EXTRACTING PATTERNS FROM LARGE MOVEMENT DATA SETS USING HYBRID SPATIOTEMPORAL FILTERING:

A CASE STUDY OF GEOVISUAL ANALYTICS IN SUPPORT OF FISHERIES ENFORCEMENT ACTIVITIES

RENÉ A. ENGUEHARD

Extracting patterns from large movement data sets using Hybrid Spatio-temporal Filtering: A case study of geovisual analytics in support of fisheries enforcement activities

by

© René A. Enguehard

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Geography

Memorial University of Newfoundland

November 2011

St. John's Newfoundland

Abstract

The ubiquitous nature of location tracking technologies has resulted in an increase in movement data being collected. These data are used in many contexts, such as understanding animal migration, aiding in fisheries enforcement, or managing fleets of taxicabs. Such large volumes of data call for more efficient data visualization and analysis methods. This research provides a general approach to the analysis of movement data, named Hybrid Spatio-temporal Filtering (HSF), which allows analysts to filter data based on characteristics of movement within a geovisual analytics environment. Filtering signatures are defined by combining movement path complexity (fractal dimension) and velocity, to extract behavioural patterns from data sets. An evaluation within a fisheries enforcement case study (using VMS data), and comparison to other approaches, confirmed the approach is useful, easy to use, and superior to some other approaches. This research demonstrates the value of signature-building filtering approaches for large movement data sets.

Acknowledgements

I would like to extend thanks to my co-supervisors, Dr. Rodolphe Devillers and Dr. Orland Hoeber, for having supported me throughout this project. Their insights and comments, as well as moral support during the extent of my masters program were invaluable to its success. They helped guide my thesis topic, manage the scale of what I wanted to accomplish, and encouraged me to look at problems from various perspectives. Their responsiveness allowed me to never be stuck on any particular problem. I am also truly thankful to them for having put up with my constant struggle in the correct use of 'that' and 'which'.

Acknowledgements also go out to Jerry Black, Trevor Fradsham, Wanda Arsenault, and their respective groups at Fisheries and Ocean Canada (DFO). Without their involvement in this project, in terms of expertise, time, and goodwill, this thesis would not have been possible. Of equal importance is the involvement of all the fisheries enforcement officers who participated in the field trials. Their enthusiasm and insightful comments were crucial to the success of this project. The amount of time that they devoted to this study, particularly considering their busy schedules, was very much appreciated.

Thanks are also extended to DFO itself, for having provided the data that was used in this project, as well as the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI), and Memorial University of Newfoundland for their financial or in-kind support.

Many people reviewed and commented on multiple parts of this thesis, making it the document it is today. My sincere thanks are extended to Randal Greene, Cassandra Lee, Garnett Wilson, and Krista Jones for their valuable input. I would also like to thank Eli Gurarie for helping me understand the inner-workings of his Behavioural Change Point Analysis (BCPA) method, and providing both code and feedback via email.

My friends and family also played a crucial role in keeping me somewhat sane throughout this endeavour. For their constant encouragement, and putting up with my short fuse towards the end, I thank them. My colleagues in the Marine Geomatics Research Lab, the User Experience Lab, and the Geovisual Analytics Research Group, provided a sounding-board for ideas and fresh perspectives. Finally, the staff and faculty of the Department of Geography played a crucial role in ensuring that everything went smoothly throughout my masters program.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1. Context and problem

Understanding physical and cultural features often involves the ability to represent these features, locate them within a given coordinate system, and sometimes record their changes or movements through time (Langran, 1992). The collection and analysis of spatio-temporal data, such as GPS tracking data, population migration data, or socio-economic data, has become an integral part of many decision-making processes (Andrienko et al., 2007b; Tomaszewski et al., 2007). In particular, governments, as well as public and private organizations, have become increasingly reliant on movement data, or data about how tracked objects change through space and time. These data are routinely used to analyze traffic flows (Andrienko et al., 2007a; Chen et al., 2011; Willems et al., 2009), manage infrastructure (Bomberger et al., 2006; Mano et al., 2010), manage fleets of vehicles (Jeung et al., 2010; Lundblad et al., 2008; Pfoser et al., 2005), or better understand animals or ecosystems (Bertrand et al., 2007; Focardi et al., 1996; Mårell et al., 2002; Nams, 2005; With, 1994). Even individuals now have access to large amounts of personal movement data, through mobile phone localization, or services such as Google Latitude, FourSquare, or image geotagging (Eagle & Pentland, 2009; Hollenstein & Purves, 2010).

These data can provide a rich source of information that can help understand complex processes inherent to movement (Andrienko et al., 2008). Of particular interest are behaviours and patterns. Behaviours are defined as being "the configuration of characteristics corresponding to a given reference (sub)set" (Andrienko et al., 2008). In other words, a behaviour is a set of characteristics that, when used to filter a data set, consistently isolates related subsets of data (patterns). Andrienko & Andrienko (2007) place behaviours into three sub-categories: Individual Movement Behaviours (IMB), Momentary Collective Behaviours (MCB), and Dynamic Collective Behaviours (DCB). The sub-categories can also be viewed as a hierarchy, with IMB focusing on individual behaviours, such as transit between locations, MCB focusing on the behaviour of a pre-determined set of individuals, such as co-location of two related individuals, and DCB focusing on global behaviours, such as migration. These categories can provide an effective guide in defining a particular behaviour, and therefore how to detect it.

In contrast to behaviours, Andrienko et al. (2008) define patterns as "representations of behaviour in some language, e.g. natural, mathematical, graphical". In this sense, a pattern can be thought of as a single representation of a behaviour, with many patterns potentially representing the same behaviour and one pattern being potentially composed of multiple simpler patterns. Viewed another way, Dodge et al. (2008) define patterns as "any recognizable spatial and temporal regularity or any interesting relationship in a set of movement data". These are then divided into two sub-categories: generic patterns, such as dispersion or symmetry, and behavioural patterns, such as foraging or migration. Again, these sub-categories can be viewed as a hierarchy, with behavioural patterns being composed of generic patterns. Conceptually, generic patterns can be related back to the concept of IMB, whereas behavioural patterns are composed of MCBs or DCBs. Therefore, to understand and detect behaviours, it is crucial to start with generic patterns.

One of the issues when working with movement data is that the data sets can be quite large, and many of the traditional analysis methods do not scale well to these sizes (Andrienko & Andrienko, 2007; Jern et al., 2008). As a result, analysts trying to make sense of those data often have to sift through a large amount of data with inefficient methods, leading to potentially valuable information being missed, due to information overload. This is compounded by the representational issues associated with large amounts of data, wherein viewers can get into "needle in a haystack" types of situations (Keim et al., 2004; Ware, 2004).

By focusing on behaviours and patterns, one can begin to analyze phenomena such as group dynamics (Andersson et al., 2008; Jeung et al., 2010), temporal cycles (Andrienko et al., 2008; Eagle & Pentland, 2009; Wood et al., 2007), movement patterns (Demšar & Virrantaus, 2010; Kwan, 2000; Murawski et al., 2005), or attraction/repulsion dynamics (Gottfried, 2011), while filtering out much of the data which are not relevant to the analyst. Studying behaviours and patterns can also provide considerable insight into the external phenomena driving behaviour, as well as helping in their prediction (Andrienko & Andrienko, 2007). For instance, a better understanding of what external phenomena cause "road rage" behaviours could lead to the elaboration of a predictive model, helping urban planners that want to prevent these types of behaviours.

Many approaches have been proposed to gain a better understanding of these movement processes, with some inspired from biological behaviours. Optimal Foraging Theory (OFT), for instance, attempts to model the way in which predatory animals foraging for food might behave (Bartumeus & Catalan, 2009; Charnov, 1976). It does this by looking at the energy balance between energy spent foraging against energy derived from the prey. This energy balance often results in a form of correlated random walk model which exhibits fractal properties, such as a Lévy flight (Mårell et al., 2002). Deviations from the optimal foraging behaviour can then be analyzed to provide insight into the particular behaviour being exhibited by the animal.

Another approach is that of representing the movement data to promote its visual analysis. This is typically addressed by either traditional cartographic approaches, interactive geovisualization approaches, or automated clustering approaches, each having their benefits and drawbacks. Cartographic approaches such as Hägerstand's space-time cube (Hägerstrand, 1970), which plots two-dimensions of space against time within a three-dimensional cube, allow the viewer to quickly understand how movements and interaction took place, both through space and time. However, using traditional approaches the viewer cannot directly manipulate the data, unlike interactive geovisualization techniques (Andrienko et al., 2007a; Kraak, 2003; Turdukulov et al., 2007; Wood et al., 2007; Zhao et al., 2008). These allow for more specific questions to be studied through visual representations and interactive filtering and highlighting. However, they also require more time and effort than automated clustering techniques, such as Self Organizing Maps (SOMs) (Choi et al., 2006; Koua & M.-J. Kraak, 2004). Automated clustering enables rapid analysis of large data sets, but with a high computational cost and limited customizability or transparency.

Movement data can also be analyzed using purely mathematical modeling (Franke et al., 2004), statistical (Gurarie et al., 2009; Underwood & Chapman, 1985), or data-mining approaches (Li et al., 2006). These approaches analyze the data and report results without necessarily having to visualize the data set. In cases where the results are visualized, usually only the extracted patterns are represented. For very large data sets, this is an obvious advantage, as is the primarily algorithmic nature of these approaches. However, this can also lead to analysts missing specific aspects of behaviour, or anomalous patterns, which they may have noticed had the data been visually represented. These approaches also suffer from a lack of transparency, in that analysts may not be able to identify the effects of the automated processes on their data.

Most of these proposed approaches adopt a particular perspective, be it biological, geovisual, or mathematical. However, few approaches combine these multiple perspectives, to achieve a more integrated and holistic approach. Integrated approaches may help analysts deal with the very large amount of data that are often associated with movement data sets. Particularly, combining foraging theory (Bartumeus & Catalan, 2009) and complex filtering in an interactive geovisual analytics environment (Ho & Jern, 2008; Johansson & Jern, 2007; Lundblad et al., 2008; Tomaszewski et al., 2007) can provide a generalized hybrid approach to analyzing these large movement data sets.

Further, many of the proposed approaches in this domain have not been evaluated in terms of usability or usefulness, nor have they been compared to one another in any meaningful sense. Doing so could provide considerable insight into the situations where one approach may be superior to others. It may also identify potentials for improvement or integration of various approaches.

1.2. Questions and hypothesis

The primary research questions to addressed in this thesis are:

- Which movement characteristics can be used to build signatures that identify specific behaviours?
- How can a geovisual analytics environment be designed to effectively allow for visual exploration of large movement data sets?
- How can a geovisual analytics environment be designed to maximize usability among analysts?
- Does the approach developed improve analysts' ability to extract movement patterns from their data sets?
- How does the proposed approach compare in terms of usability and effectiveness with existing approaches?

The research hypothesis is that a geovisual analytics system allowing filtering on multiple characteristics of movement will improve analysts' abilities to both deal with large amounts of movement data and find interesting patterns within them.

1.3. Goal and objectives

The goal of this research is to elaborate, implement, and test a novel hybrid approach to the analysis of movement data, combining the characteristics of movement in a geovisual analytics environment, as well as studying the situations where this method, and others, may be most suitable for use.

To attain this goal, the specific research objectives of this thesis are:

1. Study characteristics of movement that can be used to filter the data, and their relation to movement patterns and exhibited behaviours.
2. Design an efficient geovisual analytics environment for the visual exploration of large movement data sets.
3. Design a generic approach to complex filtering of movement data such that specific movement pattern signatures can be elaborated.
4. Implement the approach using a prototype software system.
5. Validate the usability and usefulness of the approach, using fisheries enforcement as a case study.
6. Compare the approach to other approaches within this same fisheries enforcement case study setting.

1.4. Methods

The research methodology followed by this thesis is summarized in Figure 1.1, and has proceeded in a generally linear fashion, from identification of research questions, to design, implementation, validation, and finally comparison. Literature review and communication with experts, the research community, and various other interested parties was ongoing throughout the project. The information gained from these interactions, and from reviewing existing works, informed every aspect of the project and helped in identifying directions. The application of the concepts into a prototype system and the validation process both generated issues that required communication with experts. As a result, there was continual interaction between the practical and theoretical aspects of this work.

The initial literature review helped identify the current state knowledge in fields related to movement data visualization and analysis. The fields of behavioural ecology, pattern detection, and geovisual analytics, provided information as to how to support the analysis of large movement data sets. From these, the combination of fractal dimension, to characterize movement path complexity, and velocity was hypothesized to lend itself to the detection of specific behaviours. This, combined with the concepts of interactive filtering, signature building, and geovisualization, composed the core of the design.

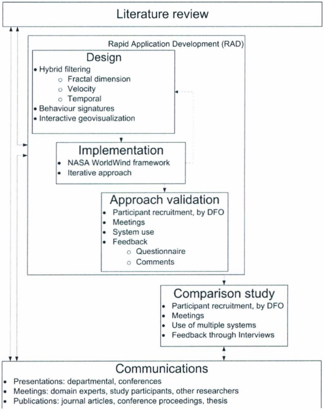


Figure 1.1. Flow of research and related methods.

The design was then implemented on top of an existing virtual globe geovisualization system, i.e. World Wind. An iterative approach similar to that of Rapid Application Development (RAD) (McConnell, 1996) was followed, with design and implementation proceeding feature by feature. Features were implemented from a user-centered viewpoint, ensuring that each element of the implementation was usable and understandable by the target user group, in this case data analysts.

The final phase of the RAD process was the validation of the approach implemented by the prototype system. This was achieved by the use of a field trial method applied to a fisheries enforcement case study. Fisheries and Oceans Canada (DFO), a partner in the larger project this thesis falls under, were asked to select a number of fisheries enforcement officers which were familiar with the analysis of vessel movement data. The data used for this project were provided by DFO, and extracted from their Vessel Monitoring System (VMS) database for the year 2009. It covered all of Atlantic Canada, from the Gulf of Saint Lawrence to the edge of the Grand Banks of Newfoundland, and from the Southern coast of Nova Scotia to Northern Labrador.

The field trials were conducted in both St. John's, NL (April, 2011), and Dartmouth, NS (June, 2011), on an individual basis. In total, nine enforcement officers participated in the approach validation. Officers were asked some background information to assess experience levels, used the prototype system, and then answered a questionnaire. This questionnaire, based on the Technology Acceptance Model (TAM) (Davis, 1989), yielded quantitative and qualitative feedback as to the usefulness and ease of use of the system.

In addition, a second field trial scheduled in parallel to the Dartmouth field trial, allowed comparison of the proposed approach to other approaches. The purpose of this field trial was to gather qualitative information as to the usefulness of the proposed approach relative to other approaches. Specifically, the comparison between an existing traditional web-mapping system, an automated approach, and the proposed approach were investigated. This was done in order to assess how the proposed approach compares to other alternatives in the context of our case study.

The communication of the research occurred throughout the project, similarly to the literature review. Multiple preliminary meetings were held with DFO, to discuss project ideas, present some early research prototypes, and to set up the field trials. The initial concept for this research was presented in the form of a research proposal to the Department of Geography (April, 2010). The main approach and research prototype were presented by the candidate at the GeoViz 2011 Workshop in Hamburg, Germany (March, 2011), and subsequently at a research seminar at the Department of Geography (April, 2011). Early findings about the comparison of the approach with other systems were presented at the Maritime Anomaly Detection (MAD) Workshop in Tilburg, Netherlands (June, 2011). During these presentations, many interesting discussions were generated with other researchers, highlighting useful literature and suggesting potential improvements that could be made to the overall approach.

1.5. Thesis organization

This thesis uses a manuscript format, with chapters two and three being journal papers that have been submitted to international peer-review journals. Chapter two details the elaboration of the proposed approach, including a literature review, a description of the approach, as well as its field trial validation, and has been submitted to the journal *Information Visualization*. It describes a novel approach for visualizing and analyzing movement data, called Hybrid Spatio-temporal Filtering (HSF), that exploits the physical, fractal, and temporal characteristics of movement to isolate specific types of behaviours. These filtering settings compose a behavioural signature, which can be combined with other signatures, or re-used on different data. The field trial validation of the approach showed both the usefulness and ease of use of this approach.

Chapter three focuses on the comparison of the HSF approach to two other existing approaches, and has been submitted to the *Journal of Ocean and Coastal Management*. Specifically, a currently used analysis system (DFO's VUE system), an automated system implementing Behavioural Change Point Analysis (BCPA), and the interactive approach of HSF, were compared by means of a field trial. DFO's experts were asked to use each system and provide feedback regarding their individual benefits and limitations, as well as any possibilities for combining approaches in specific contexts. This real-world comparison of the usefulness of the HSF approach, in comparison to other methods, showed that this approach is superior to existing and recently developed approaches, and could be applied to a number of different domains. It also identified a balance of parameters (ease-of-use, transparency, functionality, and speed of analysis) that are critical in the usefulness and acceptance of approaches in specific domains.

Chapter four summarizes how each of the research questions have been addressed and whether the hypotheses were validated. In addition, the chapter highlights the main contributions of this thesis, and explores future opportunities for research.

1.6. Co-authorship statement

This project is part of a broader Natural Sciences and Engineering Research Council of Canada (NSERC) Strategic Projects Grant (STPGP 365189-08), headed by Dr. Orland Hoeber, in Computer Science, and involving Dr. Rodolphe Devillers, in Geography, among other researchers. Fisheries and Oceans Canada (DFO) are industry partners on the grant. Early meetings with DFO identified a number of potential research projects that would be of interest to them, one of these being related to the geovisualization of complex fisheries movement data. While the general topic was constrained to dealing with fisheries data, the candidate independently decided to research the combination of fractal dimension, velocity, temporal filtering, and geovisualization, to design a flexible geovisual analytics approach to analyzing movement data. These general concepts were agreed upon by both co-supervisors, and they were formalized through a thesis proposal.

The practical aspects of the research, including literature review, design of the specific approach, attending meetings with DFO officials, development of the prototype system, chairing of the field trials in both St. John's and Dartmouth, were undertaken by the candidate. An application for joint ethical review of the field trials associated with this work, along with another project under the umbrella of the larger grant was submitted by a post-doctoral researcher working on the other project. The candidate designed the field trial methodology used in this thesis, in collaboration with my co-supervisors, conducted the field trials, and also executed the data compilation and analysis of the field trial results. This included both the quantitative results of the first round of field trials, and the interview transcriptions and analysis of the second round of field trials.

The two journal articles included in this thesis, chapters two and three, were initially developed as outlines by the candidate. These were later modified upon receiving recommendations from both co-supervisors, to make certain that the division between both papers was acceptable and that their content would fit with the targeted journals. The first drafts of each paper were written by the candidate, after which point the revisions proceeded in an iterative review process, wherein the candidate revised the manuscripts based on the comments of the co-supervisors. The candidate is the primary author on both papers, with both co-supervisors being co-authors. The candidate is also the author of this manuscript, integrating the papers into a coherent thesis.

1.7. References

Andersson, M., Gudmundsson, J., Laube, P., & Wolle, T. (2008). Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, *12*(4), 497-528. doi:10.1007/s10707-007-0037-9

Andrienko, G., Andrienko, N., & Wrobel, S. (2007a). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46. doi:10.1145/1345448.1345455

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D. A., Kraak, M.-J., MacEachren, Alan M., & Wrobel, S. (2007b). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839-857. doi:10.1080/13658810701349011

Andrienko, N., & Andrienko, G. (2007). Designing visual analytics methods for massive collections of movement data. *Cartographica*, *42*(2), 117-138. doi:10.3138/carto.42.2.117

Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic concepts of movement data. In F. Giannotti & D. Pedreschi (Eds.), *Mobility Data Mining and Privacy* (pp. 15-38). Springer. doi:10.1007/978-3-540-75177-9

Bartumeus, F., & Catalan, J. (2009). Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, *42*(43), 1-12. doi:10.1088/1751-8113/42/43/434002

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., & Gerlotto, F. (2007). Scale-invariant movements of fishermen: the same foraging strategy as natural predators. *Ecological Applications*, *17*(2), 331-337.

Bomberger, N., Rhodes, B., Seibert, M., & Waxman, A. (2006). Associative learning of vessel motion patterns for maritime situation awareness. *Proceedings of the 9th International Conference on Information Fusion* (pp. 1-8). Florence, Italy: IEEE. doi:10.1109/ICIF.2006.301661

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129-136.

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. *Proceedings of the 27th International Conference on Data Engineering* (Vol. 4, pp. 900-911). Hannover, Germany: IEEE. doi:10.1109/ICDE.2011.5767890

Choi, K.-H., Kim, J.-S., Kim, Y. S., Yoo, M.-A., & Chon, T.-S. (2006). Pattern detection of movement behaviors in genotype variation of *Drosophila melanogaster* by using self-organizing map. *Ecological Informatics*, *1*(3), 219-228. doi:10.1016/j.ecoinf.2005.12.002

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems*, *13*(3), 319-340.

Demšar, U., & Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, *24*(10), 1527-1542. doi:10.1080/13658816.2010.511223

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, *7*(3-4), 240-252. doi:10.1057/palgrave.ivs.9500182

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, *63*(7), 1057-1066. doi:10.1007/s00265-009-0739-0

Focardi, S., Marcellini, P., & Montanaro, P. (1996). Do ungulates exhibit a food density threshold ? A field study of optimal foraging and movement patterns. *The Journal of Animal Ecology*, *65*(5), 606-620. doi:10.2307/5740

Franke, A., Caelli, T., & Hudson, R. J. (2004). Analysis of movements and behavior of caribou (*Rangifer tarandus*) using hidden Markov models. *Ecological Modelling*, *173*(2-3), 259-270. doi:10.1016/j.ecolmodel.2003.06.004

Gottfried, B. (2011). Interpreting motion events of pairs of moving objects. *GeoInformatica*, *15*(2), 247-271. doi:10.1007/s10707-009-0095-2

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395-408. doi:10.1111/j.1461-0248.2009.01293.x

Ho, Q., & Jern, M. (2008). Exploratory 3D geovisual analytics. *Proceedings of the 2008 International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies* (Vol. d, pp. 276-283). Ho Chi Minh City, Vietnam: IEEE. doi:10.1109/RIVF.2008.4586367

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*, *1*(1), 21-48. doi:10.5311/JOSIS.2010.1.3

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, *24*(1), 6–21.

Jern, M., Åström, T., & Johansson, S. (2008). GeoAnalytics tools applied to large geospatial datasets. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 362-372). Columbus, USA: IEEE. doi:10.1109/IV.2008.27

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., & Shen, H. T. (2010). Discovery of convoys in trajectory databases. *Proceedings of the 34th International Conference on Very Large Data Bases* (pp. 1068-1080). Auckland, New Zealand: VLDB Endowment.

Johansson, S., & Jern, M. (2007). GeoAnalytics visual inquiry and filtering tools in parallel coordinates plots. *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems* (pp. 1-8). Seattle, USA: ACM Press. doi:10.1145/1341012.1341055

Keim, D. A., Panse, C., Sips, M., & North, S. C. (2004). Visual data mining in large geospatial point sets. *IEEE Computer Graphics and Applications*, *24*(5), 36–44. IEEE. doi:10.1109/MCG.2004.41

Koua, E. L., & Kraak, M.-J. (2004). A usability framework for the design and evaluation of an exploratory geovisualization environment. *Proceedings of the 8th International Conference on Information Visualisation* (pp. 153-158). Austin, USA: IEEE. doi:10.1109/IV.2004.1320138

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference* (pp. 1988–1995). Durban, South Africa: International Cartographic Association.

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, *8*(1-6), 185-203. doi:10.1016/S0968-090X(00)00017-6

Langran, G. (1992). *Time in Geographical Information Systems* (p. 180). London, UK: CRC Press.

Li, X., Han, J., & Kim, S. (2006). Motion-Alert : automatic anomaly detection in massive moving objects. *Proceedings of the 2006 IEEE Intelligence and Security Informatics Conference* (pp. 166-177). San Diego, USA: IEEE.

Lundblad, P., Jern, M., & Forsell, C. (2008). Voyage analysis applied to geovisual analytics. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 381-388). Columbus, USA: IEEE. doi:10.1109/IV.2008.32

Mano, J.-P., Georgé, J.-P., & Gleizes, M.-P. (2010). Adaptive multi-agent system for multi-sensor maritime surveillance. *Advances in Practical Applications of Agents and Multiagent Systems*, *70*, 285–290. doi:10.1007/978-3-642-12384-9_34

McConnell, S. M. (1996). *Rapid development: Taming wild software schedules* (1st ed., p. 680). Redmond, USA: Microsoft Press.

