A Study of Geovisual Analytics for Exploring Event Anomalies Over Multiple Geospatial Datasets

by

©Md Monjur Ul Hasan

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

M.Sc. in Computer Science

Department of Computer Science

Memorial University of Newfoundland

April, 2015

St. John's Newfoundland

Abstract

Events performed by moving entities (e.g., shoppers purchasing products from stores, tourists visiting historical places, and fishing vessels trawling) are often described by geospatial data sets. When such events are independently collected in multiple data sets, comparing the same events for positional discrepancies with their spatial and temporal contexts may reveal important insights, such as data entry, instrumental, intentional, and/or processing errors. In this work, two independently collected data sets are considered: geospatial point data describing event locations and movement data describing movement activities of the entities that performed these events. For analyzing the anomalies within these data sets a geovisual analytics approach is presented, which extracts events, identifies event anomalies, represents these anomalies on a map, and allows analysts to perform exploratory analysis to make sense of the data. This approach makes extensive use of spatial and temporal filtering, geovisualization, colour encoding, and anomaly threshold filtering. It is highly interactive, supporting analytical reasoning and knowledge discovery through visual exploration and analysis of the data sets. This approach has been implemented as a prototype system for analyzing event anomalies within two real world data sets related to fishing activities. Field trials were performed with five expert fisheries analysts to evaluate the system. Results from this study confirm the value of the approach and its potential for supporting geospatial anomaly analysis of commercial fishing events.

Acknowledgements

First of all, I would like to thank my supervisors sincerely, Dr. Orland Hoeber and Dr. Wolfgang Banzhaf, who demonstrated how research is done, provided insightful and encouraging advice, and most importantly constantly challenged me for improvement.

I would also like to extend my gratitude to the Department of Computer Science, Memorial University. Without their generous resources this thesis would not have been possible. My research is financially supported by my supervisor's NSERC Strategic Projects Grant (Geo-Visual Analytics of Capture Fisheries Statistical Data) and scholarships from the University's School of Graduate Studies.

I am also thankful to Enamul Hoque and René Enguehard, former students within our research group, for their active support during my research work. Thanks also go to the other graduate students in the Department of Computer Science who provided insightful comments and suggestions.

My dear friends in St. John's always made my life interesting and enjoyable while I am far away from my home. I am especially grateful to all of them.

I owe my deepest gratitude to my loving parents, Md. Sarfaraj Sheik and Marina Sultana, who raised me with unconditional affection and patience. They are my friends, strongest supporters, and the reason I am me. Thanks also go to my wife, Tanzira Hosen, who supported me for the decision to study in Canada.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Recent developments on low cost GPS technologies and their miniaturization have enabled the tracking of various moving entities [4], such as vehicle, animal, and human being. These tracking systems allow the independent capture of events performed by the entities from different perspectives. Thus, multiple spatio-temporal data sets are often found describing the same events. Examples include fishing events performed by commercial fishing vessels described by the vessels' log book data and the fishing locations recorded in their catch reports, and shopping events performed by individuals described by movement data collected from their smart phones and the locations of the point-of-sale machines where they used their credit cards.

While the term *event* is widely used in many domains with their specific meanings, events in the context of spatio-temporal data sets are defined as spatial objects that are undergoing existential changes on their spatial, temporal, or thematic parameters, and explain specific activities of entities [10]. Though events in this context can take multiple forms of varying complexities, this research focuses on spatial events [4] performed by moving entities in certain geographical regions within given temporal extents. Two independently collected data sets are considered: movement data describing movement of entities over time and geographical point data describing where these entities performed notable actions.

At the most basic level, the movement data is composed of a series of data points where each data point consists of a latitude, longitude, timestamp, and entity identifier [7]. Ideally, events in the movement data show movement activities of entities in the area where these events occurred and between the time periods when the events were performed. For example, the voyages of a cargo ship show trajectories between different ports and daily fishing events performed by commercial fishing vessels represented by their movement paths (generated from their log book entries) show movement activities within the fishing regions and back and forth movements between these regions and the port.

Geographical point data [50] can consist of spatial points (latitude, longitude), temporal information, entity identifier, and domain specific information. Such data points can be used to represent events. Many events are performed in a precise geographical location within a very short period of time, such as point-of-sale transactions or road accidents, while many other events are performed over a wide geographical region within a longer period of time, such as football games or daily fishing events. Geographical point data describing the latter type of events, the positional information of the data points are the representative of the geographical regions and the temporal information are the representative of the time periods of events. When these data sets do not contain information about the spatial and temporal extents (the size and shape of the area where a given event occurred and the event duration) of these events, complete understanding of these events might not be possible from the data set since these extents might change over the events' spatial, temporal, and domain specific contexts. For example, a data point representing a fishing event of a particular day may consist of a spatial point within the fishing region and the fishing execution date. However, the size of the fishing region and the duration of the fishing event cannot be defined as they depend on the location of fishing (spatial context), the time of year when the event was executed (temporal context), the type of fishing (domain specific context), and the type of vessels used for that fishing (another domain specific context).

Since these data sets can describe the same events, they are expected to be describing the same event locations. Unexpected positional discrepancies such as differences in these event locations among these data sets are often present, which are referred to as event anomalies. Identifying and analyzing these event anomalies reveals important insights within and between the data, such as potential problems with the data collection and processing methods, instrumental errors, and intentional misreporting of information. However, challenges are associated with both identifying and analyzing event anomalies from data sets that are collected independently.

When both data sets are collected with temporal synchronization and have the same level of granularity (e.g., movement data collected at one-minute intervals and geospatial point data collected with a timestamp to one-minute accuracy), the identification of the event anomalies can be done by measuring the distance between the spatial points collected at same instant of time using computational methods. However, such conditions are not guaranteed when the data sets are collected independently, thus computational methods are difficult to tune for finding anomalies within the data sets, and often generate large number of false-positive results. While identifying obvious positional discrepancies (i.e., invalid locations for the given data sets) may be simple, discovering more subtle differences in the data sets that arise from various geospatial and domain specific contexts can be challenging. Such discovery processes are often guided by the analysts' experience and expertise in that domain. Hence, an effective mechanism is required for the analysts to identify the event anomalies from the data sets.

Once the event anomalies are identified, further analyses are required for understanding their behaviour and trends. These analyses involve detailed visual inspection and comparison of the data that constitute these event anomalies. These analyses also require incorporation of the knowledge about spatial and temporal contexts of these events. At the same time, these types of analyses are particularly challenging as opposed to data analysis in general because of the geospatial features of data sets [80]. Representing different types of information are important in this regard. However, owing to the size and complexity of the data sets this is challenging for analysts to visually decode all of the information [24] [44] [49]. Therefore, a human centred analysis system is highly desirable.

For example, Figure 1.1 shows two shopping events (marked as A and B) performed by a shopper during a particular one-hour period of a day. The pathway taken by the shopper during that hour is shown by a red line. The locations of the point of sales (POS) machines in which the shopper's credit card were charged during that hour is shown by red dots. While the distances between the movement path and the locations of the POS machines seems similar for both of the events, the shopping event B may require further investigation. The map shows a river between the POS location and the path taken by the shopper, and no path is possible that would allow the shopper to cross the river. Detailed analysis that incorporates the analysts' experience about the location and the type of merchant related to this event may reveal fraudulent use of the credit card, insufficient granularity of the movement data, or incorrect data regarding the location of the pos machine.

In this context, geovisual analytics can be a promising approach for identifying and

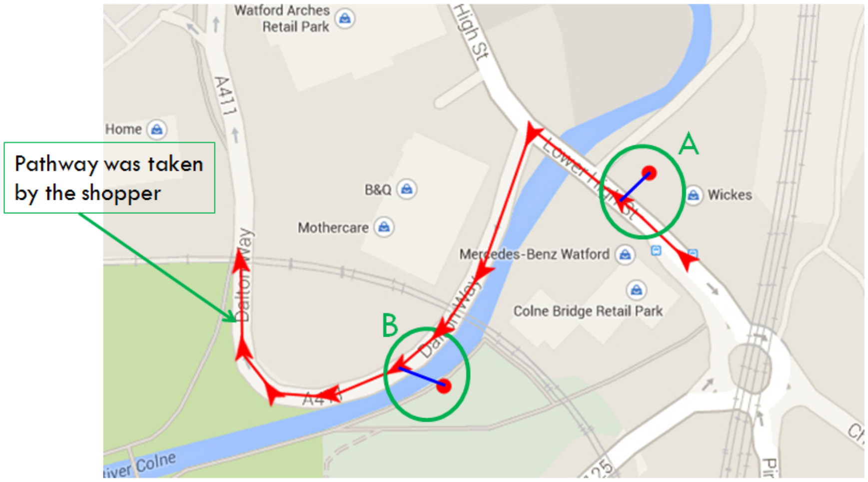


Figure 1.1: Two purchase events performed by a hypothetical shopper. The red line shows the trajectory extracted from the GPS devices of his/her car, arrowheads indicate the way points. The red dots show the locations of point of sales machines where this shopper used credit cards and the blue lines connect these locations with the nearest way points.

analyzing event anomalies related to the mismatching of event locations. Geovisual analytics is the science of analytical reasoning about spatio-temporal data sets, facilitated by interactive visual interfaces [31]. Geovisual analytics systems are designed to support highly-interactive exploration of spatio-temporal data sets by taking synergistic advantages of computer processing power and human information processing capabilities [35]. These systems have been shown to work very well for discovering and exploring unknown complex patterns, and displaying information in a manner that supports human interpretation, analytical reasoning, and decision making [12] [28]. While geovisual analytics have been studied in a wide range of domains in recent years, the notion of event analysis remains relatively less addressed. Most of the existing works and commercial GIS tools [26] [28] [47] [50] [57] [73] consider individual data elements as the unit of analysis and then compare the entire data sets together. In the context of this work events consist of a group of data points. While individual data points provide low level information about the entities (e.g., location of the entities in a particular instance of time, their direction, and speed), events provide a higher level information about them (i.e., the geographical context of the entire event and the area covered by the events). Thus, for the purpose of event anomaly analysis, considering events as the unit of analysis is more appropriate. The objective of this research is to study geovisual analytics system using events as the units of analysis, which supports analyzing the anomalies through visualization and interactive exploration of the data sets.

To alleviate the potential problems of anomaly analysis with events as units of analysis, it is important to introduce a new geovisual analytics system that would be able to process the various data sets and enable analysts to incorporate their expertise and experience into the knowledge discovery process. This new system should extract event anomalies from the data sets, present them in an interactive visualization, and allow human interactions for supporting the analytic reasoning about the data sets. The goal is to take the synergistic advantage of a computer's processing power and a human's decision-making processes to separate potential anomalies from those that are a result of non-synchronized temporal scale or errors in the data sets, and discover their trends and patterns.

1.2 Research Questions

The primary aim of this research is to develop and study a geovisual analytics approach for interactively exploring event anomalies. This approach leads to some fundamental research questions as follows:

Is it possible to design a geovisual analytics system for analyzing event anomalies?

Numerous geovisual analytics systems have been developed for analyzing spatio-temporal data sets with the incorporation of spatial, temporal, and analysts' domain specific knowledge [7] [12] [25]. In these systems, individual data elements of the data sets are used as the units of analysis. In contrast, this work requires the use of events as the units of analysis. Since the possibility of designing this type of system was not explored previously, this fundamental research question focuses on the possibility of designing such system. The anticipation is, it would be possible to design a system based on the theoretical model and common practices of geovisual analytics systems that support analytical tasks. Thus, the theoretical model and common practices of designing geovisual analytics systems are required to be studied.

Is the geovisual analytics system useful for the analysts to identify and analyze potential event anomalies?

The geovisual analytics system should visualize the event anomalies effectively, and allow interactive filtering and highlighting to support analysts' reasoning about the data sets. To fulfill this goal, the geovisual analytics system is required to overcome different challenges, such as providing meaningful representations of all the important elements of the data sets without visual clutter. To make the matter challenging, event anomalies are ill-defined. Thus, a human centred exploration process is required to make the system useful for the analysts. Therefore, the designed geovisual analytics system is required to be evaluated in order to understand the usefulness of this system for analyzing event anomalies.

Does the geovisual analytics system make it easy for the analysts to explore the data sets for extracting knowledge about event anomalies and entities' activities?

The system was designed by following the best practices and using the common user controls. Thus, the expectation is that the proposed system will be found easy for the analysts to learn and use, which will help them to extract and analyze potential event anomalies with less effort. While the event analysis activities are complex or exploratory in nature and no prior techniques are available, subjective measures such as ease of use can be helpful for validating the claim.

Does the system presented in this research enhance the analysts' abilities to make sense of the data and discover anomalies that are both known and previously unknown among the event anomalies within the data sets? The proposed system will extract potential event anomalies, represent them in a visualization, and allow analysts to explore them. The expectation is that the proposed visualizations and interaction techniques may effectively support the analysts' knowledge discovery process. This also may lead to the verification of known knowledge and discovery of new knowledge about the data sets. With appropriate evaluation such claims should be validated.

