as non-positives in the mesh (see Figure 5.2). As non-positive values are less computationally demanding, this enabled us to run the analysis.

The model has the following general form:

$$E[log(Y_{s,t})] = \eta(s,t) = \xi(s,t) + \Sigma_{j=1}^{n} f_j\{c_j(s,t)\}, \tag{5.1}$$

where $E(Y_{s,t})$ is the mean of the expected response, Y, at location s at time t, $\xi(s,t)$ is the spatio-temporal latent GMRF (a random effect), while the $f_j\{c_j(s,t)$ are smoothed functions of the covariates, where j refers to jth of a total $n = 2$ covariates (depth and temperature). The mean of the response variable, $E(Y_{s,t})$, is mapped by a canonical link function (i.e., log link) to a linear predictor, $\eta(s,t)$, by the Generalized Additive Model framework (Hastie and Tibshirani, 1990). The spatio-temporal latent field, $\xi(s,t)$, represents the cumulative effect of all unmeasured latent factors. The characteristics of this spatio-temporal random effect comprise the spatial and temporal covariance structure of the model (Rue et al., 2009b).

Once we have defined the domain and built the mesh we are able to run a model; it then becomes necessary to determine an appropriate statistical distribution for the response errors, select important covariates and establish the importance of the latent variable (the GMRF) to model performance. The error distributions for the response explored were the Poisson, the negative binomial and the Gaussian. The temporal structure for the latent GMRF was also examined, with temporally invariant, autocorrelated or independent structures examined, see Cosandey-Godin et al. (2015).

Goodness of fit

The goodness of fit for the various candidate models was compared using Bayesian diagnostic criteria; the Deviance Information Criterion (DIC) (Spiegelhalter et al.,

2002) and the Logarithm of the Pseudo-Marginal Likelihood (LPML) (Geisser and Eddy, 1979). Smaller values of DIC and larger values of LPML indicate better goodness of fit. Each of these diagnostic criteria are readily calculated by R-INLA.

Connectivity

Connectivity, or ρ, is the scale parameter in the Matérn covariance function, and is conventionally interpreted as the distance at which covariance decays to ≈ 0.13 (Cameletti et al., 2011, 2013). A large value for ρ means that the covariance decays slowly in space, (abundances at points widely separated are related). Units of this parameter are in degrees, which we convert to km for interpretability.

Spatial Variance

Spatial variance, σ^2, is an index of the differences in amplitude which exist across the seascape. This is scaled to the response, (i.e. log juvenile halibut abundance), and a large value indicates a large amplitude in the overall field.

Persistence

A temporally invariant structure for the latent variable (the GMRF) requires only 2 parameters, ρ and σ. If there is structure in time a third parameter, a, a first order autoregressive coefficient, or AR(1) term may be added. This term will have values from -1 to 1. It will quantify the degree of dependence between time periods in the model, (here $t = 3$, for the 3 management regimes). If the abundance at location s at time t is dependent on the abundance at s at time $t - 1$, that is, if locations of high/low abundance consistently remain areas of high/low abundance across time periods, a will be positive. A value for a close to 1 would indicate a strong relationship of this type. We interpret this parameter as persistence, see Carson et al. (2017).

When ρ and σ^2 are reported together as parameters of the model they describe the multivariate normal distribution of the mean of the response, after accounting for variables explicitly included. Our eventual model, incorporating 3 time periods, (fisheries management regimes), has one ρ, one σ^2, and an a. Bayesian credibility intervals were also examined and reported. To examine the distribution of

abundance of juvenile halibut these parameters were combined and the resultant random fields plotted as maps, see Figure 5.4.

5.4 Results

Time Series Abundance Indices

Halibut were often present in the GSL trawls, and most often along the edge of the continental shelf, see Figure 5.1. Stratified mean abundance of juvenile halibut decreased through the first two of our fisheries management regimes, 1989-2003,but have been increasing thereafter, Figure 5.3. In general, halibut have low catchability in trawl surveys; the stratified mean abundance ranged from 0.05 to 0.25 halibut per trawl. Looking only at positive sets, (trawls in which at least 1 halibut was captured), the mean ranged from 1.19 to 2.05, see Figure 5.3 A.
Stratified mean bottom temperatures were generally higher, and less variable during the second regime; however temperatures throughout the study have remained between 2 C^o and 5 C^o. Looking only at positive sets, mean temperatures were warmer, ranging between 3.4 C^o and 6.75 C^o, Figure 5.3 B.
The stratified mean depths for all survey sample sets ranged from 140 to 210 meters, with the deeper samplings occurring most frequently and consistently during the second management regime. In the positive sets, halibut were observed in sets with mean depths from 129.2 to 196.2 meters, Figure 5.3 C. Sets in Nova Scotian waters had the highest percentage of positives at 16%, while those in US waters had the lowest at 2%, see Table 5.1

Model Selection

We fit models using several different alternatives for likelihood family, temporal structure, covariate and inclusion of random effect. The best performing models included autoregressive spatio-temporal random effects, AR(1), which yielded a persistence parameter, a, see Tables 5.2, 5.3, implying that the value of the latent field for each time period. $\eta(s,t)$, was correlated with, (i.e. a function of), the previous one, $\eta(s,t-1)$. The DIC and CPO diagnostic criteria for models using Poisson error distributions were consistently better than those using any other error

distribution considered. The covariates were largely insignificant, likely because we focussed on only positive responses. Accordingly, we chose as the best performing model the one with the form $\eta(s,t) = \xi(s,t) +$ Bottom Temperature. A smoothing function was not applied to Temperature.

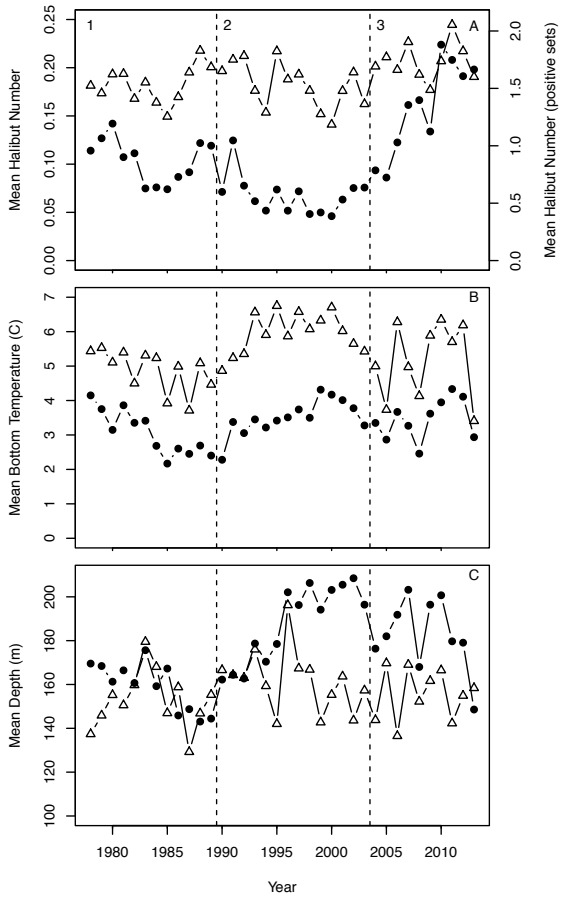

Figure 5.3: **Time series trends** Stratified mean trends in juvenile Atlantic halibut from the trawl surveys by year using all data, presence ≥ 0 (closed circles), and positive sets only (open triangles). A: the stratified mean number of halibut from all sets on the left y-axis, and calculated for the positive sets only on the right-hand y-axis, B: stratified mean bottom temperature (^{o}C), and, C: stratified mean depth (m). Time periods are: (1) 1978-1989, (2) 1990-2003, and (3) 2004-2013.

Visual examination of areas of high abundance over the three management regimes indicate that juvenile habitat distributions have changed over time, see Figure 5.4.

Covariates	DIC	CPO
$\xi(s,t)$ + Temperature	**13433.31**	**-6735.08**
$\xi(s,t)$ + Temperature + Depth	13435.71	**-6735.08**
$\xi(s,t)$	13446.01	-6741.99
$\xi(s,t)$ + Depth	13447.82	-6742.17

Table 5.2: **Covariates, DIC and CPO Values** Selection criteria for various candidate models with an AR(1) spatio-temporal covariance structure and a Poisson likelihood. Models are listed in order of lowest to highest DIC. Bolded values indicate the lowest DIC and highest CPO an are the better fitting models.

Considering the parameter values, connectivity ($\rho = 2.25^o$) shows that significant correlations in abundance are seen at distances up to 250km. The spatial variance was also quite low, $\sigma^2 = 0.12$, indicating a relatively flat field. Persistence ($a = 0.77$) was quite high, indicating that juvenile halibut were present in about the same locations in each time period and that abundance is correlated in time, Table 5.3.

Parameter	Mean	SD	0.025	0.5	0.975
Connectivity(ρ)	2.25	0.44	1.55	2.19	3.26
Spatial Variance(σ^2)	0.12	0.02	0.09	0.12	0.16
Persistence(a)	0.77	0.06	0.64	0.78	0.88

Table 5.3: **Parameter Posterior Estimates** Mean, standard deviation and 95% credible interval for the penultimate model, with AR(1) spatio-temporal structure, Poisson likelihood, a latent field and a bottom temperature covariate.

During the first regime, 1978-1989, there appear to be five concentrations of high abundance on the Scotian Shelf and in Southeastern Newfoundland, following the edge of the continental shelf, with low abundance in the GSL, Figure 5.4 A. Under the second regime, 1990-2003, there is a general increase in the GSL and a general decrease Southeast of Newfoundland. High abundance is seen in two areas of the Scotian shelf; the Gully, and off Southwest Nova Scotia (SWNS), Figure 5.4 B. In the most recent period, 2004-2013, juvenile halibut abundance has increased and the high abundance seen in SWNS and at the Gully has persisted. Halibut abundance has again increased in the GSL, but has remained low off Southeast Newfoundland, Figure 5.4 C.

Spatio-Temporal Model Results

This analysis is a broad look at at the spatio-temporal trends in juvenile halibut abundance in the NW Atlantic. Two areas of persistent abundance were identified, both on the Scotian Shelf: (1) South West of Nova Scotia, and (2) the Gully. Less persistent areas of high abundance were also seen in the GSL and SE of NL, although these transient results may be confounded with differences in surveys.

5.5 Discussion

Here we demonstrate that halibut exhibit a spatial structure in the Northwest Atlantic at a smaller scale than the current halibut stock management units. Our analysis showing statistically independent areas of persistent juvenile halibut abundance provides important evidence that halibut spatial structure is more complex than previously identified, and has varied spatially over the last 35 years. Two areas of high abundance persisted on the Scotian Shelf throughout the study period, SWNS and the Gully. Despite the widespread increase in halibut recruitment during the last decade, previously identified areas of high juvenile abundance have not re-established themselves off Southeastern Newfoundland. As a measure of structure, connectivity of juvenile halibut abundance is an order of magnitude smaller than the scale of stock structure assumed by the stock management units ($\approx$2000 kms; DFO (2015a)). Halibut have a long history of being removed as bycatch in cod fisheries, (Grasso, 2008), and persistent patches of abundance along the edge of the shelf may identify spatial refuges from cod directed fishing. Coinciding with the reduction of cod trawling fisheries in the early 1990s, and the introduction of minimum legal size limits for halibut landings, halibut have increased since the mid 2000s and catches in the Canadian trawl surveys have been well above the long term average, (DFO, 2015a,b); see also Figures 5.3, 5.4.

While significant knowledge is required to identify discreet population units; stock assessments, sustainable food classifications (i.e. Marine Stewardship Council), and vulnerable species classifications depend on knowing demographic rates for identified stocks or sub-populations and the impacts of harvesting thereupon. Sustainable management and facilitated recovery of overexploited stocks is further

complicated when fine scale distinctions are observed in local populations such as the Skagerrak cod (*Gadus morhua l.*), (Olsen et al., 2008), and yellowfin sole (*Limanda aspera*) in the Bering Sea, (Bartolino et al., 2012). In the Northeastern Atlantic, evidence of genetic differentiation supports the contention that there are local subpopulations of Atlantic halibut, (Foss et al., 1998; Haug and Fevolden, 1986; Mork and Haug, 1983). In the Northwest Atlantic, Reid et al. (2005) found no evidence of genetic differentiation supporting the existence of subpopulations but acknowledged that, in the absence of spawning information and with a limited sample size, the power of the study to detect such distinctions was not ideal. Genetically distinguishable spawning populations that mix after spawning may not be easily assigned to a location (Reid et al., 2005). More research will be required to determine the genetic structure of the Atlantic halibut population.

Even if genetic distinctions cannot be made between areas of high abundance to establish them as genetic subpopulations, these high abundance areas are vulnerable to overfishing if connectivity or mixing between them us low and fishing pressure is high, (Shackell et al., 2005). Stock collapses in harvested species have been observed on the Scotian Shelf, (Reuchlin-Hugenholtz et al., 2015), and off Newfoundland,(Hutchings, 1996). Notably, one of the areas of persistent high abundance, (SWNS), is adjacent to US waters, where the absence of recovery since the 1800s is evidence of a structural change wrought by local overfishing, (Seitz et al., 2016; Shackell et al., 2016).

Tagging of halibut off Eastern North America provides evidence of exchange between Canadian and US waters with almost 30% of recaptures of halibut tagged in New England occurring in Canadian waters, (Kanwit, 2007). In general however, studies of halibut suggest that halibut are resident, or return to, particular areas, with the majority of recaptures occurring within 200km of the point of release, and a small proportion travelling large distance across the management unit and beyond, (den Heyer et al., 2012; Stobo et al., 1988). In the GSL a study employing satellite tags over 2013-2015 found that tag "pop-off" locations were from 55-423 km from tagging location and inferred local residency from this evidence, (Murphy et al., 2017). Of note, discrete spawning units for Pacific halibut have been inferred from satellite tagging, (Seitz et al., 2017).

The statistical independence of locations separated by more than $\approx$ 250 km implies a more complex population structure than previously supposed, but does not explicitly incorporate the two distinct stock management approaches present in the GSL and the Scotian Shelf-Southern Grand Banks. The analysis does however demonstrate changing abundance under both management approaches, in all three regime periods, (DFO, 2015a,b; Trzcinski et al., 2011), see also Figure 5.4. During the 1980s and 1990s, the Northwest Atlantic ecosystem underwent a shift, with the predominant life from changing from large bodied fish to invertebrates; the shift being attributed to overexploitation of the fish stocks, (Frank et al., 2005). The observed decline in both the abundance, (Frank et al., 2005; Shackell and Frank, 2007), and size, (Shackell et al., 2010), of groundfish in these areas was followed by a large increase in the populations of benthic decapods and other prey species, likely due to the release from predation pressure, (Boudreau and Worm, 2010; Stenek et al., 2004; Worm and Myers, 2003). Recently, there have been some signs of groundfish recovery in haddock (*Melanogrammus aeglefinus*), (DFO, 2012), cod, (Cadigan, 2016; Rose and Rowe, 2015), and halibut, (DFO, 2015a,b; Trzcinski and Bowen, 2016).

We postulate that persistent areas of juvenile halibut abundance on the Scotian Shelf have, of late, been afforded protection from commercial fishing and trawl gear; and that this protection has contributed to the rebounding of this stock. One area of persistent high juvenile abundance overlaps with a seasonal groundfish spawning closure, while the other occurs in the Gully, a deep water canyon and marine protected area adjacent to an area closed to protect juvenile haddock, (O'Boyle, 2011). We note that some longline fishing is permitted in the outer areas of the Gully, (DFO, 2008), but that this fishery is pursued with a minimum catch size of 81cm. Notably, the Gully persistent area occurs on the Eastern Scotian Shelf, where commercial groundfish fisheries have been drastically curtailed, (O'Boyle, 2011), and that this contributes additional protection to juvenile halibut from capture as bycatch in other fisheries. Unlike cod, (Hutchings, 1996), halibut do not spawn in groups, (Trumble et al., 1993), which too serves to protect against localized overfishing.

Persistent locations of high juvenile halibut abundance have remained in place

through 35 years on the Scotian Shelf, suggesting that these areas are core high abundance refuges and that density dependent habitat selection is occurring. As this preferred habitat reaches resource limits, halibut may begin to occupy less ideal habitat, (Fisher and Frank, 2004; Gaston, 2003; Gaston et al., 2000). Density dependent habitat selection is generally associated with preferred habitats rich in prey. Juvenile halibut up to 30cm in length feed almost exclusively on invertebrates, those from 30cm to 80cm in length feed both on invertebrates and fish, while halibut larger than 80cm feed almost exclusively on fish, (Kohler, 1967). It has yet to be examined how juvenile halibut abundance and distribution may have co-varied with the abundance and distribution of their preferred pre species as the Northwest Atlantic ecosystem shift transpired. Building on the analyses presented here, a closer examination of regions with persistent high juvenile halibut abundance or historically high juvenile abundance, such as Southeast of Newfoundland, could find evidence of local overfishing, divergent demographic rates, or other fine scale dynamics of prey species.

5.6 Conclusion

Using a powerful statistical tool and more than 3 decades of groundfish survey trawl data from Cape Cod to the Grand Banks, we have identified areas of persistently high juvenile halibut abundance that have remained so through three different fisheries regimes. Other high abundance areas have been diminished, and as yet, have not re-established abundance. As more research will be required to define subpopulations, and to prevent serial local overfishing in the future, we propose a pragmatic interim approach; one that focusses on protecting important patches where abundance is persistently high, indicating preferred habitat rich in prey or sheltered from predators. These areas may serve as refuges from fishing pressure, both naturally, such as a deep water canyon like the Gully, or intentionally, as part of a fisheries management strategy protecting juveniles. Atlantic halibut are experiencing population growth which is quite unique for a commercially valuable species in the Northwest Atlantic. There exists an opportunity to manage the stock in an precautionary way, rather than the amelioration that has become the

norm. The history of serial local depletion of the Atlantic halibut particularly emphasizes the importance of stock structure to any management effort. Given the consequences of localized overfishing, local concentrations of halibut need protection until it becomes clear to what extent they represent distinct subpopulations or a connected common stock.

While this work on Atlantic halibut uses similar methodology to the two papers which will follow it is important to point out the very basic difference in purpose. In this paper the purpose was to look at the actual distribution of the halibut; where they were abundant, how big were the aggregations of abundance, and were there identifiable sub-groups within the larger halibut population. Having described the nature of the halibut distribution we compare what we learned to what we know of how the halibut are managed and examine whether the scale of the problem and the scale of the response are in keeping. This is very different from what follows; in the next paper the failure of the management is a given. What we focus upon is not characterizing the size and shape of the cod distribution, but we try to capture the effect of the exploitation on the population by examining how the descriptive parameters move. The parameter set used to describe the halibut can also describe the cod, but changes in the parameters reveal changes in the population of the cod. One can see the de-structuring (destruction) of the cod population as the crisis develops. The next paper is an effort to place these species and models into their spots in a larger scale. By attempting to develop a spatio-temporal model for both predator and prey we begin to model the interactions between species, and some of the pressures upon them, to understand something of why the distributions appear as they do. This effort to see the pressure exerted by one species on another, the ecological footprint, is the main point of Chapter 7.

5.7 Addendum

This analysis used the default priors provided in the R-INLA package. The article does not explicitly state what the priors were, so for clarity, and to assist the reader, they are listed here.

1. The ar(1) persistence parameter a was given a Gaussian prior with mean 0 and precision 1

2. The range parameter ρ, reparameterized in INLA as $log(\kappa)$ was given a Log Gamma prior (1, .01)

3. The variance parameter σ^2, reparameterized in INLA as $log(\tau)$ was given a Log Gamma prior (1, .00005)

4. β_{Temp}, the regression coefficient for the covariate, was given a Gaussian prior with mean 0 and precision 0.01

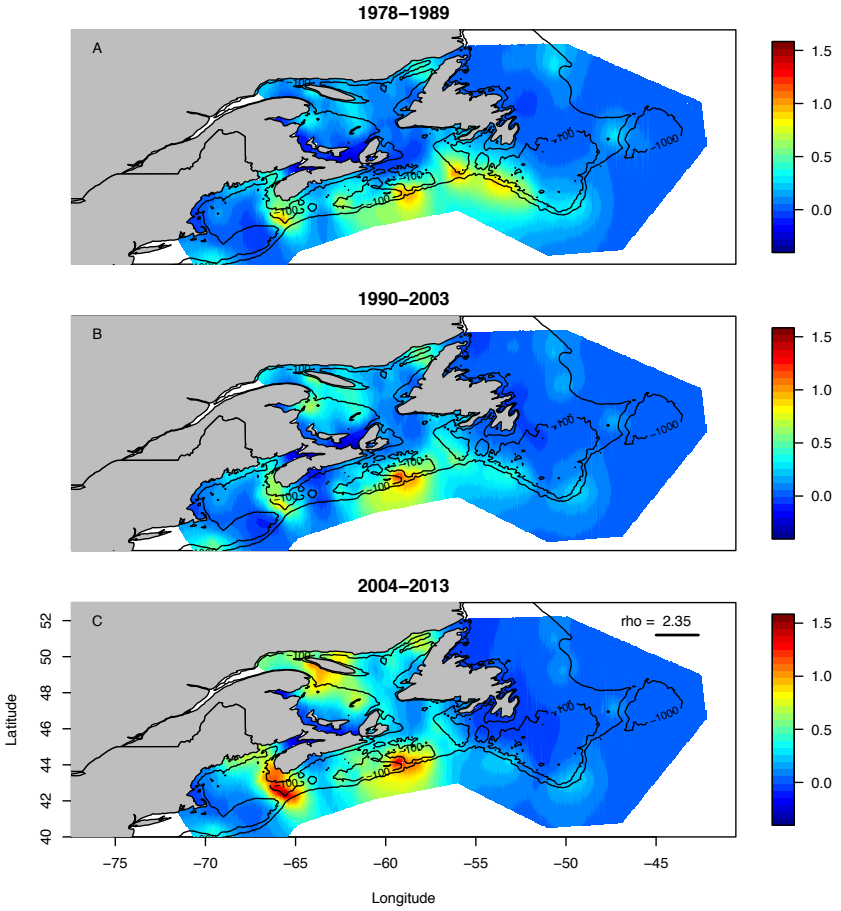

Figure 5.4: **Random fields of juvenile halibut abundance** the penultimate model by fisheries management time period, the model parameter connectivity, ρ (rho) is drawn and printed on the bottom map, units are degrees Latitude. The colour scale illustrates the random latent field (log abundance) and all three panels are the same scale. The log abundance for panel A: time period (1) 1978-1989 ranged from -0.30 to 1.07 with a mean of 0.16, for B: (2) 1990 to 2003; ranged from -0.37 to 1.27 with a mean of 0.15, and C: (3) 2004-2013; ranged from -0.25 to 1.55 with a mean of 0.21. Bathymetric contours illustrate the 100 m and 1000 m depths representing the continental shelf and larger banks.

Chapter 6

Local overfishing may be avoided by examining parameters of a spatio-temporal model

6.1 Abstract

Spatial erosion of stock structure through local overfishing can lead to stock collapse because fish often prefer certain locations, and fisheries tend to focus on those locations. Fishery managers are challenged to maintain the integrity of the entire stock and require scientific approaches that provide them with sound advice. Here we propose a Bayesian hierarchical spatio-temporal modelling framework for fish abundance data to estimate key parameters that define spatial stock structure: persistence (similarity of spatial structure over time), connectivity (coherence of temporal pattern over space), and spatial variance (variation across the seascape). The consideration of these spatial parameters in the stock assessment process can help identify the erosion of structure and assist in preventing local overfishing. We use Atlantic cod (Gadus morhua) in eastern Canada as a case study an examine the behaviour of these parameters from the height of the fishery through its collapse. We identify clear signals in parameter behaviour under circumstances of destructive stock erosion as well as for recovery of spatial structure even when combined with a non-recovery in abundance. Further, our model reveals the spatial pattern of areas of high and low density persists over the 41 years of available data and identifies the remnant patches. Models of this sort are crucial to recovery plans if we are to identify and protect remaining sources of recolonization for Atlantic cod. Our method is immediately applicable to other exploited species.

6.2 Introduction

Fish are not randomly distributed across a seascape. They occur at higher concentrations in habitat where resources can support higher densities. The geographic

distribution of fish may either expand in proportion to an increase in abundance, or it may exhibit a density-dependent habitat selection response, in which fish preferentially occupy a preferred area until it reaches maximum density, at which point they disperse into more marginal areas (Fretwell and Lucas, 1970; Gaston and Blackburn, 2000; Gaston et al., 2000). Under either conception of the range expansion process, core areas are occupied at low population size (Gaston and Blackburn, 2000). The series of 'core' or high-density areas across the seascape can be considered a meta-population, a series of sub-populations that are connected to a greater or lesser degree, and where geographically closer, sub-populations are relatively more connected (Moilanen and Hanski, 1998).

Fishing boats naturally tend to focus on high density core areas, in order to minimize effort and maximize catch. For fish species that select or occupy habitat based on density, any core area that gets depleted by fishing will fill up with fish from surrounding areas. Fishing can continue, until there are insufficiently many fish to move in and maintain density, the area then becomes locally depleted. This process has been referred to as local overfishing (Maury and Gascuel, 2001) and if it happens often, the species experiences spatial erosion across the seascape. The consequences of local overfishing can lead to stock collapse (Frank and Brickman, 2000; Hauser and Carvalho, 2008; Kerr et al., 2010a; Maury and Gascuel, 2001). Sufficient evidence of spatial erosion and non-recovery has accrued (Cianelli et al., 2013; Corten, 2013; Ruzzante et al., 2006; Safina et al., 2005) that fishery managers are becoming interested in how to maintain the integrity of a stock's spatial pattern. Here we propose a Bayesian hierarchical spatio-temporal model that involves 3 key parameters to describe spatial structure: persistence (similarity of spatial pattern over time), connectivity (degree of coherent structure present) and spatial variance (variation across the seascape). Our goal is to show that these parameters can be well estimated to provide a useful picture of stock structure on both long-term and annual scales. The long-term model parameter estimates (persistence, connectivity, spatial variance) can be interpreted as the climatological, or average spatial structure. We examine the behaviour of these parameters from the height of the fishery through the collapse. Further, our model framework can be used on an annual scale to monitor and potentially maintain a stock's spatial

structure. We use Atlantic cod (Gadus morhua) as a case study, a well known fish species with a long history as a commercially valuable and widely consumed food fish.

The fishery for Atlantic cod in eastern Canadian waters has a history dating back several centuries but has perhaps been most widely recognized in recent years for the closure of the fishery due to the collapse of the exploited stocks that occurred in the early 1990s (Fu et al., 2001; Hutchings and Myers, 1994; Myers et al., 1997; Rose et al., 2000; Walters and Maguire, 1996). One feature of these stock collapses was the very late realization that the stocks were in peril; catches, and inferred stock levels, remained high right up until the seemingly sudden and precipitous collapse (Rose and Dutka, 1999). Stock assessments may have missed the signs of the impending collapse because the distribution of cod throughout the northwest Atlantic can be density dependent (Hutchings, 1996; Shackell et al., 2005; Swain and Wade, 1993a; Tamdrari et al., 2010) making cod susceptible to being locally overfished. Specifically, it is suggested that there exists a region of prime habitat or 'core range' and that this prime range is used preferentially and that the stock's total range extends out from the core into less preferred areas under population pressure in what is termed an Occupancy-Abundance relationship; range is positively related to abundance (Gaston and Blackburn, 2000; Gaston et al., 2000), and the species is said to display Density Dependent Habitat Selection (DDHS) (Fisher and Frank, 2004; Swain and Wade, 1993b). The actual density of the species of interest in the prime habitat may remain relatively constant even as the total abundance reduces if there is in-migration from the less preferred range. In the case of the Northern cod, the fishing industry's standardized measure, catch per unit effort (CPUE) remained high, but in reality these cod were becoming spatially concentrated (Hutchings, 1996; Hutchings and Myers, 1994; Rose and Dutka, 1999) as abundance decreased. The effort expended to obtain a profitable trawl remained fairly constant but the area of ocean where these profitable trawls were being found was becoming smaller and smaller (Rose et al., 2000). Our study contributes to the growing effort to develop spatial indices that will help to maintain stock spatial integrity (Reuchlin-Hugenholtz et al., 2015).

