

SPATIAL POINT PROCESS ANALYSIS OF AT SEA SEAL
ENCOUNTERS

by

Stuart Carson

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2011

© Copyright by Stuart Carson, 2011

DALHOUSIE UNIVERSITY

DEPARTMENT OF MATHEMATICS AND STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "SPATIAL POINT PROCESS ANALYSIS OF AT SEA SEAL ENCOUNTERS" by Stuart Carson in partial fulfillment of the requirements for the degree of Master of Science.

Dated: December 9, 2011

Supervisor:

Readers:

DALHOUSIE UNIVERSITY

Date: December 9, 2011

Author: Stuart Carson

Title: SPATIAL POINT PROCESS ANALYSIS OF AT SEA SEAL
ENCOUNTERS

Department or School: Department of Mathematics and Statistics

Degree: M.Sc.

Convocation: May

Year: 2012

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions. I understand that my thesis will be electronically available to the public.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in the thesis (other than brief excerpts requiring only proper acknowledgement in scholarly writing) and that all such use is clearly acknowledged.

Signature of Author

To Evie, who has changed everything.

Table of Contents

List of Tables	vii
List of Figures	viii
Abstract	xiii
List of Abbreviations and Symbols Used	xiv
Acknowledgements	xv
Chapter 1 Introduction	1
Chapter 2 Statistical Methods for Spatial Data	4
2.1 The Literature	4
2.2 Types of Spatial Data	5
2.3 Spatial Point Processes	7
2.4 Estimation and Inference	12
2.5 Model Fitting	15
Chapter 3 Analysis of At Sea Seal Encounters	26
3.1 Data	26
3.2 Exploratory Analysis	28
3.3 The Homogeneous Poisson Model	31
3.4 Inhomogeneous Poisson Models	36
3.5 Cluster Processes	47
3.6 Including Mark Information	51
3.7 Interpretation	57
3.8 Validation	63

Chapter 4	Conclusions	70
4.1	Future Work	72
Appendix A	Candidate Inhomogeneous Models	73
Appendix B	Code	75
Bibliography		86

List of Tables

Table 3.1	Summary of candidate models for $\hat{\lambda}(u)$	42
Table 3.2	AIC Table of AIC values for each evaluated model.	44
Table 3.3	Summary of candidate models for $\hat{\kappa}(u)$	55
Table 3.4	AIC Table of AIC values for parent point models.	57
Table A.1	Models Summary of models for $\hat{\lambda}(u)$	73
Table A.2	AIC Table of AIC values for each evaluated model.	74

List of Figures

Figure 2.1	Homogeneous Vs. Inhomogeneous Two Poisson process realizations. The left panel is homogeneous, the right has $\lambda(x) = 1000^x$. Each has 140 points.	8
Figure 2.2	Distance measures Empty space distances (left panel) and a Steinen diagram showing each point surrounded by a circle equal to its nearest neighbour distance (right panel).	18
Figure 2.3	Diagnostic functions F, G, J and K for the seal encounter data.	21
Figure 2.4	Inhomogeneous K function Plot of $K_{Inhom}(r)$ for the seal encounter data. The plot is indicative of the presence of clustering in the parent points at distances less than .17 degrees, (approximately 18 km).	23
Figure 2.5	Envelopes for Inhomogeneous K function Plots of $K_{Inhom}(r)$ versus envelopes for these seal encounter data against the point-wise (left panel), and global (right panel), envelopes generated by 39 and 19 realizations of CSR respectively, $\alpha=.05$ for each.	25
Figure 3.1	Seal encounters off Sable Island	27
Figure 3.2	Window Choice. Illustration of the progressively smaller candidate windows discussed in the text.	29
Figure 3.3	Impact of Window choice. Ripley's K: Top Left, Large rectangle. Top Right, Bounding Box Inflated by 1.3. Bottom Left, Bounding Box. Bottom Left, Convex Hull. All show clustering. Effect is magnified in larger windows, the K function increases faster in a larger window for any given data set.	30
Figure 3.4	Estimated window Degree of inflation is small in relation to a simple convex hull. Both are overlaid on the data.	31

Figure 3.5	Complete Spatial Randomness The left panel displays the actual at sea seal encounter pattern to compare with a realization of a homogeneous Poisson pattern with the same number of points. There is no fit on the right.	32
Figure 3.6	Complete Spatial Randomness The left panel displays the actual seal A novel approach to studying at-sea associations in marine predators pattern within its window superimposed upon a homogeneous Poisson intensity (the blue); λ is a constant. The right panel shows residuals.	33
Figure 3.7	Quadrat Counts for test of CSR. The test rejects the hypothesis of a homogeneous Poisson process.	34
Figure 3.8	Global Envelope for Ripley's K function Plot of $\hat{K}(r)$ for the seal encounter data set against the global envelope generated by 19 realizations of CSR. This corresponds to the rejection of a null hypothesis of CSR.	35
Figure 3.9	Kernel Estimated Density The left panel displays the intensity estimate for the complete data set using a kernel smoother. The right panel has the actual locations superimposed.	37
Figure 3.10	The Inhomogeneous Poisson The left panel displays the actual seal to seal encounter pattern within its window to compare with a realization of a Inhomogeneous Poisson pattern with the same number of points in the same window and the intensity function defined as the kernel estimated density. There is much better fit.	37
Figure 3.11	Global Envelope for Inhomogeneous K function The plot of $K_{Inhom}(r)$ for the seal encounter data against the global envelope generated by 99 realizations of an Inhomogeneous Poisson process with the kernel estimate as the intensity function. There is no reason to reject the hypothesis that the data was generated by a process with the stated intensity. $\alpha = .01$	38

Figure 3.12	Quadrats determined by bathymetry and distance for test of CSR	41
Figure 3.13	Fitted Intensity Models The predicted intensity functions for the fitted models.	43
Figure 3.14	Inhomogeneous Poisson The left panel displays the actual seal to seal encounter pattern within its window to compare with a realization of a Inhomogeneous Poisson pattern with the same number of points in the same window. The intensity function of the realization is the fitted intensity of model 03.	45
Figure 3.15	Parametric Inhomogeneous Poisson Intensity Estimate The left panel displays the actual seal to seal encounter pattern within its window superimposed upon an Inhomogeneous Poisson intensity function; the predicted intensity of the parametric model 03. The right panel illustrates a remaining lack of fit, particularly between $r=.02$ and $r=.10$	46
Figure 3.16	Thomas Inhomogeneous Intensity Estimate The left panel displays the actual seal to seal encounter pattern within its window superimposed upon the Inhomogeneous mean intensity function for the Thomas process; that is, the intensity of the parent points. The right panel shows a simulation from the model. The fit is poor.	50
Figure 3.17	Matérn Inhomogeneous Intensity Estimate The left panel displays the actual seal to seal encounter pattern within its window superimposed upon the Inhomogeneous mean intensity function for the Matérn process; that is, the intensity of the parent points. The right panel shows a simulation from the model. The fit is poor.	50

Figure 3.18	Simultaneous Confidence Envelopes for the Thoams and Matérn Inhomogeneous cluster models The left panel displays the global envelope for the Thomas model. The right panel that of the Matérn model fit by minimum contrast. Both display a lack of fit in that they predict too few points at short separations and too many at long separations.	51
Figure 3.19	Histogram of Time Separation between Points Difference between observation times (in minutes). The 2 minute refresh rate of the VMT gives several data for each encounter.	54
Figure 3.20	Temporal Clusters. The circles are proportional to the number of re-detections associated with each Distinct Encounter	54
Figure 3.21	Initial Encounters. Locations of the 122 parent points making up the reduced data set.	55
Figure 3.22	Fitted Intensity Model The fitted models, Model 12, Top left. Model 13, Top right. Model 14, Bottom left. Model 15, Bottom right.	56
Figure 3.23	Fitted Intensity of Parent process Plot of our parametric inhomogeneous Poisson model for the intensity of the parent points. The right panel has the 122 points identified as parents superimposed.	58
Figure 3.24	Simulated Patterns from our Thomas and Matérn models The Thomas and Matérn models created using the new parameters.	58
Figure 3.25	Simulated Patterns from our Thomas and Matérn models The Thomas and Matérn models created using the new parameters.	59
Figure 3.26	Effect Function and Histogram of depth Left Panel. Effect function. Right Panel. Histogram of water depth at point of encounter, overlaid is the effect function.	60

Figure 3.27	Lurking Variable plot for depth	Lurking variable plot for depth, shows cumulative residuals as a function of the covariate. Compare to locations versus the 40, 60 and 80 foot contours. . . .	61
Figure 3.28	Histogram of distance	Histogram of distance from point of encounter to Sable Island. The effect function for distance (smooth line) is overlaid. Correspondence is good. Scale is 10 minutes of arc.	62
Figure 3.29	Overall view of the 2010 locations of Seal to Seal encounters off Sable Island	65
Figure 3.30	The 2010 data and the 2009 Window	Left Panel. The 2010 data require a larger window. Right Panel. The 2009 data and 2010 subset in the 2009 Window. 2009 locations in black.	65
Figure 3.31	Simulated Patterns from our Matérn model	The Matérn model is not plausible for the 2010 data.	66
Figure 3.32	Effect Function and Histogram of depth	Left Panel. Effect function and Histogram of water depth. Right Panel. Effect function and Histogram of distance.	67
Figure 3.33	Point Patterns of Female Encounters	Left Panel, The 2009 Female Encounters. Center Panel, The 2010 Female Encounters. Right Panel, the Superimposed Patterns	68
Figure 3.34	Simulated Patterns from our Matérn model	The male and female seals display different clustering.	69

Abstract

We consider a data set comprising encounters between grey seals off Sable Island in the autumn of 2009. We apply Spatial Point Pattern data analysis approaches to this data set and demonstrate the effectiveness of these techniques. Specifically we propose using point processes for clustered data; the Inhomogeneous Poisson process, and the subset of cluster processes called Cox processes. We utilize maximum likelihood methods to produce a model for the mean intensity of the seal encounters which includes spatial covariates. We then consider the impact of inhomogeneity and review minimum contrast methods for fitting the Cox process models, specifically the Thomas and Matérn cluster processes. We demonstrate that these methods allow us to gain valuable insight into the nature and extent of grey seal encounters at sea. Finally we validate both our conclusions and the approach taken using the most recent data available, that of autumn 2010.

List of Abbreviations and Symbols Used

OTN	Ocean Tracking Network
VMT	VEMCO Mobile Transceiver
λ	First Order Intensity
$\lambda(u)$	First Order Intensity
$\hat{}$	Indicates an Estimate
$\lambda_2(u)$	Second Order Intensity
W	Spatial Sampling Window
κ	First Order Parent Intensity of a Cluster Process
$\kappa(u)$	First Order Parent Intensity of a Cluster Process
μ	Expected Number of offspring of a Cluster Process
$\mu(u)$	Expected Number of offspring of a Cluster Process
σ	Standard Deviation
$F(r)$	Empty Space Function
$G(r)$	Nearest Neighbour Function
$K(r)$	Ripley's K Function
$\rho(r)$	Pair Correlation Function
CSR	Complete Spatial Randomness
α	Significance Level
AIC	Akaike Information Criterion

Acknowledgements

This thesis owes a tremendous debt to many who lent me their support. Dr. Joanna Mills Flemming, my supervisor, who first proposed the topic to me, whose energy and enthusiasm got me moving, and whose weekly meetings kept me moving. Dr. Sara Iverson, for her valued insight into grey seal biology and her patience as a reader of the eventual work.

Dr. Don Bowen's wealth of knowledge about the seals of Sable Island, what they do and where they go, added greatly to the analysis herein and provided an excellent target at which to aim. Ian Jonsen and Damian Lidgard are owed my gratitude for their assistance and encouragement throughout. Dr. Bruce Smith, who, in addition to providing me access to his personal library, also kindly acted as reader for the finished thesis.

I must also acknowledge the Ocean Tracking Network (OTN), which generously provided funding for the last leg of this Masters degree.

The greatest credit of all however goes to my wife, Pam Woodman, whose belief in my ability to finish this project eventually proved contagious, and to my daughter Evelyn, age 1, still young enough to believe that her Daddy is made of awesome.

Chapter 1

Introduction

This thesis discusses some of the more important recent applications of spatial statistics in general, and point pattern data analysis in particular, appearing in the literature; with particular emphasis on those appearing in the marine realm. A thorough review of techniques is presented followed by an illustrative analysis of a data set comprised of at sea encounters between grey seals off Sable Island, Nova Scotia.

The Ocean Tracking Network (OTN), (www.oceantrackingnetwork.org), is a major research effort headquartered at Dalhousie University. The OTN has an overall objective of fostering the conservation and sustainable use of the world's oceans. To effect this objective requires knowledge of the oceans, and knowledge of the movements and habitats of marine animals and their relationships to the ocean environment. While the scope of the OTN project is global, one specific area of exploration is the biology and behaviour of marine life. Under the auspices of the OTN, many marine species are being equipped with acoustic transmitters in order to track their movements and behaviour at sea. Lines and arrays of oceanic sensors are also being deployed as part of this effort to monitor environmental factors such as temperature and salinity in the oceans and to understand the ocean and its animal populations.

Also, as part of the OTN, new technologies are being explored to gather information heretofore unavailable, one example being the deployment of a new type of acoustic tag, a mobile transceiver called a VEMCO Mobile Transceiver (VMT). Instead of acting as a recorder of animal movement the VMT records instances of proximity to other tagged animals. In a prototype project the OTN has deployed VMTs on the Sable Island grey seals (*Halichoerus Grypus*). This choice is due in part to the need to recover the VMTs; the data is archival and the tag must be

retrieved in order to obtain the recorded information. Since the seals return to Sable Island to breed they constitute a viable option. Of immediate interest here are the 14 grey seals on Sable Island that were randomly selected and fitted with VMT tags in 2009. Seals so instrumented are referred to as "bioprobes", reflecting the fact that the organism itself is being used as the means of interrogating its environment.

These bioprobes record a contact whenever one comes within approximately 500m of another. Thus they deliver a data set comprising the times and locations of the pair-wise proximities of the tagged animals. Later it is hoped that by tagging fish, e.g. Atlantic Cod (*Gadus Morhua*), a record of the bio-probe's encounters with other species may be recorded. Encounters with such fish, (as potential prey), may yield further insight into the roles of, and relationships between, the species of marine life in the study area. Data of this type, that is, times and locations of encounters between animals, has simply not been available previously due to the infeasibility of collecting it prior to the availability of technology such as the VMT.

Traditionally, tagging efforts interested in behaviour have provided data on the location of the tagged animal. Often utilizing satellite technology to track the location of the marked animal over time, these tags have also recorded additional parameters such as depth, salinity, speed etc. From these data models of the animal movement and behaviour have been developed [17].

Mark-recapture studies capture and tag animals at a specific location and time and then attempt to relocate or recapture the animal at future time(s). Data acquired via these means can then be used to draw inference about many aspects of the study population such as survival, mortality and movement. The statistical techniques and models for describing and analyzing data of this type are well developed [27], but what has not been readily accessible before is animal behaviour between capture and recapture. The bioprobes offer the potential to capture, in the form of these encounters, some of that behaviour.

The deployment of VMT technology is yielding a new data type that lends itself to a different means of analytic approach. Bio-probe data consists of locations,

(points), distributed across a space, through time. There is either an encounter at a given location or there is not, and of course, we observe only the encounters. Data of this spatial (or spatial-temporal) type may be viewed as a point pattern. If we ignore the time component then the encounter locations become a pure point pattern in the 2D plane. The methods of point pattern analysis are then appropriate. We note that ignoring the time component is not unreasonable, as here the time frame of the experiment is essentially one season and we are interested in where and how the seals relate to one another and to the ocean environment. That is; where do the seals encounter one another during the season, how often, and what oceanographic or environmental factor(s) may be related? Some illustrative examples of point processes are given by Baddeley [14].

1. Earthquake aftershocks. The latitude, longitude and time of earthquake aftershocks. Author notes that by ignoring the time component this becomes a spatial point pattern.
2. Locations of petty crime. Plotted on a map these can be analysed as a point pattern; there is an issue as to whether any location is valid or whether the points exist only on the street grid, perhaps making this lattice type data.
3. Locations of tree types. Tree locations in a natural stand. Are the locations random by species?

Our encounter data are of this general form, points in the plane, and to consider the structure of these encounters we will demonstrate the utility of spatial methods. Chapter 2 will discuss in detail the means and methods contained in the works above. Chapter 3 is the heart of this thesis, containing a description of our data set followed by analysis and modelling, and concluding with a discussion of the results and their biological context. Chapter 4 provides conclusions and discusses further work that is possible and desirable.

Chapter 2

Statistical Methods for Spatial Data

2.1 The Literature

When considering data consisting of a point pattern we seek statistical approaches that can address the questions *how many points are where*, and *how do the points influence one another*? In the context of our particular application we seek to answer not only whether and how many encounters the seals may have but also to determine if there is a pattern to the encounters. That is, are the encounters randomly distributed spatially and/or temporally or can we discern structure? Further, can we relate this structure to factors of the physical environment?

The field of spatial statistics is an area of study that presents some interesting aspects, drawing together as it does elements from many varied disciplines, both from within and without traditional statistics. The examples come from everywhere; geology, botany, physics or meteorology, it doesn't seem to matter, spatial statistics have a role [26].

In order to gain insight into spatial statistics in general, and spatial point pattern analysis in particular, several comprehensive works were consulted. Cressie [8] provides an overview of the subject and in no particular order the books by Schabenberger and Gotway [26], Daley and Vere-Jones [11] and Møller and Waagepetersen [20] were often consulted. The field of *stochastic geometry*, as surveyed by Stoyan, Kendall and Mecke [10], shares theoretical aspects with the study of point processes and proved useful. The theory of *Random fields*, upon which the *doubly stochastic* processes of Cox [7] depend, are to be found in Vanmarcke [29].

Ripley's work [24] provides, in its modified form, the analytic and diagnostic tool for point processes central to the analysis herein. Discussed by all above except Vanmarcke, (who is not discussing point patterns), Ripley's method of *second order analysis* has become a popular tool for the analysis of point pattern

data. Ripley's original work was modified and extended significantly by Baddeley, Møller and Waagepetersen [1]. This extension figures prominently herein.

These fundamental writings on the topic of point patterns and point pattern data analysis led to consideration of the ideas of Cox [7] and Matérn, [21] as likely candidates for our particular data. Interesting, informative and illuminative examples of modern applications of these ideas include the identification and tracking of spatial and temporal lightning clusters [31], using spatial point processes to model forest fire outbreaks [19]. As was briefly mentioned in the introduction, pattern data in the marine biology 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 [22]. Indeed O'Driscoll [22] 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 [3] where locations were available and were examined in relation to the spatial distribution of the fishery activities as a whole. The only attempt to link the spatial distribution of a seagoing animal to the oceanic environment itself is found in Royer, Fromentin and Gaspar [15]. 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.

The remainder of Chapter 2 will discuss in more detail the means and methods contained in the works above.

2.2 Types of Spatial Data

Cressie [8] classifies Spatial Data as being of 3 main types: geostatistical, lattice and point pattern.

Geostatistical 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 example given in Schabenberger [26] is that of air temperature over the Continental US. The domain is fixed by the geographical boundary, and the response is continuous and could be measured anywhere, at least theoretically. *Where* it is measured does not change this property of the response.

Lattice or regional data

Spatially aggregated data falls into this category. 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 i th site, is an average for the i th 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 geostatistical 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 investigation of these data concerns the properties of (D^*) itself and may require the estimation of (D) if unknown. A data set of this type could be as simple as one where the indicator, $I(x) = 1$, if $x \in D^*$. 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*.

The seal to seal encounters are of the third type, point pattern data; each season is a single realization of the process. If the experiment of 2009 is repeated, as has been done in 2010, and is intended for 2011, then these and subsequent realizations will each have a different (random) number of seal to seal encounters and a different set of locations (D^*). The attributes, such as which pair of seals

are present as well as the descriptive attributes of the seals themselves, will differ, but attributes will be included. Thus each set of seal encounter data constitutes a *marked* point pattern, a separate realization, generated by a (common) point pattern process.

2.3 Spatial Point Processes

A point pattern data set is one composed of the locations of events or objects in the region of study. In a two-dimensional point pattern, (\mathbf{X}) would comprise points (x) in the (2D) plane. The addition of a time component yields a spatial-temporal or space-time pattern, (x,t) . The 2D locations of the events are often referred simply as $x_i (i = 1, \dots, n)$, (\mathbf{X}) referencing a 2D process.