1.3 Approach Overview

The primary objectives of this thesis are to study geovisual analytics approaches for analyzing event anomalies, and to design, develop, and study a new system by which the analysts can interact with the data sets for exploring and understanding the event anomalies and their patterns and trends. To address the problem of event anomaly analysis, a computational method is proposed for automatically extracting events and anomalies from the data sets. Methods are also proposed for separating the meaningful anomalies from those that are the result of temporally non-synchronized data collection or are within a reasonable distance from one another. Such anomalies are then visualized on a geographical map. Interactive filtering and highlighting tools are provided to allow analysts to focus on particular events that are of interest. The overall process is divided into four steps as discussed in the rest of this section.

The event extraction is a computational procedure that identifies and divides the data points into groups that are related to individual events. This procedure is designed to address the difference in temporal granularity of the data sets. Once the events are extracted, all further procedures consider these events as the units of analysis.

Next, the event anomalies are identified. For this purpose, two parameters are used: amount of time the entities spend around the event locations, and distances between the event locations and the movement paths. Choosing appropriate values for these parameters cannot be done automatically, since they require not only knowledge of the data sets, but also domain-specific knowledge regarding the actual activities of the entities. Thus, analyst-adjustable threshold values are used for identifying the event anomalies.

Once the set of event anomalies are identified, they are visually represented on a map and hieratically listed group by the entity identifiers. The map representation provides spatial context of event anomalies, which plays important role for the analyses. The hierarchal view provides additional information about events and entities, such as: ancillary data associated with events and statistical data about entities. The coordinated interactions (linked brushing) between these visualizations allow analysts to explore and understand the event anomalies with different perspectives, such as: event anomalies from a certain region or performed by a group of entities.

A set of interactive tools are also provided to facilitate opportunities for the analysts to select and analyze specific features of the event anomalies or entities that are of particular interest. Filtering tools are provided to the system to filter the event anomalies based on their spatial, temporal, and ancillary data. These filters reduce the data in the visualization to a manageable size for the analysts and interactively update the visualizations allowing refinement of filter parameters. Tools are also provided for investigating the details of individual event anomalies by isolating and tracking vessels of interest and the correlations between surrounding event anomalies/vessels, which allows analysts to discover more important knowledge about the data sets, such as the region where a higher number of event anomalies are found, and underlying behaviour of the entities that have event anomalies.

Within the course of this research, a prototype system has been developed where all these components fit together. This prototype system confirms the possibility of designing a geovisual analytics system to this problem domain. A user evaluation was performed with real world data sets, domain specific tasks, and domain experts to understand the usefulness, complexity of using the system, and effectiveness of this approach [17] [62]. For this evaluation data related to commercial fishing events are used, where professional fisheries data analysts explored and analyzed the fishing event anomalies by using this prototype system. This evaluation validated, the usefulness of the system, the complexity of using the system, and it's capability of enhancing analysts' ability to explore the events anomalies among the data sets.

1.4 Organization of the Thesis

The remainder of the thesis is organized as follows. In Chapter 2, an overview of information visualization, visual analytics, geovisual analytics, visualization of multiple data sets, event visualization, visual analytics approaches of movement and point-sample data, and anomaly analysis are provided. The proposed geovisual analytics system for exploring event anomalies over multiple geospatial data sets is explained in

Chapter 3. The prototype system and a case study as an example of exploratory data analysis are presented in Chapter 4. The evaluation of the system using field trials is explained in Chapter 5, which includes the hypothesis that guided the field trials, the study design, and the analysis of the data collected from the field trials. The thesis concludes with a summary of the contributions, limitations, generalizability of this work, and an outline of future works in Chapter 6.

Chapter 2

Related Work

This chapter will review the works related to the geovisual analytics of geographical point data and movement data. For a better understanding of this domain and the different aspects of the problem, this review will begin with an overview of information visualization. After this, visual analytics will be reviewed, since visual analytics uses visualization as one of its components. Next, geovisual analytics will be reviewed, which is a subfield of visual analytics that focuses on the analysis of spatio-temporal data sets. Since movement data is a special kind of spatio-temporal data set having additional complexities, work related to movement data will be reviewed separately after reviewing geovisual analytics in general. The review will be concluded by exploring three works that take advantage of events. A discussion is provided at the end of this chapter explaining how these existing works motivated the work have presented in this thesis.

2.1 Information Visualization

Information visualization is the field of study that focuses on visually representing abstract, semi-structured, and/or hierarchical data within interactive visualizations, and allows the viewer to see the data without having to read details [81]. Information visualization systems exploit the enormous bandwidth and processing power of the human visual system to amplify the cognition of the data, and bridge the gap between the human and the machine [72]. In this field of study, theories and principles that help in understanding the works of the human mind when perceiving objects, interpreting data, and inferencing logical conclusions out of the visual representation of the presented information are studied [41].

Information visualization systems provide different options for understanding and discovering new insights about the data that are in question. The analysis process follows Shneiderman's information seeking mantra [66], which suggests to start with showing an overview of the entire data and then apply zooming and filtering on it. These operations reduce the data into a manageable size by extracting a subset of the data that are of interest. For detailed investigations, individual data elements can also be selected, which shows additional information and allows comparisons with other data elements. Thus, information visualization systems enhance the exploration and understanding of the relationships among the data elements.

In information visualization systems, data details are visually encoded using different graphical symbols, such as point, line, area, surface, volume, icons, and glyphs. Such symbols are called visual variables [81], which are further modified using position, size, shape, colour, orientation, and texture to establish a distinct correspondence of data elements with respect to some reference map. Different data elements are also discriminated from each other by these modifiers. As such, the data can be navigated through the visual variables without losing referential focus [65].

Visual variables are also designed to support pre-attentive processing [29] of the most interesting parts of the data, which means the cognitive operations are performed on these parts of the data prior to focusing on any particular region of the visualization. The information that is required to be processed pre-attentively are represented carefully to stand out from other information. The relative visual difference is the key factor in this regard. Thus, the visual variables are encoded with different form, colour, motion, or spatial position [81]. However, only a small number of items can be visually different. For example, it is easier to spot a single hawk in a sky full of pigeons than from a sky that contains a greater variety of birds. Therefore, two important considerations for designing visual variables for pre-attentive processing are: (a) the degree of difference between the representation of the targets and non-targets, and (b) the degree of differences of the representation of the non-targets [23] [63].

Effective information visualization systems also consider how humans perceive patterns and interpret information. In this direction, the Gestalt Laws [42] are widely used theories. The Gestalt Laws fundamentally deal with two types of concepts: perception of relationships, and perception of foreground from background. However, the Gestalt Laws that focus on how humans automatically infer the existence of a relationship between things when they are connected to one another [60], is of greater relevance and importance for this research. This law asserts that the simplest and most effective way of expressing relationships between two graphical objects is connecting them through a line. This design principle has been used and extended into many graphical forms and structures that are used to show relationships between connected objects (e.g., graphs, trees, node-link diagrams).

Another important theory used for developing effective information visualization systems is Opponent Process Theory of Colour [53], which explains that there are six elementary colours arranged perceptually as opponent pairs along three axes: black–white, red–green, and yellow–blue [30]. Humans can see the differences between colours on these channels much more easily than with arbitrary colours. For example, yellow objects are easier to see within a majority of blue objects than green objects. According to this theory, objects that are coloured using these six primary colours are automatically perceived as different, and can be carefully used to represent different kinds of information on the same screen. Thus, this theory is used for both the differentiating and the pre-attentive processing of information in information visualization system [81].

Supporting user interaction is another important aspect of information visualization. Many of today's information visualization systems integrate basic interaction features, such as zooming, filtering, and focusing [83]. These interactions allow users to manipulate the presented information based on their understanding and requirements. Therefore, these features improve the user's ability to process and investigate the data at a deeper level to make relevant decisions. These interactions are also used for reducing visual clutter. Therefore, the value of information visualization systems are widely determined by its user interaction techniques [22]. Information visualization systems without interaction are basically static images, which limit its usability by the number of supported tasks.

Although information visualization is a very useful technique to understand the important insights of the data, visualization of a large number of items on a computer screen and presenting multi-dimensional data on a two dimensional screen are the main challenges in this field. These challenges are divided into two groups [28]: the computational efficiency and the visual effectiveness. The computational efficiency concerns the time needed to process the data and render views, so that an information visualization system becomes computationally efficient and scalable to allow human interactions. The visual effectiveness concerns the usefulness of data visualization to reveal insights. Therefore, the effectiveness of an information visualization system depends on the efficiency of both the computation and the visual methods.

2.2 Visual Analytics

Visual analytics is the science of analytical reasoning supported by interactive visual interfaces [70]. Visual analytics extends analysts perceptual and cognitive abilities with the help of automatic data analysis techniques and interactive visualizations [39] [40]. Visual analytics also allows optimizing complicated resolving process of a complex problem by combining analysts' background knowledge, flexibility, and creativity with the enormous storage and processing capacities of today's computers. The goals of a visual analytics system are to derive insights from the data by detecting expected and unexpected information and to effectively communicate assessments for action. Such goals are often not reachable using a simple information visualization system. These limitations of information visualization systems are overcome in visual analytics systems by tightly integrated human intelligence with computational algorithms, visualization, and interaction techniques to tune the underlying analytical processes [39].

The mantra for the visual analytics process is to analyze first, show the important, zoom, filter, analyze further, and details on demand, which is an extension of Shneiderman's information seeking mantra of information visualization [66] proposed by Keim et al.[40]. A visual analytics process starts with the data. A choice for an initial representation and interactions of the data within a visual analytics system can be made after applying different computational techniques, such as data mining, statistical analysis, parsing, reduction, enrichment, and aggregation. The process then enters a loop that includes the analysts. The knowledge is discovered from the data within this loop by driving the system toward more focused analytical techniques. The user interactions of these systems allow analysts to understand the data and configure the system for different views according to their requirements. These steps help analysts to go beyond the visualization and ultimately confirm the hypotheses built from previous iterations [39]. Figure 2.1 illustrates how a visual analytics system generates knowledge from the data sets.

While visualizations of any visual analytics system show information to the analysts, user interactions are the key features that include the analysts into that analysis loop. The types of user interactions depend on the data and the analysis tasks. Filtering, selecting, and querying the data sets are the most common types of user interactions found in visual analytics systems [1] [4] [26] and are relevant to this research. During the analysis of the data sets, dynamic queries [2] are built at every stage with these user interactions.

For large and complex data sets, it is sometime useful to limit the range of data values that are visible and mapped to the display [81] and to allow analysts to explore them interactively. The filtering techniques extract a portion of the data the analysts wish to visualize. These filters involve placing bounds on the attributes of the data. For example, temporal data can be filtered based on the time period [46] and spatial data can be filtered based on the geographical boundary [8].

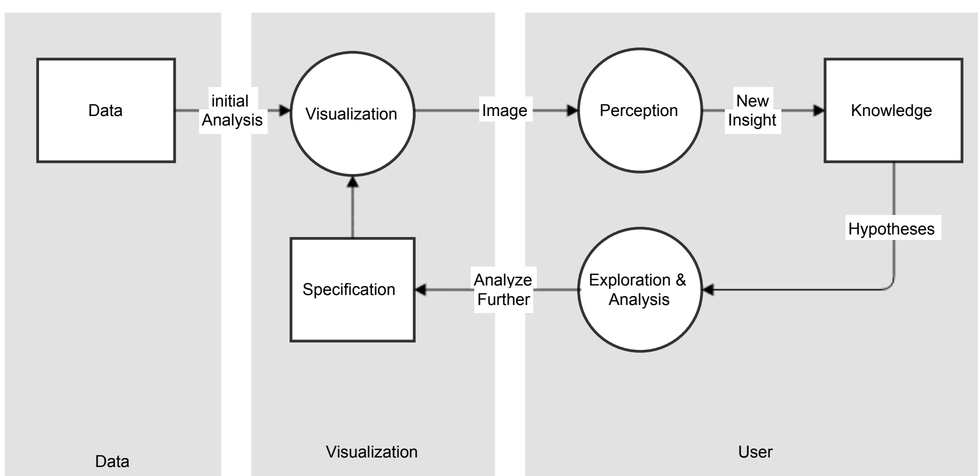


Figure 2.1: The knowledge discovery process loop of visual analytics systems proposed by Wijk [75].

Another interactive technique that is often used in visual analytics systems is called linked brushing [13]. In many visual analytics systems data objects or their different attributes are simultaneously appear in more than one display window, or different attributes can be distributed spatially within a single window. The linked brushing enables subsets of these data elements to be highlighted interactively. When the data elements appear in multiple visualizations, a group of elements selected through one visualization becomes highlighted in all the visualizations in which they appear. This feature enables the visual linking of components of heterogeneous complex data objects. For example, a visual link can be established between data elements represented in a scatter plot, a sorted list, and a 3D map by highlighting them simultaneously.

Hover queries [81] are another commonly used interaction technique for visual analytics systems. This is a selection action often performed by hovering the mouse on the data object. This action reveals extra information about the selected data sets.