6.3 Data

The Department of Fisheries and Oceans Canada (DFO hereafter) Maritimes Region has conducted an annual summer ground fish research trawl study each year since 1970. Originally designed to measure distribution and abundance of commercial species these data also incorporate information on non-commercial species. Focussed upon the Scotian Shelf the DFO survey utilizes a stratified sampling plan using the three relevant North American Fisheries Organization (NAFO) zones, 4V, 4W, and 4X demarcating the Scotian Shelf. Figure 4.1 presents the general geographical location and shows the boundaries of these NAFO areas and their associated sub-divisions, referred to as sub-zones. Each of the three NAFO zones has sampling effort (the number of sample trawls) proportional to their area. The catch is sorted by species, weighed and measured for individual weight, maturity status and age. The data have been summarized in various reports (Horsman and Shackell, 2009; Ricard and Shackell, 2013; Smith et al., 2013), stored, and are publicly available in the Ocean Biogeographic Information System (OBIS) (OBIS, 2014). OBIS is the DFO - Maritimes Region database for ground fish research trawl surveys and includes information on some 263 distinct species found on the Scotian shelf. It includes descriptive data for each cruise or mission resulting in about 200 fishing sets per year. For each set there exists trawl information: date, latitude, longitude, distance towed(km), as well as physical/water characteristics at the location and depth of the trawl; temperature(C), salinity(ppm), nitrate(ppm), phosphate(ppm) and silicate(ppm). For all species captured: genus, species, common name, total weight(kg) and count, (and total weight and count standardized by distance towed) are recorded.

Here we consider a single species, Atlantic cod, but stress that our methodology can be routinely applied to any species. We take as our response variable Atlantic cod abundance with the objective of demonstrating how a Bayesian hierarchical spatio-temporal model brings forward, in a novel and yet easy to visualize way, what, and when, the cod population did with respect to distribution and abundance during the critical years of 1986 through 2003. These are the years for which we have both the OBIS trawl data and the best available fishing data for Atlantic cod which are annual landings. These landings are not spatially indexed, that is, the locations

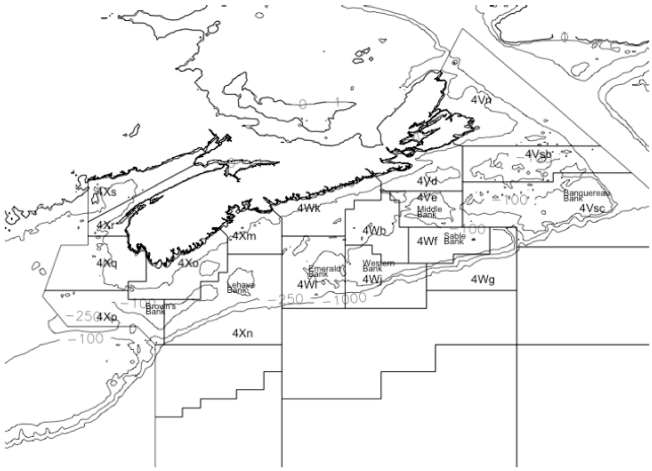

Figure 6.1: **NAFO Areas and Named Features** Designations of the NAFO zones and sub-zones on and around the Scotian shelf; 4X, 4W and 4V. The area within the contour lines marks the approximates the Scotian Shelf.

where the cod were harvested are not known, only the NAFO sub-zone (in some cases only zone) was recorded. We therefore calculate landings by sub-zone by year in an effort to assess the impact of fishing directly. Specifically we utilize a single number summary for cod landings by year and sub-zone, where for sub-zone(i), the landings for that year are calculated ($Landings_{i,t}$) as well as the total landings for the year ($Landings_t$). These are combined with our measure of abundance, the OBIS trawl data, which too is summed by sub-zone ($OBIS_{i,t}$) and by year ($OBIS_t$). The *relative exploitation* rate (RE) in that sub-zone is then calculated as,

$$RE(i,t) = \frac{\frac{Landings_{i,t}}{Landings_t}}{\frac{OBIS_{i,t}}{OBIS_t}}. \tag{6.1}$$

These data allow us to explore the relationship between the response and the available cod landings data since it is commonly held that the root cause of the cod stock collapse was overexploitation (Hutchings and Myers, 1994; Myers et al., 1997; Rose et al., 2000; Walters and Maguire, 1996). We have quite a long time series of data, 1970 to 2014, and some knowledge of what the nature of the fishing pressure was on Atlantic cod. Broadly this timespan may be separated into four distinct periods, based on the nature of the fishing pressure:

1. In the first period, 1970-1977, the main fishing effort was by the foreign fleets. It is widely held that overfishing by these foreign flagged vessels was responsible for overfishing the Atlantic cod precipitating the 'first collapse' in 1975.

2. In the second period, 1978-1985 the Cod experience a rebound in abundance as, after the imposition of a Canadian 200 nautical mile Exclusive Economic Zone(EEZ), the Atlantic cod stock was under lessened fishing pressure since the 'foreign fleet' was no longer operating in the new EEZ.

3. In the third period, 1986-1992, the Canadian domestic fleet ramped up to fill the void left by the departing foreign vessels and the fishing pressure upon the Atlantic cod stock re-intensifies, leading to another, this time even more pronounced, 'second collapse', followed by the eventual imposition of the moratorium.

4. In the fourth period, 1993-Present, the Atlantic cod stock remains at very low historical levels, and, despite the cessation of fishing, has 'failed to recover' its former abundance.

There are several other covariates worthy of consideration. In addition to the covariates measured at the time of the trawl (e.g. temperature) we also consider an oceanographic covariate (Bathymetry). Bathymetry for the area of the Scotian Shelf has been obtained from the U.S. National Oceanographic Atmospheric Administration (NOAA) (Amante and Eakins, 2009).

6.4 Methods

Spatial models depend on Tobler's Law of geography, which states that all locations are related but neighbouring locations are more related than distant locations, and estimate a statistical correlation in the residuals, after accounting for the effect of covariates (Cressie and Wilkie, 2011). We, and others, find Gaussian random fields (GRFs) (Thorson et al., 2014) to be the simplest full implementation of spatial modelling. GRFs can be efficiently used to simulate spatial dependencies in order to estimate spatial correlations in a statistical model (Shelton et al., 2014), i.e. the covariance matrix, Σ, and express these with a simple and interpretable set of parameters ρ, σ, which we interpret as connectivity and spatial variance respectively. A third parameter, a, here referred to as persistence, arises if a temporal structure is desired. For small sample sizes Σ can be calculated to estimate these parameters directly. However, this requires inverting Σ, which becomes computationally infeasible for a large number of points. INLA, (for Integrated Nested Laplace Approximation) (Lindgren and Rue, 2015; Rue et al., 2009a), approximates the inverse-covariance matrix, (Σ^{-1}), of the GRF using sparse matrices calculated using the stochastic partial differential equation approach (Illian et al., 2012b; Lindgren et al., 2011). This approximation is extremely fast, and is easily implemented using R-INLA (Rue et al., 2009a) in the R statistical platform (R Core Team, 2013). Given the ease, efficiency, and generality of the R-INLA package, we propose a Bayesian Hierarchical Spatio-Temporal model framework for the Atlantic Cod abundance data. This approach has been used in animal tracking (Carson and Flemming, 2014; Illian et al., 2012b) and more recently in the marine context (Munoz et al., 2013; Quiroz et al., 2014), modelling habitat (Pennino et al., 2013), nurseries (Paradinas et al., 2015), and bycatch (Cosandey-Godin et al., 2015; Pennino et al., 2014; Thorson et al., 2014).

Spatio-temporal model structure

The response variable $y(s,t)$ is the total number of cod captured in a single trawl (1 to a maximum observed count of 12189) at location s at time point t, $t \in (1986, ..., 2003)$. Since these data are counts, we consider suitable candidate

distributions including the Poisson and negative binomial distributions (each with their respective canonical link functions (Zuur et al., 2009)).

The mean of our response, $E[Y(s,t)] = \mu(s,t)$, is mapped by a link function to a linear predictor $\eta(s,t)$ as in the Generalized Additive Model framework (Hastie and Tibshirani, 1990). That is,

$$\eta(s,t) = \xi(s,t) + \sum_{j=1}^{n_f} f_j\{c_j(s,t)\}, \tag{6.2}$$

where the linear predictor is the sum of parts; a spatio-temporal random effect $\xi(s,t)$, and smoothed functions of covariates $f_j\{c_j(s,t)\}$, where n_f is the number of covariates. The $f_j\{c_j(s,t)\}$ are smoothed functions of covariates rather than linear ones, where $c_j(s,t)$ is the value of the jth covariate at location s and time t. Using such functions allows the effect of the covariate to vary across its values. Several of the potential covariates are highly co-linear, such that it would be inappropriate to include all of them in our model framework simultaneously. For covariates that have pairwise correlations ≥ 0.9 (e.g., nitrate and silicate), we consider only models that contain one or the other. The spatio-temporal random effect $\xi(s,t)$ may be thought of as representing the cumulative effect of latent factors impacting the response and so can be interpreted as a latent variable (Carson and Flemming, 2014) where its characteristics compose the spatial and temporal covariance structure of the model, here that of the Atlantic cod distribution on the Scotian shelf.

GMRFs and the SPDE approach

GRFs are usually defined by a mean and a covariance function $Cov[(s,t),(s',t')]$ defined for each $(s,t),(s',t')$ in $R^2 \times R$, that is, defined between locations(s) and times(t). Modelling Gaussian fields directly is often difficult, especially for large data sets and there is some literature on this problem(Simpson et al., 2012a,b; Taylor and Diggle, 2012). The Stochastic Partial Differential Equation (SPDE) approach, in which a spatio-temporal random effect $\xi(s,t)$ is treated as a GRF and represented with a Gauss Markov Random Field (GMRF), is one attempt to surmount this difficulty with some computational simplifications. Under the SPDE

approach, the continuously indexed GRF is represented as a *discretely* indexed random process, a GMRF. The computational advantages are realized by this representation since the continuous integrals of the GRF are replaced by the discrete sums of the GMRF. A thorough explanation, proofs and theoretical details may be found in (Lindgren et al., 2011).

Let us, for explanatory purposes, consider our penultimate model. This model will incorporate a first order auto-correlated spatio-temporal effect between years with coefficient a. This means that the random field incorporates a temporal persistence parameter, ($|a| < 1$), combined with a spatial covariance function. That is,

$$\xi(s,t) = a\xi(s,t-1) + \omega(s,t), \tag{6.3}$$

and,

$$Cov[\omega(s,t), \omega(s',t')] = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C(h;\nu,\kappa) & \text{if } t = t', \end{cases}$$

where $\omega(s,t)$ has a zero mean gaussian distribution, is temporally independent, and has a spatial covariance function for $s \neq s'$ where,

$$C(h;\nu,\kappa) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^\nu K_\nu(\kappa h). \tag{6.4}$$

The parameters of this Matérn covariance function, $C(h;\nu,\kappa)$, are ν and κ, $\nu > 0, \kappa > 0$. (K_ν is the modified Bessel function of the second kind). The parameter ν determines smoothness and κ determines spatial scale. and the covariance function is a function of the distance separating the locations $h = \|s - s'\|$. In practice, the parameter ν is usually fixed, (we take $\nu = 1$), and $\rho = \frac{\sqrt{8\nu}}{\kappa}$ is reported empirically with ρ being the distance at which the spatial correlation is reduced to approximately 0.1 (Cameletti et al., 2011) (Cameletti et al., 2013).

We have a continuous GF that we want to represent as a GMRF. The GMRF is a spatial process that models spatial dependence on a grid or lattice or graph (Illian et al., 2012b). If we denote this continuously indexed GF with Matérn covariance function, defined by parameters κ and ν, as $X(s)$, the aim is to find a GMRF that best represents $X(s)$. As an alternative to using a regular grid,

the SPDE approach utilizes a triangulation of the domain (Lindgren et al., 2011). The distinction is an important one; the use of a triangulation contributes to the computational advantage of this approach since, unlike a grid, it allows for cells of different sizes, reducing the number of empty cells where data is sparse while retaining fine resolution where data is dense. The domain is subdivided into non-intersecting triangles with vertices at the data locations. Additional vertices are then added sufficient to get a useful triangulation. Some care is required in the process of creating and defining a mesh, since it is desirable to have a mesh with triangles of somewhat similar size and shape, while avoiding any excessively acute vertices, (Lindgren et al., 2011). The R-INLA package offer some tools to assist the practitioner in achieving a suitable mesh. The 'max.edge' tool allows the user to specify the maximum side length for a triangle (and thus limit the maximum triangle size and hence resolution of the mesh), while the 'cutoff' tool allows the user to treat data points within a specified distance to be treated as one point, thus preventing overly small triangles and so controlling the minimum resolution of the mesh. Our triangulation is shown in Figure 6.2.

Very simply, the SPDE will represent $X(s)$ at each vertex and interpolate values in between. More completely, the basis function representation of the original field $X(s)$ is;

$$X(s) = \sum_{l=1}^{n} \psi_l(s)\epsilon_l, \tag{6.5}$$

where n is the number of vertices in the triangulation, $\psi_l(s)$ are the basis functions and ϵ_l are gaussian weights. The basis function $\psi_l(s)$ is equal to 1 at vertex l and 0 at all other vertices. The value of the field at any vertex is given by ϵ_l and values for the interior of the triangles are determined by linear interpolation. Once written this way (Lindgren et al., 2011) show that there is a mapping of the covariance function $C(h; \nu, \kappa)$ of the Gaussian field to the covariance matrix of the GMRF, (through its inverse, the precision matrix, $Q = \Sigma^{-1}$), such that the spatio-temporal model can be rewritten in terms of a GMRF.

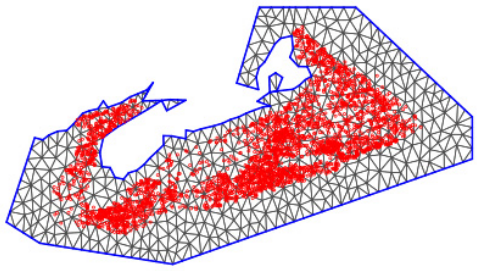

Figure 6.2: **Triangulation.** The triangulation utilized for the SPDE approach with n=548 vertices. Red symbols mark the data locations.

Model Assessment

In a Bayesian approach, the parameters that comprise our model are treated as random variables and prior information about the parameters is incorporated in a prior distribution. Recently, INLA has expanded the prior options it offers the analyst. INLA has incorporated an methodology for prior selection using 'penalized complexity priors' (pc.prior) (Simpson et al., 2015). This construction, which seeks to provide weakly informative default priors that "are useful, understandable, conservative, and better than doing nothing at all"(Simpson et al., 2015).

In the kind of model we have here, the random field is a spatial random effect; if there is no spatial random effect it is equivalent to having $\rho = \infty$ and $\sigma = 0$, that is, the effect is a constant 0 everywhere. Having a finite ρ and $\sigma > 0$ makes the model more complex, hence the rationale. The pc.prior format allows the user to control the priors by considering the problem. The user is required to supply a value for ρ_0 and a probability that $\rho < \rho_0$. By considering a reasonable lower value for the spatial effect ρ_0 is chosen. The probability chosen supplies the weight of the penalty on the more complex model. For σ, one considers a reasonable upper value for the spatial variance, the penalty shrinks the model toward $\sigma = 0$, since that is the simpler case. With no a priori expectation, we chose a values of $\rho_0 = 0.5$ with $P(\rho < 0.5) = .5$. By a similar process we chose $P(\sigma > 0.75) = .5$. All the models we subsequently report use these priors.

The various candidate models are compared using the Deviance Information Criterion (DIC), the Logarithm of the Pseudo Marginal Likelihood (LPML) and/or the Root Mean Squared Estimation Error (RMSEE).

Deviance Information Criterion: DIC

The DIC (Spiegelhalter et al., 2002) is the most common diagnostic function found in discussion of Bayesian models. It works by summing a quantity, the expected deviance $E[D(\theta, x)]$, with another, the number of effective parameters p_D. Models with *lower* sums, (lower DIC), are considered superior. The DIC is calculated by INLA by default and is found in the model output from an INLA model. To simplify interpretation, DIC measures the goodness of model fit while simultaneously penalizing complex models.

Logarithm of the Pseudo-Marginal Likelihood: LPML

Another Bayesian diagnostic model criterion is the Conditional Predictive Ordinate (CPO) (Geisser and Eddy, 1979), a *W-fold* leave one out cross validation. This is calculated by taking W equal sized samples, (typically 5 or 10 percent of observations, here 5) from the data, $x_1, \ldots, x_w$, calculating an estimate for each location (s, t), and for each location averaging over the samples to find $\widehat{CPO}$ as:

$$\widehat{CPO}_{(s,t)} = \left(\frac{1}{W} \sum_{w=1}^{W} \frac{1}{\pi(y_{(s,t)}|x_w, \theta_w)} \right)^{-1}, \tag{6.6}$$

Now, $CPO_{(s,t)}$ is a goodness of fit measure for each observation - it can be summarized for the entire data set as a single value, LPML, with $y_{-(s,t)}$ being y without the (s,t) st element.

$$LPML = \sum^{n_{obs}} log[\pi(y_{(s,t)}|y_{-(s,t)})] \approx \sum^{n_{obs}} log[\widehat{CPO}_{(s,t)}]. \tag{6.7}$$

In this way the LPML acts as a comparator of the predictive quality of the models. The *larger* the CPO, the better the model. INLA ordinarily calculates the CPO as part of the default output.

Root Mean Squared Estimation Error: RMSEE

The closeness of the estimation can be checked via the Root Mean Squared Estimation Error (RMSEE). The RMSEE is not calculated by R-INLA, but is readily calculable from the observations ($y_{(s,t)}$) and the linear predictor from equation 1.

$$RMSEE = \sqrt{\frac{1}{n_{obs}} \sum d^2_{(s,t)}}; \quad d_{(s,t)} = y_{(s,t)} - E[Y_{(s,t)}|x, \theta] \tag{6.8}$$

Clearly, *smaller* is better.

6.5 Results

Our model framework involves a spatio-temporal covariance structure and a set of covariates that best describe the response (the $\xi(s,t)$ and c_j of equation 1 respectively). We consider models that include no spatial or temporal effect at all; this amounts to simply modelling the response either as a mean (without covariates) or as a function of the covariates (only). We also consider models with a single spatial effect constant through time, as well as those with temporally varying effects. Temporally varying models considered are those with spatial effects replicated each year, that is, a single spatial effect for each year (without temporal correlation), and models with spatial effects correlated between years via an AR(1) relationship. We select the best spatio-temporal structure the same way as we

select the best distribution for $y(s,t)$ and the same way as we choose our eventual covariates. That is, we run models using the various alternative constructions and compare them using the model assessment criteria discussed in the previous section. For all spatio-temporal model formulations considered, the DIC for the negative binomial response distribution was always lower than that of the Poisson (LPML is higher, RMSEE is lower). Hence the negative binomial distribution is to be preferred and for brevity we display results in Table 1.1 for only for the negative binomial response and the three best performing physical covariates. Amongst these models the inclusion of the AR(1) temporal structure results in the lowest observed DIC (highest LPML, lowest RMSEE). On this basis we choose the AR(1) spatio-temporal structure. This leads to the following model formulation:

$$\eta(s,t) = \xi(s,t) + f(RE(i,t)) + f(Temperature(s,t)), \tag{6.9}$$

where $\eta(s,t)$ is modelled as in Equation 3, and with smoothed functions of Relative Exploitation and the Temperature at the trawl. This model performed best according to both the DIC and the RMSEE criteria. There was a slight improvement in LPML when including a Bathymetry covariate but this resulted in poorer DIC and RMSEE measures and including Bathymetry along with RE and Temperature did not improve estimation.

Covariate(s)	DIC	LPML	RMSEE
ξ(s,t) Alone	22342.11	-18244.10	245.58
ξ(s,t)+f(Temperature)	22277.74	-18192.46	246.04
ξ(s,t)+f(Bathymetry)	22236.59	-18203.59	246.48
ξ(s,t)+f(RE)	22166.12	-18173.36	246.39
ξ(s,t)+f(Temperature)+f(Bathymetry)	22206.37	**-16145.51**	244.52
ξ(s,t)+f(RE)+f(Temperature)	**22131.31**	-16245.66	**240.68**
ξ(s,t)+f(Bathymetry)+f(RE)	22138.84	-17520.52	245.77
ξ(s,t)+f(RE)+f(Temperature)+f(Bathymetry)	22155.41	-176742.23	251.08

Table 6.1: **DIC values.** DIC values for the various candidate models with an AR(1) spatio-temporal covariance structure.

The parameters of the model specify the spatio-temporal random effect. Considering each of these estimates one at a time, ρ is the spatial connectivity parameter, the range at which the spatial correlation is reduced to approximately 0.13. That

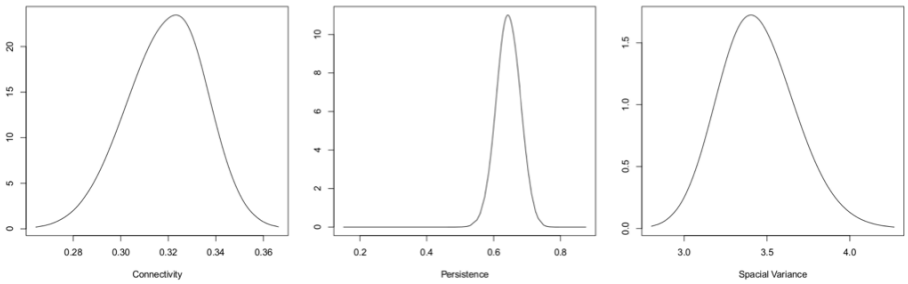

Figure 6.3: **Results** Posterior distributions for the model parameters.

is, the value $\hat{\rho} = 0.321$ is the estimated distance (in degrees - so approximately 34 km) at which this occurs. The estimated spatial variance of the GMRF is $\hat{\sigma} = 3.39$. The coefficient for the AR(1) process (the persistence parameter, a) in Equation 2 is estimated at 0.627. For the negative binomial distribution assumed for the responses, $\hat{\phi} = log(n) = 0.986$ (n is the size parameter, $\sigma^2_{nbin} = \mu(s,t)(1 + \frac{\mu(s,t)}{n})$). Plots of the posterior distributions of the parameters are provided in Figure 6.3.

By assembling the above components we can nicely display the results. Combining the effects of the fixed covariates with the model output for the random field gives us the mean of the model for each year. A plot of this mean for 1986 (pre-collapse), is shown in the upper left panel of Figure 6.4 and consists of the sum of the linear predictors in the model and the elements of the random field. A similarly constructed plot for the year 1992 (post-collapse) is shown in the upper right panel of Figure 6.4. The estimated functions for the effects of the two covariates are shown in Figure 6.5.

Relative Exploitation and Temperature are both seen to have some significant effect on cod abundance over portions of their ranges. The dashed lines in Figure 6.5 indicate the 95% credible intervals for the estimated effects. The covariates have significant effects over those portions of their ranges where this interval does not include zero. When the water is cold ($< 2^oC$) the effect is positive and trends toward negative as water temperatures rise to 10^oC. Low relative exploitation levels, (< 100), positively impact abundance. While not significantly non-zero, the trend in these two covariates are in the expected direction, and, these results are entirely

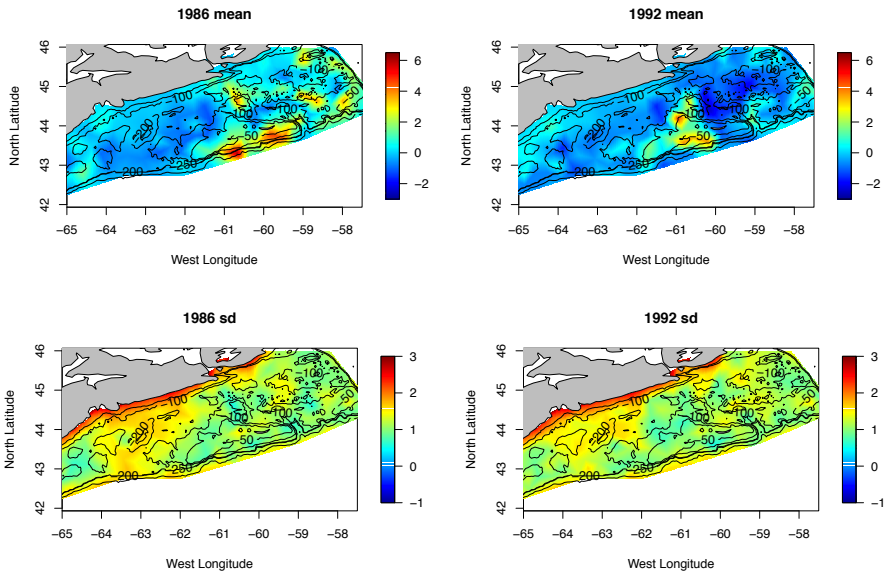

Figure 6.4: **Results** Plot of the posterior mean for the years 1986 and 1992 (pre and post collapse). The lower panels show the corresponding plots of uncertainty (the response SD). The scale is log(predicted count).

consistent with expectations and what is previously known about cod (Campana et al., 1995; Hedger et al., 2004).

We also fit our model to each year of data individually (and consequently without the persistence parameter) in order to obtain annual estimates of both the spatial connectivity parameter ρ and the abundance parameter σ so as to look for patterns potentially related to levels of exploitation. A direction of future research is to consider a single model that incorporates autoregressive relationships (for example) between these parameters, but this generality is not presently available using INLA.

6.6 Discussion

Local over-fishing (serially fishing out concentrations that do not replenish) has been inferred for Northern cod (Hutchings, 1996; Hutchings and Myers, 1994; Rose and Dutka, 1999). Our spatio-temporal model approach makes it evident that cod became spatially concentrated as abundance decreased, until they became so

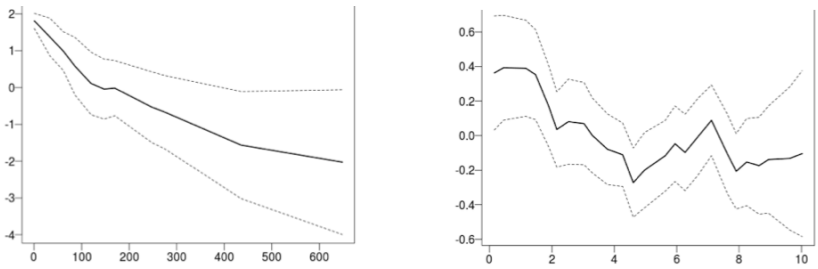

Figure 6.5: **Relative Exploitation and Temperature Effects.** The covariate plots show the value of the covariate on the X axis, and the impact of the covariate on the response on the Y axis and their 95% credible intervals (the dashed lines). Viewed as functions of the covariates, lower Relative exploitation and colder temperatures produce higher predicted abundance.

depleted that abundance even in the core areas collapsed.