Characterization of a Spatial Point Process

A Spatial Point Pattern consists of the locations of the events (x_i) , in a bounded area on the plane R^2 , and is characterized by its first and second order intensities. The pattern of points, or patterns if more than one is available, is a realization of a spatial point process. So, for example, consider events in the plane that are homogeneously, independently and identically distributed. The points constitute a pattern, where the underlying process is said to be a homogeneous Poisson process on the plane. A second pattern could be distributed the same as the first and would be a second realization of the same process. The locations of the points would be different for this second realization; but the underlying process would still be the homogeneous Poisson.

The *first order intensity*, λ of a point pattern is the expected number of points per unit area. Often called simply the *intensity*, this may be constant or it may vary from place to place, and is referred to as being *homogeneous* or *inhomogeneous* respectively. See Fig. 2.1 for an illustration of this important distinction.

The *second order intensity*, $\lambda_2(x, y)$, is the intensity associated with a second point of the process being at y , given that one at x exists. Examining the nature of this second order intensity may reveal the nature of the influence between points, that is, the *interaction*.

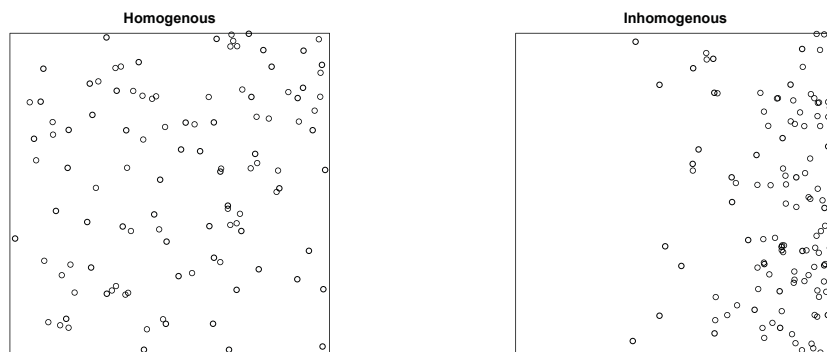


Figure 2.1: **Homogeneous Vs. Inhomogeneous** Two Poisson process realizations. The left panel is homogeneous, the right has $\lambda(x) = 1000^x$. Each has 140 points.

In the context of spatial point processes, *interaction* refers to the relationship, or dependence, between points. The points in the pattern may be *independent*, *regular* or *inhibitive*, in which the points tend to "repel" one another, or *clustered*, in which the points are "attracted" to one another. Attraction and repulsion are used to characterize situations in which the probability of a second point appearing within a given distance from a given point in the pattern is higher or lower, respectively, than if purely random.

Marks are additional information attached to the points comprising the pattern. A mark can be categorical or continuous; an example could be which pair of tagged seals are associated with a given encounter.

A point process is *stationary* if, $\lambda(u) = \lambda(u + h)$, that is, it depends only on the distance between the two points.

A point process is *anisotropic* if, $\lambda(u + h)$ is a function of both the magnitude and the direction of h . That is, the process varies in space differently dependent upon direction. A process lacking such directional dependence is referred to as *isotropic*.

The *intensity measure*; the expected number of points in the pattern, is also

often used.

$$\mu(W) = \int_W \lambda(u) du$$

The standard model for a spatial point process (X) generating the pattern assumes that the process extends throughout some 2D space, $R \subseteq R^2$, but is observed in a sampling *window*, (W). In the basic model W is fixed and known. If W is not known it must either be defined or estimated since W is in fact part of the process.

Poisson Point Processes

Poisson point processes are the simplest and most important class of point process models. Homogeneous Poisson point processes represent "complete spatial randomness" with "no interaction" between points [14]. The study of summary statistics for point processes also relies frequently upon the Homogeneous Poisson process as reference.

If the *intensity*, or *intensity function* of the Poisson process is a constant, $\lambda(x) = \lambda$, then the process is a *homogeneous* Poisson process with intensity λ . The special case $\lambda(x) = 1$ is the *standard* or *unit rate* Poisson process. If the intensity function $\lambda(x)$ is not a constant then the Poisson process is *inhomogeneous*. It is with these simple models that we begin the analysis of the seal encounter data set in Chapter 3, first looking at, and easily rejecting, Complete Spatial Randomness (CSR). We then examine the possibility of modelling the seal encounters using an inhomogeneous Poisson model; that is, proposing that the data can be modelled as independent points drawn from a Poisson process with a spatially varying mean or rate parameter. This is a much more realistic model than CSR.

Poisson Cluster Processes

More sophisticated yet are the *cluster processes* which introduce the notion of *parents* and *offspring*. Cluster Processes are two stage process with *parent* points, a spatial point process for the cluster centers, and *offspring*, a spatial point process for additional points associated with each cluster center. The complete point

pattern that is observed is the *superposition* of the two patterns. We consider the cluster centers as points of a Poisson process. When the parent points are points of a Poisson process, homogeneous or inhomogeneous, the result is termed a *Poisson cluster process*. Normally it is not possible to distinguish between a cluster process and an inhomogeneous Poisson process with only a single realization. Neither is it normally possible to determine which points in the superposition constitute parents and which are offspring. Making this distinction requires some special feature(s) of the point(s), inherent in the marks. This approach is taken in Chapter 3.

Cox Processes

Cox [7], described *doubly stochastic Poisson processes*, where he considered the intensity function $\lambda(x)$ as the realization of a *random field*. These processes are now known as Cox processes.

A random field may be viewed as an indexed set of random variables, $X(t)$. The collective outcome is denoted $x(t)$, this is a realization of the field. In the discrete case the outcome, $x(t)$, is the discrete state of the response at t . A *random point process* results if the $x(t)$ are points (or, in our case, small objects), located in a random pattern in some space.

Cox took a random field $\Lambda = \Lambda(x) : x \in D$, where X , the points, given Λ , is a Poisson process on D . X is said to be a *Cox process driven by Λ* .

The intensity function for a Cox process then is $\lambda(x) = \mathbb{E}\Lambda(x)$, and the intensity measure is $\mu(D) = \int_D \Lambda(x)dx$. If Λ is deterministic then X is a Poisson process and $\lambda = \Lambda$ and the Poisson process above is a special case.

Shot Noise Cox Processes are so named following Rice's "shot noise" model [23] for electronics, where shot noise consists of randomly timed and randomly sized impulses. The parent points have an intensity function as above, and the offspring points, x_i , form a spatial process, $U(X)$, conditional on a set of parent points, $c_j, j = (1, 2, \dots)$:

$$U(x) = \sum_j \gamma K(c_j, x)$$

The γ parameter is the expected number of offspring points associated with each parent and K is a kernel defining the distribution of the offspring about the parent.

A *Matérn* process [21] defines γ as a $\text{Poisson}(\mu)$ variable and K as the uniform distribution on the ball (disk) of radius R , $b(c_j, R)$. So, conditional on the parent points (c_j) each cluster has μ expected points distributed uniformly over the disk of radius R centered on the parent.

A *Thomas* process [28] defines γ as a $\text{Poisson}(\mu)$ variable and K as the $N_2(0, \sigma^2 I_2)$ density, an *isotropic* bivariate normal.

The Poisson cluster processes (Thomas, Matérn), can accommodate spatial inhomogeneity by generating the parent points from some inhomogeneous Poisson process whose intensity function is conventionally noted as $\kappa(u)$. The expected number of offspring in each cluster is denoted μ . There is also a need for a scale parameter, the standard deviation (σ) for the Gaussian distribution in the Thomas process, or R , the radius of the disk in the Matérn process. Inhomogeneity may also be modelled by making the distribution required for the offspring points inhomogeneous, $\mu(u)$. The overall intensity of the process, $\lambda(u)$ is then $\kappa(u)\mu$ or $\kappa\mu(u)$ respectively.

We consider these processes as candidates for describing the process that generated the seal encounter data and in Chapter 3 we discuss these data and the reasons for believing that the cluster centers, the *parents*, can be identified. Once parents identities are established the cluster models are natural choices.

Marked Point Processes

The point pattern data set may contain more information than simply the locations of the points. For instance, the points may have an identifiable characteristic; these characteristics may be of either of two types.

1. Continuous. A "real number". Each point in the pattern has a value attached for some measurement. Such patterns have *continuous marks*.
2. Categorical. A factor. The value of the factor denotes the type of point within the pattern. Such patterns are referred to as *multitype point patterns*.

Marks can be considered as a coordinate. For example, a data set of the form (longitude, latitude, time) may be considered a spatio-temporal pattern, or it may be considered as a spatial pattern in the plane with a continuous mark, time.

Previously we discussed a point pattern dataset as one composed of the locations of events or objects in the region of study. In a two-dimensional point pattern (\mathbf{X}) would comprise points (x) in the $(2D)$ plane. The $2D$ locations of the events are often referred simply as $x_i (i = 1, \dots, n)$, (\mathbf{X}) referencing a $2D$ process with some domain R , the space in which the process occurs. A marked pattern has the spatial domain R but also has the marks belonging to some set M . Such a pattern is a point pattern in $R \times M$. We now have $X = \{(x_1, m_1), \dots, (x_n, m_n)\}, x_i \in R, m_i \in M$, where the x_i are locations and the m_i are the marks.

Marks are to be distinguished from covariates. Each mark is a property of the datum itself and is measured at each datum. Covariates may be introduced when modelling a point pattern. A covariate must exist for each point in the domain D , not simply at each datum location (x_i) .

The seal encounter data are a marked point pattern. Additional information included with each location datum includes time of occurrence and the identities of the pair of seals involved. Since we treat the seal encounters as a spatial point process and not a spatio-temporal point process we regard the time of encounter as a mark and not as a co-ordinate.

2.4 Estimation and Inference

Estimating the Window

As mentioned above, a point pattern is observed within a *window*. If all that is observed is the points in the pattern then the *window*, W , must be determined or estimated. Ripley and Rasson [25] suggest a method for estimating an unknown window. Their method is to inflate either a bounding box or a convex hull, according to the number of data points and the number of vertices the convex hull will require. The Ripley-Rasson scaling factor (RS) for the convex hull is:

$$RS = \frac{1}{\sqrt{1 - \frac{m}{n}}}$$

where n is the number of data points and m is the number of vertices on the convex hull. Alternatively, for a bounding box the scaling factor is:

$$RS = \frac{1}{\sqrt{1 - \frac{4}{n}}}$$

We will use the method of Ripley and Rasson to estimate (W) for the seal to seal encounters.

Estimation of the Mean; uniform intensity.

The *intensity* of a point pattern is the expected number of points per unit area. This is analogous to the mean or expected value of a random variable. If the point process (X) is homogeneous then the expected number of points is proportional to the area of the window (W) ,

$$E[N(X \cap W)] = \lambda \cdot \text{area}(W)$$

where N is a counting operator. This is equally true for any subset of W .

$$E[N(X \cap V)] = \lambda \cdot \text{area}(V), V \subseteq W$$

and,

$$\hat{\lambda} = \frac{N(X)}{\text{area}(W)}$$

is an unbiased estimator for λ .

The simplest point pattern is the homogeneous Poisson process. The Poisson process has no interaction between points. Rejecting the Poisson null hypothesis is often seen as the first step towards fitting a more realistic model, and acts as a reference process for exploratory and diagnostic tools since a Poisson process with uniform intensity, $\lambda(u) = \lambda$, and no interaction (between points), connotes "Complete Spatial Randomness".

Estimation of the Mean; varying intensity.

If the intensity function $\lambda(u)$ is not a constant then a Poisson process is said to be *inhomogeneous*. If we assume that the expected number of points in a small area du around a location u equals $\lambda(u)du$ then $\lambda(u)$ is the *intensity function* of the process (X) [14] [26] and the expected number of points in some subset of the window (V) is,

$$E[N(X \cap V)] = \int_V \lambda(u)du, V \subseteq W$$

The homogeneous case is the special case where $\lambda(u)$ is a constant. Estimating the intensity at location u is required; in practise this is done by methods of quadrat counting or by methods of density estimation.

Kernel Estimation

A spatially varying estimate of intensity at location u can be found from an examination of the process in the area of u . Kernel estimation of this type yields a local estimate of the intensity function $\hat{\lambda}(u)$. If the neighbourhood h is small, $\hat{\lambda}(u)$ has small bias but large variance. Increasing h decreases variance but introduces bias. The exact form of the kernel is of less importance to the performance of the smoothing function than is the value specified for H [14]. Since the form of the kernel is not influential [14] we choose the Gaussian kernel for reasons of familiarity and ease of use:

$$k(t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right)$$

In this specific case the kernel will be bivariate normal with variance covariance matrix H , where H has the form:

$$H = \begin{pmatrix} \sigma^2x & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma^2y \end{pmatrix}$$

Appropriate values of the parameter H must be determined. Spherical contours (circles in $2D$) are common, but choosing different bandwidths is possible. A

bivariate normal kernel will yield elliptical contours if the variances are unequal. A non-zero co-variance will yield a rotation [26].

Bandwidth Selection

Methods for selecting the bandwidth parameter H are required. Options for estimating H are *Least Squares Cross Validation*, common in the literature of animal Range Utilization [18]. The value of H is chosen to minimize,

$$M(h) = \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{j=1}^n K^* \left(\frac{X_i - X_j}{h} \right) + \frac{2K(0)}{nh}$$

where $K^* = K^{(2)} - 2K$, and $K^{(2)}$ is a bivariate normal with variance-covariance matrix . Note that if the matrix is diagonal the kernels will be aligned with the axes of the data. That is, there is no rotation due to covariance.

Plug-in Bandwidth Selectors are also available for Bivariate Kernel Estimation. This method yields a non diagonal bandwidth matrix H [13]. Using the plug-in algorithm yields a variance-covariance bandwidth matrix identical to the *Least Squares Cross Validation* (LSCV) estimate.

2.5 Model Fitting

The R package **spatstat** [4] includes a model fitting algorithm [5] to fit a *point process model* to point process data. It allows the fitting of models which employ spatial covariates so long as the covariate $Z(u)$ is measured at every location u . The algorithm fits an inhomogeneous Poisson model for the intensity function in which intensity is either a loglinear function of the covariate(s):

$$\hat{\lambda}(u) = \exp(\beta_0 + \beta_1 Z(u))$$

or proportional to the covariates:

$$\hat{\lambda}(u) = \beta Z(u)$$

The homogeneous Poisson process above is a simple special case where,

$$\hat{\lambda}(u) = \hat{\lambda} = \exp(\beta_0)$$

Interaction: Inhibition, Independence and Clustering

Investigation of the pattern with the aim of determining if the points are inhibited (tending to avoid one another), independent or clustered (closer together than expected), can be achieved using means of distance based methods. Worthy of note is the issue of *edge effects*, which has substantial impact on distance methods. Recall that the basic point process model has the process continuing indefinitely in R but only viewed in the window W . If the nearest point of the process to a given point is not in the window then the distance based calculation will not take it into account, thereby introducing bias. In the seal encounter data we have encounters between instrumented seals and (leaving aside failure to detect on part of the tags themselves) since presumably we have them all, there is no pattern outside of the window and so we need not account for edge effects, That is, the nearest neighbour of any given point is also in the observed pattern. For the cluster models this implies that there are no points observed that are offspring of unobserved parents. This alternative construction views the point process as *finite*, existing only within a defined boundary or domain [11].

Diagnostic functions for Spatial Data

Distance Methods

Classical distance based measures include;

- Empty space distance

$$d(u) = \min_i \|u - x_i\| \text{ distance from a fixed reference } u \text{ to the nearest datum.}$$

- Nearest neighbour distance

$$t_i = \min_{i \neq j} s_{ij} \text{ for each point } x_i \text{ to its nearest neighbour.}$$

Each is discussed in detail below.

The empty space function F

The *empty space function* F , also called the *spherical contact distribution function* is a cumulative distribution function for the empty space distance being less than or equal to some arbitrary distance r .

$$F(r) = \mathbb{P}\{d(u, \mathbf{X}) \leq r\}$$

The empirical distribution is,

$$F^*(r) = \frac{1}{m} \sum_j \mathbf{1}\{d(u, \mathbf{X}) \leq r\}$$

The estimated F function is compared to a benchmark, the (homogeneous) Poisson process which has an empty space function

$$F_{Pois}(r) = 1 - \exp(-\lambda\pi r^2)$$

Using $\hat{\lambda} = \frac{n(x)}{\text{area}(W)}$ values of $\hat{F}(r) > F_{Pois}(r)$ suggests regularity and $\hat{F}(r) < F_{Pois}(r)$ suggests clustering.

The nearest neighbour function G

The *nearest neighbour function* G , is a cumulative distribution function for the nearest neighbour distance of a point in the pattern, u is an arbitrary point (the origin is a common choice), $\{d(u, \mathbf{X} \setminus \{u\})\}$ is the distance from u to a point in the pattern excluding u .

$$G(r) = \mathbb{P}\{d(u, \mathbf{X} \setminus \{u\}) \leq r \mid u \in \mathbf{X}\}$$

The empirical distribution is

$$G^*(r) = \frac{1}{n(x)} \sum_i \mathbf{1}\{t_i \leq r\}$$

The estimated F function is compared to a benchmark, the (homogeneous) Poisson process which has an empty space function

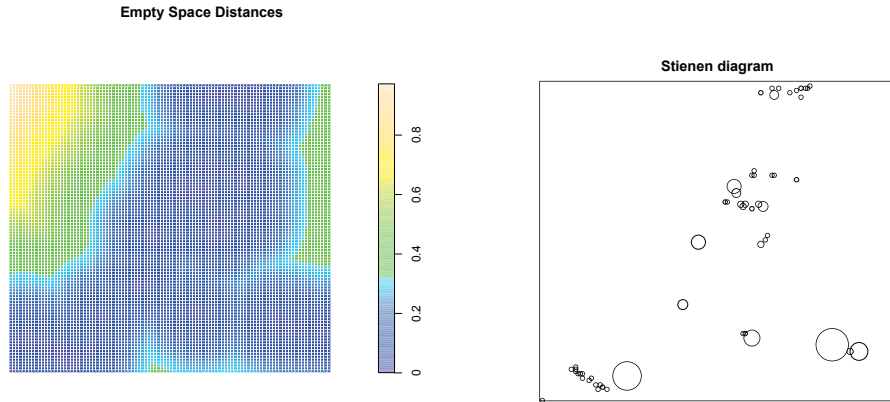


Figure 2.2: **Distance measures** Empty space distances (left panel) and a Steinen diagram showing each point surrounded by a circle equal to its nearest neighbour distance (right panel).

$$G_{Pois}(r) = 1 - \exp(-\lambda\pi r^2)$$

The result for $G(r)$ is the same as for $F(r)$ since the two are equivalent, however values of $\hat{G}(r) > G_{Pois}(r)$ suggests clustering and $\hat{G}(r) < G_{Pois}(r)$ suggests regularity, opposite to interpreting $\hat{F}(r)$. See Fig. 2.2 for a comparative illustration using our data.

Ripley's K function

Second order analysis

Ripley's K function, originally defined for stationary and isotropic point processes such that $\lambda K(r)$ is the expected number of points of the process not including u within distance r of the given point u ;

$$K(r) = \frac{1}{\lambda} \mathbb{E}\{n(\mathbf{X} \cap b(u, r) \setminus \{u\}) \mid u \in \mathbf{X}\}$$

$$K(r) = \frac{2\pi}{\lambda^2} \int_0^r x \lambda_2(x) dx$$

is a popular tool for analysis of spatial point data. The term $\lambda_2(x)$ is the *second order intensity*, the intensity of a second point of the process given that there is a point at x . If the process is homogeneous this depends only on the distance r . Intuitively, $K(r)$ is the expected number of further events within a distance r of an event x , divided by the intensity at that event. The quantity $\lambda K(r)$ then is the expected number of extra points within distance r of an arbitrary point in the process. For example, if the process is homogeneous with intensity λ , then the expected number of points is $\lambda\pi r^2$, and $K_{Pois}(r) = \pi r^2$.

Interpretation of the K function is by means of comparison to the known K function for the Poisson process.

$$K_{Pois}(r) = \pi r^2$$

Values of $\hat{K}(r) > K_{Pois}(r) = \pi r^2$ suggest clustering and $\hat{K}(r) < K_{Pois}(r) = \pi r^2$ suggest regularity.

Pair correlation function ρ

Also common is the *pair-correlation function*;

$$\rho(r) = \frac{\lambda_2(r)}{\lambda^2} = \frac{1}{2\pi r} \frac{dK(r)}{dr}$$

For a Poisson process $\rho(r) = 1$, $\rho(r) > 1$ indicates that pairs of points are more likely to occur in proximity (jointly) than for a Poisson process. $\rho(r) < 1$ indicates a repulsion.