Visual analytics systems are built upon the understanding of the human reasoning process, which maximizes our capacity to perceive, understand, and reason about the data. These systems provide the framework for analytical reasoning about the data, which allows synthesizing the information, deriving insights, detecting the expected, and discovering the unexpected [70]. These systems also enable analysts to understand scenarios and their trends from the historical and/or present perspective. Thus, visual analytics systems permit the creation of hypotheses and scenarios, and examine them in the light of available evidence. Analysts can apply human judgements to reach conclusions from a combination of assumptions and evidence [71].

Visual analytics systems also need to be easy to use for the analysts. The analysts should not be distracted from their task by overly technical or complex user interfaces. However, due to the complex and heterogeneous problem domains addressed by visual analytics, ease of use is difficult to validate [39]. Thus, a theoretically founded evaluation framework needs to be developed which allows assessing the contribution of any visual analytics system toward the level of effectiveness and efficiency [40] [70].

2.3 Geovisual Analytics

Geovisual analytics is a subclass of visual analytics designed to support explorative analysis of spatio-temporal data sets [35]. Geovisual analytics systems represent spatio-temporal data sets in graphical forms in order to facilitate the analyst's understanding of the underlying phenomena [64]. As a subclass of visual analytics, geovisual analytics is also a combined approach of visualizations, human factors, and data analysis that involves analysts in the problem solving process. The interactive visual displays of geovisual analytics systems are effective means for the analysts to reason about the data sets in many domains [5].

Geovisual analytics systems are designed for the analysts to assimilate two different sets of data attributes of spatio-temporal data sets: the geographic space and the data linked to a geographic location within that geographic space. Data elements linked with spatial locations are mapped to their geographic locations using visual variables [81] or retinal variables [14]. These variables are generally treated as graphic expressions, which contain information in the form of visual appearance [48], and allow analysts to understand the spatial context of the data elements.

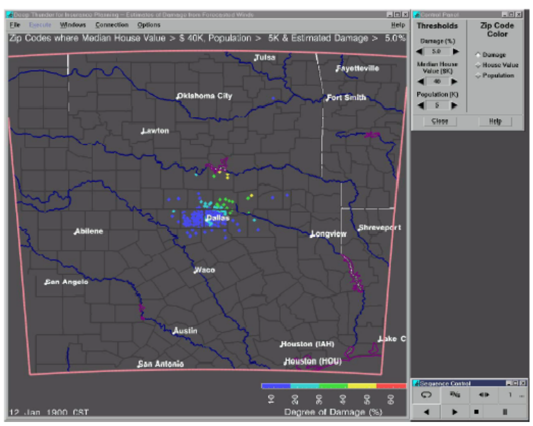
While the set of methods for creating such geographical visualizations and representing data on them is rather broad, the type of data attributes influence the effectiveness of a method for encoding the data features [14] [51]. For example, using color hue for representing categorical information on a geographical map can be very effective, but the same approach is less effective for representing numerical data [54]. To complicate matters further, the reuse of a chosen method is very difficult and often causes misinterpretation and confusion. As a result, visualization systems often use both optimal and sub-optimal choices among the available data encoding system [25].

While a wide range of symbols are used in geographic visualizations depending on the context of the analysis [50] [82], dots, circles, and icons are commonly used for representing geographical point data [50] [82]. At the same time, lines with direction glyphs are often used for representing movement data [6]. The size and colour of these visual variables are used to encode the associated numeric and categorical attributes [81]. Map colouring is commonly used for representing surface data [28]. The colour intensity of a given region is used to encode the numerical data associated with that region. Different colour hues are used to present categorical information [81].

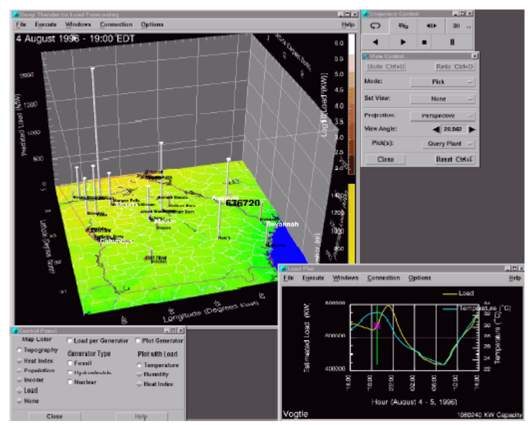
Spatio-temporal data sets often contain non-spatial attributes, such as numerical values, textual information, or images. Such attributes also play important roles in the analysis. In such cases, a combination of spatial visualization and non-spatial visualizations are used in geovisual analytics systems. These non-spatial visualizations represent the non-spatial attributes of the data sets that support analysts to explore and reason about the spatial data. For the single dimensional attributes simple static visualization such as graphs, or lists are used [72]. For multi-dimensional attributes, complex non-spatial visualizations are used, example includes Parallel Control Plots [33] [37] and Self Organized Map [68].

Geovisual analytics systems are often used to analyze multiple data sets [46] [50]. Three main streams of visualization techniques have been found [28] for analyzing multiple data sets. Figure 2.2 shows examples of these three different visualizations. The first technique uses single visualization for all the data sets. Data elements from different data sets are transformed into a common viewing framework [72] and are represented on this visualization using different visual variables [36] [50] [82] (see Figure 2.2a). Since data sets are visualized on the same visualization, qualitative analysis, such as understanding the co-relations between data sets, are clearly supported. User interactions are provided for supporting quantitative analysis, such as numerical querying of the data and parameter-based filtering [72].

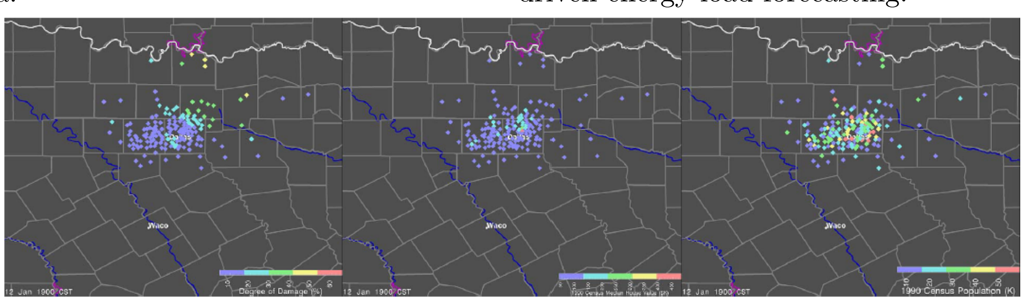
Multiple visualizations, either in adjacent windows or mosaiced together, are also used for analyzing multiple data sets. As opposed to the single view visualization technique, in this approach data sets are not required to transform for visualizing on the map. The geographical map is added as an underlying layer on the visualizations (see Figure 2.2c). One of the visualization can also have non-map based visualization (see Figure 2.2b). Visual variables are used in each visualization for displaying the data details. User interactions are linked among the visualizations, thus, manipulations in



(a) Common view visualization representing demographic data and weather forecast data.



(b) Multiple linked view using map view and graph for representing weather-model-driven energy load forecasting.



(c) Multiple linked map visualization mosaiced together representing demographic data and different attributes of weather forecast data.

Figure 2.2: Examples of different types of geographic visualizations [72].

one visualization interactively reflect on others [36] [50].

Like visual analytics, geovisual analytics is also not about simply presenting information, but an analysis session is more of a dialogue between the analysts and the data, where the visual representations are simply the interfaces or views into the data [70]. During this dialog, different perspectives on the data will be needed in each step, which are extracted by applying different user interactions. One of the widely used user interaction in this regard is filtering. In the field of geovisual analytics, zoom and pan of the geovisualization are often used for filtering the data based on their spatial attributes. Furthermore, temporal scales are used for limiting the data based on their temporal attributes. These operations extract a subset of data to a manageable range and represent them on the visualizations. These also allow analysts to see an overview of the data at a large level and identify the subset of the data of interest at every step of their analysis. In addition, the temporal scale allows analysts to understand the timing of events at very different scales.

Being an application of visual analytics, geovisual analytics also focuses on different location-related patterns and relationships within the data sets, with the express intent to support data analysis tasks. The fundamental requirement of analyzing the geospatial data is the incorporation of different contextual data within the analysis. The analytical reasoning framework provided by geovisual analytics systems allows analysts to explore the data sets with their geographical, temporal, and other domain-specific context to support specific analysis, exploration, and decision-making tasks. The goal is to support the analysts' decision-making capabilities by allowing them to assimilate complicated spatially oriented situations and reach informed decisions. In this regards, geovisual analytics allows analysts to reason about the data by creating assumptions and validating them with the evidence found in the data sets [40].

2.4 Geovisual Analytics of Movement Data

Movement data sets are of two types [6]: movement of an individual entity, and movement of multiple distinct entities. Both of these types of movement data can be analyzed with geovisual analytics systems. Geovisual analytics of movement data allows analysts to understand the movement patterns and characteristics of the moving entities. Several works have been conducted in different domains for exploring movement behaviors of an individual entity. A specific focus of this field is the analysis of very long trajectories [52]. Animated maps [9] [11], interactive cubes [38] [43], time-time plots [11], and temporal histograms [9] have been used to analyze such data sets.

Analyzing movement data for describing the behaviour of multiple entities is the other active field of study in geovisual analytics. A wide range of work has been analyzed within these types of data sets, ranging from the movement of tourists [8], animals [4], cars and trucks [7], and ships [47]. Analyzing these movement data sets reveals the patterns and trends of moving entities and explains many of their underlying phenomena and interactions.

Analyzing both of these two types of movement data is a challenging task because of their inherent complexities, which do not necessarily exist with other types of data [44]. The temporal resolution of the data sets or the time intervals of data points play an important role in the analysis. Temporal resolution directly affects the size of the data sets. Temporal and spatial accuracy are other key factors for many movement data analyses. The spatial distribution also adds complexity in these analyses. For example, analyzing the movement data sets representing tortuous paths that cross over themselves in relatively compact spatial regions are more complex than analyzing movement data representing straight paths because of the difficulties of visually tracing such paths [25]. In addition, the presence of random and systematic errors in the data sets adds difficulty in understanding the movement paths and patterns.

Considering all these challenges, pure visual methods are not sufficient for analyzing movement data [6]. Trajectories are often represented with flow lines by connecting the data points in geovisual analytics systems [10]. Simultaneously representing multiple movement activities from many entities, or large trajectories of a single entity, on a map adds additional visual complexity. In such cases, the visual decoding of information becomes difficult for the analysts [32]. Some alternatives for addressing this problem have been proposed, such as the use of animation [9] and space-time cubes [38], [43] [24] [55]. Figure 2.3 shows an example of a geovisual analytics system for analyzing ship voyages with weather information.

While many different approaches exist such as aggregating, filtering, and generating graphs and diagrams for analyzing the geospatial data sets, spatial view visualiza-

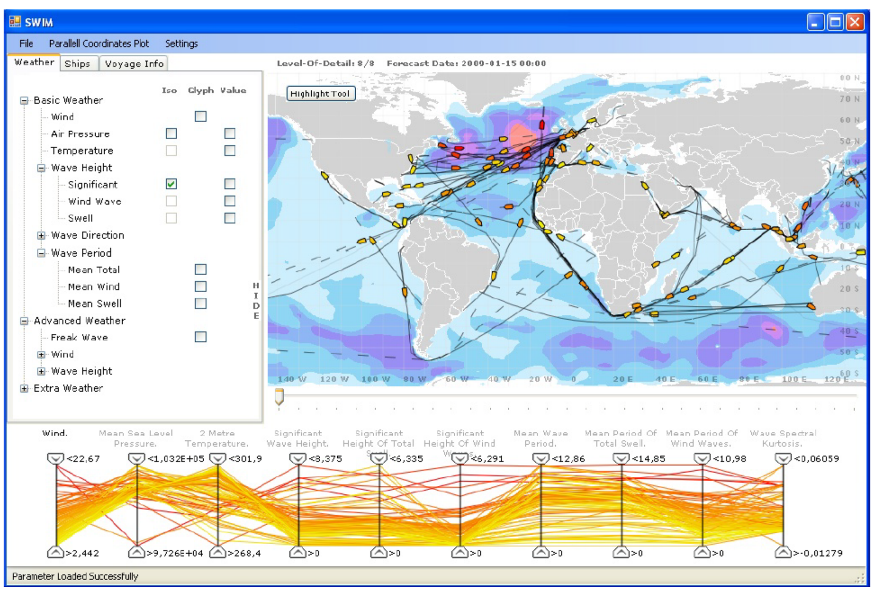


Figure 2.3: A geovisual analytics system for analyzing ship voyages with weather data. Ships and their routes are visualized using glyphs and lines on a world map. Significant wave height is displayed using an iso-surface. A temporal slider is used for the temporal filter. Weather parameters are plotted in the parallel coordinated plot [46].

tions always play the key role in the analysis process. Theoretical models have been proposed to guide the development of geovisual analytics systems to combine these approaches. A theoretical model [8] has been found for analyzing the large movement data of multiple discrete entities for patterns, which is relevant to this research. In this model, the possible types of patterns are defined as mathematical functions. These functions map entities and time to spatial positions, after which data transformations, computations, and visualization techniques facilitate the pattern detection. This model is suitable for large data sets, including those possibly too large for a computer's memory.