We begin by considering out results, pre-collapse vs. post-collapse, as in Figure 6.4; estimated abundance was very high in 1986 but by 1992 had collapsed precipitously. Looking at the posterior mean of the model for these two years we see that the largest forecast values in 1986 and 1992 (the red areas) are located in (approximately) the same location, (around 60-61W, 43N) and, moreover, have (again approximately) the same predicted value despite the precipitous decline in overall abundance, see Figure 6.7. Indeed the maxima of the observations and the maxima of the predicted values are similar in value and the value does **not** decline along with the overall decline in abundance seen over this period. What does appear to change is the spatial extent of the moderate values. Away from the red there is a general decline in the predicted values; seen as the areas of pale red/yellow in 1986 appearing as blue in 1992. The decline is seen as a reduction of the spatial extent of high and moderate values, not as a decline in the maximum values. This is entirely consistent with the hypothesized hyper-stability (Rose and

Dutka, 1999) of abundance in the preferred, or *core range* (Shackell et al., 2005). In order to more fully illustrate what is happening we present two more plots in Figure 6.6. In this figure only the areas of highest predicted abundance are shown. To emphasize our point, i.e. to highlight those areas where abundance is 'high', we chose an (arbitrary) value equal to the 75th percentile of estimated cod abundance values and then plot the locations where cod abundance was predicted to equal or exceed this value. We note that a high AR(1) term tells us that the spatial distribution of biomass stays the same year after year, and we have confidence that high density areas persist and are important. Areas that are always occupied, during periods of high and low abundance are interpreted as 'core' areas, but in a collapsed stock, even core areas will disappear(Shackell et al., 2005). We see that between the left panel and the right panel the area where abundant cod are predicted to be decreases markedly, disappearing completely from previously abundant Banquereau bank, even though peak abundance remains constant. This is interesting as it certainly appears that the cod are contracting towards the areas of highest density as the overall abundance diminishes, another result consistent with previous cod studies, e.g. (Hutchings, 1996).

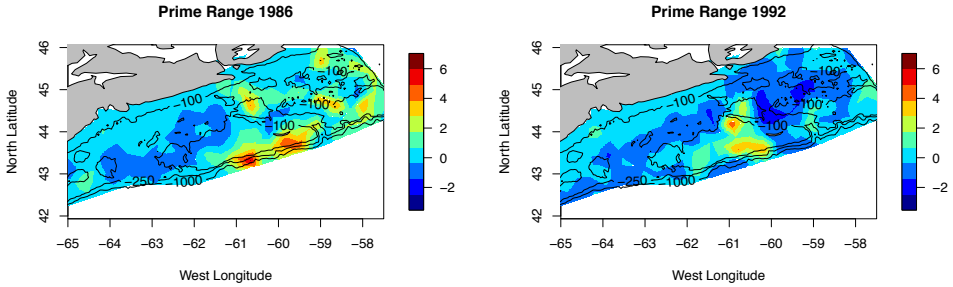

Figure 6.6: **Core Range** Plot of the posterior mean for the years 1986 and 1992, showing only those areas where the predicted mean is > 75th percentile. Viewed as 'core range'. The species range has contracted with the reduced abundance but maximum density in the aggregations has not changed, scale is log(abundance).

In summary our methodology suggests a spatio-temporal model for mean abundance that is entirely consistent with the occupancy-abundance hypothesis. In fact, Figures 6.4 and 6.6 illustrate the phenomenon of stable abundance in key areas of the range combined with a contraction of spatial distribution under the

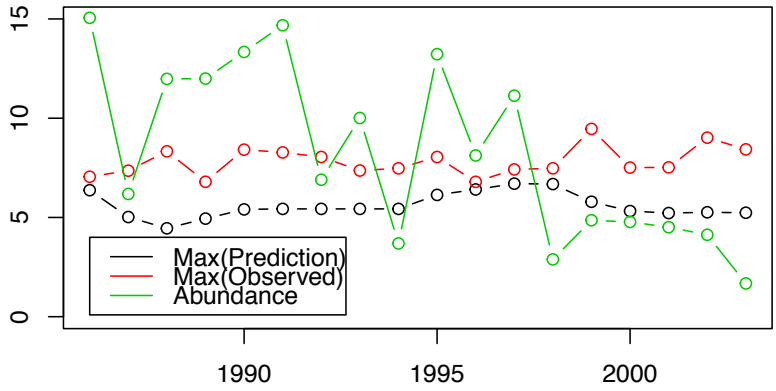

Figure 6.7: **Hyperstability** The maximum *observed* count (LogY), the maximum *predicted* count (η) are nearly constant, even though the stock is collapsing.

circumstances of an overall decline (in this case large) of the population as a whole. Interpretation of the model parameters themselves is also interesting and indeed entirely consistent with theory. The relatively large value of the persistence parameter in the AR(1) construction, (a), suggests a strong connection between the observed cod abundance from year to year; in other words the cod are to be found, or not found, in the same places year after year. Thus the areas of consistently high abundance may be thought to be important to the stock, congruent to the idea of core range posited earlier. The range or scale parameter, ρ, is interpretable since it is the distance at which covariance is considered to become insignificant (<0.13). The value $\hat{\rho} = 0.321$ is the distance (in degrees) giving us an idea of the physical scale of the cod core range or ranges, 34 km. One way of looking at the meaning of $\rho = 0.321$ is to consider the implication of independence at distances greater than ρ. If we have 2 or more areas of consistently high abundance separated by some distance greater than ρ then statistically they are independent, that is, one could consider them separately. The 3 distinct areas of high abundance remaining in the lower panel of Figure 6.6 are all separated by at least 2ρ, quite a large value given what we know of cod mobility (Hedger et al., 2004; Svedäng and Bardon, 2003; Wright et al., 2006). One might expect an element of isolation by distance

(Cadrin et al., 2014) and therefore divergence - hence an argument for separate sub-populations. The physical separation argues for the treatment of these 3 remnants as distinct putative sub-populations. From a conservancy perspective we would argue for the desirability of treatment of the aggregations as biologically independent unless other information comes to light. It is important to remember that the model does not capture within year movement patterns (that is, is based only on the July survey data: the fish from these spatial aggregations present in July could mix at a different time of year) so this is not definitive, only suggestive, but does concur with previous categorizations and it is known that cod display high degrees of site fidelity (Hedger et al., 2004; Reubens et al., 2013; Svedäng and Bardon, 2003; Wright et al., 2006). As an item for further work one could postulate that overfishing has resulted in the removal of cod from the Banquereau Bank (between the eastern 2 areas in Figure 6.6) resulting in the division of the previous population into 2 distinct remnants (Shackell et al., 2005). Previous studies (Shackell et al., 1997) found that median distance travelled to recapture for cod in this area of the Scotian shelf was 36 kms; our work supports the contention that the remnant patches are even less connected that they once were due to the erosion and elimination of some subpopulations, notably Banquereau bank. Recolonization of such a vacant, yet previously dense, portion of the range would be a hallmark of any recovery. Indeed, the spatial distribution of the cod throughout the 1990's shows little variation and the stock remains at low abundance, that is 'fails to recover'.

The premise here is that the range of a species is density dependent - that is, they only spread out when their prime territory reaches maximum capacity, or, conversely, that the density of the species will remain relatively constant in the most suitable habitat and that increases in total abundance will increase the total range and *not* the density (Rose and Dutka, 1999). If the total abundance of the species is reduced the total range may contract but density sampled in the prime territory may not change at all since there is an in-migration effect. The reaction in abundance through the years of collapse should be seen in our models posterior mean not as a reduction of the maximum level but as a shrinking of the area of maximal (or simply high) abundance. That is, we should see range shrink, not

maxima. Examining Figure 6.6 and Figure 6.8, this is exactly what we do see. We do not conclude directly that these areas of remaining high relative abundance are therefore prime habitat for cod. Since we believe the cod have been removed through overfishing (Hutchings, 1996; Hutchings and Myers, 1994; Myers et al., 1997) and since we do not know the rate at which the cod will in-migrate to fill their now vacant former habitat (Erisman et al., 2011; Shackell et al., 2005) we conservatively interpret ρ as the range of spatial aggregation of the remnants of the original population. These remnants are, now, the sole potential source for recolonization of any formerly important habitat vacated by overfishing and, this recolonization will be seen as a reversal of the trends noted herein; an increase in the area of moderate density for cod. This suggests that one indicator of recovery for the Scotian shelf Atlantic cod would be an increase in their abundance *outside* of the areas noted as containing the remnant sub-populations and argues strongly for the managerial efforts to sustain cod recovery protect these areas important to the remnant sub-populations. Indeed, the recolonization of these areas by cod and the recovery of the stock are synonymous. The failure to recover seen in Atlantic cod (Fu et al., 2001) is evident in Figure 6.6, the cod do not expand from their remnant sub-populations. In any event, the survival of the Atlantic cod depends on the future of these three surviving remnant sub-populations and knowing their location and extent is valuable information to any management plan.

Ideally, a well managed stock should not suffer changes in distribution or structure due to exploitation. In the case of Atlantic cod this was definitely not the case; measuring CPUE only in places of relatively high abundance failed to detect the contraction of a depleting stock onto core range until it was too late to prevent the collapse of the stock, resulting in the near disappearance of cod in parts of the Scotian Shelf such as Banquereau bank. Both the distribution, and structure were changed. This leads us to wonder how we may detect such changes in structure using our techniques. To do so we fit our model to the data on a year to year basis, and our findings are displayed in Figure 6.9, in which we display the joint behaviour of ρ and σ in four panels; one for each of the periods identified above, with some years of notable change highlighted. Small values of ρ are at the top, indicating high structure, small values of ρ are at the bottom, indicating lack of

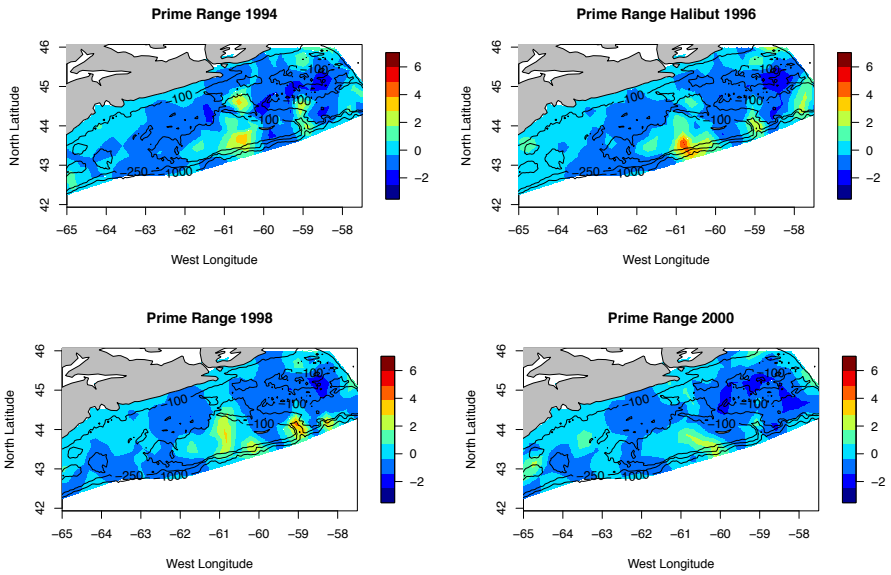

Figure 6.8: **Failure to recover** Plot of the posterior mean for the years 1994 through 2000, showing only those areas where the predicted mean is greater than the 75th percentile. Viewed as 'core range'. The species range fails to recover abundance and the habitat remains unchanging.

structure (flatness). Examining this figure we note the following:

1. In the first panel, 1970-1977, in the period which we term the 'first collapse' we see the very large shift of the parameters to the lower left marking the partial collapse of 1975.

2. The second panel, 1978-1985, might be termed the 'first recovery'. Canada imposes a 200 mile EEZ and the cod see some relief from fishing pressure. We see the cod regain first structure, 1978-1979, then start to regain abundance, 1979-1980 and 1980-1981. the re-establishment of structure is what we might expect under conditions of DDHS, a return to prime range. The recovery of numbers seems to lag re-establishment of structure.

3. The third panel, the 'second collapse'. From 1988 to 1989, and again from 1989 to 1990, there is an even stronger movement to the lower left, i.e. a simultaneous increase in ρ and decrease in σ. While the moratorium was imposed in the early 1990s few would that it was imposed too late to avert

significant degradation of the stock and Atlantic cod suffered a more profound collapse than that of 1975. Our analysis shows that the real damage was inflicted 1988-1991.

4. Panel four. After the imposition of the moratorium in the early 1990s we see, not recovery, but a period of what might be termed stable non-recovery. We see perhaps an effect of the imposition of the moratorium on Atlantic Cod, but, a re-establishment of structure without an increase in numbers. An expression of DDHS, the remnant fish re-aligning themselves onto the available habitat. This doesn't really constitute a 'recovery' however. It only reaches the top centre of the plot and σ remains small. Contrast this to the recovery of the early 80s where there is bias to the right of the plot. This top centre position is the new reality for cod, stable non-recovery.

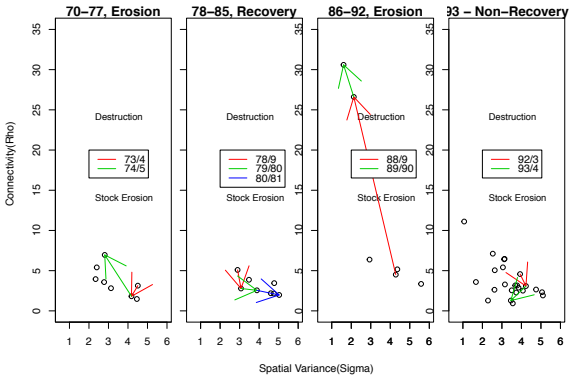

Figure 6.9: **Observed Parameter Movements** What the parameters were observed to do. The coloured arrows correspond to the years discussed in Figures 10,11.

We look first at the 'first collapse', particularly the period 1973-1975. Focussing on the first collapse, the predicted spatial mean from 1973-1975 shows erasure of areas of high density and a corresponding flattened RF, increase in ρ and decrease in σ. This is a clear picture of spatial erosion. While there is uncertainty in the parameter values during the 'first collapse', the pattern of parameter behaviour is repeated even more strongly in the 'second collapse'. Figure 6.11 displays the predicted spatial mean over the period 1988-1990. Instead of the partial collapse seen in the 1970s the Atlantic cod hits historic lows across the Scotian Shelf. The

erosion, perhaps destruction is not too strong a word, of the cod is seen as an even more extreme flattening of the RF, with correspondingly larger increase in ρ and decrease in σ. Examined in detail these collapses display common trends in parameters revealing erosion of the spatial structure present in the Atlantic cod.

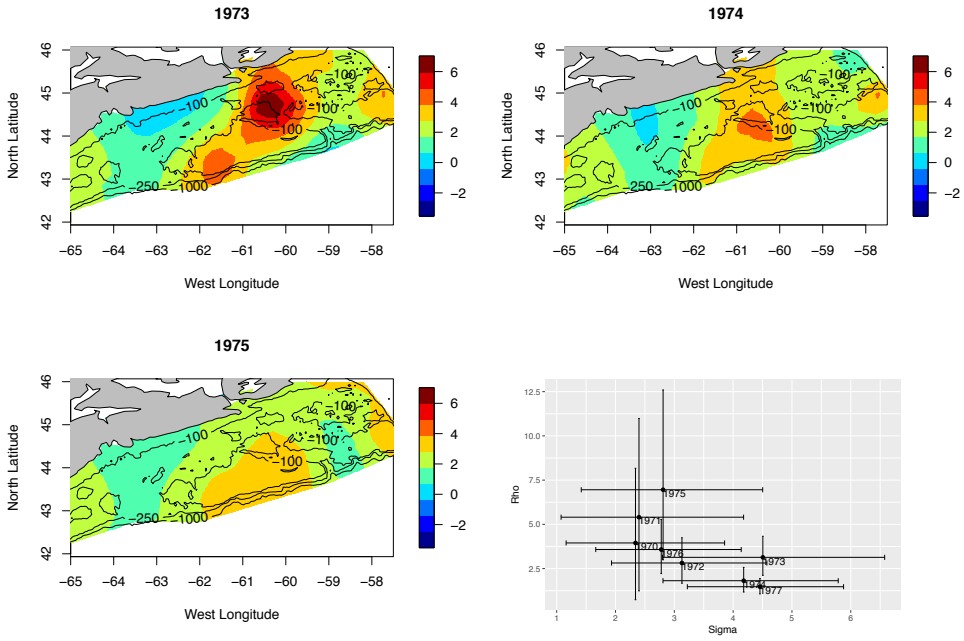

Figure 6.10: **The 'first collapse'** Plot of the posterior mean for the years 1973 through 1975, showing a dramatic flattening of the RF. This is indicative of circumstance in which heaving fishing was *eroding* the structure of the cod distribution leading to a partial collapse. In this period, $\rho = 3.1 \rightarrow \rho = 2.61 \rightarrow \rho = 7.50$ $\sigma = 3.7 \rightarrow \sigma = 2.8 \rightarrow \sigma = 2.2$. A simultaneous large increase in ρ and decrease in σ.

Under the conditions prevailing in the different periods discussed here, what should we have *expected* our parameters to do? Suppose there were no fish, that is, abundance was 0 everywhere. What would our parameters show? Our spatial connectivity parameter, ρ, would be $+\infty$ since the field is 0 everywhere, no matter how far separated. Now, practically, our estimate will be some finite number since we are estimating in a finite space, but $\hat{\rho}$ will be large. On the other hand, σ, the variance, would be 0 since the field is everywhere 0. Taken *in isolation*, σ is fairly easy to interpret. Since in our modelling framework our model of the spatial covariance structure is the RF; if there is wide separation between areas of high predicted Cod abundance and low predicted abundance, σ will be large.

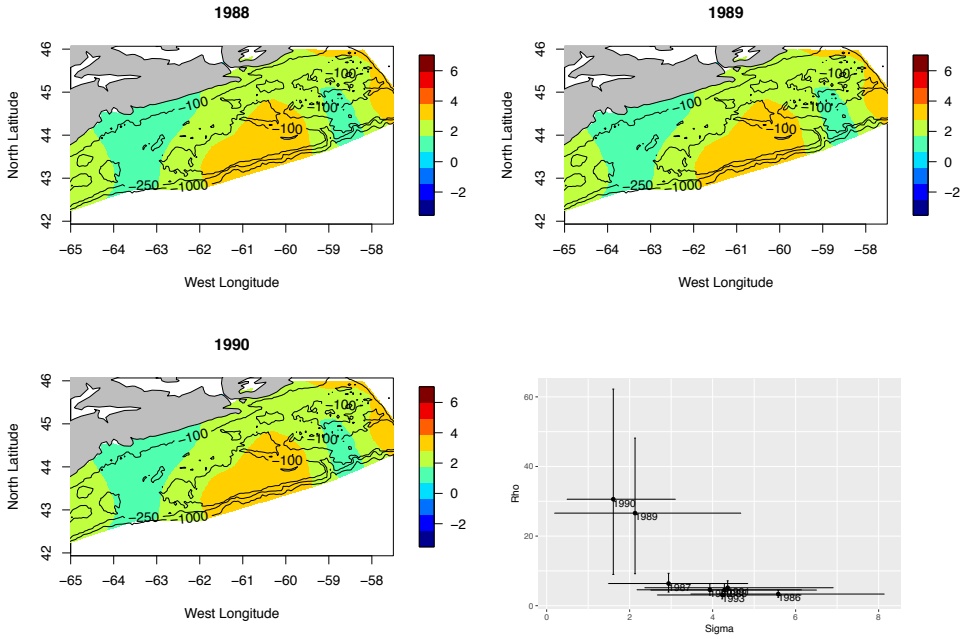

Figure 6.11: **The 'second collapse'** Plot of the posterior mean for the years 1988 through 1990, showing a dramatic flattening of the RF. This is indicative of circumstance in which heavy fishing was *effectively destroying* the structure of the cod distribution leading to near total collapse. In this period the flattening is more severe than what was seen in the 1970s, $\rho = 4.40 \rightarrow \rho = 26.43 \rightarrow \rho = 28.6$, $\sigma = 4.2 \rightarrow \sigma = 2.1 \rightarrow \sigma = 1.6$. A simultaneous *very* large increase in ρ and decrease in σ. the collapse of the early 1990s is much more pronounced.

So, *in isolation*, a large σ is needed when there is lots of contrast between areas of high fish density and low density, a small σ will mean that density is constant or nearly so over the space. In Figure 6.12 we provide a schematic view of expected parameter behaviour under differing conditions:

- We expect to see some fluctuation in the parameter values. Horizontal noise in σ is normal, year to year, fluctuation in fish numbers. Vertical noise in ρ is analogous variation in our estimation of structure.
- We expect an erosion of, or loss of, structure in the stock to express as an increase in the value of ρ, a flattened structure results in an increase in the spatial measure of covariance.
- A species experiencing moderate, or well managed, fishing might be expected

to see a small reduction in σ compared to the unexploited state, with little change in ρ.

- A species experiencing 'recovery', will see a simultaneous re-establishment of structure and increase in numbers; this would imply a movement to the top right of Figure 6.12.

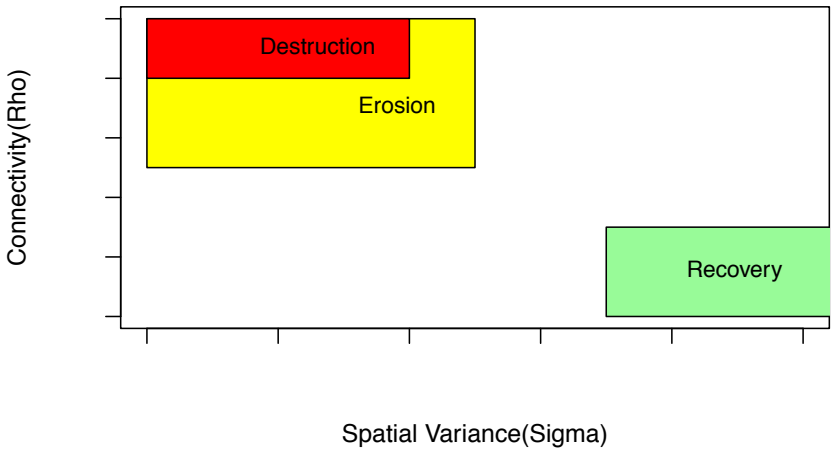

Figure 6.12: **Schematic of parameter behaviour** A schematic diagram of expected parameter values under different conditions in the fishery. Note that large ρ values are at the bottom of the schematic diagram.

- Large movement to lower left correlates to stock **destruction** as the densest areas of fish are removed and the stock structure is removed. Essentially, the field we are modelling is being flattened under stress due to over-fishing (i.e. the extreme depletion, or utter removal, of fish in high density areas), resulting in the *simultaneous* reduction of σ and increase in ρ, that is a large movement to the lower left on Figure 6.12.

Future Directions

This analysis opens up a number of possibilities, and questions. From the practical point of view of stock manager models like this suggest that, in addition to traditional means of monitoring the health of a subject stock one might routinely

examine these parameters looking for large movements or trends. Notably, for instance, an exploited stock should seek to avoid movement to the lower left of the plot, when depletion begins to erode the stock structure there is damage being done. Conversely, stock which one hopes to see recover might be monitored for movement to the upper right as being encouraging. Now these parameters are always going to be subject to some fluctuation from normal fluctuation in the spatial distribution and abundance of the subject stock, but, one can imagine that there exists a 'natural value' for them, i.e. that they will have some sort of true mean upon which they will centre in a population free from external disturbances. If you will, a box at the top centre of the ρ, σ plot will bound 'good' combinations of parameters. For the manager, excursions from the box require explanations.

6.7 Acknowledgements

We are indebted to Dr. Ken Frank for generously providing us with the Landings data for Atlantic cod. We would also extend our thanks to Dr. Shelley Lang for reading earlier versions of this manuscript and for her thoughtful commentary and wise council.

Chapter 7

An integrated modelling approach for examining spatio-temporal behaviour of a marine generalist predator

7.1 Abstract

Understanding the ecological footprint of a predator begins with an understanding of how that species uses space through time. Space use by predators is driven by many factors including the availability of prey and the nature of the physical landscape. Understanding spatial behaviour is particularly challenging in marine environments where animals and their interactions with other species are rarely observable. Recently the wide-spread use of location telemetry and data loggers for the study of marine animal movement has made large datasets (big data) available that can be used to model how marine animals use space and the factors underlying these patterns. Here we develop an integrated model framework to describe diving behaviour of grey seals (*Halichoerus grypus*) on the Scotian Shelf of eastern Canada. Geo-referenced metrics of diving behaviour (time spent at the bottom of dives, a proxy for foraging effort) for 89 adult seals are integrated with environmental factors (e.g. bottom temperature, bathymetry, and bottom sediment) and prey information derived from stratified-random, bottom trawl ecological surveys. We model the predator and prey simultaneously using a Bayesian hierarchical model incorporating latent spatial random effects represented using random fields. The resulting models reveal that the amount of bottom time exhibited by seals is relatively stable in both time and space, but is spatially differentiated by sex. Similarly, overall patterns seem not to be tied to the presence of particular prey species for the grey seal species as a whole, but males and females do appear to respond differently to certain species and prey assemblages. The pattern of habitat use for the seals showed little local spatial pattern, unlike that of the prey

themselves, which tend to be more highly structured with respect to their physical environment.

7.2 Introduction

The marine environment is a particularly challenging place for studying the animals which inhabit it. Unlike on land, where the animals go, and the behaviour they exhibit, is usually unobservable. Efforts to learn about behaviour often rely upon the use of electronic tags, which record, or transmit, information about an animal, i.e. where it goes, and perhaps some ancillary information. For these reasons, animal tracking data are key for researchers studying free ranging marine animals. Tag records are often analyzed as movement data. Successive locations from a single animal are serially correlated, and environmental covariates, along with metrics associated with that animal, are usually observed only at the animal's location. In a movement analysis these data are typically treated as time series (Breed et al., 2009; Jonsen et al., 2003) with covariates not often observed for locations, other than those occupied by the tagged animals. Focus may be upon finding the most probable actual path of movement, describing (parameterizing) a model for movement as a function of simultaneously observed covariates, or, in a switching model, categorizing discrete behavioural states (e.g. foraging vs. transiting) along the observed track based on the observed movement (Jonsen et al., 2005). Such studies are usually expensive, and the number of tagged or captured animals small. As a result, a large gap may exist between knowing where the sample of animals studied went and understanding distribution and behaviour at the level of the population.

One means of bridging this gap is to develop models that use the set of discrete locations generated by the tagged animals to model the number of individuals and/or the behaviour expected at unobserved locations. While there has been substantial development and broad scale use of location telemetry and data loggers in the study of animal movement (Hussey et al., 2015; Kays et al., 2015), models scaling inferences on the use of habitat from the tracking of individuals to a population; (to the ecological footprint), can be problematic. This is especially true of predictive models in the marine environment where the behaviours of the

animals cannot be directly observed. Thus, models predicting how species use the environment, or where animals are to be found in the environment, are important to plan effective conservation and management measures (Bailey and Thompson, 2009).

Recent advances in spatio-temporal methods (Lindgren and Rue, 2015; Lindgren et al., 2011; Rue et al., 2009b) permit the integration of telemetry data with other data sources (ones not tied to animal locations) into models incorporating spatio-temporal effects. The flexibility afforded by this modelling framework allows for the integration of different types of spatial data, measured at different locations, without the loss of the original spatial information.

Here we use spatio-temporal models to investigate the ecological footprint of a large marine predator, the grey seal, (*Halichoerus grypus*). Previous studies have attempted to model the impact of grey seal predation on commercial fish stocks in the northwest Atlantic (Mohn and Bowen, 1996; Swain et al., 2015; Trzcinski et al., 2006), with the significant limitation being the lack of direct knowledge of the distribution, in time and space, of grey seal foraging. We exploit the flexibility offered by spatio-temporal models to incorporate different data types to integrate spatial information about both predator and prey into the same model, each with their separate spatial structures. This allows us to not only directly model the spatio-temporal distribution of the seal's observed behaviour, but also to directly model any dependence between this observed behaviour and potential prey species.