Murawski, S. A., Wigley, S. E., Fogarty, M. J., Rago, P. J., & Mountain, D. G. (2005). Effort distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of Marine Science*, *62*(6), 1150-1167. doi:10.1016/j.icesjms.2005.04.005

Mårell, A., Ball, J. P., & Hofgaard, A. (2002). Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology*, *80*(5), 854-865. doi:10.1139/Z02-061

Nams, V. O. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia*, *143*(2), 179-88. doi:10.1007/s00442-004-1804-z

Pfoser, D., Tryfona, N., & Jensen, C. S. (2005). Indeterminacy and spatiotemporal data: basic definitions and case study. *GeoInformatica*, *9*(3), 211-236. doi:10.1007/s10707-005-1282-4

Tomaszewski, B. M., Robinson, A. C., Weaver, C., Stryker, M., & MacEachren, A.M. (2007). Geovisual analytics and crisis management. *Proceedings of the 4th International Information Systems for Crisis Response and Management (ISCRAM) Conference* (pp. 1-8). Delft, Netherlands.

Turdukulov, U. D., Kraak, M.-J., & Blok, C. A. (2007). Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds. *Computers & Graphics*, *31*(3), 370-379. doi:10.1016/j.cag.2007.01.028

Underwood, A. J., & Chapman, M. G. (1985). Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology*, *91*(1-2), 17–43. Elsevier. doi:10.1016/0022-0981(85)90218-7

Ware, C. (2004). *Information visualization: perception for design* (2nd ed., p. 486). San Francisco, USA: Morgan Kaufmann.

Willems, N., van de Wetering, H., & van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics Forum*, *28*(3), 959-966. doi:10.1111/j.1467-8659.2009.01440.x

With, K. A. (1994). Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology*, *9*(1), 25-36. doi:10.1007/BF00135076

Wood, J., Dykes, J. A., Slingsby, A., & Clarke, K. (2007). Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1176-83. doi:10.1109/TVCG.2007.70570

Zhao, J., Forer, P., & Harvey, A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, *7*(3-4), 198-209. doi:10.1057/palgrave.ivs.9500184

Chapter 2 Interactive exploration of movement data: A case study of geovisual analytics for fishing vessel analysis

Abstract

The analysis of large movement data sets is a challenging task, due to their size and complexity. This paper presents an interactive geovisual analytics approach named Hybrid Spatio-temporal Filtering (HSF) that integrates filtering of multiple movement characteristics, geovisualization, and multiple coordinated views to enable analysts to focus on movement patterns that are of interest. This study proposes a novel technique that combines the fractal dimension and velocity of movement paths to effectively filter out uninteresting records, through an iterative signature-building process. In addition, the fractal dimension estimation is performed using a moving-window technique, which allows investigations at multiple temporal scales. These tools are used in conjunction with a probability-based zonal incursion tool to visually represent when the movement nears areas of interest, and help identify specific types of behaviors. These features were built to deal with data sets having low and uneven sample rates. Field trials with fishing vessel movement data illustrate the utility of the interactive features and visual representations of the movement patterns. Within a geovisual analytics framework, the approach allows analysts to explore large movement data sets easily and efficiently. The combination of velocity, fractal dimension, and temporal filtering helped analysts effectively identify subsets of data that conformed to particular behavioral patterns of interest.

2.1. Introduction

The analysis of large movement data sets is a challenging task, due to the size and complexity of these data sets (Andrienko et al., 2007a; Dykes & Mountain, 2003; Kraak & van de Vlag, 2007). However, such analyses can provide useful insights into the behavior of moving targets, helping analysts to identify trends, patterns, or outliers within the data. While movement data can be visualized in a non-spatial manner (Eagle & Pentland, 2009; Underwood & Chapman, 1985; Zhao et al., 2008), geovisualization systems can provide effective ways for dealing with various aspects of the complexity of movement data. Many systems have been proposed in the past, with varying degrees of success (Andrienko et al., 2007a; Eagle & Pentland, 2009; Höferlin et al., 2011; Jern & Franzen, 2007; Kwan, 2000). None, however, have exploited the movement's complexity and physical bounds to reduce visual complexity and provide insight into the data.

To this end, this research focuses on the design and evaluation of a method that helps reduce the amount of data an analyst needs to investigate. The Hybrid Spatio-temporal Filtering (HSF) system presented provides analysts with an interface that employs a novel filtering mechanism and uses visual representations to address the inherent complexity of the data. The approach taken exploits physical characteristics of the data (i.e. velocity, heading), temporal constraints, and the complexity of the movement path, quantified using fractal dimension, to increase the effectiveness of an interactive filtering system. An interactive geovisualization interface is provided to graphically show the aspects of the data that match the filter, as well as to illustrate the possible extent of travel between the logged locations.

Although others have explored methods for filtering movement data using velocity and time, a novel aspect of HSF is the use of fractal dimension to filter the data. Fractal dimension gives a quantitative measure of how much an object, or a movement path, fills its theoretical space (Mandelbrot, 1967; Theiler, 1990). It is essentially a numerical measure of path complexity, sometimes referred to as tortuosity. Fractal dimension is scale invariant, which affords it robustness to movements of different scales and gaps in data sets. This allows the comparison of movements that differ not only in scale but also in geometry.

A hybrid filtering feature using physical characteristics and fractal dimension, together with temporal and object of interest information, provides analysts with a high degree of control on how the data are filtered. One of the reasons for combining the physical characteristics of movement with fractal dimension is that physical characteristics of movement are very sensitive to errors or omissions in the data, while the fractal dimension estimation is not. However, physical characteristics are easier to understand as they have a direct analogue in the real world, whereas the concept of fractal dimension is more abstract.

The iterative modification of filter settings to extract specific types of patterns allows analysts to build signatures to match specific movement behaviors that are being sought. For instance, transit and migration patterns, where objects are moving from one place to another in a straight line, would tend to have a high velocity and a low fractal dimension. In contrast, foraging and similar searching patterns would manifest themselves as fairly complex lines, tending towards a lower velocity and higher fractal dimension. Multiple different signatures may also be used together, in order to identify differences or similarities in spatial or temporal distributions, for example.

Due to the large size of movement data sets, implementing this approach in a visual fashion requires an interactive geovisual analytics system with support for large spatio-temporal data sets. HSF uses a virtual globe representation, with individual data points displayed as chevron glyphs and lines connecting subsequent records. This glyph-based representation lends itself well to filtering, and is straightforward to understand. The virtual globe allows for spatio-temporal data from anywhere in the world to be represented, without having to deal with projection zones or distortion when zooming in to local views or out to global views.

To test this approach, a prototype implementation of HSF was evaluated through a case study using Vessel Monitoring System (VMS) movement data collected in 2009 from vessels fishing in the Northwest Atlantic region. These data were provided by Fisheries and Oceans Canada (DFO). The evaluation was conducted as a field trial using experts at DFO whose primary duties include fisheries enforcement. The purpose of this evaluation was to establish whether the experts found this geovisual analytics approach to data analysis useful and easy to use, in comparison to their current practices.

2.2. Movement data complexities

There exist a number of inherent complexities associated with movement data, which do not necessarily exist with other types of data (Kraak & van de Vlag, 2007; Rodighiero, 2010). Generally speaking, movement data are composed of a latitude, longitude, and timestamp, with ancillary data such as velocity, heading, and altitude commonly included. The temporal resolution of these data, or the amount of time between collecting data points, plays an important role in how the data can be used. For instance, data collected on an hourly basis will not lend themselves to the same type of analyses as data collected every 10 seconds. The spatial accuracies of the data also play an important role as they relate to the spatial and temporal scales at which data can be analyzed. Similarly, temporal accuracy can play an important role in the accuracy of a data set. For instance, some positioning systems used to locate birds derive the geographic coordinates from the apparent elevation of the sun, which acts as a proxy for time (Schaefer & Fuller, 2006). Significant inaccuracies in these time estimates would lead to large inaccuracies in reported positions.

Added to these factors is the volume of movement data collected. Temporal resolution directly affects the amount of data that are recorded, since for equal spans of time, doubling the temporal resolution of the data collection system doubles the amount of data collected. Moreover, while increasing the rate of data acquisition may be useful in some cases, it can have drawbacks when trying to analyze data at certain scales. However, these are not the only factors to consider. For instance, the amount of ancillary data recorded, as well as the number of objects of interest being tracked, increases the volume of data to manage and visualize.

Other important issues to consider are the systematic and random errors within the data. Certain positioning systems, such as GPS, may not always be able to acquire a position (e.g. not enough satellites in view). In such circumstances, they may systematically send out a default position, such as where the system was initially calibrated, or an impossible position (e.g. 95°N). Added to these are random errors, such as when the system is unable to record a data point due to computer malfunctions or transmission errors. These errors may appear as random gaps within a data set and can complicate analyses that depend on regular sampling intervals.

2.3. Related works

Visualization of movement data has been an active topic of research and innovation for hundreds of years, with famous examples such as Charles Minard's 1869 flow map of Napoleon's Russian campaign of 1812 (Tufte, 2001). Like most flow maps, Minard's map relied on cartographic generalization and data summarization to help the viewer make sense of the data. While this works well for static mapping of small data sets, it requires either the cartographer to understand the entire data set before it can be generalized, or the use of automated generalization tools. Due to the size of today's movement data sets, that can often range into multi-gigabyte or terabyte sizes, full understanding of data sets often requires too much time and manual effort to be practical; automated tools have become a necessity (Andrienko et al., 2008a). Moreover, the threshold of what is considered a "large" data set is likely to increase over time, given a commensurate increase in computer processing power and data storage density. As a result, there is a continuing need for progressively better techniques to deal with large data sets.

Some of the more recent key works exploring the visualization of movement data include Hägerstrand's space-time cube, which represents two-dimensional movements through time in a three-dimensional cube (Hägerstrand, 1970). This method has been expanded and enhanced in a number of ways, such as standardizing the paths to a common origin (Kwan, 2000), adding activity or pattern classes (Ren & Kwan, 2007), and integrating the concept of the space-time prism (Kraak, 2003). These incremental improvements, while beneficial, have not overcome one of the limitations of this approach: it relies on direct depiction of data, rather than summarization or pattern extraction (Andrienko et al., 2008a). As a result, analysts are often overwhelmed by the large volume of complex data they need to deal with. Direct depiction, while faster, only offers an overview of the entire data set, while cartographic summarization of data sets of this size can be difficult. For this reason, some form of computer-aided synthesis, through categorization or pattern extraction, is an attractive option.

Approaches for condensing movement data to its most important components has been explored in a number of different ways (Andrienko et al., 2007a; Kwan, 2000; Rinzivillo et al., 2008; Willems et al., 2009). It can be achieved using, for instance, data classification or clustering techniques. Identifying clusters or patterns within the data and merging similar ones can greatly reduce the amount of data to display (Andrienko et al., 2008a). This identification of movement clusters can follow various approaches, including the identification of behavioral patterns (Andersson et al., 2008; Dodge et al., 2008), movement characteristics (Gottfried, 2011; Pelot & Wu, 2007), or levels of traffic density (Laxhammar et al., 2009; Willems et al., 2009).

The visualization of these clusters of movement data, and the merging of similar clusters, can then happen either automatically (Andrienko et al., 2007a; Gurarie et al., 2009; Rinzivillo et al., 2008), or manually (Andrienko et al., 2011; Pelot & Wu, 2007). Both approaches have relative strengths and weaknesses. For instance, automated pattern detection reduces the amount of work required for an analyst to acquire knowledge, but also removes a certain amount of control, which is retained with interactive summarization techniques (Andrienko et al., 2008a). Interactive techniques allow for a greater exploration of the data, since they do not rely on default settings, but also require more time and work to acquire useful knowledge.

Much research has been done within the biology community to analyze and understand movement data. Tracking animals using devices such as collars or archival tags is a common procedure to help understand or model migration patterns and population distributions (Focardi et al., 1996; Franke et al., 2004). Optimal foraging theory (Bartumeus & Catalan, 2009; Charnov, 1976), correlated random walks (Mårell et al., 2002), or other techniques (Choi et al., 2006; Franke et al., 2004; Mårell et al., 2002), are methods that were developed to extract useful knowledge from the patterns with such data. However, movement data can also be analyzed in other ways, such as using fractal dimension to investigate whether animals of different sizes perceive landscapes at different scales (Nams, 2005; With, 1994). It has also been shown that these same techniques can be applied to some human movements such as fishing vessels, as they behave similarly to natural predators (Bertrand et al., 2007).

The concept of fractal dimension was originally presented by Mandelbrot in 1967 (Mandelbrot, 1967), as 'fractional dimension', which essentially describes the self-similarity, or complexity, within features under study (e.g. lines, polygons, polyhedra). In the context of movement data, it allows for the differentiation between simple straight-line movements, such as walking from a house to a bus stop, to complex jagged movement, such as searching for someone in a crowd. It can be thought of as the ratio between the actual distance covered by a movement path and the straight-line distance between its start and end points.

Calculating the exact fractal dimension value for a given feature requires that it be defined mathematically. When dealing with data that were collected through some form of measurement or sampling, this approach is not feasible. However, it is possible to estimate the fractal dimension of features using a variety of methods, such as using the correlation dimension or box-counting dimension (Theiler, 1990). Many of these methods, however, suffer from performance issues related to their reliance on mathematical differentiation. As a result, a number of methods have been recently proposed to address this issue (Füchslin et al., 2001; Sevcik, 1998). These advances, coupled with increases in processing power, make the estimation of fractal dimension from large movement data sets possible in near real-time.

Another approach for dealing with large volumes of movement data is that of anomaly detection. A number of different techniques can be used to detect anomalies in movement patterns, such as rule-based systems (Li et al., 2006), statistical methods (Laxhammar et al., 2009), or pre-determined criteria sets (Sage, 2005). The common feature among these approaches is the specification of what are considered 'normal' behaviors, thus allowing the extraction of instances of behaviors that do not fit these profiles.

Such an approach of building up of profiles or models for movement patterns is not limited to anomaly detection. It can be used to predict where and when tracked targets will be or have been (Bomberger et al., 2006). The accurate modeling of the movements of vehicles, animals, or people, using data from position recording system, can be quite useful due to the insight provided into what might happen between two consecutive data points (Chang et al., 2010). This additional information can help make tasks such as estimation of fishing effort (Deng et al., 2005; Mills et al., 2006; Murawski et al., 2005), species modeling (Mullowney & Dawe, 2009), or the visualization of the data more accurate (Lundblad et al., 2008; Rodighiero, 2010).

This paper presents the HSF approach and the associated complexities for using a visual approach to analyzing the data, such as hybrid filtering and interactivity. The HSF prototype system can be considered an example of a geovisual analytics system. Geovisual analytics is the sub-field of visual analytics that deals with the challenges of exploring and analyzing spatio-temporal data (Andrienko et al., 2007b; Thomas & Cook, 2005). The main goal of visual analytics is to promote the visual analysis of data to acquire some form of knowledge which could otherwise be missed by visual inspection or entirely automated approaches. It relies on interactive representations of the data, as well as comprehensive data analysis tools, to provide a rich environment for users to explore and manipulate complex data sets, and to support decision-making.

2.4. Hybrid Spatio-Temporal Filtering (HSF)

The method proposed in this paper takes a geovisual analytics approach to solving the problems of visualizing and analyzing large movement data sets that may have low and un-even sample rates. Interactive filtering, multiple coordinated views, and details on demand give considerable analytical power to this approach. It allows for the identification of specific types of movement behaviors, as well as other types of features of interest, such as group dynamics, congested areas, and temporal cycles (e.g. day-night cycles or seasonal cycles).

Formally, HSF is the approach of integrating velocity, fractal dimension, and temporal filtering, into an interactive geovisualization environment, supported by multiple coordinated views, to aid in the analysis of large movement data sets. Particularly, the integration of velocity and fractal dimension filtering allows for the elaboration of specific movement pattern 'signatures' based on filter settings, which can be used to highlight different patterns within the data.

Movement data typically do not incorporate ancillary parameters, such as velocity. As a result, the HSF approach does not rely on these data being present, and proceeds to an automated estimation of velocity values, based on the straight-line distance between two subsequent data points. By using the ratio of the distance between two subsequent data points and the amount of time elapsed between these, an average velocity is estimated (Equation 1). These velocities are then represented visually as a histogram. This both helps analysts understand the distribution of velocity values within a data set, and select the range of velocities that represent some behavior in which the analyst is interested.

$$\vec{v} = \frac{\Delta d}{\Delta t} \qquad (1)$$

Fractal dimension estimation can be accomplished using a method presented by Sevcik (Sevcik, 1998). Fractal dimension, as a measure of path complexity, has been shown to be useful in identifying the types of movement behaviors an individual is undertaking (Bartumeus & Catalan, 2009; Bertrand et al., 2007; Mårell et al., 2002). The identification of such behavioral patterns is important for analysts in that it allows them to focus on specific types of activities, or sets of related activities, and provides them with an idea of their spatial distribution.

The fractal dimension measure is a real number based on the inherent dimensionality of the data, such as one dimensional for a straight line, or two dimensional for a complex vessel path on a plane. As such, the fractal dimension (D) for two-dimensional movement data can range from 1.0 to 2.0, with 1.0 being a perfectly straight line and 2.0 being a path so complex that it covers the entire two-dimensional plane (Figure 2.1). It can be estimated using a ratio of the unit-square total path length to the number of data points (Equation 2).

$$D_{Sevcik} = 1 + \frac{\log(L)}{\log(2N)} \qquad (2)$$

An interesting property of fractal dimension is that it is scale invariant. This means that any two paths with geometrical similarity have a similar fractal dimension value, regardless of the spatial dimension they occupy. For instance, two moving targets that are exhibiting similar paths, but with one going twice as fast or twice as far as the other, and thus covering twice the distance, would still have the same fractal dimension value. This is obviously a desirable property when attempting to define generalized filters to identify specific types of behaviors. Another advantage of fractal dimension filtering is that it is insensitive to gaps in the data, as the geometry of the paths is what is being used to estimate the fractal dimension, rather than individual data point locations. This is an important feature when used with data that includes a significant number of dropped data points or errors.

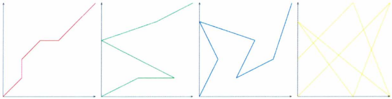


Figure 2.1. A visual representation of movement paths of varying complexities. Scale and direction of travel are irrelevant when estimating fractal dimension. From left to right, fractal dimension = 1.161, 1.393, 1.423, and 1.721.

Estimation of fractal dimension is typically done over full data sets, in order to increase accuracy. While this works well for identifying prevailing behaviors over large data sets, this makes it unsuitable for estimating changes in behavior over time. In order to support filtering of the data, a modification of the original method (Sevcik, 1998) was done, by allowing multiple sub-sets of the data to be evaluated and compared, using a moving window. This moving-window fractal dimension, with a variable window size, allows the estimation of fractal dimension to be tailored to specific types of behaviors. For instance, if an analyst knows that a specific behavior occurs over a span of two to four hours, a window size of three hours can be used. The alternative of using the full path would automatically smooth out these small-scale behaviors, making them difficult to discover. Using the moving window technique, it is also possible to locate sub-behaviors within large-scale behaviors, filtering out those that do not match this pattern.

Finally, temporal filtering is achieved through the plotting of data blocks (trips) along a time-line. Analysts can then visually interpret sequences of data, making temporal patterns such as periodicity, sequence, or synchronization, easier to detect. This also allows analysts to quickly identify those subsets of the data in which they may be interested, filtering out the rest of the data.

These three filtering modes, velocity, fractal dimension, and temporal, can function separately or in conjunction with one another in the HSF approach. Analysts can select filter settings to isolate specific behaviors, based on the velocity and complexity characteristics inherent to the patterns that make up those behaviors. Multiple behavioral signatures can be used simultaneously, to explore phenomena such as transitions from one behavior to the other, behavioral anisotropy, or related behaviors.

2.5. Prototype system

The above Hybrid Spatio-temporal Filtering (HSF) approach has been implemented within a prototype geovisual analytics system, using fisheries enforcement as a case study in order to test its feasibility and effectiveness. As a result, the prototype focused on Vessel Monitoring System (VMS) data produced by fishing vessels. The prototype includes two major components: the geovisualization component, and the interactive hybrid filtering component, integrated through multiple coordinated views (Figure 2.2). The geovisualization component represents the target movement paths and includes a feature wherein ellipses can be displayed for each set of data points to indicate a probability of incursion into zones of interest. HSF integrates three types of filtering (velocity, fractal, and temporal), and allows for the creation of reloadable signatures based on the velocity and fractal dimension filter settings.

Multiple coordinated views (Wang Baldonado et al., 2000) are used to allow changes in one of these components to be automatically reflected in all others. This effectively enables users to investigate multiple aspects of their movement data set at the same time, which is beneficial not only for the analysis of the data, but also during the construction of filtering signatures for patterns of interest.

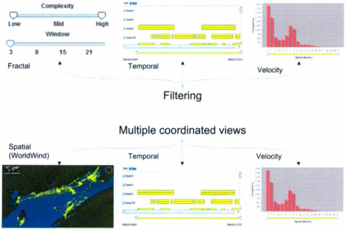


Figure 2.2. Complex filtering based on velocity, fractal, and temporal properties is used to isolate patterns; multiple coordinated views of the data support the iterative visual exploration of the data by analysts.

2.5.1. Geovisualization component

The geovisualization component used as the foundation of the HSF system is based on the Java version of NASA World Wind, an open-source 3D mapping system similar to Google Earth (Figure 2.3). This system allows for the standard view manipulation operations available in most three-dimensional mapping software, such as pan, tilt, and zoom. Additionally, HSF extends this system to support a number of other modes of interaction, such as providing hover-based details on demand regarding a particular target, and the highlighting of complete target paths via waypoint selection.

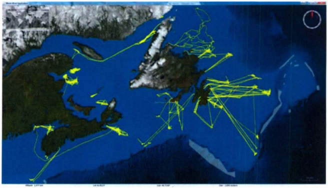


Figure 2.3. World Wind-based virtual globe displaying a month of fishing vessel movement paths in Eastern Canada.

The data points themselves are displayed using chevron-shaped glyphs for each record (Figure 2.4). Chevrons were chosen since they imply the directionality of movement, which helps users understand the temporal flow within a data set. The set of chevrons representing the path of one vessel are linked using lines, to provide analysts a general idea of the areas potentially travelled by the vessel.

Due to the temporal resolution of the VMS data used in our case study (one hour), simple straight lines produce a significant error component (Bertrand et al., 2007; Tremblay et al., 2006). As a result, a cubic Hermite spline interpolation method was used, which generates smooth curves of higher accuracy, and are assured to go through every actual data point (Tremblay et al., 2006). An accuracy assessment was run on both linear and cubic Hermite spline interpolation for our specific data, to verify that the results were consistent with those presented by Tremblay et al. (2006), in that cubic Hermite spline interpolation significantly decreases the error of the path estimation. The secondary benefit of using this spline interpolation technique is that curved paths are easy for the human eye to follow, leading to less cognitive strain on the users (Ware, 2004).

Both the chevrons and the lines are filled with a solid yellow color. Yellow was chosen primarily due to its high contrast with the blue background color of the water. This use of yellow on a blue background can readily be pre-attentively processed by viewers, allowing the paths to be identified without conscious attention (Hering, 1964; Ware, 2004). When complete paths are selected for highlighting, these are colored in cyan, which provides a chromatic contrast with the yellow paths, as well as a luminance contrast with the blue background.

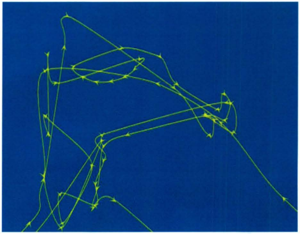


Figure 2.4. Example of fishing vessel path: the yellow curves represent the interpolated vessel tracks and the chevrons indicate the position of collected data points as well as the direction of movement.