Similarity analysis of trajectories is one of the more interesting pattern analysis in this field. Extensive research has been done for supporting such analysis. Measures and algorithms are used for querying trajectory databases and clustering trajectories based on their similarities [7] [61] [79]. Multiple methods are also available for analyzing relationships between moving objects and their spatial context, such as visualization of the dynamics of the distances of moving objects to selected locations [19], computational detection of co-locations of moving objects in space, time, and space-time [85], and the occurrence of proximity between moving objects [59].

2.5 Geovisual Analytics of Events

Despite the extensive research on analyzing both movement and geographic point data, little attention has been given to linking these data sets together via events. Some works have been found to detect events from geographical point data [4] [32] [50]. Research also has been conducted detecting events from very long trajectories based on their spatial context [4]. The rest of this section will explain these approaches.

The LAHVA system [50] was designed to analyze and interpret data about biological agents, diseases, risk factors, and other health events in different geographical locations. This system uses two sets of geographical point data: human emergency room data and veterinary hospital data. Various statistical analyses have been applied to these data sets to identify events where human and animal diseases are correlated, and displays the result in a spatio-temporal view in conjunction with a statistical view for early identification of disease outbreaks with fewer false alarms. The spatial view of this system provides the ability to visually search the data sets for spatial locations from where higher numbers of people and/or animals are seeking medical help. To support visual decoding and understanding of the clusters, data aggregation is applied before representing them in the spatial view. Icons with varying shapes, sizes and colours are used to encode different attributes of these events, such as the number of data points, and types of diseases. The linked statistical window complements the visual search by representing results of different statistical analyses. A set of user interactions are also provided for supporting analysts' reasoning about the data sets. These user interactions manipulate the spatial window and the statistical analysis. The spatial view of LAHVA system is shown in Figure 2.4. This work considers the data sets as accurate and demonstrates a geovisual analytics system that uses computational methods, data aggregation, visual variables, and multiple linked views for detecting events from two geographical data sets.

In more recent work, GTdiff [32] dynamically explores a large geographical point data set for the geo-temporal changes of ancillary attributes over spatial and temporal ranges. This system makes an extensive use of spatial and temporal filtering, data aggregation using spatial and temporal binning, and visualizations. This system divides the temporal extent of the data into multiple temporal bins of different lengths, and the spatial extent into multiple spatial bins. The difference of the ancillary data of individual spatial bins over different temporal bins are calculated and represented in the form of a difference graph organized in an inverted pyramid. This unique visualization technique allows analysts to navigate the events of increment or decrement of the ancillary data from different temporal scales. Colour encoding is used in the geospatial view to represent spatial locations where changes are found. Events of positive and negative changes are separated using opposite colour hues. The difference view and geospatial view are linked as multiple coordinated views. The main components of the GTdiff interface are shown in Figure 2.5. While this approach deals with the differences among geographical point data, the notion of the event was not the focus of this work.

In another work, Andrienko et al. [4] suggests a conceptual model, in which trajectory data is considered as a combination of special events of diverse types and extents in space and time. They consider events as independent objects extracted from long trajectories and spatio-temporal context data. Using this model, events can be extracted from the trajectories based on the spatial and temporal context (entities

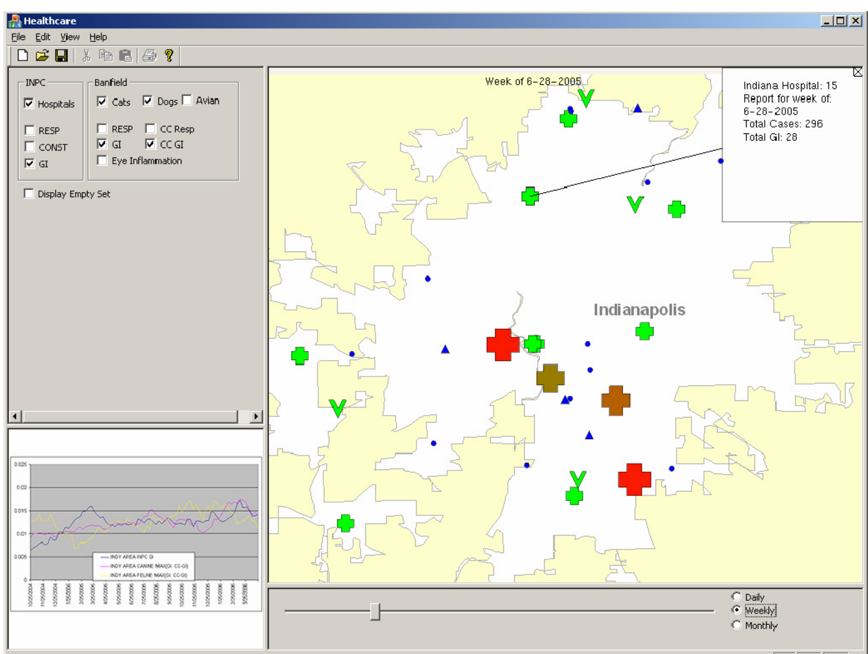


Figure 2.4: LAHVA geospatial-temporal view [50].

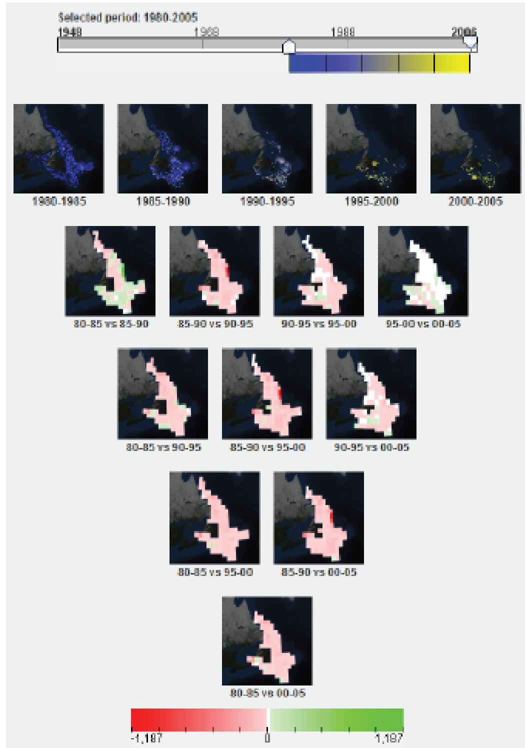


Figure 2.5: Main components of GTdiff: temporal view (top), and difference view (bottom). Data is divided into 5-year temporal bins and is shown in the difference view; the the red-green colour encoding represents the positive and negative changes [32].

moving around a certain area within a certain time) and proximity with other moving entities (one entity meets another). Interactive query tools are used to allow analysts to define movement events that are of interest. The analysis process is iterative; thus the query results are added as new objects for further analysis. This work considers the data sets as accurate and using the process of matching the movement data to the contextual data, or another movement data, as the means for extracting events. An example of proximity-based events detection is shown in Figure 2.6. The main focus of this work was to extract events using a diverse type of filtering on the data sets and represent these events on the visualization.

In more recent work [76], semantic events are extracted and analyzed to enable a

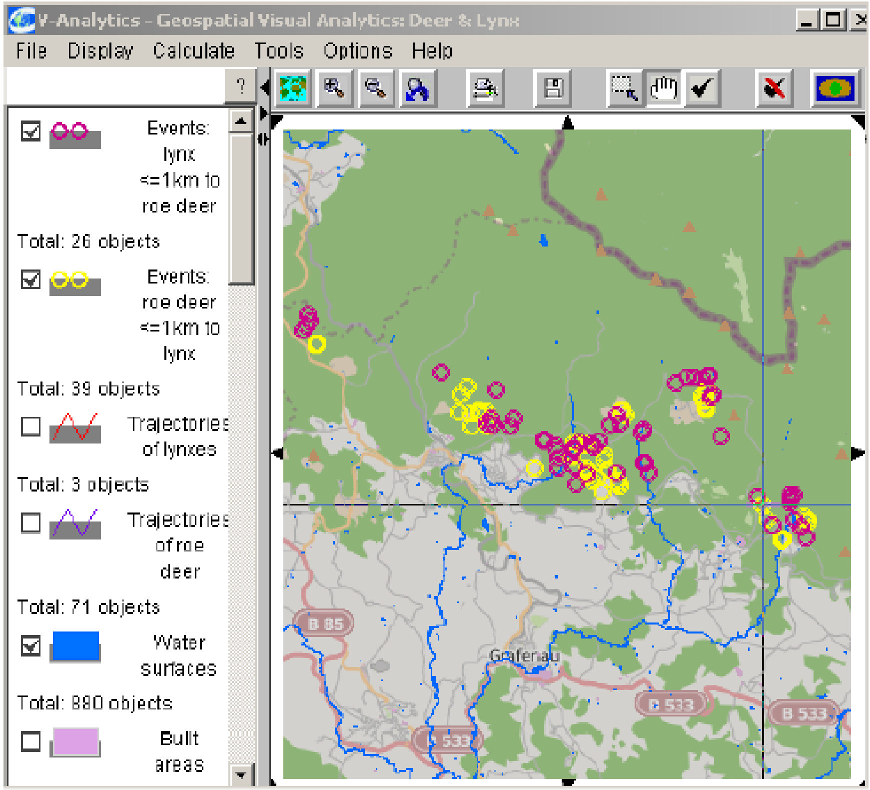


Figure 2.6: The events of spatial proximity between roe deer and lynxes are visualized on the map. The yellow circles represent the events of the roe deer and the pink circles represent the events of the lynxes [4].

better understanding of maritime events. This work proposed a semantic trajectory event framework based on the concept using an ontological approach. As an extension of Simple Event Model (SEM) [74], this framework computes raw spatio-temporal data points to obtain a semantic trajectory [67]. It includes a four level framework that applies filtering, segmentation, and optimization to the trajectory data, and finally enriching them with ancillary information. After this, using different ontologies and the maritime operators' knowledge stored in the form of semantic rules, events are automatically identified. This work offers a consistent way of modeling both semantic trajectories and semantic events. The identified events are also shown in a geographic visualization. Figure 2.7 shows the visualization of speed up and slow down events of vessels on a 3D web mapping interface. While this work was intended to identify and analyze the movement events in the maritime data, the concept of identifying and analyzing the positional event anomalies and linking multiple data sets through events were not the focus of this work.

2.6 Discussion

In this chapter, information visualization, visual and geovisual analytics, and different research domains that correspond with analysis of geospatial point and movement data were briefly reviewed. Some of the different approaches were also surveyed that explore events in geospatial point data and movement data. In particular, a number of different practices were discussed that are relevant to this research: geovisual analytics, multiple coordinated views, user interactions, event extraction, and event visualization.

From this literature review, a number of difficulties related to developing a geovisual analytics system were explained, such as information overload, visual decoding of

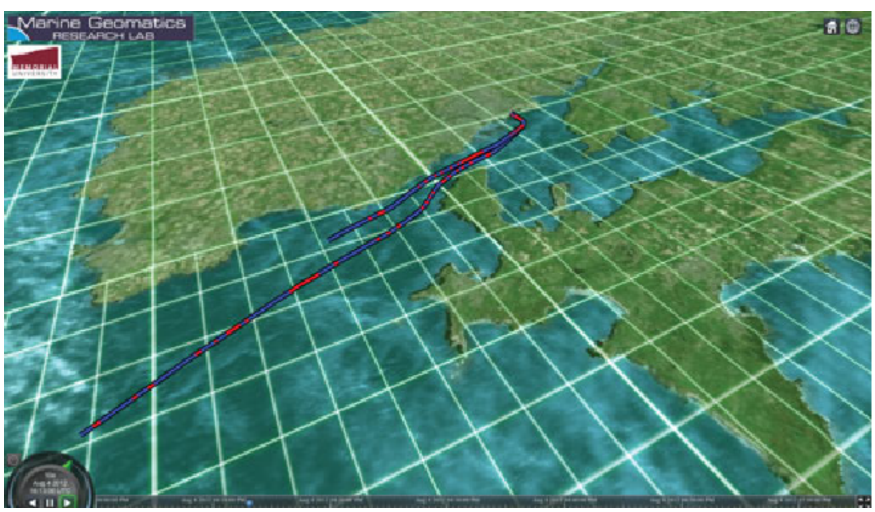


Figure 2.7: Vessels speed up and slow down events on a 3D mapping interface. The blue segments of the trajectories representing the decreases in the vessels' speed. The red segments of the trajectories representing the increases in the vessels' speed [76].

information, and support for analysts' analytical reasoning about the data. Different approaches were reviewed that address these difficulties and analyze these data sets in different domains. Few of them directly addressed the representation and exploration of events. However, all of these works considered the data sets as accurate and complete. Thus, analyzing the data sets that are not accurate or have missing data remains challenging.

Although, no existing work addressed the inaccuracy and incompleteness of the data, this literature review surveys the current state of art for designing a geovisual analytics system and its components. The knowledge discovery loop of visual analytics systems was the key inspiration of the developed system for analyzing event anomalies. The computational methods, visual variables, and user interaction techniques used in this system were also designed from the guidance of existing works.