Grey seals are size-dimorphic, marine carnivores in which adult males are approximately 50% heavier than adult females throughout the year (Beck et al., 2003a). They are a long-lived marine predator in the Family Phocidae with life expectancies for males and females of 30 and 40 years, respectively. In the Northwest Atlantic population, movement data from satellite tags indicate a broad overlap in the foraging distributions of adult males and females during the summer and fall, with the distributions limited to the continental shelf (Breed et al., 2006). They are a generalist predator feeding on a large variety of mainly fish species ($\sim$ 40 species), although individual animals often obtain most of their energy intake from a small subset of these species (Beck et al., 2007; Bowen and Harrison, 1994).

Males and females exhibit significantly different patterns of diving behaviour, energy storage (Beck et al., 2003a,b,c) and foraging tactics (Austin et al., 2003; Breed et al., 2009), which may indicate sex specific ecological footprints for the species within the same general foraging region.

Although gradients in habitat features are common, a more prevalent situation is a patchy environment, creating a mosaic of different habitats. In the marine environment patchiness is often expressed in bottom surface sediments, bottom topography, and the distribution and abundance of associated prey assemblages (Steele, 1985). How animals exploit such patchy habitats can provide insights into how individuals and populations may respond to environmental variability (Wiens, 1976). Over the course of our study grey seal abundance at Sable Island increased continuously as evidenced by estimates of the number of pups born on the island (den Heyer et al., 2017) and from population model estimates of (Hammill et al., 2017) to 345,000 grey seals in 2016. The fish stocks of the Eastern Scotian Shelf (ESS) have also undergone substantial changes over the same period with a marked shift in the ecosystem (Frank et al., 2005; Shackell et al., 2010). How, or if, the ecological footprint of the grey seal may have altered on the ESS in response to these changes is not understood. Behavioural changes in the use of space by animals are influenced by many factors including habitat suitability, prey quality and availability, competition and predation risk (Andruskiw et al., 2008; Kerfoot and , Editors; Lima, 1998). Thus, it is with the goal of improving our understanding of a major marine predator, the grey seal, and evaluating the ecological footprint of these wide-ranging predators and the risk of predation to prey species occupying different habitats in the presence of environmental variability (Mohn and Bowen, 1996), that we present this study.

To better understand this ecological footprint, we fit our spatio-temporal model to a proxy for foraging intensity, recorded contemporaneously with tracking data for 89 grey seals. We model this proxy both over time and as a function of indices of abundance for several potential prey species. We use the model to examine our expectation that, as a generalist predator, the grey seal may not key on any one specific prey by also modelling foraging behaviour as a function of an assemblage of the prey species we examined. Based on previous studies we expect to see sex

specific differences in diet (Beck et al., 2007) and distribution (Breed et al., 2006).

7.3 Materials and Methods

Our study area is the eastern Scotian Shelf (ESS). This region of the continental shelf, off Canada's east coast covers some 108,000 km^2 encompassing Sable Island ($43^o55'N$, $60^o00'W$) as well as several offshore shallow banks and basins separated by deep gullies and canyons (Figure 7.1). This continental shelf region is an important foraging area for the grey seals which breed on Sable Island (Breed et al. 2006, 2009). This island, home to a very large grey seal colony, is a crescent-shaped sandbar with a surface area of about 34 km^2, and is centrally located in the ESS. Our proposed spatio-temporal model framework integrates a collection of diverse data sources relevant to this study area, (Figure 7.1).

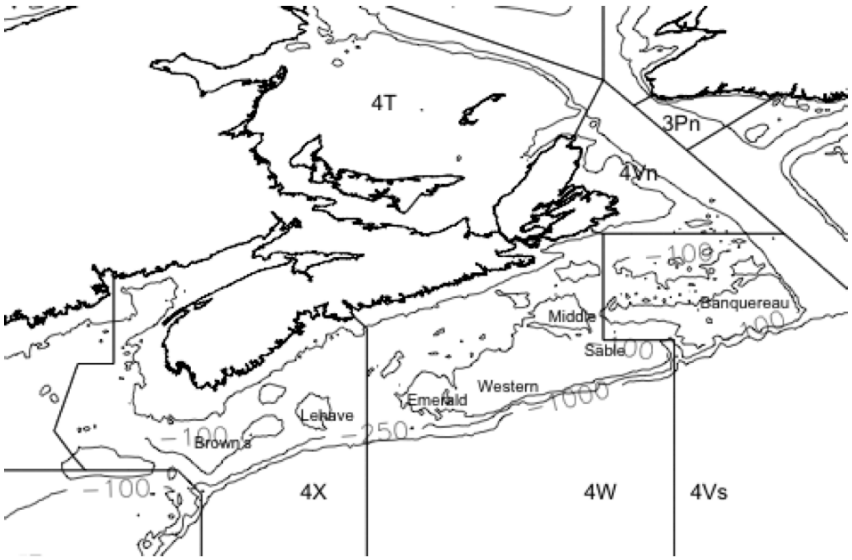

Figure 7.1: **Named Features on the Scotian Shelf** Designations of the North American Fisheries Organization zones and sub-zones on and around the Scotian shelf; 4X, 4W and 4V. The area within the contour lines marks the approximates the Scotian Shelf.

Seal Location Data

Grey seals were instrumented with electronic tags as described below from 1995 to 1999 and again from 2009 to 2015 to study their at sea distribution and diving behaviour. The 2009 to 2015 electronic tags used Fastloc GPS to record locations, resulting in fine resolution data in both space and time. The 1995 to 1999 electronic tags used Service Argos, an older system, which yields data which are less precise. Although the accuracy of locations provided by these tags differ, and the GPS recorded locations more frequently than Argos, both are considered here to provide the longest time series of observations possible. These differences also serve to illustrate the influence of sample size and resolution on predictions arising from our spatio-temporal model framework. In both time periods, adult seals were captured using a hand-held net and anaesthetized to permit the animals to be measured (body length and mass) and fitted with electronic tags (Beck et al., 2003b; Breed et al., 2006; Lidgard et al., 2012). Individuals were recaptured during the subsequent breeding season and the tags recovered and data downloaded. All field procedures were conducted in accordance with the requirements of the Canadian Council on Animal Care and were approved by Dalhousie University's Committee on Laboratory Animals and by the animal care committee of Fisheries and Oceans Canada (DFO).

GPS Tracking Data: 2009-2015

Between 2009 and 2015, 69 seals (15 males, 54 females; Table 1) were instrumented with Wildlife Computers Mk10-AF Fastloc GPS tags, (**www.wildlifecomputers.com**). The MK10-AF tag was programmed to archive GPS data that were downloaded on recovery of the tag. Tags were programmed to record a GPS location every 15 minutes and depth every 10 seconds when the seal was at sea and to suspend location attempts when the unit was dry for more than 20 min and a location had been obtained (Lidgard et al., 2012). To obtain GPS locations the archived GPS data from each MK10-AF were analyzed using propriety software from the manufacturer (WCDAP, **www.wildlifecomputers.com**) and archival ephemeris data (**www.cddis.gsfc.nasa.gov**). Locations acquired from fewer than 5 satellites and/or with a residual error of more than 30 meters were removed from the

data set due to their lower accuracy (Hazel, 2009).

Service ARGOS Tracking Data: 1995-1999

Between 1995 and 1999, 20 seals (6 males, 14 females; Table 1) were instrumented with either an SDR (Wildlife Computers, Redman, WA, USA) or an ST-18 (Telonics, Mesa, AZ, USA) satellite transmitter. These satellite transmitters were duty cycled to transmit for 8 h every day or every second day. To record diving behaviour, each animal was also instrumented with a time-depth recorder (TDR; Mk3e, Mk5, Mk6 or Mk7, Wildlife Computers, Redmond, WA, USA. TDRs were programmed to record depth every 20 s. Recording was halted while the animal was hauled out.

ARGOS satellite tracking data are similar to GPS tracking data in that location data can later be aligned to dive data and other environmental information collected by the seal. ARGOS is however, less precise. Argos location data may be subject to substantial measurement error (Austin et al., 2003), and it is not uncommon to require positional correction in order to be usable. Therefore, to obtain better and more usable location estimates from these data, we refined the locations by fitting a Discrete Correlated Random Walk (DRCW) state space model (SSM) to the Argos locations of each seal (Breed et al., 2006; Jonsen et al., 2003). The DCRW was fitted using Template Model Builder (TMB) (Kristensen et al., 2016).

Dive Data

Proprietary software, Wildlife Computers Dive Analysis Package (WCDAP), extracts depth data from the tags and uses sea surface values to separate depths into individual dives. This software also generates metrics to describe each dive, including dive depth, total dive duration, and time spent at the bottom (of the dive). Dives with a total duration of more than 30 minutes were excluded as they likely represent the erroneous concatenation of several dives (see Beck et al. (2003c)). There were 127 records so excluded, resulting in a total of 194,631 dives for the 1995-1999 period and 1,991,211 dives for the 2009-2015 period. Time spent at the bottom (bottom time) has previously been linked to time spent foraging; grey seals forage at or near the seafloor where they are more likely to encounter

prey (Beck et al., 2003c). Here we defined bottom time as time spent within 85% of the maximum depth for each dive and totalled it for all dives occurring between sequential GPS, (or ARGOS), locations. We use this single number summary as a useful proxy for time spent foraging (Beck et al., 2003c). A summary of the total number of instrumented seals by year, along with the total number of locations and dives is provided in Table 7.1.

Year	M	F	Total	n Loc	n Dive
1995	1	3	4	985	55894
1996	1	3	4	1627	46632
1997	1	2	3	576	13929
1998	1	3	4	2732	30595
1999	2	3	5	1624	47581
2009	5	8	13	87111	228243
2010	5	14	19	115253	300111
2011	0	13	13	118736	318460
2013	2	6	8	72769	211236
2014	2	4	6	72417	160752
2015	1	9	10	113268	277551
Total	21	68	89	587098	1690984

Table 7.1: **Seal location and dive data**. Summary statistics for instrumented grey seals. Reported are the year of data collection (Year), number of males (M) and females (F) and corresponding total (Total). The total number of locations (n Loc) obtained from the respective tracking method (ARGOS for 1995-1999, GPS for 2009-2011 and 2013-2015) and the total number of dives (n Dive) as calculated by WCDAP are also shown.

Prey Distribution and Abundance Data

The Department of Fisheries and Oceans Canada (DFO) has conducted an annual summer ground-fish research vessel (RV) trawl survey each year since 1970. Originally designed to measure distribution and abundance of commercial species, these data also incorporate information on non-commercial species. Focused on the Scotian Shelf and Bay of Fundy (NAFO unit areas 4X, 4W and 4V; Figure 1), the DFO survey uses a stratified random sampling design to sample fish and invertebrates. The data from these surveys have been summarized in various reports (Horsman and Shackell, 2009; Ricard and Shackell, 2013; Smith et al., 2013), stored, and are publicly available in the Ocean Biogeographic Information System

Species	Scientific Name	Years	Trawl Total	Total Abundance
Forage Fish				
Capelin	*Mallotus villosus*	1970-2015	1142	449659
		1995-1999	292	181887
		2009-2015	131	10987
Northern Sandlance	*Ammodytes dubius*	1970-2015	1873	664358
		1995-1999	237	102974
		2009-2015	542	134027
Gadids				
Atlantic Cod	*Gadus morhua*	1970-2015	10447	353944
		1995-1999	1070	40226
		2009-2015	1391	29273
Pollock	*Pollachius virens*	1970-2015	4648	197839
		1995-1999	494	8564
		2009-2015	738	22491
White Hake	*Urophycis tenuis*	1970-2015	5594	89361
		1995-1999	545	8261
		2009-2015	795	6638
Others				
Redfish	*Sebastes sp.*	1970-2015	6312	1835965
		1995-1999	638	170595
		2009-2015	1163	512445
Longhorn Sculpin	*Myoxocephalus octodecemspinosus*	1970-2015	7378	321397
		1995-1999	938	70075
		2009-2015	1131	23510
Assemblage of 7 Total		1995-1999		582582
		2009-2015	2512	739368

Table 7.2: **Prey distribution and abundance data**. Summary statistics for potential grey seal prey species. Trawl Total represents the total number of OBIS trawls in which the species was present. Total Abundance is the sum of individuals present in each of those trawls.

(OBIS) (OBIS, 2014). For each trawl the date, latitude, longitude, depth (m), distance towed(km), as well as the physical characteristics at the location (e.g. water temperature(^{o}C) are recorded.

Grey seals are a generalist predator feeding on a large variety of mainly fish species ($\sim$ 40 species), although, individual animals often obtain most of their energy intake from a small subset of these species (Beck et al., 2007; Bowen and Harrison, 1994). While we did run models on and examine the 20 most likely prey species based upon the diet analyses, we narrowed the list to the seven species by looking to see which produced the best models (via WAIC) for at least some of the time periods examined. Of these 7, all are species regularly preyed upon by grey seals (Longhorn Sculpin, Capelin, Northern Sandlance, Atlantic cod, Pollock, White Hake, Redfish sp.), some have undergone significant changes in abundance over the time period due to environmental changes (Atlantic cod, Capelin). Based on this we selected a set of likely predictive species, see Table 7.2.

In addition to single species data, we aggregated abundance data for the set of likely prey species (Capelin, Northern Sandlance, Atlantic cod, Pollock, White Hake, Redfish sp., Longhorn Sculpin) to obtain an estimate of total available prey assemblage on the Scotian shelf. Catchability is the term used when discussing the relative vulnerability of each species to being caught. Catchability is not constant, we therefore corrected the observed values of each species in the dataset using the species specific correction factors supplied in Harvey et al. (2001), see also (Pinhorn, 1988; Winters and Wheeler, 1985). For all species the abundance data were scaled by subtracting the mean and dividing by the standard deviation.

Physical and Environmental Data

We identified physical covariates thought to be useful for determining how seals use space. These are bathymetry of the Scotian Shelf, bottom substrate composition (particulate size), and water temperature. Bathymetry data are from the US National Oceanographic and Atmospheric Administration (NOAA), ETOPO1 global relief resource (Amante and Eakins, 2009). Bottom composition data (sediment size, log(cm)) was sourced from DFO, temperature data comes from the bottom trawl data set (OBIS), and was recorded at the time and depth of trawl.

7.4 Model Framework

Recall that our response variable of interest is bottom time, which we consider to be a proxy for foraging intensity. We define this as $Y_{s,t}$, the total bottom time at location s at time point t. That is, $y_{s,t}$ is the time the grey seal spends at or near the sea bottom during each dive, aggregated over the period between successive GPS locations, so that at each location s in time t we know how long the seal spent at or near the bottom during the preceding interval. The spatial component of our model requires estimating statistical correlation in the response across the habitat, while accounting for the effects of covariates (Cressie and Wilkie, 2011). We, and others, find Gaussian random fields (GRFs) (Thorson et al., 2014) to be the simplest approach to estimating these spatial effects. GRFs can be efficiently used to estimate spatial correlation (Shelton et al., 2014), i.e. a covariance matrix, Σ, and express it with a simple and interpretable set of parameters (ρ, σ) which

quantify connectivity and spatial variance. For small sample sizes Σ can be calculated so as to estimate these parameters directly. However, this requires inverting Σ, which becomes computationally infeasible with large sample sizes.

We fit our integrated spatio-temporal model for describing grey seal habitat use using INLA, for Integrated Nested Laplace Approximation (Lindgren and Rue, 2015; Rue et al., 2009a) with the aid of the freely available R (R Core Team, 2013) package R-INLA (**www.r-inla.org**). INLA (Lindgren and Rue, 2015; Rue et al., 2009a), approximates the inverse-covariance matrix, (Σ^{-1}), of the GRF using sparse matrices calculated via the stochastic partial differential equation approach (Illian et al., 2012b; Lindgren et al., 2011). This approximation is extremely fast. The approach has been used in animal tracking (Carson and Flemming, 2014; Illian et al., 2012b) and more recently in the marine context (Carson et al., 2017; Munoz et al., 2013; Quiroz et al., 2014), modelling habitat (Boudreau et al., 2017; Pennino et al., 2013), nurseries (Paradinas et al., 2015), and bycatch (Cosandey-Godin et al., 2015; Pennino et al., 2014; Thorson et al., 2014). We extend upon the habitat modelling seen in (Pennino et al., 2013), where the habitat of a single species was examined with environmental covariates by integrating both environmental and prey species data, and modelling both the habitat use and prey distributions simultaneously.

GRFs are usually defined by a mean and a covariance function $Cov[(s),(s')]$ where s and s' represent two locations and h the distance between them ($h = |s - s'|$):

$$C(h;\nu,\kappa) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^{\nu} K_{\nu}(\kappa h). \tag{7.1}$$

The parameters of this Matérn covariance function $C(h;\nu,\kappa)$, are ν and κ, $\nu > 0, \kappa > 0$. (K_{ν} is the modified Bessel function of the second kind). The parameter ν determines smoothness and κ determines spatial scale, and the covariance function depends on the distance separating the locations h. In practice, the parameter ν is usually fixed, (we take $\nu = 1$), and the more easily interpreted form $\rho = \frac{\sqrt{8\nu}}{\kappa}$ is reported with ρ being the distance at which the spatial correlation is reduced to approximately 0.1 (Cameletti et al., 2011, 2013). We will show that formulating the GRF with a Matern covariance function also bears fruit when interpreting results. That is, the parameters defining the fields (ρ, σ), (which are spatial random effects),

have reasonable biological interpretations.

Model Structure

We wish to integrate prey distribution(s) into our model but we do not have the abundance of the prey at the locations of the seals, we have them instead from an oceanic research trawl, and the trawl locations where the prey abundance are observed are different from the locations at which the seal behaviours are observed. Fortunately, our modelling approach is flexible enough to enable this integration, a particularly appealing aspect of our formulation. We begin with a model component for seal bottom time, observed at the ARGOS/GPS locations:

$$y_{s,t} = \alpha_y + \beta_c c_{s,t} + x_{s,t} + e_{s,t}, \tag{7.2}$$

where seal bottom time is modelled as a function of a spatial random effect $x_{s,t}$, and a potential prey (or assemblage) distribution, $\beta_c c_{s,t}$. Bottom time is a continuous random variable and on the natural scale is approximately gaussian distributed. The mean of the response is mapped to a linear predictor η as in the Generalized Additive Model framework Hastie and Tibshirani (1990). That is,

$$E[(Y_{s,t}] = \eta_{s,t} = \alpha_y + \beta_c c_{s,t} + x_{s,t}, \tag{7.3}$$

where the linear predictor is the sum of parts; an intercept (α_y), the prey field and coefficient ($\beta_c c_{s,t}$), and a spatial random effect, ($x_{s,t}$).
The prey distribution data that we wish to integrate into our model requires a second model component:

$$c_{j,t} = \alpha_c + \beta_w w_{j,t} + m_{j,t}, \tag{7.4}$$

where j indexes the locations at which the potential prey species are measured, α_c is an intercept, $w_{j,t}$ is a possible physical covariate (and there may be none or multiple of these), and $m_{j,t}$ is a GRF representing the structure of the covariate(s). This model component (habitat use by the prey species) provides an estimate of prey abundance at all locations, and can in turn be used to estimate prey abundance at the observed locations of the response (s) (seals). If one were instead to simply

estimate values of the prey abundance at the seal locations (s) and plug it in to Equation 7.3 in place of $c_{s,t}$ (equivalent to setting $\sigma_m = 0$ for the prey field) one would not be incorporating the uncertainty associated with these estimates into the integrated model.

Other Covariates

To recap j indexes the locations at which our prey species were observed in the trawl survey and c_j those observations, w_j is a covariate, (Bathymetry, for example), upon which the prey distribution has some dependence. Covariates evaluated in this way were bathymetry, bottom sediment size, and bottom temperature.
The $m_{j,t}$ and $x_{s,t}$ terms are the spatial components of the model. Each of these components, one for the prey species spatial structure and one for the seal's bottom time spatial structure, is a GRF, and represents the effect of latent factors impacting the response not explicitly included in the model, or, their characteristics which compose the spatial and temporal covariance structure. Table 7.3 provides a summary of our integrated spatio-temporal model framework as fully described by Equations 7.3 & 7.4.

α_c	Intercept in the covariate (prey) model
$\beta_w w_j$	Regression coefficient for prey on covariate
m_j	This is a random field, models the spatio-temporal structure for the prey species, has parameters ρ_m and σ_m^2
α_y	Intercept seal bottom time
$\beta_c c_j$	Regression coefficient for dive behaviour on prey
x_i	Another random field, structure in dive behaviour, has parameters ρ_x and σ_x^2

Table 7.3: **Parameters of the Integrated Model**. List of the parameters required for the integrated model.

Priors and Model Assessment

In a Bayesian approach, the parameters that comprise the model are treated as random variables and prior information about the parameters is incorporated by way of a prior distribution. For this study we utilize one of the methods of prior selection enabled in the R-INLA package, "penalized complexity priors" (pc.prior) (Simpson et al., 2015). This method of prior selection seeks to provide weakly informative default priors that "are useful, understandable, conservative, and better

than doing nothing at all" (Simpson et al., 2015). For a spatial random effect; if there is no spatial random effect it is equivalent to having $\rho = \infty$ and $\sigma = 0$, that is, the effect is a constant 0 everywhere. Having a finite ρ and $\sigma > 0$ makes the model more complex, hence the rationale for priors which penalize small values of ρ and large values of σ. To compare and select between different model formulations, various diagnostic measure are possible. We calculated and compared the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), the Watanabe-Aikake Information Criterion (WAIC) (Watanabe, 2009, 2010), and the Logarithm of the Pseudo Marginal Likelihood (LPML) (Geisser and Eddy, 1979). Only the WAIC is presented in the body of this paper, the others are provided in the Supplemental Information.

Subsets

All of the tracking data cannot be considered in a single model. This is partially due to the differing resolution of the Argos and GPS data. When the dive records are summarized over the locations in the Argos data sets, the resulting data are of a coarser nature than the GPS equivalents; instead of locations every 15 minutes, Argos locations are obtained every 6-8 hours. (The DRCW fitted above regularizes the data time). Thus, there are far fewer observations and a concomitant reduction in resolution. We still present Argos results together with GPS based results, since the addition of the Argos results enables an examination of the temporal stability of the ecological footprint. We can run the same model (i.e. an identically parameterized model with the same components; spatial random effect, prey field, physical covariates, etc.) on all the data, but we need to subset our seals due to the resolution difference; the *structure* of the models remains the same, only the resolution changes.

We also experienced memory issues with our computing cluster, (Dalhousie University, Department of Mathematics and Statistics, 2 x Quad Core Xeon Nehalem X5550 2.66Ghz processors with 512GB of RAM) when attempting to run a model with more than 500,000 locations. Thus, we elected to model the seals in five subsets; subdivided by time and sex. That is, we model male and female seals separately, males in two subsets, 1995-1999 (Argos) and 2009-2015 (GPS); females in

three subsets, 1995-1999 (Argos), 2009-2011 (GPS) and 2013-2015 (GPS). (There is no suitable 2012 data). We note that we have explicitly split the male and female seals into separate models; because we fully expect the male and female seals to be distinct in their behaviour based on previous tracking studies (Breed et al., 2006), and here seek to examine and interpret the differences we know *a priori* to exist, rather than attempt to establish such sex based distinctions.

Models were run sequentially, each of the seven candidate prey species with each of the three possible physical covariates, as well as the assemblage of prey with each of the covariates. This was repeated for each of the subsets for a total of 120 separate model runs. Model performance was compared using our diagnostic criterion, WAIC. We also tried including covariates in Equations 7.3 & 7.4, but in all cases, once the effect of the covariation the prey field was accounted for, the covariates had no further effect on the response.

Parameter interpretation

Our modelling framework involves two spatial covariance structures (one for seal bottom time x, one for prey abundance m) and a set of covariates that best describe the response. Each spatial effect has a range parameter (ρ) and a standard deviation (σ). Additionally the coefficient (β_c) quantifies the dependence between these effects; the estimated dependence of seal bottom time on prey abundance. Recall, a null hypothesis of no spatial effect implies a very long range and a small standard deviation, (infinite range and zero variance is perfectly flat).

The intercepts, (α_y, α_c) and the regression coefficients for prey on covariates, $(\beta_w w_j)$, have conventional interpretations but the regression coefficient for dive behaviour on prey, $(\beta_c c_j)$, merits some explanation. This coefficient is expressing the contribution of the estimated abundance of the prey species at the location where the seal was observed, $\beta_c c_s$, which was estimated using prey numbers observed at a different set of locations j. The model handles this smoothly, and an estimate of the mean of the response or the abundance of prey is available at any location in the study area, not just at the observed locations of either. Equation 3indicates that the value which goes into Equation 2 depends on an intercept, a covariate, and a random effect. Leaving aside the fixed effects of the intercept and

all else being equal with respect to the covariate and time, the contribution to the expected response is $\beta_c m_j$, where m_j is the random effect in the prey model. This value is from the multivariate normal distribution which models the spatial effect. At location j it will be $\sigma_m \Sigma_j$, i.e. a standard normal score times the standard deviation of the prey field. Since we standardized both response and prey, $\beta_c m_j$ has an interpretation of standard deviation in the response per standard deviation of the prey field, all else being equal.

Results

Females

Of the 7 models fitted to single species of prey, the sandlance model had the lowest WAIC for females in all time periods. However, for all three time periods the assemblage of all 7 species had the lowest WAIC compared to any of the individual prey (Table 7.4).

	Females 95-99	Females 09-11	Females 13-15	Males 95-99	Males 09-15
Source	Argos	GPS	GPS	Argos	GPS
Capelin	-57411	-2307267	Null	-45903	-1066265
Cod	-24915	-2756027	Null	-29760	**-1549705**
Haddock	-32098	-437853	-1256635	**-59726**	3602362
Pollock	-23264	-2742607	-23033156	-56988	-1547987
Redfish	-35785	-2745993	-2303971	-23794	-1538432
Sandlance	-1511537	-2754107	-2312186	-50415	-57420
Longhorn Sculpin	-49758	-2751057	-2309671	-48906	-1516762
Assemblage 7	**-60768**	**-2759966**	**-2317144**	-52186	-1537513

Table 7.4: **Diagnostic Criteria**. Comparison of WAIC values. Other diagnostics may be found in supplements

	ρ_m	σ_m	ρ_x	σ_x	$\beta_c c_j$
Argos Females 1995-1999 (S/L)	.34(.28,.41)	7.7(6.61,8.85)	7.38(3.05,17.7)	.36(.16,.79)	.023(.007,.036)
Argos Females 1995-1999 (A7)	.38(.29,.55	7.45(6.29,7.29)	4.7(2.34,9.31)	.63(.33,1.2)	.001(-.0011,.0013)
GPS Females 2009-2011 (S/L)	.15(.14,.16)	55.53(55.48,55.57)	.78(.62,1.01)	1.08(.88,1.34)	.002(.001,.002)
GPS Females 2009-2011((A7)	.17(.16,.18)	51.31(49.16,54.25)	.49(.43,.56)	.74(.67,.84)	.001(.0095,.0014)
GPS Females 2013-2015 (S/L)	.17(.16,.19)	36.5(35.1,38.4)	.60(.27,.92)	.69(.44,.94)	.002(.001,.004)
GPS Females 2013-2015 (A7)	**.14(.12,.16)**	**36.5(34.9,37.8)**	**.46(.39,.55)**	**.59(.53,.68)**	**.002(.001,.003)**
Argos Males 1995-1999 (Had)	.18(.17,.19)	70.66(66.85,73.67)	5.25(2.48,10.22)	.49(.24,.92)	.002(.0005,.003)
GPS Males 2009-2015 (Cod)	.11(.10,.12)	55.1(52.4,57.1)	.57(.46,.67)	.75(.65,.86)	.002(.001,.002)

Table 7.5: **Random Field Parameters**. Values characterizing the spatial effects for predator and prey. (95% Credible intervals)

For the sandlance only models based on GPS data (2009-2011 and 2013-2015), the range parameter for prey abundance (ρ_m, distance in degrees) was similar in

both time periods (Table 7.5) and represented distances of approximately 16 and 18 km for 2009-2011 and 2013-2015, respectively. The estimated spatial standard deviation (σ_m) was greater for 2009-2011 compared to 2013-2015 (Table 7.5) suggesting that there was greater variation in sandlance abundance between areas of high and low abundance in 2009-2011 compared to 2013-2015. Considered together the values for the range parameters and spatial standard deviations for sandlance abundance for these two time periods describe a spatial effect with appreciable structure (short range and large amplitude). Such a field is consistent with small areas of high abundance and very large differences in abundance between areas of high and low sandlance occurrence (see Figure 7.2 c and Figure 7.2 e).