The geovisualization component implemented includes an option of displaying semi-transparent ellipses (Figure 2.5) estimating the potential area travelled by a vessel between two consecutive data points. Due to the known temporal resolution of the data and the maximum velocity of each vessel, it is possible to estimate the maximum spatial area covered by a vessel. This visual tool can then be used to inform the analyst as to whether it is possible for a vessel to have entered into a particular zone. In the case of this study, we used zones that are closed to fishing, and therefore where most fishing vessels should not venture. The ellipses are calculated by finding the maximum recorded vessel velocity for each individual vessel within the data set. The maximum recorded vessel velocity is then increased by 10%, to account for a broad range of factors, such as tail winds, a calm sea, or maximum throttle, that could provide a higher maximum velocity than the data would suggest. This provides a conservative estimate of the true maximum velocity of a vessel, since the exact figure is generally unavailable.

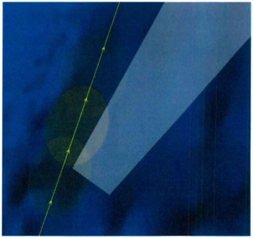


Figure 2.5. Ellipses showing a vessel track with a low probability of zonal incursion, shown by the opacity of the ellipses and their amount of overlap with a specific zone (shown here in grey).

This velocity is then used to calculate the maximum distance that could be travelled within the span of time between each pair of data points. The ratio of the distance between each pair of data points and the maximum distance that could be travelled determine the sizes of the semi-major and semi-minor axes of the ellipses, using the method described by Formula 1-3 and Figure 2 in Mills et al. (2006). In other words, the farther apart two data points, the closer the vessel is to its maximum velocity, and therefore the smaller its semi-minor axis.

Once the ellipses are calculated, the area of spatial overlap of each ellipse onto each zone is calculated. The ratio of this area to the total area of each ellipse gives a probability of incursion into each particular zone. For the special case where either of the data points is actually within a zone, the probability of incursion is set to 100%. These incursion probabilities are represented in the geovisualization view using a slightly darker color than the path line. This allows the visual association of each ellipse to each path, while allowing for increased contrast between the chevrons, paths, and ellipses. Finally, the opacity of each ellipse is set to their percent probability of incursion, with multiple overlapping ellipses adding their opacities (Figure 2.5).

2.5.2. Interactive filtering

Velocity filtering was implemented as a separate component in our prototype system. It supports the visualization of the distribution of vessel velocities within a data set, through the use of a histogram, as well as filtering of data based on a velocity range. Such information can be interesting for the analysts, as fishing vessels will display specific ranges of speed for different activities (e.g. steaming to fishing grounds, trawling). The histogram representation is also coordinated with the rest of the system, so that removal of data through other components causes the recalculation of the histogram; likewise, the filtering of data using velocity filtering causes the redrawing of the remaining data on the other components.

The velocity value for each data point is estimated by dividing the straight-line distance between each data point and its successor, and dividing by the interval of time between their timestamps. While splines were selected for the display paths between data points, performance limitations prevent them from being used to estimate velocity. With smaller data sets or faster processors, using spline interpolation would be preferable in order to increase estimation accuracy.

Once the velocities are estimated for each data point, they are grouped into one nautical mile per hour (one knot) bins and represented using a histogram. This histogram provides the analyst with an overview of the velocity-based behavior of a group of fishing vessels (Figure 2.6). This frequency distribution is generally bi-modal, due to vessels spending much time at high velocity (going to and from the fishing grounds), and much time at low velocity (fishing). This histogram can therefore provide some insight into which velocity values may provide a filter for a meaningful subset of the data. Such a filter can be set by specifying upper and lower bounds for the velocity, resulting in a removal of all movement data that does not meet this criterion from within the geovisualization component.

The fractal filtering settings operate in a similar manner to velocity filtering (Figure 2.7). However, to make the fractal filtering process more understandable to analysts that may not have a solid understanding of fractal dimension, the filter values are labeled from 'low complexity' (D=1.0) to 'high complexity' (D=2.0). The settings of the window size for the fractal dimension filtering can easily be manipulated, the effect of which can be observed in (Figure 2.8).

Temporal filtering has also been implemented through upper and lower bounds controls. In order to illustrate the temporal features of the movement data, a temporal visualization system was developed to allow analysts to see the temporal distribution of trips for individual vessels, as well as the groups of vessels (Figure 2.9). Due to the number of individual vessels within the data set, they cannot all be visualized at once, so vertical scrolling functionality is provided.

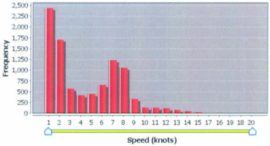


Figure 2.6. Velocity filtering histogram showing the number of data points for each velocity value. The filter at the bottom allows for the selection of a subset of data, based on velocity thresholds.

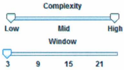


Figure 2.7. Fractal filter selection using a set of sliders controlling the complexity level (top) and window size (bottom).

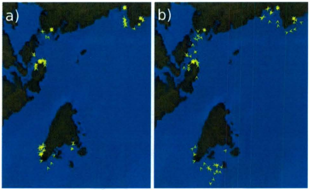


Figure 2.8. Data points filtered using fractal filtering, with 1.25 <D<1.5 and window sizes of (a) three data points and (b) 25 data points.

Trip blocks are estimated by finding gaps of over six hours within a vessel's individual data set. It was assumed that if a vessel stopped transmitting its position for over six hours, it was highly probable that it was in port, as doing so at sea would be a serious and easy to identify infraction of the fishing regulations. Selecting a shorter length, such as two hours, was found to yield too many false positives, due to the aforementioned missing data points. Similarly, selecting a longer length, such as 12 hours, was found to merge some trips together where the stop in port was shorter than 12 hours.

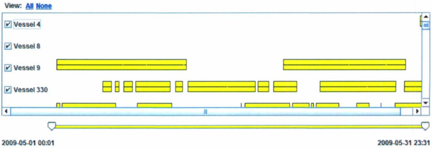


Figure 2.9. Temporal visualization view with individual vessels listed vertically and trip blocks displayed horizontally in yellow. The time slider allows for the filtering of the data based on temporal thresholds.

Each trip block is also linked into the geovisualization view, such that analysts may click on a trip block and have it highlighted in both the temporal and geovisualization views. The blocks were also provided with a continuity display, showing any interruptions in data recording as gaps within a black line that runs through the block. The advantage of this technique is that larger gaps are automatically more visible to users, while the habitual one-hour gaps are almost invisible. Finally, the functionality for filtering individual vessels was provided by a simple checkbox next to each vessel identifier.

2.5.3. Signature building

One of the novel interactive features of this system is the ability to build signatures for particular patterns and behaviors based on a combination of velocity and fractal dimension filters. The signature building interface (Figure 2.10), allows users to create and modify multiple concurrent pattern signatures, in an iterative and unstructure fashion, and then have them displayed in the geovisualization view using distinct colors.

Additionally, each signature can be turned on and off, so that analysts can focus on a subset of signatures as required by the specific data analysis or exploration task.

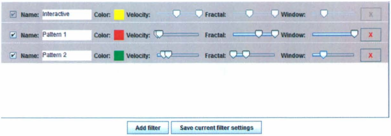


Figure 2.10. Signature building and management interface which allows users to extract multiple patterns of interest.

Each signature is built in an iterative fashion by the analysts, through modification of the interactive filter. The interactive filter is always visible and colored yellow to prevent confusion. Once the analyst has configured an interactive filter setting that isolates a particular pattern or behavior of interest, it can be added to the signature management view, given a name for future use, and assigned a color for easy identification in the geovisualization view.

The geovisualization view colors each data point according to its membership in either the interactive filter or each user-defined signature, with the interactive filter taking precedence where ties occur. In some cases, it may happen that a data point is within the filter parameters of two or more signatures. This leads to a situation where a data point could be colored a number of different ways. To resolve this issue, it was decided that for each signature its two-dimensional centroid would be calculated using the velocity and fractal dimension ranges as axes. When a data point could belong to two different signatures due to the signatures overlapping, the distance from the data point to the centroid of each one is calculated, with the shortest distance taking precedence. The result of this scheme can be seen in Figure 2.11.

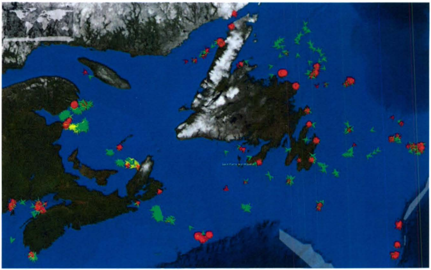


Figure 2.11. Visualization of three concurrent signatures based on different settings for velocity and fractal dimension. Green, red, and yellow are associated with these different movement patterns.

2.6. Case study evaluation

The evaluation of the prototype system was done in the context of fisheries enforcement. Fisheries enforcement officers working for the Canadian government monitor fishing vessel movements on a daily basis. This is done to better understand where vessels fish, as well as to help identify vessels that are not following the rules and regulations surrounding the fishery. To do this, enforcement officers rely on VMS data, which is GPS data produced and transmitted in this case by fishing vessels on an hourly basis. These data consist of a vessel identifier, latitude, longitude, and timestamp. The monthly aggregate datasets often include more than a million records, with over two thousand vessels to monitor, spanning an area of over four million square kilometers, covering most of the Northwest Atlantic. In addition to these data, fisheries enforcement officers often rely on ancillary data sources, such as vessel characteristics, licenses, and history. Due to privacy issues, these ancillary data were not available to participants during this case study, and were not included as part of the overall approach.

The case study evaluation was performed through direct use of the prototype system by domain experts. The field trial methodology followed used a small amount of participants, nine in this case, who are domain experts and potential users of the prototype software. These types of field trials produce qualitative data, rather than quantitative data, and help assess the potential usefulness and usability of new techniques (Shneiderman & Plaisant, 2006).

The field trials occurred in Fisheries and Oceans Canada's offices at two different locations: St. John's, NL, and Halifax, NS. The participants were fisheries enforcement officers who had experience with the spatio-temporal data generated by VMS units, as well as the patterns and behaviors of fishing vessels conducting both legal and illegal activities.

2.6.1. Field trial methodology

The field trials were divided into three phases. First, a pre-study questionnaire was administered, to obtain information such as the participant's experience level with VMS data, their familiarity with geovisualization systems, and how many years of enforcement experience they had. This was followed by the system use phase, where participants were invited to use the prototype system with actual VMS data, to perform open-ended data analysis tasks of their choosing, similar to those they would perform during the course of their regular enforcement-based data analysis of VMS data. The length of the system use phase was left up to each participant, with an upper limit of 1.5 hours. This time limit was imposed primarily due to the need for the scheduling of sessions with experts during their regular work hours.

Finally, a post-study questionnaire was administered to each participant. This questionnaire was based on the Technology Acceptance Model (TAM) (Davis, 1989), which provides measures of the perceived ease of use (PEU) and perceived usefulness (PU) of software systems. Responses to the statements were measured using a five-point Likert scale, ranging from 'strongly disagree' to 'strongly agree'. A 'not applicable' option was also provided. In addition to questions relating to PEU and PU, a number of questions were provided to gauge the relative benefits of this approach against the current data analysis techniques used by the experts, as well as a section where participants were encouraged to provide open-ended comments.

2.6.2. Hypothesis

The hypothesis guiding this field trial was that the proposed approach of providing velocity, fractal dimension, and temporal filtering within a complex interactive geovisualization interface would be beneficial in terms of the ability to find anomalous or interesting patterns by expert users, reducing the effort required to sift through large amounts of data. This reduction in effort could be further promoted in situations where the analyst could select from a set of pre-existing signatures. These benefits should become apparent in the measurements of PU, the responses to statements regarding the specific features, and the open-ended comments. Further, it was expected that the ease of use of this system would be similar to the current system used by enforcement officers (DFO's VUE system), resulting in moderate scores for PEU.

2.6.3. Participants

The participants in the field trials were fisheries enforcement officers who were required to have experience in identifying fishing vessel patterns, an understanding of the use of VMS in fisheries enforcement, and familiarity with visualizing these data. Nine participants were purposefully sampled (i.e. chosen by their respective manager), according to their knowledge of vessel movement data, and their familiarity with the VUE system and existing practice for analyzing such data.

2.7. Evaluation results

2.7.1. Pre-study questionnaire

The pre-study questionnaire provided some insight into the composition of the participant pool for this study. The median experience of the participants in fisheries enforcement was 14 years. All but one participant had experience with spatio-temporal data, specifically with VMS data. The one participant that did not was in a management position and had extensive experience in the policy and regulatory aspects of fisheries enforcement.

Of the participants who did have experience with VMS data, most had worked with these in a number of different capacities, for multiple separate projects. Moreover, the questionnaires suggested that participants had a medium level of familiarity with geovisualization and virtual globes, and medium to high level of understanding of Northwest Atlantic fisheries and fisheries infractions.

Among the participants, two main groups were identified. One group (five participants), had an average of 10 years experience with fisheries enforcement and less experience with geovisualization. The other group (four participants) was composed of more experienced individuals, who had spent an average of 30 years working with fisheries enforcement and fishing vessel data. However, responses to the post-questionnaires and use of the prototypes by these groups of participants suggested no differences between the two groups, as their responses did not appear to differ.

2.7.2. Observations

Participants were encouraged to use the prototype system in a manner consistent with their normal VMS data analysis activities. Some participants chose to perform the kinds of exploratory analyses they perform regularly, focusing on their particular management region. Others decided to look for behaviors on a more global level, evaluating the effectiveness of the prototype system at discerning specific movement patterns. A few participants were able to find previously unknown anomalous movement patterns, which they noted and followed up upon.

One such pattern was readily extracted both visually and through the use of a low velocity/high fractal dimension filter, with a large (13+ hour) window size. The analyst could not initially identify what type of vessel this was, and found that the behavior was not at all consistent with the rest of the fishing behaviors in the same zone. Using the extra information contained in the VUE system, such as vessel names, licenses, and characteristics, the vessel was identified as a dredger (underwater excavator) which the analyst was previously unaware of.

2.7.3. Post-study questionnaire

Once the participants decided to stop using the prototype system, they were asked a series of questions relating to a number of different categories, specifically their PU (Figure 2.12), PEU (Figure 2.13), and the overall usefulness of specific components of the system (Figure 2.14). They were also asked to determine which system (HSF or VUE) they preferred for a set of specific analysis tasks (Figure 2.15).

The responses for the perceived usefulness of HSF were very consistent among participants, with the exception of participant four (Figure 2.12). This participant, which works in a more management-oriented capacity, declined to answer this section, citing that 'working with these systems are not part of my regular duties'. As such, the participant felt unable to accurately gauge the usefulness of this approach. Excluding this response set, six out of eight participants either agreed or strongly agreed that HSF was useful, with the two remaining participants leaning towards neutral or agreeing. The responses for the perceived ease-of-use section were similar to that of perceived usefulness (Figure 2.13). Seven out of nine participants either agreed or strongly agreed that the prototype system was easy to use, with the two remaining participants leaning towards neutral or agreeing.

Seven individual components of the HSF system were also evaluated in terms of usefulness (Figure 2.14). These included the hovering feature (providing details on demand for the data points or trip blocks), the velocity histogram, the automatic rescaling of the histogram according to filtered data, the velocity filtering, the fractal dimension filtering, the temporal filtering, and the hybrid filtering (by combining the filter types). All of these features were shown to have a high degree of usefulness.

Finally, participants were asked to choose their preferred fisheries data exploration method under a set of specific use cases: understanding the spatial distribution of fishing vessels, understanding the temporal distribution of fishing vessels, locating areas where vessels fish, understanding the velocity distribution of vessels, finding infractions, and the participants' overall preferred approach. These results shown in Figure 2.15 illustrate that the HSF system was generally preferred for most tasks. Overall five out of nine participants would rather use HSF than the VUE system, three participants preferred the reverse, and one participant provided a mixed response. However, the presence of some preference for their current system can be accounted for by the fact that participants are more used to it, and that VUE is more full-featured than our prototype system. For instance, the ability to export data, cross-reference with other databases, and generate reports, are all tools available with VUE which are valuable to analysts but were not included as part of the prototype implementation of HSF due to data privacy concerns.

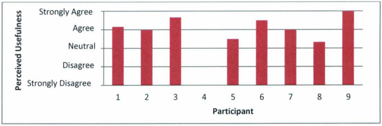


Figure 2.12. Mean Perceived Usefulness (PU) responses by participant.

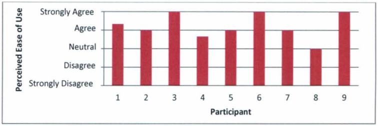


Figure 2.13. Mean Perceived Ease of Use (PEU) responses by participant.

2.7.4. Participants comments

A number of useful comments were made during the course of the field trials. Most often, comparisons were made with the VUE system. Participants mentioned that they would like to see the HSF approach integrated into their current system, as it would save them a lot of time and effort. They also seemed rather enthusiastic at being able to 'look at their data from a different angle'.

Participants also cited the lack of features in the prototype system as a detriment. These features, such as displaying a vessel's licenses, length, and capacity, are available in DFO's VUE system, but not in the prototype system. While these are obviously valid concerns when using the software for real-world data analysis, from a research perspective we decided to not re-implement any existing features unless directly required by the approach we set out to evaluate. Moreover, the data sets provided by DFO for this project were purposefully limited, due to privacy issues, making most of these features impossible to implement.

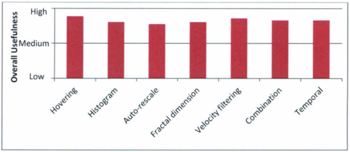


Figure 2.14. Mean Overall Usefulness value of specific features.

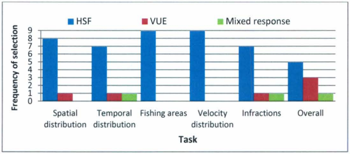


Figure 2.15. Preferred approach for specific tasks.

Participants also mentioned the idea that creating a set of known 'good' signatures for patterns and having them available to all users of a system would save a significant amount of time and effort for beginning users. This would also remove some of the burden of understanding how the moving window fractal dimension filter functions, since complex signatures could be pre-determined by more experienced users.

Finally, velocity filtering was often mentioned as being very useful and intuitive. A number of participants noted that this type of filtering would save them a lot of time if used daily. This is consistent with the responses regarding specific features (seen in Figure 2.14), which lists it as the highest rated filtering method for the prototype.

2.8. Conclusion

Our goal in this research was to develop and study an approach to dealing with the complexities of analyzing movement data. Although using visualization to analyze such data is not new, existing approaches that focus almost exclusively on visually representing the data are not effective when the data include a large number of objects that need to be tracked, cover a large spatial extent, or span a large temporal range. Rather than directly depicting the data, our approach has been to develop tools that allow analysts to extract patterns from the data. Patterns extracted are subsets of the overall data set that match specific characteristics of interest of movement data. Although there exist many automatic pattern extraction methods, we focused on the development of interactive pattern extraction methods since they support a more flexible approach to exploring the data.

In particular, we presented a novel Hybrid Spatio-temporal Filtering (HSF) approach that allows the analyst to filter large spatio-temporal data sets, within an interactive geovisual analytics environment. HSF's approach of combining velocity, fractal dimension, and temporal filters allowed analysts to extract specific behavioral signatures from large volumes of data. Although velocity and temporal filtering are not uncommon, the fractal dimension filtering and its multi-scale moving window implementation is a novel approach within this context. Further, the combination of the three types of filtering provides a powerful tool for hiding the uninteresting aspects of the data, supporting a greater degree of exploration than would be possible otherwise.

We have implemented this approach as a prototype geovisual analytics system. In addition to being able to filter the data using velocity, fractal dimension, and time, the system also supports the ability to save such filters as pattern signatures. Multiple such signatures can be displayed simultaneously, using color to visually identify the different patterns within the data and hiding all other data that do not match these signatures. The end result is a powerful tool for exploring and analyzing large (spatially, temporally, and object-wise) and sparse movement data sets.

We studied the prototype HSF system within a field trial setting in the domain of fisheries enforcement, using a real fishing vessel movement dataset, consisting of over one million records, and containing over one thousand vessels. The results of these field trials, involving nine fisheries enforcement officers working for DFO, showed the ability of this approach to find anomalous and interesting patterns. While HSF would need some additional features to replace the normal data analysis tool used by experts in an operational setting, there was strong support for the usefulness of the specific data filtering and analysis features implemented in the prototype.

The velocity filter and the building up of complex signatures based on velocity and fractal dimension were mentioned by participants as having the potential to make data analysis task much more efficient. The idea of being able to look at specific patterns or behaviors in isolation, providing insight regarding the spatial or temporal distributions, as well as being able to look at multiple patterns together, was appealing to participants. In addition, the usefulness and ease-of-use of this approach were shown to be high, as measured through a TAM questionnaire, with six out of nine participants agreeing or strongly agreeing that the prototype is useful, and seven out of nine participants agreeing or strongly agreeing that the prototype is easy to use. Since the participants in the field trials were experts in analyzing this type of data, and had access to existing software tools to support their activities, the feedback received from them can be considered as being extremely reliable. The positive outcomes from the field trials provide strong evidence regarding the value of providing analysts with interactive tools that support the filtering and highlighting of data based on speed, path complexity, and time.

The approach of combining velocity, fractal dimension, and temporal filtering in a complex geovisual analytics system has been shown to be beneficial for visualizing large movement data sets. Particularly, this approach is robust against gaps in the data, due to its use of a modified fractal dimension estimation method that uses a moving window. This allows the approach to perform well even with data that have a low temporal resolution, such as the hourly data used in the case study, and also be tailored to specific types of movement patterns, at multiple different temporal scales. This is particularly important in situations where high temporal resolution data is expensive or even impossible to acquire.

This work adds to the growing body of literature showing the potential utility of geovisual analytics approaches for exploring complex spatio-temporal datasets. The HSF has the potential to save significant amounts of time and allow a more thorough analysis of these types of data. In the context of fisheries enforcement, this approach can help in finding, and ultimately preventing, illegal fishing, as well as the management of the complex marine ecosystem. However, this same approach could be applied to virtually any movement data, such as pedestrian, cars, or animal movements, enabling a wide range of patterns and behaviors to be investigated.

Further testing is required to determine how HSF compares to other approaches currently in use in other jurisdictions and other movement data analysis contexts. In addition, it may be beneficial to evaluate the performance of this approach with data that have a higher temporal resolution, such as 15 minute VMS data or even three minute Automated Identification System (AIS) data. More work is also required to evaluate whether incorporating a vessel's reported heading information, and cubic Hermite spline interpolation, into the fractal dimension calculation would make any significant difference in the quality of the estimation.

2.9. Acknowledgements

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for the funding of this project through the Strategic Projects Grant STPGP 365189-08, and the Canada Foundation for Innovation (CFI) and Memorial University of Newfoundland for providing the laboratory infrastructure. We also would like to thank Fisheries and Oceans Canada, and especially Jerry Black and Trevor Fradsham and their groups, for their collaboration and having provided access to the data, as well as the fisheries enforcement officers who participated in this study.

2.10. Funding

This work was funded by a Strategic Projects Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) held by the second and third authors (grant number: STPGP 365189-08), and an NSERC Discovery Grant held by the third author.

2.11. References

Andersson, M., Gudmundsson, J., Laube, P., & Wolle, T. (2008). Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, *12*(4), 497-528. doi:10.1007/s10707-007-0037-9

Andrienko, G., Andrienko, N., & Wrobel, S. (2007a). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46. doi:10.1145/1345448.1345455

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I., & Wachowicz, M. (2008a). Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization*, *7*(3-4), 173-180. doi:10.1057/ivs.2008.23

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S., & Wrobel, S. (2011). From movement tracks through events to places : Extracting and characterizing significant places from mobility data. *Proceedings of the 2011 IEEE Symposium on Visual Analytics Science and Technology* (pp. 159-168). Providence, USA: IEEE.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D. A., Kraak, M.-J., MacEachren, Alan M., & Wrobel, S. (2007b). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839-857. doi:10.1080/13658810701349011

Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008b). Basic concepts of movement data. In F. Giannotti & D. Pedreschi (Eds.), *Mobility Data Mining and Privacy* (pp. 15-38). Springer. doi:10.1007/978-3-540-75177-9

Bartumeus, F., & Catalan, J. (2009). Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, *42*(43), 1-12. doi:10.1088/1751-8113/42/43/434002

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., & Gerlotto, F. (2007). Scale-invariant movements of fishermen: the same foraging strategy as natural predators. *Ecological Applications*, *17*(2), 331-337.