Other related forms

Other functions that are used to study dependence in point patterns exist. They are related to $K(r)$ as follows:

The *radial distribution function*, or *empty space function* $F(r)$, reviewed already, may be related to $K(r)$:

$$F(r) = \lambda \frac{dK(r)}{dr}$$

and measures the rate of change of the expected number of extra points as r increases.

The L function;

$$L(r) = \sqrt{\frac{K(r)}{\pi}}$$

is commonly used instead of the K function. The L function and the K function are in one-to-one correspondence but the L function is the identity, $L(r) = I(r) = r$, for a Poisson process. For small values of r , $L(r) - r > 0$ indicates aggregation or clustering at distances less than r , $L(r) - r < 0$ indicates regularity or repulsion.

The J function combines F and G so that,

$$J(r) = \frac{1 - F(r)}{1 - G(r)}$$

Since $F_{Pois}(r) = G_{Pois}(r)$, $J_{Pois}(r)$ is always unity. Values of $J_{Pois}(r) < 1$ suggest clustering, values of $J_{Pois}(r) > 1$ suggest regularity.

Regrettably, both these distance based methods, $F(r)$ and $G(r)$ require the pattern to be both stationary and homogeneous, as does $K(r)$ in its original form above, meaning none can be applied to our data.

Only the 2D forms of these functions, $F(r)$, $G(r)$, $K(r)$, $J(r)$, $L(r)$ and $\rho(r)$ are provided for simplicity, although more complex general forms do exist for higher dimensions. See Fig. 2.3 for an example and comparison of the $F(r)$, $G(r)$, $K(r)$, and $J(r)$ functions. All represent the same data and each exhibits strong indication of clustering although the shape of the curve characteristic of clustering is different in each case.

Comparison of diagnostic functions

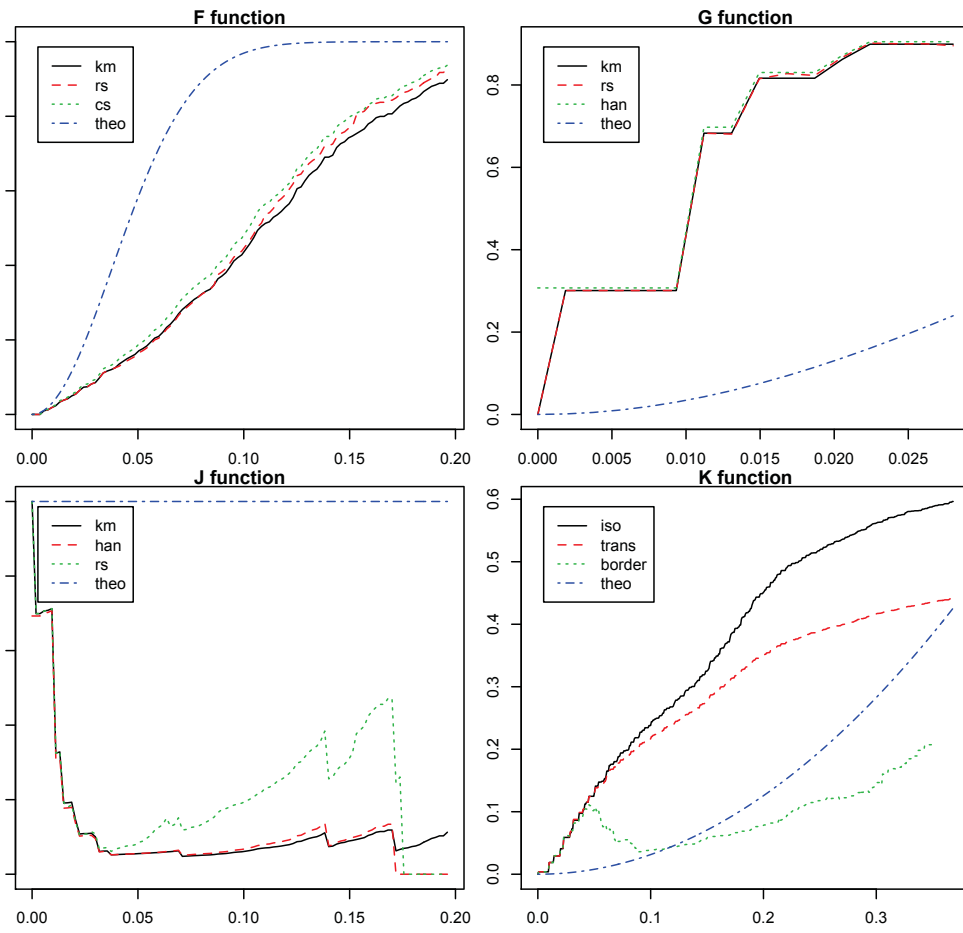


Figure 2.3: Diagnostic functions F, G, J and K for the seal encounter data.

Dealing with Inhomogeneity

The Inhomogeneous K function

Baddeley, Moller and Waagepetersen derived a modified form of *Ripley's K function* for inhomogeneous processes [1]. By reweighting each point in the pattern (x_i) by the inverse of the intensity at that point so that $w_i = \frac{1}{\lambda_i}$ they define the *inhomogeneous K function*. This results in a condition termed *second order reweighted stationary* and permits the use of the modified form of Ripley's K , the Inhomogeneous K function, to be used for analysis of a pattern.

$$K_{Inhom}(r) = \mathbb{E} \left(\frac{1}{\lambda(u)} \sum_{x_j \in \mathbf{X}} \frac{1}{\lambda(x_j)} \mathbf{1}\{0 < \|u - x_j\| \leq r \mid u \in \mathbf{X}\} \right)$$

Conveniently $K_{Inhom}(r)$ reduces to $K(r)$ and is a constant λ if applied to a homogeneous Poisson process. Furthermore, $K_{Inhom}(r)$ for an inhomogeneous Poisson process is πr^2 , which is identical to the $K(r)$ for a homogeneous Poisson process.

$$\hat{K}_{Inhom}(r) = \frac{1}{area(W)} \sum_i \sum_{i \neq j} \frac{\mathbf{1}(\|x_i - x_j\| \leq r)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)} e(x_i, x_j, r)$$

Baddeley [14] recommends the use of a parametric model such as will be seen in Chapter 3 to provide the estimate of $\hat{\lambda}(u)$ required to estimate $K_{Inhom}(r)$. The $e(x_i, x_j, r)$ term is an edge correction if needed. Plotting the $K_{Inhom}(r)$ against the theoretical $K_{Inhom,Pois}(r) = \pi r^2$ shows evidence of clustering. There are "too many" observed points within r at small values of r ; $K_{Inhom}(r) > K_{Inhom,Pois}(r) = \pi r^2$, and "too few" for larger values of r . The impact of the inhomogeneity correction may be seen by comparison of the uncorrected $K(r)$ in Fig. 2.3 above to the Inhomogeneous K function in 2.4 below.

Simulation Envelopes

The observed pattern is a single realization of the underlying process. There may be substantial variation displayed between separate realizations of the same process. Comparing the estimated $\hat{K}(r)$ to the theoretical $K_{Pois}(r)$ and looking

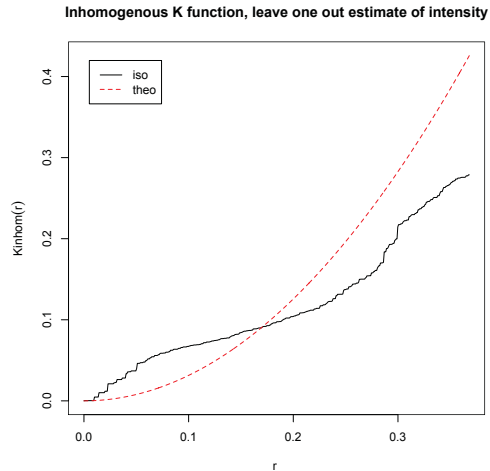


Figure 2.4: **Inhomogeneous K function** Plot of $K_{Inhom}(r)$ for the seal encounter data. The plot is indicative of the presence of clustering in the parent points at distances less than .17 degrees, (approximately 18 km).

for divergence poses the question; how much difference due to divergence from the theoretical process is there in these data's generating process, and how much is due to simple random variation in the data?

Pointwise Monte Carlo Tests

It is possible to address this question of divergence from the theoretical process in the data's generating process by simulating realizations from a given intensity function. The maximum and minimum values of $\hat{K}(r)$ for each value of r form the boundaries of an *envelope*. A $\hat{K}(r)$ lying outside of the envelope for a given intensity is evidence that these data did not arise from a process driven by that intensity. The language of hypothesis testing is used. The null hypothesis is some reference point process model, say CSR. The alternative is not specified beyond "the null is not true". Baddeley gives the alternative as "the pattern is a realization of another process" [14].

In general, if we generate M realizations of CSR and then calculate $\hat{K}^{(j)}(r)$ for $j = 1, 2, \dots, M$, we determine the upper and lower bounds for our envelope.

$$LB(r) = \min_j \hat{K}^{(j)}(r)$$

$$UB(r) = \max_j \hat{K}^{(j)}(r)$$

Interpretation

Under the null hypothesis of CSR the probability that $\hat{K}(r)$ is not in $[LB(r), UB(r)]$ is $\frac{2}{M+1}$ so the test has exact significance $\alpha = \frac{2}{M+1}$. For any previously specified value of r we would reject the null at $\alpha = .05$, if $\hat{K}(r)$ lies outside the boundaries for an envelope using $M=39$ simulations.

Simultaneous Monte Carlo Testing

The weakness of the pointwise method is that the value of r to be tested must be specified in advance. It is not appropriate to plot the envelopes and simply look to see if $\hat{K}(r)$ ever goes outside the bounds. It is possible to construct *simultaneous critical bounds* which do have that interpretation, i.e. such that under the null the probability of $\hat{K}(r)$ leaving the bounds is exactly α . This is done by computing the maximum deviation for the Poisson $K(r)$ for each of the M simulations:

$$D_j = \max_r |\hat{K}(r) - K_{Pois}(r)|$$

The maximum value for D is found and the boundaries of the envelope become,

$$LB(r) = \pi r^2 - D_{max}$$

$$UB(r) = \pi r^2 + D_{max}$$

and $\hat{K}(r)$ leaves this envelope, called a *global* envelope, when D_{data} exceeds D_{max} ; this happens with probability $\frac{1}{M+1}$ under the null.

It is also possible to construct envelopes of this form from null hypotheses other than CSR. For example one may construct an envelope from simulations of a parametric model such as will be constructed in Chapter 3. We utilize this

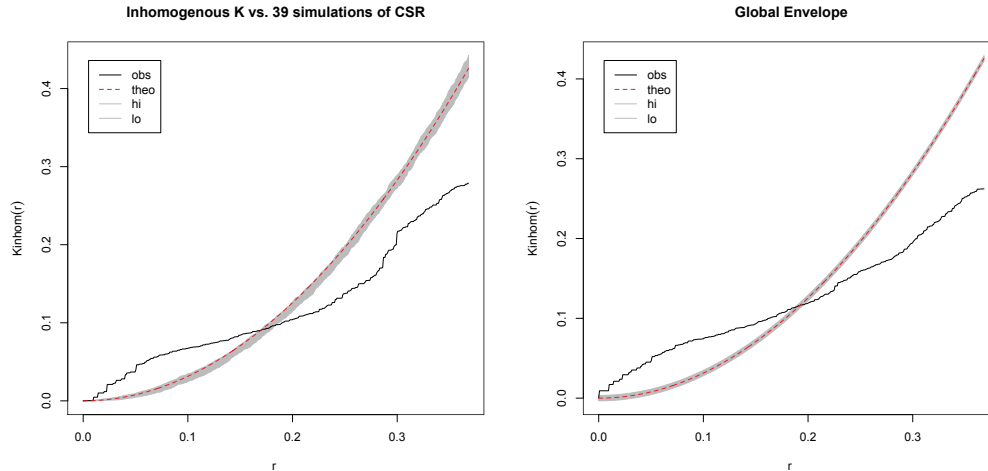


Figure 2.5: **Envelopes for Inhomogeneous K function** Plots of $K_{Inhom}(r)$ versus envelopes for these seal encounter data against the pointwise (left panel), and global (right panel), envelopes generated by 39 and 19 realizations of CSR respectively, $\alpha=.05$ for each.

method to test our models. The null hypothesis is the reference point process model. The alternative is not specified beyond "the pattern is a realization of another process" as above [14].

Fig. 2.5 shows a comparison of the pointwise envelope to the global envelope for these seal encounter data. The pointwise envelope would reject the idea of CSR for any previously chosen value of r except in a small region about $r = .17$. The global envelope can be looked at for any departure of $\hat{K}(r)$ from the boundaries. Since it does the idea of CSR is rejected. Since we have no prior expectations about the seal encounter distances or their distribution, and hence no specific values of r to test, we will use the global envelopes to check the fit of the models that follow in Chapter 3 by simulating envelopes from those models and seeking any departure of $\hat{K}(r)$ from the boundaries.

Chapter 3

Analysis of At Sea Seal Encounters

In this Chapter we describe our at sea seal encounter data and then perform some exploratory analysis, first determining the appropriate window and then considering how we may model these data as a point process within this window. Starting with basic and formal ideas like the homogeneous Poisson model we go on to develop more sophisticated parametric models for the point process. Eventually two stage models called Cox models are investigated. These require us to model the point process in two stages; first modelling the mean for a set of parent points and then modelling the offspring separately. These models are checked using simultaneous critical bounds for Ripley's K function produced by Monte Carlo simulation from the model.

3.1 Data

The data set we examine is comprised of at sea seal to seal encounters recorded as part of the OTN bio-probe project. The first data set was recorded between October and December of 2009. It is the expressed intention of the OTN program to repeat this experiment in subsequent years. The 2009 data was received in the form of 10 Excel spreadsheets, one each for 10 of the 14 tagged seals. Detections of the 4 remaining seals appeared in these 10 files even though these 4 did not have data files of their own. This would indicate that while these 4 seals were detected by the tags of the animals represented by the 10 files, the VMTs on these 4 seals did not themselves record the presence of the animal detecting them. An individual record in the data set consists of the date, time, latitude and longitude of the encounter and the identity(s) of the seal(s) involved. In 2009 there were 1444 encounters initially, ranging from a low of 64 to a high of 346 per seal. Some of these so called encounters were removed on the advice of the tag manufacturer

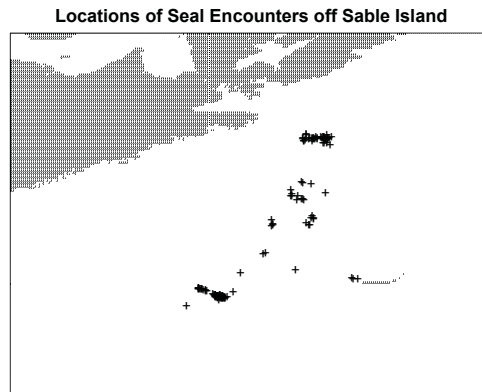


Figure 3.1: Seal encounters off Sable Island

VEMCO (www.vemco.com); namely those in which the transmitting seal ID and the receiving seal ID were one and the same, indicating a spurious self detection. After cleaning in this way there remained 892 total encounters in the 2009 set.

Next, these data were rearranged in a more useful form. Instead of aggregating the encounters by seal ID number, these data were resorted to show the *pair* of seals involved. This produced a set that still contained 892 data points but indexed by 22 pairs of seals rather than the original sorting by the 10 recording tag IDs. The number 22 arises from the 22 distinct pairs of seals that were observed to have encounters. It is notable that 22 is far fewer than the 91 ($14 \text{ choose } 2$) that were theoretically possible. The frequency with which these encounters occurred is quite variable; 4 distinct pairs constitute 63 percent of the total observations. Some pairs are represented by a single observation only.

Figure 3.1 serves to locate these data geographically, placing it in context of the coast of Nova Scotia and Sable Island.

3.2 Exploratory Analysis

Choice of Window

The Window within which a spatial point process is modelled is an integral part of the model. The basic presumption is that the process is ongoing throughout some space and that the window is a subset; implying that the process continues indefinitely beyond the bounds of the window. Summary statistics and descriptions based on distance must therefore take into consideration edge effects, correcting for instances in which the nearest neighbour to a point lies outside the window. The alternative construction is to view the point process as finite, existing only within a defined boundary or domain [11]. The specific instance we are discussing is a finite process since the at sea seal encounters have some geographical limit, even if we don't know what it is. This relieves us of the need to account for edge effects but still requires that we choose a window appropriately. While the window we choose must contain all of the data, too large a window will incorporate too much "white space", and overstate the degree of clustering present. Consequently, we have a statistical problem of estimating the unknown domain from which our data are drawn.

There are several options for the window. For example, one could use the boundaries of the plot of the data in Fig 3.1 as a window. However, there is no data in the majority of the window although it does serve to illustrate the geography nicely. For analytical purposes a much smaller window is desired. The literature suggests the bounding box and the convex hull as likely candidates, and we have adopted the inflated bounding box and inflated convex hull of the method of Ripley and Rasson [25] for estimating an unknown window. Fig. 3.2, shows the bounding box and convex hull estimated by Ripley and Rasson's method and inflated by a factor of $\frac{1}{\sqrt{1-\frac{4}{892}}}$ and $\frac{1}{\sqrt{1-\frac{7}{892}}}$ respectively. We also include a larger bounding box inflated by a factor of 1.3 simply to show the effect of the inflation on the K function; all shown within the same arbitrary rectangle used above to locate the data geographically.

The four corresponding plots of the estimated Ripley's K function follow in Fig.

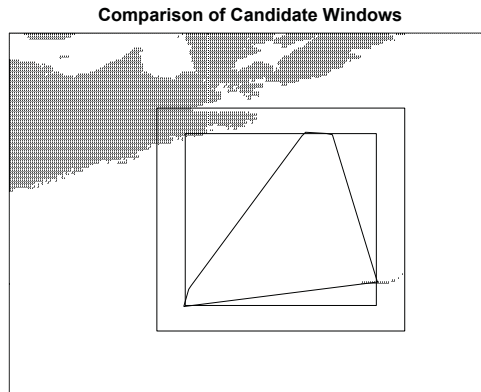


Figure 3.2: **Window Choice.** Illustration of the progressively smaller candidate windows discussed in the text.

3.3 in order to demonstrate the importance of window selection. Note that the larger the window the larger the value of $\hat{K}(r)$ is for any given radius r . All show evidence of clustering since $\hat{K}(r) > \pi r^2$, (the dashed red lines), but the same data in a larger window show stronger clustering since $\hat{K}(r)$ exceeds πr^2 by a larger amount at a like r . This is the effect of the "empty space" in the window. The pattern in the larger windows appears tightly clustered *in relation to the window* than does the identical pattern in a smaller window.

In keeping with the principle that the window should include as little white space as possible [4] [26] the inflated convex hull would seem to present a viable choice. Such irregular polygons however present difficulties in calculating many of the distance methods used in summarizing and analyzing point patterns and methods for using them are not fully developed [4]. Given this we select the bounding box as estimated by the method of Ripley and Rasson. The window is illustrated overlaid over a simple, uninflated, bounding box in Fig. 3.4.

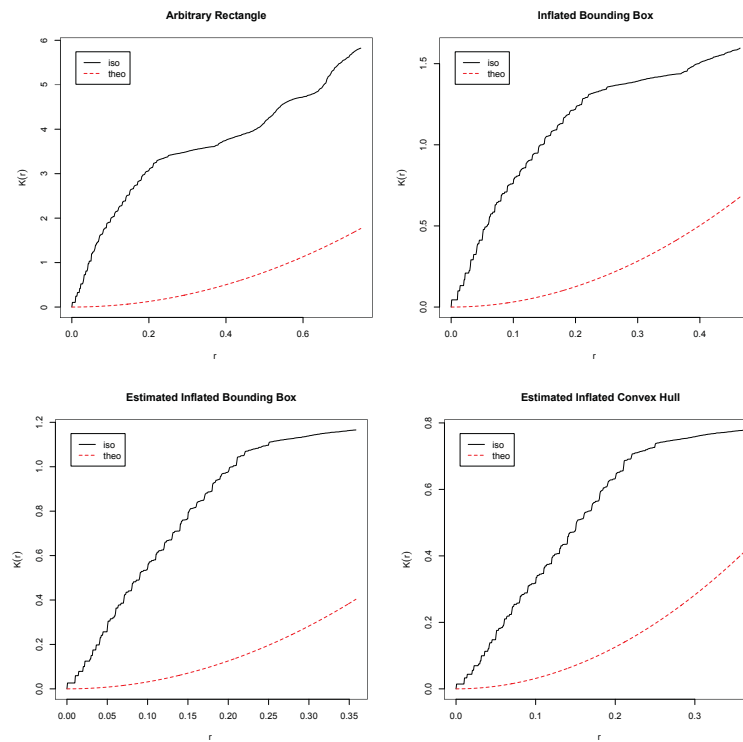


Figure 3.3: **Impact of Window choice.** Ripley's K: Top Left, Large rectangle. Top Right, Bounding Box Inflated by 1.3. Bottom Left, Bounding Box. Bottom Left, Convex Hull. All show clustering. Effect is magnified in larger windows, the K function increases faster in a larger window for any given data set.