The literature review shows that the design of visual variables for representing different aspects of the data requires symbols, modifiers, and proper use of colours. These modifiers and colour encoding data detail differentiate data elements from each other. In particular, the symbol and modifier for representing movement data and point sample data, and the theory for using colour for representing spatial information on the map are taken into account while designing visual variables for representing event anomalies.

A number of different approaches were also discussed for visualizing multiple data sets. This survey was used to provide useful insights for designing effective visualizations for event anomalies analysis. Multiple coordinated views were found particularly appealing in this regard. This type of visualization has been showing promise for exploring relationships among data. Most of the geovisual analytics systems from different decision support domains take the advantage of multiple coordinated views. Therefore, this approach of visualization was used for this work.

The visual analytics information seeking mantra also suggested that user interaction incorporate analysts into the knowledge discovery loop. While diverse types of user interactions are available, three main types of user interactions were found relevant for analyzing event anomalies: filtering, linked brushing, and details on demand. Thus, following the practices of using these interactions from existing geovisual analytics systems, the user interactions for the new system was designed.

The details of each of these components and how they are organized to create the knowledge discovery loop within the new system is shown in Chapter 3. A prototype system has been developed that fits these components together and will be discussed in Chapter 4, along with a case study to demonstrate a real world data analysis example.

Chapter 3

Geovisual Analytics for Event Anomalies

This research is focused on developing geovisual analytics system for exploring event anomalies from independently collected geographical point data and movement data representing same entities. The goal of this research can be divided into two parts: first, to identify methods for extracting events anomalies from the data sets and second, to visually represent them within an interactive interface for analysts' exploration and analytic reasoning. Thus, four elements are the key components of this system: event extraction, event anomaly detection and thresholding, anomaly representation, and interactive filtering and exploration. The remainder of this chapter will present each of these elements with the specific design decisions that fit them with the features of the data and the analysis tasks.

3.1 Event Extraction

Events are defined by the change of existential, spatial, or thematic properties over time. These changes also define the boundary of events [10]. Andrienko et al. have explored methods for extracting events from movement data based on these changes [4]. Their work was focused on extracting events from accurate and complete data sets, and their objective was to understand the underlying phenomenon of such events based on their spatial context or surrounding events. Data elements in the data sets associated with same events were not linked in the data sets. The lengths of the events were also unknown. Thus, identifying these links based on certain conditions was the key task of their event extraction process.

In this work, events are extracted from the data sets where data elements associated with same events are linked among the data sets with entity identifiers. The lengths of the events are also defined prior to the analysis (e.g., hourly event, daily event, and weekly events). Furthermore, all the data elements from both of the data sets are associated with specific events. Having such considerations, the approach taken in this work for extracting events is much simpler. However, the work considers the existence of errors in the data sets, and the objective is to understand the underlying phenomenon of such errors based on the context of their spatial and temporal extent and/or surrounding events.

The event extraction process starts with segmenting both of the data sets based on the event length. In the next step, segments from both of the data sets representing the same temporal extent (e.g., same day) are paired. Finally, data elements describing the same entities within these pairs constitute events. Thus, each event has a movement trajectory and position information about one entity. Figure 3.1 illustrates the entire process.

For example, in the fisheries domain, daily fishing events can be extracted from fishing

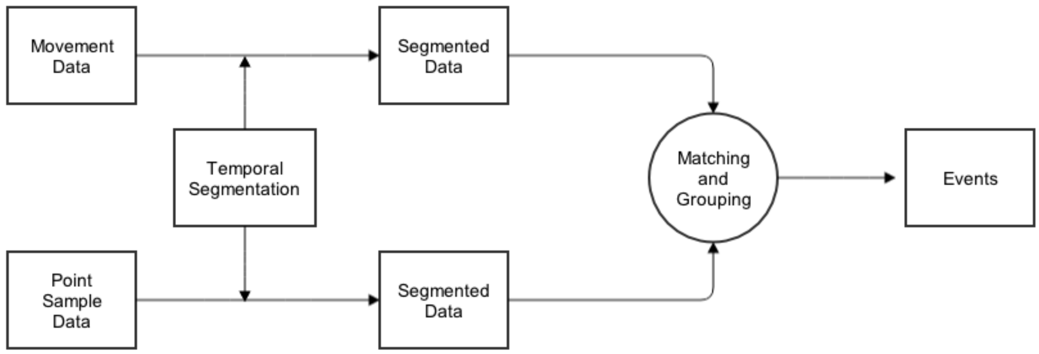


Figure 3.1: The process of event extraction from movement data and geographical point data describing the same set of spatial events.

vessels log book data and their daily catch report. Fishing vessels log book data contain vessels' multi-days voyages where data points are taken at a regular interval (i.e., positional information in every hour) and the daily catch reports contain the locations where the fishing were executed in each day. Using the approach mentioned in the previous paragraph, first, these data sets are segmented where each segment contains data elements from a specific day. Then the trajectory information from a given vessel for a particular day are combined with the catch report from that vessel for the same day. These combined data elements constitute the daily fishing events.

Extracting events using the approach explained in the above two paragraphs may be a trivial task for temporally synchronized data sets. However, without the temporal synchronization further challenges are needed to be addressed. When the level of granularity for movement data is higher than the geographical point data, two movement data points are associated with each of the events; describing where the entity was before and after the event. Alternatively, a series of movement data points are associated with each of the events when movement data has a lower level of granularity than the geographical point data. However, in both cases a degree of uncertainty is present for the matching of data elements. This uncertainty must be addressed when seeking anomalies in the data sets for avoiding false-positive results.

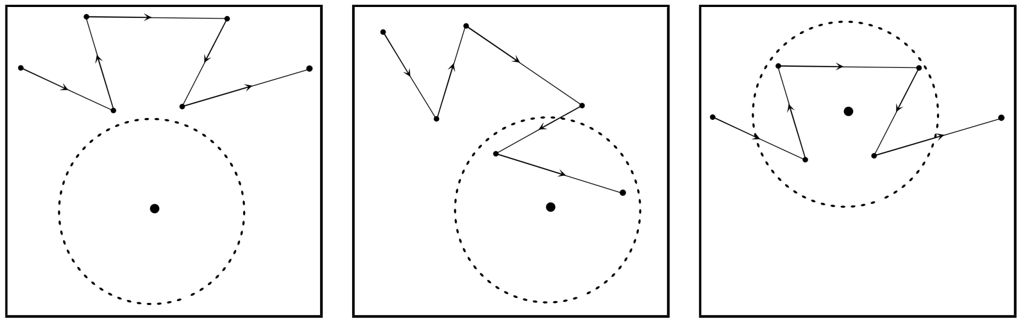
3.2 Anomaly Detection and Thresholding

Once data points related to individual events are identified, the next task is to detect events that have anomalies. When data sets are temporally synchronized, this detection is a trivial task. The distance between the event location and the entity at the time of the event can easily be determined. If the distance exceeds an accepted value, then the event is considered to have an anomaly. For example, an event occurred on a particular day at 14:00 hours. By identifying the location of the entity that was performing the event on that instance of time and then comparing the distance with the acceptable value is sufficient for detecting the anomaly.

Like the event extraction, the real world event anomaly detection process also experiences additional challenges because of different factors: likelihood of temporarily synchronized data sets are very low; mismatch in the temporal granularity between the data sets; events are performed over geographical regions; and events are performed for a certain period of time such as an hour, day, or week. In many cases the length and the starting time of events are also not known. For example, when fishermen report geographical points in the ocean as their location of daily fishing activities, the actual fishing events may be performed over a geographical area of several square kilometers and the fishing events might lasted only for few hours starting any time during that day. However, their fishing vessels VMS data (the movement data set in this case) shows not only the movement activities within the fishing region, but also other movement activities from the same day, such as their travel to and from the location of fishing. They also may have multiple fishing sessions at different locations at the same day.

Considering all these facts, a more complex method must be employed for determining event anomalies. The approach taken in this work considers the geographic distances between event locations and the data points on the associated movement paths, along with the amount of time entities spent within a certain distance from event locations. Two threshold values are used to determine the acceptable distance and time. A subset of the movement data points is determined, which consists of those data points that are within the distance threshold from their event locations. Next, based on the granularity of movement data points and the number of data elements in the subset, the time an entity spent around the event location is determined. For a given event, if this time is more than the time threshold the event is considered normal, and an anomaly otherwise.

For example, fishermen report their fishing events and record vessels' positions in the vessels' log book for every hour, where a typical fishing event has a length of three hours, and the fishing area is of one kilometer in radius. Thus, if the vessel's log book entry associated with a given fishing event showing at least three data points within a distance of one kilometer from the a reported fishing location, the event is normal. Figure 3.2 shows such three examples; each has six movement data points. The threshold values here are one kilometer for distance and three hours for time. The movement paths are represented with flow-lines where the dots on the lines shows the movement data points and the arrows shows the movement direction. The event locations are shown with a circle indicating one kilometer distance from the reported geographical point at any time. In Figure 3.2d, any data point of the movement path was not within the acceptable distance from the fishing location at any time. In Figure 3.2e, the movement path was close to the fishing location but only for two hours. In Figure 3.2f, the movement path was within the acceptable distance for four hours. As a result, the first and second events may have potential event anomalies,



(d) (e) (f)

Figure 3.2: Examples of potential anomalies between movement paths and event locations.

while the third one is not.

While this approach considers many aspects of event anomalies, choosing appropriate threshold values cannot be done automatically. Determining these values not only depends on the knowledge of the data sets, but also requires domain-specific knowledge regarding the actual activities of the entities. Thus, these values should be adjusted by the analysts. Simple slider controls can be employed for this purpose that supports independent adjustments of these values.

This anomaly detection procedure was designed in consideration of data sets that explain events in two different ways with a focus on anomalies that represent a mismatch in the event locations. However, event anomalies in different domains may occur for different reasons, such as abrupt changes in velocity and visiting restricted regions. In such cases, different event anomaly detection techniques will be required to identify event anomalies. Nevertheless, once the event anomalies are extracted, the remainder of this system will be able to show the relations between the data sets in the context of these anomalies.

3.3 Anomaly Representation

After detecting potential event anomalies, the next step is to represent them in a meaningful way for further analysis. This work represents anomalies in a map-based and a non-map-based visualization, namely map representation and tree representation, respectively. A visual variable is introduced for representing event anomalies on the map-based visualization. In the non-map-based visualization the event anomalies are represented as a hierarchical list grouped by the event entity identifiers. Each of these views are discussed in detail in rest of this section.

3.3.1 Map Representation

Representing multiple event anomalies on the map requires representing movement paths, event locations, and positional discrepancies of all the events. Thus, the visual variables should allow analysts to visually group data associated with individual events easily and quickly. At the same time, the information about the data sources (i.e., attributes from the geographical point data and attributes from movement data) should be communicated. Thus, by looking at the visualization, analysts are able to identify individual events, their movement paths, event locations, and positional discrepancies. Furthermore, the visualization should not overwhelm the analysts making it difficult for them to interpret.

The visual variable for representing events in this work uses flow lines for showing the movement paths. Chevrons show the actual movement data points on that path and the directions. Points are used for representing event locations. Finally, the positional discrepancies are represented using lines connecting the event locations with the closest point in the movement path associated with those events. An event anomaly represented with this approach is shown in Figure 3.3. Here the white flow

Figure 3.3: Example of the event anomaly representation of an oceanic event performed by a vessel. Dark blue represents the ocean and dark green represents the land. The white flow line represents the trajectory of the vessel, with the arrowheads represent the locations of the the data points of the trajectory. The yellow line connects the nearest trajectory data point with the event location found in the geographical point data.

line shows the movement path of the entity and the yellow line connects this movement path to the event location.

One disadvantage of joining the event locations with the movement paths is the added visual complexity when showing multiple events simultaneously. However, this approach has other advantages that outweigh the disadvantages. The first advantage of this visual variable is the line providing perceptional grouping of event data elements as described by the principal of connectedness [60]. This law states that connecting different graphical objects by lines is a very powerful way for expressing the relationships. Thus, the additional cognitive load for analysts is reduced.

The other advantage of this approach is the pre-attentive processing of larger positional discrepancies and clusters of event anomalies. Connecting the event locations with the movement paths will carry more visual weight in the display for those events that have larger discrepancy because of their longer connecting lines. Many such lines close to each other will also create a visual cluster. Thus, they will be perceived pre-

Figure 3.4: Example of pre-attentive processing of event anomalies within a set of oceanic events performed by vessels. A cluster of event anomalies is visible at the top left part of the figure. Three other positional discrepancies can also be pre-attentively identified in the figure.

attentively, which may guide analysts to quickly choose a smaller geographical region or specific events for further investigation.

An example of pre-attentive processing of event anomalies is shown in Figure 3.4. Event anomalies in this figure are represented using the same colour scheme as in Figure 3.3. In this figure two geographical locations are easily identified where most of the event anomalies occurred. Two event anomalies having larger discrepancies also stand out where the event locations are on the left of the figure and the movement paths are on the right.