Bomberger, N., Rhodes, B., Seibert, M., & Waxman, A. (2006). Associative learning of vessel motion patterns for maritime situation awareness. *Proceedings of the 9th International Conference on Information Fusion* (pp. 1-8). Florence, Italy: IEEE. doi:10.1109/ICIF.2006.301661

Chang, S.-K., Liu, K.-Y., & Song, Y.-H. (2010). Distant water fisheries development and vessel monitoring system implementation in Taiwan—History and driving forces. *Marine Policy*, *34*(3), 541-548. Elsevier. doi:10.1016/j.marpol.2009.11.001

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129-136.

Choi, K.-H., Kim, J.-S., Kim, Y. S., Yoo, M.-A., & Chon, T.-S. (2006). Pattern detection of movement behaviors in genotype variation of *Drosophila melanogaster* by using self-organizing map. *Ecological Informatics*, *1*(3), 219-228. doi:10.1016/j.ecoinf.2005.12.002

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems*, *13*(3), 319-340.

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., & Die, D. (2005). Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(3), 611-622. doi:10.1139/f04-219

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, *7*(3-4), 240-252. doi:10.1057/palgrave.ivs.9500182

Dykes, J. A., & Mountain, D. A. (2003). Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, *43*(4), 581-603. doi:10.1016/S0167-9473(02)00294-3

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, *63*(7), 1057-1066. doi:10.1007/s00265-009-0739-0

Focardi, S., Marcellini, P., & Montanaro, P. (1996). Do ungulates exhibit a food density threshold ? A field study of optimal foraging and movement patterns. *The Journal of Animal Ecology*, *65*(5), 606-620. doi:10.2307/5740

Franke, A., Caelli, T., & Hudson, R. J. (2004). Analysis of movements and behavior of caribou (*Rangifer tarandus*) using hidden Markov models. *Ecological Modelling*, *173*(2-3), 259-270. doi:10.1016/j.ecolmodel.2003.06.004

Füchslin, R. M., Shen, Y., & Meier, P. F. (2001). An efficient algorithm to determine fractal dimensions of point sets. *Physics Letters A*, *285*(1-2), 69-75. doi:10.1016/S0375-9601(01)00339-5

Gottfried, B. (2011). Interpreting motion events of pairs of moving objects. *GeoInformatica*, *15*(2), 247-271. doi:10.1007/s10707-009-0095-2

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395-408. doi:10.1111/j.1461-0248.2009.01293.x

Hering, E. (1964). *Outlines of a theory of the light sense* (p. 317). Cambridge, USA: Harvard University Press.

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, *24*(1), 6–21.

Höferlin, M., Höferlin, B., Weiskopf, D., & Heidemann, G. (2011). Uncertainty-aware video visual analytics of tracked moving objects. *Journal of Spatial Information Science*, *2*(2), 87-117. doi:10.5311/JOSIS.2010.2.1

Jern, M., & Franzen, J. (2007). Integrating InfoVis and GeoVis components. *Proceedings of the 11th International Conference on Information Visualization (IV '07)*, (pp. 511-520). Zürich, Switzerland: IEEE. doi:10.1109/IV.2007.70

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference* (pp. 1988–1995). Durban, South Africa: International Cartographic Association.

Kraak, M.-J., & van de Vlag, D. E. (2007). Understanding spatiotemporal patterns: visual ordering of space and time. *Cartographica*, *42*(2), 153-161. doi:10.3138/carto.42.2.153

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, *8*(1-6), 185-203. doi:10.1016/S0968-090X(00)00017-6

Laxhammar, R., Falkman, G., & Sviestins, E. (2009). Anomaly detection in sea traffic - a comparison of the Gaussian Mixture Model and the Kernel Density Estimator. *Proceedings of the 12th International Conference on Information Fusion* (pp. 756-763). Seattle, USA: IEEE.

Li, X., Han, J., & Kim, S. (2006). Motion-Alert : automatic anomaly detection in massive moving objects. *Proceedings of the 2006 IEEE Intelligence and Security Informatics Conference* (pp. 166-177). San Diego, USA: IEEE.

Lundblad, P., Jern, M., & Forsell, C. (2008). Voyage analysis applied to geovisual analytics. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 381-388). Columbus, USA: IEEE. doi:10.1109/IV.2008.32

Mandelbrot, B. B. (1967). How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science, 156*(3775), 636-638. doi:10.1126/science.156.3775.636

Mills, C. M., Townsend, S. E., Jennings, S., Eastwood, P. D., & Houghton, C. A. (2006). Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science*, *64*(2), 248-255. doi:10.1093/icesjms/fsl026

Mullowney, D. R., & Dawe, E. G. (2009). Development of performance indices for the Newfoundland and Labrador snow crab (*Chionoecetes opilio*) fishery using data from a vessel monitoring system. *Fisheries Research, 100*(3), 248-254. doi:10.1016/j.fishres.2009.08.006

Murawski, S. A., Wigley, S. E., Fogarty, M. J., Rago, P. J., & Mountain, D. G. (2005). Effort distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of Marine Science*, *62*(6), 1150-1167. doi:10.1016/j.icesjms.2005.04.005

Mårell, A., Ball, J. P., & Hofgaard, A. (2002). Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology*, *80*(5), 854-865. doi:10.1139/Z02-061

Nams, V. O. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia*, *143*(2), 179-88. doi:10.1007/s00442-004-1804-z

Pelot, R., & Wu, Y. (2007). Classification of recreational boat types based on trajectory patterns. *Pattern Recognition Letters*, *28*(15), 1987-1994. doi:10.1016/j.patrec.2007.05.014

Ren, F., & Kwan, M.-P. (2007). Geovisualization of human hybrid activity-travel patterns. *Transactions in GIS*, *11*(5), 721-744. doi:10.1111/j.1467-9671.2007.01069.x

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Information Visualization*, *7*(3-4), 225-239. doi:10.1057/palgrave.ivs.9500183

Rodighiero, D. (2010). Guidelines to visualize vessels in a geographic information system. *Proceedings of the 14th International Conference on Information Visualisation* (pp. 455-459). Salt Lake City, USA: IEEE. doi:10.1109/IV.2010.70

Sage, B. (2005). Identification of "high risk vessels" in coastal waters. *Marine Policy, 29*(4), 349-355. doi:10.1016/j.marpol.2004.05.008

Schaefer, K. M., & Fuller, D. W. (2006). Comparative performance of current-generation geolocating archival tags. *Marine Technology Society Journal, 40*(1), 15-28. doi:10.4031/002533206787353673

Sevcik, C. (1998). A procedure to estimate the fractal dimension of waveforms. *Complexity International, 5*(19), 1-20.

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools. *Proceedings of the 2006 AVI workshop on Beyond Time and Errors Novel Evaluation Methods for Information Visualization* (pp. 1-7). Venice, Italy: ACM Press. doi:10.1145/1168149.1168158

Theiler, J. (1990). Estimating fractal dimension. *Journal of the Optical Society of America A, 7*(6), 1055-1073. doi:10.1364/JOSAA.7.001055

Thomas, J., & Cook, K. (2005). *Illuminating the path: research and development agenda for visual analytics* (p. 194). Los Alamitos, USA: IEEE Computer Society.

Tremblay, Y., Shaffer, S. A., Fowler, S. L., Kuhn, C. E., McDonald, B. I., Weise, M. J., Bost, C.-A., et al. (2006). Interpolation of animal tracking data in a fluid environment. *The Journal of Experimental Biology, 209*(1), 128-40. doi:10.1242/jeb.01970

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed., p. 200). Cheshire, CT, USA: Graphics Press.

Underwood, A. J., & Chapman, M. G. (1985). Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology, 91*(1-2), 17–43. Elsevier. doi:10.1016/0022-0981(85)90218-7

Wang Baldonado, M. Q., Woodruff, A., & Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '00* (pp. 110-119). New York, USA: ACM. doi:10.1145/345513.345271

Ware, C. (2004). *Information visualization: perception for design* (2nd ed., p. 486). San Francisco, USA: Morgan Kaufmann.

Willems, N., van de Wetering, H., & van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics Forum, 28*(3), 959-966. doi:10.1111/j.1467-8659.2009.01440.x

With, K. A. (1994). Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology*, *9*(1), 25-36. doi:10.1007/BF00135076

Zhao, J., Forer, P., & Harvey, A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, *7*(3-4), 198-209. doi:10.1057/palgrave.ivs.9500184

Chapter 3 Comparing interactive and automated methods for analyzing large vessel movement data set

Abstract

The ever-increasing traffic on the world's oceans led the adoption of new technologies for improving maritime safety, helping manage activities taking place in the maritime environment, and ensuring compliance with national and international regulations. This had led to a commensurate increase in the volume of movement data collected to support ocean management activities. As a result, new methods are required to help extract useful knowledge from these large data sets. Various approaches have been proposed for visualizing and exploring movement data and detecting patterns within these data, but those approaches have generally not been tested in a real-world context or compared together, making their actual usability and utility unclear. This paper describes, compares, and assesses three such approaches in the context of fisheries enforcement: an existing system used for fisheries enforcement operations in Canada (VUE), a novel Hybrid Spatio-temporal Filtering (HSF) system developed by the authors, and an automated Behavioural Change Point Analysis (BCPA) system. A field trial was conducted with experienced fisheries enforcement officers to compare and contrast the benefits and drawbacks of the three approaches. While all three presented advantages and disadvantages, the interactivity of VUE and HSF were identified as desirable features, as they provide analysts with more control over the data, while allowing flexible data exploration. BCPA, while providing an automated approach to the data analysis, was

pointed out as being too much of a "black box", causing unease among the experts who require a level of transparency similar to that of legally admissible evidence. In the end, the experts suggested that the best approach would be to merge the analytical power of their existing VUE system with the exploratory power of the HSF system. This study provides some insight into the value of using interactive mapping and filtering approaches in support of data analysis in the context of fisheries enforcement.

3.1. Introduction

Monitoring vessels movement and enforcing commercial fishing regulations in marine environments increasingly relies on the use of positioning systems aboard vessels, such as Vessel Monitoring Systems (VMS) or Automatic Identification Systems (AIS). Such systems transmit the position of vessels at regular time intervals and typically generate very large movement databases. Those databases can be analyzed by organizations in charge of ensuring maritime safety or enforcing regulations. In many cases, a manual analysis of those data sets is challenging, due to the large volume of data. In such a data-rich environment, data exploration can become a crucial step in extracting knowledge (Jern et al., 2008). Geovisual analytics is an approach that aims at exploring large data sets through the use of interactive visualization of spatial or spatio-temporal data. It can be defined as the sub-area of visual analytics that focuses on space and time, their related research problems, and using these approaches to solving more general problems (Andrienko et al., 2007b). Many geovisual analytics systems and approaches currently exist (Andrienko & Andrienko, 2007; Tomaszewski et al., 2007), and focus on a variety of different domains. However, they all share the common goal of providing an interactive environment in which one can explore and interrogate spatial data sets for the purposes of data analysis and decision making.

Movement data, often recorded as spatial coordinates with a time stamp, can be a powerful and complex data source for understanding behaviours. They can help with the analysis of the movements of individuals or groups, finding movement patterns, or identifying outliers. They also allow for the recording and replay of the movements of a target, such as a fishing vessel or an animal. This, in turn, allows researchers to better understand patterns of movement and their associated behaviours (Dodge et al., 2008). Such data can also be used to ascertain group dynamics, such as identifying the leader of a group or whether certain groups attract or repel each other (Gottfried, 2011).

Due to the high complexity and information density of movement data, it often is difficult to formulate good questions or hypotheses about phenomena of interest without first exploring them. This preliminary exploration often helps determine the proper selection of tools and methods that will support further analysis (Andrienko et al., 2006). In addition, such exploratory tasks may lead to the analysis of the data from different perspectives than what might have been done with only *a priori* knowledge.

One common method used during the exploration of movement data is pattern extraction. Patterns, in this sense, can be thought of as "any recognizable spatial and temporal regularity or any interesting relationship in a set of movement data" (Dodge et al., 2008). Often, patterns within a data set can be automatically extracted through mathematical or statistical means. Doing so essentially filters out the uninteresting aspects of the data, providing a partial solution to the problem of information overload in the analysis of large data sets. Such a filtering activity allows analysts to focus only on specific patterns in the data, helping them to gain a better understanding of the spatial distribution, composition, and characteristics of these patterns. In the context of movement data, patterns can also indicate some types of behaviour of a particular target, such as a movement path becoming very complex when a vessel is fishing with a long-line, or dredging for scallops.

In addition to the automated pattern extraction methods, it is also possible for analysts to detect and extract patterns visually and interactively. Interactive pattern extraction gives more control to the analyst as to which patterns are being extracted, and allows analysts to explore data sets more easily. A number of methods and systems have been proposed to help visually and interactively identify patterns within large movement data sets (Andrienko et al., 2007a; Andrienko & Andrienko, 2008; Enguehard et al., submitted; Zhao et al., 2008). However, little research has been devoted to evaluating the usability or utility of these methods within real-world settings, nor to identifying and studying synergistic approaches.

The goal of this study is to assess the effectiveness of different visualization and exploration approaches for spatio-temporal movement data. Three systems, implementing three different approaches, were chosen: (1) the VUE system currently used by Fisheries and Oceans Canada (formerly Department of Fisheries and Oceans, or DFO), which provides limited support for interactive analysis, but includes access to a rich set of analysis tools and metadata regarding the fishing vessels, their fishing licenses, etc.; (2) a Hybrid Spatio-temporal Filtering (HSF) system developed by the authors (Enguehard et al., submitted), which provides a high degree of support for interactive filtering and exploration; and (3) a Behavioural Change Point Analysis (BCPA) system that provides an automated method for clustering movement patterns (Gurarie et al., 2009). Each of these three approaches includes a geovisualization interface with a level of interactivity appropriate to their degree of support for interactive data analysis (i.e. high interactivity in HSF, moderate interactivity in VUE, and low interactivity in BCPA). The assessment of the relative strengths and weaknesses of each system was conducted through field trials, wherein fisheries enforcement officers conducted self-chosen data analysis tasks that are common in their normal work activities and that take advantage of their expert knowledge of fisheries movement data and fisheries enforcement concerns.

While the approaches explored in this study could be used to analyze many different types of movement data, experimentation and testing were done using fishing vessel movement data. The focus of the data analysis tasks was on identifying movement patterns that are indicative of illegal fishing behaviour. DFO is the Canadian governmental agency in charge of monitoring and implementing policies within Canadian waters. Like many countries, they maintain a program of Vessel Monitoring Systems (VMS), where certain fishing vessels must be equipped with a system that records and transmits the vessel's location at pre-determined time intervals. These VMS positions feed in near real-time to a centralized database, which fisheries enforcement officers can use to detect abnormal or potentially illegal activity. Such activities can range from fishing in an area closed to specific fishing activities, fishing after the closure of a given fishery, fishing for a species other than the one authorized by the license, or illicitly transferring fish from one vessel to another. Enforcement officers can then collect evidence to determine whether it is warranted to investigate further the individuals connected to these activities once they return to port. The timeliness of such investigations can have a significant impact on their success, as well as on the ability to deter others from engaging in the same types of activities.

A challenging aspect of fisheries enforcement is the fact that while the volume of VMS data to be analyzed keeps increasing, the number of enforcement officers has remained relatively constant. As a result, the enforcement officers often focus their attention on "problem" vessels that are suspected of prior illegal activities. It is expected that through the use of focused data analysis and visualization tools, more powerful filtering, and/or automated approaches, previously unidentified movement patterns within the data may be brought to light. This could significantly increase the efficiency and effectiveness of enforcement officers, leading to better management of the ocean environment and preservation of fisheries resources.

The remainder of this paper will be structured as follows. Section 2 will relate this work to previous research. Section 3 will present the three data analysis approaches studied in this paper. Section 4 will present the details of the field trial methods used to compare and assess the approaches. Section 5 will present the results of the field trials. Section 6 will discuss these results and offer a comparative analysis. Finally, Section 7 will summarize the key contributions and outline future work.

3.2. Related Works

Representing and analyzing the movement of people, animals, or objects has been the focus of studies in various disciplines for centuries. Early flow maps from the 19th century cartographer Minard are examples of how cartographic visualization can be a powerful way to represent movements, but also to derive knowledge about the underlying behaviours resulting in those movements (Tufte, 2001). With the advent of GPS technology, and the increasing pervasiveness of GPS-enabled mobile devices, movement data are becoming a prevalent source of information. It is now common to track animals to study their behaviour, to keep track of mobile phone movements for location-based services, or to track fleets of buses, taxis, or vessels to improve their management. These data sets are often very large and complex, and as a result extracting useful knowledge from them has become a scientific challenge.

As a consequence, there is an increasing need for the design, development, and study of novel methods that allow analysts to discover new knowledge amid the complexity within the data. One such approach that is gaining popularity is geovisual analytics (Andrienko et al., 2008). Visual analytics focuses on the acquisition of knowledge through visual analysis of data, to support data exploration and decision-making processes (Thomas & Cook, 2005). Geovisualization distinguishes itself from visual analytics by having to take into account the complexities involved with space (and often time), multiple levels of decision-making actors, and the need to allow for human intervention into the decision making process using implicit, rather than explicit, knowledge (Andrienko et al., 2007b).

The complexities involved in spatio-temporal movement data can quickly grow due to the various levels of data density and additional information with which they can be associated. For instance, AIS, which is commonly used to prevent collisions between large ships in fog or bad weather, transmits GPS coordinates and a number of other parameters, such as vessel name, heading, and velocity at least every three minutes, and usually much more often (every few seconds) (Schwehr & McGillivary, 2007). In contrast, VMS, which is used by DFO and other organizations around the world to monitor the locations of fishing vessels, transmits GPS coordinates, a timestamp, and a vessel identifier at a much lower rate (e.g. once an hour in the Newfoundland maritime region).

The wide range of differences in movement data has led to a wide range of geovisual analytic approaches to deal with them. Many of these approaches (Eagle & Pentland, 2009; Kwan, 2000; Lundblad et al., 2008; Zhao et al., 2008) are specific to certain types of spatio-temporal data (commuting data, vessel movement data, daily activity data, and mobile phone data, respectively), while others (Andrienko et al., 2007a; Johansson & Jern, 2007) are more generic. Some are based on data clustering (Andrienko et al., 2011b; Rocha et al., 2010), others use data filtering (Andrienko et al., 2011a; Enguehard et al., submitted; Lundblad et al., 2008), and still others use path-based aggregation (Demšar & Virrantaus, 2010; Laxhammar et al., 2009; Willems et al., 2009). While this study focuses on vessel movement data, approaches have been devised for movement data from a number of different perspectives, such as shipping (Laxhammar et al., 2009; Lundblad et al., 2008; Willems et al., 2009), or ecology (Bertrand et al., 2007; Gurarie et al., 2009; Raymond & Hosie, 2009).

Multiple aspects of VMS have also been studied extensively, such as system implementation issues (Chang et al., 2010; Molenaar & Tsamenyi, 2000), its use for the estimation of performance indices (Deng et al., 2005; Mullowney & Dawe, 2009), its use to estimate fishing activity (Mills et al., 2006; Murawski et al., 2005; Witt & Godley, 2007), or how VMS locations can be interpolated (Hintzen et al., 2010).While fisheries movement data have been the focus of a number of studies, little research has been devoted to geovisual analytics approaches that could support fisheries enforcement work. For instance, (Rodighiero, 2010) presents a generalized guideline to be used when visualizing VMS data, but with little emphasis on the analytical aspects or how to use these guidelines to support enforcement operations. The behavioural aspects of VMS data have also been studied, such as how the velocity or geometric patterns of movement can be used to ascertain certain activities. These approaches can be applied to detect areas where vessels are fishing in order to help manage fish stocks (Bertrand et al., 2007; Saitoh et al., 2011). While some studies used geovisual analytics approaches for movement data in the maritime domain (Lundblad et al., 2008; Willems et al., 2009), none addressed questions of fisheries enforcement. Studies exploring methods for fisheries enforcement either study the analytical or the geovisual aspects of fisheries movement data, but not both of these aspects together. This research attempts to address this issue through field trials of geovisual analytics systems that focus on real-world data analysis tasks in the fisheries enforcement domain.

3.3. Geovisual Analytics Systems

The three geovisual analytics systems compared in this study were designed and developed in different contexts for exploring movement data, including two that were developed or adapted by the authors for this study. The VUE system is currently used in day-to-day enforcement operations by DFO. It combines a web-based GIS package and a large relational database. It was designed to be highly flexible, allowing for many different kinds of analyses using diverse data sets, but due to the constraints of the web-based framework, provides limited real-time interaction.

The HSF prototype system was designed from a research perspective, with a goal of studying the value of different visual and interactive approaches to filtering movement data (Enguehard et al., submitted). Although an emphasis was placed on exploring large data sets, rather than answering specific questions, fisheries enforcement informed some of the key design decisions of the prototype. The system provides powerful interactive filtering tools based on the spatial and temporal aspects of the data, and effective visual representations of the movement data.

The core approach used in the BCPA system was developed in behavioural ecology (Gurarie et al., 2009), and places a heavy emphasis on automation. The system attempts to identify statistically homogenous behavioural patterns, and uses this information to show where and when movement patterns change. These change points are used to segment the movement data, which are then represented in a visual manner. As an automated approach, the support for interactive exploration is limited to standard pan and zoom operations within the geovisual representation.

Each of these systems (VUE, HSF, and BCPA) has advantages and disadvantages, as well as varying degrees of interactive support for data analysis. The relative differences between these systems in terms of automation and interactivity are summarized in Figure 3.1.

3.3.1. VUE System

VUE is the visualization system currently used by DFO to access and visualize VMS-based fishing vessel movement data. It is a web-based application that allows for the interactive loading and querying of large data sets, through the selection of temporal ranges, spatial extents, and vessels of interest. The data are then downloaded from a central server and displayed on a static map. Data points are displayed as coloured dots (Figure 3.2), with the colour indicating the identity of each vessel. These dots are then connected with straight lines, with an arrowhead indicating direction of travel. Other visual elements include coastlines, bathymetry, international borders, and fishing closures.

The VUE system is used by enforcement officers and fisheries analysts within DFO to support a number of tasks. It allows for the assessment of where fishing vessels go and when they move. This can be used to calculate Catch Per Unit Effort (CPUE), or as a proxy for species distributions (Mills et al., 2006; Mullowney & Dawe, 2009). Another important use of the VUE system is to help document evidence (VMS logs) for court cases involving illegal fishing activities, such as fishing in closed regions or after the season has ended. Some of these uses require a quick analysis of near real-time data, whereas others require the analysis of very large historical data sets. As a result, the system is designed to be versatile with respect to what it allows analysts to accomplish.

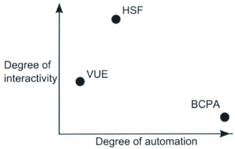


Figure 3.1. VUE, HSF, and BCPA approaches with respect to automation and interactivity.

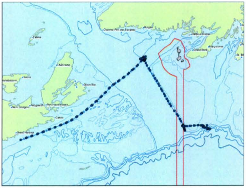


Figure 3.2. VUE's map display showing a fishing vessel's movement path in Eastern Canada.

The system itself integrates two separate components: a custom web-mapping application, and a large database system which manages all the VMS data and associated data sets, such as license information, quotas, and vessel characteristics. The mapping system incorporates static pan and zoom operations, which operate in a non-real-time fashion, as well as complex data management using layers. Since the mapping system does not allow for real-time interactive use, panning is a click-and-wait operation, while zooming occurs by selecting an area of interest and reloading the map.

The process of pattern extraction or anomaly detection using the VUE system is most efficient when analysts know what they are specifically looking for, such as a type of vessel with a specific species license in a specific area of interest. When these types of information are not known, analysts need to proceed in a more exploratory manner, investigating large amounts of data visually. While the VUE system does support this, its efficiency is severely limited, both in terms of visualization effectiveness and functionality. Typically, the VUE system only loads vessel data for vessels within a specified area of interest. To load other data, the VUE system takes a significant amount of time, which varies and is based on the amount of data requested. This wait time can lead to analysts losing interest in their specific task, or may break their train of thought. Moreover, there is a maximum number of vessels that may be displayed on-screen at any given time due to the way vessels are visually represented with distinct colours.

The database system used with VUE manages a very large amount of data. Vessels in Eastern Canada typically communicate their positions on an hourly basis. As a consequence, VMS often generates more than 50,000 data points every day. These data contain information about where the vessel is located, when it was there, and the vessel's identification. The vessel identification can then be used to find more information about it, such as crew information, licensing restrictions, and many more attributes that provide a complete historical overview of each vessel. This enables enforcement officers to not only analyze vessel movement patterns, but also inspect a vessel's history, such as past infractions, incursion into fishing zones for which the vessel does not have licenses, and the off-loading of fish species that are not consistent with the areas where they reportedly fished.