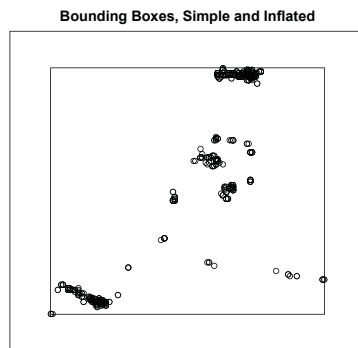


Figure 3.4: **Estimated window** Degree of inflation is small in relation to a simple convex hull. Both are overlaid on the data.

3.3 The Homogeneous Poisson Model

When speaking of a pattern we are speaking of the points themselves as arranged in the space, either the original pattern (the data), or a simulated pattern, called a realization. When speaking of a process we are speaking of the properties that give rise to the pattern, for example a Poisson process. We reserve the term model for the models we are constructing for the mean, or *intensity*, of the process. These models are the parametric or non-parametric estimates of the intensity of the process.

As discussed in Chapter 2, one common initial approach to spatial data is to model them as distributed uniformly and randomly in the space. More formally, if the *intensity*, or *intensity function* of a process at point u is a constant, $\lambda(u) = \lambda$, then the process is a *homogeneous* Poisson process with intensity λ . The special case $\lambda(x) = 1$ is the *standard* or *unit rate* Poisson process. Such Poisson point processes are simultaneously the simplest and most important class of point process models. A Poisson process with uniform intensity, $\lambda(u) = \lambda$, and no interaction (between points) connotes "complete spatial randomness".

Recall, the *intensity* of a point pattern is the expected number of points per

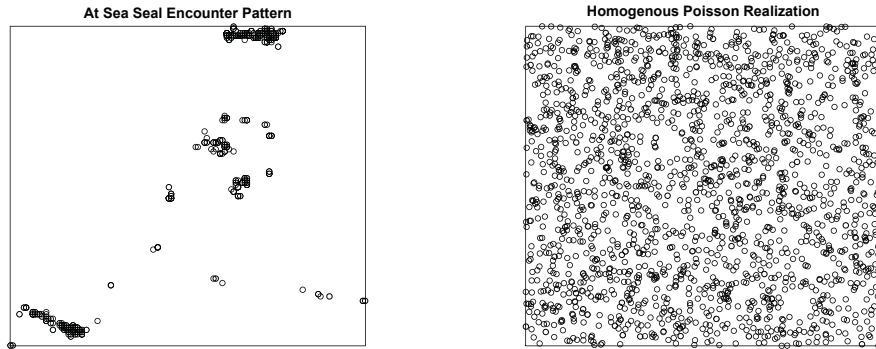


Figure 3.5: **Complete Spatial Randomness** The left panel displays the actual at sea seal encounter pattern to compare with a realization of a homogeneous Poisson pattern with the same number of points. There is no fit on the right.

unit area. This is analogous to the mean or expected value of a random variable. If the point process (X) is homogeneous then the expected number of points (N) is proportional to the area of the window (W). That is,

$$E[N(X \cap W)] = \lambda \text{ area}(W)$$

and,

$$\hat{\lambda} = \frac{n(X)}{\text{area}(W)}$$

is an unbiased estimator for λ .

The homogeneous Poisson model with no interaction between points is not a serious candidate model for our data. Even a cursory glance at the seal encounters reveals a gross departure. However, rejecting the Poisson null hypothesis is often seen as the first step towards fitting a more realistic model, insofar as the study of summary statistics for point processes also relies frequently upon the Poisson process as reference [14] [20]. We do this in the next section.

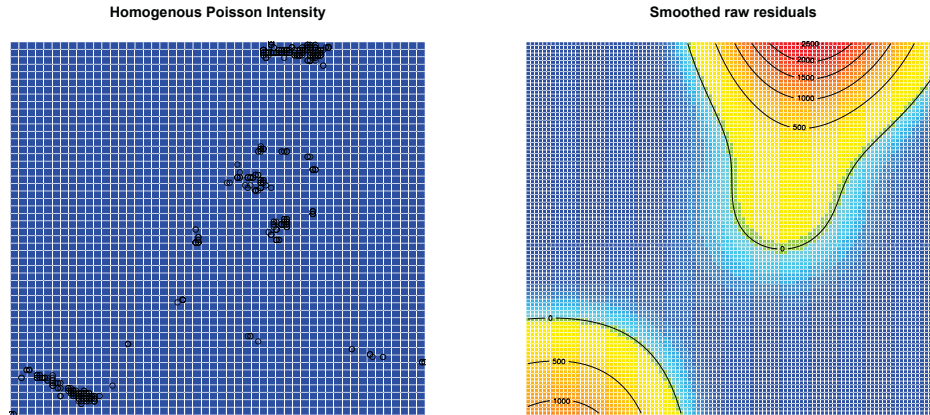


Figure 3.6: **Complete Spatial Randomness** The left panel displays the actual seal A novel approach to studying at-sea associations in marine predators pattern within its window superimposed upon a homogeneous Poisson intensity (the blue); λ is a constant. The right panel shows residuals.

Tests of Complete Spatial Randomness

Quadrat Count Test of CSR

The first question is whether or not our set of encounters could plausibly have been generated by a homogeneous Poisson process. If so, then we have Complete Spatial Randomness (CSR). The hypothesis that these data arose from a *homogeneous* Poisson point process in the plane can be formally tested using a quadrat count test. We begin by dividing the window into equal area subregions (quadrats) and count the number of points lying in each. The homogeneous Poisson process of intensity $\lambda > 0$ has the properties [14]:

1. The number $N(X \cap B)$ of points falling in any region B is a Poisson random variable,
2. The expected number of points falling in B is $E[N(X \cap B)] = \lambda \cdot \text{area}(B)$,
3. If B_1, B_2 are disjoint then $N(X \cap B_1)$ and $N(X \cap B_2)$ are independent random variables.

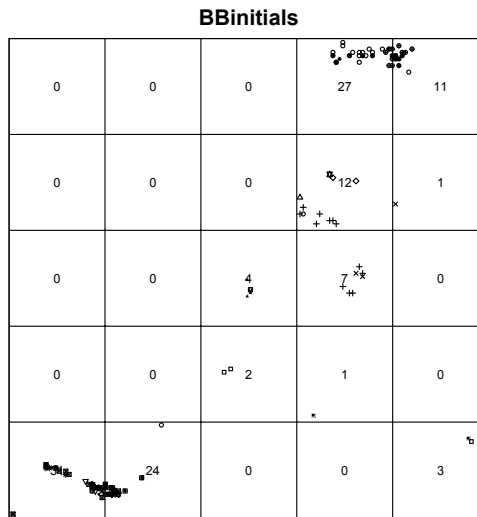


Figure 3.7: **Quadrat Counts** for test of CSR. The test rejects the hypothesis of a homogeneous Poisson process.

If the window W is divided into subregions (quadrats) B_1, \dots, B_m of equal area the counts in each quadrat, $n_j = n(x \cap B_j)$ for $j = 1, \dots, m$, are, under the null hypothesis of CSR, independent, identically distributed Poisson random variables with the same expected value. Therefore the Pearson χ^2 goodness-of-fit test can be used. Performing this test yields a tiny p-value; less than $2.2e-16$, hence we can confidently proceed on the assumption that the points are not the realization of a homogeneous Poisson process.

We can also look at a plot of residuals for validation. The right panel of Fig. 3.6 contains a plot of residuals left by the homogeneous Poisson model in the left panel. These residuals display a pattern, or shape, which is undesirable. While the residual plot does not quantify the disparity like the quadrat count test, it does provide an easy, quick and intuitive validation method.

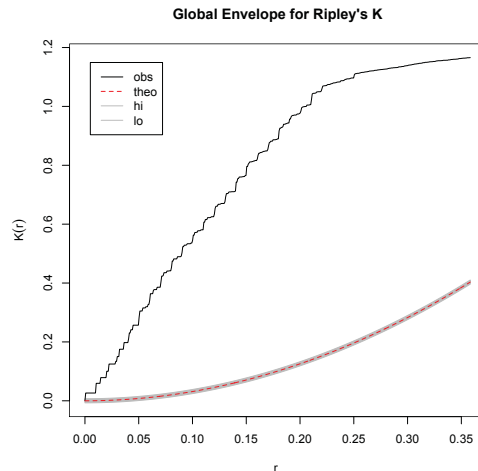


Figure 3.8: **Global Envelope for Ripley's K function** Plot of $\hat{K}(r)$ for the seal encounter data set against the global envelope generated by 19 realizations of CSR. This corresponds to the rejection of a null hypothesis of CSR.

Simulation Testing

Simulation testing (as discussed in §2.5) supports a similar conclusion. Simultaneous Monte Carlo critical bounds, or envelopes, are shown in Fig. 3.8. $\hat{K}(r)$ clearly strays beyond the boundaries of the envelope confirming our rejection of CSR at the $\frac{1}{M+1}$ significance. Here $M = 19$ and so $\alpha = .05$.

This result is hardly surprising; indeed given that we are modelling animal behaviour a result of CSR would have been quite unexpected. The animals are presumably acting in accordance with some imperative of their environment such as food sources and are likely to be found where such food is to be had. Any model we propose may even be seen to represent a proxy for the distribution of food sources for the seals if that is seen as the primary driver behind their movements. Naturally we would like to have a much better model. In the next section we continue our search for a more meaningful model by considering Poisson models with spatially varying (*inhomogeneous*) intensity.

3.4 Inhomogeneous Poisson Models

We are now at the point where we have found substantial evidence to suggest that a spatially varying intensity function, $\lambda(u)$ is required for our seal encounter process (X). That is, we have found substantial evidence to suggest the intensity function $\lambda(u)$ is not a constant and hence our process is not homogeneous but instead is *inhomogeneous*. If we assume that the expected number of points in a small area du around a location u equals $\lambda(u)du$ then $\lambda(u)$ is the *intensity function* of the process (X) [14] [26], and the expected number of points in some subset of the window (V) is:

$$E[N(X \cap V)] = \int_V \lambda(u)du, V \subseteq W$$

Estimating the intensity at location u is required; in practise this is done by using either non-parametric smoothers (e.g. kernel estimation) or by building a parametric model for the intensity.

Kernel Estimation of the Intensity

A kernel estimate of intensity can be used as our estimate of the intensity function for the process. Applying a kernel smoother to the seal encounters yields an estimate of the intensity function as shown in Fig. 3.9. Superimposing data (as in the right panel), shows good correspondence, however, this is expected since all we have done is simply smooth the data.

Simulating the Process using the Kernel Estimate

The realization of an Inhomogeneous Poisson process with the kernel estimated intensity function improves conformity to these data as can be seen in Fig. 3.10. The simulated pattern (right panel), much more closely resembles these data (left panel) than did the homogeneous Poisson simulation seen in Fig. 3.5 (right panel).

We can use the kernel estimate of intensity to simulate patterns generated by the inhomogeneous process model defined by this intensity estimate. Having done this we can also generate envelopes, such as the one seen in the right panel

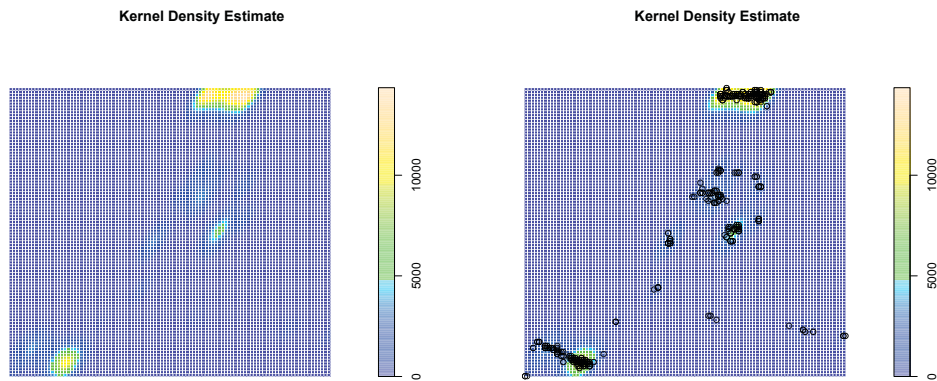


Figure 3.9: **Kernel Estimated Density** The left panel displays the intensity estimate for the complete data set using a kernel smoother. The right panel has the actual locations superimposed.

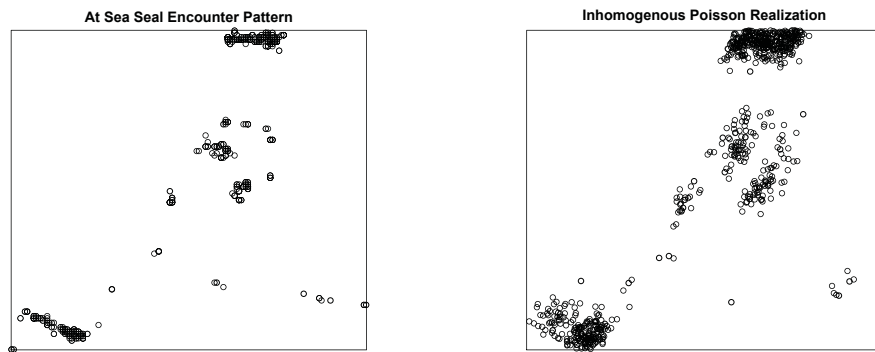


Figure 3.10: **The Inhomogeneous Poisson** The left panel displays the actual seal to seal encounter pattern within its window to compare with a realization of a Inhomogeneous Poisson pattern with the same number of points in the same window and the intensity function defined as the kernel estimated density. There is much better fit.

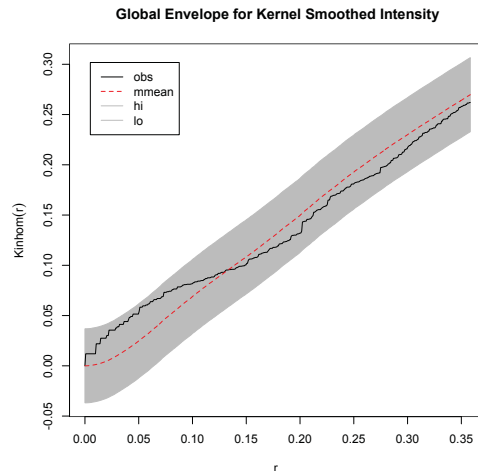


Figure 3.11: **Global Envelope for Inhomogeneous K function** The plot of $K_{Inhom}(r)$ for the seal encounter data against the global envelope generated by 99 realizations of an Inhomogeneous Poisson process with the kernel estimate as the intensity function. There is no reason to reject the hypothesis that the data was generated by a process with the stated intensity. $\alpha = .01$

of Fig. 3.10 and gives a global envelope as seen in Fig. 3.11. Clearly a reasonable fit is achieved as we fail to reject the notion that the process is other than as stated at $\alpha=.01$. However, this isn't really a meaningful model, since the intensity function, $\lambda(u)$ that we used is simply a smooth of these data. In this instance all we have shown is that an inhomogeneous process with an intensity equal to our sample mean fits the data well. In addition to this shortcoming there is no way to link the kernel smoothed estimate to any covariates or to make any explanatory inference from it. More informatively, in the next section we will model the intensity function using an alternative, parametric method.

Model Fitting; Parametric Estimation of the Intensity

The R package **spatstat** [4] includes a model fitting algorithm [5] to fit a *point process model* to point pattern data. It allows the fitting of models which employ spatial covariates so long as the covariate $Z(u)$ is measured at every location u . The algorithm fits an inhomogeneous Poisson model for the intensity function in

which intensity is either a loglinear function of the covariate(s):

$$\hat{\lambda}(u) = \exp(\beta_0 + \beta_1 Z(u))$$

or proportional to the covariates:

$$\hat{\lambda}(u) = \beta Z(u)$$

and the homogeneous Poisson process is a simple special case where,

$$\hat{\lambda}(u) = \hat{\lambda} = \exp(\beta_0).$$

Available Covariates

Modelling spatial intensity implies that the intensity is expected to vary across the space (window), so it would seem reasonable to use spatial covariates that potentially directly influence intensity. Scientists who study seal behaviour have noted that the seals display a preference for the banks off of Sable Island; areas where the seabed rises to form underwater bluffs or banks. Further, they suggest that their foraging behaviour is governed by the need to maximize their (caloric) benefit whilst minimizing their energy expended to achieve the benefit [9]. The bathymetry of the ocean area where the seals interact used was obtained from the US National Oceanic and Atmospheric Administration (NOAA) (www.ngdc.noaa.gov/mgg/global/global.html) [2] with a resolution of 1 minute of longitude/latitude and depth in feet. These data, together with the distance from Sable Island, the seal's base location, to forage will be used to quantify any preference for the banks, thereby quantifying the seal's imperative to minimize energy expended while foraging.

Quadrat (Tessellation) Test of Preference by Depth

A glance at the seal encounter locations superimposed on the bathymetry in Fig. 3.12 seems to confirm the idea that the seals prefer the banks. We can test this formally by defining our quadrats according to depth, creating divisions of our

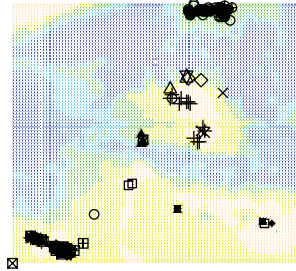
window by bathymetry, that is a tessellation by depth with each tile being a range of depths and each tile being of equal area. Tessellating in this way retains all of the properties enumerated in §3.3. Performing a χ^2 test we obtain a p-value of less than $2.2e-16$ and can conclude that the seals show a strong locational preference due to water depth and water depth is a likely covariate.

Quadrat (Tesselation) Test of Preference by Distance

A glance at the seal encounter locations superimposed on the plot of distance from Sable Island seems to show that there is a preference by distance. Similarly we can test this formally by defining our quadrats according to distance and creating a tessellation of our window by distance, each tile being of equal area. Performing a χ^2 test we again see a p-value of less than $2.2e-16$. We conclude that the seals show a strong preference according to distance from Sable Island and this distance is a likely covariate.

Having shown clearly that the seal encounter locations are not randomly distributed with respect to either of our proposed covariates we proceed to construct parametric models using them as predictors.

Tessellation by Depth



Tessellation by Distance

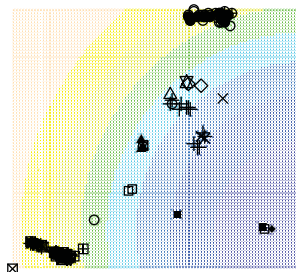


Figure 3.12: Quadrats determined by bathymetry and distance for test of CSR.

Candidate Models

Letting $B(u)$ be the depth at location u , and $D(u)$ be the distance to Sable Island from point u , we will now fit parametric models to these data using these covariates. These models produce an estimate for the inhomogeneous intensity of the Poisson model of the process $\lambda(u)$. As stated earlier, encounters are expected to occur more frequently in shallower areas, (the banks), however the relationship is not expected to be linear. Distance is expected to act as a penalty from the seals perspective [9]. In keeping with the penalty interpretation of depth we may expect distance to appear so as to heavily penalize sites with long travel distances. Given the objective we will favour models that have a meaningful biological interpretation as opposed to simple explanatory efficacy. Many models were fitted, examined and rejected. See Appendix for details. We will discuss the choice of the best model and it's properties.

A summary of the candidate models is given in Table 3.1. All of these models possess some amount of predictive power, with the exception of Model 00, which is the homogeneous Poisson model (shown in it's parametric form for comparison).

Model	Form
Model 00	$\hat{\lambda}(u) = \exp(\hat{\beta}_0)$
Model 01	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 D(u))$
Model 02	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 D(u) + \hat{\beta}_4 D(u)^2)$
Model 03	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2 + \hat{\beta}_6 D(u)^3)$
Model 04	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2 + \hat{\beta}_6 D(u)^3)$
Model 05	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2)$

Table 3.1: **Summary** of candidate models for $\hat{\lambda}(u)$

In all models all parameters are significant at $\alpha=.01$ except the $B(u)^2$ term in model 04 which is significant at $\alpha=.1$. This gives weight to the results of the previous section where we showed that the seal's encounters were not randomly located with respect to either of the covariates, see Fig. 3.12.