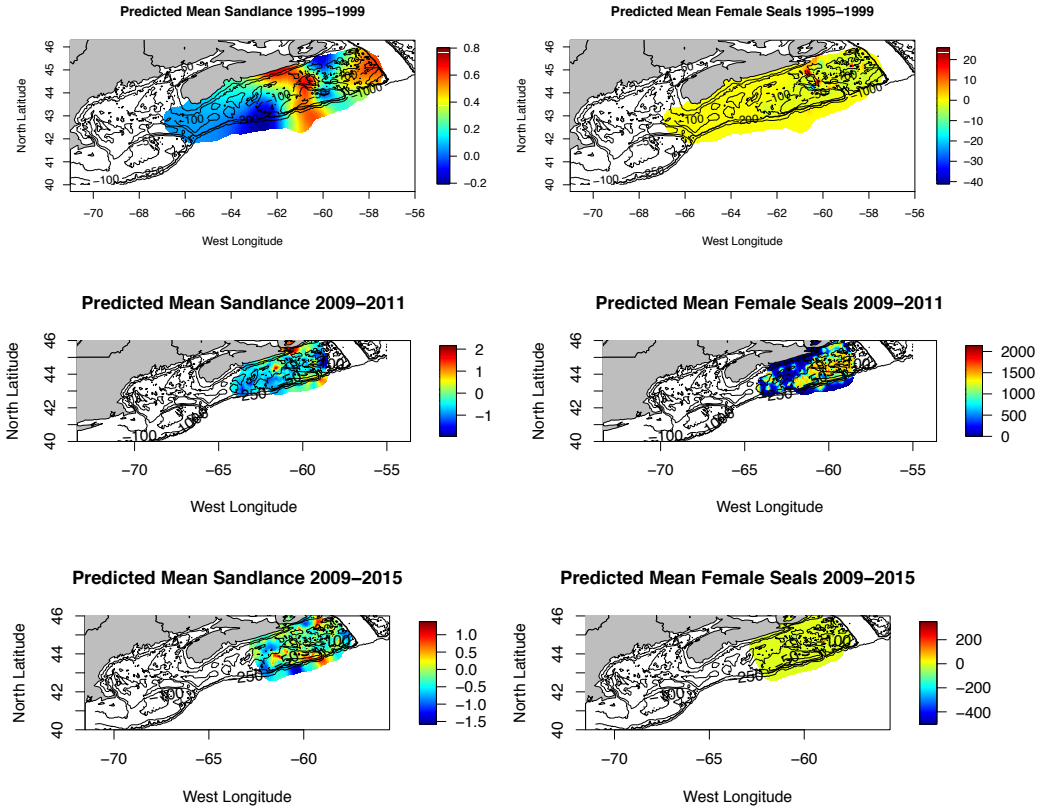

Figure 7.2: **Random Fields: Females and Sandlance** Random field for Sandlance Abundance (left) and Bottom Time (Right).

In contrast, the parameters for the spatial effect for the seals' bottom time for the

sandlance only models based on GPS data (2009-2011 and 2013-2015) describe a spatial structure which is flatter and smoother than the field for sandlance abundance. The range parameters for bottom time (ρ_x) were larger than those for prey abundance, representing distances of approximately 86 and 65 km for 2009-2011 and 2013-2015, respectively (Table 7.5). For both time periods the spatial standard deviation was considerably smaller than those for prey abundance (Table 7.5), Considered together the values for the range parameters and spatial standard deviations for bottom time for these two time periods suggest that bottom time is dependent across wider distances on the Scotian Shelf compared to sandlance abundance, and the high and low areas are more similar (Figure 7.2d and Figure 7.2e).

The coefficient for prey abundance ($\beta_c c_j$) for the sandlance only models based on GPS data (2009-2011 and 2013-2015) were credibly non-zero for both time periods (Table 7.5), indicating a positive relationship between the bottom time of grey seal females and sandlance abundance. Letting all else be equal a value of the prey random effect 1 standard deviation above the mean would give an expected increase in the response of about 0.111 (2009-2011) and .07 (2013-2015) standard deviations, an effect size of an additional 1.2 (0.6, 1.8) and 0.7 (0.35, 1.03) minutes of bottom time per 15-minute GPS observation interval, respectively.

Both bottom substrate and trawl temperature had a significant impact on the abundance of sandlance; the negative signs of the respective coefficients for bottom substrate ($\beta_w w_j$), , -.1123 (-.1150,-.1095), and for temperature, -0.0322 (-.032,-.0317), suggest that sandlance prefer colder water and small substrate sizes (sediment is measured as the log(diameter(cm)) of the particle). Once the effect of sediment on the sandlance abundance is taken into account, substrate has no further contribution to seal bottom time, and the coefficient limits include zero, 0.0001 (-.0002,.0005).

While the random field for sandlance is similar between 2009-2011 and 2013-2015 (see Figure 7.2c and Figure 7.2e), the latent variable for the seal bottom time does show some differences between periods (see Figure 7.2d and Figure 7.2f). The field is highest close to Sable Island in 2009-2011, suggesting that the seals were foraging closer in the earlier period.

Comparing the models based on the assemblage of all 7 prey species to the sandlance only models for the 2009-2011 and 2013-2015 GPS data, the results for the random field parameters were very similar (Table 7.5).

Because of the lower resolution in the Argos data (1995-1999) the parameter ranges and standard deviations are not directly comparable to the values obtained from the models which used GPS data. However, consistent with the results from the GPS models, the range parameter for bottom time was larger and the spatial standard deviation smaller for bottom time compared to the values for prey abundance for both the sandlance only model and the prey assemblage model (Table 7.5). Similar to the results for the GPS data this indicates that spatial field for prey abundance had more structure (shorter range and larger amplitude) than the field for seal bottom time (Figure 7.2a and Figure 7.2b).

Consistent with the GPS results, the coefficient for prey abundance ($\beta_c c_j$) was non-zero for both Argos models (Table 7.5). Letting all else be equal, a value of a sandlance random effect 1 standard deviation above the mean would give an expected increase in the response of about .177 standard deviations, an effect size of an additional 66 (21, 110) minutes of bottom time per 8 hour Argos observation interval, or 2.06 (0.67, 3.42) minutes of bottom time per 15 minutes.

As in the GPS models, both bottom substrate and trawl temperature had a significant impact on the abundance of sandlance in the Argos model; the negative signs of the respective coefficients, -0.0168 (- 0.0182, -0.0154) for bottom substrate, and, -0.1212 (-0.1216, -0.1208) for temperature suggest that sandlance prefer colder water and small substrate sizes. Once the effect of sediment on the sandlance abundance is taken into account, substrate had no further contribution to seal bottom time, and the coefficient was not non zero, -0.0087 (-0.0399, 0.0224).

The spatial random effects for the Argos models show distinct structure in the sandlance in the years 1995-1999 (Figure 7.2a). After accounting for temperature and sediment the prey field reveals high concentrations of sandlance around Sable Island and extending NNW from it. There is a second area of higher concentration to the ENE, on the Banqureau bank. The spatial random effect for the seal bottom time aligns closely with this pattern (Figure 7.2d), especially the band extending NNW from Sable.

Males

Of the 7 models fitted to single species of prey, the models with the lowest WAIC for males in in the 2009-2015 GPS period and 1995-1999 Argos were those with cod and haddock respectively (Table 7.4).

For the cod only model based on GPS data (2009-2015), the range parameter for prey abundance had a short range (12 km) and a relatively large spatial standard deviation consistent with small areas of high abundance and very large differences between areas of high and low cod abundance (Figure 7.3c). Like the results for the females with sandlance, the spatial effect for the male seals seals bottom time was smoother and flatter than the field for cod abundance and showed a range consistent with that of the females over the same time period (59 km). This suggests that the bottom time of the male seals, like the females, is similar over large distances, and that the areas of high and low bottom time are similar (Figure 7.3d).

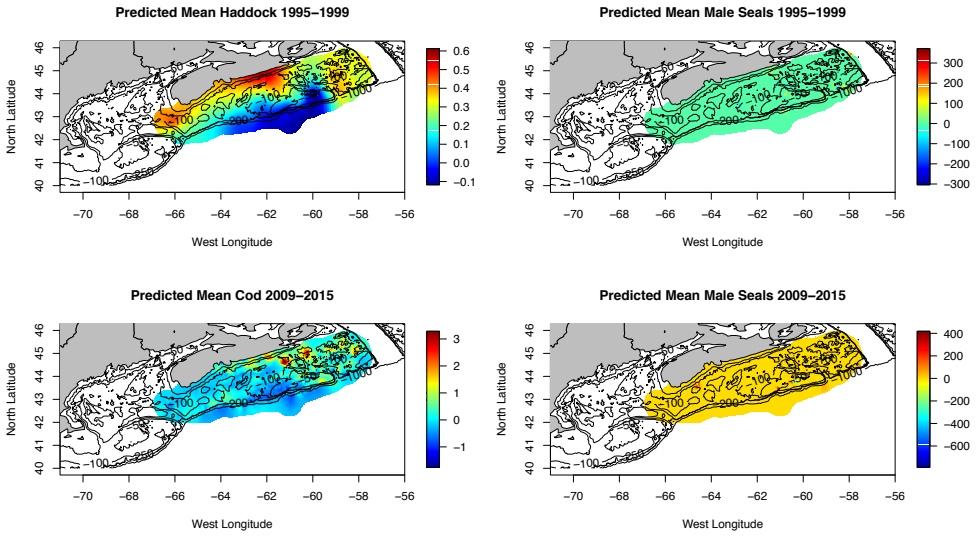

Figure 7.3: **Random Fields: Males and Cod and/or Haddock** Random field for Cod and/or Haddock Abundance (Left) and Bottom Time (Right).

The coefficient for prey abundance ($\beta_c c_j$) for the cod only models based on GPS data (2009-2015) were credibly non-zero (Table 5). A value for the cod random effect 1 standard deviation above the mean would give an expected increase the

response of about .110 standard deviations, an effect size of an additional 1.28 (0.6, 1.37) minutes of bottom time per 15-minute GPS observation interval.

Both bottom substrate and trawl temperature had a significant impact on the abundance of cod in the model for the GPS data. The coefficient for substrate, .135 (.132,.139), suggested that cod preferred coarser substrates. The coefficient for temperature ,0.0153 (.0152,.0154), suggests a positive relationship between cod abundance and temperature. Once the effect of sediment on the cod abundance was taken into account, substrate had no further contribution to seal bottom time, and the coefficient limits included zero, ,0.0015 (-.0046,.0075).

Because of differences in the data resolution, the parameter ranges and standard deviations for the males modelled with Argos data (1995-1999) are not directly comparable to the values obtained from the models which used GPS data (2009-2015). However, like the results for the GPS model, the range parameter for bottom time was larger and the spatial standard deviation smaller compared to the values for prey abundance for the favoured haddock only model (Table 7.5). These results indicate that, similar to the GPS model, the spatial field for prey abundance had more structure (shorter range and larger amplitude) than the field for male seal bottom time in the Argos model (Figure 7.3a and Figure 7.3b).

Consistent with the GPS results, the coefficient for prey abundance ($\beta_c c_j$) was non-zero for the Argos model (Table 7.5). Letting all else be equal, a value of the haddock random effect 1 standard deviation above the mean would give an expected increase the response of about .141 standard deviations, an effect size of an additional 52 (13, 78) minutes of bottom time per 8-hour Argos observation interval, (1.63 (0.4, 2.43) minutes of bottom time per 15 minutes).

Both bottom substrate and trawl temperature had a significant impact on the abundance of haddock in the model for the Argos data. The coefficient for substrate, -.1276 (-1277.,-.1274), suggested that haddock preferred smaller substrates,. The coefficient for temperature ,-0.0309 (-.031,-.0308), suggests a negative relationship between haddock abundance and temperature. Once the effect of sediment on the haddock abundance was considered, substrate had no further contribution to seal bottom time, and the coefficient was not non zero, ,-0.0058 (-.0036,.0247).

In summary, we find that, regardless of the model, the preferred prey are distributed across the ESS as a function of the physical covariates, but that there is still a substantial latent variable effect. Sediment and temperature are informative, but do not entirely determine the prey distribution and so other factors, here subsumed into the latent variable, must also play a role. For the seals, the bottom time responds to the distribution of available prey, and once this has been accounted for very little latency remains.

Discussion

Several key findings emerge from our analysis of the spatio-temporal distribution of grey seal diving behaviour. First is the stability of the footprint of the grey seal behaviour in both time and space. Despite both changes in the structure and functioning of the ESS ecosystem used by seals and the large increase in population size over the course of our study, the seals displayed little change in their pattern of foraging, as represented by the distribution of bottom time. Second, this pattern of bottom time appears to depend mostly upon the distribution of available prey, although the most influential prey varies by sex. Females show a strong association with the distribution of Northern sandlance, although the association with an assemblage of available prey is even stronger. By contrast, males display associations with two Gadid species (cod or haddock) depending on the time period examined. Furthermore, in the case of males, and unlike females, this single species association was stronger than that of the assemblage of prey. Third, for both sexes and times periods, he association of the bottom time is influenced by the available prey and not environmental covariates. However, environmental covariates did strongly influence the prey species distribution. Thus, the physical environment had only an indirect influence on the spatio-temporal footprint of the seals by influencing the distribution of prey. Finally, the integrated spatio-temporal modelling approach we propose here demonstrates that we can explicitly visualize and quantify the behaviour (in this case diving, proxy for foraging) of predators, throughout the species distribution and estimate how this behaviour is influenced by the distributions of prey species. Of course, there are some caveats. The seals in this study were all captured on Sable Island, and all foraged on the

ESS. There are other seal colonies in eastern Canada, along the south shore of Nova Scotia and in the Gulf of St. Lawrence, and seals which breed in these colonies may not be fully represented here, even if they were to leave these areas and forage on the ESS. Further, seals which breed on Sable, but which leave the ESS, are also not represented. Nevertheless, were data available from these other colonies they could be incorporated into the model.

Given the wide range of prey species consumed by grey seals (Bowen and Lawson, 1993), they are regarded as a generalist predator, but there is a clear difference between the ecological footprint of the male and female seals. Although sex-specific differences in diet and distribution have been previously demonstrated (Beck et al., 2003b; Breed et al., 2006), our results provide a better understanding of the ecological factors driving those differences by revealing how physical environmental covariates influence prey distributions which in turn influence the distribution of seal diving behaviour.

As a generalist predator, grey seal foraging might be expected to show relationships with the abundance of prey, without regard to type of prey; that is, to the number of prey items encountered. Our results indicate support for this prediction in the case of female grey seals, where the prey assemblage seems to influence the distribution of diving more then any single species. However this was not the case in males. Conversely, some suggest that a preference may be seen in a generalist for the most abundant type of prey, (Jaworski et al., 2013), since that will be the prey most commonly encountered. Again, females provide some support for this prediction given that the sandlance model was a close second to the assemblage and is numerically the most abundant prey of those we investigated (see below).

In females the most predictive prey species were the forage fish species (Sandlance/Capelin), with Sandlance the single best, for all 3 periods examined. Sandlance is also the most abundant prey species (counts) for all 3 periods. The mean amount of bottom time predicted by our model for female seals increases with the increasing abundance of Northern sandlance and does so across the entire time span covered in this study. As noted above we saw increases in our forage proxy behaviour per given increase of sandlance of 8.1 minutes/hour 1995-1999, 4.8 minutes per hour 2009- 2011, and 2.8 minutes per hour in 2013-2015. This suggests

that Northern sandlance has been, and remains an important prey species for female grey seals, even as the numbers of grey seals exploiting the sandlance prey has increased greatly. Sandlance seem to have benefited from the ecosystem shift on the Scotian shelf (Frank et al., 2005; Shackell et al., 2010), and their numbers have increased. However, the strength of the association seems to be declining over time, +8.1 m/hr=>+4.8 m/hr=>2.8 m/hr: while far from conclusive, this is suggestive that the relative importance of sandlance as prey for the female grey seal may be reduced. This is particularly interesting when examining the relationship between these female seals and the assemblage of all 7 species over the same time period. The strength of the association with the assemblage shows no trend over time, +4.4 m/hr=>+2 m/hr=>2.8 m/hr: overall the effect size is slightly smaller than that of sandlance even though goodness of fit is better. This impression is reinforced by examining the spatial random effects (see Figure 7.2). In the earliest 1995-1999 period, there is a notable coincidence of the latent effect for prey and the latent effect for seal bottom time. The female seals were showing high levels of bottom time where the sandlance were abundant. In fact the bottom time was high even after accounting for the sandlance abundance being high, hence both random effects are high simultaneously and coincidentally. The same held true in the second 2009-2011 time period, but not to such a high degree. One could take this as meaning the female seals are still cueing on the high levels of sandlance, but perhaps not so exclusively as previously. By the third 2013-2015 time period the seal spatial random effect is almost completely flat. Bottom time still depends on sandlance abundance, but through the modelled dependency (β). It certainly looks like the dependence or preference for sandlance is reducing and the female seals are cueing less exclusively upon them. Again, far from conclusive, but suggestive that the relative importance of sandlance as prey for the female grey seal may be reduced. This would be consistent with the seals widening their foraging as their numbers increase away from the heavy competition where sandlance is most abundant, and younger seals forming preferences for other, abundant, prey species.

In males, the most predictive prey species are the members of the Gadids, (Haddock/Pollock/Cod). Exactly which species is best varies by time period. The

male seals show less bottom time where there are more Sandlance. The same is true for the females, the Cod has an almost opposite effect on the females as does Sandlance, although in the female model it just misses significance. It has been shown from dietary and tracking studies that grey seals have sex differentiated diets (Beck et al., 2003b,c, 2005, 2007; Breed et al., 2006, 2009), with females preferring the smaller forage and out competing the males. Here we can relate the spatial variation of foraging behaviour directly to the spatial distribution of prey abundance and quantify the effect of prey density on the foraging behaviour. This supports this contention since our model shows a sex-based difference in foraging behaviour that is prey species dependent; supporting the suggestion that females prefer the smaller prey of higher energy density) (e.g., Sandlance), whereas males prefer larger lower energy density prey, given their larger body size and digestive physiology (Beck et al., 2003b). We can see from this study the increased foraging effort expended by the seals as a function of these exact prey species.

In all cases the predator field is very flat and smooth, indicating little structure. Most of the structure that is present is in the assemblage (Sandlance predominating) for the females, and in the Gadid fishes for the males. This means that, once Sandlance and the physical covariates are accounted for, there is relatively little residual covariance structure left in the females space use distribution (the latent variable component of the model contributes relatively little), and, once the the physical covariates and a best Gadid is accounted for (which Gadid is time variant) there is relatively little residual structure left in the male space use distribution. This means that the seals respond to the available prey distributions, and that these prey distributions respond to the physical environment. The seals are not responding to depth or temperature when they dive on their preferred forage spots, they dive there because there is prey there.

The nature of these relationships has persisted over time; the predator field remained relatively smooth, for both sexes, for all time periods, despite the profound changes in structure and relative abundance of prey on the ESS (Shackell et al., 2012). The ESS experienced a profound shift away from large bodied predatory fish towards forage fish and invertebrates (Bundy, 2005). The biomass of the forage fish increased by as much as 900%, while the large predatory fish (cod being the prime

example) collapsed to as little as 5% of previous abundance (Frank et al., 2011). This shift in the ecosystem brought with it instability: the collapse of the large fish was triggered by overfishing, the forage fish, released from predation boomed to well beyond carrying capacity (Frank et al., 2011) and then oscillated as new equilibria were achieved. This shift to forage fish favoured the seals as they (especially females) clearly forage upon them, but the pattern of spatial usage displayed by the seals themselves remained unaffected. To some extent this stability may reflect the way in which grey seals, and other phocid seal and perhaps other taxa of marine predators dive. Grey seals dive to the bottom while travelling and while foraging. Thus, by using bottom time as a metric we might expect a smoother distribution of diving than if we had a better measure of when seals were foraging. By contrast, fur seals and sea lion, marine birds, and many cetacean species dive deeply only mainly during foraging otherwise travelling near the surface. In such species, we might predict that the patterns of diving to more closely reflect the spatio-temporal variability in the distribution of prey. However, this relatively smooth pattern of diving behaviour shown by grey seals may also suggest a form of bet-hedging, whereby potential new patches of prey are continually being search for on the way to known patches. Indeed the female patterns show more variation than the males, even though the females preferred forage food, the Sandlance, was increasing; again implying that the males may be even more general in their prey choices.

This modelling approach itself is quite attractive. We have here integrated a substantial quantity of data; each of the GPS models incorporates several hundred thousand seal locations, each with associated dive measures, multiple environmental covariates and, the survey trawl data for multiple prey species. None of these data sets were reduced by binning . The original locations were retained for both the seals and the prey, even though these locations were different. The resulting models are relatively simple to interpret; the latent effects being specified by only two parameters. Perhaps the most appealing aspect of our modelling framework is its ability to incorporate spatial information pertaining to both the predator and the prey into the same model without the need for the binning or averaging. All the spatial information contained in the tracks is retained in this model.

Contrast that with models using grid cells to regularize space. For example, in (Matthiopoulos et al., 2004) such a grid is used, and inference is performed at a finite set of locations; the centres of each cell. This is not such a large concern with Argos locations, where the observed locations are uncertain, but loses much when the resolution of GPS is available. Modelling a behaviour, such as bottom time, by grid cell would not seem optimal; the behaviour, like the maritime environment itself, is patchy and constitutes a local phenomena. Other methods, such as kernel smoothers (Matthiopoulos, 2003), could estimate a continuous surface, (much like our model does; a response at every location), but the resulting smooth is not so easily interpreted. Had we, for example, simply used some sort of average for the abundance of the prey species, (by grid cell, or a locally weighted average by kernel smoother), the variability inherent in the prey measurements would be lost. Here, the uncertainty is retained and the GMRF representation has parameters which may be interpreted to estimate just how smooth the surface is, how variable it is, or to compare surfaces if more than one exists in the model. In this way, we showed that the prey and assemblage fields in our models are much more highly structured than those of the predator. The connection between the prey fields and the physical covariates and between seal behaviour and prey distributions is made explicit. This suggests that models attempting to use physical oceanographic covariates may not provide the best predictions of habitat quality or the spatio-temporal distribution of foraging hotspots.

Acknowledgments

We extend our thanks to Dr. Greg Breed and Dr. Cornelia den Heyer for reading earlier versions of this manuscript and for their thoughtful commentary and wise council.

7.A Supplementary Model Details: Diagnostics

GPS Females 2013-2015

Prey Species	DIC	LPML	WAIC
Capelin	-2019218	Inf	NULL
Cod	-2258356	Inf	NULL
Haddock	-1751547	Inf	-1256635
Pollock	-2252016	-1000531	-2303156
Redfish	-2257741	-1003836	-2303971
Sandlance	-2260635	-1004637	**-2312186**
Longhorn Sculpin	-2260563	-1005138	-2309671
Assemblage 7	-2272870	-1011186	**-2317144**

Table 7.A.1: **Diagnostic Criteria**. Comparison of diagnostic criteria values for various prey candidate species for the 2013-2015 female seals

GPS Females 2009-2011

Prey Species	DIC	LPML	WAIC
Capelin	-2469206	Inf	-2307267
Cod	-2697883	-1199833	-2756027
Haddock	-1558759	Inf	-437853
Pollock	-2685706	-1194336	-2742607
Redfish	-2694577	Inf	-2745993
Sandlance	-2694727	-1198115	**-2754107**
Longhorn Sculpin	-2694065	-1198508	-2751057
Assemblage 7	-2708216	-1205412	**-2759966**

Table 7.A.2: **Diagnostic Criteria**. Comparison of diagnostic criteria values for various prey candidate species for the 2009-2011 female seals

Argos Females 1995-1999

Prey Species	DIC	LPML	WAIC
Capelin	-60631.31	-27582.81	-57411.19
Cod	-56142.53	Inf	-24915.42
Haddock	-63651.34	Inf	-32098.68
Pollock	-47409.36	Inf	-23264.2
Redfish	-55929.3	Inf	-35785.76
Sandlance	-59806.11	-26995.37	-57420.63
Longhorn Sculpin	-68311.21	Inf	-49757.98
Assemblage 7	-59441.20	NA	**-60768.52**

Table 7.A.3: **Diagnostic Criteria**. Comparison of diagnostic criteria values for various prey candidate species for the 1995-1999 female seals

GPS Males 2009-2015

Prey Species	DIC	LPML	WAIC
Capelin	-1269458	Inf	-1066265
Cod	-1514642	-677113	-1549705
Haddock	991347.5	Inf	3602362
Pollock	-1515061	-674287	-1547987
Redfish	-1517155	-676782	-1538432
Sandlance	-1511446	-672926	-1511537
Longhorn Sculpin	-1516016	-674761	-1516762
Assemblage 7	-1530178	Inf	-1537513

Table 7.A.4: **Diagnostic Criteria**. Comparison of diagnostic criteria values for various prey candidate species for the 2009-2015 male seals

Argos Males 1995-1999

Prey Species	DIC	LPML	WAIC
Capelin	-51254.52	-23746.30	-45903.61
Cod	-47860.03	Inf	-29760.71
Haddock	-60986.47	-27305.52	-59726.56
Pollock	-55787.41	-24837.93	-56988.77
Redfish	-43974.10	Inf	-23794.10
Sandlance	-51601.89	-23196.77	-50415.15
Longhorn Sculpin	-55604.69	-25561.29	-48906.02
Assemblage 7	-51034.63	NA	-52186.07

Table 7.A.5: **Diagnostic Criteria**. Comparison of diagnostic criteria values for various prey candidate species for the 1995-1999 male seals

7.B Supplementary Model Details: Female Seal Models

GPS Females 2013-2015: Sandlance

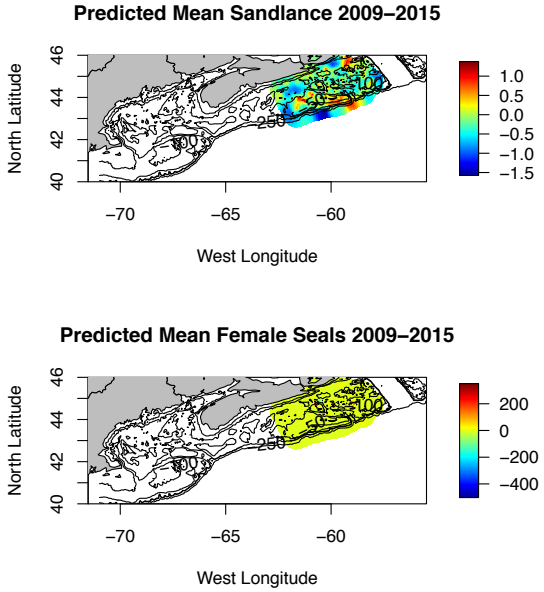

Figure 7.B.1: **GPS Females 2013-2015: Sandlance** Random field for Sandlance and GPS Females 2013-2015.