These functions, as well as many more, have been integrated as selectable tools surrounding the mapping system in VUE (Figure 3.3). The system itself operates over the Internet, but is only accessible through the DFO Virtual Private Network (VPN), ensuring the security and confidentiality of all the data. None of the data are stored locally, which means that there is currently no possibility of using the VUE system in an offline fashion.

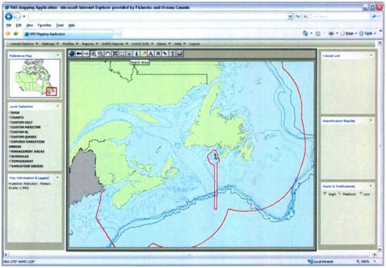


Figure 3.3. The data analysis interface surrounding VUE's mapping system.

3.3.2. Hybrid Spatio-temporal Filtering

An interactive hybrid spatio-temporal filtering (HSF) approach for the exploration of fishing vessels movement was proposed by (Enguehard et al., 2011) and presented in detail and evaluated in (Enguehard et al., submitted). It combines filtering on the path complexity of the vessel tracks, the vessel's velocity, and temporal constraints. This combination allows analysts to build specific filtering signatures to identify behaviours of interest within large data sets. The HSF system provides a multiple-coordinated-views geovisualization interface, allowing the analyst to instantly see the outcomes of these filtering activities. This allows analysts to look at their data in an exploratory fashion, but also to define re-usable signatures with specific settings, to match and extract specific patterns or behaviours. The geovisualization component of the prototype tool was developed on top of NASA World Wind (Kim & Hogan, 2011), an open-source virtual globe similar to Google Earth. It allows the display of geospatial data in a virtual three-dimensional environment, instead of a standard top-down cartographic representation. The HSF system also automates a number of processes, such as data interpolation (through the use of a cubic Hermite spline), heading and velocity estimations, and potential range of travel between data points.

Taking advantage of the features of human colour perception (Ware, 2004), the data points are displayed using yellow arrowheads (Figure 3.4), to contrast strongly with the blue ocean. The heading of each arrowhead is determined by the forward great-circle azimuth to the next data point in each trip, indicating the direction of the travel. A trip, in the case of this fisheries data, was defined as a set of data points for one individual vessel with no gaps in time larger than six hours. Six hours was selected since vessels usually spend at least six hours at port when finishing a trip. Moreover, experts indicated that vessels undertaking illegal fishing activities may turn off their VMS system for one or two hours but not for six hours, as that would make the gap in the VMS data too obvious.

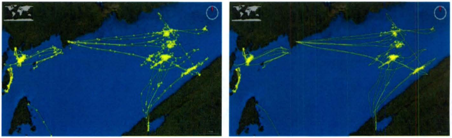


Figure 3.4. Movement data representation implemented in the HSF system. Left: visualization of all the data. Right: same data filtered to only show arrowheads where fishing activities take place.

The filtering features use the fractal dimension of movement paths, which is a measure of path complexity, along with the velocity of the target (see Enguehard et al. (submitted) for details on the method). In ecology, fractal dimension has been shown to be a reasonably good indicator of a predator's behaviour, such as foraging or moving between food patches (Mårell et al., 2002; Nams, 2005; With, 1994). Similarly, the complexity of fishing vessel movements is believed to be indicative of specific behaviours (Bertrand et al., 2007). For instance, a fishing vessel using a given gear type may display a much larger path complexity than other vessels using another gear type.

Velocity, however, is a physical characteristic of movement, directly related to the vessel type, the environmental conditions, and the decisions made by the individual being tracked. Velocity is also a good indicator of fishing vessel behaviours, as fishing activities typically occur within certain speed ranges (Saitoh et al., 2011). This velocity will vary depending on the type of fishing gear used as it is related to various physical limitations such as the tensile strength of nets or lines, or safety considerations when throwing or pulling crab pots.

The hybrid filter within this approach allows analysts to combine specific ranges of velocities and fractal dimensions to extract particular patterns indicative of behaviours of interest. The patterns extracted by these types of filters can be designed to be much more specific than those extracted by either velocity or complexity filters. A simple velocity filter, for instance, could identify whether a vessel is fishing or steaming. Likewise, a complexity filter could identify when a vessel is undertaking complex behaviours, such as maneuvering close to shore, or trawling. However, combining these filters could provide information about the types of behavioural patterns exhibited during specific types of fishing, such as long-line fishing (low velocity, high fractal dimension), fishing along a depth contour (low velocity, low fractal dimension), or steaming (high velocity, low fractal dimension).

Added to the hybrid behavioural filtering, explicit filtering modes using start and end dates, and vessel selection are available. When data points are removed using filters, their arrowheads disappear from the map (Figure 3.4), but the underlying path remains, to help the analyst maintain awareness of the movements of each vessel. The resulting map has a significantly lower level of complexity and allows the analysts to focus on patterns of interest. The hybrid filter can be saved as a signature and reused at another time, with the possibility to use multiple concurrent signatures simultaneously. In this case, each signature is assigned a unique colour, which is used to represent both the arrowheads matching that signature and the vessel paths.

3.3.3. Behavioural Change Point Analysis

Behavioural Change Point Analysis (BCPA) is a recent approach for automatically analyzing movement data (Gurarie et al., 2009). It was tested and shown to be effective using northern fur seal (*Callorhinus* ursinus) movement data. Moreover, according to a study by Bertrand et al. (2007), the behaviour of fishing vessels can be accurately likened to that of other natural predators. As a result, the BCPA method is applicable to fishing vessel movement data and can be considered a meaningful comparison point.

It was implemented in this study such that it can analyze fishing vessels movements. BCPA automatically identifies the point, or points, at which the statistical distribution of a set of time-series data significantly changes, according to a pre-defined significance level. It uses a moving window to sweep over all the data, segmenting the window into two parts, with the segmentation point being the candidate change point. The strength of the statistical difference between the two parts is measured, and all the candidate break points with a difference greater than the threshold significance level are determined to be significant change points. The drawback of the approach is that it is computationally intensive, particularly over large data sets. As a result, while BCPA is an automated approach, it cannot detect the patterns in real-time.

BCPA measures the change in movement patterns over time, and attempts to link large changes in movement patterns with a change in behaviour. However, it does not provide a justification as to what the behaviours are or if there is any meaningful difference between the behaviours beyond a statistical difference in the velocity vectors. In that respect, BCPA operates like most unsupervised classification techniques, grouping similar data but not providing an explanation for the similarity.

A visual interface for the BCPA technique was developed using NASA World Wind, with the geovisual representation following a similar approach to what was used for the HSF system. Since the BCPA technique is automated, the entire data set was processed once and the behavioural change points were simply loaded into the system. As such, the interactive features were limited to the selection or filtering of vessels, and the standard pan and zoom operations within the map-based representation.

A set of visually distinct colours was used to represent segments of different behaviours (Figure 3.5). As a result, the change points determined by BCPA are represented as the transition points between two colours, making them easy to identify.

As such, an analyst interested in one particular type of behaviour may locate a section where it seems that a related behavioural pattern is under way. The analyst could then visually determine where and when that behaviour started and stopped, based on colour changes. For example, this may help in situations where it is unclear whether some fishing started or stopped within or near an area closed to fishing.

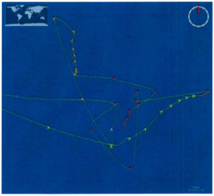


Figure 3.5. Segmentation of movement path done by BCPA with the different patterns identified displayed using different colours.

3.4. Comparison of approaches

A field trial using expert participants was conducted to assess the relative strengths and weaknesses of each of the three approaches. The advantage of field trials over other types of formal studies (e.g., laboratory studies), is that they provides a realistic test of the approaches by allowing expert participants to use real data to perform their usual tasks (Shneiderman & Plaisant, 2006). This allows the participants to provide informed opinions as to the value of each approach within the context of their real-world work activities and environments, leading to a more reliable assessment of the systems under investigation.

According to Plaisant (2004): "Usability of information visualization tools can be measured in a laboratory, however to be convincing, utility needs to be demonstrated in a real setting". Field trials are not intended to provide reliable or comparable quantitative data, but rather in-depth qualitative information as to the real-world utility of a particular approach or tool. These types of studies generally have small sample sizes due to the need for real users and the high degree of involvement required by the participants. The evaluation methodology is intentionally unstructured, and generally left up to the participant. As a result, forcing a quantitative analysis method onto data collected in an inherently qualitative manner could have limited value. Moreover, the insightful qualitative feedback from a small number of real-world users has the potential to be much more valuable than the statistical analysis of quantitative data measured for contrived tasks (Shneiderman & Plaisant, 2006).

The main objective of our field trial was to gain a better understanding of the benefits and drawbacks of the three systems presented, in order to better understand which approaches work well for various situations . Specifically, the tasks of analyzing and visualizing large movement data sets in the context of fisheries enforcement were investigated within this study. Of particular interest are the questions of how to deal with coarse or irregular sampling intervals, interactivity vs. automation, filtering and analytical power, usability, efficiency, and effectiveness. This information can then be used to help devise more refined or useful systems, not just limited to a fisheries context but to any movement data, as well as potentially highlighting opportunities for integration of different approaches.

3.4.1. Participants

Four fisheries enforcement officers from DFO who analyze VMS data on a regular basis were selected by their respective managers to perform an evaluation of each of the three systems. The main duty of these enforcement officers is to ensure that the Canadian regulations related to fisheries are not circumvented. Although they have a number of tools at their disposal, one of the primary data analysis tasks is to identify anomalous vessel movement patterns within the VMS data.

All the experts selected for the field trial had over ten years of experience in enforcement operations, and had frequently used the VUE system to perform their regular duties. As a result, they were familiar with the patterns and movements of fishing vessels, as well as the functionality, benefits, and drawbacks of the VUE system's approach to data analysis.

3.4.2. Methodology

The field trial was conducted in the DFO's offices in Dartmouth, NS, with each participant using the various software systems individually. All participants in the study had extensive prior experience in using the VUE system. For the two other systems (HSF and BCPA), participants were provided with detailed training sessions guided by the researcher conducting the field trials. Furthermore, the researcher remained available to assist the participants with their data analysis tasks, allowing them to operate at an expert level even though two of the approaches were new to them. Their interactions with the systems were recorded, as well as any comments they made, through the use of video and audio recording.

Participants were asked to analyze data from the 2009 VMS database using each of the three approaches in turn. These data had been previously sanitized, and all personally identifying information were removed for use with the HSF and BCPA systems, due to privacy concerns. As such, fields like vessel names, licenses, and gear were unavailable. Participants were given full control of the systems, and were allowed to work with them for up to two-hours, with the exact amount of time being determined by them. The two hour time limit was enacted due to the limited availability of participants; however, the length provided the officers with ample time to assess the systems.

Finally, all participants were interviewed to examine their opinions and experiences with using each of the systems, and whether they expected any potential benefit to combining some of the features from the various approaches. The interviews followed a semi-structured format, ensuring that certain key questions were answered, but allowed full liberty in terms of follow-up questions and specific wording. In all, seven key questions were asked:

- What is your opinion of the interactive Hybrid Spatio-temporal Filtering (HSF) technique?
- What is your opinion of the automated Behaviour Change Point Analysis (BCPA) technique?
- Were the segments identified by the BCPA technique meaningful?
- In terms of ease of use, would you rather use one of the three techniques (VUE, HSF, BCPA) over the others?

- Would either of the new techniques (HSF, BCPA) help you during your day-to-day activities and tasks?
- Would combining any of the techniques be beneficial, and if so, which combinations?
- Overall, do you prefer one technique over the others, and if so, why?

3.5. Results

Seven key issues were commented upon by each participant after having used each of the three systems to support their self-selected enforcement tasks. This section summarizes the results related to these seven topics, as well as any pertinent comments made by the participants. Most of the comparisons drawn by the participants were between the HSF or BCPA systems and the VUE system (due to their high degree of familiarity with VUE), although some comparisons were made between the two new systems.

It is worthwhile to note that although the participants were more familiar with the VUE system, they were able to interact with each of the systems at an expert level. This was achieved by helping the users achieve their intended tasks if they did not know how to operate the various tools of the HSF and BCPA systems. This was done to allow each of the three systems to prevent bias from being introduced by lack of training or knowledge of any particular system, and to allow a fair comparison of each system.

3.5.1. Opinions of the Hybrid Spatio-temporal Filtering (HSF) system

All four participants stated that they found the interactive HSF system to be useful and easy to use. On the topic of usefulness, one participant said "I find that it's something that can be useful, because when people do something illegal, sometimes they'll follow certain patterns, and it gives us a better understanding of activities that are not normal". Other participants noted that "you could drill down and get more information, ask more questions, filter it or make it more complex, if you wanted", or that it was simply "more in depth".

Its ability to highlight patterns was also found to be superior to that of VUE, with one participant stating, after having found an anomalous pattern: "I was quite surprised at how well you could see the actual fishing patterns, those back and forth and back and forth, all over the place. I'm going to find out what boat that was! You've got me quite intrigued, and it's something I would have not noticed otherwise". The relative ease exhibited by participants in finding patterns in which they were interested was linked to their effective use of the interactive hybrid filtering tools. They intuitively understood how the path complexity and velocity threshold controls functioned, which enabled them to focus on their intended task, rather than how to accomplish it.

Two of the participants also noted that most of the things that could be done using the HSF system could also be done using the VUE system, but with one noting "it would take me much much longer to do the same with the VUE system". Likewise, all the participants brought up the issue that there was some functionality within the VUE system that was not available in the HSF system, primarily related to external data sources, such as fishing licenses. However, they understood that the HSF system could not get access to those data sources due to the confidentiality of the data.

3.5.2. Opinions of the Behavioural Change Point Analysis (BCPA) system

Two out of four participants stated that they found the automated techniques within the BCPA system useful, with the two other participants identifying some potential but with significant reservations. The primary concerns were the perceived inaccuracies in the identification of behavioural change points and the reluctance to rely on a "black box". All participants mentioned that the technique would have to be very robust for it to be used as evidence in a court of law, and that the perceived inconsistencies, where the participants thought the BCPA system identified behavioural change points either too early or too late, did not lend credence to legal arguments. Further, one participant noted that "we need the system to be beyond doubt", which suggests that statistically-based methods may not be particularly well received in this specific field. This sentiment was echoed by another participant, stating that "I want a system that I know I can trust".

Two general issues were identified with the BCPA system's segmentation of the vessel data points. Often, the technique would appear to identify a change in the behaviour of a vessel either too early or too late compared to a visual identification of the vessel tracks available with either the VUE system or the HSF system. This problem was due to the statistical method by which data points are considered similar or different from the ones that precede or follow them, resulting in some inconsistencies in accurately identifying where the change in behaviour begins. This problem could potentially be reduced with data of higher temporal resolution. Another issue was that, by its very nature, the BCPA algorithm does not have the ability to merge related patterns that are disconnected into a common cluster. For instance, it does not classify a vessel coming into port, and going back into port, as the same patterns, even though they are related. Instead, its conceptual operation partitions the connected waypoints in a movement path into multiple segments, based on automatically selected waypoints that divide these segments. This led to some confusion among the participants.

The precision of the visual identification of activity patterns using the HSF system was also generally seen as higher than that of the BCPA system. One participant stated that "although they were close, they weren't more precise". Other participants were more open to the legal robustness of the interactive method, commenting that "if you were going to court, you might want to say 'I picked these ones'", instead of having a system that automated the process.

3.5.3. BCPA segment relevance

In addition to the free-form data analysis done with the three systems, participants were also asked to identify, where possible, the points at which the exhibited behaviour of a vessel changed in what they deemed was a significant fashion, using the HSF system. The break points identified interactively were then looked at using the BCPA system. This was done to verify whether that approach consistently found meaningful change points, and also to gauge whether it might have found other change points that the experts may have missed. The intent was to help participants gain a better understanding of the strengths and weaknesses of interactivity and automation.

The segments identified by the BCPA method were generally found to be meaningful by users, but with a significant amount of re-adjustment required in their ways of thinking. For instance, one participant mentioned “at first they didn’t make sense, because I was focusing on the total amount of vessels that I saw in the groupings”. However, when examining them individually, the same participant noted that “it made a lot more sense”, and another participant mentioned that the identified segments were meaningful “in most cases”. In the few cases where the segments identified were not meaningful to the participants, they noted that this was very worrisome, from an enforcement perspective. The ramifications of this in a court case were brought up, for instance, saying that “If there’s any flaw that can be identified, in any system, they [the defense] would jump right at it”.

3.5.4. Preferred method in terms of ease-of-use

In terms of ease-of-use, half of the participants said they would rather use the HSF system, whereas the other half stated that they preferred the VUE system, mainly due to familiarity. The participants were experts in the use of the VUE system, and understandably were much more proficient in its use than of the two new prototype systems. The BCPA system was not preferred by any participant, even though its interface was very similar to that of HSF, suggesting that the participants’ perceptions of HSF’s ease-of-use was positively influenced by the interactive nature by which the data could be explored, rather than the interface. The seeming dislike for BCPA can be explained by some of the participants' trouble in understanding how this method functions, and their concerns as to the legal issues surrounding it. One participant noted that "it seems fairly accurate, but like everything, it needs some fine tuning", suggesting that if the identified break points were more precise and understandable it may become more appealing.

A particularly interesting outcome here is the fact that participants' preferences in terms of ease-of-use were split between the existing VUE system and the highly interactive HSF system, even though both had interactive controls that would need to be manipulated to explore and filter the data. The automated methods in BCPA were not viewed as being easy to use, even though the degree of interactivity in the interface was significantly lower than in the other methods. It is probable that the perceived lower usefulness of this technique, in contrast to VUE and HSF, affected their opinions as to its ease-of-use.

3.5.5. Helpfulness of the techniques

Participants generally found that the interactive analysis approach supported by the HSF system was more powerful than the approach taken by VUE, due to its ability to visually filter and highlight patterns of interest. However, they were split as to the helpfulness of the BCPA system. One participant found BCPA's automated approach more helpful than VUE because "it would help me pin-point things and see where there were changes", whereas another participant stated that "I feel like the automated technique might confuse the user more than anything".

3.5.6. Combination of techniques

All participants expressed interest in combining some aspects of the different techniques. The idea of merging the complex interactive filtering within the HSF system with the familiarity and context-specificity of VUE was very popular. One participant stated: "if we could combine some of the features from our VUE system with the interactive [HSF] system, I think it would be a knockout". This sentiment was shared by all participants, to some degree. However, some of this opinion regarding the merging of the novel approaches of HSF within the existing framework of VUE could be attributed to resistance to change, or the need for some extra data access features in order for HSF to be a viable replacement for VUE in this fisheries enforcement setting.

Combining the automated features of the BCPA system with VUE was much more contentious. One participant stated that combining both approaches may help "if the automated technique [BCPA system] got some work, because right now I don't really see how it would be useful", whereas another stated that "for the fisheries officers that want a quick peek at what's going on, having a system that [automatically] colours different activities, I think could help a lot". This feature, of having a quick look at data, is also available in the HSF system by using a set of pre-determined or user-configured filter signatures.

In terms of combining the interactive HSF system with the automated BCPA system, the participants were fairly enthusiastic. Getting a "best of both" approach was seen as something that would be able to not only address some of the shortcomings of either technique, but as a way to provide some form of solution to the "black box" problem with the BCPA system.

3.5.7. Overall preferences

In general, the participants favoured the VUE system over the other two systems. This can be explained by the fact that the two prototype systems did not have access to the information about vessels and licenses. This technical limitation resulted from the fact that such additional data were not made available for the study due to confidentiality concerns. Participants were made aware of this fact early on in the study, as well as the fact that had the data been available, it could have been integrated with both of the prototype systems. However, it is clear that this still played a factor in which system they preferred.

In terms of the approach, participants clearly favoured the highly interactive approach promoted by the HSF system, for a number of different reasons. Due to the constraints imposed by the web-based framework upon which it was built, the VUE system appeared to be slow and cumbersome, whereas the automated approach within the BCPA system worried the experts in terms of it taking away some of their control and being a potential liability in a court of law.

3.6. Discussion

The interviews conducted with each participant revealed interesting and unexpected information. The automated pattern extraction in the BCPA system was seen as very interesting to most experts, although approached with some degree of reticence. The lack of control seemed worrisome to the experts, who are used to having complete control over the querying of their data. The "black box" nature meant that there was no easy way to explain why two segments had been identified as different behaviours.

To this was added the fact that there were a few perceived inaccuracies in the detection of behavioural change points. Although generally the break points identified were within a few records (hours, in this case) of being correct, this clearly did not seem to be accurate enough; the experts were expecting a high degree of accuracy from this automated method, similar to what they would be able to achieve themselves manually. Also, there were a few cases where the automated system identified change points that did not represent any change in the data that was meaningful from the analysts' perspective.

In addition, there was a split between the participants regarding how the identified behaviours should be displayed. Most participants were mildly annoyed at the idea that a vessel steaming to the fishing ground, and back from it, would not be identified as the same behaviour by the use of the same colour. However, one participant did point out that these are inherently different behaviours and therefore that they should not get clustered together, noting: "I actually like that, because you can tell whether it [the vessel] was coming or going". Identifying similar clusters and merging them if they are below some similarity threshold might work well to improve the usefulness of the BCPA system. However, this approach would also involve significantly more processing, and would require a formal validation to be useful in a legal capacity.

The interactive filtering techniques within the HSF system were generally seen as a good alternative to VUE's approach. Currently, the VUE system only allows for limited non-interactive filtering of data, on parameters such as vessel type, general area, and temporal range. The potential for combining the filtering of data based on velocity and the geometric complexity of movement paths, and then saving these settings as filter signatures within the HSF system was seen as a significant improvement.

Moreover, the general approach used by the HSF system that allows the analyst to focus on patterns rather the individual vessels, seemed attractive to the experts. Currently, enforcement officers know of a few vessels, ostensibly repeat offenders, which they focus on. Creating signatures for particular patterns of interest, based for instance on the activities of a particular known offender, could allow the experts to gain a better situational awareness and identify interesting or anomalous behaviour in a more general and exploratory manner. Two participants commented that having pre-determined filtering signatures would save them some time, and make the data analysis process more user friendly. It was suggested that these signatures could go through a rigorous validation process, and then be distributed to all analysts for use. This would not preclude analysts from creating their own signatures, but provide a pool of signatures to select from if required. Analyzing the velocity and complexity of the paths of a large number of vessels, while knowing the type of fishing being done, could also allow for the derivation of a set of parameters that could be used to define a signature that targets vessels exhibiting a similar behaviour.

The approach used by the VUE system puts a heavy emphasis on data analysis, rather than data exploration. Many additional data sources are available, providing access to information such as quotas, zones, permits, and amount of fish landed at shore. Moreover, due to the limited ability to interactively manipulate the data, analysts generally use the VUE system to provide answers to specific questions or hypotheses formulated primarily based on *a priori* knowledge. VUE's approach is to promote this vessel-by-vessel investigation, which seems to lead to a form of tunnel vision wherein the VUE system is used to confirm suspicions, rather than to explore all of the data to find potential offenders. In that sense, VUE's approach is considerably different from the other two systems in this study: VUE promotes analysis, based on well-defined hypotheses, through complex data manipulation and the merging of data from multiple sources, while the HSF and BCPA systems promote visual exploration through either interactivity or automated segmentation.

Table 3.1 summarizes the benefits and drawbacks of the three approaches. Generally, the VUE system's approach focused on providing maximum functionality. The BCPA system, in contrast, focused on identifying movement patterns quickly, but with limitations in terms of transparency and control over the automation process. The HSF system falls somewhere in between these approaches, offering increased speed of analysis, a small degree of automation via the saved signatures, and a more effective visualization system, but at the cost of some functionality and transparency. By moving the focus from data analysis to data exploration, through the use of complex filtering signatures, some of the more in-depth analytical functionality of the HSF system necessarily became diminished with respect to VUE's approach. This, in turn, increased the ease of use of the system and the speed of analysis. However, by virtue of having complex filtering signatures, the transparency of the approach was also diminished, and analysts may not always be aware of the impact of their filters, particularly when it comes to the less familiar path complexity (i.e., fractal dimension) filter.

