

COSINE: A tool for constraining spatial neighbourhoods in marine environments

by

César Augusto Suárez
B.Sc., Universidad Santiago de Cali, 2003

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Geography

© César Augusto Suárez, 2013
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

COSINE: A tool for constraining spatial neighbourhoods in marine environments

by

César Augusto Suárez
B.Sc., Universidad Santiago de Cali, 2003

Supervisory Committee

Dr. Trisalyn A. Nelson, Department of Geography
Supervisor

Dr. Rosaline R. Canessa, Department of Geography
Departmental Member

Abstract

Supervisory Committee

Dr. Trisalyn A. Nelson, Department of Geography
Supervisor

Dr. Rosaline R. Canessa, Department of Geography
Departmental Member

Spatial analysis methods used for detecting, interpolating or predicting local patterns require a delineation of a neighbourhood defining the extent of spatial interaction in geographic data. The most common neighbourhood delineation techniques include fixed distance bands, k-nearest neighbours, or spatial adjacency (contiguity) matrices optimized to represent spatial dependency in data. However, these standard approaches do not take into consideration the geographic or environmental constraints such as impassable mountain ranges, road networks or coastline barriers. Specifically, complex marine landscapes and coastlines present common problematic neighbourhood definitions for standard neighbourhood matrices used in the spatial analysis of marine environments. Therefore, the goal of our research is to present a new approach to constraining spatial neighbourhoods when conducting geographical analysis in marine environments. To meet this goal, we developed methods and software (COnstraining SpAtIal NEighbourhoods - COSINE) for modifying spatial neighbourhoods, and demonstrate their utility in two case studies. Our method enables delineation of neighbourhoods that are constrained by coastlines and the direction of marine currents. Our software calculates and evaluates whether neighbouring features are separated by land, or are within a user defined angle that excludes interaction based on directional processes. Using decision rules a modified spatial weight matrix is created, either in binary or row-standardized format. Within open source software (R), a graphical user interface enables users to modify the standard spatial neighbourhood definition distance, inverse distance and k-nearest neighbour. Two case studies are presented to demonstrate the usefulness of the new approach for

detecting spatial patterns: the first case study observes marine mammals' abundance and the second, oil spill observation. Our results indicate that constraining spatial neighbourhoods in marine environments is particularly important at larger spatial scales. The COSINE tool has many applications for modelling both environmental and human processes.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	v
List of Tables	vii
List of Figures	viii
Acknowledgements	x
Dedication	xi
1.0 Introduction	1
1.1 Research Context	1
1.2 Thesis themes and objectives	3
1.3 Thesis structure	4
2.0 COSINE: A tool for constraining spatial neighbourhoods in marine settings	6
2.1 Introduction	6
2.2 Objective	8
2.3 Context: Spatial neighbourhoods in marine applications	9
2.4 COSINE Tool	13
Description	14
Creating the graphical user interface (GUI)	14
Defining standard spatial neighbourhoods	14
Including constraining factors in the spatial neighbourhood delineation	15
Output for spatial analysis	19
Implementation	20
2.4.1 COSINE Findings	21
Graphical user interface (GUI)	21
2.5 Case Study	23
Marine mammals distribution and abundance	24
Oil spills occurrence	24

	vi
Study Area	25
Data description and preparation	26
Marine mammals data.....	26
Oil spills from vessels.....	28
Currents data.....	29
2.5.1 Methods.....	30
2.5.2 Results.....	33
Marine mammals abundance	33
Oil spills occurrence	36
2.6 Discussion	43
2.7 Conclusions.....	47
3.0 Conclusion	49
3.1 Summary	49
3.2 Research Contributions.....	51
3.3 Research Opportunities.....	52
Bibliography	54
Appendix A Graphical user interface (GUI).....	64
Input / Output section.....	64
Directionality section	65
Neighbourhood type section	65
Appendix B Help File	66
Appendix C R Code	69
Graphical user interface (GUI)	69
Neighbourhood Type: Distance	74
Neighbourhood Type: Inverse Distance	80
Neighbourhood Type: K-Nearest Neighbour (Knn).....	86

List of Tables

Table 2-1. Total number of individuals, 2004	27
Table 2-2. Number of high and low clusters for oil spills occurrence and marine mammals abundance	36

List of Figures

Figure 2-1. Common spatial neighbourhood delineations (a) fixed distance; (b) K-nearest; (c) adjacency	10
Figure 2-2. Unconstrained spatial neighbourhood delineation for: (a) fixed distance; (b) K-nearest neighbourhood.....	15
Figure 2-3. Unconstrained spatial neighbourhood delineation intersecting barriers -land- (White background represents water and colour represents land) for: (a) fixed distance; (b) K-nearest neighbourhood	15
Figure 2-4. Direction of nodes to calculate angle divergence	18
Figure 2-5. Direction of nodes to calculate angle divergence	19
Figure 2-6. Constrained spatial neighbourhood delineation (White background represents water and colour represents land) for: (a) fixed distance; (b) K-nearest neighbourhood .	20
Figure 2-7. Graphical user interface (GUI).....	23
Figure 2-8. Extent of the study area.....	26
Figure 2-9. Marine mammals sightings, 2004	28
Figure 2-10. Oil spill observations, 2008.....	29
Figure 2-11. Marine mammals abundance - Number of neighbourhood per neighbourhood delineation. 40 km – 120 km distance	34
Figure 2-12. Marine mammals abundance - G_i * Z-score per neighbourhood delineation. 40 km – 120 km distance	35
Figure 2-13. Oil spills - Number of neighbourhood per neighbourhood delineation. 40 km – 120 km distance	37
Figure 2-14. Oil spills - G_i * Z-score per neighbourhood delineation. 40 km – 120 km distance	38
Figure 2-15. Marine mammals abundance - clusters of high and low values.....	39
Figure 2-16. Oil spills occurrence - clusters of high and low values.....	40
Figure 2-17. Marine mammals abundance - Open waters - clusters of high and low values	41

Figure 2-18. Marine mammals abundance - Inlets - clusters of high and low values 41

Figure 2-19. Oil spills occurrence - Open waters - clusters of high and low values 42

Acknowledgements

I would like to express my deepest gratitude to my advisor and baby coach Dr. Trisalyn Nelson, for her mentoring, support, guidance and foremost patience, throughout the steep learning curve. I would also thank Dr. Rosaline Canessa for her valuable insight and suggestion during this research.

I would like to thank Raincoast Conservation Foundation, Dr. Chris Darimont and Caroline Fox for providing marine mammals' data as well as insight information for my research. Thank to Norma Serra-Sogas for and National Aerial Surveillance Program (NASP) for providing data on oil spills.

Special thanks to Dr. John Verzani for his support during the graphical user interface development, Felipe Ardila and Jed Long for their support during the wonderful world of programming in R.

Special thanks to my colleagues in the SPAR lab (Go SPAR), Jess, Shanley, Matthew, Keith and Jed, for their countless hours of sharing ideas, troubleshooting and laughing. And finally, to the Geography department staff for their assistance and advice.

To my father Jorge and mother Alicia for their support and positive attitude towards this endeavour. To my lovely wife for her patience and advice during difficult times and to my Enano (a.k.a. Munchkin or Lilliputien) Nicholas, who taught me that the best alarm (with no snoozing option) in the world is a baby's laugh, chat, cry or scream.

Dedication

Esta tesis está dedicada a todos aquellos que estuvieron a mi lado en buenos y malos momentos, para aquellos que no están ya con nosotros pero nunca olvidados, muy pronto todos estaremos reunidos nuevamente. A la luz de mis ojos y verdadero norte, Nicholas y Liliana, los amo demasiado, ustedes son la razón que me hace ser cada vez (poco a poco) más un mejor hombre y padre, eres la bendición infinita de nuestro creador. A mis padres, sobrino y a mis suegros y cuñado, todos ustedes han sido y serán una gran e importante parte en mi vida.

This thesis is dedicated to all the people who stand by my side for better or for worse, for those who are gone but not forgotten, soon we all will be sharing the same destiny. To the lights of my eyes and true north, Nicholas and Liliana, I love you so much, you are the reason I'm becoming (one step at a time) a better man and father, you are the truly blessing I was waiting for in my life. To my parents, nephew, sister, and in-laws, you all are an important part in my life.

“Algún día en cualquier parte, en cualquier lugar indefectiblemente te encontrarás a ti mismo, y ésa, sólo ésa, puede ser la más feliz o la más amarga de tus horas...”

Pablo Neruda

"Someday, somewhere - anywhere, unfailingly, you'll find yourself, and that, and only that, can be the happiest or bitterest hour of your life."

Pablo Neruda

1.0 Introduction

1.1 Research Context

Geographical information systems have revolutionized the way we manage, visualize, and interpret geospatial data leading to unparalleled advancements in spatial analysis. Spatial analysis considers location, distance, area and interaction aspects of geospatial data to quantify patterns in phenomena, with the objective to understand and predict complex human or environmental processes (Anselin, 1989; Anselin & Getis, 2010; Wong & Lee, 2005). Routed in spatial analysis is the concept of geographic dependency where phenomena in close proximity are more likely to be related than phenomena further apart (Tobler, 1970). Quantifying the spatial extent of this geographic dependency is important for understanding a wide variety of processes with applications for fisheries (Booth, 2000; Lorenzen et al., 2010), pollution (Jerrett et al., 2005), land use (Chomitz & Thomas, 2003), human behaviour (Rushton, 1993) research.

A key component and often misunderstood step in spatial analysis is the delineation of a spatial neighbourhood to represent the geographical extent of spatial dependency within and between geospatial datasets. Techniques such as interpolation (Gold, 1994a; Isaaks & Srivastava, 1989), clustering detection (Fotheringham & Zhan, 2010), spatial autocorrelation analysis (Cliff & Ord, 1970), and spatial regression (Brunsdon et al., 1996) rely on intelligent selections of spatial neighbourhoods to perform analysis. Take, for instance, a theoretical example where a researcher has gone into the field to collect sea surface temperatures readings for a geographic area at discrete locations within a 20 mile extent off shore. Naturally, the researcher will want to create a continuous

representation of sea surface temperature over the whole study area; however, currents and proximity to land may affect the variability in temperature readings. In order to extrapolate the discrete temperature readings into a model of sea surface temperature, there needs to be a quantification of the similarity between the readings in geographic space. The geographic space where pattern is quantified is defined by a delineation of a spatial neighbourhood used to optimize the representation of phenomena similarity in geographic space.

Often commercial GIS software provides limited tools and methods for spatial neighbourhood selection. The focus has been to delineate neighbourhoods by an Euclidean distance (Anselin & Getis, 2010) or spatial contiguity (O'Sullivan & Unwin, 2002), which ignore the geographic context such as natural land barriers in the case of environmental studies, or movement restrictions such as road networks in human research. Common spatial neighbourhood delineations are implemented due to their relative simplicity, in part because the additional dimensions (x, y) already add mathematical and computational complexities.

Marine environments are particularly complex for spatial analysis due to irregular study area extents and structure. The most common spatial neighbourhoods used in spatial studies of marine environments are delineated by Euclidean distance (Grech et al., 2011; Magalhaes et al., 2002; ODriscoll, 1998). The study of species and environmental interactions in both terrestrial and aquatic environments is increasing (Anselin & Getis, 2010; Wong & Lee, 2005). Therefore, the need for tools that delineates intelligent spatial neighbourhoods, which take into account natural restrictions on spatial dependence of processes, is evident.

The concept of constraining spatial neighbourhoods is not new. Examples, include contiguity or adjacency spatial neighbourhood delineations, where the use of tessellated polygons, either by Voronoi or Constrained Delaunay triangulation, are implemented in point patterns distribution by altering (adding, changing or removing) a neighbour (Gold & Condal, 1995; Gold, 1992, 1994a; Nordvik & Harding, 2008). In commercial software (i.e., ArcGIS10[®]), approaches to constrain spatial neighbourhoods are implemented through network analysis tools, where features such as road closures or blocked intersections can be added to generate spatially weighted neighbourhoods (Esri, 2011). Moreover, interpolation techniques, such as Inverse Distance Weighted (IDW) or spline with barriers can limit the area of influence in the interpolation. Despite previous approaches and specified tools for restricting analysis by natural barriers, there is still a lack of accessibility to tools for constrained neighbourhood definitions that restrict the utility of spatial analysis across disciplines.

The constraint requirements for spatial analysis success and the increase in availability of geographic data require that users have the ability to identify and select the appropriate spatial neighbourhoods for analysis. Research has shown that failure to correctly delineate neighbourhoods can lead to misinterpretation in the results (Nelson & Robertson, 2012).

1.2 Thesis themes and objectives

The goal of this thesis is to propose a new method and open source tool to modify existing neighbourhood definitions to incorporate geographic or topographic contexts by undertaking the following objectives:

- 1) Identify and describe standard spatial neighbourhoods used in marine research and summarize two approaches for constraining spatial neighbourhood definitions to improve spatial analysis in marine environments.
- 2) Develop open source software that will limit the influence of standard neighbourhood definitions by implementing constraint approaches (e.g., land and/or direction).
- 3) Demonstrate the advantages of our new method of spatial neighbourhood delineation on two marine case studies: marine mammals' abundance and oil spill occurrence.

1.3 Thesis structure

This thesis is divided into three chapters, and three appendices. Chapter 1, this chapter is an introduction on the subject of spatial neighbourhood definitions and issues with implementation in marine environments. Chapter 2 is the core of the thesis, and it has been prepared as a manuscript for peer-review, with the introduction to the problem of spatial neighbourhoods in marine environments and a proposed method (COSINE) to delineate them, as well as two case studies to demonstrate impacts of analysis when using unconstrained or constrained spatial neighbourhoods. Chapter 3 provides a summary of the main findings, contributions and research opportunities of this research. The last section, the Appendix, contains description and information of COSINE software that will allow users to integrate the final output in GIS software or spatial analysis packages. COSINE output and software are flexible enough to be manipulated by end users at any time. The appendix section has three components: i) a description of the graphical user interface (GUI), ii) a help file with useful information for the correct use of COSINE

software, including procedures on how to integrate COSINE output with common spatial analysis software and an explanation of fields for the COSINE output (shapefiles), iii) R code for COSINE GUI, and COSINE spatial neighbourhood definitions (distance, inverse distance and k-nearest neighbours).

2.0 COSINE: A tool for constraining spatial neighbourhoods in marine settings

2.1 Introduction

Geographic information systems (GIS) and spatial analysis have improved the way we map, analyze and model a wide variety of geographic data with applications for multiple disciplines, such as, human behaviour (Funk et al., 2010; Rushton, 1993), land use (Chomitz & Thomas, 2003), pollution (Dionisio et al., 2010; Jerrett et al., 2005), and ecology (Fortin & Dale, 2005). Briefly, spatial analysis research identifies spatial patterns of association caused by, or affected by, any kind of feature or phenomena. In turn, GIS enables datasets to be spatially integrated to conduct spatial analysis, statistical modelling and probability mapping of environments. An essential step in any spatial analysis technique is to consider phenomena similarity by their location in space (Anselin, 1989; Fotheringham & Rogerson, 2009).

In ecology and spatial analysis, the region of interaction between features is clearly identified. For instance, in ecology, the term ecological neighbourhoods refers to areas that are scaled to a particular ecological process, time period and an organism's mobility or activity (Wiens, 1989). For spatial analysis, the spatial neighbourhood is the area where geographic features influence one another (O'Sullivan & Unwin, 2002). In ecology or spatial analysis, the "space" or region of spatial interaction must be delineated in what is called a spatial neighbourhood (Nelson & Robertson, 2012). Whether performing an interpolation, quantifying spatial autocorrelation, or conducting spatially explicit regression methods, spatial neighbourhoods are delineated by the analyst and should incorporate some fundamental understanding of spatial dependence

inherent in the dataset (O'Sullivan & Unwin, 2010). For instance, during spatial autocorrelation analysis, the analyst quantifies the level of attribute similarity in a geographical group of data (Redfern et al., 2013; Richardson & Schoeman, 2004; Russell et al., 1992); in this case, spatial neighbourhood is what it is used to define the geography.

The most common spatial neighbourhood delineations are distance, k-nearest neighbours, spatial adjacency (contiguity), and spatial dependency. Though types of neighbourhoods are discussed below it is helpful to understand their utility in spatial analysis. Delineations based on distance or adjacency imply there is a specific distance or contiguity level beyond which the interaction among features is no longer present. A spatial neighbourhood delineated by distance assumes interaction among features occurs within a distance given. For example, Windle et al., (2010) analyzed the distribution and abundance of Atlantic cod (*Gadus morhua*) using spatial neighbourhood delineated by distance (Euclidean distance).

Attempts have been made to constrain spatial neighbourhoods in spatial analysis. For example in terrestrial ecology, Nelson and Robertson (2012) used spatial neighbourhoods delineations, such as distance and adjacency (watershed), to identify infestation hot spots in epidemic mountain pine beetle populations in British Columbia, Canada. They found that neighbourhood selection is impacted by the use of topographic barriers, specifically watersheds, and the use of an unconstrained approach can alter the cluster results by combining unconnected populations as well as detecting spatial relationship across scales. Another approach used to constrain spatial neighbourhoods is the use of Voronoi or Constrained Delaunay triangulation for proximity polygons using

contiguity or adjacency spatial neighbourhood definitions (Gold & Condal, 1995; Gold, 1992, 1994a).

Marine researchers are increasingly using GIS and spatial analysis to investigate abundance (Mellin et al., 2010), distribution (Grech et al., 2011) and impacts on different marine species on diverse ecosystems (West & van Woesik, 2001). Similarly, spatial analysis has been used to predict spatial patterns of illegal or accidental oil discharge from marine commercial vessels (Serra-Sogas et al., 2008). Marine research typically applies simple and standard neighbourhoods and may benefit from novel and ecological based neighbourhood delineations. Whether to perform analysis of hot spots (LeDrew et al., 2004; Serra-Sogas et al., 2008) or to predict probability of species' presence/absence (Overholtz, 2002; Windle et al., 2010), spatial analysis requires a spatial neighbourhood.

Due to complex marine landscape and coastlines, using standard neighbourhood definitions for spatial analysis of marine environments is problematic. Standard approaches to delineate spatial neighbourhoods do not take into account geographic or topographic context. Failure to do so, may lead to assume that unrelated features are linked (Nelson & Robertson, 2012). Not taking into account geographic or topographic context as constraints, may identify patterns or relationships across impassable environmental or manmade boundaries.

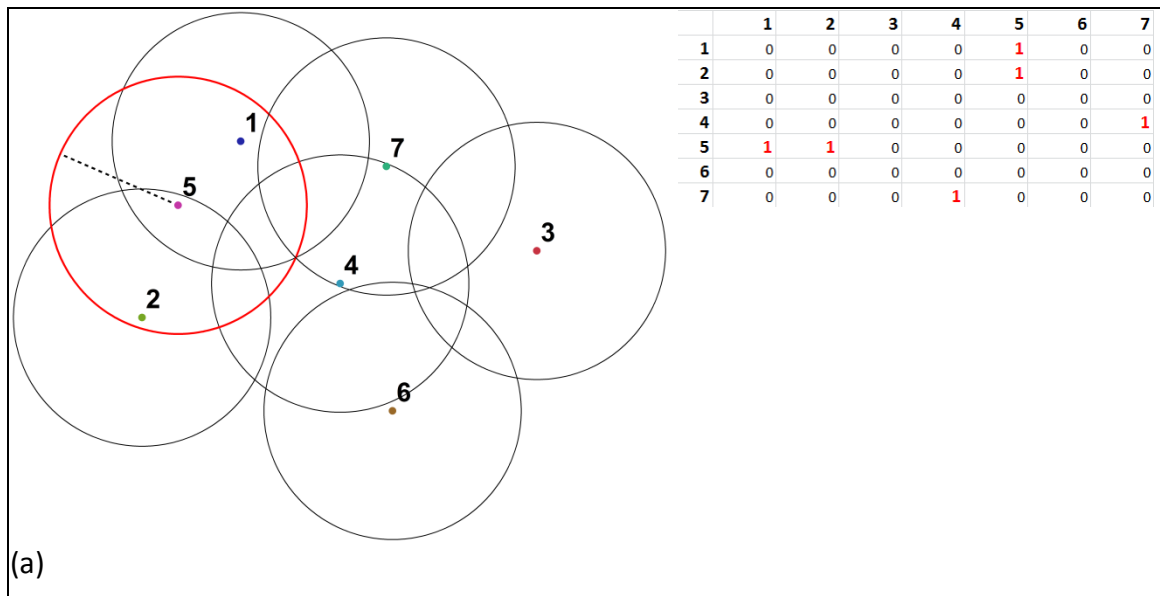
2.2 Objective

The goal of this research is to propose methods for constraining spatial neighbourhoods in marine environments. To meet this goal we describe common spatial neighbourhood definitions and outline two approaches for constraining neighbourhoods in marine environments, the first incorporating land as constraint and the second current

directions. Following, we demonstrate the benefit of our approaches on two case studies using data on marine mammals' abundance and oil spill occurrence.

2.3 Context: Spatial neighbourhoods in marine applications

When conducting spatial analysis, the level of influence or interaction of geographic features is defined by the user as a spatial neighbourhood and often operates behind the scene as a roving window. The choice of a spatial neighbourhood is somewhat subjective, but should be carefully considered as the inappropriate selection can generate misleading analysis results (Nelson & Robertson, 2012).