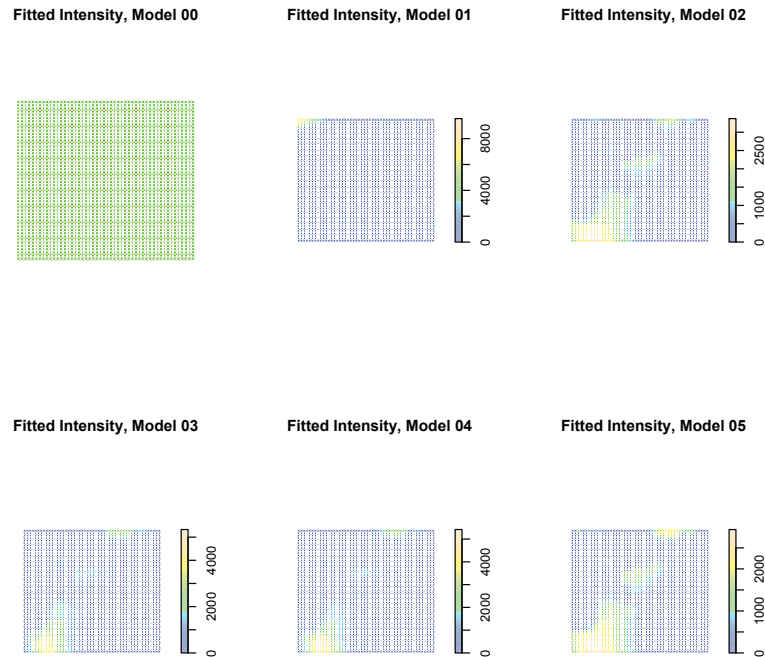


Figure 3.13: **Fitted Intensity Models** The predicted intensity functions for the fitted models.

The fitted intensities of these models are displayed graphically in Fig 3.13. The fitted intensity functions are visually appealing in that they more closely resemble the earlier kernel estimate and allow the spatial variation which we know is required. See Fig. 3.9 for comparison to the kernel smooth of intensity derived from these encounter data. Model selection is examined formally in the next section.

For emphasis we repeat here that these parametric models are attempting to model the entire process as a function of a single, spatially varying, mean; the intensity of the process. This is in contrast *cluster models* which we examine in §3.5.

Model Selection

Model selection based on a likelihood ratio test (LRT) is possible for nested models like we consider here. Performing likelihood ratio tests rejects any of the models, 00, 01, 02, 04 or 05 as subsets of Model 03 with a p-value of essentially 0. Based on the LRT Model 03 is to be preferred.

Alternatively, one may also utilize the *Akaike Information Criterion* (AIC) as a decision criterion. In cases where the models to be compared are not nested Baddeley [14] suggests the use of AIC.

AIC

The Akaike Information Criterion is a means of model selection that imposes a penalty on the number of parameters [16]. If p is the number of parameters then the AIC is defined;

$$AIC = -2\text{maximum log likelihood} + 2p$$

and so small values of AIC indicate better fit. Table 3.1 illustrates that Model 03 best captures the spatial variation. The AIC values for Models 00-05 are shown here.

Model	AIC
Model 00	-8860.133
Model 01	-9866.568
Model 02	-10822.893
Model 03	-11211.269
Model 04	-11144.699
Model 05	-10878.207

Table 3.2: **AIC** Table of AIC values for each evaluated model.

By either of the selection criteria Model 03 has the best predictive properties and is the best choice. All of the coefficients in model 03 are interpretable in terms of the seal's behaviour and, desirably, the coefficients of model 03 are in line

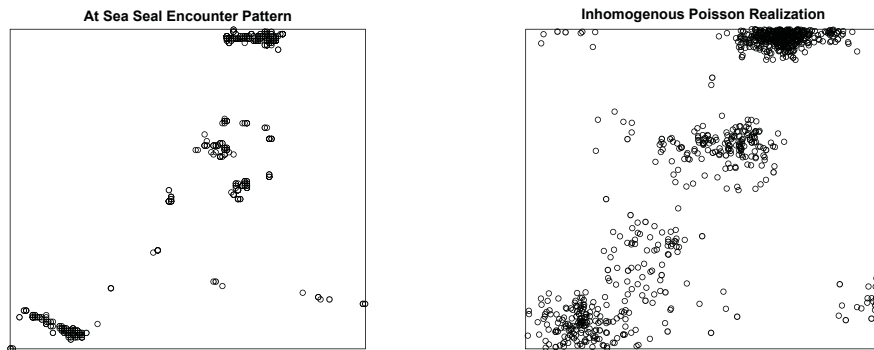


Figure 3.14: **Inhomogeneous Poisson** The left panel displays the actual seal to seal encounter pattern within its window to compare with a realization of a Inhomogeneous Poisson pattern with the same number of points in the same window. The intensity function of the realization is the fitted intensity of model 03.

with our a priori expectations. We have adopted Model 03 as the best choice for modelling the mean of an inhomogeneous Poisson process for the seal encounters. The fitted parameters of Model 03 are shown below.

Model 03

$$\hat{\lambda}(u) = \exp(\beta_0 + \beta_1 B(u) + \beta_2 B(u)^2 + \beta_3 B(u)^3 + \beta_4 D(u) + \beta_5 D(u)^2 + \beta_6 D(u)^3)$$

$$\hat{\lambda}(u) = \exp(8.37 + 0.181B(u) + 0.00314B(u)^2 + 0.0000202B(u)^3$$

$$- 2.61D(u) + 0.794D(u)^2 - 0.0516D(u)^3)$$

Simulating the Process from the Model

Figure 3.14 shows the seal encounters in the left panel and a realization of our Model 03 in the right. The realization of the process, while clearly better than the homogeneous Poisson realization seen in Fig. 3.5 is not as good as that of the

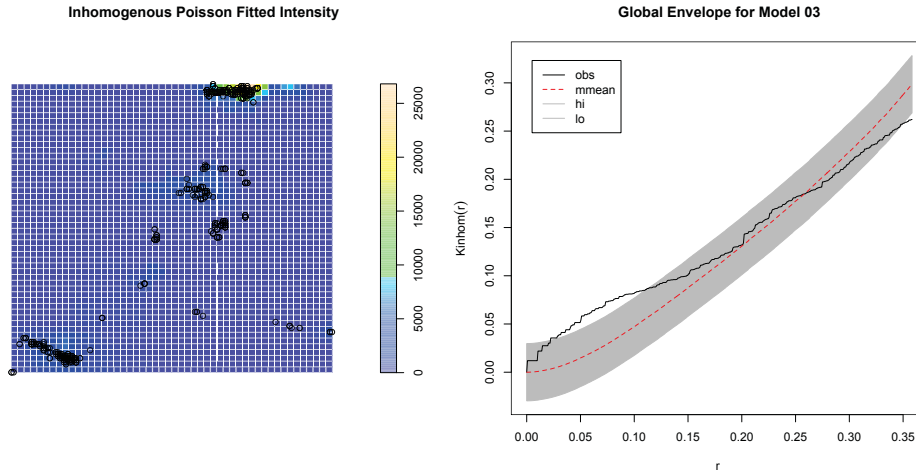


Figure 3.15: **Parametric Inhomogeneous Poisson Intensity Estimate** The left panel displays the actual seal to seal encounter pattern within its window superimposed upon an Inhomogeneous Poisson intensity function; the predicted intensity of the parametric model 03. The right panel illustrates a remaining lack of fit, particularly between $r=.02$ and $r=.10$

inhomogeneous Poisson process defined via the kernel intensity estimate, see Fig. 3.10. Superimposing the point pattern over the fitted intensity as in Fig. 3.15 (left panel) reveals no obvious issues but the Global simulation envelope (right panel) still displays a lack of fit; the $\hat{K}_{inhom}(r)$ function strays out of bounds. It does so between $r=.02$ and $r=.1$, indicating that even after allowing for the effects of the covariates, the seal encounters are still clustered at these distances. Comparing the simulation to the actual points in Fig. 3.14 shows the model to be blurring or scattering the points more than is desirable, the impact of which is detected in the behaviour of Ripley's K function at these distances, as seen in the Global envelope, Fig. 3.15 (right panel). We are forced to reject the null hypothesis that the seal encounters are well modelled by the inhomogeneous Poisson parametric model.

The Inhomogeneous Poisson intensity model fails to equal the performance of that obtained via the kernel estimate. When simulated, a lack of fit is perceived in the range between $r=.02$ and $r=.1$. The data are more clustered, or more tightly clustered, than the model predicts. We now turn to the aptly named

cluster processes in an attempt to further improve our ability to model the seal encounters.

3.5 Cluster Processes

The foregoing models all attempt to model the process using a single mean function. That is, each point in the realization is considered as an event of a Poisson process with a single, spatial mean, $\lambda(u)$. The Ripley's K function compared to the best performing parametric model, Model 03, clearly shows (see Fig 3.15) evidence of clustering in these data not accounted for by the model. The Cox processes are a convenient way to model clustered data [20]. Accordingly we examine them further here.

Effectively, a Cox process is a Poisson process with a random intensity function. The Thomas and Matérn processes are both Cox processes. Desirably, both the Thomas and Matérn processes have K functions of known analytic form. This makes possible the fitting of models for these processes using the method of *Minimum Contrast* [12].

The K function for a Thomas process is

$$K_{\theta}(r) = \pi r^2 + \frac{1}{\kappa} \left(1 - \exp\left(-\frac{r^2}{4\sigma^2}\right)\right),$$

and that of the Matérn Cluster process is

$$K_{\theta}(r) = \pi r^2 + \frac{h\left(\frac{r}{2R}\right)}{\kappa}$$

where $h(z) = 2 + \left(\frac{1}{\pi}\right) \left((8z^2 - 4) \arccos(z) - 2 \arcsin(z) + 4z \sqrt{((1 - z^2)^3)} - 6z \sqrt{(1 - z^2)} \right)$ for $z \leq 1$, and $h(z) = 1$ for $z > 1$.

The theoretical intensity of the Cluster process $\hat{\lambda}_{\theta} = \kappa * \mu$, is a function of the location u , where $\theta = (\kappa, \mu, \sigma)$ is the parameter for the Thomas process and $\theta = (\kappa, \mu, R)$ that of the Matérn process. In both cases κ is an intensity function for the *parent points*, and μ is the expected number of *offspring* points per cluster. The Thomas process has parameter σ , the standard deviation of a bivariate normal distribution; this distribution describes the offspring about each parent point. The

Matérn process has parameter R , the radius of a disk, centered on each parent, in which the offspring are uniformly distributed. By estimating these parameters we can fit these Cox processes to these data.

The values of the parameter θ that best match $K_\theta(r)$, the theoretical K function under the model, to $\hat{K}_{Inhom}(r)$, the K function estimated from these data, are desired. In this fashion the properties of the model best describe the clustering characteristics of these data. This is achieved by the method of *minimum contrast* [12] which minimizes the difference between the 2 functions over some range:

$$D(\theta) = \int_a^b \|\hat{K}_{inhom}(r)^q - K_\theta(r)^q\|^p dr$$

where $0 \leq a \leq b$, and $p, q > 0$ are indices. It is usual to choose $p = \frac{1}{4}$ and $q = 2$ so that the contrast criterion $D(\theta)$ is the integrated squared difference between the fourth roots of the two functions [30].

These models can be fitted as either homogeneous or inhomogeneous. Given the results of the analysis thus far inhomogeneous is the obvious choice. There are two ways to incorporate the inhomogeneity into the model, either, generate the parent points homogeneously with a constant κ and a spatially varying μ , that is $\mu(u)$, or, generate the parents inhomogeneously using a spatially varying κ , $\kappa(u)$, and a constant μ . Here we fit the model with spatially varying offspring:

$$\hat{\lambda}(u) = \kappa\mu(u).$$

When fitting a cluster model by minimum contrast one can still specify an intensity function for the mean that is a function of the covariate(s). The same, best performing, mean function as was found earlier (with the inhomogeneous process) parametric model ($depth + depth^2 + depth^3 + dist + dist^2 + dist^3$) is used because $\hat{\lambda}(u)$, the overall intensity at point u , is the best estimate available based on the covariates. What we are doing here is allowing $\hat{\lambda}(u)$ to be decomposed into κ , the intensity of the parents, and $\mu(u)$, an inhomogeneous intensity for the clusters. In this way a model of the form, $\hat{\lambda}(u) = \kappa\mu(u)$ is used and the parameters κ and σ or R are estimated for the Thomas and Matérn processes respectively. Thus the

model has an inhomogeneous overall intensity, $\hat{\lambda}(u)$ and inhomogeneous clusters with the expected number of offspring $\mu(u)$. One should interpret this model as the seals spending *more or less time* in proximity to one another depending on *where* the encounter occurs.

For the Thomas cluster process we obtain $\kappa=1.458$ and $\sigma=0.0004$ and for the Matérn cluster process the respective fitted values of the parameters are $\kappa=1.458$ and $R=0.00068$.

For the Thomas and the Matérn Inhomogeneous cluster point process models give similar results. The κ parameters are very similar. Simulating from our Thomas and Matérn cluster models we can see, as in Figs. 3.16 and 3.17 respectively, that the simulated processes from the two models are very similar to one another but are not terribly good in comparison to the actual data.

Calculating simultaneous envelopes for the two models shows the lack of fit, the K function leaves the boundaries in both cases. Fig. 3.18 also reveals the nature of the lack of fit, the K function is well above the means for the models (the red lines) at $r < .15$ and falls significantly below at $r > .25$. In other words, there are not enough points at short separations in either model and there are too many at longer separations in both. Looking to the simulations in Figs. 3.16 and 3.17 we see that both appear overly dispersed, or spread out, in comparison to the actual data, consistent with the interpretation of the envelopes.

It is important to note that when fit in this way it is the expected *cluster size* that is varying over the window to create the inhomogeneity. Contrast this to the results of the following section.

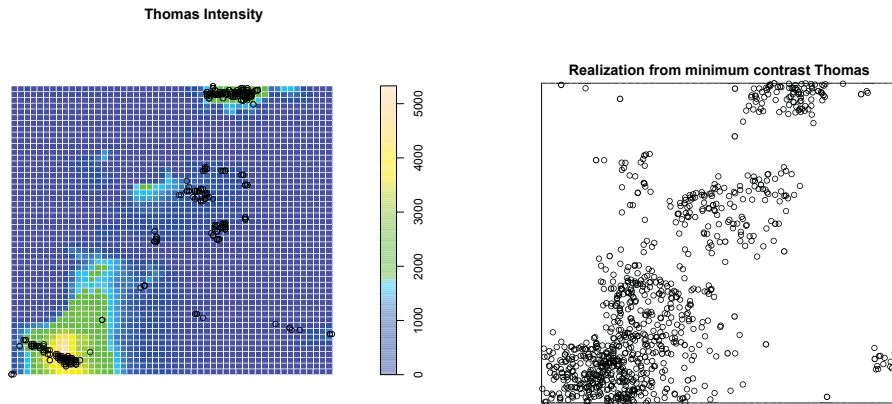


Figure 3.16: **Thomas Inhomogeneous Intensity Estimate** The left panel displays the actual seal to seal encounter pattern within its window superimposed upon the Inhomogeneous mean intensity function for the Thomas process; that is, the intensity of the **parent** points. The right panel shows a simulation from the model. The fit is poor.

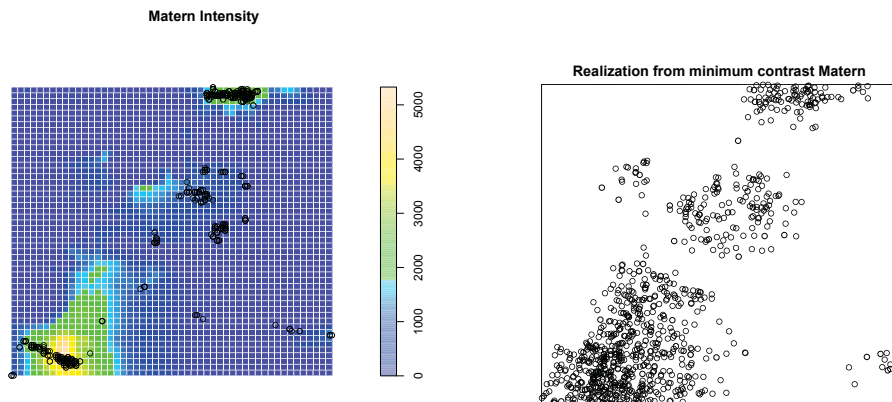


Figure 3.17: **Matérn Inhomogeneous Intensity Estimate** The left panel displays the actual seal to seal encounter pattern within its window superimposed upon the Inhomogeneous mean intensity function for the Matérn process; that is, the intensity of the **parent** points. The right panel shows a simulation from the model. The fit is poor.

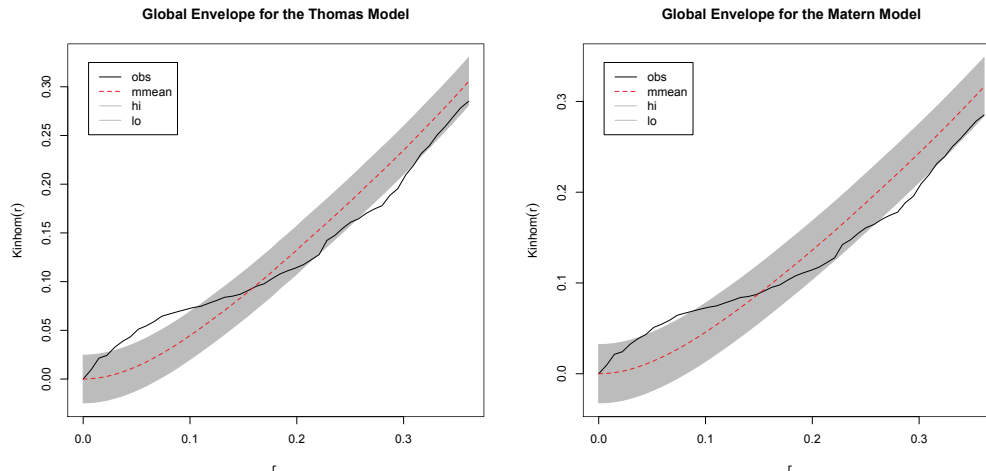


Figure 3.18: **Simultaneous Confidence Envelopes for the Thoams and Matérn Inhomogeneous cluster models** The left panel displays the global envelope for the Thomas model. The right panel that of the Matérn model fit by minimum contrast. Both display a lack of fit in that they predict too few points at short separations and too many at long separations.

3.6 Including Mark Information

Temporal Clusters

There are properties of the seal encounter data that may allow us further refinement. Specifically, there are groupings in the seal encounters related to the VMT technology and how it works. Poisson cluster processes are two stage process with *parents*, a spatial point process for the cluster centers, and *offspring*, a spatial point process for the points associated with each cluster center. The complete point pattern that is observed is the *superposition* of the two patterns. When the cluster centers are points of a Poisson process it is termed a *Poisson cluster process*. Normally it is not possible to distinguish between a Cox process (below) and an inhomogeneous Poisson process with only a single realization. Neither is it normally possible to determine which points in the superposition constitute parents and which are offspring. Making this distinction requires some special feature of the point(s), inherent in the marks. The seal encounter data include additional

information, specifically the time at which the encounter occurred. While, up to now, we have not examined the temporal component we can make use of the time to refine our ability to model these data.

A VMT has a signal pulse rate of approximately two minutes. That is, it sends out a signal, a pulse, every two minutes on average. This creates an a priori expectation that the observations will be correlated in time. Since, once in proximity a pair of seals are likely to be re-observed at the data sample rate of the tag there will likely be subsequent encounters over the ensuing time at the 2 minute refresh rate of the tags. Eventually the seals separate and the repeated series of contacts ends. This short series of contacts can be viewed together as a cluster event. The same pair of seals may have a subsequent encounter, so the question arises, for what interval between sequential contacts will we judge a contact to be of the same encounter cluster, and alternatively, when will we adjudge the contact to be the initial contact of a new cluster.

In order to explore this feature of these data further we difference the observation times. That is, we calculate the set of times between each observation of a given pair of seals and their next observation. A quick glance at the histogram of these differences (in Fig. 3.19) reveals a preponderance of short separations and a small number of much larger lags with separations of many hours or days. It seems a reasonable assumption to treat the very long separations as new encounters rather than assuming the two seals have been in contact for that long without a detection. Accordingly these are designated as initial encounters of new clusters and the summary statistics are recalculated. The mean time of proximity, that is the time from one pairwise encounter to the next encounter between the same pair, *excluding these long lags*, was 8:28 (eight minutes-twenty eight seconds) and the standard deviation was 22:44. Approximately 77 percent of the lags are less than 10 minutes, 95 percent of all lags are less than 46 minutes, 98 percent are less than 60 minutes. This is entirely in keeping with our a priori expectation. Setting the threshold time higher results in more clusters; fewer clusters simplifies analysis at the cost of information, so the threshold is kept high. Using a cutoff of 1 hour, results in 122 initial encounter points for their respective clusters.