Table 3.1. A general overview of the characteristic of each approach, based on participant responses.

	VUE	HSF	BCPA
Access to data	High	Medium	Medium
Ease-of-use	Medium	High	Low
Transparency	High	Medium	Low
Functionality	High	Medium	Low
Speed of analysis	Low	Medium	High

3.7. Conclusions

The ever-increasing traffic on the world's oceans led the adoption of new technologies for improving maritime safety, helping manage activities taking place in the maritime environment, and ensuring compliance with national and international regulations. According to the United Nations Conference on Trade and Development, world seaborne trade in 2011 was 46.6% higher than in 2000, and 220% higher than in 1990 (UNCTAD, 2011). In fisheries, many countries around the word have adopted the VMS system to support fisheries enforcement. While such vessel tracking systems have now become central to many fisheries enforcement activities, they generate very large sets of movement data that can be challenging to analyze. Different approaches can be used to derive information from these large data sets, ranging from interactive approaches to fully automated ones. Some approaches focus on the analysis or visualization of individual vessels, others on groups of vessels, and others on movement patterns. The main challenge with all of these approaches is how to identify features that are of interest to an analyst in an efficient manner.

This paper compared three approaches: (1) an existing system named VUE, currently used for fisheries enforcement, which focuses on individual vessels; (2) an interactive Hybrid Spatio-temporal Filtering (HSF) system, which allows filtering on velocity and fractal dimension to highlight specific movement patterns with user-specified characteristics; and (3) a Behavioural Change Point Analysis (BCPA) system, that extracts patterns based on statistical changes in behaviour. All three techniques were rigorously tested through a field trial, and presented the benefits and drawbacks of each method when used by fisheries enforcement officers in the context of their daily work. VUE was seen as a slower system, but with a much greater analytical power and control due to its ability to connect to external databases, and the specific analytical functions it offers. The HSF system was faster at exploring the data, providing considerable user control, but somewhat constrained analytical power in the context of this prototype implementation. BCPA allowed very fast interaction with the system once the off-line data processing was complete, although this speed of analysis was indicative of the limited functionality available in terms of analytical tools or user controls.

Results indicate that the approach used by the VUE system, which is representative of many similar systems based on web interfaces, does not scale well to the size of the data sets that analysts need to work with on a daily basis. While focusing on individual vessels can work well for small datasets, it quickly becomes ineffective for larger data sets that can include hundreds or thousands of vessels. This study confirmed that approaches based on data geovisualization, filtering, and segmentation are much better in these situations, as they do not require the analyst to visualize all of the data at once, but rather filter out most of the data to allow the analyst to focus on patterns of interest. Such results are consistent with other studies of the field of geovisual analytics that have indicated the superiority of these systems compared to more traditional mapping approaches when exploring large data volumes (Kraak, 2003; Wisniewski et al., 2009).

However, this study also confirms that enforcement officers still require rich analytical functionalities, such as the ability to get details about licenses, reported fish catches, or cross-reference vessel information. As a result, most experts were open to a synergistic approach that could combine HSF's intuitive filtering capabilities, interactivity, and speed with the analytical power provided by VUE. Participants were not in favour of having the BCPA approach integrated into a day-to-day system mostly due it its lack of transparency, but thought it could be appropriate for first pass analyses or as part of an automated warning system.

Multiple parameters were identified as having an impact on analysts' decisions as to the applicability of a particular approach, or combination of approaches, in the fisheries enforcement domain. In particular, ease-of-use, transparency, functionality, and speed of analysis were primary concerns. VUE ranked highly on each of these parameters (Table 3.1) except speed of analysis, due to a combination of the experience the participants had with the system and the specific data analysis activities it was designed to support. While HSF was somewhat less transparent and provided less functionality due to limitations in terms of data it could access, it allowed for quicker data analysis and was viewed as being as easier to use even in light of the highly interactive features for filtering and exploring the data. Although the automatic pattern extraction features in BCPA made it very fast in terms of data analysis, the functionality was limited, the participants did not report it as being particularly easy to use, and the low transparency made analysts uneasy as to the liability it might pose in legal proceedings. In the end, a synergistic approach that combined the positive features from each system was identified as having the potential for being particularly effective for data analysis in the context of fisheries enforcement. It is likely that for other domains, such as animal tracking or container shipping, the combination of data analysis features and the balance between the parameters of ease-of-use, transparency, functionality, and speed of analysis might be significantly different. As such, further research is required to analyze and understand the specific data analysis needs in different domains, and to study how geovisual analytics tools can be optimized to support the needs of data analysts.

3.8. Acknowledgements

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for the funding of this project through the Strategic Projects Grant STPGP 365189-08, and the Canada Foundation for Innovation (CFI) and Memorial University of Newfoundland for providing the laboratory infrastructure. We also would like to thank Fisheries and Oceans Canada, and especially Jerry Black and Trevor Fradsham and their groups, for their collaboration and having provided access to the data, as well as the fisheries enforcement officers who participated in this study.

3.9. References

Andrienko, G., Andrienko, N., & Heurich, M. (2011a). An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, *25*(9), 1347-1370. doi:10.1080/13658816.2011.556120

Andrienko, G., Andrienko, N., & Wrobel, S. (2007a). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46. doi:10.1145/1345448.1345455

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I., & Wachowicz, M. (2008). Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization*, *7*(3-4), 173-180. doi:10.1057/ivs.2008.23

Andrienko, G., Andrienko, N., Fischer, R., Mues, V., & Schuck, A. (2006). Reactions to geovisualization: an experience from a European project. *International Journal of Geographical Information Science*, *20*(10), 1149-1171. doi:10.1080/13658810600816524

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S., & Wrobel, S. (2011b). From movement tracks through events to places : Extracting and characterizing significant places from mobility data. *Proceedings of the 2011 IEEE Symposium on Visual Analytics Science and Technology* (pp. 159-168). Providence, USA: IEEE.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D. A., Kraak, M.-J., MacEachren, Alan M., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839-857. doi:10.1080/13658810701349011

Andrienko, N., & Andrienko, G. (2007). Designing visual analytics methods for massive collections of movement data. *Cartographica*, *42*(2), 117-138. doi:10.3138/carto.42.2.117

Andrienko, N., & Andrienko, G. (2008b). Supporting visual exploration of massive movement data. *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '08)* (pp. 474-475). Napoli, Italy: ACM Press. doi:10.1145/1385569.1385663

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., & Gerlotto, F. (2007). Scale-invariant movements of fishermen: the same foraging strategy as natural predators. *Ecological Applications*, *17*(2), 331-337.

Chang, S.-K., Liu, K.-Y., & Song, Y.-H. (2010). Distant water fisheries development and vessel monitoring system implementation in Taiwan—History and driving forces. *Marine Policy*, *34*(3), 541-548. Elsevier. doi:10.1016/j.marpol.2009.11.001

Demšar, U., & Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, *24*(10), 1527-1542. doi:10.1080/13658816.2010.511223

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., & Die, D. (2005). Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(3), 611-622. doi:10.1139/f04-219

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, *7*(3-4), 240-252. doi:10.1057/palgrave.ivs.9500182

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, *63*(7), 1057-1066. doi:10.1007/s00265-009-0739-0

Enguehard, R. A., Devillers, R., & Hoeber, O. (2011). Geovisualization of fishing vessel movement patterns using hybrid fractal / velocity signatures. *Proceedings of the First International GeoViz Workshop* (pp. 1-2). Hamburg, Germany.

Enguehard, R. A., Hoeber, O., & Devillers, R. (submitted). Interactive exploration of movement data : A case study of geovisual analytics for fishing vessel analysis. *Information Visualization*, 30p.

Gottfried, B. (2011). Interpreting motion events of pairs of moving objects. *GeoInformatica*, *15*(2), 247-271. doi:10.1007/s10707-009-0095-2

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395-408. doi:10.1111/j.1461-0248.2009.01293.x

Hintzen, N. T., Piet, G. J., & Brunel, T. (2010). Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data. *Fisheries Research*, *101*(1-2), 108-115. doi:10.1016/j.fishres.2009.09.014

Jern, M., Åström, T., & Johansson, S. (2008). GeoAnalytics tools applied to large geospatial datasets. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 362-372). Columbus, USA: IEEE. doi:10.1109/IV.2008.27

Johansson, S., & Jern, M. (2007). GeoAnalytics visual inquiry and filtering tools in parallel coordinates plots. *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems* (pp. 1-8). Seattle, USA: ACM Press. doi:10.1145/1341012.1341055

Kim, R., & Hogan, P. (2011). World Wind JAVA SDK. Retrieved from http://worldwind.arc.nasa.gov/java/

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference* (pp. 1988-1995). Durban, South Africa: ICA.

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, *8*(1-6), 185-203. doi:10.1016/S0968-090X(00)00017-6

Laxhammar, R., Falkman, G., & Sviestins, E. (2009). Anomaly detection in sea traffic - a comparison of the Gaussian Mixture Model and the Kernel Density Estimator. *Proceedings of the 12th International Conference on Information Fusion* (pp. 756-763). Seattle, USA: IEEE.

Lundblad, P., Jern, M., & Forsell, C. (2008). Voyage analysis applied to geovisual analytics. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 381-388). Columbus, USA: IEEE. doi:10.1109/IV.2008.32

Mills, C. M., Townsend, S. E., Jennings, S., Eastwood, P. D., & Houghton, C. A. (2006). Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science*, *64*(2), 248-255. doi:10.1093/icesjms/fsl026

Molenaar, E. J., & Tsamenyi, M. (2000). Satellite-based vessel monitoring systems for fisheries management: international legal aspects. *The International Journal of Marine and Coastal Law*, *15*(1), 65-110. doi:10.1163/157180800020492643

Mullowney, D. R., & Dawe, E. G. (2009). Development of performance indices for the Newfoundland and Labrador snow crab (*Chionoecetes opilio*) fishery using data from a vessel monitoring system. *Fisheries Research*, *100*(3), 248-254. doi:10.1016/j.fishres.2009.08.006

Murawski, S. A., Wigley, S. E., Fogarty, M. J., Rago, P. J., & Mountain, D. G. (2005). Effort distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of Marine Science*, *62*(6), 1150-1167. doi:10.1016/j.icesjms.2005.04.005

Mårell, A., Ball, J. P., & Hofgaard, A. (2002). Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology*, *80*(5), 854-865. doi:10.1139/Z02-061

Nams, V. O. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia*, *143*(2), 179-88. doi:10.1007/s00442-004-1804-z

Plaisant, C. (2004). The Challenge of Information Visualization Evaluation. *Proceedings of the Conference on Advanced Visual Interfaces* (pp. 109-116). Gallipoli (Lecce), Italy. IEEE. doi: 10.1145/989863.989880

Raymond, B., & Hosie, G. (2009). Network-based exploration and visualisation of ecological data. *Ecological Modelling*, *220*(5), 673-683. doi:10.1016/j.ecolmodel.2008.12.011

Rocha, J. A. M. R., Times, V. C., Oliveira, G., Alvares, L. O., & Bogorny, V. (2010). DB-SMoT: A direction-based spatio-temporal clustering method. *Proceedings of the 5th IEEE International Conference Intelligent Systems* (pp. 114-119). London, UK: IEEE. doi:10.1109/IS.2010.5548396

Rodighiero, D. (2010). Guidelines to visualize vessels in a geographic information system. *Proceedings of the 14th International Conference on Information Visualisation* (pp. 455-459). Salt Lake City, USA: IEEE. doi:10.1109/IV.2010.70

Saitoh, S.-I., Mugo, R., Radiarta, I. N., Asaga, S., Takahashi, F., Hirawake, T., Ishikawa, Y., Awaji, T., In, T. & Shima, S. (2011). Some operational uses of satellite remote sensing and marine GIS for sustainable fisheries and aquaculture. *ICES Journal of Marine Science*, *68*(4), 687-695. doi:10.1093/icesjms/fsq190

Schwehr, K. D., & McGillivary, P. A. (2007). Marine ship automatic identification system (AIS) for enhanced coastal security capabilities: an oil spill tracking application. *Proceedings of the 2007 Oceans Conference* (pp. 1-9). Vancouver, Canada: IEEE. doi:10.1109/OCEANS.2007.4449285

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools. *Proceedings of the 2006 AVI workshop on Beyond Time and Errors Novel Evaluation Methods for Information Visualization* (pp. 1-7). Venice, Italy: ACM Press. doi:10.1145/1168149.1168158

Thomas, J., & Cook, K. (2005). *Illuminating the path: research and development agenda for visual analytics* (p. 194). Los Alamitos, USA: IEEE Computer Society.

Tomaszewski, B. M., Robinson, A. C., Weaver, C., Stryker, M., & MacEachren, A.M. (2007). Geovisual analytics and crisis management. *Proceedings of the 4th International Information Systems for Crisis Response and Management (ISCRAM) Conference* (pp. 1-8). Delft, Netherlands.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed., p. 200). Cheshire, CT, USA: Graphics Press.

UNCTAD (2011). Review of Maritime Transport 2011. *United Nations Conference on Trade and Development.* 229p. Retrieved from: http://www.unctad.org/en/docs/rmt2011_en.pdf

Ware, C. (2004). *Information visualization: perception for design* (2nd ed., p. 486). San Francisco, USA: Morgan Kaufmann.

Willems, N., van de Wetering, H., & van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics Forum*, *28*(3), 959-966. doi:10.1111/j.1467-8659.2009.01440.x

Wisniewski, P. K., Pala, O., Lipford, H. R., & Wilson, D. C. (2009). Grounding geovisualization interface design: a study of interactive map use. *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 3757-3762). Boston, USA: ACM. doi: 10.1145/1520340.1520567

With, K. A. (1994). Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology*, *9*(1), 25-36. doi:10.1007/BF00135076

Witt, M. J., & Godley, B. J. (2007). A step towards seascape scale conservation: using vessel monitoring systems (VMS) to map fishing activity. *PloS ONE*, *2*(10), e1111, 1-5. doi:10.1371/journal.pone.0001111

Zhao, J., Forer, P., & Harvey, A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, *7*(3-4), 198-209. doi:10.1057/palgrave.ivs.9500184

Chapter 4 Conclusions

4.1. Summary

The effective analysis of large movement data sets has become more important since location tracking devices have started being used in a variety of national and international programs, as well as everyday life. Gaining useful knowledge from these large and complex data sets can be a challenging task. Analyzing these data sets using traditional methods is often difficult, or impossible, and can lead to significant patterns or anomalies being missed altogether by analysts. This can be due to a number of issues, such as information overload, the inability to visualize all the information at once due to the size of these data sets, or having the data being too complex to fully understand.

Various approaches have been proposed in the literature to address these issues, such as geovisual analytics (Ho & Jern, 2008; Jern et al., 2008; Lundblad et al., 2008; Tomaszewski et al., 2007), filtering (Andrienko et al., 2011; Saitoh et al., 2011), clustering (Chen et al., 2011; Rinzivillo et al., 2008; Rocha et al., 2010), pattern detection (Focardi et al., 1996; Kwan, 2000; Pelot & Wu, 2007), statistical analysis (Bertrand et al., 2007; Gurarie et al., 2009; Mårell et al., 2002; Underwood & Chapman, 1985), or artificial intelligence (Bomberger et al., 2006; Choi et al., 2006; Mano et al., 2010).

This thesis proposed an approach for movement data analysis that focuses on the identification and visualization of patterns, called Hybrid Spatio-temporal Filtering (HSF). HSF relies on the concept of fractal dimension as a proxy for behaviour (Bertrand et al., 2007; With, 1994), building on existing methods for the estimation of fractal dimension (Mandelbrot, 1967; Sevcik, 1998; Theiler, 1990), and combining it with velocity filtering (Saitoh et al., 2011), into an interactive filtering interface allowing analysts to create behavioural signatures. The HSF approach is both visually effective and interactive, relying on core concepts of geovisual analytics (Andrienko et al., 2007), such as multiple coordinated views (Andrienko & Andrienko, 2007; Wang Baldonado et al., 2000), data filtering, and details on demand. This approach was implemented as a software prototype and evaluated through two separate field trials for a case study in fisheries enforcement.

The outcome of the field trials confirmed that the approach is useful, easy to use, and superior to some existing or recently proposed approaches. Moreover, it was found that there is a balance, likely specific to each domain, which must be met between ease of use, transparency, functionality, and speed of analysis for an approach to be acceptable for a particular analysis activity. In the case of fisheries enforcement, transparency must remain high due to legal concerns, which can be achieved through the use of interactive controls, but some level of automation is also required to enable the analysis of the large volumes of data generated by the fishing fleet. The identification of this balance of parameters, and the approach presented to meet this balance for the field of fisheries enforcement, are important results that may be applied to many other fields that deal with movement data.

Five research questions were presented in section 1.3, and have been answered as follows:

- **Which movement characteristics can be used to build signatures that identify specific behaviours?**

This thesis suggested combining the fractal dimension and velocity parameters of movement data to characterize movement patterns. The field trials presented in section 2.6, and the results presented in section 2.7, confirmed the value of such an approach. Analysts were able to extract patterns of interest by modifying the filter parameters (fractal dimension thresholds, fractal dimension window size, and velocity thresholds) in an iterative fashion. The pattern signatures were then assessed by analysts to identify specific types of behaviours which they were interested in, such as types of fishing, or travel between ports and fishing grounds. Moreover, analysts found that these filter signatures were easy to use, providing the results which they expected.

- **How can a geovisual analytics environment be designed to effectively allow for visual exploration of large movement data sets?**

A geovisual analytics environment was designed, prototyped on top of the NASA World Wind virtual globe, and presented in section 2.5. This visualization and analysis environment displayed the data in an interactive and visually intuitive fashion, following the general principles of information visualization, such as perceptually distinct colour encodings (Ware, 2004), the reduction of chartjunk (Tufte, 2001), and real-time interaction. The standard virtual globe interactions (pan, tilt, zoom) were supplemented by data filtering, drill-down, details on demand, and multiple coordinated views. This combination of highly interactive tools allowed analysts to deal with the very large data sets produced by Vessel Monitoring Systems (VMS), consisting of millions of records per month.

HSF's approach of combining temporal, velocity, and fractal dimension filtering in a multiple coordinated views geovisual analytics environment was shown to be effective for allowing the exploration of large movement data sets in the field trial results of section 2.7. In this field trial, analysts used a prototype implementation of this approach and were able to effectively explore VMS data sets of their choice, to perform similar kinds of analyses that they would undertake during their regular duties. The usefulness of the approach was demonstrated through the high perceived and overall usefulness (Figures 2.12 and 2.14, respectively) of the approach as assessed during the field trial.

- **How can a geovisual analytics environment be designed to maximize usability among analysts?**

A number of aspects were found to play important roles in the usability of systems for exploring movement data. In particular, providing graphical representations of the data, as well as allowing the filtering of data, enabled analysts to explore large data sets with more ease, and isolate the patterns that were interesting to them faster than with other systems. This was demonstrated in both the mean perceived ease-of-use of Figure 2.13, in section 2.7, and the participant responses reported in section 3.5.1. The use of multiple coordinated views, with spatial, temporal, and velocity views, aided in the understanding of the impact of modifying filter settings. Adherence to a single mode of filter operation, namely the slider paradigm, also helped prevent analysts from needing to learn different ways to filter data. Finally, for the more unusual filtering type of fractal dimension filtering, analysts were isolated from much of the quantitative complexity by providing them with sliders labeled from low complexity to high complexity, instead of the actual fractal dimension values.

This approach proved beneficial, as demonstrated by the ease of use of the approach being on par with the VUE system with which participants were very familiar, as seen in section 3.5.4. During this field trial, analysts were only provided with a short training session, after which they were free to ask for help. Participants generally showed little difficulty in making use of the various filtering mechanisms to explore their data.

- **Does the approach developed improve analysts' ability to extract movement patterns from their data sets?**

The validation of the HSF approach was performed using two separate field trials. The first field trial looked at the overall usefulness and ease of use of the approach and of some of its components. This validation was achieved through an extended use of the prototype system by analysts performing their regular tasks with actual data, followed by a questionnaire adhering to the Technology Acceptance Model (TAM) (Davis, 1989). The main purpose was to provide a quantitative indication as to whether the approach itself was valid, but not to make statements as to its effectiveness compared to other approaches. Figure 2.14 in section 2.7 clearly demonstrates the validity of the HSF approach, with participants rating the various components of the approach at levels exceeding that of their usual tools. When asked to take into account all of the various factors, participants still preferred the HSF approach overall.

To provide some context as to how this approach compares to other alternatives, a second field trial was presented in section 3.4. It used some of the same participants from the first field trial, and compared HSF with two other approaches: one was the system currently used by DFO to analyze vessel movement data, and the other approach was based on the Behavioural Change Point Analysis (BCPA) method. As presented in section 3.5.7, this field trial showed that analysts preferred the HSF approach over the other tested approaches, for the case study of fisheries enforcement. However, it was also suggested that this would not necessarily be the case for all analysis activities and contexts; an important outcome of this research is a guideline as to how to design geovisual analytics systems to match the particularities of specific fields, particularly with respect to the balance of ease-of-use, transparency, functionality, and speed of analysis seen in Table 1 of section 3.6.

- **How does the proposed approach compare in terms of usability and effectiveness with existing approaches?**

Section 2.3 presented many approaches that already exist for dealing with movement data in a visual fashion. Section 3.4 then compared three approaches: the proposed HSF approach, a system named VUE based on traditional mapping methods and currently in use by DFO, and another one based on the BCPA method. Interactivity of these approaches ranges from very limited (BCPA), to somewhat interactive (VUE), to very interactive (HSF). Transparency, or the ability for analysts unfamiliar with the inner workings of the systems to determine how the data were manipulated, ranges from very transparent (VUE), to somewhat transparent (HSF), and to a low level of transparency

(BCPA). The usability of the HSF approach was confirmed, in section 3.5.4, as being on par, and potentially superior, to that of other systems. The effectiveness of the HSF approach was also confirmed, in section 3.5.5, as being higher than other alternatives.

It was noticed that the acceptance of particular approaches was directly related to the balance of interactivity, transparency, and automation. In the context of this case study, legal concerns, such as admissibility of evidence in court, were very important. The lack of transparency provided by automated methods, and lack of interactivity provided by traditional methods, made HSF a preferred choice. However, in a different context, such as animal tracking or automated alert systems, a different balance between these elements may need to be found (Lloyd & J. Dykes, 2011). The identification of these elements, which affect the acceptance of geovisual analytics systems, is one of the key contributions of this study.

The research hypothesis for this thesis was that "**a geovisual analytics system allowing filtering on multiple characteristics of movement will improve analysts' abilities to both deal with large amounts of movement data and find interesting patterns within them**". This hypothesis can be confirmed based on the validations presented in the Chapters 2 and 3. In section 2.6, analysts were asked to explore a set of VMS data consisting of over a million records. Participants accomplished this task with relative ease, as reported in Figure 2.13 of section 2.7. In section 3.5.1, it was shown that participants were able to identify interesting movement patterns using the same approach. Further, section 3.6 validated the proposed approach within a more generalized context, provided that the balance for ease-of-use, transparency, functionality, and speed of analysis remains the same. Understanding this balance, and designing geovisual analytics approaches to match it, is a key aspect in creating successful systems for specific disciplines and geo-temporal data analysis contexts.

4.2. Future works

This research demonstrated the value of signature-building approaches for filtering large amounts of movement data. However, in doing so, it also highlighted limitations of such approaches, as well as issues with user acceptance of these approaches. There are also a number of improvements that could be made to the HSF approach, requiring further study and validation, particularly with respect to the performance of the underlying system. As suggested in Chapter 3, a better understanding of the balance between ease-of-use, transparency, functionality, and speed of analysis would be beneficial. Moreover, a more in depth study into the impact of changing the temporal resolution of the data would be of great interest, both in terms of the validity of the HSF approach and of pure scientific interest. It is probable that an increased temporal resolution would result in more accurate fractal dimension, velocity, and heading estimates, but would also increase the computational load. Studying these effects could be done using AIS data down-sampled to various resolutions, for instance.

Additionally, the HSF approach could potentially be applied to a number of other domains. The movement patterns of herds or flocks of animals, migration patterns of birds, mammals or fish, or the predatory patterns of large carnivores, could all be studied using HSF. However, the HSF approach is not only limited to ecological data, the movement patterns of students throughout a university campus, using mobile phone localization, could be another potential use. The only real requirement for a data set to be compatible with the HSF approach is that it be movement data with a behavioural component.