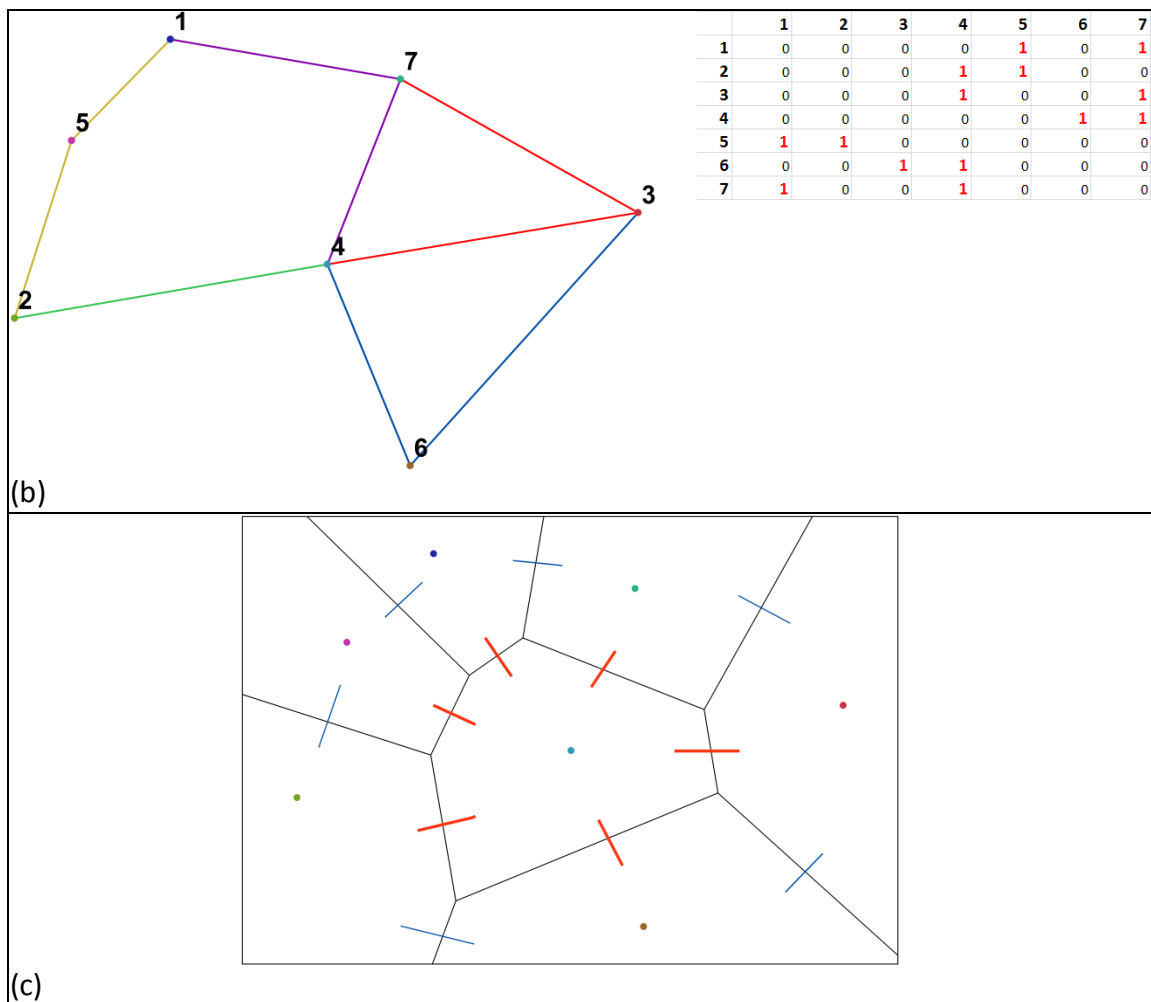


Figure 2-1. Common spatial neighbourhood delineations (a) fixed distance; (b) K-nearest; (c) adjacency

Marine studies often consider the influence of space as a fixed distance to delineate spatial neighbourhoods, ignoring barriers, such as coastlines or current directions. For example, spatial neighbourhood delineations by fixed distance have been used in different types of spatial analysis; to analyze the spatial patterns of groundfish distribution (Wigand et al., 2013) and to monitor the change in coral reef structures across space (LeDrew et al., 2004). Ultimately, the choice and geographic extent of the neighbourhood, which defines spatial relationships between objects, is determined subjectively by the analyst. Typically, methods employed are simplistic in terms of

implementing real world constraints on species movements and environmental processes. Types of neighbourhoods include: fixed and adaptive distance bands, k-nearest neighbour, and contiguity between areal-units (Figure 2-1). Some neighbourhood delineations are binary while others rely on a weighting scheme for valuing data importance on the region of interest for local analysis. The most common spatial neighbourhood delineation used in marine studies is fixed distance, which represents the region of space (delineated by a buffer of a given distance) associated around an object under study (O'Sullivan & Unwin, 2002). Distance neighbourhoods have been used to explore the relationship between seagrass, fish and nutrients in temperate seagrass systems (White et al., 2011) and characterize change in coral reefs (LeDrew et al., 2004).

K-nearest neighbours is also a standard approach for neighbourhood creation that delineates neighbours based on a constant number (k) of locations in each neighbourhood (Nelson & Robertson, 2012). Marine studies have employed this method to detect indicators for the assessment of marine environmental conditions using methods such as inverse distance weighted (IDW) (Chang et al., 2006).

Spatial neighbourhoods may also be delineated by adjacency or contiguity. In this approach, the influence of neighbouring features is assumed to occur only when they share a common boundary (O'Sullivan & Unwin, 2002). One example of a marine study that delineated spatial neighbourhood by adjacency is Chandrasekharan et al. (2008), who conducted a spatial-temporal analysis of sample data to understand the suitability of land for agriculture and the reclamation period of the Tsunami of 2004 affected coastal areas of Nagapattinam district of Tamilnadu state in India.

Spatial neighbourhoods delineated by spatial dependency or interaction, assume that locations in close proximity tend to have more similar attributes than locations further apart, similar to Tobler's first law of geography (Fotheringham, 2009). Spatial dependence delineations of spatial neighbourhoods have an inherent weighting factor, typically based on a semivariogram (Myers, 1991; O'Sullivan & Unwin, 2002). Some examples of spatial neighbourhoods delineated by spatial dependency are spatial distribution analysis of reef-associated fish species using stratified sampling procedure (Aguilar-Perera & Appeldoorn, 2008) and spatial interaction of predator-prey between oceanic birds and their prey (Russell et al., 1992).

Spatial neighbourhoods can be categorized as binary or weighted (Aldstadt & Getis, 2006; O'Sullivan & Unwin, 2002). Binary neighbourhoods are the most common and simple to implement as they do not require previous knowledge of the process involving delineation of neighbours, that means, entities are in or out of the neighbourhood (if distance definition is used). Likewise, if definition by adjacency or contiguity is used as binary, the binary selection is whether entities are next to a neighbourhood or not (O'Sullivan & Unwin, 2002). An example of the use of binary neighbourhoods is the identification of the variability of abundance estimate of pelagic fish stocks derived by acoustic surveys (Marchal & Petitgas, 1993). This study characterized the spatial variability of pelagic biomass by separating fish density into two components, the number of schools per sea surface unit and the biomass in the schools. Weighted neighbourhoods assign "weights" to neighbours based on proximity, that is, neighbourhoods that are close to the location of observation will receive more weight than those further apart (Nelson & Robertson, 2012). An investigation of spatial distribution of

reef-associated fish species using a stratified sampling procedure is one example (Aguilar-Perera & Appeldoorn, 2008) of a study using weighted neighbourhoods.

Another aspect of all spatial neighbourhoods is that they can be fixed or adaptive. Fixed neighbourhoods are applied consistently across study areas (Fotheringham, 2009; Nelson & Robertson, 2012). Adaptive neighbourhoods allow the neighbourhood definition to vary over the study area based on data clusters (Nelson & Robertson, 2012). Chang et al., (2006) used fixed distance spatial neighbourhood to detect indicators for the assessment of marine environmental conditions using spatial analysis methods. Adaptive neighbourhoods have been used, for example, to quantify the spatial extent of the summer runoff–seawater mixing zone of the Great Barrier Reef (GBR) lagoon (Wooldridge et al., 2006). Identification of spatial distribution patterns and determination of degree of trace element contamination in the estuarine (Magesh et al., 2011) is another example of a study using an adaptive neighbourhood.

2.4 COSINE Tool

Standard approaches to delineate spatial neighbourhoods are not appropriate to identify relationships among observations as the geographic or topographic component is missing. Current GIS software do not present solution to this issue, even though analyses such as, network analysis or spatial interpolations have this option. We present COSINE, a software to constrain existing neighbourhood delineations using ancillary data on coastlines and current directions.

Description

COSINE was developed within open source software (R) via a graphical user interface (GUI). The software enables users to select a standard delineation of neighbourhood: fixed distance, inverse distance (with power 1 or 2) and k-nearest neighbourhood (Knn) and apply constraints based on an input coastline polygon and/or a raster grid representing direction of flow and a user defined angle. Modified spatial neighbourhoods can be output in either binary (default) or row-standardized weight matrix format, which is the final output used to perform spatial analysis. For instance, users can use the final output (spatial weighting matrix) to perform local statistical analysis.

Creating the graphical user interface (GUI)

The initial step was to create the GUI to allow the user interact with the software. Using open source software (R), users may change the location of the neighbourhood type function. The GUI has three sections: Input/Output, Direction and Neighbourhood type that will be explained in more detail later in this section.

Defining standard spatial neighbourhoods

Standard neighbourhoods are delineated by creating connections among points (which could be a polygon centroid) (Figure 2-2). Here, we term connections between points “links”. A unique identifier is assigned to each link, and the distance of each link is calculated and saved in a geodatabase. A binary value (i.e., one) is given to links that are within a threshold distance given by the user.

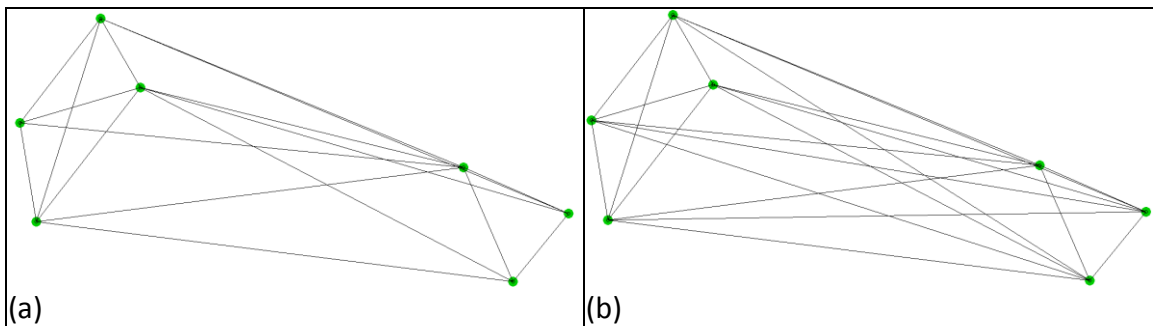


Figure 2-2. Unconstrained spatial neighbourhood delineation for: (a) fixed distance; (b) K-nearest neighbourhood

Including constraining factors in the spatial neighbourhood delineation

To account for the influence of coastlines on spatial interaction, features separated by land are excluded from sharing a neighbourhood. Neighbourhoods are defined first and a spatial statistic used to quantify the interaction within the neighbourhood. To delineate neighbourhoods with consideration of the barriers imposed by land, all links are intersected with land and a binary value (i.e., one) is given to only those links that do not intersect land, hence, they can be considered neighbours (Figure 2-3).

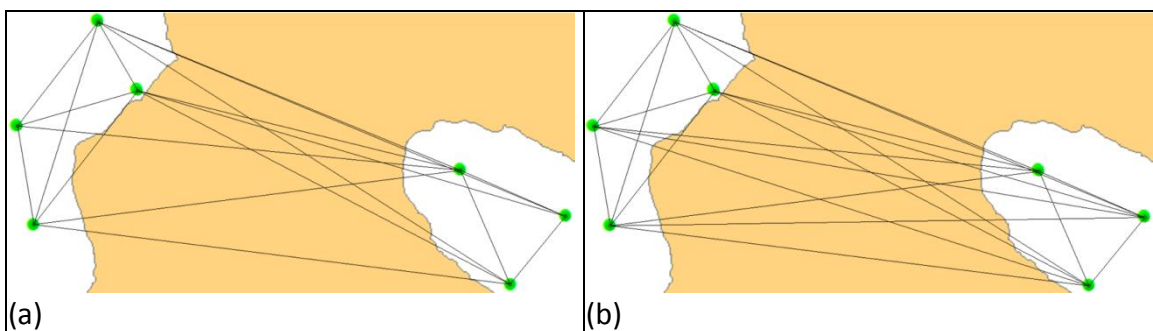


Figure 2-3. Unconstrained spatial neighbourhood delineation intersecting barriers -land- (White background represents water and colour represents land) for: (a) fixed distance; (b) K-nearest neighbourhood

Following, a directional relationship (optional) is calculated. If the user chooses to use a directionality approach, a new binary value (i.e., one) is assigned to the angles that are less than the threshold angle given by the user. Ancillary data (Meridional and Zonal

ocean currents) are used in the directionality constrain: Meridional currents provide information about sea water velocity (m/s) in a north/south direction and Zonal currents provide information about sea water velocity (m/s) as well, but in an east/west direction, whenever there is a negative value for Meridional or Zonal currents, means there is a change in direction to the South or West, respectively. Meridional or Zonal has the same implication for terrestrial environments, and can be applied, for example, to wind directions. For instance, studying the impact of pollution over locations with significant importance, like schools or hospitals.

To explain how the directionality approach is used in COSINE software, we will assign a letter (v) to Meridional currents and letter (u) to Zonal currents. Whenever each observation is overlapped on raster currents file the cell values for v and u are extracted (and saved in the geodatabase). For each observation point, a node is created x or y (x_u, x_v or y_u, y_v), called node at each point at the end of the link, and by implementing the Cauchy-Schwarz inequality (Kheruntsyan et al., 2012; Li & Heap, 2008; Masjed-Jamei, 2009):

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\| \quad (1)$$

the vector divergence is calculated in order to identify if the two nodes are going in the same direction (or parallel in a geometrical sense). The equality happens when $v = 0$. As the norm for each node is positive, it is possible to use it as follows:

$$0 \leq \frac{|\langle x, y \rangle|}{\|x\| \cdot \|y\|} \leq 1 \quad (2)$$

Or

$$\frac{x_u y_u + x_v y_v}{\sqrt{x_u^2 + x_v^2} \times \sqrt{y_u^2 + y_v^2}} \quad (3)$$

The interpretation of equation 2 in R in a geographic space is as follows:

$$\frac{x_u y_u + x_v y_v}{|x_u, x_v| \times |y_u, y_v|} \quad (4)$$

To illustrate the directionality approach, a theoretical example could be that two observations are considered neighbours if both are within a specific angle assigned by the researcher, meaning both observations could be heading to the same direction or could be parallel. Figure 2-4 shows the theoretical example displaying all components required for calculating the Cauchy-Schwarz inequality: Link_id = '136' represents the connection between Node 1 (or Point 1) = '16' and Node 2 (or Point 2) = '17', each node has information about meridional and zonal currents. Node 1 has direction north-west and Node 2 south-west, and by calculating the Cauchy-Schwarz inequality, it generates an angle divergence of 41°. The angle divergence for each link is then used to identify if it is under the threshold value for the angle provided by the user.

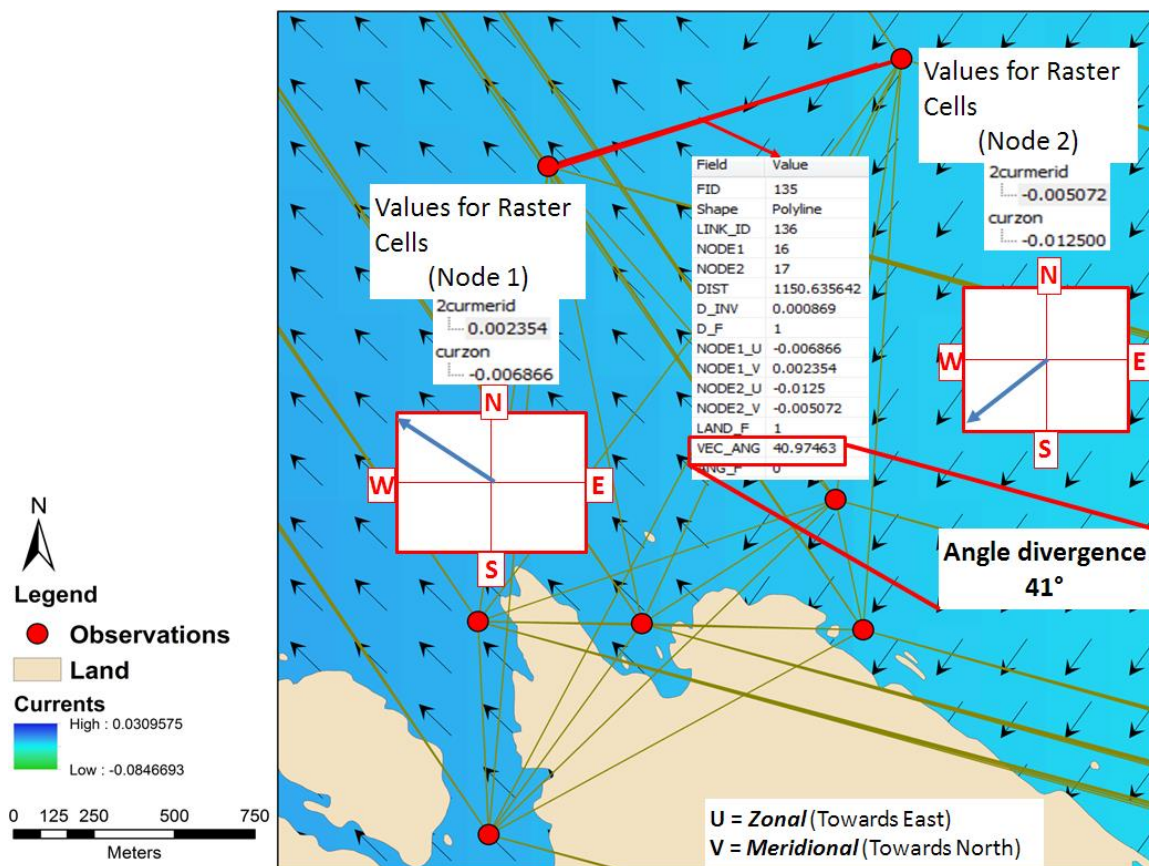


Figure 2-4. Direction of nodes to calculate angle divergence

Figure 2-5 displays another theoretical example where direction for both nodes is similar, with the same Link_id but for the purpose of this example, the direction of meridional and zonal currents for Node 1 and Node 2 are south-west. The Cauchy-Schwarz inequality calculates an angle divergence of 4°.

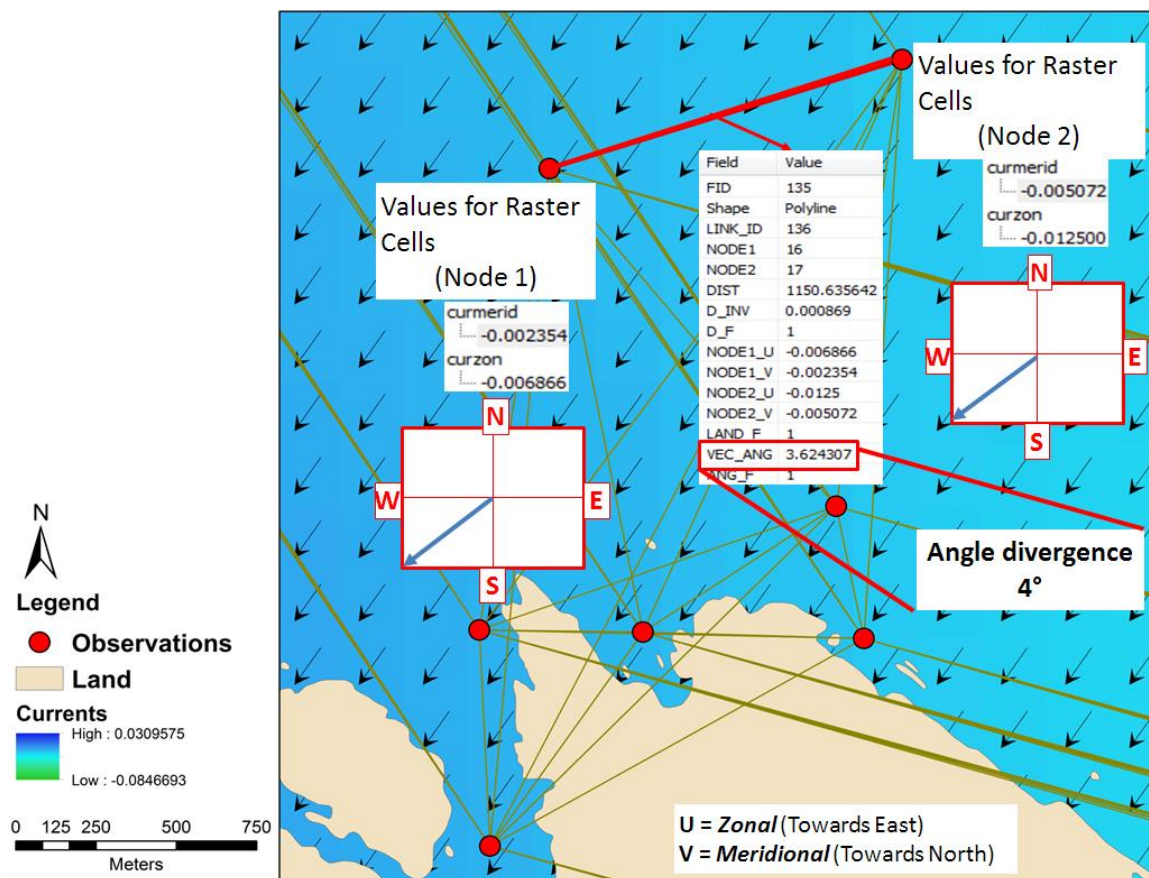


Figure 2-5. Direction of nodes to calculate angle divergence

Output for spatial analysis

The last step of the process is to create a Boolean selection, where only the links that satisfy all requirements given by the user to delineate a spatial neighbourhood (i.e., binary value of one) are selected (Figure 2-6). Depending on the selection made by the user, whether selecting the directionality approach or not, the output file (weight matrix) may be either in binary or row-standardized format.