Here we make use of the fact that the cluster generating process is known; we know that the seals encounter one another and that the refresh rate of the tags will continue to report the same two seals for so long as they remain in proximity. It seems natural to view the initial encounters, i.e. the first observation in each cluster, as the parent points of our cluster process. The subsequent re-observations of the same pair of seals will then be offspring.

We now construct Poisson Cluster Models using the 122 initial points as our parents. This is again a two step process: 1) create a parametric model for the mean using the 122 parent points and our covariates, and 2) model the offspring by estimating their distribution according to the Thomas and Matérn processes. Unlike the previous section we are specifying an inhomogeneous intensity for the designated parents κ , meaning that the spatial inhomogeneity is modelled by $\kappa(u)$ and therefore there is a constant expected number of offspring μ . The inhomogeneity in the model is therefore a function of the inhomogeneous intensity of the parent points, the offspring's distribution does not vary spatially. Such models have the interpretation that the seals encounter one another *more or less frequently* depending on location, but then, having encountered one another, the time they spend in proximity is independent of location. That is:

$$\hat{\lambda}(u) = \kappa(u)\mu.$$

Plotting only the set of initial encounters yields a display that is similar to the full data set in appearance, see Fig. 3.21. This is not surprising since the points not shown are those that are strongly dependent in time on the initial encounters and are therefore proximate spatially. The density estimate derived from only the initial encounter is not so strongly peaked. Again, this is not surprising, since removing multiple proximately located data reduces the weight attached to those locations, spreading out the density estimate.

Intensity Estimation for the Initial Encounters

Having identified the parent points we now need to estimate their mean $\kappa(u)$. This parallels estimation of the mean for the inhomogeneous model earlier with

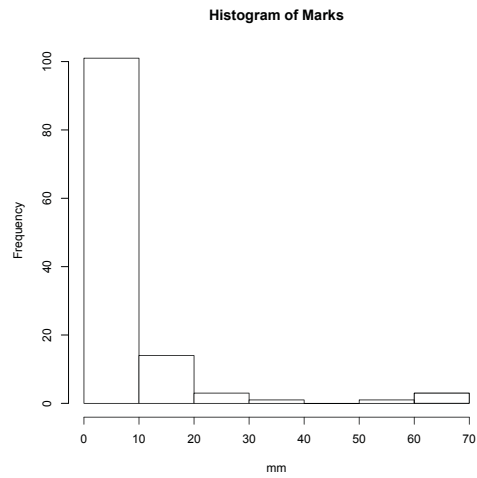


Figure 3.19: **Histogram of Time Separation between Points** Difference between observation times (in minutes). The 2 minute refresh rate of the VMT gives several data for each encounter.

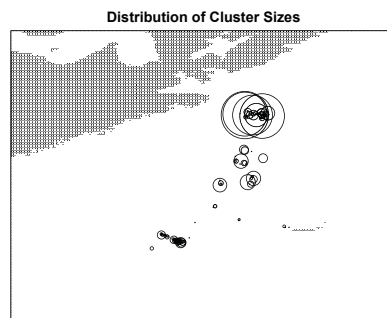


Figure 3.20: **Temporal Clusters.** The circles are proportional to the number of re-detections associated with each Distinct Encounter

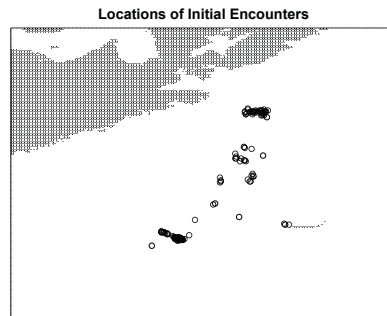


Figure 3.21: **Initial Encounters.** Locations of the 122 parent points making up the reduced data set.

the distinction that only the 122 parent points are used. This can be done by methods of kernel smoothing or by use of a parametric model. The limitations of the kernel smoother are as they were when we were modelling the entire process. We therefore proceed directly to parametric estimation.

Parametric Estimate of the Mean for the Parent Points

In order to estimate the intensity of the 122 *parent* points only, we fit parametric models using the same covariates as we did earlier.

Model	Form
Model 00	$\hat{\kappa}(u) = \exp(\hat{\beta}_0)$
Model 11	$\hat{\kappa}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 D(u))$
Model 12	$\hat{\kappa}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 D(u) + \hat{\beta}_4 D(u)^2)$
Model 13	$\hat{\kappa}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2 + \hat{\beta}_6 D(u)^3)$
Model 14	$\hat{\kappa}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2 + \hat{\beta}_6 D(u)^3)$
Model 15	$\hat{\kappa}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2)$

Table 3.3: **Summary** of candidate models for $\hat{\kappa}(u)$

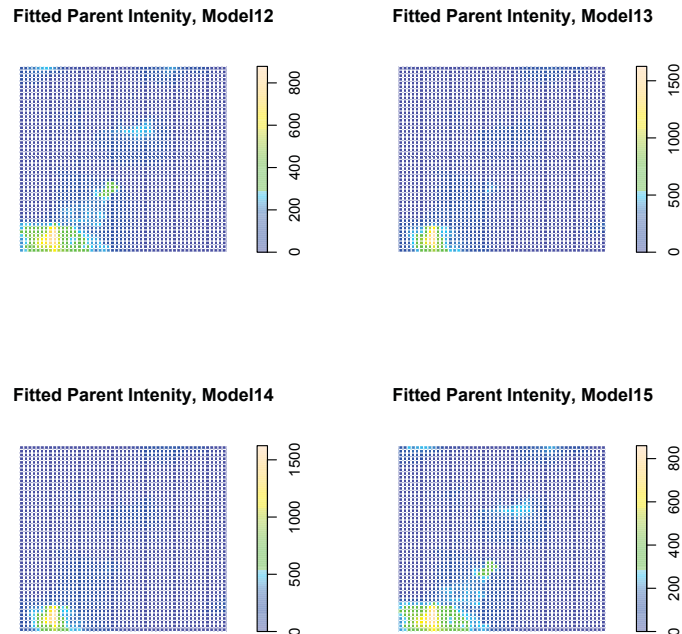


Figure 3.22: **Fitted Intensity Model** The fitted models, Model 12, Top left. Model 13, Top right. Model 14, Bottom left. Model 15, Bottom right.

The fitted intensities of these models are displayed graphically in Fig 3.22.

Model Selection

As previously, model selection based on a likelihood ratio test is possible if the models are nested, i.e., if one model is a subset of the other. In cases where the models to be compared are not nested then the *Akaike Information Criterion (AIC)*, is used [14] [16]. The AIC for the parametric models for parent point are seen in Table 3.3. Model 13 is the best model for the 122 parent points.

Views of the fitted intensity of Model 13 and and the superimposed (parent) pattern are shown in Fig. 3.23.

The parametric model for the mean of the parent points is κ in the parameter $\theta = (\kappa, \mu, \sigma)$ To build a Thomas or Matérn model using this mean function $\kappa(u)$

Model	AIC
Model 11	-914.9445
Model 12	-1049.4194
Model 13	-1113.9677
Model 14	-1113.5837
Model 15	-1047.5419

Table 3.4: **AIC** Table of AIC values for parent point models.

we need to estimate the other parameters in θ . μ is straightforward since the expected number of offspring per cluster, μ , is,

$$\mu = \frac{892 - 122}{122} = 6.31$$

as there are 892 total points and 122 identified as parents, leaving 892-122 offspring to be divided into 122 clusters.

For the Thomas process an estimate of σ^2 is obtained by minimum contrast as above in §3.5 but now specifying $\kappa(u)$ as the predicted fit of Model 13, our parametric model for the parents. This produces an estimate for σ^2 of 0.00365. For the Matérn process an estimate of R is obtained by minimum contrast in the same way. This produces an estimate for R of 0.06985. Having parameterized our models we can now simulate from them.

Simulating the respective Thomas and Matérn cluster processes with these new parameter estimates yields realizations like those in Fig. 3.24. These realizations much better duplicate the original pattern. Calculating the envelopes results in Fig. 3.25. The Thomas model is rejected since the K function leaves the bounds. The Matérn model on the other hand stays in. We cannot reject the Matérn model.

3.7 Interpretation

Here we examine the eventual Matérn model with respect to our biological conjectures. Taking first depth, what is the impact of bathymetry on the seal's

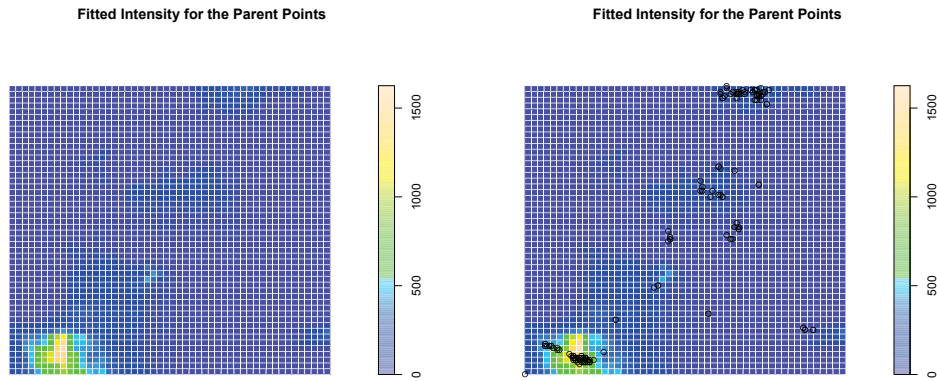


Figure 3.23: **Fitted Intensity of Parent process** Plot of our parametric inhomogeneous Poisson model for the intensity of the parent points. The right panel has the 122 points identified as parents superimposed.

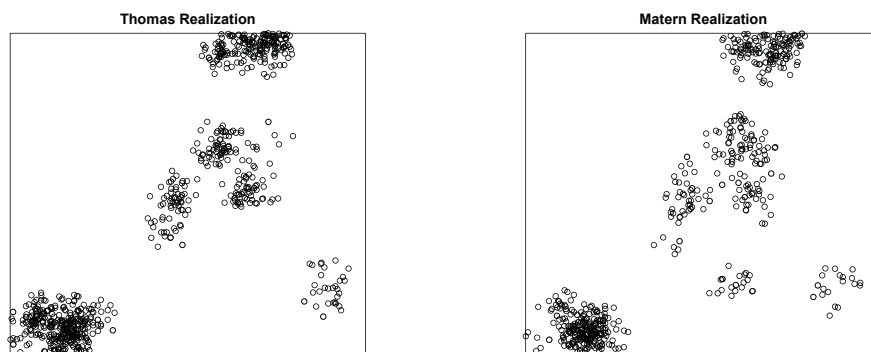


Figure 3.24: **Simulated Patterns from our Thomas and Matérn models** The Thomas and Matérn models created using the new parameters.

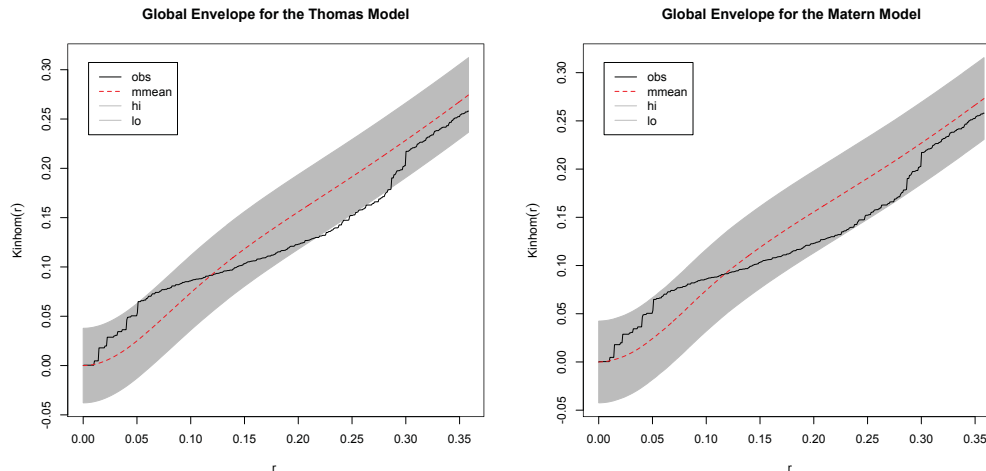


Figure 3.25: **Simulated Patterns from our Thomas and Matérn models** The Thomas and Matérn models created using the new parameters.

encounters? Since our inhomogeneous Poisson model and our fitted inhomogeneous Matérn have the same overall intensity $\hat{\lambda}(u)$ the argument here applies to both. The left panel of Fig. 3.26 shows the shape of the effect of depth on $\hat{\lambda}$ (the effect function). Shallow water has a very large, positive impact on the intensity $\hat{\lambda}$, consistent with the idea that the seals prefer the shallow water areas (the banks). What appears inconsistent is the monotonicity of the effect function. The shallowest water available to the seals is the immediate vicinity of Sable, where encounters are scarce. The banks are shallow *relative* to the surrounding seafloor, not absolutely. We might have expected a maximum, or at least a local maximum, to have occurred in the effect function at a depth corresponding to that found on the banks. Fig. 3.28 is a *lurking variable plot*, used to determine at which values of a covariate the model fails to capture the impact of the covariate. The lurking variable plot reveals some significant dynamic at depths of 40-80 feet. The model under predicts from approximately 40 to 50 feet, then dramatically over predicts around 60 feet, returns to under predicting from 70 to 80 feet and reaches equilibrium by 100 feet. There are very few encounters in water deeper than 100 feet. This pattern is indicative of a bi-modal distribution of water depth at point of

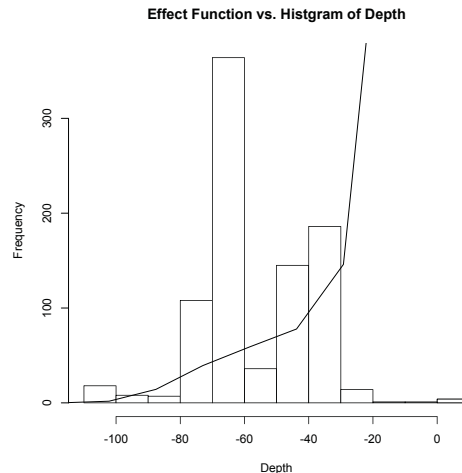


Figure 3.26: **Effect Function and Histogram of depth** Left Panel. Effect function. Right Panel. Histogram of water depth at point of encounter, overlaid is the effect function.

encounter. The histogram in Fig. 3.27 (right panel) confirms the bi-modality and confirms the information in the lurking variable plot; under prediction, then over prediction then return to under prediction.

Reexamination of Fig. 3.1 reveals that there are two small areas of very high intensity in the seal encounter pattern, in the north-center of the pattern and in the south-west corner. These are the locations of the banks discussed previously. Not coincidentally the depths of these banks are approximately 70 feet and approximately 40 feet. Plotting the pattern together with the bathymetry as in Fig 3.27 (right panel) elucidates the point more fully. A higher order polynomial may have been able to capture the effect of depth (the bimodality) better by but orders higher than 3 are not supported by **spatstat**, the R package available with which to fit point process models [14]. Higher order polynomials are promised for the forthcoming Version 2.0 [14], and investigation is warranted in future. Alternatively an additive modelling approach would be very worthwhile and potentially an avenue of future research.

Comparison of the effect function for distance to the histogram of distances from Sable Island to point of encounter shows no such discrepancy, indeed the effect

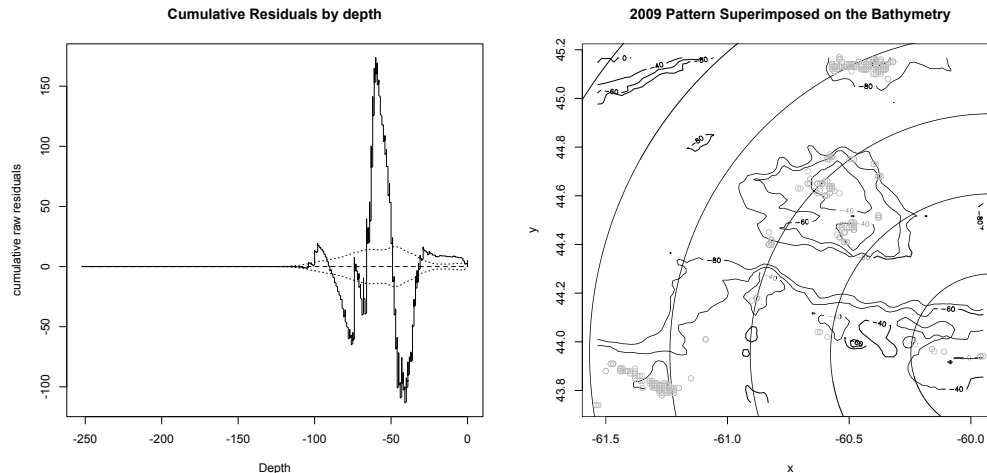


Figure 3.27: **Lurking Variable plot for depth** Lurking variable plot for depth, shows cumulative residuals as a function of the covariate. Compare to locations versus the 40, 60 and 80 foot contours.

of distance seems to be well captured. The unimodal histogram is consistent with the interpretation that the seals use the *closest* area of suitable depth, travelling to it directly as evinced by the few encounters near Sable Island, and no farther; note the absence of encounters at distance greater than required.

While not statistically significant it does appear that there is some shape remaining, the shape of $\hat{K}_{Inhom}(u)$ does not follow the mean trend of the model perfectly, the black and red lines in Fig 3.25 diverge, hinting that the seal encounters may still be exhibiting clustering after conditioning on the covariates. The model still seems to under predict slightly at short separations, (small r), and over predict slightly at long separations, (large r), even after conditioning on the model covariates. The ocean seafloor topography is, at best, a proxy for whatever is actually driving the seals behaviour, presumably feeding. So the interpretation of the foregoing must be that the spatial distribution of the food source is a function of and clustered with reference to the depth of water. The clustering is at the scale of approximately .067 degrees of longitude, or approximately 7 km. The covariates are an issue, the resolution of the bathymetry is one sounding every minute of arc, or approximately every 2 km. Given this fact the ability of the model to resolve

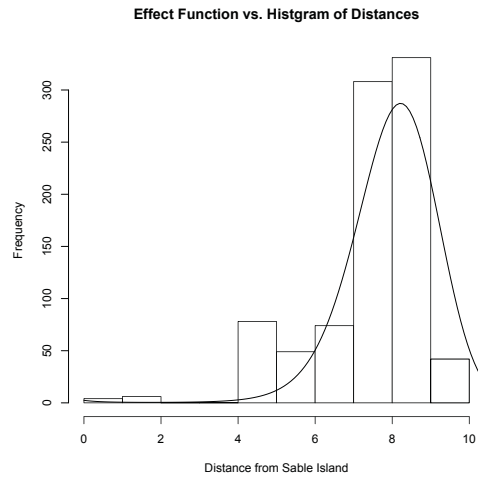


Figure 3.28: **Histogram of distance** Histogram of distance from point of encounter to Sable Island. The effect function for distance (smooth line) is overlaid. Correspondence is good. Scale is 10 minutes of arc.

detail below that scale is unreliable. The Matérn parameter R , the radius of the disk about each parent on which the offspring are distributed, which we estimated to be 0.069, is further evidence of clustering in the approximately 7km scale.

3.8 Validation

In an effort to validate the preceding analysis we now turn to the 2010 data. Fig. 3.26 locates the 2010 encounter locations in relation to the coast of Nova Scotia and Sable Island. The 2010 data set consists of 1115 seal to seal encounters recorded in a repeat of the experiment that generated the 2009 encounter data with some emendations. The sample size was increased, instead of 14 seals as in 2009 there were 20 seals in the 2010 data set. Notwithstanding the aforementioned changes Fig. 3.26 places the encounters in the same general area as the 2009 data.

We consider first the need to estimate the window (W) for a point process if unknown, as is the case here. Plotting the 2010 encounters and the window (W), estimated from the 2009 data (using the method of Ripley and Rassin as in Fig. 3.27 (left panel)) reveals a problem. There are 40 encounters from the 2010 encounter set lying *outside* the estimated window. Clearly, the estimate of the window derived from the 2009 data is too small to contain the entire process, even though it was entirely suitable for the 2009 observations.