The HSF approach functions primarily due to its combination of fractal dimension and velocity. Neither form of filtering is novel, with velocity being used in a wide range of systems (Bomberger et al., 2006; Gurarie et al., 2009; Saitoh et al., 2011), and fractal dimension being a common metric used in ecology (Mårell et al., 2002; Nams, 2005; With, 1994), but the combination of both had never been explored within a geovisual analytics framework. Unfortunately, fractal dimension estimation, particularly using a moving window, is computationally expensive. Calculating the cubic Hermite spline of movement paths, to provide accurate interpolation, is also computationally expensive (Hintzen et al., 2010; Tremblay et al., 2006). As a result, a number of potential improvements were not incorporated into the HSF prototype implementation, such as estimating velocity based on spline interpolation instead of linear interpolation, simulating a higher VMS sample rate using spline points to provide a more accurate fractal dimension estimation, or creating graphical representations of the range of fractal dimension values for the data set in a similar manner to the other coordinated views. It is probable that by implementing these improvements, the prototype would be more useful, but this would require a significant increase in processing power, perhaps exceeding what is commonly available.

Additionally, it would be beneficial to study the factors affecting acceptance of new approaches to movement data exploration, to gain a better understanding of the issues of balance between the various parameters identified in Chapter 3. Some work has been done on this topic (Lloyd & J. Dykes, 2011), but more needs to be done to understand the factors affecting this balance. Gaining a better understanding of these factors would enable the results from this research, as well as other similar studies, to be applied to a greater number of situations, as well as providing potential avenues for more effective approaches.

The signature-building aspect of this research could also be extended to accept a wide range of inputs, beyond that of manipulating the filter controls. For instance, it may be possible to design a system to control the signature-building process based on available pattern data. A number of different approaches have already been explored, such as using neural networks, in the form of Self-Organizing Maps (SOMs) (Bomberger et al., 2006; Choi et al., 2006). A genetic algorithm could also be used to present analysts with a set of progressively refined patterns to choose from. Alternatively, computer vision systems and pattern recognition algorithms could be used to determine filtering signatures based on patterns sketched by analysts. Each of these different approaches would involve analysts at different levels, and provide them with varying degrees of control. As a result, they would have to be carefully selected to match the specific acceptance criteria of the domain in which they would be used.

4.3. References

Andrienko, G., & Andrienko, N. (2007). Coordinated multiple views: a critical view. *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization* (pp. 72-74). Zürich, Switzerland: IEEE. doi:10.1109/CMV.2007.4

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S., & Wrobel, S. (2011). From movement tracks through events to places : Extracting and characterizing significant places from mobility data. *Proceedings of the 2011 IEEE Symposium on Visual Analytics Science and Technology* (pp. 159-168). Providence, USA: IEEE.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D. A., Kraak, M.-J., MacEachren, Alan M., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839-857. doi:10.1080/13658810701349011

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., & Gerlotto, F. (2007). Scale-invariant movements of fishermen: the same foraging strategy as natural predators. *Ecological Applications*, *17*(2), 331-337.

Bomberger, N., Rhodes, B., Seibert, M., & Waxman, A. (2006). Associative learning of vessel motion patterns for maritime situation awareness. *Proceedings of the 9th International Conference on Information Fusion* (pp. 1-8). Florence, Italy: IEEE. doi:10.1109/ICIF.2006.301661

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. *Proceedings of the 27th International Conference on Data Engineering* (Vol. 4, pp. 900-911). Hannover, Germany: IEEE. doi:10.1109/ICDE.2011.5767890

Choi, K.-H., Kim, J.-S., Kim, Y. S., Yoo, M.-A., & Chon, T.-S. (2006). Pattern detection of movement behaviors in genotype variation of *Drosophila melanogaster* by using self-organizing map. *Ecological Informatics*, *1*(3), 219-228. doi:10.1016/j.ecoinf.2005.12.002

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems*, *13*(3), 319-340.

Focardi, S., Marcellini, P., & Montanaro, P. (1996). Do ungulates exhibit a food density threshold ? A field study of optimal foraging and movement patterns. *The Journal of Animal Ecology*, *65*(5), 606-620. doi:10.2307/5740

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395-408. doi:10.1111/j.1461-0248.2009.01293.x

Hintzen, N. T., Piet, G. J., & Brunel, T. (2010). Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data. *Fisheries Research*, *101*(1-2), 108-115. doi:10.1016/j.fishres.2009.09.014

Ho, Q., & Jern, M. (2008). Exploratory 3D geovisual analytics. *Proceedings of the 2008 International Conference on Research, Innovation and Vision for the Future in*

Computing and Communication Technologies (Vol. d, pp. 276-283). Ho Chi Minh City, Vietnam: IEEE. doi:10.1109/RIVF.2008.4586367

Jern, M., Åström, T., & Johansson, S. (2008). GeoAnalytics tools applied to large geospatial datasets. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 362-372). Columbus, USA: IEEE. doi:10.1109/IV.2008.27

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, *8*(1-6), 185-203. doi:10.1016/S0968-090X(00)00017-6

Lloyd, D., & Dykes, J. (2011). Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2498-2507. doi:10.1109/TVCG.2011.209

Lundblad, P., Jern, M., & Forsell, C. (2008). Voyage analysis applied to geovisual analytics. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 381-388). Columbus, USA: IEEE. doi:10.1109/IV.2008.32

Mandelbrot, B. B. (1967). How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science*, *156*(3775), 636-638. doi:10.1126/science.156.3775.636

Mano, J.-P., Georgé, J.-P., & Gleizes, M.-P. (2010). Adaptive multi-agent system for multi-sensor maritime surveillance. *Advances in Practical Applications of Agents and Multiagent Systems*, *70*, 285–290. doi:10.1007/978-3-642-12384-9_34

Mårell, A., Ball, J. P., & Hofgaard, A. (2002). Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology*, *80*(5), 854-865. doi:10.1139/Z02-061

Nams, V. O. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia*, *143*(2), 179-88. doi:10.1007/s00442-004-1804-z

Pelot, R., & Wu, Y. (2007). Classification of recreational boat types based on trajectory patterns. *Pattern Recognition Letters*, *28*(15), 1987-1994. doi:10.1016/j.patrec.2007.05.014

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Information Visualization*, *7*(3-4), 225-239. doi:10.1057/palgrave.ivs.9500183

Rocha, J. A. M. R., Times, V. C., Oliveira, G., Alvares, L. O., & Bogorny, V. (2010). DB-SMoT: A direction-based spatio-temporal clustering method. *Proceedings of the 5th IEEE International Conference Intelligent Systems* (pp. 114-119). London, UK: IEEE. doi:10.1109/IS.2010.5548396

Saitoh, S.-I., Mugo, R., Radiarta, I. N., Asaga, S., Takahashi, F., Hirawake, T., Ishikawa, Y., Awaji, T., In, T. & Shima, S. (2011). Some operational uses of satellite remote sensing and marine GIS for sustainable fisheries and aquaculture. *ICES Journal of Marine Science*, *68*(4), 687-695. doi:10.1093/icesjms/fsq190

Sevcik, C. (1998). A procedure to estimate the fractal dimension of waveforms. *Complexity International*, *5*(19), 1-20.

Theiler, J. (1990). Estimating fractal dimension. *Journal of the Optical Society of America A*, *7*(6), 1055-1073. doi:10.1364/JOSAA.7.001055

Tomaszewski, B. M., Robinson, A. C., Weaver, C., Stryker, M., & MacEachren, A.M. (2007). Geovisual analytics and crisis management. *Proceedings of the 4th International Information Systems for Crisis Response and Management (ISCRAM) Conference* (pp. 1-8). Delft, Netherlands.

Tremblay, Y., Shaffer, S. A., Fowler, S. L., Kuhn, C. E., McDonald, B. I., Weise, M. J., Bost, C.-A., et al. (2006). Interpolation of animal tracking data in a fluid environment. *The Journal of Experimental Biology*, *209*(1), 128-40. doi:10.1242/jeb.01970

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed., p. 200). Cheshire, CT, USA: Graphics Press.

Underwood, A. J., & Chapman, M. G. (1985). Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology*, *91*(1-2), 17–43. Elsevier. doi:10.1016/0022-0981(85)90218-7

Wang Baldonado, M. Q., Woodruff, A., & Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '00* (pp. 110-119). New York, USA: ACM. doi:10.1145/345513.345271

Ware, C. (2004). *Information visualization: perception for design* (2nd ed., p. 486). San Francisco, USA: Morgan Kaufmann.

With, K. A. (1994). Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology*, *9*(1), 25-36. doi:10.1007/BF00135076

Bibliography

Andersson, M., Gudmundsson, J., Laube, P., & Wolle, T. (2008). Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, *12*(4), 497-528. doi:10.1007/s10707-007-0037-9

Andrienko, G., & Andrienko, N. (2007). Coordinated multiple views: a critical view. *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization* (pp. 72-74). Zürich, Switzerland: IEEE. doi:10.1109/CMV.2007.4

Andrienko, G., Andrienko, N., & Heurich, M. (2011). An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, *25*(9), 1347-1370. doi:10.1080/13658816.2011.556120

Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46. doi:10.1145/1345448.1345455

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I., & Wachowicz, M. (2008). Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization*, *7*(3-4), 173-180. doi:10.1057/ivs.2008.23

Andrienko, G., Andrienko, N., Fischer, R., Mues, V., & Schuck, A. (2006). Reactions to geovisualization: an experience from a European project. *International Journal of Geographical Information Science*, *20*(10), 1149-1171. doi:10.1080/13658810600816524

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S., & Wrobel, S. (2011). From movement tracks through events to places : Extracting and characterizing significant places from mobility data. *Proceedings of the 2011 IEEE Symposium on Visual Analytics Science and Technology* (pp. 159-168). Providence, USA: IEEE.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D. A., Kraak, M.-J., MacEachren, Alan M., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839-857. doi:10.1080/13658810701349011

Andrienko, N., & Andrienko, G. (2007). Designing visual analytics methods for massive collections of movement data. *Cartographica*, *42*(2), 117-138. doi:10.3138/carto.42.2.117

Andrienko, N., & Andrienko, G. (2008). Supporting visual exploration of massive movement data. *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '08)* (pp. 474-475). Napoli, Italy: ACM Press. doi:10.1145/1385569.1385663

Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic concepts of movement data. In F. Giannotti & D. Pedreschi (Eds.), *Mobility Data Mining and Privacy* (pp. 15-38). Springer. doi:10.1007/978-3-540-75177-9

Bartumeus, F., & Catalan, J. (2009). Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, *42*(43), 1-12. doi:10.1088/1751-8113/42/43/434002

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., & Gerlotto, F. (2007). Scale-invariant movements of fishermen: the same foraging strategy as natural predators. *Ecological Applications*, *17*(2), 331-337.

Bomberger, N., Rhodes, B., Seibert, M., & Waxman, A. (2006). Associative learning of vessel motion patterns for maritime situation awareness. *Proceedings of the 9th International Conference on Information Fusion* (pp. 1-8). Florence, Italy: IEEE. doi:10.1109/ICIF.2006.301661

Chang, S.-K., Liu, K.-Y., & Song, Y.-H. (2010). Distant water fisheries development and vessel monitoring system implementation in Taiwan—History and driving forces. *Marine Policy*, *34*(3), 541-548. Elsevier. doi:10.1016/j.marpol.2009.11.001

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129-136.

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. *Proceedings of the 27th International Conference on Data Engineering* (Vol. 4, pp. 900-911). Hannover, Germany: IEEE. doi:10.1109/ICDE.2011.5767890

Choi, K.-H., Kim, J.-S., Kim, Y. S., Yoo, M.-A., & Chon, T.-S. (2006). Pattern detection of movement behaviors in genotype variation of *Drosophila melanogaster* by using self-organizing map. *Ecological Informatics*, *1*(3), 219-228. doi:10.1016/j.ecoinf.2005.12.002

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems*, *13*(3), 319-340.

Demšar, U., & Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, *24*(10), 1527-1542. doi:10.1080/13658816.2010.511223

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., & Die, D. (2005). Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(3), 611-622. doi:10.1139/f04-219

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, *7*(3-4), 240-252. doi:10.1057/palgrave.ivs.9500182

Dykes, J. A., & Mountain, D. A. (2003). Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, *43*(4), 581-603. doi:10.1016/S0167-9473(02)00294-3

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, *63*(7), 1057-1066. doi:10.1007/s00265-009-0739-0

Enguehard, R. A., Devillers, R., & Hoeber, O. (2011). An evaluation of fractal/velocity pattern extraction. *Proceedings of the First International Maritime Anomaly Detection (MAD) Workshop* (pp. 1-2). Tilburg, Netherlands.

Enguehard, R. A., Devillers, R., & Hoeber, O. (2011). Geovisualization of fishing vessel movement patterns using hybrid fractal / velocity signatures. *Proceedings of the First International GeoViz Workshop* (pp. 1-2). Hamburg, Germany.

Enguehard, R. A., Devillers, R., & Hoeber, O. (submitted). Comparing interactive and automated methods for analyzing large vessel movement data set. *Ocean & Coastal Management,* 26p.

Enguehard, R. A., Hoeber, O., & Devillers, R. (submitted). Interactive exploration of movement data : A case study of geovisual analytics for fishing vessel analysis. *Information Visualization*, 30p.

Focardi, S., Marcellini, P., & Montanaro, P. (1996). Do ungulates exhibit a food density threshold ? A field study of optimal foraging and movement patterns. *The Journal of Animal Ecology*, *65*(5), 606-620. doi:10.2307/5740

Franke, A., Caelli, T., & Hudson, R. J. (2004). Analysis of movements and behavior of caribou (*Rangifer tarandus*) using hidden Markov models. *Ecological Modelling*, *173*(2-3), 259-270. doi:10.1016/j.ecolmodel.2003.06.004

Füchslin, R. M., Shen, Y., & Meier, P. F. (2001). An efficient algorithm to determine fractal dimensions of point sets. *Physics Letters A*, *285*(1-2), 69-75. doi:10.1016/S0375-9601(01)00339-5

Gottfried, B. (2011). Interpreting motion events of pairs of moving objects. *GeoInformatica*, *15*(2), 247-271. doi:10.1007/s10707-009-0095-2

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395-408. doi:10.1111/j.1461-0248.2009.01293.x

Hering, E. (1964). *Outlines of a theory of the light sense* (p. 317). Cambridge, USA: Harvard University Press.

Hintzen, N. T., Piet, G. J., & Brunel, T. (2010). Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data. *Fisheries Research*, *101*(1-2), 108-115. doi:10.1016/j.fishres.2009.09.014

Ho, Q., & Jern, M. (2008). Exploratory 3D geovisual analytics. *Proceedings of the 2008 International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies* (Vol. d, pp. 276-283). Ho Chi Minh City, Vietnam: IEEE. doi:10.1109/RIVF.2008.4586367

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*, *1*(1), 21-48. doi:10.5311/JOSIS.2010.1.3

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, *24*(1), 6–21.

Höferlin, M., Höferlin, B., Weiskopf, D., & Heidemann, G. (2011). Uncertainty-aware video visual analytics of tracked moving objects. *Journal of Spatial Information Science*, *2*(2), 87-117. doi:10.5311/JOSIS.2010.2.1

Jern, M., & Franzen, J. (2007). Integrating InfoVis and GeoVis components. *Proceedings of the 11th International Conference on Information Visualization (IV '07)*, (pp. 511-520). Zürich, Switzerland: IEEE. doi:10.1109/IV.2007.70

Jern, M., Åström, T., & Johansson, S. (2008). GeoAnalytics tools applied to large geospatial datasets. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 362-372). Columbus, USA: IEEE. doi:10.1109/IV.2008.27

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., & Shen, H. T. (2010). Discovery of convoys in trajectory databases. *Proceedings of the 34th International Conference on Very Large Data Bases* (pp. 1068-1080). Auckland, New Zealand: VLDB Endowment.

Johansson, S., & Jern, M. (2007). GeoAnalytics visual inquiry and filtering tools in parallel coordinates plots. *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems* (pp. 1-8). Seattle, USA: ACM Press. doi:10.1145/1341012.1341055

Keim, D. A., Panse, C., Sips, M., & North, S. C. (2004). Visual data mining in large geospatial point sets. *IEEE Computer Graphics and Applications*, *24*(5), 36–44. IEEE. doi:10.1109/MCG.2004.41

Kim, R., & Hogan, P. (2011). World Wind JAVA SDK. Retrieved from http://worldwind.arc.nasa.gov/java/

Koua, E. L., & Kraak, M.-J. (2004). A usability framework for the design and evaluation of an exploratory geovisualization environment. *Proceedings of the 8th International Conference on Information Visualisation* (pp. 153-158). Austin, USA: IEEE. doi:10.1109/IV.2004.1320138

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference* (pp. 1988–1995). Durban, South Africa: International Cartographic Association.

Kraak, M.-J., & van de Vlag, D. E. (2007). Understanding spatiotemporal patterns: visual ordering of space and time. *Cartographica*, *42*(2), 153-161. doi:10.3138/carto.42.2.153

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, *8*(1-6), 185-203. doi:10.1016/S0968-090X(00)00017-6

Langran, G. (1992). *Time in Geographical Information Systems* (p. 180). London, UK: CRC Press.

Laxhammar, R., Falkman, G., & Sviestins, E. (2009). Anomaly detection in sea traffic - a comparison of the Gaussian Mixture Model and the Kernel Density Estimator. *Proceedings of the 12th International Conference on Information Fusion* (pp. 756-763). Seattle, USA: IEEE.

Li, X., Han, J., & Kim, S. (2006). Motion-Alert : automatic anomaly detection in massive moving objects. *Proceedings of the 2006 IEEE Intelligence and Security Informatics Conference* (pp. 166-177). San Diego, USA: IEEE.

Lloyd, D., & Dykes, J. (2011). Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study. *IEEE Transactions* *on Visualization and Computer Graphics*, *17*(12), 2498-2507. doi:10.1109/TVCG.2011.209

Lundblad, P., Jern, M., & Forsell, C. (2008). Voyage analysis applied to geovisual analytics. *Proceedings of the 12th International Conference on Information Visualisation* (pp. 381-388). Columbus, USA: IEEE. doi:10.1109/IV.2008.32

Mandelbrot, B. B. (1967). How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science*, *156*(3775), 636-638. doi:10.1126/science.156.3775.636

Mano, J.-P., Georgé, J.-P., & Gleizes, M.-P. (2010). Adaptive multi-agent system for multi-sensor maritime surveillance. *Advances in Practical Applications of Agents and Multiagent Systems*, *70*, 285–290. doi:10.1007/978-3-642-12384-9_34

McConnell, S. M. (1996). *Rapid development: Taming wild software schedules* (1st ed., p. 680). Redmond, USA: Microsoft Press.

Mills, C. M., Townsend, S. E., Jennings, S., Eastwood, P. D., & Houghton, C. A. (2006). Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science*, *64*(2), 248-255. doi:10.1093/icesjms/fsl026

Molenaar, E. J., & Tsamenyi, M. (2000). Satellite-based vessel monitoring systems for fisheries management: international legal aspects. *The International Journal of Marine and Coastal Law*, *15*(1), 65-110. doi:10.1163/157180800020492643

Mullowney, D. R., & Dawe, E. G. (2009). Development of performance indices for the Newfoundland and Labrador snow crab (*Chionoecetes opilio*) fishery using data from a vessel monitoring system. *Fisheries Research*, *100*(3), 248-254. doi:10.1016/j.fishres.2009.08.006

Murawski, S. A., Wigley, S. E., Fogarty, M. J., Rago, P. J., & Mountain, D. G. (2005). Effort distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of Marine Science*, *62*(6), 1150-1167. doi:10.1016/j.icesjms.2005.04.005

Mårell, A., Ball, J. P., & Hofgaard, A. (2002). Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology*, *80*(5), 854-865. doi:10.1139/Z02-061

Nams, V. O. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia*, *143*(2), 179-88. doi:10.1007/s00442-004-1804-z

Pelot, R., & Wu, Y. (2007). Classification of recreational boat types based on trajectory patterns. *Pattern Recognition Letters*, *28*(15), 1987-1994. doi:10.1016/j.patrec.2007.05.014

Pfoser, D., Tryfona, N., & Jensen, C. S. (2005). Indeterminacy and spatiotemporal data: basic definitions and case study. *GeoInformatica*, *9*(3), 211-236. doi:10.1007/s10707-005-1282-4

Raymond, B., & Hosie, G. (2009). Network-based exploration and visualisation of ecological data. *Ecological Modelling*, *220*(5), 673-683. doi:10.1016/j.ecolmodel.2008.12.011

Ren, F., & Kwan, M.-P. (2007). Geovisualization of human hybrid activity-travel patterns. *Transactions in GIS*, *11*(5), 721-744. doi:10.1111/j.1467-9671.2007.01069.x

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Information Visualization*, *7*(3-4), 225-239. doi:10.1057/palgrave.ivs.9500183

Rocha, J. A. M. R., Times, V. C., Oliveira, G., Alvares, L. O., & Bogorny, V. (2010). DB-SMoT: A direction-based spatio-temporal clustering method. *Proceedings of the 5th IEEE International Conference Intelligent Systems* (pp. 114-119). London, UK: IEEE. doi:10.1109/IS.2010.5548396

Rodighiero, D. (2010). Guidelines to visualize vessels in a geographic information system. *Proceedings of the 14th International Conference on Information Visualisation* (pp. 455-459). Salt Lake City, USA: IEEE. doi:10.1109/IV.2010.70

Sage, B. (2005). Identification of "high risk vessels" in coastal waters. *Marine Policy*, *29*(4), 349-355. doi:10.1016/j.marpol.2004.05.008

Saitoh, S.-I., Mugo, R., Radiarta, I. N., Asaga, S., Takahashi, F., Hirawake, T., Ishikawa, Y., Awaji, T., In, T. & Shima, S. (2011). Some operational uses of satellite remote sensing and marine GIS for sustainable fisheries and aquaculture. *ICES Journal of Marine Science*, *68*(4), 687-695. doi:10.1093/icesjms/fsq190

Schaefer, K. M., & Fuller, D. W. (2006). Comparative performance of current-generation geolocating archival tags. *Marine Technology Society Journal*, *40*(1), 15-28. doi:10.4031/002533206787353673

Schwehr, K. D., & McGillivary, P. A. (2007). Marine ship automatic identification system (AIS) for enhanced coastal security capabilities: an oil spill tracking application. *Proceedings of the 2007 Oceans Conference* (pp. 1-9). Vancouver, Canada: IEEE. doi:10.1109/OCEANS.2007.4449285

Sevcik, C. (1998). A procedure to estimate the fractal dimension of waveforms. *Complexity International*, *5*(19), 1-20.

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools. *Proceedings of the 2006 AVI workshop on Beyond Time and Errors Novel Evaluation Methods for Information Visualization* (pp. 1-7). Venice, Italy: ACM Press. doi:10.1145/1168149.1168158

Theiler, J. (1990). Estimating fractal dimension. *Journal of the Optical Society of America A*, *7*(6), 1055-1073. doi:10.1364/JOSAA.7.001055

Thomas, J., & Cook, K. (2005). *Illuminating the path: research and development agenda for visual analytics* (p. 194). Los Alamitos, USA: IEEE Computer Society.

Tomaszewski, B. M., Robinson, A. C., Weaver, C., Stryker, M., & MacEachren, A.M. (2007). Geovisual analytics and crisis management. *Proceedings of the 4th International Information Systems for Crisis Response and Management (ISCRAM) Conference* (pp. 1-8). Delft, Netherlands.

Tremblay, Y., Shaffer, S. A., Fowler, S. L., Kuhn, C. E., McDonald, B. I., Weise, M. J., Bost, C.-A., et al. (2006). Interpolation of animal tracking data in a fluid environment. *The Journal of Experimental Biology*, *209*(1), 128-40. doi:10.1242/jeb.01970

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed., p. 200). Cheshire, CT, USA: Graphics Press.

Turdukulov, U. D., Kraak, M.-J., & Blok, C. A. (2007). Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds. *Computers & Graphics*, *31*(3), 370-379. doi:10.1016/j.cag.2007.01.028

Underwood, A. J., & Chapman, M. G. (1985). Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology*, *91*(1-2), 17–43. Elsevier. doi:10.1016/0022-0981(85)90218-7

Wang Baldonado, M. Q., Woodruff, A., & Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '00* (pp. 110-119). New York, USA: ACM. doi:10.1145/345513.345271

Ware, C. (2004). *Information visualization: perception for design* (2nd ed., p. 486). San Francisco, USA: Morgan Kaufmann.