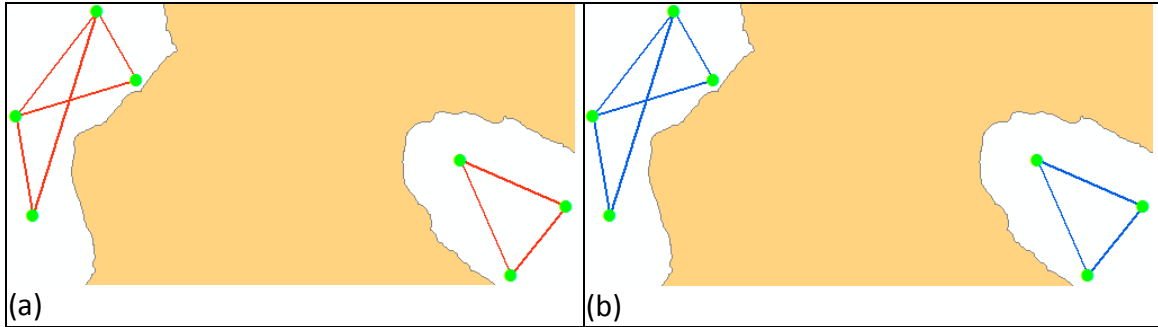


Figure 2-6. Constrained spatial neighbourhood delineation (White background represents water and colour represents land) for: (a) fixed distance; (b) K-nearest neighbourhood

Implementation

The output for COSINE can be used to perform spatial analysis using ArcGIS or OpenGeoDa. The source code will be placed on our website

(<http://www.geog.uvic.ca/spar/>). Here are some useful steps to work with COSINE:

- An OBJECTID column is required, if it does not exist, the user needs to create an OBJECTID column, with LONG (Data type), if it already exists, it is appropriate to delete current OBJECTID and create a new one.
- Calculate new OBJECTID column as follow: [FID] +1

To work with ArcGIS with the spatial weights matrix:

1. Open the .DBF file (in Excel) from the newly create shape file
2. From the first row, Delete: NID and WEIGHT column names, **leave only** OBJECTID
3. Select "Save As type" and choose Other Formats. From the Other Formats choices, select Formatted Text (Space delimited) (*.prn).
4. Choose a location for the file and click SAVE.
5. If a warning message is displayed, click YES

6. In ArcGIS, open your analysis method, and Under Conceptualization of Spatial Relationships, select "GET_SPATIAL_WEIGHTS_FROM_FILE"

7. Under "Weights Matrix File (optional)" parameter, locate and select the file with (.prn) extension you have created in steps 1-5

To work with OpenGeoDa using the spatial weights matrix, note that OpenGeoDa structure for spatial weight matrix requires having in the first line of the matrix the following structure:

- "0" "Total Number of Features" "Source Feature Class (Name of the point shapefile)" "Unique ID Field Name (OBJECTID)"

Example: 0 131 Oil_Spill OBJECTID

2.4.1 COSINE Findings

In order to implement the proposed neighbourhood delineation approach, this research required the development of a customized GUI and subsequently test by using two case studies.

Graphical user interface (GUI)

The GUI has three sections: Input/Output, Direction and Neighbourhood type (Figure 2-7). Input/output section is where the user has the option to select the data inputs: points that neighbourhoods are delineated for and a polygon or polyline to define the land or barriers. The output file name will create two outputs. First, a feature (line) file will be the neighbourhood file, and is the line feature class containing the weights in tabular form (dbf file). Second, a feature (line) file with all data described in the methods section, such as, ID of each node, Link ID, Link measure distance, Meridional and Zonal

values of each node, and binary value for Land constrain, spatial neighbourhood selection (Distance, Inverse Distance, K-nearest neighbourhood) and angle constraint. The shapefile with all data is useful to identify and troubleshoot all the parameters specified and acts as an informative feature for the user. As spatial analysis in marine environment is different to terrestrial environment, the input/output section has an additional component that permits the user to select whether the analysis is in terrestrial or marine (default) environments. For example, identification of the predominant food source for grizzly bears (*Ursus arctos horribilis*) could determine the habitat range, hence, areas with no food sources can be considered constraints (Rode et al., 2001). A similar example is the identification of ecological constraints (i.e., water) to identify demographic characteristics of the caribou (*Rangifer tarandus*) (Mallory & Hillis, 1998).

Directionality (currents) section allows the user to specify the path to Meridional or Zonal raster files, as well as the maximum angle that two observations can be apart, taking into consideration velocity and direction in the raster file.

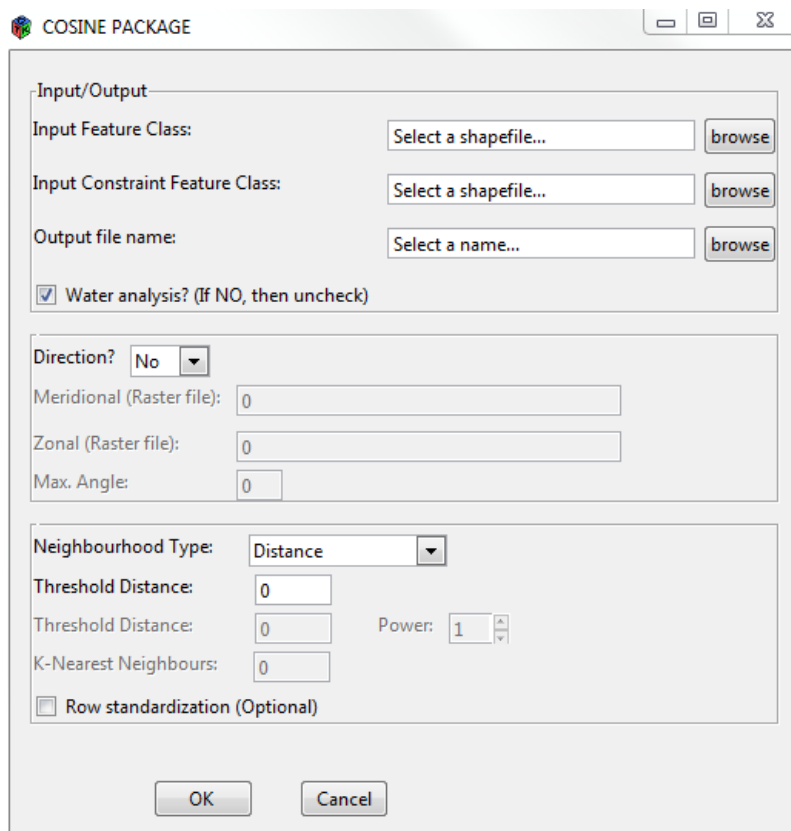


Figure 2-7. Graphical user interface (GUI)

2.5 Case Study

In this study, we use the measure of spatial autocorrelation to describe the spatial pattern of phenomena. Spatial autocorrelation can be defined as the notion that “all things are related and near things more than far” (Tobler 1970). Quantitative measures of spatial autocorrelation are used to characterize how similar or dissimilar nearby events are. When nearby events are similar the spatial pattern is considered to have been generated from a clustering process. Clustered events that have high values are often termed hot spots; whereas, clustered events with low values are termed cold spots. Measures of spatial autocorrelation can be evaluated statistically to determine if the level of clustering, or hot spot, is statistically unexpected, given a null hypothesis of randomness (O’Sullivan & Unwin, 2002).

To demonstrate the benefit of COSINE, we introduced spatial analysis on two marine applications. More specifically, we detected clusters of high or low values using local measures of autocorrelation, in this case local Getis-Ord G_i^* statistics. The goal of this case study is to determine how different neighborhood delineations impact detection of clusters of high or low values using local Getis-Ord G_i^* statistics. To achieve this goal, we perform local G_i^* statistics on marine mammals' abundance and oil spill occurrence data and compared results obtained using unconstrained and constrained neighbourhood delineations.

Marine mammals distribution and abundance

Spatial distribution and abundance data of marine mammals are crucial to identify locations of high primary productivity (Moore & DeMaster, 1997; Whitehead et al., 2010). Such information is essential to wildlife conservation and management, as it can be used to recognise different factors that can affect the normal distribution and abundance of marine mammals, such as incidental by-catch, traffic vessels collisions or pollution (Kaschner et al., 2006; Williams & Thomas, 2007).

Oil spills occurrence

Illegal oil discharge is a major source of contamination in marine environments and a major factor in the mortality of seabirds and marine mammals and their impact can last for a long period of time (Hooker & Gerber, 2004; O'Hara & Morgan, 2006; O'Hara, et al., 2013; Serra-Sogas et al., 2008). Most of oil discharges made by marine vessels are the result of accidental or intentional dumping during normal operations (O'Hara & Morgan, 2006). Worldwide increasing awareness has been given to the fact that small scale vessel-

source oil spills have greater impact than larger scale spills in marine environments (Serra-Sogas et al., 2008).

Study Area

The study area is located in the inshore western Canadian waters between BC - Washington and BC-Alaska borders (Figure 2-8). BC's coast is surrounded by small islands, narrow straits, and channels, with a coastline of approximate 1,600 km, the mainland coast is characterized by deep indentations from fjords, with smooth indentations on west side of Vancouver Island (Farley, 2011). The majority of the terrain is composed of rocky shorelines with narrow, deep inlets (Johannessen et al., 2007). Most of the climate in BC is associated with El Niño-Southern Oscillation (ENSO), and the mild winter climate is a product of constant coastal storms and low-pressure systems with strong winds during the season (Johannessen et al., 2007). During summer months subtropical high-pressure systems dominates the region with mild winds coming from the northwest (Walker & Sydneysmith, 2008). The average annual precipitation for coastal BC is greater than 1000 mm (Daly et al., 2002). Average sea surface temperature is 9-15°C during summer months and 6-9°C during winter months (Crawford, 2001). Tides are primarily semi-diurnal and move in a counter-clockwise direction, marine currents are driven by wind conditions, freshwater input and topographic terrain (Johannessen et al., 2007).

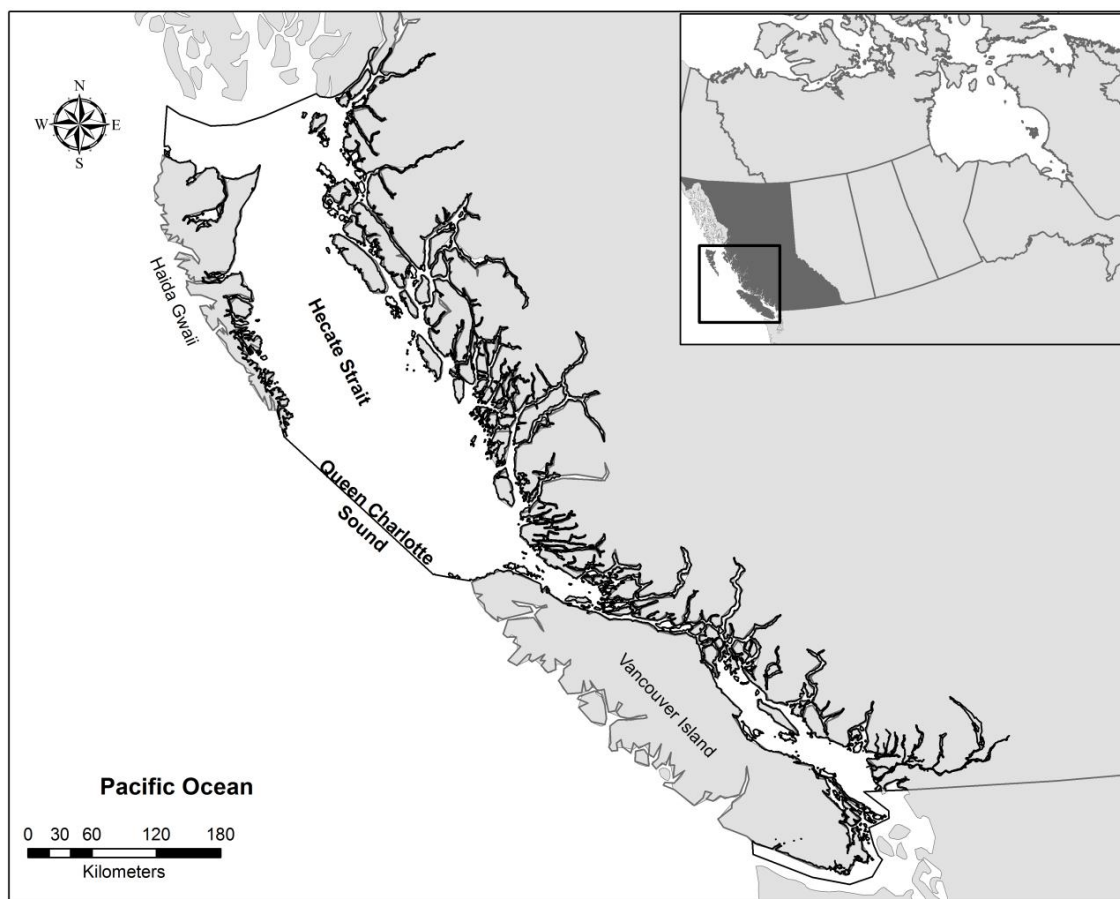


Figure 2-8. Extent of the study area

Data description and preparation

Data for this research are divided into two case studies: marine mammals' abundance and oil spill occurrence.

Marine mammals data

Marine mammal data were acquired using surveys to estimate the abundance of marine mammals' species in the inner waters of coastal BC during the summer months of 2004 (Table 2-1). Marine mammals were sighted along line transects divided into four strata. Each strata is a geographic area that was created for survey sampling to ensure all dominant geographic and environmental conditions were included in abundance

estimates (Best & Halpin, 2009). Each strata was sampled using equal-spacing along either in a zig-zag or parallel formation of transect lines with a random starting point to ensure equal coverage probability within strata ((Williams & Thomas, 2007), see (Thomas et al., 2007) for description of the survey). Following the transect lines, using binoculars and an angle board mounted on the deck railing of the boat, the radial angle to sighted animal location was measured. A unique identifier was assigned to each marine mammal sighting with additional attributes such as distance, angle to the mammal, bathymetry, and sea conditions also recorded. To generate abundance data for marine mammals' sightings, observation data were aggregated in a polygon grid of 4km by 4km cell size using Spatial Join in ArcGIS 10 (ESRI®). Though grid cell is typically selected, the 4km by 4km cell size retained the detail of the pattern without resulting in multiple grids with only one point. Grid cell sizes that are too large lead to over-smoothing and loss of pattern (Bailey & Gatrell, 1995). Figure 2-9 shows the location of marine mammals sightings made in 2004.

Table 2-1. Total number of individuals, 2004

<i>Species</i>	<i>Number of Individuals</i>
Humpback whale (<i>Megaptera novaeangliae</i>)	42
Minke whale (<i>Balaenoptera acutorostrata scammonii</i>)	8
Fin whale (<i>Balaenoptera physalus</i>)	7
Gray whale (<i>Eschrichtius robustus</i>)	3
Harbour porpoise (<i>Phocoena phocoena</i>)	46
Dall's porpoise (<i>Phocoenoides dalli</i>)	77
Killer whale (<i>Orcinus orca</i>)	15
Pacific white-sided dolphin (<i>Lagenorhynchus obliquidens</i>)	70
Harbour seal (<i>Phoca vitulina richardsi</i>)	318
Steller sea lion (<i>Eumetopias jubatus</i>)	13
Northern Elephant seal (<i>Mirounga angustirostris</i>)	5
<i>Total Species Per Year</i>	604

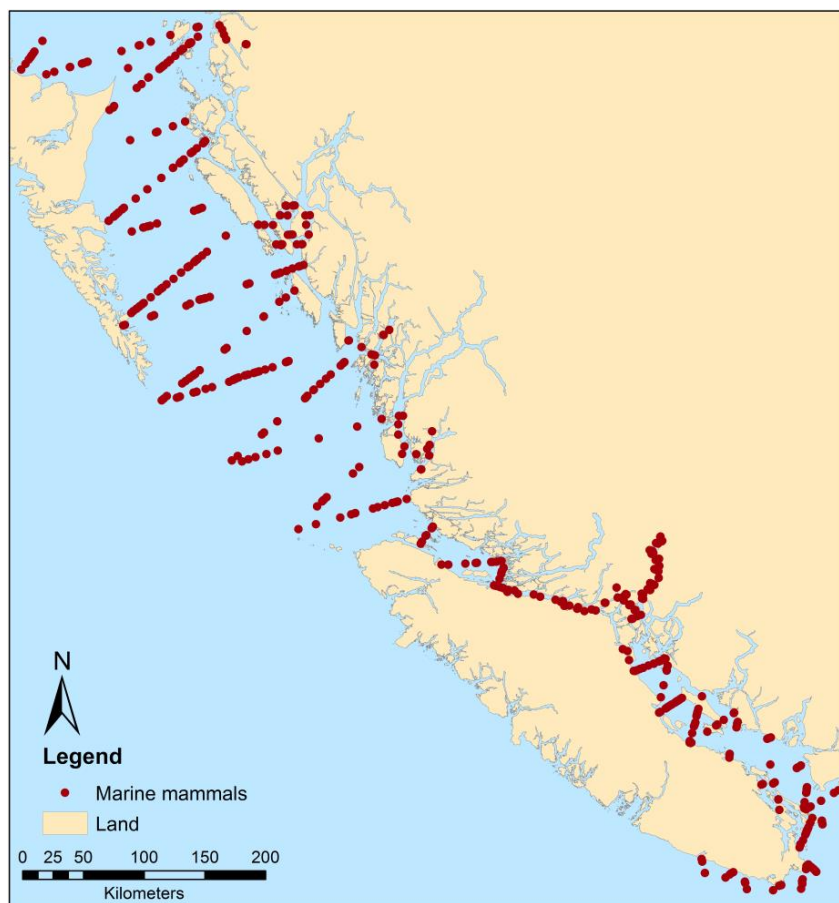


Figure 2-9. Marine mammals sightings, 2004

Oil spills from vessels

Oil spill data were provided by National Aerial Surveillance Program (NASP), operated by Transport Canada, which patrols on a regular basis the Canadian Exclusive Economic Zones (EEZs), in this case, the Pacific EEZ as part of mechanism to monitor oil spill pollution based on MARPOL conventions (Serra-Sogas et al., 2008). Oil spills detected by NASP flight crew were collected on a visual search using global position system (GPS). Time and position were assigned for each oil spill detected and the volume was calculated by remote sensors, such as synthetic aperture radar (SAR) that allowed monitoring during day and night. For the purpose of this case study, only oil spill

observations from 2008 will be considered since there are the largest number of observations for that year (124). Figure 2-10 shows the location of oil spill observations made in 2008.

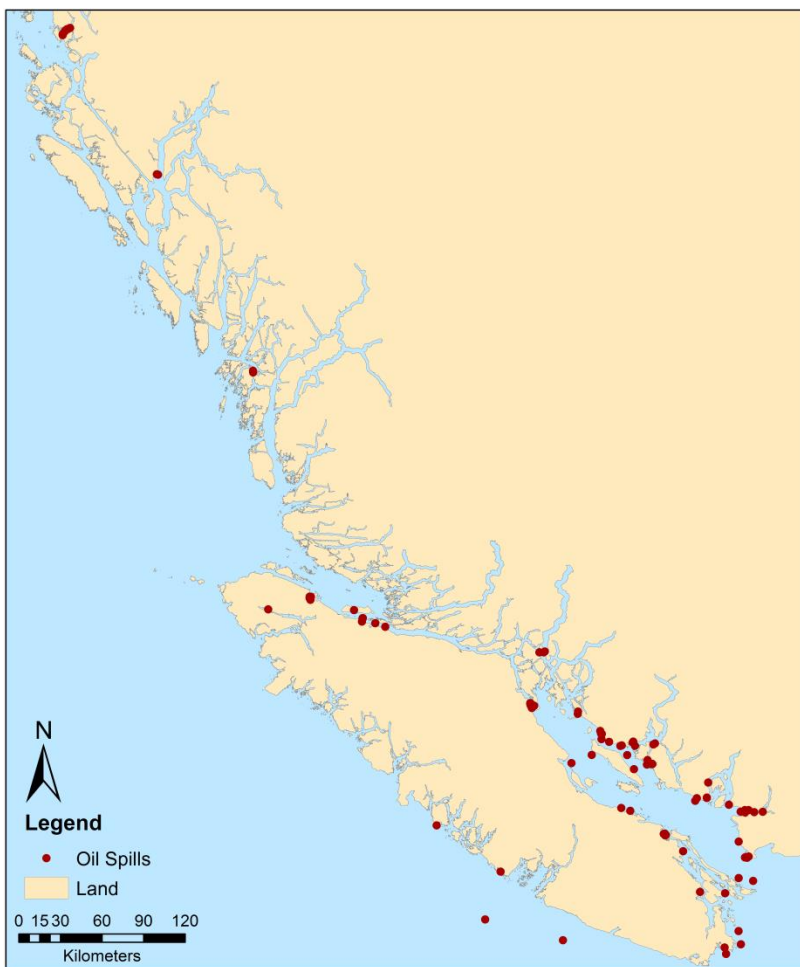


Figure 2-10. Oil spill observations, 2008

Currents data

Environmental data were composed of dynamic data for ocean currents (Meridional and Zonal). Currents are a method of dispersion for ocean pollutants such as oil spills (Samuels et al., 2013). Meridional currents data provide information about sea water

velocity (m/s) in a north/south direction (Bonjean & Lagerloef, 2002). Larger values in magnitude imply stronger (positive) equatorial divergence. Positive values correspond to currents moving towards north. Negative values of velocity represent currents going in a south direction. Similar to Meridional currents, Zonal currents, are geostrophic currents which are the result of the pressure gradient and the Coriolis force (Hwang et al., 2008). Data corresponding to Zonal currents explain (in a diagnostic sense) the east/west deflection due to the equatorial undercurrent (EUC) and the south equatorial current (SEC) in m/s (Bonjean & Lagerloef, 2002). Positive values indicate east current directions, while negative values represent currents moving towards west.