Using the 2010 observations to validate our models based on those of 2009 implies we regard the 2010 data as a second realization of the same process. In order to consider the 2010 data as a second realization we must compare them on a like basis. Here are three options;

- (W_{2010})** Consider only the subset of 2010 encounters lying *within* the 2009 window. Doing this does require some data loss, (40 observations), but has the advantage of validating models based *only* on the 2009 data.

- (W_{2010})** Calculate a new window calculated using the 2010 data only.

- ($W_{2009-2010}$)** Calculate a new window using all of the available 2009 and 2010 data.

With the caveat that ($W_{2009-2010}$) is now our current best estimate of the true domain for the seal encounters we choose to validate our 2009 based model using

the first option, (W), therefore including only the subset of 2010 encounters lying within our originally calculated window. We do this because the window (W) of a point process model is *part* of the model, and this leaves us to believe that this uses these data to best mimic a second realization of our model. Expanding the window on the basis of the additional information received in the 2010 data is best viewed as a refinement of the 2009 model rather than a validation exercise.

Fig. 3.27 (right panel) presents the 2009 data and the subset of the 2010 data that lies within the original window (W).

We calculate simultaneous Monte Carlo envelopes from this model and examine the Inhomogeneous K function for the subset of the 2010 data. This allows the assessment of the degree of clustering present in the 2010 data in reference to the 2009 data and in reference to the model. Fig. 3.28 shows the results. The envelope is the global simulation envelope derived from our inhomogenous Matérn cluster Model; κ is the parametric model of our 122 designated parent points, $\mu=6.31$ and $R=0.06985$. The dashed red line is the mean of the model prediction. The black line is $\hat{K}_{Inhom2009}(r)$ from the 2009 data and the solid red line is $\hat{K}_{Inhom2010}(r)$ for the 2010 data.

$\hat{K}_{Inhom2009}(r)$ lies entirely within the envelope as it did before, $\hat{K}_{Inhom2010}(r)$ does not. The nature of the divergence shows that the 2010 data are more highly clustered than the 2009 data. $\hat{K}_{Inhom2010}(r) > \hat{K}_{Inhom2009}(r)$ for values of $r < .25$ and $\hat{K}_{Inhom2010}(r) < \hat{K}_{Inhom2009}(r)$ for values of $r > .25$. In comparison to the 2009 based Matérn model the model significantly under predicts the 2010 clustering between approximately $r = .017$, (the resolution limit) and $r = .07$. The 2010 data are more clustered than the 2009 data, and significantly so.

We are left with the caveat of whether and perhaps most importantly how we can consider this 2010 data a second realization of the same process. One must ask *why* and in *what manner* is the 2010 pattern different. One possibility is that the relationship between the seal's behaviour and the covariates has altered. Perhaps depth and distance did not influence seal behaviour in 2010, or influenced it differently.

To gain some insight we essentially redid our analysis of the 2009 data but

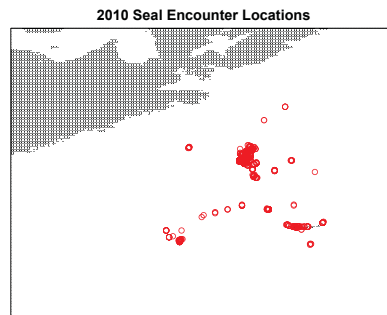


Figure 3.29: Overall view of the 2010 locations of Seal to Seal encounters off Sable Island

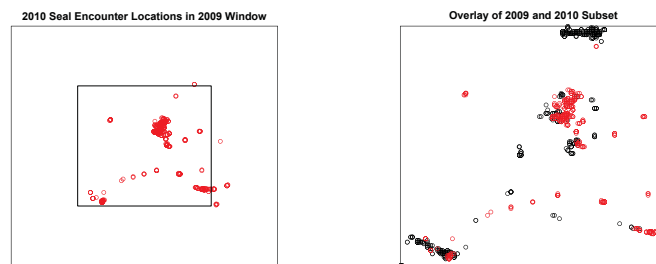


Figure 3.30: The 2010 data and the 2009 Window Left Panel. The 2010 data require a larger window. Right Panel. The 2009 data and 2010 subset in the 2009 Window. 2009 locations in black.

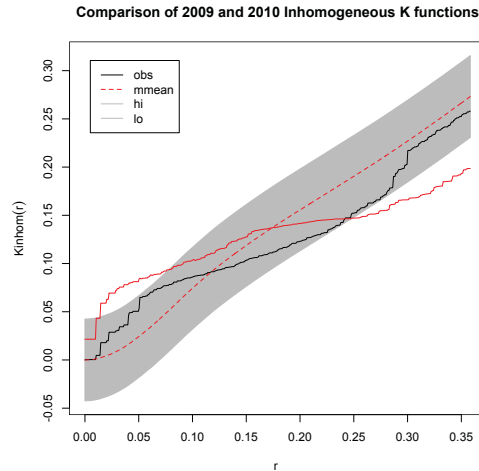


Figure 3.31: **Simulated Patterns from our Matérn model** The Matérn model is not plausible for the 2010 data.

this time using only the 2010 data, and hence (W_{2010}) . The model for $\lambda_{2010}(u)$ is very similar to the previous parametric model for $\lambda_{2009}(u)$. It contains each of the covariates, depth and distance, at the third order. All of the coefficients are significant at $\alpha = .01$ except the depth^3 term, which is significant at $\alpha = .1$.

Using this model $\lambda_{2010}(u)$ we examine the impact of depth and distance in the same manner as we did previously, compare Fig 3.32 to Fig.3.26. Each shows a histogram of the water depths at point of encounter. Superimposed on each is the effect function for depth from an inhomogeneous model for the mean, $\lambda(u)$. While not identical the shape of the effect function is similar, and is subject to the same caveats raised during our discussion of Fig. 3.26. The histograms display some difference, the bimodality seen in 2009 is missing and there is a notable difference in the frequency of encounter in the 60-80 foot range; the reduction in the 60-80 foot encounters reflecting the near absence of encounters on the bank in the north center of the window.

Repeating with the effect function of distance, as seen in the right panel of Fig. 3.32 gives similar insight. The fit isn't quite as good but nothing in Fig. 3.32 contradicts the findings we gave for the 2009 data. The 2010 encounters may

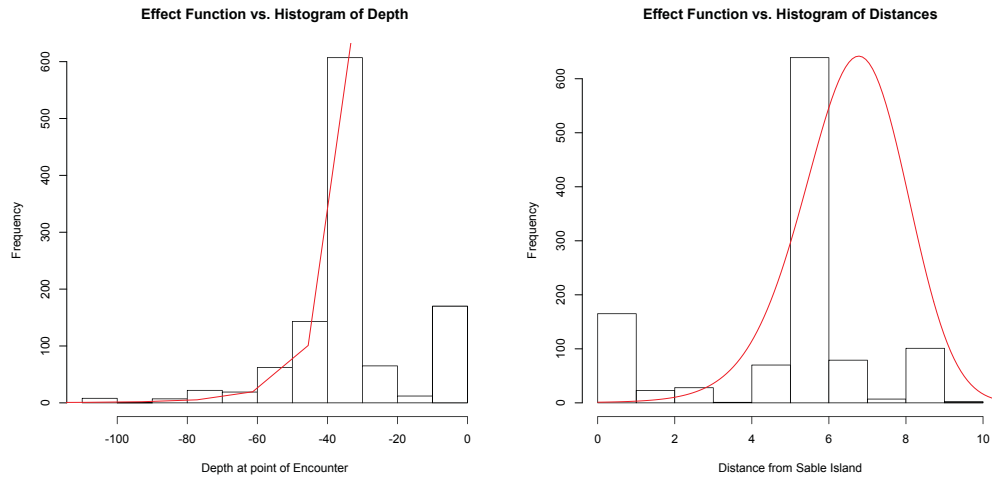


Figure 3.32: **Effect Function and Histogram of depth** Left Panel. Effect function and Histogram of water depth. Right Panel. Effect function and Histogram of distance.

have tended to be closer to Sable, in somewhat shallower water, but there is no explanation for the difference in clustering to be found here. The conjectures that the seals prefer certain water depths, the banks (between about 40 to about 70 feet), and forage in the closest places of suitable depth, (the penalty interpretation of distance), remain intact.

As an aside, the histograms for 2010 display an appreciable number of encounters in shallower water and short distance (0-10 feet and 0-10 minutes of arc), corresponding to an obvious grouping of encounters in the immediate vicinity of Sable Island. Encounters of this type, close and shallow, are conspicuously absent from the 2009 data. The VMTs have a feature that is intended to shut off the transmitter when the device is dry, i.e. the seal is ashore. Since we are concerned with at sea encounters, and the at sea status of this group is open to question, further scrutiny of these points and the influence may be warranted.

Having cast doubt on the idea that influence of the covariates upon the seals has altered we must look further to explain the difference in clustering present in the patterns. A recent discussion with Don Bowen, (one of the scientists carrying out the tagging project [6]) elicited the fact that the *composition* of the sampled

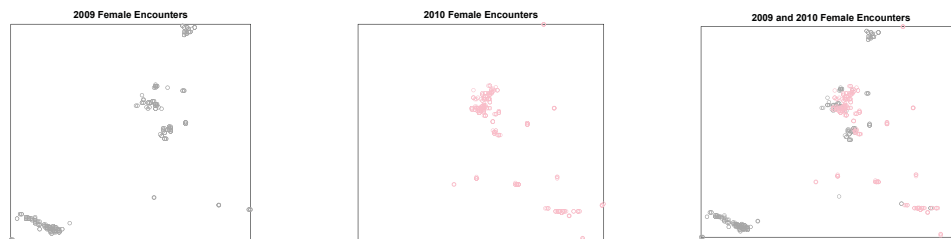


Figure 3.33: **Point Patterns of Female Encounters** Left Panel, The 2009 Female Encounters. Center Panel, The 2010 Female Encounters. Right Panel, the Superimposed Patterns

seals was not constant. As we described in §3.1 there were 14 seals in the 2009 sample, 8 females and 6 males. In 2010, based in part on the desire to record encounters with tagged Atlantic Cod, the researchers tagged 20 seals, 14 females and 6 males. If we conjecture that the movement patterns of seals are different for males and females [9] it is reasonable to question if the increased proportion of females in our 2010 data group is responsible for the difference in the encounter clustering suggested by our analyses.

To explore the idea of sex differences further we subset both the 2009 and 2010 data to include only those encounters in which the recording seal is a female. Females were chosen since there are more of them and hence, more data. Figure 3.33 shows these 2009, and 2010 encounter patterns.

We now construct a parametric model for $\hat{\lambda}(u)$ just as before but using only the 2009 female encounters data subset. Perhaps unsurprisingly at this point this proves to be cubic in both depth and distance with all parameters significant at $\alpha = .01$ except that of depth¹. We will now calculate the simultaneous Monte Carlo envelopes from this model and compare the result to the Inhomogeneous K function for the female subset of the 2010. This allows the assessment of the degree of clustering present in the 2010 female encounter data in reference to the 2009 female encounter data and in reference to the model based on the 2009 female encounters.

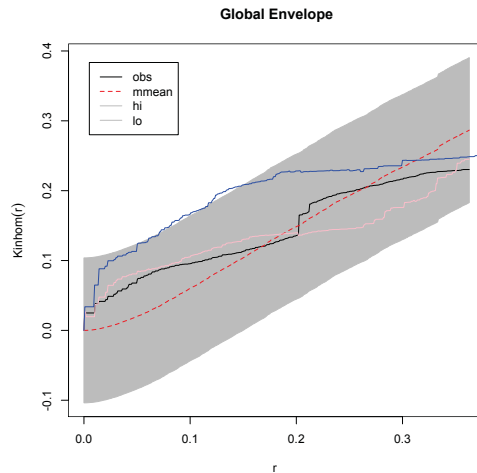


Figure 3.34: **Simulated Patterns from our Matérn model** The male and female seals display different clustering.

Fig. 3.34 shows the results. The envelope is the global simulation envelope derived from our inhomogenous model based on the 2009 female encounters. The dashed red line is the mean of the model prediction. The black line is $\hat{K}_{InhomFem2009}(r)$ from the 2009 female encounter data and the pink line is $\hat{K}_{InhomFem2010}(r)$ for the 2010 female encounter data.

Both $\hat{K}_{InhomFem2009}(r)$, the black line, and $\hat{K}_{InhomFem2010}(r)$, the pink, lie entirely within the global envelope. The nature of the divergence shows that the 2010 female data are not significantly different than the 2009 female data. The blue line is the estimate $\hat{K}_{InhomMale2010}(r)$. It leaves the global envelope and is therefore significantly (at $\alpha = .05$) different from the clustering pattern for the 2009 females while the clustering for the 2010 females is not. The male and female seals show different clustering and we conclude that changing the gender composition of the sample is likely responsible for the difference in the pattern from 2009 to 2010.

Chapter 4

Conclusions

In this study we have succeeded in providing statistical evidence of close association between the locations where seals encounter one another, the distance from their land habitat and the topography of the ocean floor. Indeed, conditioned on the seafloor topography and distance, the seal encounters, highly clustered when viewed in isolation, no longer exhibit the characteristic shape in their K function betraying clustering to a significant degree. Point pattern data analysis techniques enabled us to link the spatial distribution of these discrete events, the seal encounters, to water depth and to distance from Sable Island. This result relied not only on the point pattern of seal encounters, but additionally on information contained in the marks. That is, without the marks and the attendant ability to distinguish the parent points, efforts to model these data were markedly less successful. While this inability in distinguishing cluster processes from inhomogeneous processes is well known and our difficulties in producing an effective Poisson cluster model from the points alone was to some degree expected it does serve to underscore the utility of mark data. The fact that the cluster model proved the better model in comparison to the inhomogeneous Poisson is likely due to similar causes. The seals behaviour, and the workings of the VMT technology, argue for a cluster process. The seal's behaviour 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. Any model almost certainly had to be of inhomogeneous intensity. The VMT technology similarly argues for a cluster process since the approximate 2 minute refresh rate gives rise to multiple data points closely associated in time. Thus the eventual form of the model and the offspring status of the many subsequent points in the model.

The eventual model conforms closely to what was expected based on previous

knowledge of seal behaviour with the caveat that such encounters had not previously been directly 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 marine mammals, as well as other animals, postulate that their foraging behaviour is directed by the imperative of maximizing feeding benefit while simultaneously minimizing energy expenditure to reap said benefit. 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.

Spatial distribution of the seal's prey species would be a valuable augmentation. At best the use of seafloor topography stands as a proxy for the spatial distribution of the seal's preferred or available prey. While practical difficulties have thus far stymied efforts to build a data set from the tagged fish that constitute potential prey it is hoped that eventually this effort will succeed. Should it succeed and produce a point pattern of fish to seal encounters then the logical approach would be to condition the seal to seal point pattern on the seal to fish pattern. Evidence of clustering, in terms of elevated second order intensities for seals given the presence of fish, could be interpreted as seals seeking the fish. It would also be possible to condition the seal encounters on a continuous covariate such as a prey density. If the prey density were estimated at every point in some (D) then the pattern of seal to seal encounters (D^*) could be conditioned upon it as we have done here with the bathymetry. If such patterns or density estimates were available for more than one potential prey species, e.g. Atlantic Cod, *Gadus Morhua* and Atlantic Halibut, *Hippoglossus Hippoglossus*, then conditioning on both patterns or densities could be used to infer preference by prey species.

Data being forthcoming from this or other studies is not unlikely given the continued availability of this type of tag. In point of fact it is an explicit goal of the bioprobe project. Atlantic Cod, *Gadus Morhua*, were tagged on the Eastern Scotian Shelf (202 fish), and in the Southern Gulf of St. Lawrence (200 fish), however no encounters between these tagged Cod and the grey seal bioprobes were detected, suggesting either habitat overlap is not as expected, and hence the

highly politicized predator prey relationship unlikely, or that larger numbers of fish or tagging of other alternative stocks are required for reasonable probabilities of detection by the bioprobes.

4.1 Future Work

Time. In this study we have confined ourselves to a seasonal time scale. This is equivalent to assuming that the intensity of the process is constant with respect to time, $\lambda(t) = \lambda$. With the limited data available investigating the time component was unprofitable, however, given the stated intention of OTN to carry on collecting this data in future, this may change. With additional seasons of data the spatial analysis presented here can be expanded to a spatio-temporal analysis; we can look for intra season variation in the seal's usage of the space.

Simulated densities for the hypothetical prey species could be created. The bathymetry is a proxy for the distribution of the prey species. Replacing the bathymetry covariate with a density surface estimate of prey stock to directly model seal response to the prey distribution is a sensible refinement.

Investigate other non or semi-parametric methods for estimating intensity. Instead of the rather complex parametric models for the intensity a spatial GAM could be attempted. Using a smoother to approximate the functions of depth, distance or other covariate(s) is a logical extension.

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

Appendix A

Candidate Inhomogeneous Models

Model	Form
Model 00	$\hat{\lambda}(u) = \exp(\hat{\beta}_0)$
Model 01	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 D(u))$
Model 02	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 D(u) + \hat{\beta}_4 D(u)^2)$
Model 03	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2 + \hat{\beta}_6 D(u)^3)$
Model 04	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2 + \hat{\beta}_6 D(u)^3)$
Model 05	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3 + \hat{\beta}_4 D(u) + \hat{\beta}_5 D(u)^2)$
Model 06	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u))$
Model 07	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2)$
Model 08	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 B(u) + \hat{\beta}_2 B(u)^2 + \hat{\beta}_3 B(u)^3)$
Model 09	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 D(u))$
Model 10	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 D(u) + \hat{\beta}_2 D(u)^2)$
Model 11	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 D(u) + \hat{\beta}_2 D(u)^2 + \hat{\beta}_3 D(u)^3)$
Model 12	$\hat{\lambda}(u) = \hat{\beta}_0 \sqrt{-B(u)}$
Model 13	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \sqrt{-B(u)})$
Model 14	$\hat{\lambda}(u) = \hat{\beta}_0 \log B(u)$
Model 15	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log B(u))$
Model 16	$\hat{\lambda}(u) = \hat{\beta}_0 \sqrt{D(u)}$
Model 17	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \sqrt{D(u)})$
Model 18	$\hat{\lambda}(u) = \hat{\beta}_0 \log D(u)$
Model 19	$\hat{\lambda}(u) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log D(u))$

Table A.1: **Models** Summary of models for $\hat{\lambda}(\cdot)$

Model	AIC
Model 00	-8860.133
Model 01	-9887.9941
Model 02	-10822.7681
Model 03	-11218.7282
Model 04	-11144.8554
Model 05	-10903.8939
Model 06	-9553.4682
Model 07	-9921.7116
Model 08	-10007.0152
Model 09	-8976.8247
Model 10	-9468.3184
Model 11	-9745.0766
Model 12	-926.2856
Model 13	-9370.4628
Model 14	-9012.7676
Model 15	-9015.9637
Model 16	-9012.7676
Model 17	-9015.9637
Model 10	-9046.7771
Model 11	-9045.7829

Table A.2: **AIC** Table of AIC values for each evaluated model.