Another important aspect of the event anomaly analysis is the detection and representation of missing data points within the movement data. Some data points may be missing in the movement data sets because of instrumental failure or communication error when collecting them remotely. In order to maintain a consistent data set, interpolated data are added to the data set. This interpolation considers straight line between the known data points before and after the missing data points. Anomalies can be detected from the computational method because of the interpolated data

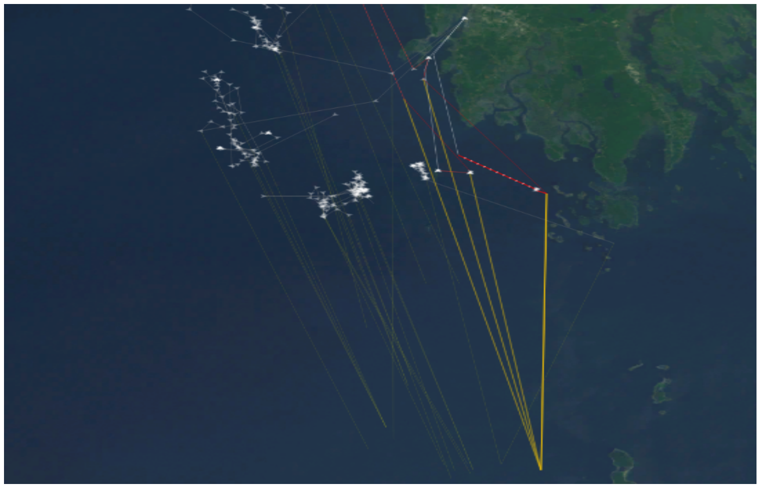


Figure 3.5: Example of representing missing data points within oceanic events. The dark red line shows the interpolated segments of a trajectory where the data points are missing. The white empty circles on the line represent the estimated locations of the data points along that straight path.

points. Thus, it is important to visually convey this interpolation back to the analysts. For representing such data points, empty circles are used (see Figure 3.5).

Representing event anomalies on the geographical map without overpowering the visual representation so that they can be readily perceived and correctly interpreted is a challenging task [49]. The Opponent Process Theory of Colours is utilized in this regards for representing features of the data sets on a map. The colours used for representing event anomalies are perceptually distinct from the base map. For example, oceans and land on a map are represented using shades of blue and green, and other features are represented using shades of black. Thus, based on the Opponent Process Theory of Colour [30], yellow, red, and white are available for representing the aspects of event anomalies.

While considering all the facts described in this section, the chevrons and empty circles on the flow lines may cause visual clutter when showing a large number of events all together. This information also may not add much value to the analysts at some instance of the analysis, in particular when showing a large number of movement paths. Thus, a feature that interactively shows and hides the chevrons may improve the situation. Similar options can be provided for the empty circles. Missing the movement data points may not have an effect on the analysis when they are lower in number or not consecutive. Thus, the interactive features may hide them when they are not consecutive for an analysts' chosen period of time.

3.3.2 Tree Representation

Another visualization is used in this work to complement the map representation of event anomalies. In this visualization, the list of event anomalies are grouped by their entity identifiers and represented as a tree. This listing also represents the statistical and ancillary attributes of event anomalies.

In this tree structure, entity identifiers that have at least one event anomaly added as the child of the root node. Statistical information about each of the entities such as the number of event anomalies and the total events performed by that entity within the chosen spatial and temporal extents are also shown beside the entity identifier. Event anomalies are added as child nodes of their entity identifier, which is the second level of the tree. Multiple information of events are represented in this level of the tree, such as date and time of events, and ancillary data that is associated with those events. The root node also represents statistical information such as the total number of event anomalies vs. the total number of events.

Figure 3.6 shows an example of a tree representation of event anomalies. The root node indicates that there are 28 potential event anomalies out of 51 events. The first level of nodes shows the list of entity identifiers represented by integer numbers. The bracketed numbers beside each of the identifier represent the number of event anomalies and the total number of events performed by that entity, respectively. The second level of this tree shows individual events represented by the event data. The bracketed integer number beside each of the event date represents ancillary attributes associated with individual events.

The tree representation is extended with checkboxes. By clicking the check boxes individual event anomalies or all the events anomalies performed by a particular entity can be hidden and revealed from the map representation. This feature allows analysts to remove the event anomalies from the map view that are not of interest.

3.4 Interactive Filtering

Interactive filtering is the key user interaction that any geovisual analytics system provides. Interactive filtering the data often used to reduce the visual clutter that arises when many data elements are displayed in close spatial proximity [34]. In this work, interactive filtering is provided by four concurrent filtering parameters: temporal extent, spatial extent, anomaly threshold, and ancillary data. The goals of these filters are twofold: first, to filter out uninteresting events and second, to allow ana-

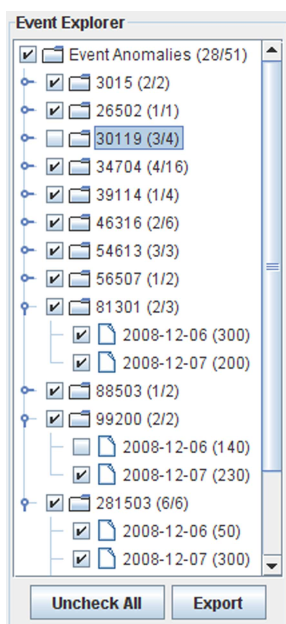


Figure 3.6: Example of the tree representation of event anomalies.

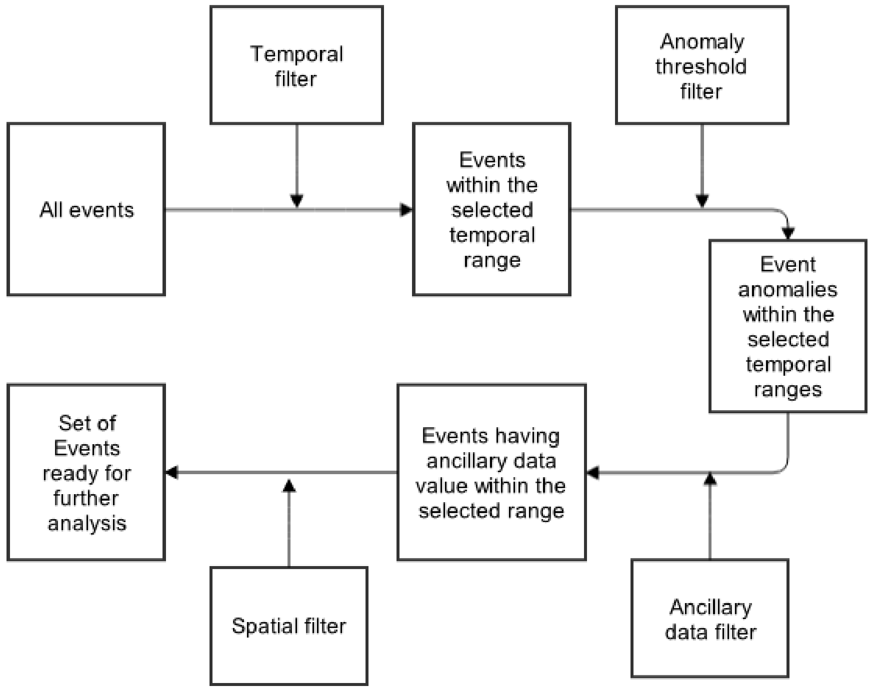


Figure 3.7: Interactive filters applied to events for extracting a smaller number of potential event anomalies.

lysts to isolate event anomalies of particular interest. These filters direct analysts' attention to potentially interesting event anomalies as required by their specific analysis activities, without relying on the automatic discovery of event anomalies based on a set of predetermined parameters. Figure 3.7 shows how these filters applied on the events and extract smaller number of event anomalies for the analysts for further analysis.

3.4.1 Temporal Filter

Temporal filtering is implemented through a time line and a window slider. The time line shows the temporal range of the data sets that is calculated dynamically. Analysts can adjust the lower and upper bound controls to select a temporal window. By adjusting this temporal window, all the event anomalies from this time period are interactively visualized in the map and tree representations. Thus, this filtering allows analysts to visualize event anomalies only from a specific period of time that is of interest.

While two date inputs can also perform the same filtering, this window slider supports another important aspect of event anomaly analysis. By sliding the temporal window, analysts can understand the timing of events at very different scales and compare event anomalies for their trends over the different periods of the time line. For example, while analyzing anomalies in the reports of commercial fishing activities, setting a temporal window for four months and sliding them over a multi-year time line will show the trends of incorrect reporting in different seasons. Figure 3.8 shows a typical temporal window slider control over a two-year time line.

3.4.2 Spatial Filter

While a temporal slider allows analysts to choose the temporal extent of the data that they are interested in, they might also be interested in a specific geographic region when they are analyzing data sets covering a large spatial extent. Panning and zooming in the geovisualization allow analysts to change the area of focus. These panning and zooming operations are used for filtering the data from both of the visualizations. Using this filtering technique, analysts can start with a larger spatial extent of the data (e.g., data from the entire country). After this, they can choose a smaller extent derived by different aspects of the data (e.g., a particular city that has more event anomalies than others).

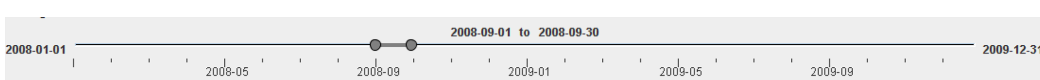


Figure 3.8: Example of the window slider. The time span represented by this slider is two years, January 01, 2008 to December 31, 2009. The timeline is marked for each months and labeled in four month intervals.

The advantages of applying these filters are twofold. Removing the event from the visualizations will provide a smaller set of the event anomalies. Thus, analysts will not deviate by the other event anomalies outside of that region. The other advantage is that removing events from outside of the selected spatial extent will reduce the memory load for the computer system. Since many of the spatio-temporal data sets are too large for the computer memory, this is an effective technique for building a computationally efficient geovisual analytics system for analyzing such large data sets.

3.4.3 Anomaly Threshold Filter

Distance and time threshold parameters were introduced earlier in this chapter explaining detection of event anomalies. These values can also be adjusted for filtering event anomalies. For example, initially a geographical region that has many positional discrepancies can be selected for detail analysis. Once the analysts zoom and pan to that region, they might be interested only on those event anomalies that have longer positional discrepancies. Thus, changing the threshold parameters, event anomalies with smaller positional discrepancies are removed from both of the visualizations. An example of this filtering approach is shown in the Figure 3.9.

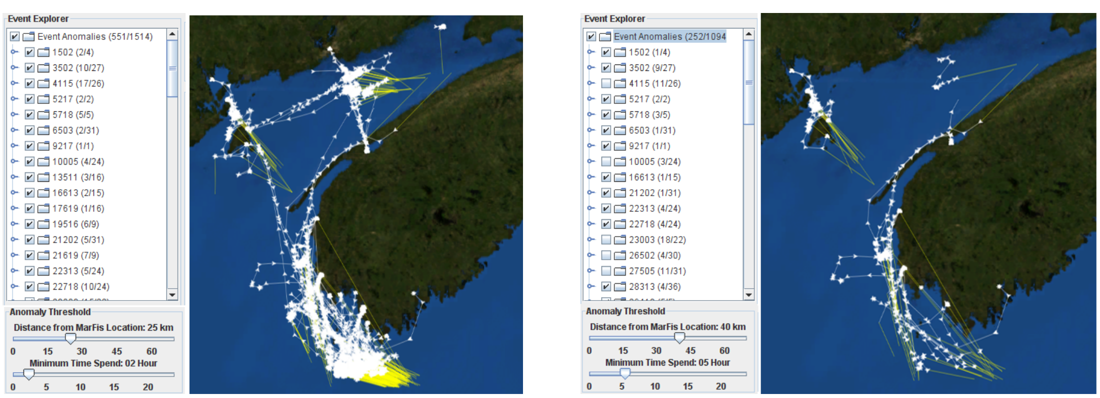
3.4.4 Ancillary Data Filter

The ancillary data represents the non spatio-temporal aspect of events. For many anomaly analysis tasks this data plays a significant role. For example, analyzing anomalies in the reports of commercial fishing events, the catch amount may play an important role. Analysts may pay higher attention to the suspicious fishing events that reported higher catch amount, and may ignore if the amount is very low. Thus, an ancillary data filter is introduced in the system for filtering the event anomalies that exceeded an analysts' selected value. This filter is implemented using a spinner control for filtering events based on the integer type ancillary data. However, this filter can also be implemented for other types of data depending on the specifics of the events.

3.5 Interactive Highlighting

Interactive highlighting is a widely used interactive tool for visual analytics systems. In this work, event anomalies are represented on a map and complementary information is shown within a tree structure. The map representation shows the locations of event anomalies and the tree representation shows the entity identifiers information, event identification, and associated statistical and ancillary data. The interactive highlighting feature is provided in this system by using linked brushing, which allows analysts to visually link all the data associated with individual events.

Analysts can highlight an event anomaly from the map representation by clicking on the chevrons, flow line, or the line connecting the event location and the movement



(a) Before. (b) After.

Figure 3.9: Example of anomaly threshold filtering shows a reduced number of events after changing the threshold values. The anomaly threshold sliders (lower left) are linked with both the tree representation of the events (upper left) and the geographic visualizations (right). The before and after configurations show how manipulating the threshold values changes which event anomalies are shown in the visualizations.

path. The event can also be highlighted from the tree representation by clicking on the corresponding node. Selecting the event from either of these visualizations highlights it in both. Analysts can also select and deselect multiple events by holding the control key from the keyboard while clicking the mouse.