Sandlance F 13-15	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.174	0.008	0.159	0.173	0.191	0.172
Stdev for m	36.507	0.862	35.096	36.401	38.437	36.102
Range for x	0.602	0.176	0.271	0.605	0.921	0.606
Stdev for x	0.694	0.130	0.439	0.698	0.937	0.718
Beta for o.c	0.002	0.001	0.001	0.002	0.004	0.002

Table 7.B.1: **Parameters Values of the Latent Fields**. GPS tracked Female seals 2013-2015 using Sandlance as the prey species.

GPS Females 2009-2011: Sandlance

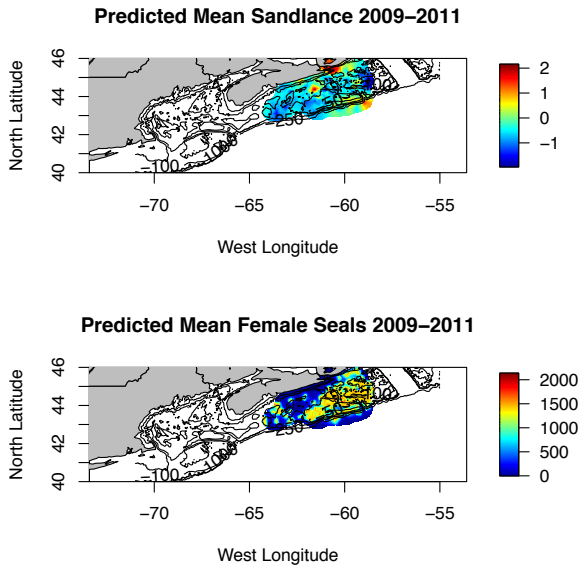

Figure 7.B.2: **GPS Females 2009-2011: Sandlance** Random field for Sandlance (top) and GPS Females 2009-2011 (bottom).

Sandlance F 9-11	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.153	0.000	0.153	0.153	0.153	0.153
Stdev for m	55.525	0.024	55.481	55.524	55.573	55.522
Range for x	0.782	0.101	0.616	0.770	1.010	0.742
Stdev for x	1.082	0.118	0.881	1.071	1.344	1.042
Beta for o.c	0.002	0.000	0.001	0.002	0.002	0.002

Table 7.B.2: **Parameters Values of the Latent Fields**. GPS tracked Female seals 2009-2011 using Sandlance as the prey species.

Argos Females 1995-1999: Sandlance

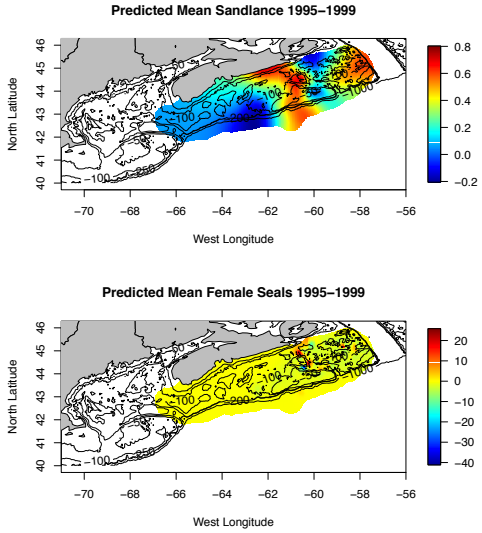

Figure 7.B.3: **Argos Females 1995-1999: Sandlance** Random field for Sandlance (top) and Argos Females (bottom). Hot spots for the seals along the line from Sable to the NW. Aligns well with Sandlance abundance, but ignores the Sandlance on Banquereau Bank ENE of Sable

Sandlance F 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.342	0.034	0.279	0.341	0.413	0.340
Stdev for m	7.695	0.569	6.609	7.686	8.846	7.683
Range for x	7.383	3.940	3.048	6.327	17.722	4.841
Stdev for x	0.356	0.167	0.158	0.315	0.789	0.252
Beta for o.c	0.023	0.007	0.007	0.023	0.036	0.024

Table 7.B.3: **Parameters Values of the Latent Fields**. Argos tracked Female seals 1995-1999 using Sandlance as the prey species.

GPS Females 2013-2015: Cod

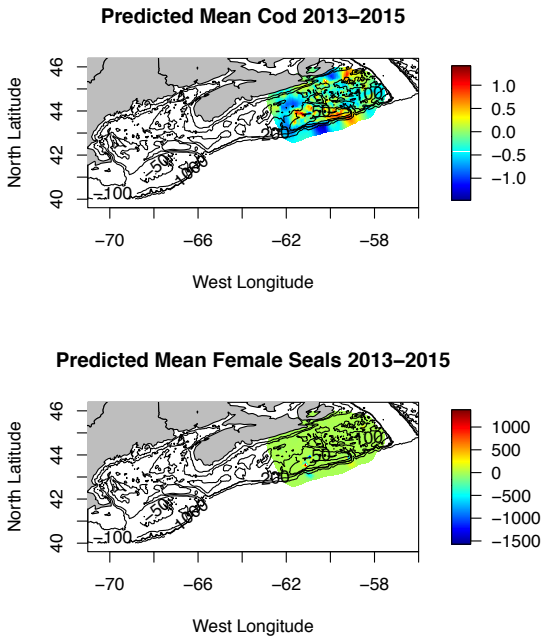

Figure 7.B.4: **GPS Females 2013-2015: Cod** Random field for Cod (top) and GPS Females 2013-2015 (bottom).

Cod F 13-15	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.137	0.004	0.130	0.137	0.146	0.136
Stdev for m	118.570	2.574	113.690	118.501	123.798	118.323
Range for x	0.542	0.095	0.388	0.531	0.760	0.506
Stdev for x	0.794	0.104	0.613	0.785	1.021	0.767
Beta for o.c	-0.001	0.000	-0.002	-0.001	0.000	-0.001

Table 7.B.4: **Parameters Values of the Latent Fields**. GPS tracked Female seals 2013-2015 using Cod as the prey species. Provided for comparison to Sandlance

GPS Females 2009-2011: Cod

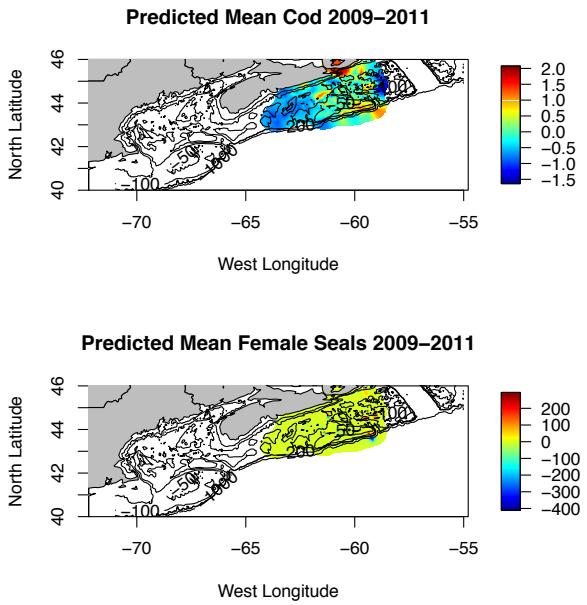

Figure 7.B.5: **GPS Females 2009-2011: Cod** Random field for Cod (top) and GPS Females 2009-2011 (bottom).

Cod F 9-11	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.145	0.000	0.145	0.145	0.145	0.145
Stdev for m	88.644	0.037	88.563	88.648	88.706	88.663
Range for x	0.569	0.133	0.310	0.573	0.811	0.590
Stdev for x	0.832	0.157	0.521	0.838	1.121	0.866
Beta for o.c	0.000	0.000	0.000	0.000	0.001	0.000

Table 7.B.5: **Parameters Values of the Latent Fields**. GPS tracked Female seals 2009-2011 using Cod as the prey species. Provided for comparison to Sandlance

Argos Females 1995-1999: Cod

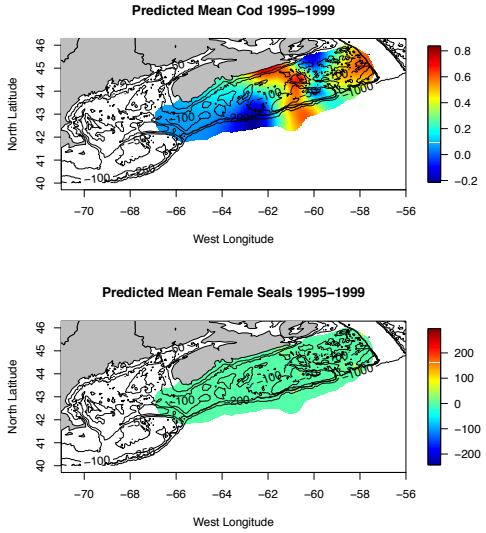

Figure 7.B.6: **Argos Females 1995-1999: Cod** Random field for Cod (top) and Argos Females (bottom).

Cod F 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.106	0.021	0.064	0.108	0.140	0.116
Stdev for m	29.026	1.207	26.391	29.157	31.024	29.714
Range for x	4.588	1.839	2.274	4.163	9.311	3.465
Stdev for x	0.651	0.207	0.368	0.608	1.165	0.532
Beta for o.c	0.000	0.003	-0.005	0.000	0.007	0.000

Table 7.B.6: **Parameters Values of the Latent Fields**. GPS tracked Female seals 1995-1999 using Cod as the prey species. Provided for comparison to Sandlance

GPS Females 2013-2015: Assemblage of 7

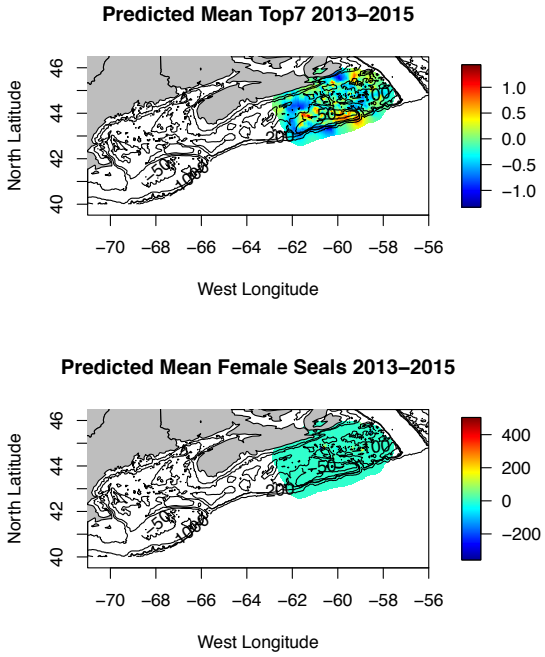

Figure 7.B.7: **GPS Females 2013-2015: Assemblage of 7** Random field for Assemblage and GPS Females 2013-2015.

Assemblage7 F 13-15	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.138	0.009	0.123	0.137	0.157	0.134
Stdev for m	36.494	0.742	34.904	36.553	37.798	36.781
Range for x	0.456	0.041	0.388	0.452	0.549	0.439
Stdev for x	0.594	0.040	0.525	0.591	0.681	0.582
Beta for o.c	0.002	0.001	0.001	0.002	0.003	0.003

Table 7.B.7: **Parameters Values of the Latent Fields**. GPS tracked Female seals 2013-2015 using Assemblage as the prey species.

GPS Females 2009-2011: Assemblage of 7

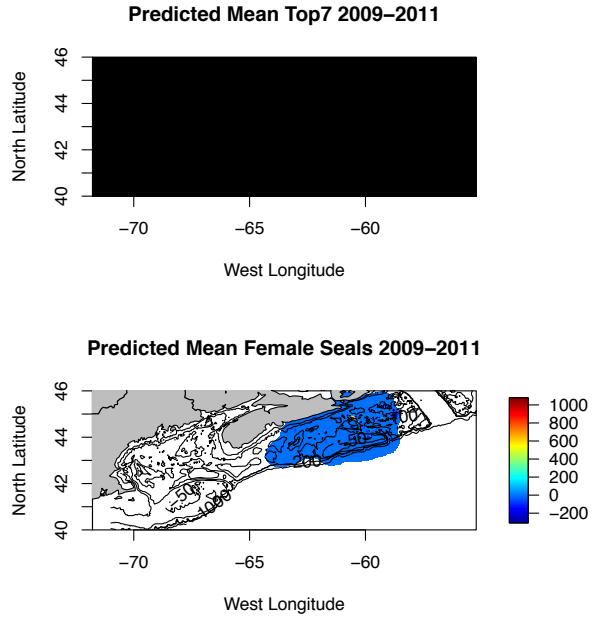

Figure 7.B.8: **GPS Females 2009-2011: Assemblage of 7** Random field for Assemblage and GPS Females 2009-2011.

Assemblage7 F 09-11	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.170	0.004	0.160	0.170	0.177	0.172
Stdev for m	51.308	1.298	49.159	51.147	54.245	50.601
Range for x	0.487	0.032	0.433	0.484	0.557	0.475
Stdev for x	0.744	0.042	0.672	0.740	0.837	0.728
Beta for o.c	0.001	0.000	0.000	0.001	0.001	0.001

Table 7.B.8: **Parameters Values of the Latent Fields**. GPS tracked Female seals 2009-2011 using Assemblage as the prey species.

Argos Females 1995-1999: Assemblage of 7

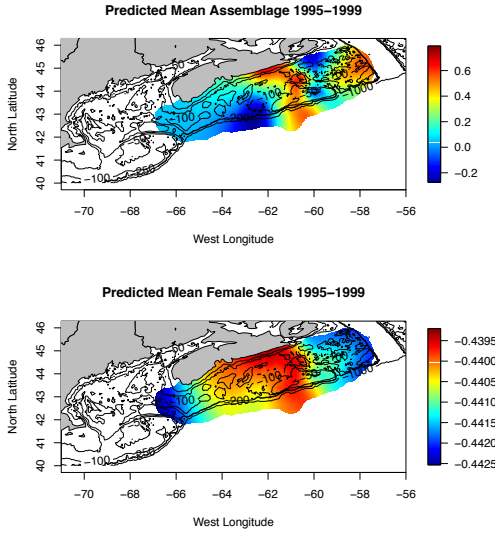

Figure 7.B.9: **Argos Females 1995-1999: Assemblage of 7** Random field for Assemblage (top) and Argos Females (bottom).

Assemblage7 F 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.300	0.290	0.022	0.215	1.061	0.060
Stdev for m	61.129	103.154	6.382	32.093	294.991	14.039
Range for x	4.578	1.762	2.260	4.199	9.025	3.567
Stdev for x	0.648	0.227	0.338	0.602	1.215	0.522
Beta for o.c	0.002	0.315	-0.623	0.005	0.614	0.017

Table 7.B.9: **Parameters Values of the Latent Fields**. Argos tracked Female seals 1995-1999 using Assemblage as the prey species.

7.C Supplementary Model Details: Male Seal Models

GPS Males 2009-2015: Sandlance

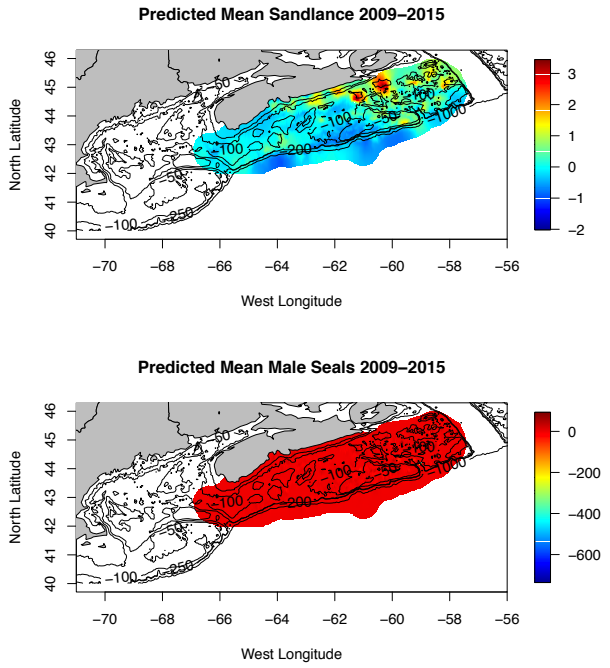

Figure 7.C.1: **GPS Males 2009-2015: Sandlance** Random field for Sandlance and GPS Males 2009-2015.

Sandlance M 9-15	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.107	0.003	0.103	0.107	0.113	0.106
Stdev for m	219.130	2.851	212.757	219.498	223.73	220.635
Range for x	0.236	0.015	0.208	0.235	0.265	0.235
Stdev for x	0.980	0.038	0.906	0.980	1.056	0.980
Beta for o.c	-0.002	0.000	-0.003	-0.002	-0.001	-0.002

Table 7.C.1: **Parameters Values of the Latent Fields**. GPS tracked Male seals 2009-2015 using Sandlance as the prey species. Provided for comparison to Cod

Argos Males 1995-1999: Sandlance

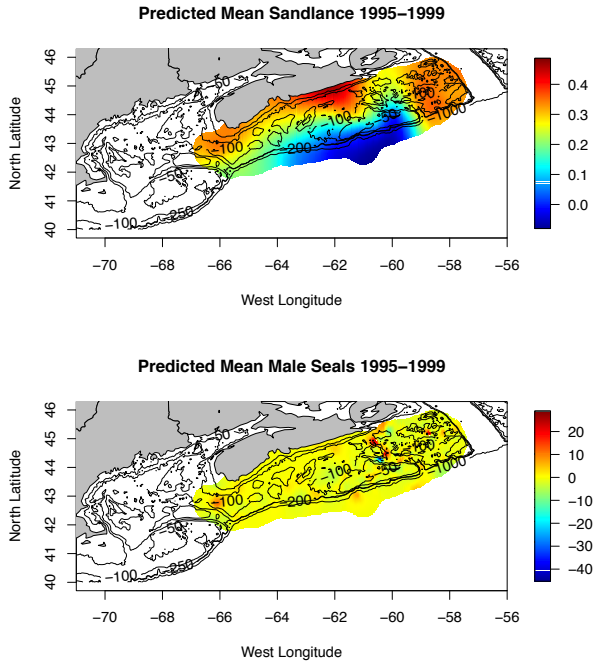

Figure 7.C.2: **Argos Males 1995-1999: Sandlance** Random field for Sandlance and GPS Males 1995-1999.

Sandlance M 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.342	0.034	0.279	0.341	0.413	0.340
Stdev for m	7.695	0.569	6.609	7.686	8.846	7.683
Range for x	7.383	3.940	3.048	6.327	17.722	4.841
Stdev for x	0.356	0.167	0.158	0.315	0.789	0.252
Beta for o.c	0.023	0.007	0.007	0.023	0.036	0.024

Table 7.C.2: **Parameters Values of the Latent Fields**. Argos tracked Male seals 1995-1999 using Sandlance as the prey species. Provided for comparison to Cod

GPS Males 2009-2015: Cod

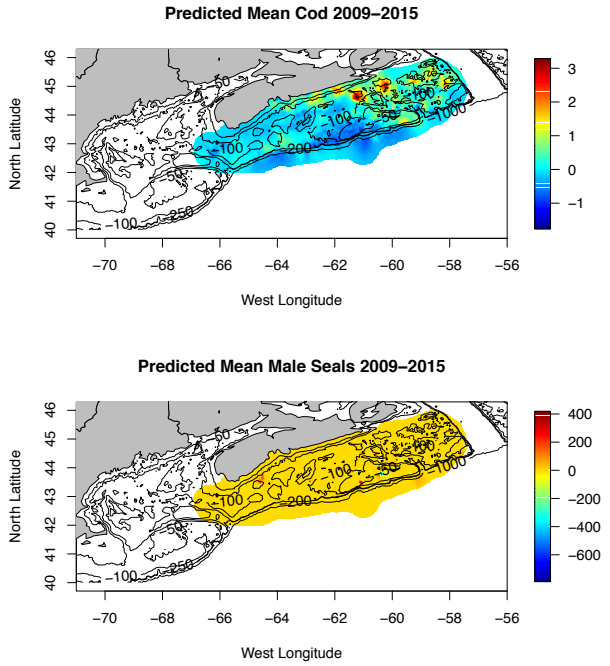

Figure 7.C.3: **GPS Males 2009-2015: Cod** Random field for Cod and GPS Males 2009-2015.

Cod M 09-15	mean	sd	0.025quant	0.5quant	0.975quant	mode
a.c	-0.2451	0.0019	-0.2488	-0.2451	-0.2414	-0.2451
a.y	-0.0427	0.0215	-0.0850	-0.0427	-0.0004	-0.0427
sed	0.0135	0.0002	0.0132	0.0135	0.0139	0.0135
tem	0.0153	0.0000	0.0152	0.0153	0.0154	0.0153
sed2	0.0015	0.0031	-0.0046	0.0015	0.0075	0.0015

Table 7.C.3: **Parameters Values for fixed effects**. GPS tracked Male seals 2009-2015 using Cod as the prey species.

Argos Males 1995-1999: Cod

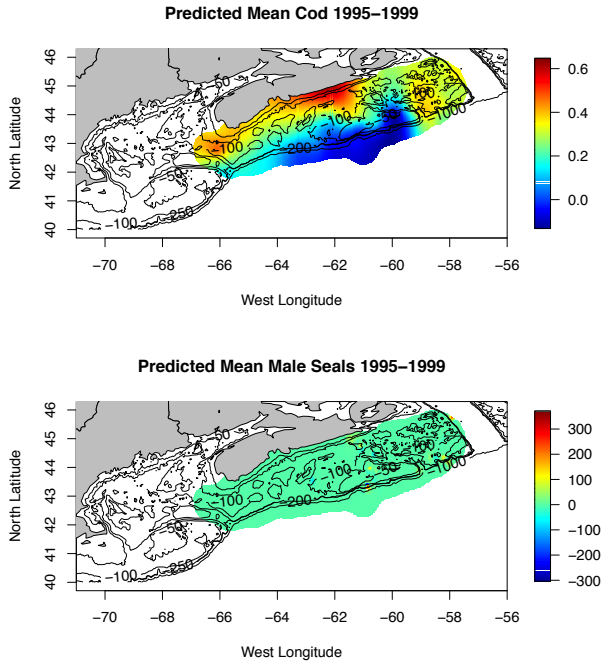

Figure 7.C.4: **Argos Males 1995-1999: Cod** Random field for Cod and GPS Males 1995-1999.

Cod M 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.128	0.005	0.117	0.128	0.138	0.130
Stdev for m	35.868	0.804	34.167	35.924	37.313	36.138
Range for x	5.044	2.488	2.160	4.409	11.539	3.472
Stdev for x	0.382	0.103	0.227	0.367	0.627	0.336
Beta for o.c	0.000	0.001	-0.002	0.000	0.002	-0.001

Table 7.C.4: **Parameters Values of the Latent Fields**. Argos tracked Male seals 1995-1999 using Sandlance as the prey species. Provided for comparison to Cod

GPS Males 2009-2015: Haddock

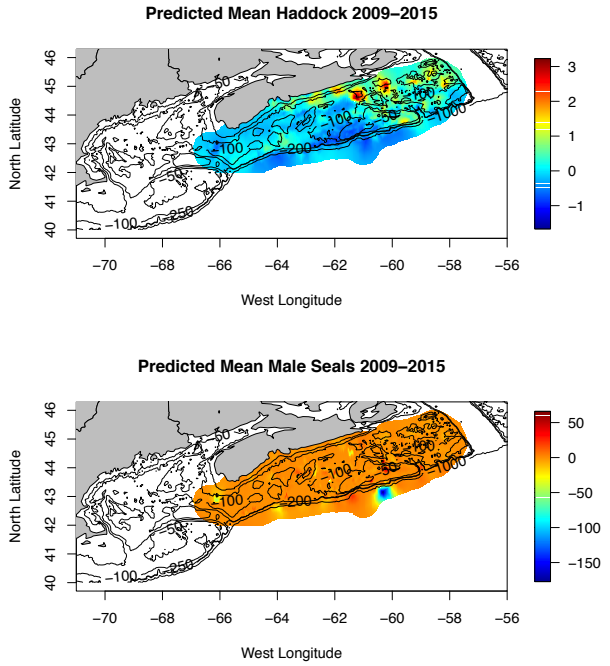

Figure 7.C.5: **GPS Males 2009-2015: Haddock** Random field for Haddock and GPS Males 2009-2015.

Haddock M 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.485	0.034	0.412	0.488	0.544	0.501
Stdev for m	5.890	0.853	14.054	15.969	17.362	16.298
Range for x	0.678	0.080	0.556	0.665	0.864	0.631
Stdev for x	0.808	0.082	0.683	0.795	0.999	0.758
Beta for o.c	-0.013	0.002	-0.015	-0.013	-0.009	-0.014

Table 7.C.5: **Parameters Values of the Latent Fields**. GPS tracked Male seals 2009-2015 using Haddock as the prey species.

Argos Males 1995-1999: Haddock

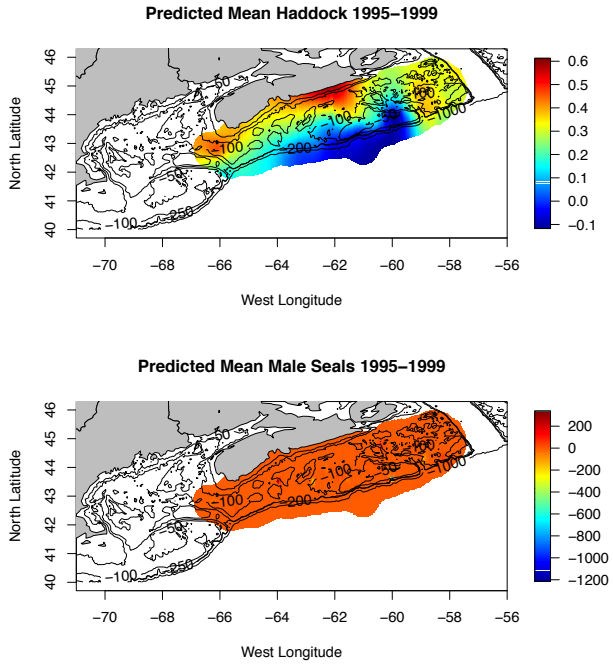

Figure 7.C.6: **Argos Males 1995-1999: Haddock** Random field for Haddock and GPS Males 1995-1999.

Haddock M 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.180	0.007	0.165	0.181	0.191	0.183
Stdev for m	70.660	1.760	66.846	70.827	73.675	71.470
Range for x	5.253	2.002	2.485	4.869	10.220	4.208
Stdev for x	0.491	0.176	0.242	0.459	0.924	0.403
Beta for o.c	0.002	0.002	0.0005	0.002	0.003	0.004

Table 7.C.6: **Parameters Values of the Latent Fields**. Argos tracked Male seals 1995-1999 using Haddock as the prey species.