Appendix B

Code

```
library(spatstat)
  Creates polygons
p1<-read.table("isadora's shapefiles/stu005/poly1",sep="", header=FALSE)
p1x<-p1[,1]
p1y<-p1[,2]
p2<-read.table("isadora's shapefiles/stu005/poly2",sep="", header=FALSE)
p2x<-p2[,1]
p2y<-p2[,2]
p3<-read.table("isadora's shapefiles/stu005/poly3",sep="", header=FALSE)
p3x<-p3[,1]
p3y<-p3[,2]
p4<-read.table("isadora's shapefiles/stu005/poly4",sep="", header=FALSE)
p5x<-p5[,1]
p5y<-p5[,2]
p5<-read.table("isadora's shapefiles/stu005/poly5",sep="", header=FALSE)
p5x<-p5[,1]
p5y<-p5[,2]
p6<-read.table("isadora's shapefiles/stu005/poly6",sep="", header=FALSE)
p6x<-p6[,1]
p6y<-p6[,2]
p7<-read.table("isadora's shapefiles/stu005/poly7",sep="", header=FALSE)
p7x<-p7[,1]
p7y<-p7[,2]
p8<-read.table("isadora's shapefiles/stu005/poly8",sep="", header=FALSE)
p8x<-p8[,1]
```

```
p8y<-p8[,2]
p9<-read.table("isadora's shapefiles/stu005/poly9",sep="", header=FALSE)
p9x<-p9[,1]
p9y<-p9[,2]
p10<-read.table("isadora's shapefiles/stu005/poly10",sep="", header=FALSE)
p10x<-p10[,1]
p10y<-p10[,2]
p11<-read.table("isadora's shapefiles/stu005/poly11",sep="", header=FALSE)
p11x<-p11[,1]
p11y<-p11[,2]
p12<-read.table("isadora's shapefiles/stu005/poly12",sep="", header=FALSE)
p12x<-p12[,1]
p12y<-p12[,2]
p13<-read.table("isadora's shapefiles/stu005/poly13",sep="", header=FALSE)
p13x<-p13[,1]
p13y<-p13[,2]
p21<-read.table("isadora's shapefiles/stu005/poly21",sep="", header=FALSE)
p21x<-p21[,1]
p21y<-p21[,2]
p31<-read.table("isadora's shapefiles/stu005/poly31",sep="", header=FALSE)
p31x<-p31[,1]
p31y<-p31[,2]
p41<-read.table("isadora's shapefiles/stu005/poly41",sep="", header=FALSE)
p41x<-p41[,1]
p41y<-p41[,2]
p51<-read.table("isadora's shapefiles/stu005/poly51",sep="", header=FALSE)
p51x<-p51[,1]
p51y<-p51[,2]
p61<-read.table("isadora's shapefiles/stu005/poly61",sep="", header=FALSE)
p61x<-p61[,1]
p61y<-p61[,2]
```

```
p71<-read.table("isadora's shapefiles/stu005/poly71",sep="", header=FALSE)
p71x<-p71[,1]
p71y<-p71[,2]
p81<-read.table("isadora's shapefiles/stu005/poly81",sep="", header=FALSE)
p81x<-p81[,1]
p81y<-p81[,2]
p91<-read.table("isadora's shapefiles/stu005/poly91",sep="", header=FALSE)
p91x<-p91[,1]
p91y<-p91[,2]
p101<-read.table("isadora's shapefiles/stu005/poly101",sep="", header=FALSE)
p101x<-p101[,1]
p101y<-p101[,2]
p111<-read.table("isadora's shapefiles/stu005/poly111",sep="", header=FALSE)
p111x<-p111[,1]
p111y<-p111[,2]
p121<-read.table("isadora's shapefiles/stu005/poly121",sep="", header=FALSE)
p121x<-p121[,1]
p121y<-p121[,2]
p131<-read.table("isadora's shapefiles/stu005/poly131",sep="", header=FALSE)
p131x<-p131[,1]
p131y<-p131[,2]
p141<-read.table("isadora's shapefiles/stu005/poly141",sep="", header=FALSE)
p141x<-p141[,1]
p141y<-p141[,2]
p151<-read.table("isadora's shapefiles/stu005/poly151",sep="", header=FALSE)
p151x<-p151[,1]
p151y<-p151[,2]
p161<-read.table("isadora's shapefiles/stu005/poly161",sep="", header=FALSE)
p161x<-p161[,1]
p161y<-p161[,2]
p171<-read.table("isadora's shapefiles/stu005/poly171",sep="", header=FALSE)
```

```
p171x<-p171[,1]
p171y<-p171[,2]
p181<-read.table("isadora's shapefiles/stu005/poly181",sep="", header=FALSE)
p181x<-p181[,1]
p181y<-p181[,2]
p191<-read.table("isadora's shapefiles/stu005/poly191",sep="", header=FALSE)
p191x<-p191[,1]
p191y<-p191[,2]
p201<-read.table("isadora's shapefiles/stu005/poly201",sep="", header=FALSE)
p201x<-p201[,1]
p201y<-p201[,2]
p211<-read.table("isadora's shapefiles/stu005/poly211",sep="", header=FALSE)
p211x<-p211[,1]
p211y<-p211[,2]
p221<-read.table("isadora's shapefiles/stu005/poly221",sep="", header=FALSE)
p221x<-p221[,1]
p221y<-p221[,2]
p231<-read.table("isadora's shapefiles/stu005/poly231",sep="", header=FALSE)
p231x<-p231[,1]
p231y<-p231[,2]
p241<-read.table("isadora's shapefiles/stu005/poly241",sep="", header=FALSE)
p241x<-p241[,1]
p241y<-p241[,2]
p261<-read.table("isadora's shapefiles/stu005/poly261",sep="", header=FALSE)
p261x<-p261[,1]
p261y<-p261[,2]
p271<-read.table("isadora's shapefiles/stu005/poly271",sep="", header=FALSE)
p271x<-p271[,1]
p271y<-p271[,2]
Defines Windows
W<-owin(poly = list(list(x=c(-66.3,-66.3,-54,-54) , y=c
```

```

(48,41,41,48)), list(x=(p261x), y=(p261y)), list(x=(p31x),
y=(p31y)),list(x=(p71x),y=(p71y)),list(x=(p1x), y=(p1y)))
V<-owin(poly = list(list(x=c(-63,-63,-59,-59) , y=c(46,43,43,46)),
list(x=(p271x), y=(p271y)), list(x=(p261x), y=(p261y))))
Seal Encounters
ALLpairs <- (read.csv("RealSealtoSeal/ALLpairwise.csv"))
m<-ALLpairs[,8]
Creates Point Processes
RSALLP <-ppp(ALLpairs[,6],ALLpairs[,7], window=RS)
BBALLP <-ppp(ALLpairs[,6],ALLpairs[,7], window=BB)
Convex Hull and Bounding Box
Z <- convexhull.xy(ALLpairs[,6],ALLpairs[,7],)
Y<-bounding.box.xy(ALLpairs[,6],ALLpairs[,7],)
Ripley-Rasson
RS<-ripras(ALLpairs[,6],ALLpairs[,7], shape="convex")
BB<-ripras(ALLpairs[,6],ALLpairs[,7], shape="rectangle")
Create Bathymetry
x<-read.table("Depth2/MyVBathy-2332.xyz")
vec<-xV3
mat<-matrix(vec, nrow=181, ncol=241,byrow=TRUE)
newmat<-matrix(0,nrow=181,ncol=241)
for(i in 1:181){
newmat[i,]<-mat[181-i+1,]
}
plot1<-im(newmat, xcol=seq(-63,-59,length=241), yrow=seq(43,46,length=181))
Create Distance
sable<-NULL
sable[1]<-43.95000
sable[2]<-59.91666
dist<-dep
for(i in 1:43621){

```

```

out = sin(sable[1]/(180.00*pi))*sin(dist[i,2]/(180.00*pi))
+cos(sable[1]/(180.00*pi))*cos(dist[i,2]/(180.0*pi))
cos(abs(sable[2]-dist[i,1])/(180.0*pi));
out = (acos(out)*180.0)/(pi)*60.0;
dist[i,3]<-out
}
vec<-distV3
mat<-matrix(vec, nrow=181, ncol=241,byrow=TRUE)
newmat<-matrix(0,nrow=181,ncol=241)
for(i in 1:181) newmat[i,]<-mat[181-i+1,]
plot2<-im(newmat, xcol=seq(-63,-59,length=241), yrow=seq(43,46,length=181))
Quadrat Test
Q <- quadratcount(BBALLP, nx = 5, ny = 5)
plot(BBinitials, cex = 0.5, pch = "+")
plot(Q, add = TRUE, cex = .7)
quadrat.test(BBinitials, nx = 5, ny = 5)
Quadrats by Covariate
chd<-quantile(BBdepth, probs =(0:8)/8)
BBD<-cut(BBdepth, breaks=chd,labels = 1:8)
TESS<-tess(image=BBD)
plot(TESS, ribbon=FALSE, main="Tesselation by Depth")
plot(BBALLP, add=TRUE)
qb<-quadratcount(BBALLP,tess=TESS)
quadrat.test(BBALLP,tess=TESS)
chdd<-quantile(BBdist, probs =(0:8)/8)
BBD2<-cut(BBdist, breaks=chdd,labels = 1:8)
TESS2<-tess(image=BBD2)
plot(TESS2, ribbon=FALSE, main="Tesselation by Distance")
plot(BBALLP, add=TRUE)
qb<-quadratcount(BBALLP,tess=TESS)
quadrat.test(BBALLP,tess=TESS)

```

```

Homogeneous Poisson
HP<-rpoispp(795,win=BB)
plot(HP, main="Homogenous Poisson Realization")
plot(BBALLP, main="Seal Encounter Locations")
fitPois<-ppm(BBALLP, 1)
plot(predict(fitPois), col="blue", main="Homogenous Poisson Intensity")
plot(BBALLP, add=TRUE)
diagnose.ppm(fitPois, which="smooth")

InhomoPoisson
Kernel Estimate
library(ks)
yack<-cbind(ALLpairs[,6],ALLpairs[,7])
varcov<-cov(yack)
H<- Hpi(x=yack)
X <- density(BBALLP, varcov=H, edge=FALSE)
IHP<-rpoispp(X,win=BB,edge=FALSE)
plot(IHP[BB], main="Inhomogenous Poisson Realization")
JJ<-J[BB, drop=FALSE]
ee <- expression(rpoispp(J))
EE <- envelope(BBALLP, Kinhom, nsim = 49, global = TRUE, simulate = ee)

Parametric models for the process
model00<-ppm(BBALLP, 1)
model01<-ppm(BBALLP, depth+dist, covariates=list
(depth=plot3, dist=plot2))
model02<-ppm(BBALLP, polynom(depth,2)+polynom(dist,2),
covariates=list(depth=plot3, dist=plot2))
model03<-ppm(BBALLP, polynom(depth,3)+polynom(dist,3),
covariates=list(depth=plot3, dist=plot2))
model04<-ppm(BBALLP, polynom(depth,2)+polynom(dist,3),
covariates=list(depth=plot3, dist=plot2))
model05<-ppm(BBALLP, polynom(depth,3)+polynom(dist,2),

```



```

covariates=list(depth=plot3, dist=plot2))
model06<-ppm(BBALLP, depth, covariates=list(depth=plot3))
model07<-ppm(BBALLP, polynom(depth,2), covariates=list
(depth=plot3))
model08<-ppm(BBALLP, polynom(depth,3), covariates=list (depth=plot3))
model09<-ppm(BBALLP, dist, covariates=list(dist=plot2))
model10<-ppm(BBALLP, polynom(dist,2), covariates=list
(dist=plot2))
model11<-ppm(BBALLP, polynom(dist,3), covariates=list
(dist=plot2))
model12<-ppm(BBALLP, offset(sqrt(-depth)), covariates=list(depth=plot3))
model13<-ppm(BBALLP, sqrt(-depth), covariates=list
(depth=plot3))
model14<-ppm(BBALLP, offset(log(depth)), covariates=list(depth=plot3))
model15<-ppm(BBALLP, log(depth), covariates=list
(depth=plot3))
model16<-ppm(BBALLP, offset(sqrt(dist)), covariates=list
(dist=plot2))
model17<-ppm(BBALLP, sqrt(dist), covariates=list
(dist=plot2))
model18<-ppm(BBALLP, offset(log(dist)), covariates=list
(dist=plot2))
model19<-ppm(BBALLP, log(dist), covariates=list(dist=plot2))
AIC(model00,model01,model02,model03,model04,model05,model06,model07,
model08,model09,model10,model11,model12,model13,model14,model15,
model16,model17,model18,model19)
pp <- expression(rmh(model03))
PP <- envelope(BBALLP, Kinhom, nsim = 39, global = TRUE, simulate = pp)
PIHP<-rpoispp(predict(model03),win=BB,edge=FALSE)
plot(PIHP, main="Inhomogenous Poisson Realization")
plot(predict(model03), main="Inhomogenous Poisson Fitted Intensity")

```

```

diagnose.ppm(model03, which="smooth")
Minimum Contrast Models
modelT<-kppm(RSALLP, x+y+ploynom(depth,3)+ploynom(dist,3),
"Thomas", covariates=list(depth=plot1))
plot(rpoint(892, predict(Thomas))[RS], main="Realization from
minimum contrast Thomas")
modelM<-kppm(RSALLP, x+y+depth, "MatClust", covariates=list
(depth=plot1))
plot(rpoint(892, predict(Matern))[RS], main="Realization from
minimum contrast Matern")
Parametric Models for the Parents
model00<-ppm(BBinitials, 1)
model11<-ppm(BBinitials, depth+dist, covariates=list
(depth=plot1, dist=plot2))
model12<-ppm(BBinitials, polynom(depth,2)+polynom(dist,2),
covariates=list(depth=plot1, dist=plot2))
model13<-ppm(BBALL, polynom(depth,3)+polynom(dist,3),
covariates=list(depth=plot1, dist=plot2))
model14<-ppm(BBinitials, polynom(depth,2)+polynom(dist,3),
covariates=list(depth=plot1, dist=plot2))
model15<-ppm(BBinitials, polynom(depth,3)+polynom(dist,2),
covariates=list(depth=plot1, dist=plot2))
X<-predict(model
mm<-expression(rMatClust(X,.06985,6.31,RS))
MM <- envelope(BBALLP, Kest, nsim = 9, global = TRUE, simulate = mm)
tt<-expression(rThomas(X,.0365,6.31,BB))
TT <- envelope(BBALLP, Kest, nsim = 9, global = TRUE, simulate = tt)
Validation
BB09<-ripras(ALLpairs[,6],ALLpairs[,7], shape="rectangle")
BB10<-ripras(All2010[,3],All2010[,4], shape="rectangle")
BBALL<-ripras(G,H, shape="rectangle")

```

```

All09BB <-ppp(ALLpairs[,6],ALLpairs[,7], window=BB09)
All10BB <-ppp(All2010[,3],All2010[,4], window=BB10)
ALL<-ppp(G,H, window=BBALL)
modelBB09<-ppm(All09BB, polynom(depth,3)+polynom(dist,3),
covariates=list(depth=plot1, dist=plot2))
modelBB10<-ppm(All10BB, polynom(depth,3)+polynom(dist,3),
covariates=list(depth=plot1, dist=plot2))
modelBBALL<-ppm(ALL, polynom(depth,3)+polynom(dist,3),
covariates=list(depth=plot1, dist=plot2))
distplot09<-plot2[All09BB]
distplot10<-plot2[All10BB]
distplotALL<-plot2[ALL]
hist(distplot09,xlab="Distance from Sable Island", main="Effect
Function vs. Histogram of Distances")
plot(effectfun(modelBB09,depth=-40,"dist"), add=TRUE)
hist(distplot10,xlab="Distance from Sable Island", main="Effect
Function vs. Histogram of Distances")
plot(effectfun(modelBB10,depth=-40,"dist"), add=TRUE)
hist(distplotALL)
plot(effectfun(modelBBALL,depth=-40,"dist"), add=TRUE)
hist(distplotALL,xlab="Distance from Sable Island", main="Effect
Function vs. Histogram of Distances")
plot(effectfun(modelBBALL,depth=-95,"dist"), add=TRUE, col="blue")
plot(effectfun(modelBB09,depth=-95,"dist"), add=TRUE,)
plot(effectfun(modelBB10,depth=-65,"dist"), add=TRUE, col="red")
depthplot09<-plot3[All09BB]
depthplot10<-plot3[All10BB]
depthplotALL<-plot3[ALL]
hist(depthplot09)
plot(effectfun(modelBB09,dist=5,"depth"), add=TRUE)
hist(depthplot10,xlab="Depth at point of Encounter", main="Effect

```

```

Function vs. Histogram of Depth")
plot(effectfun(modelBB10,dist=4,"depth"), add=TRUE, col="red")
hist(depthplotALL,xlab="Depth at point of Encounter", main="Effect
Function vs. Histogram of Depth")
plot(effectfun(modelBBALL,dist=5,"depth"), add=TRUE, col="blue")
plot(effectfun(modelBB09,dist=5,"depth"), add=TRUE,)
plot(effectfun(modelBB10,dist=5,"depth"), add=TRUE, col="red"
male/female?
Females09 <- (read.csv("RealSealtoSeal/females09.csv"))
Females10 <- (read.csv("2010_detections_gps/females10.csv"))
Fem09 <-ppp(Females09[,6],Females09[,7], window=BB09)
Fem10 <-ppp(Females10[,3],Females10[,4], window=BB09)
Males09 <- (read.csv("RealSealtoSeal/males09.csv"))
Males10 <- (read.csv("2010_detections_gps/males2010.csv"))
Male09 <-ppp(Males09[,12],Males09[,13], window=BB09)
Male10 <-ppp(Males10[,3],Males10[,4], window=BB09)
MALES<-superimpose(Male09,Male10)
FEMS<-superimpose(Fem09,Fem10)
fm <- expression(rmh(Fmodel09))
FM <- envelope(Fem09, Kinhom, nsim = 19, global = TRUE, simulate = fm)
plot(FM, main="Global Envelope")
plot(Kinhom(Fem10), c(border,theo) r, add=TRUE, col="pink")
plot(Kinhom(Male10), c(border,theo) r, add=TRUE, col="purple")
plot(Kinhom(MALES), c(border,theo) r, add=TRUE, col="blue")
plot(Kinhom(FEMS), c(border,theo) r, add=TRUE, col="purple")

```

Bibliography

- [1] J. Møller A. Baddeley and R. P. Waagepetersen. Non and semi-parametric estimation of interaction in inhomogenous point patterns. *Statistica Neerlandica*, Vol. 54(3), 329-350, 2000.
- [2] C. Amante and B. W. Eakins. Etopo1 1 arc-minute global relief model: Procedures, data sources and analysis. *NOAA Technical Memorandum NESDIS NGDC-24*, 1-19, 2009.
- [3] S. J. Morreale B. Gardner, P. J. Sullivan and S. P. Epperly. 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, 2008.
- [4] A. Baddeley and R. Turner. Spatstat, an r package for analyzing spatial point patterns. *Journal of Statistical Software*, Vol. 12, 1-42, 2005.
- [5] M. Berman and T. R. Turner. Approximating point process likelihoods with glim. *Applied Statistics*, Vol. 41, 31-38, 1992.
- [6] W. D. Bowen. As personal communication.
- [7] D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Series B*, Vol. 17, No. 2, 129-164, 1955.
- [8] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, Rev. ed. 1993.
- [9] I. D. Jonsen D. C. Lidgard, W. D. Bowen and S. J. Iverson. A novel approach to studying at-sea associations in marine predators. *Unpublished Manuscript*, 2011.
- [10] W. S. Kendall D. Stoyan and J. Mecke. *Stochastic Geometry and its Applications*. Wiley, 1983.
- [11] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Vol I*. Springer, 2003.
- [12] P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Academic Press, 1983.
- [13] T. Duong and M. L. Hazelton. Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation. *Journal of Multivariate Analysis*, Vol. 93, 417-433, 2005.

- [14] A. Baddeley et. al. (Editors). *Case Studies in Spatial Point Process Modelling*. Springer, 2006.
- [15] J. M. Fromentin F. Royer and P. Gaspar. Association between bluefin tuna schools and oceanic features in the western mediterranean. *Marine Ecology Progress Series*, Vol. 269, 249-263, 2004.
- [16] J. J. Faraway. *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall, 2006.
- [17] R. A. Myers I. D. Jonsen and J. Mills Flemming. Meta-analysis of animal movement using state-space models. *Ecology*, Vol. 84(11), 3055-3063, 2003.
- [18] K. A. Keating and S. Cherry. Modeling utilization distributions in space and time. *Ecology*, Vol. 90(7), 1971-1980, 2009.
- [19] J. Møller. 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, 2010.
- [20] J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, 2004.
- [21] B. Matérn. Spatial variation. *Meddelanded Stat.*, Skogsforsk 49 (5), 1-144, 1960.
- [22] G. A. Rose R. L. O'Driscoll, D. C. Schneider and G. R. Lilly. 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, 2000.
- [23] S. O. Rice. Mathematical analysis of random noise. *Bell Systems Technical Journal*, Vol. 23, 282-332, 1944.
- [24] B. D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, Vol. 13, No. 2, 255-266, 1976.
- [25] B. D. Ripley and J. P. Rasson. Finding the edge of a poisson forest. *Journal of Applied Probability*, Vol. 14, 483-491, 1977.
- [26] Oliver Schabenberger and Carol A. Gotway. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, 2005.
- [27] C. J. Schwartz. The jolly-seber model: More than just abundance. *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 6(2), 195-205, 2001.

- [28] M. Thomas. A generalization of poisson's binomial limit for use in ecology. *Biometrika*, Vol. 36, 18-25, 1949.
- [29] E. Vanmarcke. *Random Fields, Analysis and Synthesis*. MIT Press, 1988.
- [30] R Waagepetersen. An estimating function approach to inference for inhomogeneous neymancott processes. *Biometrics*, Vol. 63(1), 252-258, 2007.
- [31] D. G. Woolford and W. J. Braun. Convergent data sharpening for the identification and tracking of spatial temporal centers of lightning activity. *Environmetrics*, Vol. 18, 461-479, 2007.