To visually pop out the highlighted events in the map representation, a chromatic contrast is maintained while choosing the colour for these lines. The colour intensity for the event anomalies from other entities is also reduced. Thus, analysts can focus on the selected event without losing the information of surrounding event anomalies. In the tree representation, highlighted events and the corresponding event entities are highlighted by adding a gray background.

Another aspect of interactive highlighting is the inspection of the events and the associated entities for understanding their trends and patterns. Thus, additional contextual information about the movement path of highlighted entities during the selected temporal range are added to the map representation to understand where the entities travelled before, after, and between the events having anomalies. The other event anomalies from the same entities, if found in the selected temporal range, are also shown as highlighted. The movement paths of the highlighted entities that are not part of any event anomalies are shown with white straight lines. To differentiate event anomalies at this mode of visualization, three different levels of visual intensities and colour encodings are used. For example, in case of oceanic data sets the movement paths of highlighted events are shown at a high level of intensity using red colour, the lines connecting the movement paths and event locations are shown at a high level of intensity using yellow colour; the movement paths of other anomalies from the same entities of the selected events are shown at the normal level of intensity using red colour, the lines connecting the movement paths with event locations are shown at the normal level of intensity using yellow colour; the path showing where the vessels travelled before, after, and between these event anomalies are shown at a high level of intensity using white colour; and all the remaining non-selected anomalies are shown at a low level of intensity using white colour for the movement paths and yellow colour for the connecting lines between movement paths and event locations. This allows the analysts to readily see what is selected, with all other contextual information about the entity.

Figure 3.10 shows an example of highlighted event anomalies. In this example, the red flow line with higher intensity is representing the movement path of the highlighted event. The red flow lines with normal intensity are representing the movement paths of the event anomalies from the same entities of the selected event. The white lines are representing the movement paths of the same entities showing where it travelled before, after and between the event anomalies. White flow lines in lower intensity are showing the movement path associated with event anomalies from other entities. The yellow lines are joining the movement path and event locations: higher intensity is

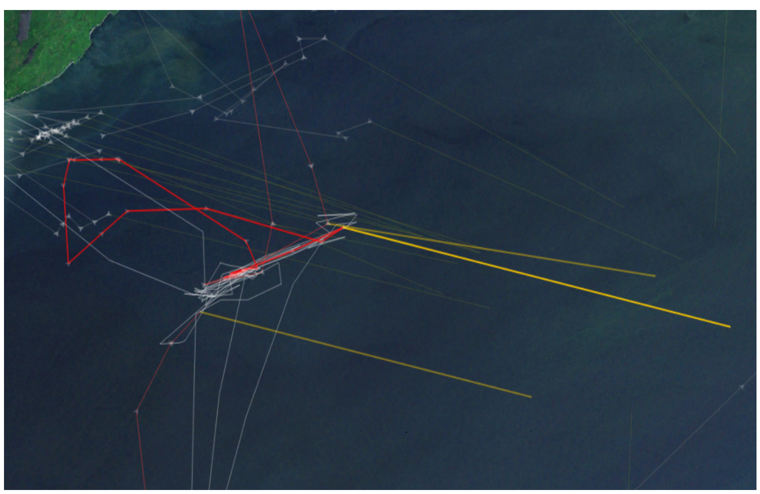


Figure 3.10: Example of highlighted event anomalies within an oceanic data set. A set of anomalous events are shown in this figure, with one of these highlighted (red flow line and yellow line with high intensity). Two event anomalies performed by the same vessel (red flow line and yellow line with normal intensity) are shown, along with the movement activities in between these event anomalies (white line with normal intensity). The event anomalies from the other vessels are shown in lower intensity.

used for the event that is selected, normal intensity is used for the events from the same entity as the selected event, and lower intensity is used for the event from other anomalies.

3.6 Analytical Reasoning

Analytical reasoning is the central of the analysts' task of applying human judgments to reach conclusions from a combination of evidence and assumptions [70]. Within a geovisual analytics system, the support for analytical reasoning is provided by incorporating the analysts into the knowledge discovery loop. Thus, the components of a geovisual analytics system are required to be organized in such a way that analysts can iteratively investigate the data and discover new insights, such as patterns and trends. Within this iterative process, each iteration generates new knowledge which may lead to new hypotheses. These hypotheses are then validated in subsequent iterations of analysis. Thus, the system is required to reconfigure its analysis process for supporting such activities.

A guideline for developing visual analytics systems is discussed in Section 2.2. Being a subclass of visual analytics, geovisual analytics systems also follow the same guideline. Figure 3.11 shows how the components described earlier in this chapter are organized in the developed geovisual analytics system for supporting analytic reasoning. This framework is developed from the motivation of visual analytics mantra proposed by Keim et al..

The system starts with extracting events from the data sets. After this, the system detects event anomalies based on initial thresholds values, which are subjected to the domain specific knowledge. Initial spatial and temporal filter parameters can also be added. This initial filtering reduces the size of the data to show on the visualizations.

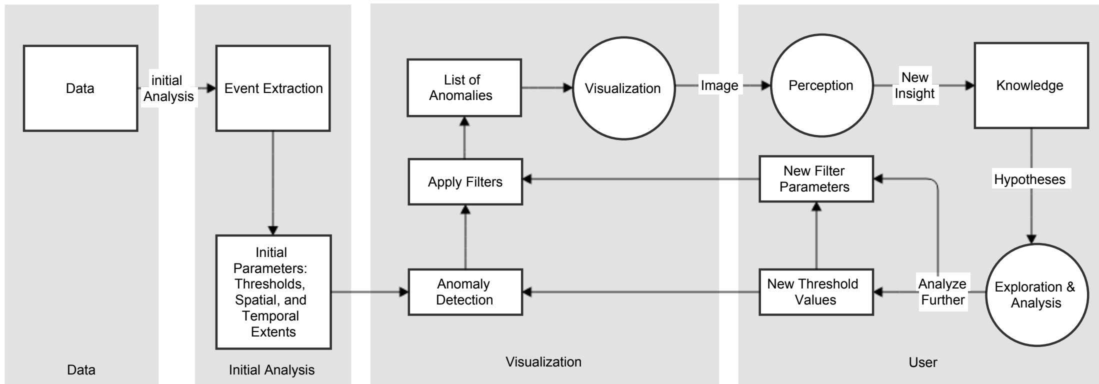


Figure 3.11: The knowledge discovery loop of the geovisual analytics system for analytic reasoning of event anomalies.

This filter was introduced based on the theoretical model proposed by Andrienko and Andrienko [8] (discussed in Section 2.4). If the system is analyzing smaller data sets this filter can be set to the full extent of the data. Once these initial actions are completed, the system shows the result on the visualizations (map representation and tree representation).

At this point, the system enters into the analysis loop. The knowledge discovered from the initial representation may drive analysts to select new parameters for filtering and thresholding. Once the parameters are set by the analysts, the system reconfigures its computational process and create a new list of event anomalies. This event anomalies are then presented in the visualizations. This loop continues until the analysts gather the knowledge they are looking for.

Since the system has the ability to reconfigure its computational process, analysis in different directions can be performed in each of the iterations. For example, an analyst may identify a cluster of event anomalies at the end of one iteration, and wish to pan and zoom to that region. In the next iteration they might be interested in understanding event anomalies from that region but with a larger temporal extent. Thus, applying a new temporal filter creates a new set of anomalies and presents them in the visualizations. At this iteration, analysts may find some interesting event anomalies and can investigate further using the interactive highlighting tools. Alternatively, they may not find any interesting anomalies and want to set new spatial, temporal, and threshold parameters.

The interactive nature of these filtering, thresholding, and highlighting tools, along with the analysts' knowledge and experience about the data, and their understanding of the domain, allow the analysts to make informed choices for setting the analysis parameters in each iteration. This interactions allow analysts to perform diverse types of analysis, such as seeking a specific type of anomalous behaviour or perform exploratory analysis of large data. Thus, analysts can explore many different scenarios in which the anomalies might be present. Therefore, the organization of visualizations and the interactive tools in this developed system will support analysts to perform analytic reasoning about the event anomalies in order to gather new knowledge.

This interactive nature also allows analysts to explore many different scenarios in which the anomalies might be present. Thus, the organization of visualizations and the interactive tools in this developed system will support analysts to perform analytic reasoning about the event anomalies to gather new knowledge.

3.7 Discussion

Positional event anomalies are ill-defined. Thus, identifying such anomalies requires the incorporation of spatial, temporal, and domain specific contexts, along with analysts' experience and expertise. Although, the use of geographical maps and time lines are well known methods for adding the spatial and temporal contexts within the analysis, incorporating analysts' experience, expertise and domain specific contexts requires the creation of innovative methods. In this work, designing the method for identifying event anomalies was challenging, since positional event anomalies in the data sets describing events from same entities is an unexplored problem in geovisual analytics. The new method is also required to be working with uncertainties in the data sets, such as data quality problem, and intentionally introduced incorrect data. To address these issues, the anomaly threshold has been designed to allow analysts to incorporate the relevant domain specific context based on their knowledge of the data and the domain in which it was collected.

A few methods have been found for representing events detected from either geographical point data sets or trajectory data sets. Many methods are also available for representing movement data. However, the visual representation of event anomalies in multiple data sets was not addressed in any of the previous works. While representing events may be an easy task, visually representing the anomalies within the data set was challenging. Two particular challenges were the perceptual grouping of associated data, and keeping the visual complexity low when representing a large number of anomalies. The perceptual grouping of the geographic point data and movement data is done by following Gestalt Laws. In addition, the principle from the Opponent Process Theory of Colour is extensively used while choosing colours for representing different components of the event anomalies. This visual variable also confirms the pre-attentive processing of severe anomalies and clusters of anomalies. Therefore, these theories can be considered as the theoretical foundation that supports the potential benefits of using the visual variable for representing two different types of data and links between them.

The user interaction is one of the fundamental features of any information visualization system. Analysts get more power over the control of the presented information because of these interactive tools. These tools also include analysts in the knowledge discovery process and allow them to perform analytic reasoning about the data sets.

A theoretical framework is shown in Section 2.2, which is followed in order to design the new system. A prototype system is designed by following this new framework will be shown in Chapter 4. A case study of real world event anomaly analysis will also be presented to illustrate the benefit of this approach.

The event anomaly detection, visualization and analytic reasoning approaches are designed based on the theories and principles from geovisual analytics literature. It is expected that these approaches will support analysis of real world event anomalies which will answer the research questions asked in Chapter 1. To validate these answers, a set of field trials with professional data analysts using real world data were conducted, which will be explained in detail in Chapter 5.

Chapter 4

Implementation and Case Study

The geovisual analytics components described in the previous chapter have been implemented within a prototype system in conjunction with other common geovisualization techniques. While the previous chapter explained these components in general, this chapter will explain how they have been implemented for a specific data sets. To illustrate a real world example, this prototype uses two data sets related to fisheries enforcement: VMS (vessel monitoring system) data as movement data, and MarFis (fishing event location data) as geospatial point data. Both of these data sets describe the same daily fishing event locations of individual vessels. Positional discrepancies among the fishing event locations of same fishing event recorded in these data sets are called fishing event anomalies, which are to be analyzed using the prototype system. The details of the data sets and different aspects of analyzing fishing event anomalies are described at the beginning of this chapter. Next, the prototype implementation is described, followed by a case study to illustrate how the prototype system supports exploration and analytic reasoning about these fishing event anomalies.

4.1 Fisheries Data Sets

The data sets used in this prototype were collected from inshore scallop fleet fisheries off the coast of Nova Scotia, Canada over the two year period of Jan 1, 2008 to Dec 31, 2009. Commercial fishing vessels are required to be equipped with VMS devices to comply with the licensing conditions. These devices record the GPS locations of vessels on hourly basis, which constitute the VMS data. The attributes of this data set that were used within the prototype system are shown in Table 4.1.

Commercial fishing vessels often perform multi-day fishing trips. Sometimes vessels execute fishing activities in multiple locations on the same day. Upon returning to the port, these vessels must report where they fished each day, as well as the approximated amount of fish caught (measured in kilograms) in each location. Each of these fishing locations is recorded by a geographical point data (latitude and longitude). Although these readings provide precise geographical locations, the fishing events actually take place over a region on the ocean. At the same time, some fishers rounded the reading to the nearest minute on their report. When fishers performed fishing in multiple locations on the same day, multiple reports are made for that day. The reports are logged in the MarFis database system. The attributes of this data set that used within the prototype system are shown in Table 4.2.

The VMS data contains a total of 744,461 data points collected from 209 fishing vessels

Attribute	**Description**
LON	Longitude
LAT	Latitude
VRN	Vessel identification number
VMSDATE	Date of the data point
VMSTIME	Timestamp of the data point

Table 4.1: List of VMS data set attributes used in this research.