Willems, N., van de Wetering, H., & van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics Forum*, *28*(3), 959-966. doi:10.1111/j.1467-8659.2009.01440.x

With, K. A. (1994). Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology*, *9*(1), 25-36. doi:10.1007/BF00135076

Witt, M. J., & Godley, B. J. (2007). A step towards seascape scale conservation: using vessel monitoring systems (VMS) to map fishing activity. *PloS ONE*, *2*(10), e1111, 1-5. doi:10.1371/journal.pone.0001111

Wood, J., Dykes, J. A., Slingsby, A., & Clarke, K. (2007). Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1176-83. doi:10.1109/TVCG.2007.70570

Zhao, J., Forer, P., & Harvey, A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, *7*(3-4), 198-209. doi:10.1057/palgrave.ivs.9500184

Appendix A - ICEHR application for ethics review

MEMORIAL UNIVERSITY OF NEWFOUNDLAND

INTERDISCIPLINARY COMMITTEE ON ETHICS IN HUMAN RESEARCH (ICEHR) APPLICATION FOR ETHICS REVIEW

Application Guidelines

Researchers must submit **5 paper copies** (1 original and 4 copies) plus an electronic copy of the proposal, including all attachments, at least 4 weeks before ethics approval is required, to:

ICEHR Co-ordinator
Office of Research
INCO Innovation Centre, Rm. 2015
230 Elizabeth Avenue
Memorial University of Newfoundland
St. John's, NL A1C 5S7

Submit electronic copy to icehr@mun.ca.

Please refer to our web page at **http://www.mun.ca/research/ethics_committee.php** for information on preparing your application.

Each application should contain:

1. **Completed Application Form** (copy attached). This document provides basic information, such as name of researcher, title of project. It consists of two pages.

2. **Summary of the Research** - should include sufficient information about the proposed research so that the ethical issues can be considered in appropriate context. But not too long?

3. **Statement of Ethical Issues** and how the researcher will deal with these issues (*e.g.* harms and benefits, free and informed consent, privacy and confidentiality).

4. **Copies of any relevant forms or documents to be used in the research,** such as recruiting materials, consent forms, information sheets for participants, questionnaires. Other documents may be included if the researcher judges that they would be helpful.

Interdisciplinary Committee on Ethics in Human Research (ICEHR)
Application for Ethics Review

Section A: *Principal Researcher*

Last Name: Wilson	First Name: Garnett	Title (Dr./Mr./Ms./etc.): Dr.
Department/Faculty/School: Department of Computer Science		
Mailing Address for correspondence, if different from department: none	Email: gwilson@mun.ca Telephone: 709-737-4891	
Position: ☐ Faculty ☐ Graduate Student ☐ Master's ☐ Ph.D. ☐ Undergraduate Student ☑ Other (specify Postdoctoral Fellow)		
NOTE: (***Students** must complete Section C and have their supervisor complete Section D*)		

Co-applicant(s):

Name	Position	Department	Email
René Enguehard	Master's Student	Geography	u25rae@mun.ca
Orland Hoeber	Assistant Professor	Computer Science	hoeber@mun.ca
Rodolphe Devillers	Associate Professor	Geography	rdeville@mun.ca

Section B: *Research Project*

Title of Project: Field Trials with Geo-Visual Analytics Prototype Software	
Start Date: Jan. 15, 2011	Anticipated Completion Date: Aug. 31, 2011
Total cost of Project: $500	Sponsor/Agency: NSERC
Will funds be administered through Memorial? ☑ Yes ☐ No	
Will project receive scholarly/academic review? ☑ Yes ☐ No If Yes, by whom (granting agency, department, etc.)? NSERC Note: ICEHR may request scholarly review of projects that will not otherwise be reviewed	
Has this proposal been submitted to any other REB? ☐ Yes ☑ No If yes, specify which one(s): ______ Approval (date approved, pending): ______	

Section C: *Signature*

Principal Researcher:

I, Garnett Wilson the Principal Researcher, have read Memorial University's Policy on Ethics of Research Involving Human Participants and agree to conduct the research subject to the policy statements in Section 3 and accept the responsibilities as outlined in Section 18.

Signature of Principal Researcher: Garnett Wilson **Date** Dec. 17, 2010

Access to Information and Protection of Privacy

The information on this form is collected under the authority of the Memorial University Act (RSNL 1990 Chapter M-7) and is needed for and will be used by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) to assess your application for ethics review and by the ICEHR Coordinator to administer ethics clearance. If you have any questions about the collection and use of this information contact the ICEHR Coordinator, Office of Research, at 709-737-8368.

Section D: Supervisor

Last Name: Devillers	First Name: Rodolphe	Title (Dr./Mr./Ms./etc.): Dr
Department/Faculty/School: Geography		
Email: rdeville@mun.ca	Telephone: 864-8412	

If this project is part of the work for a thesis or honours essay, your signature below will indicate departmental/faculty/school approval that:

the above-named research has been approved by the Department, Faculty, or School of Geography / Computer Sciences

I, Rodolphe Devillers, the ~~Supervisor~~ co-supervisor, have read the Memorial University guidelines regarding research involving human subjects and agree to conduct the research subject to the policy statements in Section 3 and accept the responsibilities as outlined in Section 18.

Signature of Supervisor:

Date: December 20, 2010

NOTE: It is the responsibility of researchers to read the ICEHR "Information for Researchers" and the Tri-Council Policy Statement "Ethical Conduct for Research Involving Humans"

Appendix B - Summary of ethical concerns

Field Trials With Geo-Visual Analytics Prototype Software

Summary of Research
In the course of our research on geo-visual analytics, we have developed two research prototype software systems with the purpose of assisting analysts with their tasks of discovering interesting patterns within geo-spatial data. Our focus has been in the domain of fisheries-related data. The two research prototypes are:

- GTdiff, which allows analysts to see how fisheries-related data are changing over space and time.
- VMS Explorer, which allows analysts to detect anomalous fishing vessel movement patterns.

The primary objective of this study is to gain insight into how the prototype systems that have been developed can be used in real-world problem solving and data exploration activities. We also wish to gain a deeper understanding of the types of problem solving and decision making activities that are being undertaken by fisheries data analysts. As such, the study will be conducted as a set of field trials.
Within a field trial, software is evaluated in as realistic a setting as possible, using real data sets and real end-users as participants. For this study, data has been provided by Fisheries and Oceans Canada (DFO), and participants will be recruited from within this organization as well. In this type of evaluation, the investigator and the participant work together with the software system under investigation in an attempt to solve a real-world data analysis or decision-making task. By being directly involved in the use of the software, the investigator allows the participant to work effectively without prior training.
For the field trials of GTdiff, data has been provided by the Oceans Habitat & Species at Risk department within DFO. Since the data is confidential, the software will be installed on a DFO computer and the field trials will be conducted within their work setting (at the DFO office in St. John's). Participants for this field trial will be recruited from this department since they are the ones who are familiar with the types of data analysis needed for fisheries data that has significant spatial and temporal aspects.
For the field trials of VMS Explorer, data has been provided by the Conservation & Protection department within DFO. The data set provided encompasses a one-year timespan of fishing vessel movement in Atlantic Canada, with all identifying information for the vessels removed. Since all confidential information has been removed, the software will be installed on a computer that we provide, and the field trial will be conducted at a location that is convenient for the participants (either the DFO office in St. John's, the DFO office in Halifax, or at Memorial University). Participants for this field trial will be recruited from this department since they are the ones who are familiar with the types of data analysis needed for detecting anomalous vessel movement patterns. Note that the costs indicated for this study on the ICEHR Application for Ethics Review are to cover possible travel costs for this field trial.

The participant recruitment will be done with the assistance of managers within DFO. The managers are aware of who among their employees are knowledgeable within the specific task domains, open to new ways of performing data analysis, and are willing to provide honest and constructive feedback. Although there will be a power relationship between the managers and the participants, once the participants have been selected all communication will be directly with them. Further, we will ensure the participants that they have the right to withdraw from participation at any time, and that no raw data collected in the course of the field trials will be shared with DFO or their managers. We are seeking 3-4 participants for each of the two software prototypes under investigation in this study. While it may be possible for the manager to determine who provided what information, even when the results are reported in anonymous and summary format, we will take all reasonable action to protect the confidentiality and anonymity of their participation.

Study Procedures

The primary steps in the field trial procedure are as follows:

1. Present the consent forms to the participant, and obtain their free and informed consent to participate in this research project.
2. Discuss the specific data analysis tasks the systems are designed to support with the participant.
3. Instruct the participant to perform their data analysis tasks with the system. Note that within field trials, the investigator assists the participant in performing the task, showing them features of the software that can be of assistance, and explaining to them how the software works.
4. Administer a post-study questionnaire with the purpose of measuring the participant's subjective reactions regarding the use of the software.
5. Conduct a brief interview to gain the participant's input on specific elements of the software and the support it provided for their data analysis activity.
6. Thank the participant for their time and comments.

It is expected that the entire study will take approximately 60 minutes.

Statement of Ethical Issues

Scholarly Review

Scholarly review of this research has occurred through the NSERC review of a Strategic Projects Grant (STPGP 365189-08) for which Dr. Hoeber is the principal investigator.

Harms & Benefits

Harms

There are no extraordinary risks or harms associated with this research project. Participants will be required to conduct knowledge discovery and exploratory tasks using the software.

If at any time the participant feels uncomfortable with the study, they may discontinue their participation without penalty. Further, no raw data or details of the participants' performance will be shared with their managers or directly with DFO.

Benefits

The primary benefit that participants may gain in this study is the discovery of interesting patterns or aspects of the data that they may not have been previously aware.

Participant Recruitment

This research is being conducted in partnership with DFO. Our contacts within the Oceans Habitat & Species at Risk, and Conservation & Protection units will assist in the recruitment of participants. The recruitment will be purposeful, in which employees of DFO that conduct data analysis within these units will be selected.

Free and Informed Consent

Competence

There are no issues with the competence of the participants to provide free and informed consent in this study.

Parental or Third-Party Consent

Not applicable.

Free Consent

There will be no coercion or special inducement used to recruit participants. Although the recruitment will be performed by managers within DFO, participants who wish to withdraw their participation prior to completing the study will be able to do so, without any ramifications.

Classroom Administration of Questionnaires

Not applicable.

Informed Consent

Prior to beginning the study, a consent form (see attached) will be provided for the participant to read and sign. The consent form will outline the activities involved in the study, the methods employed for protecting their anonymity and the confidentiality of the data collected, and will assure the participants that they can exit the study at any time.

Deception

Not applicable.

The Process of Obtaining Consent

Once the group of participants are identified, they will be provided with a copy of the consent form in advance via email in order to provide them with advance knowledge of the study. At the beginning of the study, two printed copies of the consent form will be provided and reviewed in detail with the participant. These will be signed by both the participant and the investigator. One copy will be given to the participant and the other kept for our files.

Privacy and Confidentiality

Anonymity
Knowledge of the participants' identities will be known since email will be used to coordinate participation dates and times. However, individual participants will not be required to write their names or any other identifying information on the research questionnaires or data collection forms. All information that links an individual's identity to their participation ID will be kept separate.
Although video recording will be used while the participant is using the software, the video camera will be aimed at the computer screen, keyboard, and mouse. The discussions between the researcher and the participant will be included in this recording, and later transcribed. Any details that identify the participant will be replaced with pseudonyms in the transcripts.
During the interview portion of the study, audio recordings will be made. As with the video recording, pseudonyms will be used in the transcripts to protect the anonymity of the participants.
A final step we will take to protect the anonymity of the participants is to not release any of the raw data collected in the course of this study to DFO. Any reporting of the outcomes of this research will exclude identifying information of the participants.

Confidentiality
All research materials will be held confidential by the principal investigators and kept in a secure on-campus locations. Data will be kept for a minimum of five years. When we decide to dispose of the data, all printed data collection forms will be shredded, and all digital media will be destroyed in accordance with University policy.
Within the consent form, we will ensure the participants that the data itself will only be used by researchers within our group, and will not be shared in raw format with anyone including our partners (their employer).

Conflict of Interest
Not applicable.

Inclusiveness
Not applicable.

Aboriginal Peoples
Not applicable.

Appendix C - ICEHR approval letter

MEMORIAL
UNIVERSITY

Interdisciplinary Committee on Ethics in Human Research (ICEHR)

Office of Research
St. John's, NL Canada A1C 5S7
~~Tel: [illegible]~~
www.mun.ca

January 17, 2011

<u>ICEHR No. 2010/11-070-SC</u>

Dr. Garnett Wilson
Department of Computer Science
Memorial University of Newfoundland

Dear Dr. Wilson:

Thank you for your submission to the Interdisciplinary Committee on Ethics in Human Research (ICEHR) entitled *"Field Trials with Geo-Visual Analytics Prototype Software"*.

The Committee has reviewed the proposal and appreciates the care and diligence with which you have prepared your application. We agree that the proposed project is consistent with the guidelines of the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (*TCPS*). *Full ethics clearance* is granted for <u>one year</u> from the date of this letter.

If you intend to make changes during the course of the project which may give rise to ethical concerns, please forward a description of these changes to Mrs. Brenda Lye at blye@mun.ca for the Committee's consideration.

The *TCPS* requires that you submit an annual status report on your project to ICEHR, should the research carry on beyond *January 2012*. Also, to comply with the *TCPS*, please notify us upon completion of your project.

We wish you success with your research.

Yours sincerely,

Lawrence F. Felt

Lawrence F. Felt, Ph.D.
Chair, Interdisciplinary Committee on
Ethics in Human Research

LF/en
copy: Co-Supervisors: Dr. Rodolphe Devillers, Department of Geography
Dr. Orland Hoeber, Department of Computer Science

Telephone: (709) 864 2561 / 864 2861 Fax: (709) 864 4612

Appendix D - Field trial consent form (both field trials)

Project Title:	**Field Trials with Geo-Visual Analytics Prototype Software**
Researchers:	**Dr. Orland Hoeber** **Dr. Rodolphe Devillers** **Dr. Garnett Wilson** **René Enguehard**

You are invited to participate in a research project that seeks to explore the utility of two research prototype software systems developed to assist data analysis tasks. This form is part of the process of providing you with information on the study and for the researchers to gain your informed consent to participate in the study. It will give you an overview of what the research is about and what your participation will involve. If you would like more details about something mentioned below, or further information not included here, you please feel free to ask. Take the time to read this document carefully and to understand any other information given to you by the researcher.
It is entirely up to you to decide whether to take part in this research. If you choose not to participate or if you decide to withdraw from the research once it has started, there will be no negative consequences for you, now or in the future.

Introduction:
In the course of our research on geo-visual analytics, we have developed two research prototypes with the purpose of assisting analysts with their tasks of discovering interesting patterns within geo-spatial data. Our focus has been in the domain of fisheries-related data. The two research prototypes are:
GTdiff: allowing analysts to see how fisheries-related data are changing over space and time
VMS Explorer: allowing analysts to detect anomalous vessel movement patterns
You have been selected to participate in this field trial due to your interest in analyzing fisheries data that has a significant spatial and temporal aspect (to be explored using GTdiff), or because of your interest in fishing vessel movement patterns (to be analyzed using VMS Explorer).

Purpose of study:
The primary objective of this study is for the researchers to gain insight into how the prototype systems that have been developed can be used in real-world problem solving and data exploration activities. We also wish to gain a deeper understanding of the types of problem solving and decision making activities that are being undertaken by the participants.

What you will do in this study:
In this study, you will be asked to use either GTdiff or VMS Explorer to analyze a set of data, seeking interesting patterns within the data. If you are using GTdiff, the data has been provided by the Oceans Habitat & Species at Risk unit at DFO and is related to the commercial fisheries in the Laurentian Channel over an 11-year period. If you are using VMS Explorer, the data has been provided by the Conservation & Protection unit at DFO and represents all of the fishing vessel movement in Atlantic Canada in 2009. After using the software to explore the data, you will be asked to complete a questionnaire. A short interview will also be conducted in which the researcher will ask your opinion on various aspects of the software and the types of data analysis you normally perform.
Your use of the software will be video recorded so that we can analyze your activities at a later date, and so that the researcher can focus on helping you to perform your data analysis tasks. The interview will be audio-recorded to ensure that we accurately capture your comments and discussion with the researcher.

Length of time:
The field trial is expected to take a total of 60 minutes.

Possible Benefits:
The primary benefit that you may find when participating in this study is the discovery of interesting patterns or aspects of the data that you had not previously been aware of. Further, your participation will provide the researchers with valuable information regarding how you are able to perform data analysis tasks using the prototype software. This will assist us in the further development of geo-visual analytics software.

Possible risks:
There are no extraordinary risks or harms associated with this study.

Confidentiality:
In order to maintain the privacy of your participation in this study, the data collected will be held strictly confidential by the researchers. Any identifying information will be kept separate from the details of your participation in the study. The data itself will only be used by researchers within our group, and will **not** be shared in raw format with anyone including our partners or your employer.

Anonymity:
Knowledge of your identity is not required. You will not be required to write your name or any identifying information on the research questionnaires. The data collected will only be used by researchers within our group. Any reporting of the outcomes of this research will exclude identifying information of the participants.

Recording of Data:
Your use of the software prototype will be video recorded. However, the focus of the video recording will be on what you are doing with the software system. As such, the video camera will be pointed at the computer screen, keyboard, and mouse. The audio

portion of the recording will capture the discussions between yourself and the researcher. This video recording will be captured and stored in electronic format only.
Data from the questionnaire will be collected on paper, and will subsequently be entered into an electronic format.
The interviews conducted after using the software will be audio-recorded, and will be stored in electronic format only.

Reporting of Results:
Results from this study will be published and shared with our partners. However, this will only occur after the data has been anonymized. While the raw video and audio recordings will not be included in these reports, direct quotations and images from the video recording may be used. In these cases, we will ensure that any identifying information is removed.

Storage of Data:
All research materials will be held confidential by the principal investigators. Physical material will be kept in a secure on-campus location; electronic material will be stored on password-protected computer systems. Data will be kept for a minimum of five years. When we decide to dispose of the data, all physical material will be shredded, and all digital media will be destroyed in accordance with University policy.

Questions:
You are welcome to ask questions at any time during your participation in this research. If you would like more information about this study, please contact:
Dr. Orland Hoeber
Department of Computer Science
Memorial University
hoeber@cs.mun.ca
709-737-3222
The proposal for this research has been approved by the Interdisciplinary Committee on Ethics in Human Research and found to be in compliance with Memorial University's ethics policy. If you have ethical concerns about the research (such as the way you have been treated or your rights as a participant), you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at (709) 737-2861

Consent:
Your signature on this form means that:

- You have read the information about the research.
- You have been able to ask questions about this study.
- You are satisfied with the answers to all of your questions.
- You understand what the study is about and what you will be doing.
- You understand that you are free to withdraw from the study at any time, without having to give a reason, and that doing so will not affect you now or in the future.

If you sign this form, you do not give up your legal rights, and do not release the researchers from their professional responsibilities.
The researcher will give you a copy of this form for your records.

Your Signature:
I have read and understood the description provided; I have had an opportunity to ask questions and my questions have been answered. I consent to participate in the research project, understanding that I may withdraw my consent at any time. A copy of this Consent Form has been given to me for my records.

______________________________ ________________________

Signature of Participant Date

Researcher's Signature:
I have explained this study to the best of my ability. I invited questions and gave answers. I believe that the participant fully understands what is involved in being in the study, any potential risks of the study and that he or she has freely chosen to be in the study.

______________________________ ________________________

Signature of Investigator Date

Appendix E - Participant questionnaire (both field trials)

Participant:_____

Please answer the following questions with regards to your background.

1. For how many years have you been involved in fisheries enforcement?

2. Have you ever worked with spatio-temporal data using other geovisualization systems?

   Yes No

3. If yes, how many such projects have you worked on?

   1 2 3 4 5 N/A

4. What is your level of understanding of geovisualization?

   Low Medium High

5. How familiar are you with virtual globes (Google Earth, NASA Worldwind, etc...)?

   Not at all Very familiar

   1 2 3 4 5

6. How familiar are you with the movement patterns associated with the North Atlantic fisheries?

   Not at all Very familiar

   1 2 3 4 5

7. How familiar are you with the movement patterns associated with vessel infractions?

   Not at all Very familiar

   1 2 3 4 5

Appendix F - Post-study questionnaire (first field trial)

Participant:___

The following questions relate to your experience using the prototype system for exploring movement data. Your answers to the following questions allow for a more accurate analysis of the data collected during this study.

INSTRUCTIONS: Please rate how strongly you agree or disagree with the following statements by circling the appropriate number.

Question	N/A	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
Using the prototype system enabled me to accomplish my usual tasks more quickly.	0	1	2	3	4	5
Using the prototype system improved my performance in exploring data.	0	1	2	3	4	5
Using the prototype system increased my productivity.	0	1	2	3	4	5
Using the prototype system enhanced my effectiveness at exploring data-sets.	0	1	2	3	4	5
Using the prototype system made it easier to explore data-sets.	0	1	2	3	4	5
I found the prototype system useful.	0	1	2	3	4	5

Question	N/A	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
Learning to operate the prototype system was easy for me.	0	1	2	3	4	5
I found it easy to get the prototype system to do what I wanted it to do.	0	1	2	3	4	5
My interaction with the prototype system was clear and understandable.	0	1	2	3	4	5
I found the prototype system to be flexible to interact with.	0	1	2	3	4	5
It was easy for me to become skillful at using the prototype system.	0	1	2	3	4	5
I found the prototype system easy to use.	0	1	2	3	4	5

Question	N/A	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I found the ability to access data point information by hovering over their arrows to be useful.	0	1	2	3	4	5
I found the histogram representation of vessel speeds to be useful.	0	1	2	3	4	5
I found the automatic rescaling of the histogram to be useful.	0	1	2	3	4	5
I found that the ability to filter by fractal dimension was useful.	0	1	2	3	4	5
I found that the ability to filter by vessel velocity was useful.	0	1	2	3	4	5
I found that the ability to filter by temporal range was useful.	0	1	2	3	4	5

Put an **x** below the method which you find most effective for the following:

	Prototype system	**Other systems**
Understanding the spatial distribution of vessels.		
Understanding the temporal distribution of vessels.		
Understanding where vessels fish.		
Understanding the distribution of vessel velocities.		
Identifying potential infractions.		
Overall, I prefer to use:		

If you have any further comments or suggestions, you may write them down below:

____

____

____

____

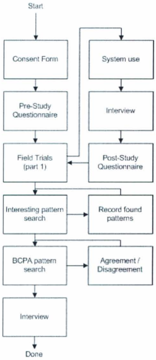
____

____

Thank you for your participation!

Appendix G - Procedure for second field trial

1. Get user to read and sign the Consent Form (ICEHR approved one, on letterhead).
2. Get the user to fill out the Pre-Study Questionnaire (same as first round of field trials).
3. Proceed to first round of field trials.
4. Get user to try and find interesting patterns, infractions, etc. within the cut-down data set (10 vessels), using the hybrid fractal/velocity filtering interface.
5. Record each found pattern's location and a brief explanation using Data Sheet A (appended).
6. Present the expert with the data segmented based on BCPA, using the simplified interface.
7. Record each found pattern's location and a brief explanation using Data Sheet A.
8. Proceed to short unstructured interview, touching on the all points listed in List A (appended).
9. Thank them for their time.



Appendix H - Change point recording sheet (second field trial)

Participant #:_____

Vessel Number	Start Timestamp	End Timestamp	Notes

Comments:

Appendix I - General interview questions

1. What did you think of the interactive filtering technique for finding interesting patterns?
2. What did you think of the automated segmentation technique for finding interesting patterns?
3. Do you think that the segments that were identified by the automated technique were meaningful?
   3.1. Were the segments you identified found by the automated technique?
   3.2. Did the automated technique find segments that you had missed?
4. Would you rather use one technique over the other, in terms of ease of use?
5. Would either visualization technique help you during your day-to-day operations?
6. Do you think that combining any of these methods would be beneficial?
   6.1. What do you think about combining the interactive technique and VUE?
   6.2. What do you think about combining the automated technique and VUE?
   6.3. What do you think about combining the interactive and automated techniques?
   6.4. What do you think about combining all three systems?
7. Do you prefer one technique over the other, and if so, why?