Current data were derived from the AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data) program with a 0.25x0.25 degrees spatial resolution and a temporal resolution of one month (average). To account for missing data due to cloud cover or proximity to shore on a daily basis, seasonal values of each current were extracted, interpolated and rasterized using ArcGIS version 10 (ESRI®). The cell value nearest to the centroid of each cell was used to extract the values of each current. For the purpose of this case study, only currents data for 2008 will be used.

2.5.1 Methods

Using local measures of spatial autocorrelation, we quantified and mapped geographic variation in the association of nearby data and identified clusters of data with values that are similar and extreme relative to the mean of cell values (Fortin & Dale, 2005; Getis & Ord, 1992). G_i^* is a measure of local spatial autocorrelation that evaluates individual features within a neighbourhood and enables detection of local concentrations of high or low values in an attribute (O'Sullivan & Unwin, 2010). A feature with a high

value surrounded by other high values is identified as a hot spot (Getis & Ord, 1992; O'Sullivan & Unwin, 2010). It can be calculated as:

$$G_i^* = \frac{\sum_j w_{ij}x_j}{\sum_{j=1}^n x_j} \quad (5)$$

Where w_{ij} is a symmetric spatial neighbourhood with ones (1) for all links or spatial neighbours defined as being within distance of a given i , the rest of the links are zero (0) including the link of point i to itself. $\sum_j w_{ij}x_j$ is the sum of all x_j within a distance of i but not including x_i . $\sum_{j=1}^n x_j$ is the sum of all x_j not including x_i (Getis & Ord, 1992). G_i^* generates output with a Z ($G_i^* Z$) score for each feature; the Z-score represents the statistical significance of clustering for a given distance. High or low Z-score indicates that the neighbours have high or low attribute values, which we define as clusters of high or low densities of marine mammals' abundance or oil spill occurrence, respectively. Clusters of high and low values were calculated with a 90% confidence interval, which means Z-score values greater than 1.65 or less than -1.65. We present the number of neighbourhoods detected for each distance range as well as the Z-score in a boxplot to show the ranges of $G_i^* Z$ -score values.

Clusters of high and low values were detected using G_i^* with neighbourhoods delineated with and without the use of COSINE software. Marine mammals' abundance data were analyzed using two options: i) standard spatial neighbourhoods (unconstrained), and ii) land constrained neighbourhood (constrained land). Oil spill occurrence data were analyzed using three options: i) standard spatial neighbourhoods (unconstrained), ii) land constrained neighbourhoods (constrained land) and iii) land and directionality (ocean currents) constrained neighbourhoods (constrained dir).

To assess the impact of contextualizing spatial neighbourhood delineations, we compared the nature of neighbourhoods and detection of clusters of high or low values. Each case study has different options to evaluate and compare unconstrained and constrained approaches. First, we assessed how the number of neighbours delineated using unconstrained and constrained neighbourhoods impacted the results. Second, using G_i^* statistic to detect locations with clusters of values that are either high or low relative to the mean (\bar{x}) value, we quantified how that detection was impacted with or without a constrained approach. Marine mammals' abundance data were used to demonstrate the influence of land constraints on neighbourhood delineations. Oil spill volume data were used to demonstrate the impact of constraining neighbourhoods by both land and directionality (ocean currents).

Identification of the number of neighbours detected per neighbourhood delineation and the G_i^* Z-score to detect spatial clusters of high or low values for marine mammals' abundance and oil spills data were delineated using binary format and distance neighbourhood type. To analyze the impact that different distances selection has over a study area, we selected a range of distances from 40 km to 120 km, in order to compare the impact of the scale in neighbourhood delineation for unconstrained, constrained and constrained with directionality (for oil spills) approaches. The following sections highlight the results, separated into marine mammals' abundance and oil spills occurrence information.

2.5.2 Results

Marine mammals abundance

Using an unconstrained approach for delineating spatial neighbourhoods, the number of neighbours in each neighbourhood increases as lag distances increase; for instance, minimum and maximum values for a distance of 40 km are two to 29 compared to 22 to 87 for a 120 km distance. For a constrained land approach, the range for number of neighbours also increases, but less so, from one to 15 (40 km) to one to 58 (120 km) (Figure 2-11). Median values for the number of neighbours ranges from 13 in an unconstrained approach, compared to seven in a constrained land approach for a 40 km distance. An increase in median values appears in a 120 km distance, with 56 median values for unconstrained and 16 for constrained approach (Figure 2-11). There is a reduction between 27% (40 km) and 33% (120 km) in the number of neighbours for constrained land to unconstrained approach. The percentage is calculated taking into account the unconstrained approach as being the 100% for all distances.

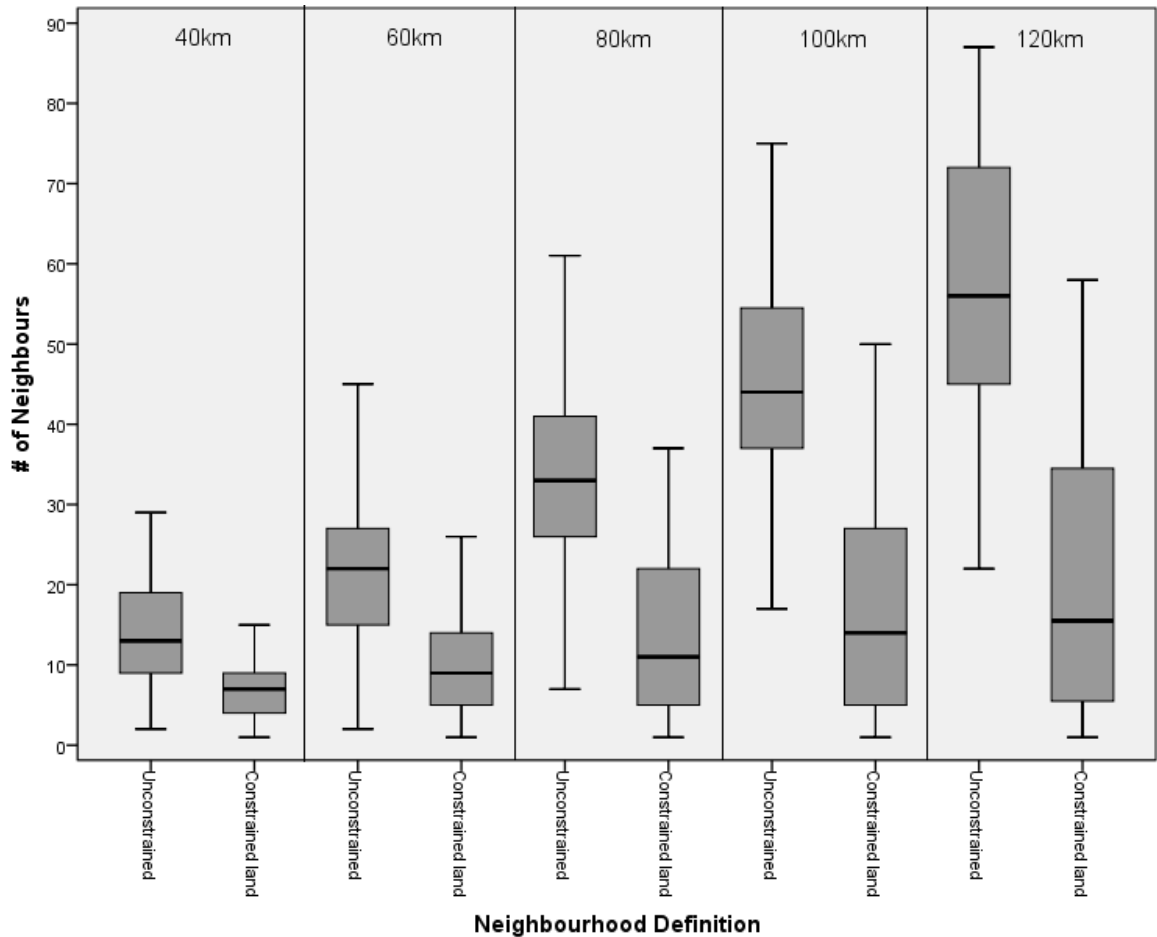


Figure 2-11. Marine mammals abundance - Number of neighbourhood per neighbourhood delineation. 40 km – 120 km distance

Ranges in minimum and maximum values for G_i^* Z-score variation can be seen in Figure 2-12. For a lag distance of 40 km, the ranges are from -1.77 to 3.75 and for a constrained land approach the minimum value is -1.65 and the maximum 2.19. For a lag of 120 km the minimum value for an unconstrained approach is -2.16 and maximum value is 3.25, for a constrained land approach the range is from -2.59 to 2.78 respectively (Figure 2-12 and 2-15).

When neighbourhoods are delineated using distance of 40 km, 60 km, 80 km, 100 km, and 120 km, spatial clusters range from 46 (40 km) to 114 (120 km) for clusters of

high values (positive autocorrelation) and 3 (40 km) to 61 (120 km) for clusters of low values in an unconstrained approach. For a constrained land approach the values remain stable in 23 clusters of high value with a minimum increase in the range of 60 km and 80 km, and for clusters of low value the ranges are from 1 (40 km) to 34 (120 km) (Table 2-2).

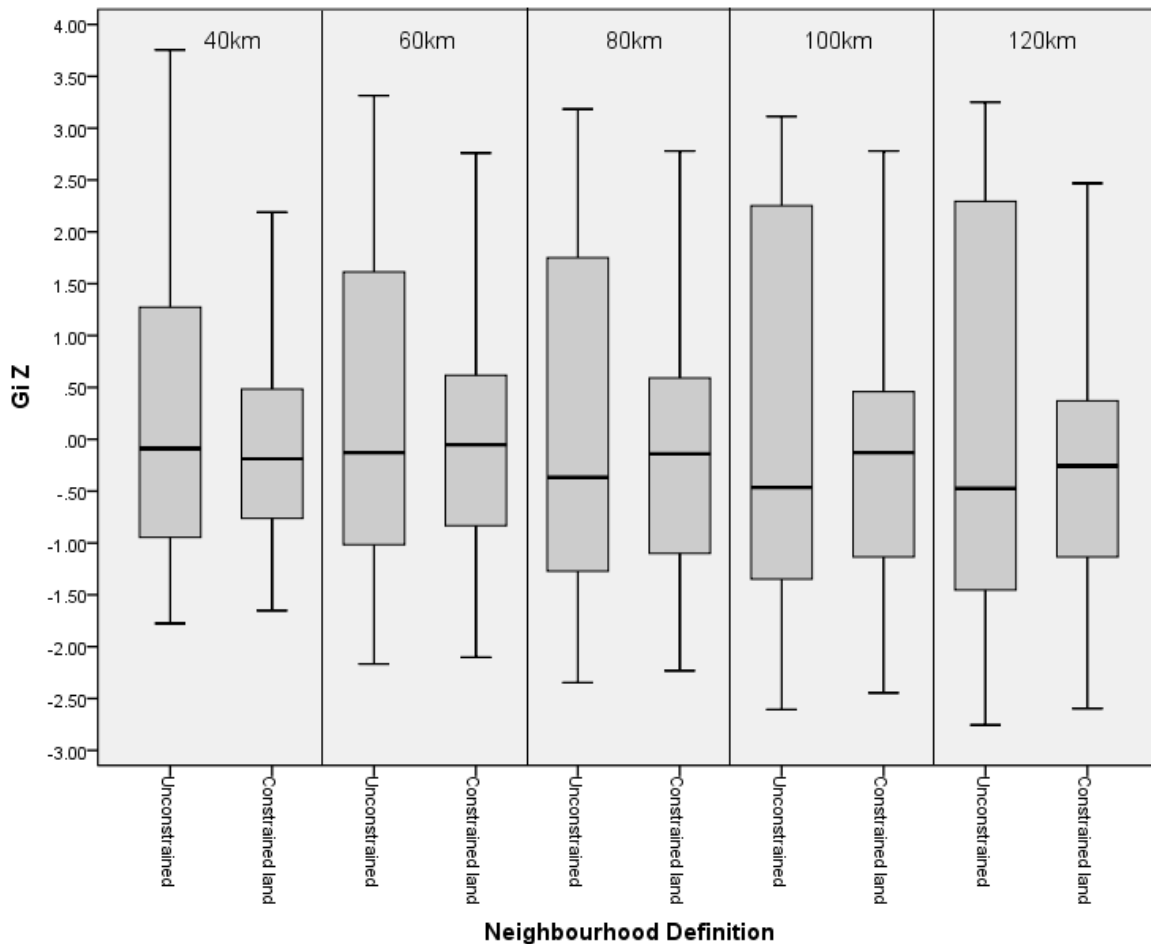


Figure 2-12. Marine mammals abundance - G_i^* Z-score per neighbourhood delineation. 40 km – 120 km distance

Table 2-2. Number of high and low clusters for oil spills occurrence and marine mammals abundance

Lag (Km)	G _i * Z-score (90% confidence)	Oil spills			Marine mammals	
		Unconstrained	Constrained land	Constrained dir	Unconstrained	Constrained land
40	High	6	7	4	46	23
	Low	18	0	0	3	1
60	High	2	7	4	68	24
	Low	4	0	0	22	8
80	High	0	5	4	82	24
	Low	2	0	0	43	30
100	High	6	5	4	104	23
	Low	0	0	0	54	27
120	High	2	5	4	114	23
	Low	1	0	0	61	34

Oil spills occurrence

There is variation in the number of neighbour counts for an unconstrained and a constrained (constrained land and constrained dir) approaches (Figure 2-13). Ranges in minimum and maximum values in the number of neighbours vary according to the distance chosen, for instance, for an unconstrained approach the values are from one to 35 in 40 km to one to 81 in 120 km. For a constrained land approach in a 40 km there are one to 14 neighbours and one to 19 in a 120 km. In a constrained dir approach, the values remained stables with 18 neighbour counts across all scales. Variation in the number of neighbourhoods per neighbourhood type can be seen in Figure 2-9. Trends in median value vary for an unconstrained approach but remain stable for constrained approaches. 15 to 54 median values in the number of neighbours for 40 km to 120 km distance and a median value of four neighbours for constrained with and without directionality (Figure 2-13).

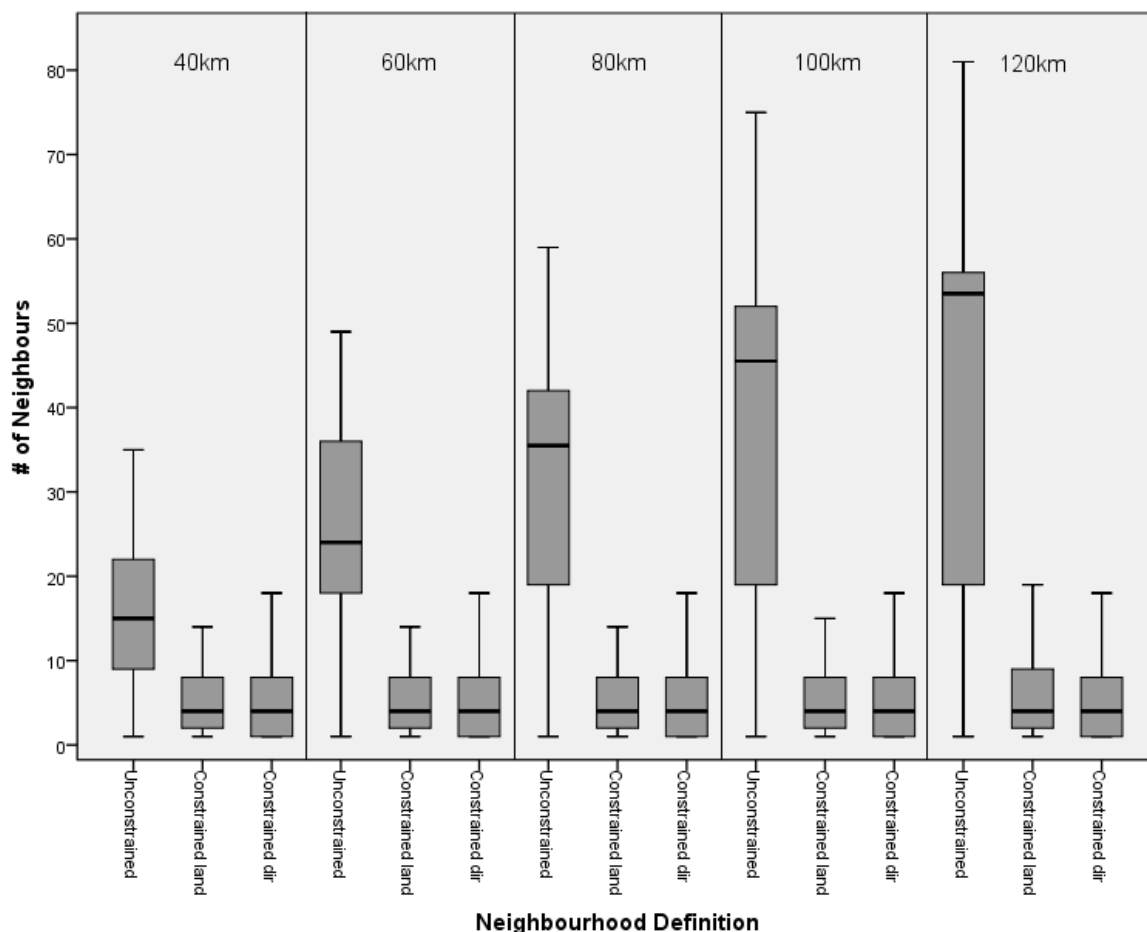


Figure 2-13. Oil spills - Number of neighbourhood per neighbourhood delineation. 40 km – 120 km distance

Variation in the G_i^* Z-score values can be seen in Figure 2-14. Ranges in minimum and maximum values vary according to the distance chosen; 40 km distance ranges for an unconstrained approach are -1.89 to 2.12. For a constrained land -1.12 to 1.49, and for a constrained dir -1.12 to 0.74. For a 120 km distance, the values are reduced, from -1.87 to 1.73 for an unconstrained approach. -1.23 to 1.17 for a constrained land and remain constant likely due to the localized position of the oil spills in the constrained dir approach with values ranging from -1.22 to 0.74 (Figure 2-14 and 2-16).

Spatial clusters of high values identified using neighbourhoods delineated by distances from 40 km to 120 km for the unconstrained approach have a range of 6 (40 km) to 2 (120 km). For the constrained land approach there are seven (40 km) to five (120 km) and for the constrained dir, four clusters across all distances (Table 2-2). For spatial clusters of low values, values ranges from 18 (40 km) to one (120 km) for an unconstrained approach, and for both constrained land and constrained dir approaches there are no clusters of low values (Table 2-2). Clusters of high values remain constant for the constrained dir likely due to geographic barriers impeding that observations have influence over distant locations. High or low values depend on high density of oil spills surrounding high or low concentrations of each occurrence.