GPS Males 2009-2015: Assemblage of 7

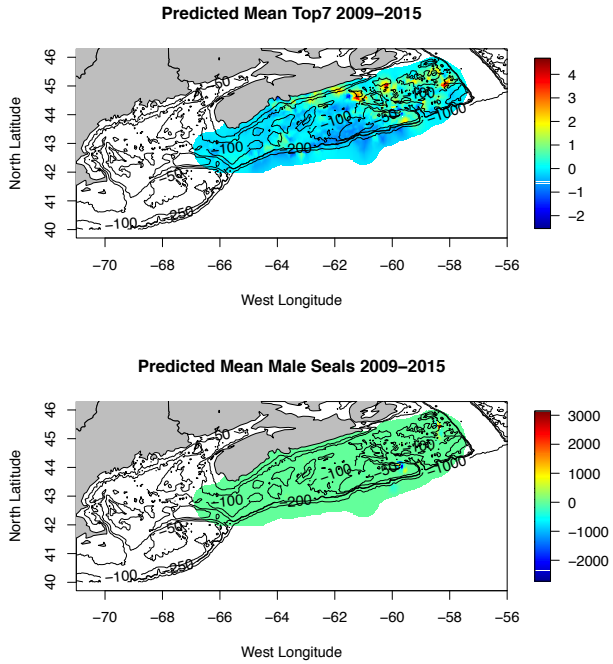

Figure 7.C.7: **GPS Males 2009-2015: Assemblage of 7** Random field for Assemblage of 7 and GPS Males 2009-2015.

Assemblage7 M 09-15	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.107	0.003	0.103	0.107	0.113	0.106
Stdev for m	218.921	2.768	212.650	219.309	223.255	220.860
Range for x	0.309	0.006	0.297	0.309	0.321	0.309
Stdev for x	1.274	0.022	1.226	1.275	1.314	1.282
Beta for o.c	-0.002	0.000	-0.003	-0.002	-0.001	-0.003

Table 7.C.7: **Parameters Values of the Latent Fields**. GPS tracked Male seals 2009-2015 using Assemblage as the prey species.

Argos Males 1995-1999: Haddock

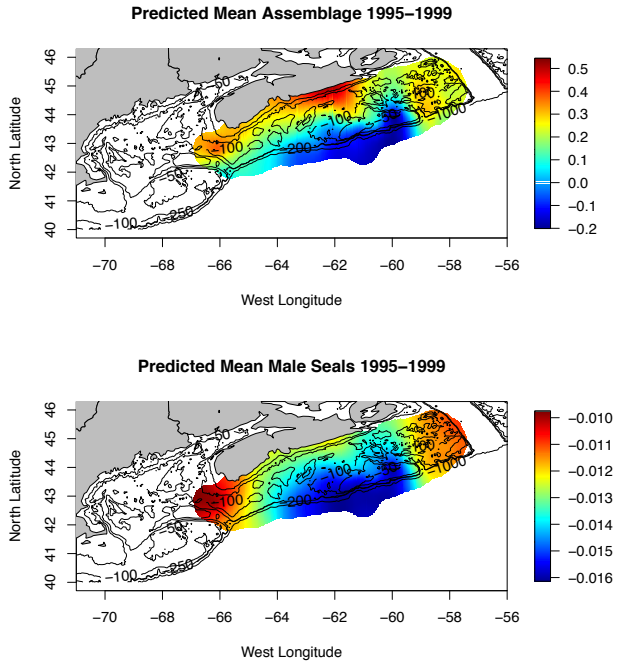

Figure 7.C.8: **Argos Males 1995-1999: Assemblage of 7** Random field for Assemblage and GPS Males 1995-1999.

Assemblage7 M 95-99	mean	sd	0.025quant	0.5quant	0.975quant	mode
Range for m	0.311	0.303	0.022	0.222	1.116	0.061
Stdev for m	58.880	91.618	6.322	32.433	274.131	14.361
Range for x	5.292	2.577	2.338	4.628	12.015	3.641
Stdev for x	0.415	0.169	0.206	0.375	0.850	0.310
Beta for o.c	0.013	0.261	-0.512	0.019	0.515	0.038

Table 7.C.8: **Parameters Values of the Latent Fields**. Argos tracked Male seals 1995-1999 using Assemblage as the prey species.

Chapter 8

Conclusions

The primary objectives of this thesis were, first, to find suitable statistical methods to model some of the novel, unique and spatially explicit data types being generated by recent tagging technologies deployed. Second, to develop models for the population distributions and/or abundances of various species, and, third, to integrate such models for multiple species, and so be capable of examining the interdependence/interrelationships between marine species, such as predator and prey. All of this in the marine environment where such modelling is essential to understanding of processes which cannot usually be directly observed.

8.1 Bayesian Hierarchical Spatio-Temporal Models for Marine Data

The Bayesian geo-statistical tool provided by R-INLA is bound to continue to find new applications in the areas of marine modelling such as I have explored here. The R-INLA interface makes using these sophisticated models accessible to a wide audience of science practitioners. A novel data form arising from the technical advances in tagging technology was neatly characterized by these models ability to model the mean of the underlying process, thus exposing the nature of the biological circumstances underlying these new data. The application of these models to fisheries stock structure gave genuine insight into both notable, historic, failings in fisheries management and also, lent new perspective into the ongoing question of how exactly to manage the potentially multiple sub-populations of a species of high commercial and conservation interest existing on North America's eastern shore. Finally, the study incorporating two, interdependent, random fields allows a characterization of the relationship between two species co-existing in the same space, with an identifiable statistical relationship characterizing their likely biological relationship, retaining all the available spatial information, that is both novel, and would be difficult to accomplish otherwise.

Encounter Data Sets

Chapter 4 illustrates the method of fitting and evaluating a spatio-temporal GMRF model, using the encounter data generated by the OTN deployed VMT tags. This encounter data reveals itself to be informative, and by treating it as geo-statistical data, in that it is modelled a spatially continuous mean rather than modelled as a single realization of a pattern, we find that INLA provides a useful tool. That is, INLA allowed us to model the mean occurrence rates of the encounters across the space, as opposed to trying to model the occurrence of encounters directly as a point process.

Single fish Species Stock Structure Models

Chapters 5 and 6 provide models for fitting abundance data from scientific research trawls. These models were necessitated by the need to provide prey information for inclusion into our eventual integrated model for predator prey relationships. Originally intended to incorporate predator-prey encounter data akin to the seal to seal encounter data seen in Chapter 4, however, the small number of predator-prey encounters actually observed precluded this original intention. Seeing the predator prey interactions in a marine environment was, and remains, one of the difficulties that led to the study. i therefore turned to other sources of data to provide our prey information, and, it was necessary to see if such ancillary data could be used. The spatio-temporal models produced for these individual prey species proved not only useful for our eventual, integrated models, but proved able to provide insight and information about the individual species, their stock structures, and their management consequences; all of which were valuable, stand alone, results. The first instance examined the Atlantic halibut, a species whose present abundance is characterized as healthy. The analysis focussed on questions of exactly where is the halibut abundant, is there one large population, or many sub populations, and how does the scale of the management effort compare to the information available about the spatial scale of Atlantic halibut abundance? These questions were engendered by the difference in the species outcome (success) under two separate political regimes. Showing that the present management paradigm, which conceives of Atlantic halibut as a single large population, is inconsistent

with the modelled population, resolves the dispute.

The second study fitting abundance data from scientific research trawls deals with Atlantic cod. This species is in a very different state than the halibut. Having experienced an ecosystem wide collapse in abundance in the 1990s the species is not recovering. Our paper examines the species not in a descriptive way, but in a dynamic way. The paper first models the cod abundance in a way similar to the treatment of halibut in the previous paper, but then discusses how the over stressing of the cod population by fishing is seen in the movement patterns of the parameters in these models over time. These parameter movements show a telling pattern as the population collapses that is informative and consistent.

Integrated, Multi-Species Interdependence Models

The integrated models presented in Chapter 7 are modelling the abundance of prey and the behaviour of a predator in reaction to that abundance, simultaneously, across the habitat available to both predator and prey, estimating the impact of the physical environment on each. Researchers cannot normally see the interplay between predator and prey in the marine environment, but know it must exist. The models of Chapter 7 give some insight into *which* species interact significantly on the Scotian Shelf, and also, *where* they interact. The *which* and the *where* are related to a *why*, insofar as the physical environment contributes to the prey distribution which determines the spatial characteristics of the predator's behaviour, the "ecological footprint" of the predator.

Our earliest efforts to model the encounter data made use of point pattern methods. That is, they were trying to model the pattern of occurrences of the encounters, treating the locations themselves as a random occurrence to be predicted. The focus there is the *intensity* of the pattern, the expected density of dots on the map, derived from the locations, and the statistical distribution of distances between locations. INLA models such as those explored here do not attempt to model the pattern, rather they model the mean of the underlying process giving rise to the encounters, treating the data as a single realization (or multiple realizations of) of the process. That is, to model an expected number of encounters at all locations, rather than model a distribution of additional encounters within a

given distance of any given observation. This approach turned out to be both more amenable to approach via recent statistical methods like INLA, but also readily interpretable in the biological sense. In fact, the biological interpretability of the parameters in the INLA models became one of the most appealing features of the approach, and forms a large part of one of the papers resulting from this research, see Chapter 6. This ability to model the mean of some behaviour associated with the locations and not the locations themselves provided the means to separate tracking data from the modelling of the tracks themselves. This in turn allowed us to reach our ultimate objective, to extend the data taken from individually tracked animals to models that delivered insight into the state, changing status, or ecological footprint of the species as a whole.

8.2 Future work

The relative ease of use of these models, combined with the usefulness of their results, indicate that these models will continue to find interested users. While the final, most complex, models dealt with here did begin to find the computational limits of the available computer technology, this limitation is temporary. Computers will continue to advance and so approaches like INLA will continue to reduce the difficulties associated with the big-n problem. (Or perhaps, more likely, n will continue to increase, keeping the difficulty constant). One area where this could apply would be the addition of what I will term dynamic covariates. In our models, all the physical covariates were static. Depth of water, bottom substrate, etc. do not vary. But the ocean is not static, it moves. Covariates such as flow or currents, if they could be input at each node of the mesh, would be an obvious first step. Of course, flows and current do not exist solely in two dimensions. All of the models we have presented are spatial, in the sense that the response is modelled over a two dimensional space. There is no technical reason why the Delauney triangulation of two dimension space (the mesh) could not be replaced by a three dimensional Delauney tetragonalization of three dimensional volume. Doing so would require (x,y,z) location data, and three dimensional covariates where possible. As the tagging technology improves, and more channels are recorded, such as accelerometer data from a tag in roll, pitch and yaw, the volume of the ocean could be modelled.

Bibliography

Alegana, V., Atkinson, P., co, C. L., Ruktanonchai, N., Bosco, C., zu Erbach-Schoenberg, E., Didier, B., Pindolia, D., Menach, A. L., Katokele, S., Uusiku, P., Tatem, A., 2016a. Advances in mapping malaria for elimination: Fine resolution modelling of *Plasmodium falciparum* incidence. Scientific Reports Vol. 6:29628.

Alegana, V., Kigozi, S., Nankabirwa, J., Arinaitwe, E., Kigozi, R., Mawejje, H., Kilama, M., Ruktanonchai, N., Ruktanonchai, C., Drakeley, C., Lindsay, S., Greenhouse, B., Kamya, M., Smith, D., Atkinson, P., Dorsey, G., Tatem, A., 2016b. Spatio-temporal analysis of malaria vector density from baseline through intervention in a high transmission setting. Parasites & Vectors Vol. 9:637.

Amante, C., Eakins, B. W., 2009. Etopo1 1 arc-minute global relief model: Procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24 1-19.

Andruskiw, A., Fryxell, J., Thompson, I., Baker, J., 2008. Habitat mediated variation in predation risk by the American marten. Ecology Vol 89, 2273-2280.

Arab, A., 2015. Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. International Journal of Environmental Research and Public Health Vol. 12(9), 10536-10548.

Arim, M., Naya, D. E., 2003. Pinniped diets inferred from scats: Analysis of biases in prey occurrence. Canadian Journal of Zoology Vol. 81, 67-73.

Austin, D., McMillan, J., Bowen, W., 2003. A three stage algorithm for filtering erronoeous argos satellite locations. Marine Mammal Science Vol. 19(2), 371-383.

Bailey, H., Thompson, P., 2009. Using marine mammal habitat modelling to identify priority conservation zones within a marine protected area. Marine Ecology Progress Series Vol. 378, 279-287.

Bannerjee, S., Carlin, B. P., Gelfand, A. E., 2004. Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall.

Bannerjee, S., Gelfand, A. E., FINLAy, A. O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society Series B, Vol. 70, 825-848.

Bartolino, V., Ciannelli, L., Spencer, P., Wilderbuer, T., Chan, K., 2012. Scale dependent detection of the effects of harvesting a marine fish population. Marine Ecology Progress Series Vol. 444, 251-261.

Baum, J. K., Myers, R. A., Kehler, D. G., Worm, B., Harley, S. J., Doherty, P. A., 2003. Collapse and conservation of shark populations in the North West Atlantic. Science 299, 389-392.

Beck, C. A., Bowen, W. D., Iverson, S. J., 2003a. Sex differences in the seasonal patterns of energy storage and expenditure in a phocid seal. Journal of Animal Ecology Vol. 72(2), 280-291.

Beck, C. A., Bowen, W. D., McMillan, J. I., Iverson, S. J., 2003b. Sex differences in diving at multiple temporal scales in a size-dimorphic capital breeder. Journal of Animal Ecology Vol. 72, 979-993.

Beck, C. A., Bowen, W. D., McMillan, J. I., Iverson, S. J., 2003c. Sex differences in the diving behavior of a size-dimorphic capital breeder: The grey seal. Animal Behaviour Vol. 66, 777-789.

Beck, C. A., Iverson, S. J., Bowen, W. D., 2005. Blubber fatty acids of grey seals reveal sex differences in the diet of a size-dimorphic marine carnivore. Canadian Journal of Zoology. Vol. 83, 377-388.

Beck, C. A., Iverson, S. J., Bowen, W. D., Blanchard, W., 2007. Sex differences in grey seal diet reflect seasonal variation in foraging behaviour and reproductive expenditure: Evidence from quantitative fatty acid signature analysis. Journal of Animal Ecology Vol. 76, 490-502.

Benoit, H. P., Swain, D. P., Bowen, W. D., Breed, G. A., Hamill, M. O., Harvey, V., 2011. Evaluating the potential for grey seal predation to explain elevated natural mortality in three fish species in the Southern Gulf of St. Lawrence. Marine Ecology Progress Series Vol. 442, 149-167.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics Vol. 43, 1-20.

Bivand, R., Gómez-Rubio, V., Rue, H., 2015. Spatial data analysis with R-INLA with some extensions. Journal of Statistical Software Vol. 63(20).

Bolker, B. M., 2008. Ecological Models and Data in R. Princeton University Press.

Bolstad, W. M., 2010. Understanding Computational Bayesian Statistics. John Wiley and Sons.

Boudreau, S., Shackell, N., Carson, S., den Heyer, C., 2017. Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the Northwest Atlantic Ocean. Ecology and Evolution Vol. 7(22), 9193–9844.

Boudreau, S., Worm, B., 2010. Top down control of lobster in the Gulf of Maine: Insights from local ecological knowledge and reseach surveys. Marine Ecology Progress Series Vol. 403, 181-191.

Bowen, W. D., 2011. Historical grey seal abundance and changes in the abundance of grey seal predators in the North West Atlantic. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/026.

Bowen, W. D., 2014. Stock assessment of Canadian grey seals (*Halichoerus grypus*). DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2014/010.

Bowen, W. D., 2017. Stock assessment of Canadian grey seals (*Halichoerus grypus*). DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2017/045.

Bowen, W. D., Carter, P., Hammill, M. O., 2011. Estimated grey seal diets near Sable Island derived from fecal samples: 1991 to 2010. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/0234.

Bowen, W. D., Harrison, G. D., 1994. Offshore diet of grey seals *Halichoerus grypus* near Sable island, Canada. Marine Ecology Progress Series Vol. 112, 1-11.

Bowen, W. D., Iverson, S. J., 2013. Methods of estimating marine mammal diets: A review of validation experiments and sources of bias and uncertainty. Marine Mammal Science Vol. 29(4), 719-754.

Bowen, W. D., Lawson, J. W., 1993. Seasonal and geographic variation in the species composition and size of prey consumed by grey seals (*Halichoerus grypus*) on the Scotian Shelf. Canadian Journal of Fisheries and Aquatic Science Vol. 50, 1768-1778.

Bowering, W., 1986. The distribution, age and growth, and sexual maturity of Atlantic halibut (*Hippoglossus hippoglossus*) in the Newfoundland and Labrador area of the Northwest Atlantic. Canadian Technical Report of Fisheries and Aquatic Sciences 1432, 1-34.

Breed, G. A., Bowen, W. D., McMillan, J. I., Leonard, M. L., 2006. Sexual segregation of seasonal foraging habitats in a non-migratory marine mammal. Proceedings of the Royal Society Series B, 273, 2319-2326.

Breed, G. A., Jonsen, I. D., Myers, R. A., Bowen, W. D., Leonard, M. L., 2009. Sex-specific, seasonal foraging tactics of adult grey seals (*Halichoerus grypus*) revealed by state space analysis. Ecology Vol. 90(11), 3209-3321.

Bris, A. L., Fisher, J., Murphy, H., Galbraith, P., Castonguay, M., Loher, T., Robert, D., 2017. Migration patterns and putative spawning habitats of Atlantic halibut (*Hippoglossus hippoglossus*) in the Gulf of St. Lawrence revealed by geolocation of pop-up satellite archival tags. ICES Journal of Marine Science, fsx098.

Bundy, A., 2005. Structure and functioning of the Eastern Scotian Shelf ecosystem before and after the collapse of groundfish stocks in the early 1990s. Canadian Journal of Fisheries and Aquatic Science Vol. 62, 1453-1473.

Cadigan, N., 2016. A state-space stock assessment model for Northern cod, including unreported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Science Vol. 73(2), 296-308.

Cadrin, S. X., Kerr, L. A., Mariani, S., 2014. Stock Identification Methods (2nd ed.). Academic Press.

Cameletti, M., Ignaccolo, R., Bande, S., 2011. Comparing spatio-temporal models for particulate matter in the Piemonte. Environmetrics Vol. 22, 985-996.

Cameletti, M., Lindgren, F., Simpson, D., Rue, H., 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. Advances in Statistical Analysis Vol. 97, No. 2, 109-131.

Campana, S. E., Mohn, R. K., Smith, S. J., Chouinard, G. A., 1995. Spatial implications of a temperature-based growth model for Atlantic cod (*Gadus morhua*) off the Eastern coast of Canada. Canadian Journal of Fisheries and Aquatic Science Vol. 52, 2445-2456.

Carson, S., Flemming, J. M., 2014. Seal encounters at sea: A contemporary spatial approach using R-INLA. Ecological Modelling, Vol. 291, 175-181.

Carson, S., Shackell, N., Flemming, J. M., 2017. Local overfishing may be avoided by examining parameters of a spatio-temporal model. PLoS ONE Vol. 12(9): e0184427.

Cianelli, L., Fisher, J., Skern-Mauritzen, M., Hunsicker, M., Hidalgo, M., Frank, K., Bailey, K., 2013. Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: A review. Marine Ecology Progress Series Vol. 480, 227-243.

Corten, A., 2013. Recruitment depressions in North Sea herring. ICES Journal of Marine Science Vol. 70(1), 1-15.

Cosandey-Godin, A., Krainski, E. T., Worm, B., Flemming, J. M., 2015. Applying Bayesian spatio-temporal models to fisheries bycatch in the Canadian Arctic. Canadian Journal of Fisheries and Aquatic Science Vol. 72(2) 186-197.

Cox, D. R., 1955. Some statistical methods connected with series of events. Journal of the Royal Statistical Society Series B, Vol. 17, No. 2, 129-164.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society Series B, Vol. 70, 209-226.

Cressie, N., Wilkie, C. K., 2011. Statistics for spatio-temporal data. Wiley.

Cressie, N. A. C., Rev. ed. 1993. Statistics for Spatial Data. Wiley.

Daley, D. J., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes, Vol I. Springer.

den Heyer, C., Armsworthy, A., Wilson, S., Wilson, G., Bajona, L., Bond, S., Trzcinski, M., 2012. Atlantic halibut all sizes tagging program summary report for 2006 to 2011. Canadian Technical Report of Fisheries and Aquatic Sciences 2992, vi+38pp.

den Heyer, C., Lang, S., Bowen, W., Hammill, M., 2017. Pup production at Scotian Shelf grey seal (*Halichoerus grypus*) colonies in 2016. DFO Can. Sci. Advis. Sec. 2017/056.

DFO, 2008. The gully marine protected area management plan. Oceans and Habitat Branch DFO/2007-1227, 80pp, ISBN: 978-0-662-45468-7.
URL `http://www.dfo-mpo.gc.ca/Library/333121.pdf`

DFO, 2011. Impacts of grey seals on fish populations in Eastern Canada. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2010/071.

DFO, 2012. Assessment of the status of 4X 5Y haddock in 2011. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2012/023.

DFO, 2015a. 2014 assessment of Atlantic halibut on the Scotian Shelf and Southern Grand Banks. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2015/012.

DFO, 2015b. Stock assessment of Atlantic halibut of the Gulf of St. Lawrence for 2013 and 2014. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2015/023.

Diggle, P. J., Ribiero, P. J., 2006. Model based geostatistics. Springer.

DofC, 1999. US Department of Commerce Northeast multispecies fishery: Amendment 9 to the Northeast multispecies fishery management plan. federal register. Rules and Regulations Vol. 64, No. 199: 55821-55828.
URL `http://archive.nefmc.org/nemulti/planamen/Amend209/finalrule.pdf`

DofC, 2012. US Department of Commerce Northeast fisheries scence center (NEFSC): Assessments or data updates of 13 northeast groundfish stocks through 2010. Northeast Fish Sci Cent Ref Doc 12-06, 789pp.
URL `http://nefsc.noaa.gov/publications`

Erisman, B. E., Allen, L. G., Claisse, J. T., II, D. J. P., Miller, E. F., Murray, J. H., 2011. The illusion of plenty: Hyperstability masks collapses in two recreational fisheries that target fish spawning aggregations. Canadian Journal of Fisheries and Aquatic Science Vol. 68, 1705-1716.

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

Foss, A., Imsland, A., Nævdal, G., 1998. Population genetic studies of the Atlantic halibut in the North Atlantic Ocean. Journal of Fish Biology Vol. 5, 901-905.

Frank, K., Petrie, B., Choi, J., Leggatt, W., 2005. Trophic cascades in a formerly cod dominated ecosystem. Science Vol 308, 1621-1623.

Frank, K., Petrie, B., Fisher, J., Leggett, W., 2011. Transient dynamics of an altered large marine ecosystem.

Frank, K. T., Brickman, D. W., 2000. Allee effects and compensatory population dynamics within a stock complex. Canadian Journal of Fisheries and Aquatic Science Vol. 57(3), 513-517.

Fretwell, S. D., Lucas, H. L., 1970. On territorial behaviour and other factors influenceing the habitat distribution of birds. I. Theoretical developement. Acta Biotheoretica Vol. 19, 16-36.

Fu, C., Mohn, R., Fanning, L. P., 2001. Why the Atlantic cod, *Gadus morhua* stock off Eastern Nova Scotia has not recovered. Canadian Journal of Fisheries and Aquatic Science Vol. 58, 1613-1623.

Gardner, B., Sullivan, P. J., Morreale, S. J., Epperly, S. P., 2008. Spatial and temporal statistical analysis of bycatch data: Patterns of sea turtle bycatch in the North Atlantic. Canadian Journal of Fisheries and Aquatic Science Vol. 65, 2461-2470.

Gaston, K. J., 2003. The structure and dynamics of geographic ranges. Oxford University Press; On Demand.

Gaston, K. J., Blackburn, T. M., 2000. Pattern and process in macroecology. Blackwell Science.

Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., Lawton, J. H., 2000. Abundance-occupancy relationships. Journal of Applied Ecology Vol. 37 S1, 39-59.

Geisser, S., Eddy, W. F., 1979. A predictive approach to model selection. Journal of the American Statistical Association Vol. 74, 153-160 (Corr: V75, p765).

Grasso, G., 2008. What appeared limitless plenty: The rise and fall of the Nineteenth-Century Atlantic halibut fishery. Environmental History Vol. 13, 66-91.

Hammill, M., den Heyer, C., Bowen, W., Lang, S., 2017. Grey seal population trends in Canadian waters, 1960-2016 and harvest advice. DFO Can. Sci. Advis. Sec. 2017/052.

Harvey, S. J., Myers, R., Barrowman, N., Bowen, K., Amiro, R., 2001. Estimation of research trawl catchability for biomass reconstruction of the Eastern Scotian Shelf. DFO Can. Sci. Advis. Sec. Res. Doc. 2001/084.

Harvey, V., Hammill, M. O., Swain, D. P., Breed, G. A., Lydersen, C., Kovacs, K. M., 2012. Winter foraging by a top predator, the grey seal *Halichoerus grypus*, in relation to the distribution of prey. Marine Ecology Progress Series Vol. 462, 273-286.

Hastie, T. J., Tibshirani, R. J., 1990. Generalized Additive Models. Chapman & Hall.

Haug, T., Fevolden, S., 1986. Morphology and biochemical genetics of Atlantic halibut, (*Hippoglossus hippoglossus*), from various spawing grounds. Journal of Fish Biology Vol. 28, 367-378.

Hauser, L., Carvalho, G., 2008. Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. Fish and Fisheries Vol. 9(4), 333-362.

Hazel, J., 2009. Evaluation of fast-acquisition GPS in stationary tests and fine-scale tracking of green turtles. Journal of Experimental Marine Biology and Ecology Vol. 374(1), 58-68.

Hedger, R., McKenzie, E., Heath, M., Wright, P., Scott, B., Gallego, A., Andrews, J., 2004. Analysis of the spatial distributions of mature cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) abundance in the North Sea (1980-1999) using generalised additive models. Fisheries Research Vol. 70, 17-25.

Horsman, T. L., Shackell, N. L., 2009. Atlas of important habitat for key fish species of the Scotian Shelf, Canada. DFO Can. Tech. Rep. Fish. Aquat. Sci. 2835.

Hu, L., Wroblewski, J., 2009. Conserving a subpopulation of the Northern Atlantic cod metapopulation with a marine protected area. Aquatic Conservation: Marine and Freshwater Ecosystems Vol. 19, 178-193.

Hu, Y., Ward, M., Xia, C., Li, R., Sun, L., Lynn, H., Gao, F., Wang, Q., Zhang, S., Xiong, C., Zhang, Z., Jiang, Q., 2016. Monitoring schistosomiasis risk in East China over space and time using a Bayesian hierarchical modeling approach. Scientific Reports Vol. 6:24173.

Hussey, N., Kessel1, S., Aarestrup, K., Cooke, S., Cowley, P., Fisk, A., Harcourt, R., Holland, K., Iverson, S., Kocik, J., Flemming, J. M., Whoriskey, F., 2015. Aquatic animal telemetry: A panoramic window into the underwater world. Science Vol. 348.

Hutchings, J., 1996. Spatial and temporal variation in the density of Northern cod and a review of hypotheses for the stock's collapse. Canadian Journal of Fisheries and Aquatic Science Vol. 53, 943-962.

Hutchings, J., Myers, R., 1994. What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, off Newfoundland and Labrador. Canadian Journal of Fisheries and Aquatic Science Vol. 51, 2126-2146.

Illian, J. B., Sørbye, S. H., Rue, H., 2012a. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Annals of Applied Statistics. Vol. 6, No. 4, 1499-1530.

Illian, J. B., Sørbye, S. H., Rue, H., Hendrichsen, D., 2012b. Using INLA to fit a complex point process model with temporally varying effects - a case study. Journal of Environmental Statistics Vol 3 , No. 7.

Iverson, S. J., Field, C., Bowen, W. D., Blanchard, W., 2004. Quantitative fatty acid signature analysis: A new method for estimating predator diets. Ecological Monographs Vol. 74, 211-235.