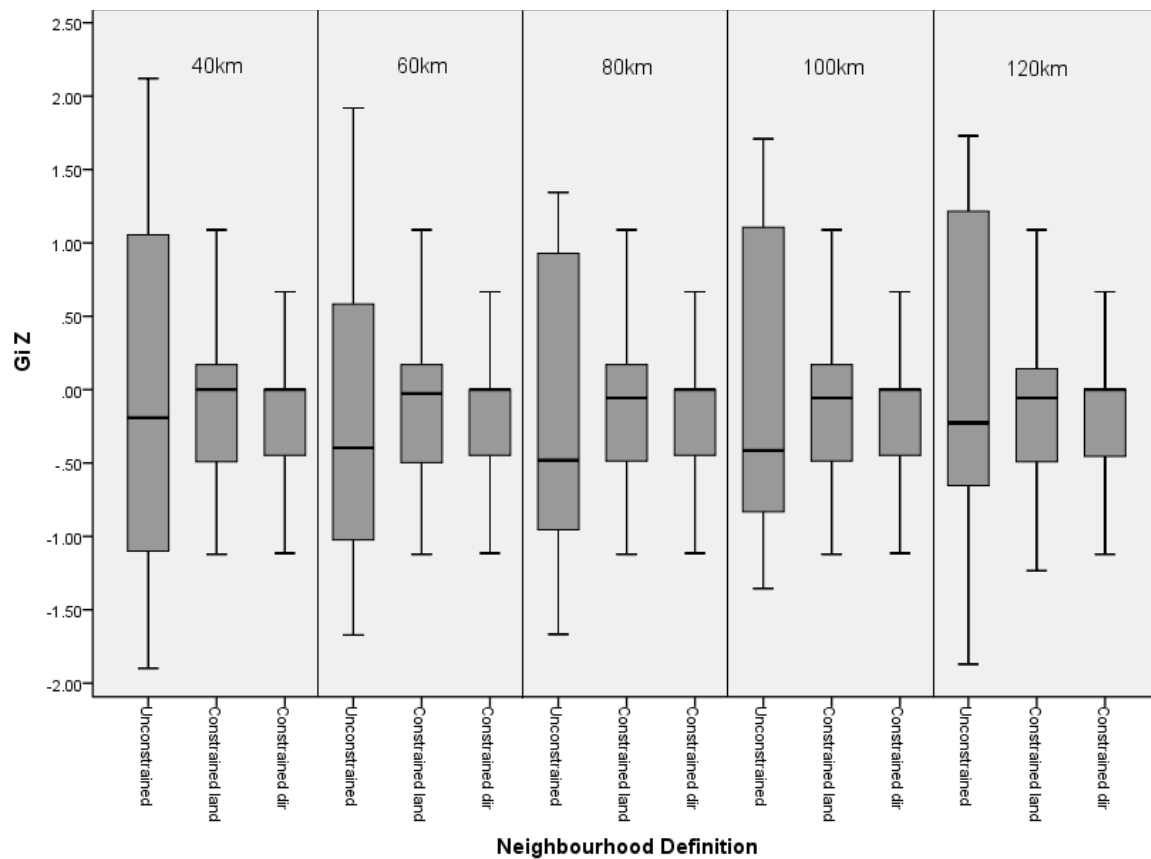


Figure 2-14. Oil spills - G_i^* Z-score per neighbourhood delineation. 40 km – 120 km distance

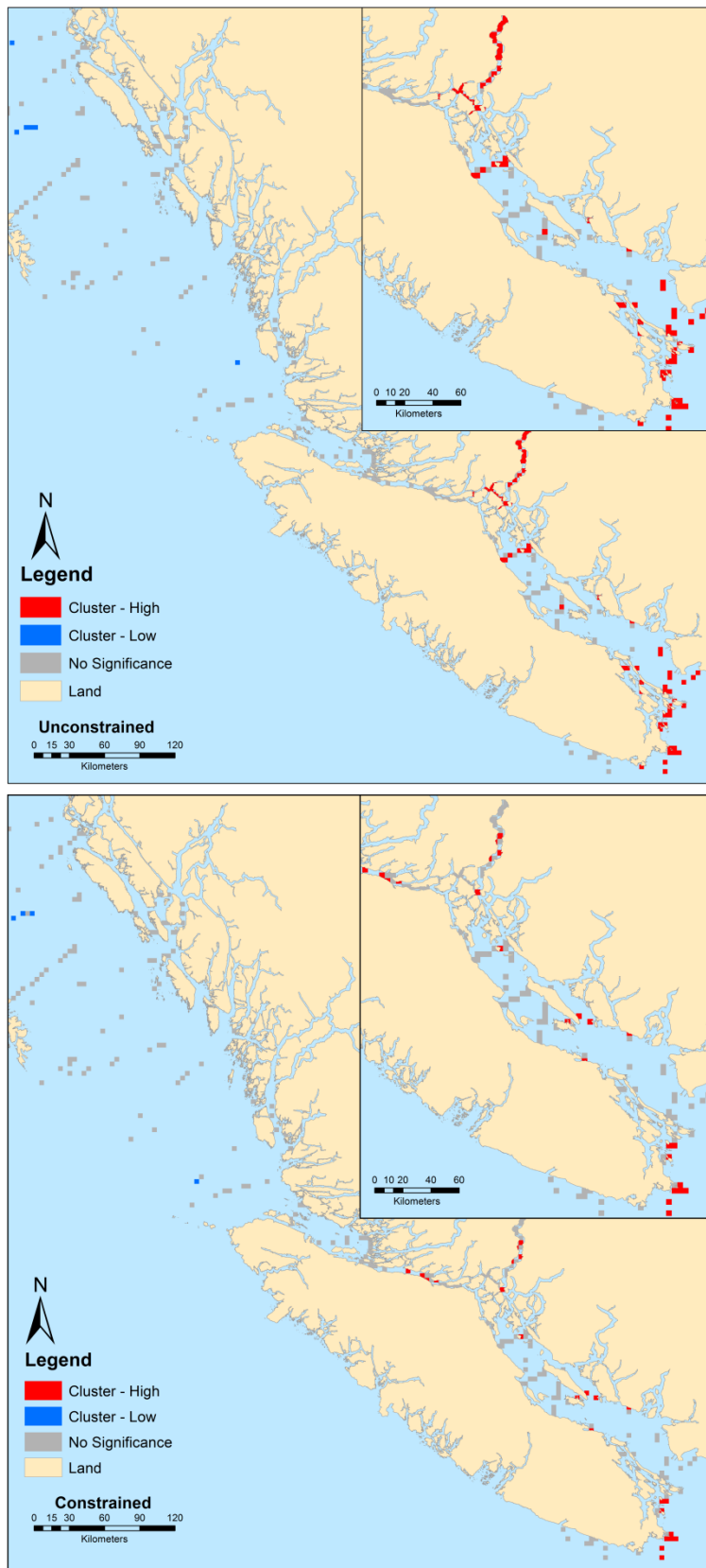


Figure 2-15. Marine mammals abundance - clusters of high and low values

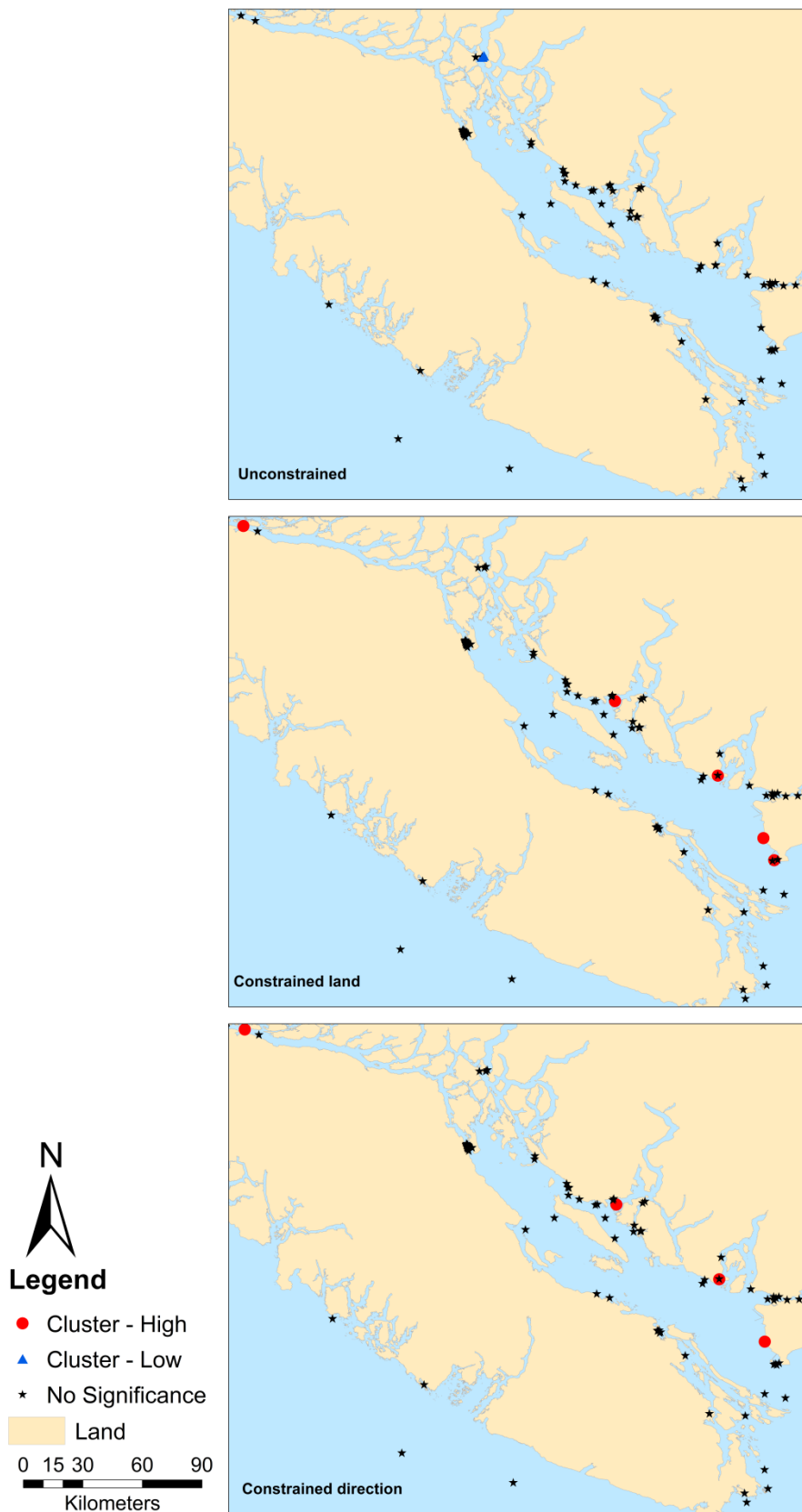


Figure 2-16. Oil spills occurrence - clusters of high and low values

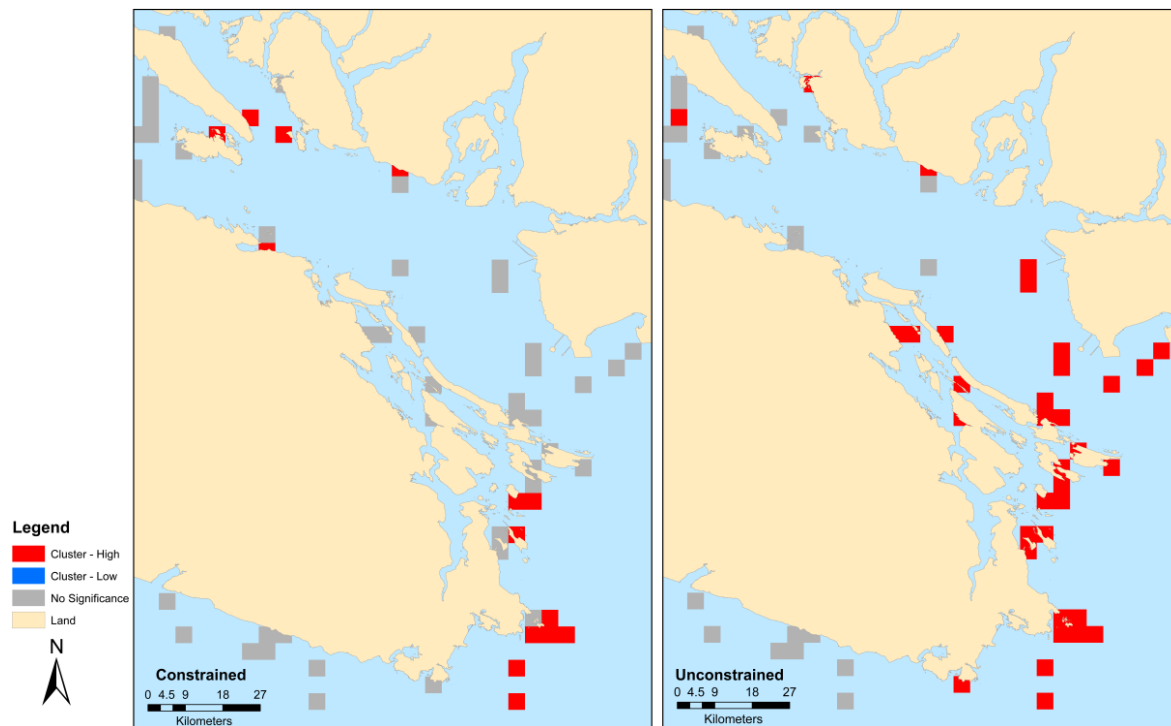


Figure 2-17. Marine mammals abundance - Open waters - clusters of high and low values

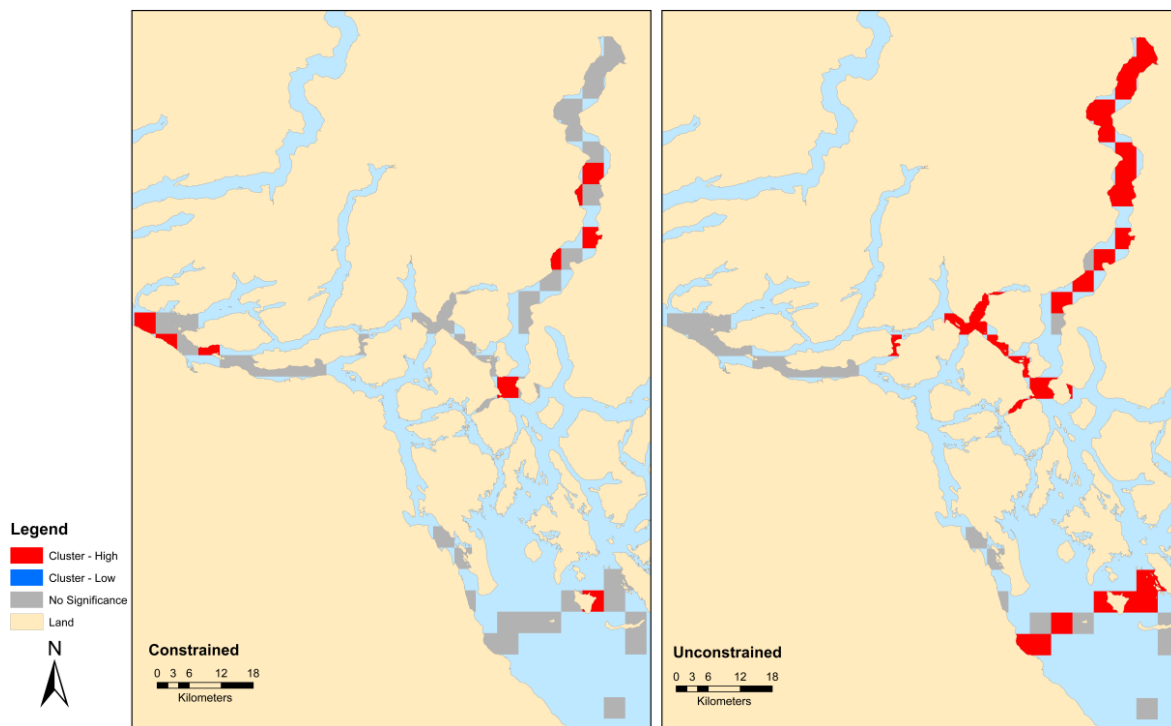


Figure 2-18. Marine mammals abundance - Inlets - clusters of high and low values

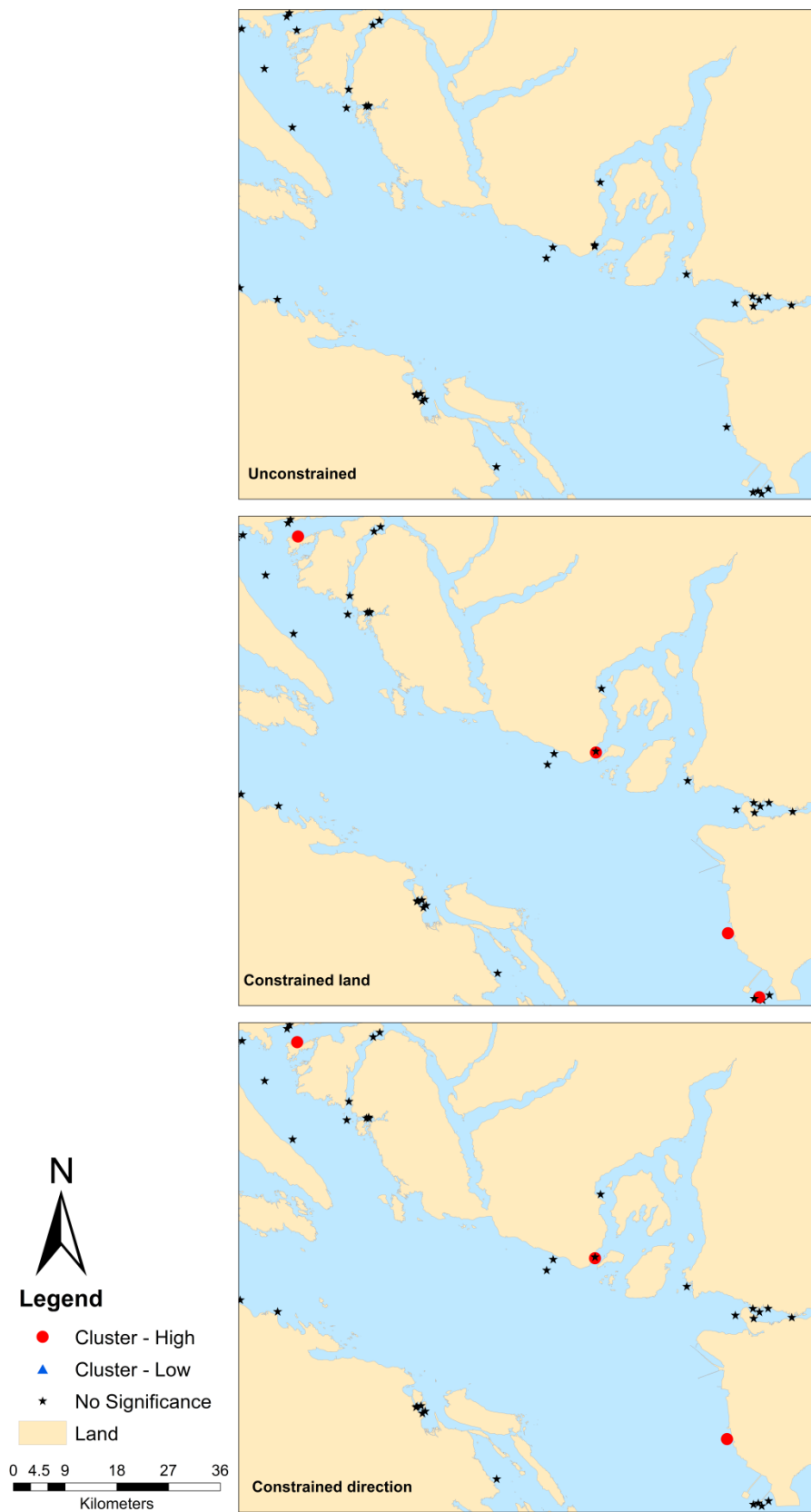


Figure 2-19. Oil spills occurrence - Open waters - clusters of high and low values

2.6 Discussion

In ecology, the term ecological neighbourhood refers to areas that are scaled to a particular ecological process, time period and an organism's mobility or activity (Wiens, 1989). In spatial analysis the area where geographic features influence one another is considered a spatial neighbourhood (O'Sullivan & Unwin, 2002). Either in ecology or spatial analysis, the neighbourhood should represent the spatial interaction and dependency among features; hence, a proper neighbourhood definition is important to determine what the outcome of the analysis will be.

Implementation of constrained approaches using Voronoi or Constrained Delaunay triangulation (CDT) in spatial neighbourhood definitions by adjacency of unconnected points have been previously applied (Gold, 1992; Nordvik & Harding, 2008). With Voronoi or CDT approaches the spatial relationship among neighbours can be altered by adding, changing or removing a neighbour (i.e, Gold and Condal, 1995; Gold, 1994a, 1994b) based on ancillary data of marine environment, in this case the coastline. Depending on the expertise of the researcher, Voronoi or CDT methods can be selected to constrain spatial neighbourhoods; however, this is not currently implemented in common GIS software options.

Another approach to intelligent delineation of spatial neighbourhoods was implemented in AMOEBA (A Multidirectional Optimum Ecotope-Based Algorithm) (Aldstadt & Getis, 2006). AMOEBA map clusters of high or low values by creating a spatial weight matrix based on local statistics such as G_i^* (Aldstadt & Getis, 2006; Jankowska et al., 2008). Spatial neighbourhoods are delineated based on a contiguity approach and are selected when G_i^* of such neighbourhood is greater than zero (0), which

means that the value at that specific location is larger than the mean of all units surrounded it (Aldstadt & Getis, 2006). The AMOEBA procedure constrains spatial neighbourhoods to limit cluster size, by incorporating control over the maximum number of observations in a cluster. The maximum number of neighbours is determined by the geometric form of spatial clusters as subregions of spatial association are identified within the study area (Aldstadt & Getis, 2006). As well, a threshold value based on a variable of the feature, such as area or population (Jankowska et al., 2008). Constraining cluster growth from data properties allows more flexibility and a more realistic representation of spatial pattern. One particular issue with AMOEBA is that it works with areal data, which normally is delineated by spatial contiguity, and does not consider internal barriers. However, both AMOEBA and COSINE aim to provide constraints to spatial neighbourhood delineations that will lead to more realistic analysis results.

Neighbourhood definitions for spatial studies of marine environments would benefit from methods which integrate additional physical data. For instance, Figure 2-17 to Figure 2-19 display examples where the implications of not using constrained approaches result in the detection of additional high values unrepresentative of the process (Figure 2-17 and 2-18). Similarly, including directionality constrained approach demonstrates different findings when comparing it to a constrained approach with no direction (Figure 2-19). The results in our analysis change by including different constraining types, hence, improving upon analysis that does not incorporate natural constraints impeding analysis results.

Taking into consideration the geographic context for the definition of spatial neighbourhoods provides realistic representations of spatial interaction. For instance, the

number of neighbourhood counts for unconstrained and constrained approaches varies, with a reduction between 27% and 33% (40 km and 120 km) for marine mammals' abundance. Similar trends are present in oil spills data, with a reduction from 60% (40 km) and 49% (40 km) for constrained with land and constrained with direction respectively, to 77% to 78% for constrained with land and constrained with direction in a 120 km distance. The reduction in the neighbourhood count is due to geographical barriers impeding the relationship among observations.

The use of local statistics such as G_i^* , acknowledges there is a variation at a local level and suggests there are processes that work at different temporal and spatial scales in a study area (Fortin & Dale, 2005). The concept of scale and its implications for spatial analyses has been the subject of many studies (Dungan et al., 2002; Garrigues et al., 2006; Woodcock & Strahler, 1987; Wu, 2004). Scale is an additional consideration to MAUP, as using the same data at different scales generates different patterns (Wong, 2009). For instance, ecological process and patterns can be identified at different scales, from pollution at the macro-scale (Allen & Rosselot, 1994) to marine mammals species distribution at the global scale (Kaschner et al., 2006), to exposure levels of metals and organic micro-pollutants in clams (*Tapes philipinarum*) at local scale (Bertazzon, et al., 2006). To identify at what scale spatial patterns occur, based on distance, functions such as Ripley's K, pair correlation densities, and second-order product densities are used (Detto & Muller-Landau, 2013).

Another aspect that has to be taken into account when performing spatial analysis of aggregated data is the edge effect. Incorporating internal edges into the study area requires the implementation of an edge correction that modifies the spatial

neighbourhood delineation. In edge effect, aggregated data that are near the center of the study area will have neighbourhoods in all direction, whereas close to the boundary neighbourhoods can only be defined in the direction towards the study area centre (Goreaud & Péliissier, 1999; O'Sullivan & Unwin, 2002). Edge corrections have been suggested. For instance, it is possible to weighted neighbours near a study area edge (Ripley, 1977) or a guard zone around the edge of the study area can include data used only for building neighbourhoods for other data features (O'Sullivan & Unwin, 2010). As well, when study areas are rectangular a toroidal “wrap” can be employed. By joining the top and bottom edges of the study area, data can be rotated within the study region to effectively borrow data from other parts of the regions for defining neighbourhoods (Yamada & Rogerson, 2003).

The edge effect literature provides additional insight into methods for adjusting neighbourhoods. However, it also highlights a limitation of COSINE, COSINE will have more edge effects than unconstrained analyses, as edges will occur both at the study area extent and within the study area. Future research could include edge corrections within COSINE by implementing a weighted edge correction using the inverse distance spatial neighbourhood delineation.

Implementation of directionality as an additional component to constrained spatial neighbourhoods is useful for research in open waters that is subject to different current directions. For the case study on oil spills, there is a reduction in the number of locations with surrounding high values when the direction approach is added. The directionality approach is useful for oil spill or any kind of study that relies on influence of data based on direction. Similarly, directionality (currents) may alter the relationship of neighbours

in inlet areas or areas where currents may change as a result of variation in topography. For instance, neighbourhood selection may be influenced by the direction of an ocean current, if both observations are within a specific angle, which means, both observations may be heading to the same direction. One example is, if the angle between two observations is greater than 10° , both observations may not be related as the speed and direction of each observation may lead them to different paths.

In an ecological framework a constrained neighbourhood can incorporate impassable barriers, such as wildfires for long-tailed finch (*Poephila acuticauda*) (i.e., Brazill-Boast et al., 2011), or limiting animal movement in both terrestrial and marine environments. Constrained spatial neighbourhoods in plant ecology (i.e., Rayburn et al., 2011) can also be used to identify spatial patterns and interaction of plants over a specific area. In the context of environmental degradation, wind direction inclusion can inform atmospheric pollution modelling (Gallo & Chasco, 2012), CONSINE, can be used to constrain the influence of features, using wind direction, in terrestrial environments. Likewise, human mobility restrictions such as road networks and urban settlements can inform the relative settlement patterns and hot spot detections central to crime research (i.e., Hipp et al., 2012; Murray & Grubestic, 2013). Applications can extend to all arenas of spatial modelling in which simply delineated contiguity or distance decay matrices of space are unsuitable for representing how human, animal or environmental processes develop spatial patterns.

2.7 Conclusions

This research aimed to implement and compare spatial neighbourhoods delineated by distance. We demonstrate that by using unconstrained or constrained approaches for

spatial analysis, outcomes are altered. Implementing constrained approaches improves the analysis of local observations, by limiting the area of influence or interaction among observations. In spatial analysis, the outcome of an analysis will be affected depending on how the data were aggregated as well as at what scale the analysis was performed. We have illustrated this specific example by identifying clusters of high or low values at different distances. Standard approaches of spatial neighbourhoods do not rely on a geographic or topographic component, rather in a planar environment where impediments in the influence of features are not present.

Spatial analysis in marine environments is a good example of how a correct definition of spatial neighbourhood is required to ensure accurate results in the analysis. Performing spatial analysis in marine environments requires implementing constraints, such as coastlines or directionality of currents, in order to identify spatial relationship among observations. Constrained approaches, such as constraining by land, will limit the influence that observations have over distance, reducing the issue with scale when aggregating data. For instance, reducing the influence of observations over distance by limiting the spatial or temporal scale will help to identify the influence that local patterns of association have over a study area. Further research on implementing terrestrial applications will be valuable to improve the existing application as well as to compare it with traditional approaches.

3.0 Conclusion

3.1 Summary

The goal of this research was to develop and implement a new method to constrain standard neighbourhood definitions in marine environments. Three objectives were addressed to meet this goal. 1) Describe and identify common spatial neighbourhood definitions and outline two methods to constrain spatial neighbourhood definitions, based on land and direction of ocean currents. 2) Develop an open source software tool (CONstraining SPAtial NEighbourhoods - COSINE), which can be used in current GIS software, and limits the influence of standard approaches of neighbourhood definitions by integrating land or direction as constraining components. 3) Demonstrate the impacts of the new methods on two case studies for marine mammals' abundance and oil spills occurrence.

The case studies of marine mammals' abundance and oil spill occurrence demonstrated the ability of COSINE software to add a new approach to constraining spatial neighbourhood definitions in marine environments. Similarly, COSINE software can be used to perform spatial analysis in terrestrial environments by changing the environment. Spatial analysis in terrestrial environments has several applications, among them species distribution, which can be improved with the implementation of a constraint approach like COSINE. Furthermore, directionality approach can be used in terrestrial environments to study impacts of pollution over population, seeds dispersion, or allergens influenced by seasonal changes.

By implementing local statistic method Getis-Ord G_i^* to detect locations with high surrounding values of high/low density of oil spills or abundance of marine mammals, we were able to identified clusters of high or low values using both unconstrained and constrained spatial neighbourhoods. While the reduction in the number of locations with high/low values using a constrained approach is instinctive, normally this is not the case in present spatial neighbourhood definitions.

The development of COSINE using open software allows spatial analysts to implement and to create further definitions of spatial neighbourhoods that may be used in conjunction with additional data types for input or constraint, in order to modify the geographic influence of spatial neighbourhoods. Though designed for the general user, the output of COSINE requires manual user editing to accommodate to the input requirements of different GIS software that employs spatial analysis. With COSINE as open software, users can improve it to accommodate the output according to a variety of GIS software.

COSINE software at the present is suitable for point data or areal data extracting the centroid of the polygon. To accomplish reliable results in spatial analysis is imperative to select an appropriate spatial neighbourhood definition (Nelson & Robertson, 2012). Implementation of constraint approaches using Voronoi or Constrained Delaunay Triangulation (CDT) in spatial neighbourhoods definitions by adjacency have been previously implemented (Gold, 1992; Nordvik and Harding, 2008). It is important to note that CDT or Voronoi methods are not always chosen by researchers and the spatial neighbourhood definition is subject to the knowledge of the researcher, hence, the

idea to propose the constrained approach to standard definitions of spatial neighbourhoods.

3.2 Research Contributions

The major contribution of this research is the development of a new method to constrain existing spatial neighbourhood definitions, with an important contribution to the fields of spatial analysis, marine and terrestrial ecology, human behaviour and GIS. This contribution followed attempts to constraint spatial neighbourhoods using the adjacency approaches like Voronoi or CDT, where the spatial relationship among neighbours can be altered by adding, changing or removing a neighbourhood (Gold & Condal, 1995; Gold, 1992, 1994a; Nordvik & Harding, 2008).

Another contribution to this thesis is the implementation of local statistics to detect values of high or low density for abundance or occurrence in two marine case studies, marine mammals and oil spills. As a case study, the implication of constraining existing delineations of spatial neighbourhood using barriers, like land, or directionality approach, like ocean currents, allowed the researches to compare the different findings in locations with high or low values for abundance or occurrence. Influence of observations for unconstrained are higher when this approach is used, compared to constrained with barrier or even less when directionality is added in the constrained with barrier and direction approach.

Current GIS software provide limited integration with tools and methods for complex spatial analysis, leading to development of specialized software that integrates spatial analysis and GIS data (Martin, 2008). There are several software for spatial analysis, some of them, SaTScan (spatial, temporal, and space-time scan statistics) (Amin

et al., 2012), SAM (spatial analysis in macroecology) (Rangel et al., 2010), PASSaGE (pattern analysis, spatial statistics and geographic exegesis) (Rosenberg & Anderson, 2011) and OpenGeoDa (Geographic Data Analysis) (Anselin et al., 2006) that can benefit from the outcome of this research, by implementing a modified spatial weight matrix with local spatial statistics analysis. Implementation of COSINE in open software, permits future improvements and implementations with commercial and open GIS software as well as spatial analysis packages

3.3 Research Opportunities

There are many areas where COSINE software can be improved, such as input data type and neighbourhoods type, at the moment, fixed distance, inverse distance (with power 1 or 2) and k-nearest neighbourhood (Knn) spatial neighbourhoods delineations can be selected, future development could include contiguity (Queen, Rook, Bishop), Voronoi or Delaunay triangulation methods to delineate spatial neighbourhoods. Additionally, the directionality approach can be improved by including a directionality variogram in order to identify the trend in observations or a moving window that allows the user to identify changes in the angle selected. As stated by Nelson & Robertson (2012), spatial neighbourhood definition is not always explained in literature, in part because it can draw attention to the method chosen and divert the goal of the research. However, neighbourhood selection is crucial in the development of correct spatial analysis research.

The development of this thesis research deals with point objects as input feature to be evaluated and polygon or polylines features as the barriers input features as well as raster in ESRI GRID format. Improvements in COSINE software could include input of

more data types, such as polygon or polylines as the features to be evaluated in the neighbourhood selection, as well as points as barriers for network analysis. Including more data types as input or barriers, will increase the implementation of spatial analysis in marine or terrestrial environments.

Integration of the geographic or topographic component into spatial neighbourhood delineations will improve research opportunities in the field of spatial analysis, as it will identify patterns at a local scale. Applications for delineating spatial neighbourhoods in marine or terrestrial environments ranges from animal behaviour (Kanagaraj et al., 2011; Smulders et al., 2010; Wigand et al., 2013), to species richness (White et al., 2010), to marine species (Embling et al., 2012), to maritime and terrestrial traffic (Erdogan, 2009; Pelot & Plummer, 2010; Wang et al., 2009). Whenever there is interaction among features, a constraint factor should be included, implementing an extra component to the first law of geography “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970), which is to identify if closer features are actually related even though there is an impedance that prevent them to be related. The ideas presented in this research offer the potential to apply a more realistic spatial analysis in a geographic context; we hope to open more opportunities in the spatial analysis research and environment.

Bibliography

- Aguilar-Perera, A., & Appeldoorn, R. S. (2008). Spatial distribution of marine fishes along a cross-shelf gradient containing a continuum of mangrove-seagrass-coral reefs off southwestern Puerto Rico. *Estuarine, Coastal and Shelf Science*, 76(2), 378–394. doi:10.1016/j.ecss.2007.07.016
- Aldstadt, J., & Getis, A. (2006). Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. *Geographical Analysis*, 38(4), 327–343. doi:10.1111/j.1538-4632.2006.00689.x
- Allen, D. T., & Rosselot, K. S. (1994). Pollution prevention at the macro scale: flows of wastes, industrial ecology and life cycle analyses. *Waste Management*, 14(3-4), 317–328. doi:10.1016/0956-053X(94)90078-7
- Amin, R., Ritter, E. K., & Cossette, L. (2012). A Geospatial Analysis of Shark Attack Rates for the Coast of California: 1994–2010. *Journal of Environment and Ecology*, 3(1), 246–255. doi:10.5296/jee.v3i1.2700
- Anselin, L. (1989). *What is special about spatial data? Alternative perspectives on spatial data analysis*. (D. A. Griffith, Ed.) *Symposium on Spatial Statistics Past Present and Future* (pp. 63–77). Institute of Mathematical Geography.
- Anselin, Luc, & Getis, A. (2010). Spatial Statistical Analysis and Geographic Information Systems. In Luc Anselin & S. J. Rey (Eds.), *Perspectives on Spatial Data Analysis* (Vol. 26, pp. 19–33). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-01976-0
- Anselin, Luc, Syabri, I., & Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, 38(1), 5–22. doi:10.1111/j.0016-7363.2005.00671.x
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. (Longman, Ed.) *Social Science Medicine 1982* (Vol. 42, p. 413). Longman. doi:10.1016/0277-9536(95)00183-2
- Bertazzon, S., Micheletti, C., Critto, A., & Marcomini, A. (2006). Spatial analysis in ecological risk assessment: Pollutant bioaccumulation in clams *Tapes philipinarum* in the Venetian lagoon (Italy). *Computers Environment and Urban Systems*, 30(6), 880–904. doi:10.1016/j.compenvurbsys.2005.09.003

- Best, B., & Halpin, P. (2009). *Predictive Marine Mammal Modeling for Queen Charlotte Basin, British Columbia* (p. 120). Sidney, BC, Canada: Raincoast Conservation Foundation.
- Bonjean, F., & Lagerloef, G. S. E. (2002). Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean. *Journal of Physical Oceanography*, 32(10), 2938–2954. doi:10.1175/1520-0485(2002)032<2938:DMAAOT>2.0.CO;2
- Booth, A. (2000). Incorporating the spatial component of fisheries data into stock assessment models. *ICES Journal of Marine Science*, 57(4), 858–865. doi:10.1006/jmsc.2000.0816
- Brazill-Boast, J., Rooij, E., Pryke, S. R., Griffith, S. C., & Van Rooij, E. (2011). Interference from long-tailed finches constrains reproduction in the endangered Gouldian finch. *The Journal of animal ecology*, 80(1), 39–48. doi:10.1111/j.1365-2656.2010.01756.x
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4).
- Chandrasekharan, H., Sarangi, A., Nagarajan, M., Singh, V. P., Rao, D. U. M., Stalin, P., ... Anbazhagan, S. (2008). Variability of soil-water quality due to Tsunami-2004 in the coastal belt of Nagapattinam district, Tamilnadu. *Journal of environmental management: Environmental Aspects of the Indian Ocean Tsunami Recovery*, 89(1), 63–72. doi:10.1016/j.jenvman.2007.01.051
- Chang, Y., Huang, Y., Chang, D.-C., & Lee, M.-T. (2006). Assessment of marine environment by spatial analysis in Kao-Ping coastal water. *Journal of Environmental Engineering and Management*, 16(5), 351–363.
- Chomitz, K. M., & Thomas, T. S. (2003). Determinants of Land Use in Amazonia: A Fine-Scale Spatial Analysis. *American Journal of Agricultural Economics*, 85(4), 1016–1028. doi:10.1111/1467-8276.00504
- Cliff, A. D., & Ord, K. (1970). Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography*, 46(1), 269–292.
- Crawford, W. R. (2001). *Oceans of the Queen Charlotte Islands. Canadian Technical Report of Fisheries and Aquatic Sciences*.
- Daly, C., Gibson, W., Taylor, G., Johnson, G., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate Research*, 22(2), 99–113. doi:10.3354/cr022099

- Detto, M., & Muller-Landau, H. C. (2013). Fitting ecological process models to spatial patterns using scalewise variances and moment equations. *The American naturalist*, *181*(4), E68–82. doi:10.1086/669678
- Dionisio, K. L., Rooney, M. S., Arku, R. E., Friedman, A. B., Hughes, A. F., Vallarino, J., ... Ezzati, M. (2010). Within-neighborhood patterns and sources of particle pollution: mobile monitoring and geographic information system analysis in four communities in Accra, Ghana. *Environmental health perspectives*, *118*(5), 607–13. doi:10.1289/ehp.0901365
- Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., ... Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*(5), 626–640. doi:10.1034/j.1600-0587.2002.250510.x
- Embling, C. B., Illian, J., Armstrong, E., van der Kooij, J., Sharples, J., Camphuysen, K. C. J., & Scott, B. E. (2012). Investigating fine-scale spatio-temporal predator-prey patterns in dynamic marine ecosystems: a functional data analysis approach. *Journal of Applied Ecology*, *49*(2), 481–492. doi:10.1111/j.1365-2664.2012.02114.x
- Erdogan, S. (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of safety research*, *40*(5), 341–51. doi:10.1016/j.jsr.2009.07.006
- Esri. (2011). ArcGIS Desktop: Release 10. Redlands CA. Environmental Systems Research Institute.
- Farley, A. L. (2011). *Atlas of British Columbia* (p. 163). UBC Press.
- Fortin, M.-J., & Dale, M. (2005). *Spatial Analysis: A Guide for Ecologists* (1st ed., p. 392). UK: Cambridge University Press.
- Fotheringham, A. S. (2009). Geographically Weighted Regression. In A. Stewart Fotheringham & P. A. Rogerson (Eds.), *The SAGE Handbook of Spatial Analysis* (1st ed., p. 528). London: SAGE Publications Ltd.
- Fotheringham, A. S., & Rogerson, P. A. (2009). Introduction. In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE Handbook of Spatial Analysis* (1st ed., p. 528). London: SAGE Publications Ltd. doi:10.4135/978-1-8570-2013-7
- Fotheringham, A. S., & Zhan, F. B. (2010). A Comparison of Three Exploratory Methods for Cluster Detection in Spatial Point Patterns. *Geographical Analysis*, *28*(3), 200–218. doi:10.1111/j.1538-4632.1996.tb00931.x
- Funk, S., Salathé, M., & Jansen, V. A. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface the Royal Society*, *7*(50), 1247–1256. Retrieved from

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2894894&tool=pmcentrez&rendertype=abstract>

- Gallo, J. Le, & Chasco, C. (2012). Dealing with Data at Various Spatial Scales and Supports: An Application on Traffic Noise and Air Pollution Effects on Housing Prices with Multilevel Models. In E. Fernández Vázquez & F. Rubiera Morollón (Eds.), *Defining the Spatial Scale in Modern Regional Analysis* (pp. 281–309). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-31994-5_14
- Garrigues, S., Allard, D., Baret, F., & Weiss, M. (2006). Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sensing of Environment*, 103(1), 81–96. doi:10.1016/j.rse.2006.03.013
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x
- Gold, C. M. (1992). The meaning of “neighbour.” In A. U. Frank, I. Campari, & U. Formentini (Eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space* (Vol. 639, pp. 220–235). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/3-540-55966-3
- Gold, C. M. (1994a). Problems with Handling Spatial Data - the Voronoi Approach. *CISM Journal ASCSGC*, 45(1), 65–80.
- Gold, C. M. (1994b). Three approaches to automated topology, and how computational geometry helps. In *Sixth International Symposium on Spatial Data Handling* (pp. 145–158). Edinburgh, Scotlan: CRC Press.
- Gold, C. M., & Condal, A. R. (1995). A spatial data structure integrating GIS and simulation in a marine environment. *Marine Geodesy*, 18(3), 213–228. doi:10.1080/15210609509379757
- Goreaud, F., & Pélissier, R. (1999). On explicit formulas of edge effect correction for Ripley’s K -function. *Journal of Vegetation Science*, 10(3), 433–438. doi:10.2307/3237072
- Grech, A., Sheppard, J., & Marsh, H. (2011). Informing species conservation at multiple scales using data collected for marine mammal stock assessments. *PloS one*, 6(3), e17993. doi:10.1371/journal.pone.0017993
- Hipp, J. R., Faris, R. W., & Boessen, A. (2012). Measuring “neighborhood”: Constructing network neighborhoods. *Social Networks*, 34(1), 128–140. doi:10.1016/j.socnet.2011.05.002

- Hooker, S. K., & Gerber, L. R. (2004). Marine Reserves as a Tool for Ecosystem-Based Management: The Potential Importance of Megafauna. *BioScience*, *54*(1), 27. doi:10.1641/0006-3568(2004)054[0027:MRAATF]2.0.CO;2
- Hwang, C., Shih, H.-C., Guo, J., & Hsiao, Y.-S. (2008). Zonal and Meridional Ocean Currents at TOPEX/Poseidon and JASON-1 Crossovers around Taiwan: Error Analysis and Limitation. *Terrestrial, Atmospheric and Oceanic Sciences*, *19*(1-2), 151. doi:10.3319/TAO.2008.19.1-2.151(SA)
- Isaaks, E. H., & Srivastava, R. M. (1989). *Applied Geostatistics*. (R. M. Srivastava, Ed.) (p. 561). Oxford: Oxford University Press.
- Jankowska, M., Aldstadt, J., Getis, A., Weeks, J., & Fraley, G. (2008). An amoeba procedure for visualizing clusters. In *GIScience*. Park City, UT.
- Jerrett, M., Burnett, R. T., Ma, R., Pope, C. A., Krewski, D., Newbold, K. B., ... Thun, M. J. (2005). Spatial Analysis of Air Pollution and Mortality in Los Angeles. *Epidemiology*, *16*(6), 727–736. doi:10.1097/01.ede.0000181630.15826.7d
- Johannessen, D. I., Harris, K. A., Macdonald, J. S., & Ross, P. S. (2007). Marine Environmental Quality in the North Coast and Queen Charlotte Islands, British Columbia, Canada: A Review of Contaminant Sources, Types, and Risks. *Canadian Technical Report of Fisheries and Aquatic Sciences*, *2717*, 101. Retrieved from <http://www.arlis.org/docs/vol1/182520446.pdf>
- Kanagaraj, R., Wiegand, T., Kramer-Schadt, S., Anwar, M., & Goyal, S. P. (2011). Assessing habitat suitability for tiger in the fragmented Terai Arc Landscape of India and Nepal. *Ecography*, *34*(6), 970–981. doi:10.1111/j.1600-0587.2010.06482.x
- Kaschner, K., Watson, R., Trites, A., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, *316*(December), 285–310. doi:10.3354/meps316285
- Kheruntsyan, K., Jaskula, J. C., Deuar, P., Bonneau, M., Partridge, G., Ruaudel, J., ... Westbrook, C. (2012). Violation of the Cauchy-Schwarz Inequality with Matter Waves. *Physical Review Letters*, *108*(26), 1–5. doi:10.1103/PhysRevLett.108.260401
- LeDrew, E. F., Holden, H., Wulder, M. A., Derksen, C., & Newman, C. (2004). A spatial statistical operator applied to multirate satellite imagery for identification of coral reef stress. *Remote Sensing of Environment*, *91*(3-4), 271–279. doi:<http://dx.doi.org/10.1016/j.rse.2003.10.007>