Jaworski, C., Bompard, A., Genies, L., Amiens-Desneux, E., Desneux, N., 2013. Preference and prey switching in a generalist predator attacking local and invasive alien pests. PLOS ONE.

Jonsen, I. D., Mills-Flemming, J., Myers, R. A., 2005. Robust state-space modelling of animal movement data. Ecology Vol. 86, Issue 11 (November 2005).

Jonsen, I. D., Myers, R. A., Flemming, J. M., 2003. Meta-analysis of animal movement using state-space models. Ecology Vol. 84(11), 3055-3063.

Jousimo, J., Ovaskainen, O., 2016. A spatio-temporally explicit random encounter model for large-scale population surveys. PLoS ONE Vol. 11(9): e0162447.

Kammann, E. E., Wand, M. P., 2003. Geoadditive models. Applied Statistics Vol. 52, 1-18.

Kanwit, J., 2007. Tagging results from the 2000-2004 federal experimental fishery for Atlantic halibut (*Hippoglossus hippoglossus*) in the Eastern Gulf of Maine. Journal of Northwest Atlantic Fishery Science Vol. 38, 37-42.

Kays, R., Crofoot, M., Jetz, W., Wikelski, M., 2015. Terrestrial animal tracking as an eye on life and planet. Science Vol. 348.

Kerfoot, W., (Editors), A. S., 1987. Predation: Direct and indirect impacts on aquatic communities. University Press of New England, Hanover.

Kerr, L., Cadrin, S., Secor, D., 2010a. The role of spatial dynamics in the stability, resilience and productivity of an estuarine fish population. Ecological Application Vol. 20(2), 497-507.

Kerr, L., Cadrin, S., Secor, D., 2010b. Simulation modelling as a tool for examining the consequences of spatial structure and connectivity on local and regional population dynamics. ICES Journal of Marine Science Vol. 67, 1631-1639.

Kohler, A., 1967. Size at maturity, spawning season, and food of Atlantic halibut. Journal of the Fisheries Research Board of Canada Vol. 24(1), 53-56.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., Bell, B., 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software Vol. 70, 1-42.

Lidgard, D. C., Bowen, W. D., Jonsen, I. D., Iverson, S. J., 2012. Animal-borne acoustic transceivers reveal patterns of at-sea associations in an upper-trophic level predator. PLoS ONE 7(11): e48962.

Lima, S., 1998. Non-lethal effects in the ecology of predator-prey interactions. BioScience Vol 48, 25-34.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software Vol. 63, No. 19.

Lindgren, F., Rue, H., Lindstrom, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). Journal of the Royal Statistical Society Series B, Vol. 73, No. 4, 423-498.

Lunn, D., Speigelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions.

Martell, S., Vincent, A., Turis, B., 2013. MSC Final report v4.0. Canada Atlantic halibut. SCS Global Services, Emeryville CA 259pp.

Matérn, B., 1960. Spatial variation. Meddelanded Stat. Skogsforsk 49 (5), 1-144.

Matthiopoulos, J., 2003. Model supervised kernel smoothing for the estimation of spatial usage. Nordic Society Oikos Vol 102, 367-377.
URL `http://www.jstor.org/stable/3548039`

Matthiopoulos, J., McConnell, B., Duck, C., Fedak, M., 2004. Using satellite telemetry and aerial counts to estimate space use by grey seals around the British isles. Journal of Applied Ecology Vol 41, 476-491.

Maury, O., Gascuel, D., 2001. 'Local overfishing' and fishing tactics: Theoretical considerations and applied consequences in stock assessment studied with a numerical simulator of fisheries. Aquatic Living Resources Vol. 14(4), 203-210.

Meehan, T., Michel, N., Rue, H., 2017. Estimating animal abundance with n-mixture models using the R-INLA package for R. https://arxiv.org/pdf/1705.01581.pdf.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation state calculations by fast computing machines. Journal of Chemical Physics 23, 1087.

Mohn, R., Bowen, W. D., 1996. Grey seal predation on the Eastern Scotian Shelf: Modelling the impact on Atlantic cod. Canadian Journal of Fisheries and Aquatic Science Vol. 53, 2722-2738.

Moilanen, A., Hanski, I., 1998. Metapolulation dynamics: Effects of habitat quality and landscape structure. Ecology Vol. 79(7), 2503-2515.

Moller, A., Thorarinsdottir, T., Lenkoski, A., Gneiting, T., 2016. Spatially adaptive, Bayesian estimation for probabilistic temperature forecasts. https://arxiv.org/pdf/1507.05066.pdf.

Møller, J., 2010. Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. Scandinavian Journal of Statistics Vol. 37, No. 2, 2-25.

Møller, J., Waagepetersen, R. P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall.

Mork, J., Haug, T., 1983. Genetic variation in halibut, (*Hippoglossus hippoglossus L.*), from Norwegian waters. Hereditas Vol. 9, 167-174.

Munoz, F., Pennino, M. G., Conesa, D., López-Quílez, A., Bellido, J. M., 2013. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. Stochastic Environmental Research and Risk Assessment Vol. 27, No. 5, 1171-1180.

Murphy, H., Fisher, J., Bris, A. L., Desgagnes, M., Castonguay, M., Loher, T., Robert, D., 2017. Characterization of depth distributions, temperature associations, and seasonal migrations of Atlantic halibut in the Gulf of St. Lawrence using pop-up satellite archival tags. Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science.

Musenge, E., Chirwa, T. F., Kahn, K., Vounatsou, P., 2012. Bayesian analysis of zero inflated spatiotemporal HIV/TB child mortality data through INLA and SPDE approaches: Applied to data observed between 1992 and 2010 in rural North East South Africa. International Journal of Applied Earth Observation and Geoinformation.

Myers, R., Hutchings, J., Barrowman, N., 1997. Why do fish stocks collapse? The example of cod in Atlantic Canada. Ecological Applications Vol. 7(1), 91-106.

Neilson, J., Bowering, W., 1989. Minimum size regulations and the implications for yield and value in the Canadian Atlantic halibut (*Hippoglossus hippoglossus*) fishery. Canadian Journal of Fisheries and Aquatic Sciences Vol 46(11), 1899-1903.

Neilson, J., Bowering, W., Frechet, A., 1987. Management concerns for Atlantic halibut (*Hippoglossus hippoglossus*) in the Canadian North Atlantic. DFO Can. Atl. Fish. Sci. Adv. Cmte. (CAFSAC) 87/73.

OBIS, 2014. Ocean Biogeographic Information System: DFO Scotian Summer Research Trawl Survey (OBIS Canada). Department of Fisheries and Oceans (DFO) Canada.
URL `http://www.gbif.org/dataset/8393f330-f762-11e1-a439-00145eb45e9a`

O'Boyle, R., 2011. Benefits of marine protected areas and fisheries closures in the Northwest Atlantic. Canadian Technical Report of Fisheries and Aquatic Sciences 2948, iii+68pp.

O'Driscoll, R. L., Schneider, D. C., Rose, G. A., Lilly, G. R., 2000. Potential contact statistics for measuring scale-dependent spatial pattern and association: An example of Northern cod (*Gadus Morhua* and capelin (*Mallotus Villosus*). Canadian Journal of Fisheries and Aquatic Science Vol. 57, 1355-1388.

Olsen, E., Knutsen, H., Gjøsæter, J., Jorde, P., Knutsen, J., Stenseth, N., 2008. Small scale biocomplexity in coastal Atlantic cod supporting a Darwinian perspective on fisheries management. Evolutionary Applications Vol. 1, 524-533.

OTN, 2013. OTN Canada phase II proposal 158749.

Paradinas, I., Conesa, D., Pennino, M. G., Munoz, F., fernandez, A. M., Lopez-Quilez, A., Bellido, J. M., 2015. Bayesian spatio-temporal approach to identifying fish nurseries by validating persistence areas. Marine Ecology Progress Series Vol. 528, 245-255.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopouos, J., 2008. State-space models of individual animal movement. Trends in Ecology and Evolution Vol. 23, 87-94.

Payne, M., 2010. Mind the gaps: A state space model for analyzing the dynamics of North Sea herring spawning components. ICES Journal of Marine Science Vol. 67, 1939-1947.

Pennino, M. G., Munoz, F., Conesa, D., López-Quílez, A., Bellido, J. M., 2013. Modelling sensitive elasmobranch habitats. Journal of Sea Research Vol. 83, 209-218.

Pennino, M. G., Munoz, F., Conesa, D., López-Quílez, A., Bellido, J. M., 2014. Bayesian spatio-temporal discard model in a demersal trawl fishery. Journal of Sea Research Vol. 90, 44-53.

Pinhorn, A. T., 1988. Catchability in some of the major groundfish fisheries of the East coast of Canada. Journal of Northwest Atlantic Fisheries Science Vol. 8, 15-23.

Plummer, M., 2009. Bayesian graphical models using MCMC.

Quiroz, Z. C., Prates, M. O., Rue, H., 2014. A Bayesian approach to estimate the biomass of anchovies off the coast of Perú. Biometrics.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org`

R. Froese and D. Pauly, eds., 2015. FishBase.
URL `www.fishbase.org`

Reich, D., Dealteris, J., 2009. A simulation study of the effects of spatially complex population structure for Gulf of Maine Atlantic cod. North American Journal of Fisheries Management Vol. 29, 116-126.

Reid, D., Pongsomboon, S., Jackson, T., McGowan, C., Murphy, C., Martin-Robichaud, D., Reith, M., 2005. Microsatellite analysis indicates an absense of population structure among (*Hippoglossus hippoglossus*) in the Northwest Atlantic. Journal of Fish Biology Vol. 67, 570-576.

Reubens, J. T., Pasotti, F., Degraer, S., Vincx, M., 2013. Residency, site fidelity and habitat use of Atlantic cod, *Gadus morhua* at an offshore wind farm using acoustic telemetry. Marine Environmental Research Vol. 90, 128-135.

Reuchlin-Hugenholtz, E., Shackell, N. L., Hutchings, J. A., 2015. The potential for spatial distribution indices to signal thresholds in marine fish biomass. PLOS ONE.

Ricard, D., Shackell, N. L., 2013. Population status (abundance/biomass, geographic extent, body size and condition), important habitat, depth, temperature and salinity preferences of marine fish and invertebrates on the Scotian Shelf and Bay of Fundy (1970-2012). DFO Can. Tech. Rep. Fish. Aquat. Sci. 3012.

Roney, N., Hutchings, J., Olsen, J., Knutsen, E., Albretson, J., Kuparinen, A., 2016. Fine scale life history structure in a highly mobile marine fish. Evolutionary Ecology Research Vol 17, 95-109.

Rose, G., deYoung, B., Kulka, D., Goddard, S., Fletcher, G., 2000. Distribution shifts and overfishing the northern cod (gadus morhua): a view from the ocean. Canadian Journal of Fisheries and Aquatic Science Vol. 57, 644-663.

Rose, G., Rowe, S., 2015. Northern cod comeback. Canadian Journal of Fisheries and Aquatic Science Vol. 72(12), 1789-1798.

Rose, G. A., Dutka, D. W., 1999. Hyper-aggregation of fish and fisheries: How catch-per-unit-effort increased as the Northern cod (*Gadus morhua*) declined. Canadian Journal of Fisheries and Aquatic Science Vol. 56(Suppl. 1), 118-127.

Ross, B. E., Hooten, M. B., Koons, D. N., 2012. An accessible method for implementing hierarchical models with spatio-temporal abundance data. PLoS ONE 7(11): e49395.

Royer, F., Fromentin, J. M., Gaspar, P., 2004. Association between bluefin tuna schools and oceanic features in the Western Mediterranean. Marine Ecology Progress Series Vol. 269, 249-263.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Theory and Applications. Chapman & Hall.

Rue, H., Martino, S., Chopin, N., 2009a. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B, Vol. 71, No. 2, 319-392.

Rue, H., Martino, S., Chopin, N., 2009b. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B, Vol. 71, No. 2, 319-392.

Ruzzante, D., Mariani, S., Bekkevold, D., Andre, C., Mosegaard, H., Clausen, L., Dahlgren, T., Hutchinson, W., Hatfield, E., Torstensen, E., Brigham, J., Simmonds, E., Laikre, L., Larsson, L., Stet, R., Ryman, N., Carvalho, G., 2006. Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. Proceedings of the Royal Society Series B-Biological Sciences 273: 2279-2284.

Sadykova, D., Scott, B., Dominicis, M. D., Wakelin, S., Sadykov, A., Wolf, J., 2017. Bayesian joint models with INLA exploring marine mobile predator- prey and competitor species habitat overlap. Ecology and Evolution 2017: 1-15.

Safina, C., Rosenberg, A., Myers, R., II, T. Q., Collie, J., 2005. US Ocean fish recovery: Staying the course. Science Vol. 309, 707-708.

Schabenberger, O., Gotway, C. A., 2005. Statistical Methods for Spatial Data Analysis. Chapman & Hall.

Schrodle, B., Held, L., 2011. A primer on disease mapping and ecological regression using INLA. Computational Statistics Vol. 26(2), 241-258.

Seitz, A., Evans, M., Courtney, M., Kanwit, J., 2016. Continental shelf residency by adult Atlantic halibut electronic tagged in the Gulf of Maine. Journal of Northwest Atlantic Fishery Science Vol. 48, 33-40.

Seitz, A., Farrugia, T., Norcross, T., Loher, B., Neilsen, J., 2017. Basin scale reproductive segregation of Pacific halibut (*Hippoglossus stenolepis*). Fisheries Management and Ecology Vol. 24, 339-346.

Shackell, N., Bundy, A., Nye, J., Link, J., 2012. Common large-scale responses to climate and fishing across Northwest Atlantic ecosystems. ICES Journal of Marine Science Vol. 69(2), 151-162.

Shackell, N., Frank, K., 2007. Compensation in exploited marine fish communities on the Scotian Shelf, Canada. Marine Ecology Progress Series Vol. 336, 235-247.

Shackell, N., Frank, K., Fisher, J., Petrie, B., Leggatt, W., 2010. Decline in top predator body size and changing climate altertrophic structure in an oceanic ecosystem. Proceedings of the Royal Society Series B-Biological Sciences 277: 1353-1360.

Shackell, N., Frank, K., Nye, J., den Heyer, C., 2016. A transboundary dilemma: Dichotomous designations of Atlantic halibut status in the Northwest Atlantic. ICES Journal of Marine Science Vol. 73, 1798-1805.

Shackell, N. L., Frank, K. T., Brickman, D. W., 2005. Range contraction may not always predict core areas: An example from marine fish. Ecological Applications Vol. 15(4), 1440-1449.

Shackell, N. L., Stobo, W. T., Frank, K. T., Brickman, D. W., 1997. Growth of cod (*Gadus morhua*) estimated from mark recapture programs on the Scotian Shelf and adjacent areas. ICES Journal of Marine Science Vol. 54, 383-398.

Shelton, A., Thorson, J., Ward, E., Feist, B., 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Science Vol. 71 1655-1666.

Shelton, P., Sinclair, A., Chouinard, G., Mohn, R., Duplisea, D., 2006. Fishing under low productivity conditions is further delaying recovery of Northwest Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences Vol 63(2), 235-238.

Simpson, D., Lindgren, F., Rue, H., 2012a. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. Environmetrics Vol. 23, 65-74.

Simpson, D., Lindgren, F., Rue, H., 2012b. Think continuous: Markovian gaussian models in spatial statistics. Spatial Statistics Vol. 1, 16-29.

Simpson, D., Rue, H., Martins, T. G., Riebler, A., Sorbye, S. H., 2015. Penalising model component complexity: A principled, practical approach to constructing priors.
URL `http://https://arxiv.org/abs/1403.4630`

Smedbol, R., Wroblewski, J., 2002. Metapopulation theory and Northern cod population structure: Interdependency of subpopulations in recovery of a groundfish population. Fisheries Research Vol. 55, 161-174.

Smith, C. D., King, M. C., Shackell, N. L., 2013. Spring, summer and fall seasonal survey maps of fish distribution on the Scotian Shelf between 1978 and 1984. Cdn. Manuscript. Rpt. of Fish. and Aquat. Sci. 3013.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B, Vol. 64, No. 4, 583-639.

Steele, J., 1985. A comparison of terrestrial and marine ecological systems. Nature Vol. 313, 355-358.

Stenek, R., Vavrinec, J., Leland, A., 2004. Accelerating trophic level dysfunction in kelp forest ecosystems of the Western North Atlantic. Ecosystems Vol. 7, 323-332.

Stephenson, R., 2002. Stock structure and management structure: An ongoing challenge for ICES. ICES Marine Science Symposia Vol. 215, 305-314.

Sterling, J. T., Springer, A. M., Iverson, S. J., Johnson, S. P., Pelland, N. A., Johnson, D. S., Lea, M.-A., Bond, N. A., 2014. The sun, moon, wind and biological imperative-shaping contrasting wintertime migration and foraging strategies of adult male and female Northern fur seals (*Callorhinus ursinus*). PLOS ONE Vol. 9, Issue 4.

Sterner, T., 2007. Unobserved diversity, depletion and irreversability: The importance of subpopulations for management of cod stocks. Ecological Economics Vol. 61, 566-574.

Stobo, W., Neilson, J., Simpson, P., 1988. Movements of Atlantic halibut (*Hippoglossus hippoglossus*) in the Canadian North Atlantic. Canadian Journal of Fisheries and Aquatic Science Vol. 45, 484-491.

Stoyan, D., Kendall, W. S., Mecke, J., 1983. Stochastic Geometry and its Applications. Wiley.

Svedäng, H., Bardon, G., 2003. Spatial and temporal aspects of the decline in cod (*Gadus morhua L.*) abundance in the Kattegat and Eastern Skagerrak. ICES Journal of Marine Science Vol. 60, 32-37.

Svedäng, H., Cardinale, M., André, C., 2001. Recovery of former fish productivity: Philopatric behaviours put depleted stocks in an unforeseen deadlock. In Ecosystem based Management of Marine Fisheries: An Evolving Perspective, A. Belgrano and C. Fowler Eds., pp 232-247. Cambridge University Press.

Svedäng, H., Svenson, A., 2006. Cod (*Gadus morhua L.*) populations as behavioural units: Inference from time series on juvenile abundance in the Eastern Skagerrak. Journal of Fish Biology Vol. 69, 151-164.

Swain, D., Wade, E., 1993a. Density-dependent geographic distribution of Atlantic cod, *Gadus morhua* in the Southern Gulf of St. Lawrence. Canadian Journal of Fisheries and Aquatic Science Vol. 50(4) 725-733.

Swain, D. P., Benoit, H. P., Hamill, M. O., 2015. Spatial distribution of fishes in a Northwest Atlantic ecosystem in relation to risk of predation by a marine mammal. Journal of Animal Ecology, 10.1111/1365-2656.12391.

Swain, D. P., Wade, E. J., 1993b. Density dependent geographic distribution of Atlantic cod, *Gadus morhua* in the Southern Gulf of St. Lawrence. Canadian Journal of Fisheries and Aquatic Science Vol. 50, 725-733.

Tamdrari, H., Castonguay, M., Brethes, J., Duplisea, D., 2010. Density-independent and -dependent habitat selection of Atlantic cod (*Gadus Morhua*) based on goeostatistical aggregation curves in the Northern Gulf of St. Lawrence. ICES Journal of Marine Science Vol. 67, 1676-1686.

Taylor, B. M., Diggle, P. J., 2012. INLA or MCMC? A tutorial and comparitive evaluation for spatial prediction in log-Gaussian Cox processes. http://arxiv.org/abs/1202.1738v2.

Thorson, J. T., Shelton, A. O., Ward, E. J., Skaug, H. J., 2014. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for west coast groundfishes. ICES Journal of Marine Science, 10.1093/icesjms/fsu243.

Trumble, R., Neilson, J., Bowering, W., McCaughran, D., 1993. Atlantic halibut (*Hippoglossus hippoglossus*) and Pacific halibut (*H. stenolepis*) and their North American fisheries. Canadian Bulletin of Fisheries and Aquatic Science Vol 227, 84pp.

Trzcinski, M., Armsworthy, S., Wilson, S., Mohn, R., Campana, S., 2011. A framework for the assessment of the Scotian Shelf and Southern Grand Banks Atlantic halibut stock. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/002.

Trzcinski, M., Bowen, W., 2016. The recovery of Atlantic halibut: A large, long-lived, and exploited marine predator. ICES Journal of Marine Science Vol. 73(4), 1104-1114.

Trzcinski, M. K., Mohn, R., Bowen, W. D., 2006. Continued decline of an Atlantic cod population: How important is grey seal predation. Ecological Applications Vol. 16(6), 2276-2292.

Trzcinski, M. K., Mohn, R., Bowen, W. D., 2009. Estimating the impact of grey seals on the Eastern Scotian Shelf and Western Scotian Shelf cod populations. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/052.

Tucker, S., Bowen, W. D., Iverson, S. J., 2008. Convergence of diet estimates derived from fatty acids and stable isotopes within individual grey seals. Marine Ecology Progress Series Vol. 354, 267-276.

Vanmarcke, E., 1988. Random Fields, Analysis and Synthesis. MIT Press.

Walters, C., Maguire, J., 1996. Lessons for stock assessment from the Northern cod collapse. Reviews in Fish Biology and Fisheries Vol. 6, 125-137.

Watanabe, S., 2009. Algebraic Geometry and Statistical Learning Theory. Cambridge University Press.

Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research Vol. 11, 3571-3591.

Whittle, P., 1953. On stationary processes in the plane. Biometrika 41(3/4), 434-449.

Whittle, P., 1963. Stochastic processes in several dimensions. Bulletin of the Institute of International Statistics 40, 974-994.

Wiens, J., 1976. Population responses to patchy environments. Annual Review of Ecology and Systematics Vol. 7, 81-120.

Winters, G. H., Wheeler, J. P., 1985. Interaction between stock area, stock abundance and catchability coefficient. Canadian Journal of Fisheries and Aquatic Science Vol. 42, 989-998.

Woolford, D. G., Braun, W. J., 2007. Convergent data sharpening for the identification and tracking of spatial temporal centers of lightning activity. Environmetrics Vol. 18, 461-479.

Worm, B., Myers, R., 2003. Meta-analysis of cod-shrimp interaction reveals top down control in oceanic food webs. Ecology Vol 84(1), 162-173.

Wright, P. J., Galley, E., Gibb, I. M., Neat, F. C., 2006. Fidelity of adult cod to spawning grounds in Scottish waters. Fisheries Research Vol. 77, 148-158.

Zuur, A., Ieno, E. N., Walker, N., Saviliev, A. A., Smith, G. M., 2009. Mixed Effects Models and Extensions in Ecology with R. Springer.

Appendix A

A Derivation of the Conditional Independence in a Gauss Markov Random Field

Let us consider a simple GMRF as in Figure A.1.

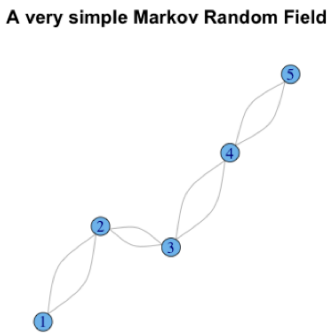

Figure A.1: **A simple MRF.** A simple graph of 5 nodes.

Given the definition of the GMRF, a mesh with n nodes implies an n-dimensional multivariate normal distribution. Since we are defining a precision matrix Q to be non-zero iff node i and node j share an edge, that is, node i and node j are neighbours. This Markov Blanket idea for our simple graph gives this precision matrix:

$$Q = \begin{bmatrix} a & f & 0 & 0 & 0 \\ f & b & g & 0 & 0 \\ 0 & g & c & h & 0 \\ 0 & 0 & h & d & i \\ 0 & 0 & 0 & i & e \end{bmatrix}$$

The covariance matrix, Σ, is dense, even though Q is not.

$$\Sigma = Q^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

The conditional independence that we want to show holds that the value of the field at location i is independent of that at j, given the values at the neighbours of i. That is $X_i \perp X_j | X_{-\{i,\delta_i\}}$, where δ_i is the set of neighbours of i.

To show conditional independence of the value of the GMRF at any node given only the neighbours, we will show that X_1 and X_3, the values of the field at nodes 1 and 3 respectively, are conditionally independent given the neighbours (2,4). That is, we want the distribution of $X_1, X_3 | X_2, X_4$. We can ignore X_5 since the distribution of any subset of variables of a multi-variate normal can be expressed by simply selecting the appropriate means, and covariances from the original matrices $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

So while our GMRF is $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we need work only with the following $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\boldsymbol{\mu} = E \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}, \Sigma = Q^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}$$

$$= \frac{1}{C} \begin{bmatrix} bcd - bci^2 - bh^2 - dg^2 & -(cd - h^2)f & dgf & -hgf \\ -(cd - h^2)f & a(cd - h^2) & -dga & hga \\ dgf & -dga & (ab - f^2)d & -h(ab - f^2) \\ -hgf & hga & -h(ab - f^2) & abc - ag^2 - cf^2 \end{bmatrix}$$

We will rearrange

Where $C = (abcd - abh^2 - adg^2 - cdf^2 + f^2h^2)$

The covariance matrix, Σ, is dense, even though Q is not.

To show conditional independence of the value of the GMRF at any node given only the neighbours, we will show that X_1 and X_3, the values of the field at nodes 1 and 3 respectively, are conditionally independent given the neighbours (2,4). That is, we want the distribution of $X_1, X_3|X_2, X_4$. We can ignore X_5 since the distribution of any subset of variables of a multi-variate normal can be expressed by simply selecting the appropriate means, and covariances from the original $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

$$\boldsymbol{\mu}_{1234} = E \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}, \Sigma_{1234} = Q_{1234}^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}$$

We want the distribution of $X_1, X_3|X_2, X_4$. We can rearrange (and partition) our matrices $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as:

$$\boldsymbol{\mu}_{1324} = E \left[\begin{array}{c} X_1 \\ X_3 \\ \hline X_2 \\ X_4 \end{array}\right] = \left[\begin{array}{c} \mu_1 \\ \mu_3 \\ \hline \mu_2 \\ \mu_4 \end{array}\right] = \left[\begin{array}{c} \mu_A \\ \hline \mu_B \end{array}\right], \Sigma_{1324} = \left[\begin{array}{cc|cc} \sigma_{11} & \sigma_{13} & \sigma_{12} & \sigma_{14} \\ \sigma_{31} & \sigma_{33} & \sigma_{32} & \sigma_{34} \\ \hline \sigma_{21} & \sigma_{23} & \sigma_{22} & \sigma_{24} \\ \sigma_{41} & \sigma_{43} & \sigma_{42} & \sigma_{44} \end{array}\right] = \left[\begin{array}{c|c} \Sigma_{AA} & \Sigma_{AB} \\ \hline \Sigma_{BA} & \Sigma_{BB} \end{array}\right]$$

Once partioned like this, the conditional distribution of $X_1, X_3|X_2, X_4$ has

$$Mean = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(\begin{bmatrix} X_2 \\ X_4 \end{bmatrix} - \mu_B),$$

and the covariance of X_1 and X_3, if conditioned is

$$\begin{aligned} Covariance &= \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \\ &= \frac{1}{C}\begin{bmatrix} bcd - bci^2 - bh^2 - dg^2 & dgf \\ dgf & (ab - f^2)d \end{bmatrix} \\ -\frac{1}{C}\begin{bmatrix} -(cd-h^2)f & -hgf \\ -dga & -h(ab-f^2) \end{bmatrix} &\begin{bmatrix} \frac{abc-ag^2-cf^2}{ac} & -\frac{hg}{c} \\ -\frac{hg}{c} & \frac{cd-h^2}{c} \end{bmatrix} \frac{1}{C}\begin{bmatrix} -h(ab-f^2) & -hgf \\ -dga & -(cd-h^2)f \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{c} \end{bmatrix} \end{aligned}$$

illustrating the conditional independence.