- Li, J., & Heap, A. D. (2008). *A Review of Spatial Interpolation Methods for Environmental Scientists* (Record 200., p. 137). Geoscience Australia.
- Lorenzen, K., Steneck, R. S., Warner, R. R., Parma, A. M., Coleman, F. C., & Leber, K. M. (2010). The spatial dimensions of fisheries: putting it all in place. *Bulletin of Marine Science*, *86*(2), 169–177.
- Magalhaes, M. F., Batalha, D. C., & Collares-Pereira, M. J. (2002). Gradients in stream fish assemblages across a Mediterranean landscape: contributions of environmental factors and spatial structure. *Freshwater Biology*, *47*(5), 1015–1031. doi:10.1046/j.1365-2427.2002.00830.x
- Magesh, N. S., Chandrasekar, N., & Vetha Roy, D. (2011). Spatial analysis of trace element contamination in sediments of Tamiraparani estuary, southeast coast of India. *Estuarine, Coastal and Shelf Science*, *92*(4), 618–628. doi:10.1016/j.ecss.2011.03.001
- Mallory, F. F., & Hillis, T. L. (1998). Demographic characteristics of circumpolar caribou populations: ecotypes, ecological constraints, releases, and population dynamics. *Rangifer*, *18*(5), 49–60. doi:10.7557/2.18.5.1541
- Marchal, E., & Petitgas, P. (1993). Precision of acoustic fish abundance estimates: separating the number of schools from the biomass in the schools. *Aquatic Living Resources*, *6*(3), 211–219. Retrieved from <http://publications.edpsciences.org/articles/alr/pdf/1993/03/alr93305.pdf>
- Martin, D. (2008). The Role of GIS. In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE Handbook of Spatial Analysis* (1st ed., p. 528). London: SAGE Publications Ltd.
- Masjed-Jamei, M. (2009). A functional generalization of the Cauchy-Schwarz inequality and some subclasses. *Applied Mathematics Letters*, *22*(9), 1335–1339. doi:10.1016/j.aml.2009.03.001
- Mellin, C., Bradshaw, C. J. A., Meekan, M. G., & Caley, M. J. (2010). Environmental and spatial predictors of species richness and abundance in coral reef fishes. *Global Ecology and Biogeography*, *19*(2), 212–222. doi:10.1111/j.1466-8238.2009.00513.x
- Moore, S. E., & DeMaster, D. P. (1997). Cetacean Habitats in the Alaskan Arctic. *Journal Of Northwest Atlantic Fishery Science*, *22*, 55–69. doi:10.2960/J.v22.a5
- Murray, A. T., & Grubestic, T. H. (2013). Crime Modeling and Mapping Using Geospatial Technologies. In M. Leitner (Ed.), *Crime Modeling and Mapping Using Geospatial Technologies, Geotechnologies and the Environment 8* (8th ed., pp. 105–124). Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-4997-9

- Myers, D. E. (1991). On variogram estimation. In *Proceedings of the First International Conference on Statistical Computing* (pp. 261–266). Cesme, Turkey: American Sciences Press.
- Nelson, T. A., & Robertson, C. (2012). Refining spatial neighbourhoods to capture terrain effects. *Ecological Processes*, 1(3), 1–11. doi:10.1186/2192-1709-1-3
- Nordvik, T., & Harding, C. (2008). Interactive Geovisualization and Geometric Modelling of 3D Data - A Case Study from the Åknes Rockslide Site, Norway. In A. Ruas & C. Gold (Eds.), *Headway in Spatial Data Handling* (pp. 367–384). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-68566-1
- O'Hara, P. D., & Morgan, K. H. (2006). Do low rates of oiled carcass recovery in beached bird surveys indicate low rates of ship-source oil spills? *Marine Ornithology*, 34(2), 133–140.
- O'Hara, P. D., Serra-Sogas, N., Canessa, R., Keller, P., & Pelot, R. (2013). Estimating discharge rates of oily wastes and deterrence based on aerial surveillance data collected in western Canadian marine waters. *Marine pollution bulletin*, 69(1-2), 157–64. doi:10.1016/j.marpolbul.2013.01.034
- O'Sullivan, D., & Unwin, D. (2002). *Geographic Information Analysis. Geographic Information Analysis* (p. 436). Wiley. Retrieved from <http://www.loc.gov/catdir/toc/wiley023/2002028837.html>
- O'Sullivan, D., & Unwin, D. (2010). *Geographic Information Analysis* (2nd ed., p. 432). New Jersey: Wiley.
- ODriscoll, R. L. (1998). Description of spatial pattern in seabird distributions along line transects using neighbour K statistics. *Marine Ecology Progress Series*, 165, 81–94.
- Overholtz, W. J. (2002). The Gulf of Maine–Georges Bank Atlantic herring (*Clupea harengus*): spatial pattern analysis of the collapse and recovery of a large marine fish complex. *Fisheries Research*, 57(3), 237–254. doi:10.1016/S0165-7836(01)00359-9
- Pelot, R., & Plummer, L. (2010). Spatial Analysis of Traffic and Risks in the Coastal Zone. In D. R. Green (Ed.), *Coastal and Marine Geospatial Technologies* (13th ed., Vol. 13, pp. 227–238). Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-9720-1
- Rangel, T. F., Diniz-Filho, J. A. F., & Bini, L. M. (2010). SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography*, 33(1), 46–50. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0587.2009.06299.x/pdf>

- Rayburn, A. P., Schiffers, K., & Schupp, E. W. (2011). Use of precise spatial data for describing spatial patterns and plant interactions in a diverse Great Basin shrub community. *Plant Ecology*, 212(4), 585–594. doi:10.1007/s11258-010-9848-0
- Redfern, J. V., McKenna, M. F., Moore, T. J., Calambokidis, J., Deangelis, M. L., Becker, E. A., ... Chivers, S. J. (2013). Assessing the risk of ships striking large whales in marine spatial planning. *Conservation biology : the journal of the Society for Conservation Biology*, 27(2), 292–302. doi:10.1111/cobi.12029
- Richardson, A. J., & Schoeman, D. S. (2004). Climate impact on plankton ecosystems in the Northeast Atlantic. *Science (New York, N.Y.)*, 305(5690), 1609–12. doi:10.1126/science.1100958
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society Series B Methodological*, 39(2), 172–212. doi:10.1046/j.1369-7412.2003.05285.x
- Rode, K. D., Robbins, C. T., & Shipley, L. A. (2001). Constraints on herbivory by grizzly bears. *Oecologia*, 128(1), 62–71. doi:10.1007/s004420100637
- Rosenberg, M. S., & Anderson, C. D. (2011). PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution*, 2(3), 229–232. doi:10.1111/j.2041-210X.2010.00081.x
- Rushton, G. (1993). Human Behaviour in Spatial Analysis. *Urban Geography*, 14(5), 447–456. doi:10.2747/0272-3638.14.5.447
- Russell, R. W., Hunt, G. L., Coyle, K. O., & Cooney, R. T. (1992). Foraging in a fractal environment: Spatial patterns in a marine predator-prey system. *Landscape Ecology*, 7(3), 195–209. doi:10.1007/BF00133310
- Samuels, W. B., Amstutz, D. E., Bahadur, R., & Ziemniak, C. (2013). Development of a Global Oil Spill Modeling System. *Earth Science Research*, 2(2). doi:10.5539/esr.v2n2p52
- Serra-Sogas, N, O'Hara, P., Canessa, R., Bertazzon, S., & Gavrilova, M. (2008). Exploratory Spatial Analysis of Illegal Oil Discharges Detected off Canada's Pacific Coast. *Lecture Notes in Computer Science*, 5072, 81–95. Retrieved from <http://www.springerlink.com/content/r2h87652j3743457/fulltext.pdf>
- Serra-Sogas, Norma, O'Hara, P. D., Canessa, R., Keller, P., & Pelot, R. (2008). Visualization of spatial patterns and temporal trends for aerial surveillance of illegal oil discharges in western Canadian marine waters. *Marine pollution bulletin*, 56(5), 825–33. doi:10.1016/j.marpolbul.2008.02.005
- Smulders, M., Nelson, T. A., Jelinski, D. E., Nielsen, S. E., & Stenhouse, G. B. (2010). A spatially explicit method for evaluating accuracy of species distribution models.

- Diversity and Distributions*, 16(6), 996–1008. doi:10.1111/j.1472-4642.2010.00707.x
- Thomas, L. E. N., Williams, R. O. B., & Sandilands, D. (2007). Designing line transect surveys for complex survey regions. *Journal of Cetacean Research and Management*, 9(1), 1–13.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234–240. doi:10.1001/archgenpsychiatry.2010.16
- Walker, I. J., & Sydneysmith, R. (2008). British Columbia. In D. S. Lemmen, F. J. Warren, J. Lacroix, & E. Bush (Eds.), *From Impacts to Adaptation: Canada in a Changing Climate 2007* (1st ed., pp. 329–386). Ottawa, ON: Government of Canada.
- Wang, C., Quddus, M. A., & Ison, S. G. (2009). Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident; analysis and prevention*, 41(4), 798–808. doi:10.1016/j.aap.2009.04.002
- West, K., & van Woesik, R. (2001). Spatial and Temporal Variance of River Discharge on Okinawa (Japan): Inferring the Temporal Impact on Adjacent Coral Reefs. *Marine Pollution Bulletin*, 42(10), 864–872. doi:10.1016/S0025-326X(01)00040-6
- White, E. P., Ernest, S. K. M., Adler, P. B., Hurlbert, A. H., & Lyons, S. K. (2010). Integrating spatial and temporal approaches to understanding species richness. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1558), 3633–43. doi:10.1098/rstb.2010.0280
- White, K. S., Westera, M. B., & Kendrick, G. A. (2011). Spatial patterns in fish herbivory in a temperate Australian seagrass meadow. *Estuarine, Coastal and Shelf Science*, 93(4), 366–374. doi:10.1016/j.ecss.2011.05.006
- Whitehead, H., O'Brien, K., & Worm, B. (2010). Diversity of deep-water cetaceans and primary productivity. *Marine Ecology Progress Series*, 408, 1–5. doi:10.3354/meps08619
- Wiens, J. A. (1989). Spatial Scaling in Ecology. *Functional Ecology*, 3(4), 385–397.
- Wigand, L. A., Klinger, T., & Logsdon, M. G. (2013). Patterns in groundfish abundance along the Eastern Bering Sea outer continental margin. *ICES Journal of Marine Science*. doi:10.1093/icesjms/fst054
- Williams, R., & Thomas, L. (2007). Distribution and abundance of marine mammals in the coastal waters of British Columbia, Canada. *Journal of Cetacean Research and Management*, 9(1), 15–28.

- Windle, M. J. S., Rose, G. A., Devillers, R., & Fortin, M.-J. (2010). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science*, 67(1), 145–154. doi:10.1093/icesjms/fsp224
- Wong, D. (2009). The Modifiable Areal Unit Problem (MAUP). In P. A. Rogerson & A. S. Fotheringham (Eds.), *The SAGE Handbook of Spatial Analysis* (1st ed., p. 528). New York, Buffalo: SAGE Publications Ltd.
- Wong, D. W. S., & Lee, J. (2005). *Statistical Analysis of Geographic Information with ArcView GIS And ArcGIS*. (Wiley, Ed.) (p. 464). US.
- Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), 311–332. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/0034425787900150>
- Wooldridge, S., Brodie, J., & Furnas, M. (2006). Exposure of inner-shelf reefs to nutrient enriched runoff entering the Great Barrier Reef Lagoon: post-European changes and the design of water quality targets. *Marine pollution bulletin*, 52(11), 1467–1479. doi:10.1016/j.marpolbul.2006.05.009
- Wu, J. (2004). Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology*, 19(2), 125–138. doi:10.1023/B:LAND.0000021711.40074.ae
- Yamada, I., & Rogerson, P. A. (2003). An Empirical Comparison of Edge Effect Correction Methods Applied to K-function Analysis. *Geographical Analysis*, 35(2), 97–109. doi:10.1111/j.1538-4632.2003.tb01103.x

Appendix A

Graphical user interface (GUI)

The GUI has three sections (Input/Output, Direction and Neighbourhood type).

The screenshot shows the COSINE PACKAGE GUI with the following sections:

- Input/Output:**
 - Input Feature Class: Select a shapefile... (browse)
 - Input Constraint Feature Class: Select a shapefile... (browse)
 - Output file name: Select a name... (browse)
 - Water analysis? (If NO, then uncheck)
- Direction?**
 - Direction? No (dropdown)
 - Meridional (Raster file): 0
 - Zonal (Raster file): 0
 - Max. Angle: 0
- Neighbourhood Type:**
 - Neighbourhood Type: Distance (dropdown)
 - Threshold Distance: 0
 - Threshold Distance: 0
 - Power: 1 (spin box)
 - K-Nearest Neighbours: 0
 - Row standardization (Optional)

Buttons: OK, Cancel

Input / Output section

Users have the option to select the input feature (points) and the feature (polygon) used to constrain the input feature. Similarly, the option to create a feature class (line) will act as weighting matrix with its dbf file, another output generated is a shapefile (lines) with all the connections between observations and the required data used to create the matrix. User has the option to use the checkmark to specify whether the analysis is in terrestrial or marine environments.

Input/Output

Input Feature Class:

Input Constraint Feature Class:

Output file name:

Water analysis? (If NO, then uncheck)

Directionality section

User has the option to select directionality approach (Meridional or Zonal currents for marine environments) and the maximum angle that two observations can be apart.

Direction? ▼

Meridional (Raster file):

Zonal (Raster file):

Max. Angle:

Neighbourhood type section

User can select between fixed distance, inverse distance and K-nearest (Knn) neighbourhood delineations. For all neighbourhood delineations, there is an option to create row standardized or binary matrix. For fixed distance, the threshold distance is giving in metres, similar to the inverse distance, with the addition of the power (one or two), which will reduce the influence of observation, and finally Knn with n number of neighbourhoods based on user selection.

Neighbourhood Type: ▼

Threshold Distance:

Threshold Distance: Power: ▲▼

K-Nearest Neighbours:

Row standardization (Optional)

Appendix B

Help File

PRE-PROCESSING

- It is highly recommended that user install RStudio
- Once installed, user can open the GUI interface, select all (ctrl+A) R code for the GUI and run it, no need to run line by line
- When asked to install GTK+, click Yes (ok)
- It is recommended that the observation (points) and constraint (polyline) shapefile be located under the same folder

User needs to create an OBJECTID column, if already exists, is appropriate to delete current OBJECTID and create one.

To do this:

1. User must create column name 'OBJECTID' with LONG (Data type) and precision 9 (no need to enter precision)
2. Calculate new OBJECTID column as follow: [FID] +1

PROCEDURE to work with ArcGIS for Spatial weights Matrix

1. Open the .DBF file (in Excel) from the newly create shape file
2. From the first row, Delete: NID and WEIGHT column names, LEAVE ONLY OBJECTID
3. Select "Save As type" and choose Other Formats. From the Other Formats choices, select Formatted Text (Space delimited) (*.prn).

4. Choose a location for the file and click SAVE.
5. If a warning message is displayed, click YES
6. In ArcGIS, Open Your analysis method, and Under Conceptualization of Spatial Relationships, select "GET_SPATIAL_WEIGHTS_FROM_FILE"
7. Under "Weights Matrix File (optional)" parameter, locate and select the file with (.prn) extension you have created in steps 1-5

More help:

[http://wiki.answers.com/Q/How_do_you_convert_a_excel_file_to_an_ASCII_file_with_f
ixed_length_fields](http://wiki.answers.com/Q/How_do_you_convert_a_excel_file_to_an_ASCII_file_with_fixed_length_fields)

PROCEDURE to work with OpenGeoDa using spatial weights matrix:

OpenGeoDa structure for spatial weight matrix requires having in the first line of the matrix the following structure:

```
"0" "Total Number of Features" "Source Feature Class (Name of the point shapefile)"
"Unique ID Field Name (OBJECTID)"
0 131 Oil_Spill_Pacific_2008 OBJECTID
```

NOTE: The requirement of OBJECTID is that will be used to link the IDs from the Source Feature Class (point shapefile) and the spatial weight matrix, OBJECTID must match on both features. R reads data.frame of the shapefile (FID in ArcGIS), and spatial statistics tools in ArcGIS and OpenGeoDa requires a Source_ID for the weight matrix (OBJECTID).

Explanation of Fields:

- Input Constraint Feature Class: Requires polygon shapefile.
- Output file name: will create a matrix weight shapefile and an additional shapefile 'CompleteLinks'.
- Water analysis? (If NO, then uncheck): This checkbox evaluates whether the analysis is for aquatic or terrestrial environment (if terrestrial, then uncheck it).
- Zonal (Raster file): Raster file - Zonal direction is towards East (if velocity is negative, then will go towards West).
- Meridional (Raster file): Raster file - Meridional direction is towards North (if velocity is negative, then will go towards South).
- Max. Angle: Maximum angle between Zonal and Meridional direction to be considered in the same flow direction.
- Threshold Distance: Select maximum (threshold) distance (fixed distance or inverse distance) for observations to be considered neighbours.
- Power: Power value for inverse distance (1,2).
- K-nearest neighbours: select number of nearest neighbours (KNN).
- Row standardization: Recommended whenever feature distribution is potentially biased due to sampling design or to an imposed aggregation scheme. Checked-Spatial weights are standardized by row. Each weight is divided by its row sum. Unchecked-No standardization of spatial weights is applied.

Appendix C

R Code

Graphical user interface (GUI)

```

#####
#
# Find out if required packages have been install, if not, then proceed to
# install them
#
#####

list.of.packages <- c("gWidgets", "gWidgetsRGtk2", "maptools", "raster",
                    "spdep", "sp", "geoR", "igraph", "spatstat", "gtools",
                    "rgeos", "SDMTools", "rgdal", "fgui")
new.packages <- list.of.packages[!(list.of.packages %in%
                                installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)

require(gWidgets)
options("guiToolkit"="RGtk2")

#####

#####-----
#
# Change the path for your neighbourhood type:
# - Distance.r
# - Inverse_Distance.r
# - Knn.r
#
#####-----

Distance_source="C:/Users/.../Rscripts/Distance.r"
InverseDistance_source="C:/Users/.../Rscripts/Inverse_Distance.r"
Knn_source="C:/Users/.../Rscripts/Knn.r"

#####
#
#FUNCTIONS
#
#####

Distance <- function(inputFileDir=character(0),
                    poligonConstrainFileDir=character(0),
                    outputFileDir=character(0), isWater=character(0),
                    fieldConstrainDir_u=character(0),
                    fieldConstrainDir_v=character(0), maxAng=character(0),
                    Dmax=character(0), Dmax1=character(0),
                    InvPower=character(0), Knn=character(0),

```

```

        dir=character(0), cb=character(0),
        RowStnd=character(0),...){
print(list(inputFileDir=inputFileDir,
           poligonConstrainFileDir=poligonConstrainFileDir,
           outputFileDir=outputFileDir, isWater=isWater,
           fieldConstrainDir_u=fieldConstrainDir_u,
           fieldConstrainDir_v=fieldConstrainDir_v, maxAng=maxAng,
           Dmax=Dmax,Dmaxl=Dmaxl, InvPower=InvPower,Knn=Knn, dir=dir,
           cb=cb, RowStnd=RowStnd))
}

Inverse_Distance <- function(inputFileDir=character(0),
                             poligonConstrainFileDir=character(0),
                             outputFileDir=character(0),
                             isWater=character(0),
                             fieldConstrainDir_u=character(0),
                             fieldConstrainDir_v=character(0),
                             maxAng=character(0), Dmaxl=character(0),
                             InvPower=character(0), Dmax=character(0),
                             Knn=character(0), dir=character(0),
                             cb=character(0), RowStnd=character(0),...){
print(list(inputFileDir=inputFileDir,
           poligonConstrainFileDir=poligonConstrainFileDir,
           outputFileDir=outputFileDir, isWater=isWater,
           fieldConstrainDir_u=fieldConstrainDir_u,
           fieldConstrainDir_v=fieldConstrainDir_v, maxAng=maxAng,
           Dmaxl=Dmaxl, InvPower=InvPower, Dmax=Dmax, Knn=Knn, dir=dir,
           cb=cb, RowStnd=RowStnd))
}

Knn <- function(inputFileDir=character(0),
                poligonConstrainFileDir=character(0),
                outputFileDir=character(0), isWater=character(0),
                fieldConstrainDir_u=character(0),
                fieldConstrainDir_v=character(0), maxAng=character(0),
                Knn=character(0), Dmax=character(0), Dmaxl=character(0),
                InvPower=character(0), dir=character(0), cb=character(0),
                RowStnd=character(0),...){
print(list(inputFileDir=inputFileDir,
           poligonConstrainFileDir=poligonConstrainFileDir,
           outputFileDir=outputFileDir, isWater=isWater,
           fieldConstrainDir_u=fieldConstrainDir_u,
           fieldConstrainDir_v=fieldConstrainDir_v, maxAng=maxAng,
           Knn=Knn, Dmax=Dmax, Dmaxl=Dmaxl, InvPower=InvPower, dir=dir,
           cb=cb, RowStnd=RowStnd))
}

l <- list()

#####
#
#Layout
#
#####

w <- gwindow("COSINE PACKAGE")
g <- ggroup(cont = w, horizontal = FALSE)
gl <- ggroup(horizontal = FALSE, cont = g, expand = TRUE)

##
fr0 <- gframe ( "Input/Output", cont = g, horizontal=FALSE )
ly0 <- glayout ( cont = fr0 , expand = TRUE )

```

```

##
ly0 [2,1] <- inputFileDir <- glabel("Input Feature Class:", cont = ly0)
ly0 [2,2] <- (l$inputFileDir <- gfilebrowse (text = "Select a shapefile...",
                                           type = "open", quote = FALSE,
                                           filter = list("Shapefiles" =
                                                         list(patterns =
c("*.shp"))),
                                           cont = ly0))

ly0 [3,1] <- poligonConstrainFileDir <- glabel("Input Constraint Feature
                                           Class:", cont = ly0)
ly0 [3,2] <- (l$poligonConstrainFileDir <- gfilebrowse (text = "Select a
shapefile...",
type = "open",
quote = FALSE,
filter =
list("Shapefiles" =
list(patterns =
c("*.shp"))),
cont = ly0))

ly0 [4,1] <- outputFileDir <- glabel("Output file name:", cont = ly0)
ly0 [4,2] <- (l$outputFileDir <- gfilebrowse (text = "Select a name...",
type = "save", quote = FALSE,
filter = list("Shapefiles" =
list(patterns =
c("*.shp"))),
cont = ly0))

ly0[5,1] <- (l$isWater <- gcheckbox("Water analysis? (If NO, then uncheck)",
checked = TRUE, cont=ly0))

##
fr1 <- gframe ( "", cont = g, horizontal=FALSE )
ly1 <- glayout ( cont = fr1 , expand = TRUE )

ly1 [1,1] <- glabel ( "Direction?", cont = ly1)
ly1 [1,2] <- l$dir <- gcombobox ( c("No","Yes"), cont = ly1 )

##
ly2 <- glayout ( cont = fr1 , expand = TRUE )

##
ly2 [1,1] <- fieldConstrainDir_v <- glabel("Meridional (Raster file):",
cont = ly2)
ly2 [1,2] <- (l$fieldConstrainDir_v <- gedit("0", initial.msg = "Paste the
path to the raster file (no
extensions)", cont = ly2))

ly2 [2,1] <- fieldConstrainDir_u <- glabel("Zonal (Raster file):", cont = ly2)
ly2 [2,2] <- (l$fieldConstrainDir_u <- gedit("0", initial.msg = "Paste the
path to the raster file (no
extensions)", cont = ly2))

ly21 <- glayout ( cont = fr1 , expand = TRUE )
ly21 [1,1,8] <- maxAng <- glabel("Max. Angle:", cont = ly21)
ly21 [1,8] <- (l$maxAng <- gedit("0", initial.msg = "Enter maximum angle
between a pair of observations",

```

```

                                cont = ly21, coerce.with = as.numeric))

enabled(ly2) <- FALSE
enabled(ly21) <- FALSE

#####-----
fr3 <- gframe ( "", cont = g, horizontal=FALSE )
ly3 <- glayout ( cont = fr3 , expand = TRUE )

ly3 [1,1,3] <- glabel("Neighbourhood Type: ", cont=ly3)
ly3 [1,3] <- l$cb <- gcombobox(c("Distance", "Inverse_Distance", "Knn"),
                             cont=ly3)

ly4 <- glayout(cont=fr3)
ly4[1,1,3] <- "Threshold Distance:"
ly4[1,5] <- (l$Dmax <- gedit("0",cont = ly4, coerce.with = as.numeric))

ly5 <- glayout(cont=fr3)
ly5[1,1,3] <- "Threshold Distance:"
ly5[1,5] <- (l$Dmax1 <- gedit("0",cont = ly5, coerce.with = as.numeric))

ly5[1,8] <- "Power:"
ly5[1,9] <- (l$InvPower <- gspinbutton(from=1, to=2, by=1, cont = ly5))

ly6 <- glayout(cont=fr3)
ly6[1,1,3] <- "K-Nearest Neighbours: "
ly6[1,3] <- (l$Knn <- gedit("0",cont = ly6, coerce.with = as.numeric))

ly7 <- glayout(cont=fr3)
ly7[1,1] <- (l$RowStnd <- gcheckbox("Row standardization (Optional)",
                                checked = FALSE, cont=ly7))

enabled(ly5) <- FALSE
enabled(ly6) <- FALSE

#####-----
#
#Size text
#
#####-----

## Input/Output
size ( l$inputFileDir ) <- c ( 200,20)
size ( l$poligonConstrainFileDir ) <- c ( 200,20)
size ( l$outputFileDir ) <- c ( 200,20)

## Direction
size ( l$fieldConstrainDir_v ) <- c ( 250,20)
size ( l$fieldConstrainDir_u ) <- c ( 250,20)
size ( l$maxAng ) <- c ( 30,20)

## Neighbourhood
size ( l$Dmax ) <- c ( 50,20)
size ( l$Dmax1 ) <- c ( 50,20)
size ( l$InvPower ) <- c ( 40,20)
size ( l$Knn ) <- c ( 50,20)

#####-----
#
#Buttons

```

```

#
#####-----

ly8 <- glayout(cont=g)

ly8[4,9,13] <- btnfn <- gbutton("      OK      ", cont=ly8)
ly8[4,12,16] <- cbtn <- gbutton(" Cancel ", container = ly8,
                               handler = function(j,...) dispose(w))

#####
#
#Handler options ("gray out")
#
#####

addHandlerChanged(l$dir, handler=function(h,...) {
  val <- svalue(h$obj)
  enabled(ly2) <- val == "Yes"
  enabled(ly21) <- val == "Yes"
})

addHandlerChanged(l$cb, handler=function(h,...) {
  val <- svalue(h$obj)
  enabled(ly4) <- val == "Distance"
  enabled(ly5) <- val == "Inverse_Distance"
  enabled(ly6) <- val == "Knn"
})

#####
#
# Function to run Constraint script
# (Function in another R script)
#
#####

addHandlerClicked(btnfn, handler=function(h,...) {
  lst <- sapply(l, svalue, simplify=FALSE) ## a list of values
  dispose(w)
  if (lst$cb == "Distance"){
    cat ("\nNeighbourhood Type: Distance", "\n\n")

    Distance <- capture.output(with(list(inputFileDir=lst$inputFileDir,
poligonConstrainFileDir=lst$poligonConstrainFileDir,
outputFileDir=lst$outputFileDir,
isWater=lst$isWater,

fieldConstrainDir_u=lst$fieldConstrainDir_u,

fieldConstrainDir_v=lst$fieldConstrainDir_v,
maxAng=lst$maxAng,
Dmax=lst$Dmax, RowStnd=lst$RowStnd),
source(Distance_source)))

do.call("Distance", lst)
} else if (lst$cb == "Inverse_Distance"){
  cat ("\nNeighbourhood Type: Inverse Distance", "\n\n")

  Inverse_Distance <- capture.output(with(list(inputFileDir=lst$inputFileDir,

poligonConstrainFileDir=lst$poligonConstrainFileDir,

```

```

outputFileDir=lst$outputFileDir,
isWater=lst$isWater,
    fieldConstrainDir_u=lst$fieldConstrainDir_u,
    fieldConstrainDir_v=lst$fieldConstrainDir_v,
    maxAng=lst$maxAng, Dmax1=lst$Dmax1,
    InvPower=lst$InvPower,
    RowStnd=lst$RowStnd),
    source(InverseDistance_source)))
do.call("Inverse_Distance", lst)
}else if (lst$cb == "Knn"){
  cat ("\nNeighbourhood Type: Knn", "\n\n")

  Knn <- capture.output(with(list(inputFileDir=lst$inputFileDir,
poligonConstrainFileDir=lst$poligonConstrainFileDir,
    outputFileDir=lst$outputFileDir,
    isWater=lst$isWater,

fieldConstrainDir_u=lst$fieldConstrainDir_u,

fieldConstrainDir_v=lst$fieldConstrainDir_v,
    maxAng=lst$maxAng,
    Knn=lst$Knn, RowStnd=lst$RowStnd),
    source(Knn_source)))
  do.call("Knn", lst)
}
)

```

Neighbourhood Type: Distance

```

rm(list = ls(all = TRUE)) #CLEAR ALL

#-----
# Load libraries to run R-code
#-----
require(maptools)
require(raster)
require(spdep)
require(sp)
require(geoR)
require(igraph)
require(spatstat)
require(gtools)
require(rgeos)
require(SDMTools)
require(rgdal)
require(fgui)

#*****
# First section is to look for the path and establish the working directory.
# From there:
# 'getDirNFile' will look for the path (regular windows path) and remove
# '.shp' extension
#*****
getDirNFile <- function(inputFileDir ){
  #GET AND SET ROOT DIR AND FILE DIR

```

```

DIR=unlist(strsplit(inputFileDir, "", fixed = TRUE))
TokenPos=1:length(DIR)
TokenPosMatch=TokenPos[DIR=="\\"]
rootDir=paste0(DIR[1:TokenPosMatch[length(TokenPosMatch)]], collapse="")
fileDir=paste0(DIR[(TokenPosMatch[length(TokenPosMatch)]+1):(length(DIR)-4)],
              collapse="") #remove '.shp' extension from the path
setwd(rootDir) #setup a working directory (i.e. where data are saved)

return(c(rootDir, fileDir))
}

#*****
# First section is to look for the path and establish the working directory.
# From there:
# 'getDirNFile1' will look for the path (regular windows path) to the raster
# file (in this case a folder from windows explorer perspective), it will not
# remove any extension
#*****
getDirNFile1 <- function(inputFileDir1 ){
  #GET AND SET ROOT DIR AND FILE DIR
  DIR=unlist(strsplit(inputFileDir1, "", fixed = TRUE))
  TokenPos=1:length(DIR)
  TokenPosMatch=TokenPos[DIR=="\\"]
  rootDir=paste0(DIR[1:TokenPosMatch[length(TokenPosMatch)]], collapse="")
  fileDir=paste0(DIR[(TokenPosMatch[length(TokenPosMatch)]+1):length(DIR)],
                collapse="")
  setwd(rootDir) #setup a working directory (i.e. where data are saved)

  return(c(rootDir, fileDir))
}

Distance <- function(inputFileDir,                               ##Input Feature Class
                    poligonConstrainFileDir,                 ##(Shape file)
                    outputDir,                                ##Input Constraint
                    isWater,                                   ##Feature
                    fieldConstrainDir_u,                       ##Class (Shape file)
                    fieldConstrainDir_v,                       ##Output File - Distance
                    maxAng,                                    ##Water or Land feature
                    Dmax,                                       ##analysis
                    RowStnd,                                    ##Input Direction File
                    Dmax1, InvPower, Knn, dir, cb               ##(Raster file - Zonal)
                    ){                                          ##Input Direction File
  ##(Raster file -
  ##Meridional)
  ##Select Max. Angle
  ##Select Max. Distance
  ##Row standardization
  ##unused argument(s)

  #*****
  # LOAD GIS Data
  #*****
  td <- "." # This option will separate name and extension apart

  # -*/-*/-*/-*/-*/-*/-*/-*/
  # LOAD SIGHTINGS
  # -*/-*/-*/-*/-*/-*/-*/-*/
  fileDir=getDirNFile(inputFileDir)[2]
  sightings <- readOGR(td, fileDir)

  # -*/-*/-*/-*/-*/-*/-*/-*/

```



```

names(OUTPUT)[2] <- paste("NID")
names(OUTPUT)[3] <- paste("WEIGHT")
}

#####-----
#   SAVE OUTPUT FILE
# -----
#Distance shapefile and csv
#####-----
fileDir=getDirNFile1(outputFileDir)[2]
writeOGR(OUTPUT, "Nb", fileDir, driver="ESRI Shapefile")

cat("\nWeight Matrix:",fileDir, "(shp file), has been created", "\n")

cat("\nNeighbourhood Type: Distance - completed", "\n")

cat("\nMinimum distance to have at least one neighbours is:", max_k1,
     "\nSmaller distances may contain 'islands'", "\n")

gc(TRUE) #- get rid of trash >.<

return(OUTPUT)
}

```

Neighbourhood Type: Inverse Distance

```

rm(list = ls(all = TRUE))#CLEAR ALL

#-----
# Load libraries to run R-code
#-----
require(maptools)
require(raster)
require(spdep)
require(sp)
require(geoR)
require(igraph)
require(spatstat)
require(gtools)
require(rgeos)
require(SDMTools)
require(rgdal)
require(fgui)

#####
# First section is to look for the path and establish the working directory.
# From there:
# 'getDirNFile' will look for the path (regular windows path) and remove
# '.shp' extension
#####
getDirNFile <- function(inputFileDir ){
  #GET AND SET ROOT DIR AND FILE DIR
  DIR=unlist(strsplit(inputFileDir,"",fixed = TRUE))
  TokenPos=1:length(DIR)
  TokenPosMatch=TokenPos[DIR=="\\"]
  rootDir=paste0(DIR[1:TokenPosMatch[length(TokenPosMatch)]],collapse="")
  fileDir=paste0(DIR[(TokenPosMatch[length(TokenPosMatch)]+1):(length(DIR)-4)],

```

```

collapse="") #remove '.shp' extension from the path
setwd(rootDir) #setup a working directory (i.e. where data are saved)

return(c(rootDir,fileDir))
}

#*****
# First section is to look for the path and establish the working directory.
# From there:
# 'getDirNFile1' will look for the path (regular windows path) to the raster
# file (in this case a folder from windows explorer perspective), it will not
# remove any extension
#*****
getDirNFile1 <- function(inputFileDir1 ){
  #GET AND SET ROOT DIR AND FILE DIR
  DIR=unlist(strsplit(inputFileDir1,"",fixed = TRUE))
  TokenPos=1:length(DIR)
  TokenPosMatch=TokenPos[DIR=="\\"]
  rootDir=paste0(DIR[1:TokenPosMatch[length(TokenPosMatch)]],collapse="")
  fileDir=paste0(DIR[(TokenPosMatch[length(TokenPosMatch)]+1):length(DIR)],
collapse="")
  setwd(rootDir) #setup a working directory (i.e. where data are saved)

  return(c(rootDir,fileDir))
}

Inverse_Distance <- function(inputFileDir, ##Input Feature Class
                             ##(Shape file)
                             poligonConstrainFileDir, ##Input Constraint
                             ##Feature Class (Shape
                             ##file)
                             outputFileDir, ##Output File -
                             ##Distance
                             isWater, ##Water or Land
                             ##feature analysis
                             fieldConstrainDir_u, ##Input Direction File
                             ##(Raster file - Zonal)
                             fieldConstrainDir_v, ##Input Direction File
                             ##(Raster file -
                             ##Meridional)
                             maxAng, ##Select Max. Angle
                             Dmax1, ##Select Max. Distance
                             ##for Inverse Distance
                             InvPower, ##Select Inverse
                             ##Distance
                             ##power (1,2)
                             RowStnd, ##Row standardization
                             Dmax, Knn, dir, cb ##unused argument(s),
                             ##only works for GUI
){
  #*****
  # LOAD GIS Data
  #*****
  td <- "." # This option will separate name and extension apart

  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  # LOAD SIGHTINGS
  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  fileDir=getDirNFile(inputFileDir)[2]
  sightings <- readOGR(td, fileDir)

  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  # LOAD CONSTRAINED

```



```

#   SAVE OUTPUT FILE
# -----
#Distance shapefile and csv
####-----
fileDir=getDirNFile1(outputFileDir)[2]
writeOGR(OUTPUT, "Nb", fileDir, driver="ESRI Shapefile")

cat("\nWeight Matrix:",fileDir, "(shp file), has been created", "\n")

cat("\nNeighbourhood Type: Inverse Distance - completed", "\n")

gc(TRUE) #- get rid of trash >.<

return(OUTPUT)
}

```

Neighbourhood Type: K-Nearest Neighbour (Knn)

```

rm(list = ls(all = TRUE))#CLEAR ALL

#-----
# Load libraries to run R-code
#-----
require(maptools)
require(raster)
require(spdep)
require(sp)
require(geor)
require(igraph)
require(spatstat)
require(gtools)
require(rgeos)
require(SDMTools)
require(rgdal)
require(fgui)

#####
# First section is to look for the path and establish the working directory.
# From there:
# 'getDirNFile' will look for the path (regular windows path) and remove
# '.shp' extension
#####
getDirNFile <- function(inputFileDir ){
  #GET AND SET ROOT DIR AND FILE DIR
  DIR=unlist(strsplit(inputFileDir,"",fixed = TRUE))
  TokenPos=1:length(DIR)
  TokenPosMatch=TokenPos[DIR=="\\"]
  rootDir=paste0(DIR[1:TokenPosMatch[length(TokenPosMatch)]],collapse="")
  fileDir=paste0(DIR[(TokenPosMatch[length(TokenPosMatch)]+1):(length(DIR)-4)],
    collapse="") #remove '.shp' extension from the path
  setwd(rootDir) #setup a working directory (i.e. where data are saved)

  return(c(rootDir,fileDir))
}

#####
# First section is to look for the path and establish the working directory.
# From there:
# 'getDirNFile1' will look for the path (regular windows path) to the raster
# file (in this case a folder from windows explorer perspective), it will not

```

```

# remove any extension
#*****
getDirNFile1 <- function(inputFileDir1 ){
  #GET AND SET ROOT DIR AND FILE DIR
  DIR=unlist(strsplit(inputFileDir1,"",fixed = TRUE))
  TokenPos=1:length(DIR)
  TokenPosMatch=TokenPos[DIR=="\\"]
  rootDir=paste0(DIR[1:TokenPosMatch[length(TokenPosMatch)]],collapse="")
  fileDir=paste0(DIR[(TokenPosMatch[length(TokenPosMatch)]+1):length(DIR)],
    collapse="")
  setwd(rootDir) #setup a working directory (i.e. where data are saved)

  return(c(rootDir,fileDir))
}

Knn <- function(inputFileDir,           ##Input Feature Class
                poligonConstrainFileDir, ##(Shape file)
                outputFileDir,          ##Input Constraint Feature
                isWater,                ##Class (Shape file)
                fieldConstrainDir_u,    ##Output File - KNN
                fieldConstrainDir_v,    ##Water or Land feature
                maxAng,                 ##analysis
                Knn,                    ##Input Direction File
                RowStnd,                ##(Raster file - Zonal)
                Dmax, Dmax1, InvPower, dir, cb ##Input Direction File
                ) {                    ##(Raster file - Meridional)
                                        ##Select Max. Angle
                                        ##Select Number of Nearest
                                        ##Neighbours (KNN)
                                        ##Row standardization
                                        ##unused argument(s),
                                        ##only works for GUI

  #*****
  # LOAD GIS Data
  #*****
  td <- "." # This option will separate name and extension apart

  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  # LOAD SIGHTINGS
  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  fileDir=getDirNFile(inputFileDir)[2]
  sightings <- readOGR(td, fileDir)

  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  # LOAD CONSTRAINED
  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  fileDir=getDirNFile(poligonConstrainFileDir)[2]
  land <- readOGR(td, fileDir)

  #*****
  # CREATE NET
  #*****

  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  # CREATE ADJACENCY MATRIX
  # -*/-*/-*/-*/-*/-*/-*/-*/-*/-*/
  sn=length(sightings)
  # everything connected to everything
  adjMat=matrix(1,nrow=sn,ncol=sn);

```



```

names(OUTPUT_3)[4] <- paste("DIST")

OUTPUT_3$WG=1

#-----
# ROW STANDARDIZED
#-----
if (RowStnd==TRUE) {

  NeighborClusters=unique(OUTPUT_3$NODE1)
  CLUSTER=matrix(0,nrow=length(OUTPUT_3$NODE1),ncol=1)
  CLUSTER_D_W=matrix(0,nrow=length(OUTPUT_3$NODE1),ncol=1)
  ones=matrix(1,nrow=length(OUTPUT_3$NODE1),ncol=1)
  # FOR EACH CLUSTER WEIGHT DISTANCE
  for(i in 1:length(NeighborClusters)){
    clusterIndex=OUTPUT_3$NODE1==NeighborClusters[i]
    CLUSTER[clusterIndex]=i
    CLUSTER_D_W[clusterIndex]=OUTPUT_3$DIST[clusterIndex]/
      sum(OUTPUT_3$DIST[clusterIndex])

  }
  # Insert columns
  OUTPUT_3$CLU=c(CLUSTER)
  OUTPUT_3$WEIGHT=c(CLUSTER_D_W)
  OUTPUT_3$CLU_D_W_I=c(ones-CLUSTER_D_W)

  OUTPUT=OUTPUT_3[c("NODE1","NODE2","WEIGHT")]
  names(OUTPUT)[1] <- paste("OBJECTID")
  names(OUTPUT)[2] <- paste("NID")

  cat("\nRow standardization computed", "\n")

} else {

  OUTPUT=OUTPUT_3[c("NODE1","NODE2","WG")]
  names(OUTPUT)[1] <- paste("OBJECTID")
  names(OUTPUT)[2] <- paste("NID")
  names(OUTPUT)[3] <- paste("WEIGHT")

}

#####-----
# SAVE OUTPUT FILE
# -----
#Distance shapefile and csv
#####-----
fileDir=getDirNFile1(outputFileDir)[2]
writeOGR(OUTPUT, "Nb", fileDir, driver="ESRI Shapefile")

cat("\nWeight Matrix:",fileDir, "(shp file), has been created", "\n")

cat("\nNeighbourhood Type: Knn - completed", "\n")

gc(TRUE) #- get rid of trash >.<

return(OUTPUT)
}

```