Attribute	Description
VRN	Vessel ID
DATE_FISHED	Fishing Date
LATITUDE	Latitude value of fishing location
LONGITUDE	Longitude value of fishing location
EST_WEIGHT	Estimated catch weight on the fishing date.

Table 4.2: List of MarFis data set attributes used in this research.

over the two years period, which indicates an average of 74.2 days at sea per vessel per year. The MarFis data contains a total of 18,030 fishing events performed by the same 209 commercial fishing vessels, which means 43.1 fishing events per vessel per year. Thus, there is a discrepancy found between the average days of fishing scallop and the average days at sea. Since the scallop fishing season is limited at Canada, the vessels obtain licenses to fish multiple species throughout the year. The VMS data contains data about vessel movements and the MarFis data contains data related to the scallop fishing only, which explain this discrepancy.

4.2 Fishing Event Anomalies

The fishing event anomalies are the positional discrepancies regarding the same fishing locations found in the above described data sets. For a given fishing event, the movement data found in the VMS data should show movement activities of that vessel for an appropriate length of time around the location reported in the MarFis data for that day. Some events are found where VMS data shows movement in a region which is significantly distant from the fishing location reported in the MarFis data. In some other cases movement activities are found around the reported locations but not long enough to explain the reported catch amounts. Both of these events are considered as event anomalies.

The reason for having such anomalies may be as simple as human error. The data entry operator or the fishers themselves may recorded incorrect readings while reporting the locations. It may also be caused by instrumental or communication errors, which caused VMS data points missing for a few hours. At the same time the underlying reason may be more severe than these, such as intentional reporting of incorrect locations or turning off the VMS devices to hide the vessel locations when performing illegal activities. Thus, the analysis of such anomalies may provide new and important insights for not only understanding and correcting problems in the data but also for fisheries enforcement.

Although the geographical points in the MarFis data are the representative of the fishing regions, these data points neither contain an estimation of the sizes of the fishing regions nor the duration of the fishing sessions. The fishing region sizes and session duration vary based on several factors associated with the fishing events, such as the location in that ocean, time of the year, type of the vessel, and the abundance of fish in the region. Because of these variabilities and lack of information, the event anomalies cannot be analyzed with automatic methods. The analysis requires incorporation of spatial, temporal, and domain specific contexts along with the analysts' prior knowledge and experience about the data. The remainder of this chapter explains the development of a geovisual analytics prototype system for analyzing these fishing event anomalies and a case study illustrating how these fishing event anomalies can be analyzed using this prototype system.

4.3 Prototype Implementation

4.3.1 Platform

The prototype system for analyzing fishing event anomalies was developed as a standard Java desktop application on the Java Development Kit 7 [58]. This prototype system also has been using a Java version of NASA World Wind [56], an open source three-dimensional mapping SDK. This mapping SDK supports the standard geospatial data visualization and view manipulation operations available in most 3D mapping software, such as pan, tilt, and zoom. All other elements of this system were implemented in custom software written by the author of this thesis.

NASA World Wind also contains the latest landscape information in great detail, which is also regularly updated by NASA [69]. In addition, being a Java SDK, the application was compatible in all environments. Thus, the expectation was that during the real world use of the prototype, the latest contextual information will be found within its underlying map, and the prototype will run on the analysts' native environments that will provide a better user experience.

4.3.2 System Architecture

In this prototype system, the geovisual analytics components described in the previous chapter were put together. The visual components of this prototype are: temporal filter (*Fishing Period*), geovisualization (*Map View*), tree visualization (*Event Tree*), and filtering tools (*Anomaly Threshold, Ancillary Data Filter, Show/Hide* controller). The prototype interface is shown in Figure 4.1.

The visual components are connected to the four main modules: event extraction, filtering, anomaly detection, and visualization. Figure 4.2 shows the system architecture diagram illustrating how these modules and visual components work together

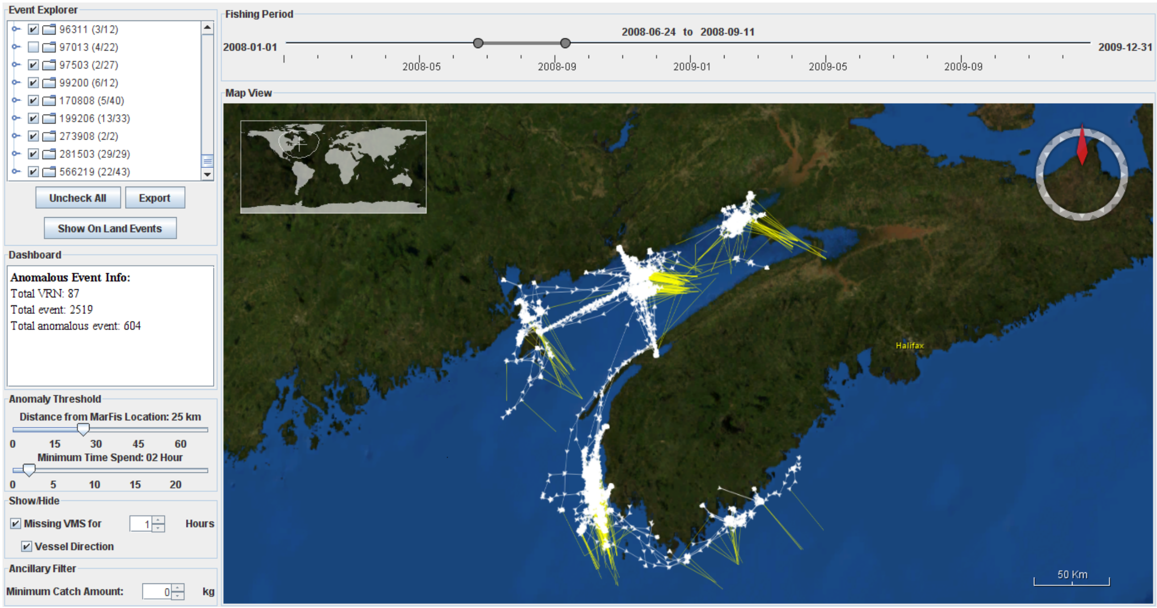


Figure 4.1: The prototype interface of the geovisual analytics system for analyzing event anomalies. The visual interfaces are: temporal filter at top-right, geovisualization at bottom-right, event tree at top-left, and filtering tools at bottom-left.

with the data sets. Colour encodings are used in the diagram for representing different types of components: data sets are shown in red; modules are shown in green; visualization components are shown in blue; and interactive filter controllers are shown in cyan.

4.3.2.1 Event Extraction

The event extraction module reads the data sets and extracts events by following the procedure described in Section 3.1. Once this module identifies events, it also finds the closest movement data points from the event locations within that event. Further, this module identifies the missing data points and interpolates them to make the data sets consistent. It also flags these missing data points, so that other modules can distinguish them. All these events are stored in groups where events from the same vessel are stored in the same group.

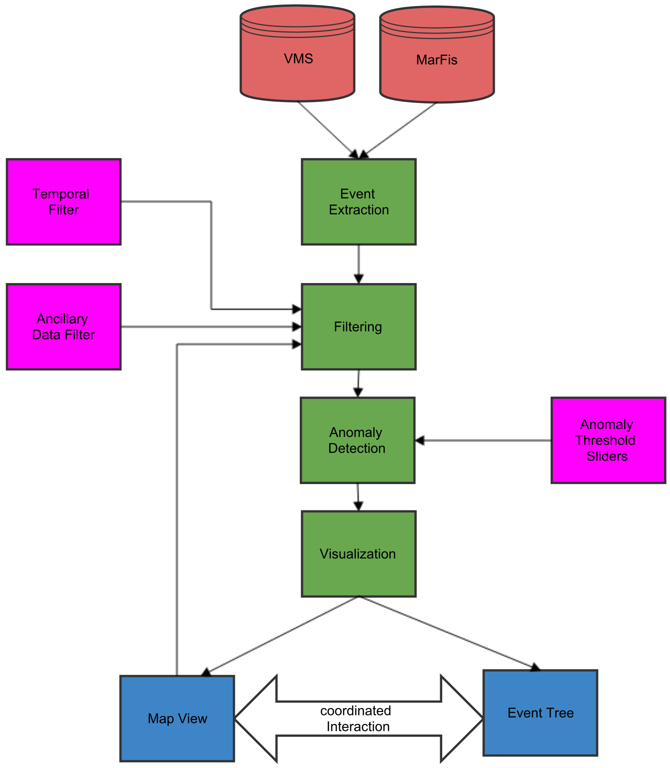


Figure 4.2: The system architecture of the event anomaly analysis system.

4.3.2.2 Filtering

The filtering module filters events based on the filter parameters. Initially, this module filters based on the default settings. During the analysis, analysts modify these parameters using the visual components (Temporal Filter, Map View, and Ancillary Data Filter). In both cases, events are checked for their spatial extent, temporal extent, and ancillary data value for filtering. Events that are within the current focus in each of these settings are marked using a hash table. This hash table is particularly useful for efficient searching in very large data sets, allowing not to have multiple copies of events in the memory, and increasing other modules' event searching efficiency [18].

4.3.2.3 Anomaly Detection

The event anomaly detection module identifies event anomalies using the threshold values. It reads the hash table generated by the filtering module and applies the anomaly detection procedure (see Section 3.2) to all the events from that table. Thus, the event detection procedure is applied to a limited number of events, which are currently under the analysts' focus. This approach increases the computational efficiency of the system. One disadvantage of this approach is this module is required to run every time after any of the filtering parameters are changed. However, careful implementation can avoid calculating event anomalies for those events that are already calculated. This module creates another hash table that lists the events that are found having anomalies. While the event detection procedure is directly related to the threshold values, this module works with the *Anomaly Threshold* sliders.

4.3.2.4 Visualization

The visualization module reads the hash table generated by the anomaly detection module, and represents them in both of the visualizations described earlier in the Section Map Representation (Section 3.3.1) and Tree Representation (Section 3.3.2). When the event anomalies are displayed on the map, this module considers the analysts' preferences regarding showing the direction glyphs, missing data points, and selection of event anomalies. Colour themes are also applied to the visual variables from this module. In Section 3.3.1, the importance of choosing the right colours was described, which explains that the colours for representing event anomalies on the map are directly related with the geographical extent of the data sets. The colour scheme for a specific data may not work for the another. For these particular data sets, events are performed at ocean. On the map, the ocean is represented with blue, which influenced the choices of the colours for representing event anomalies. Thus, the events are represented on the map using the colour scheme given in Table 4.3. Event anomalies can also be highlighted as described in Section 3.5, which stand them out from others event anomalies. Additional data is also shown for the highlighted events. As a result, a different colour scheme is required for representing both highlighted and non-highlighted event anomalies. Table 4.4 shows the colour scheme used in this prototype when the map view is showing highlighted event anomalies. Both of these colour schemes are chosen based on the Opponent Process Theory of Colour [30] and advice from the ColorBrewer [15].

This visualization module also shows event anomalies in the *Event Tree* that includes additional statistical information calculated by this module. The data are calculated using the two hash tables generated by the filtering module and anomaly detection module. Thus, these values are calculated after any changes in the filtering parameters or anomaly threshold values.

4.3.3 Work Flow

This prototype first reads a configuration file, which contains the initial parameters, then it initializes the visual components based on these values. Next, the prototype initiates the event extraction module. As described earlier in this chapter, this event extraction module extracts events from the data sets. Once all the events are extracted, this module initiates the filtering module. The filtering module identifies

Component	**Colour**	**Intensity**
Movement path	White	Normal
Chevron	White	Normal
Empty circle	White	Normal
Anomaly	Yellow	Normal

Table 4.3: Colour scheme for visual variables to show event anomalies when no event is selected.

Component	**Colour**	**Intensity**
Selected event anomalies		
Movement path	Red	High
Chevron	White	High
Empty circle	Cyan	High
Anomaly	Yellow	High
Anomalies from the vessels for which at least one anomaly is selected		
Movement path	Red	Normal
Chevron	White	Normal
Empty circle	Cyan	Normal
Anomaly	Yellow	Normal
Movement path (Path of the events that do not have an anomaly)	White	Normal
Event from other vessels		
Movement path	White	Low
Chevron	White	Low
Empty circle	White	Low
Anomaly	Yellow	Low

Table 4.4: Colour scheme for visual variable when one or more events are selected. Three different types of events exist in this case: selected events, events from the same vessel for which at least one event anomaly is selected, and event anomalies from the other vessels.

events that are within the initial focus as given in the configuration file. After this, the anomaly detection module is initiated that detects the anomalies based on the default anomaly threshold values from these identified events. Finally, the visualization module is initiated that represents the event anomalies in the *Map View* and *Event Tree*.

After completing the above steps, the prototype becomes ready for analysts to explore the data sets. A sequence of actions takes place after each of the user interactions. For instance, when analysts change the temporal extent, a sequence of operations is initiated. This operation starts with the filtering module that examines all the events and identifies those that are within the new temporal range. Next, the event anomaly detection module initiates that examines the events identified by the filtering modules to find event anomalies based on the current setting of anomaly threshold values. Finally, the visualization module updates the *Map View* and *Event Tree* to show the events anomalies.

Another series of operations are performed when analysts zoom or pan the map, or change the ancillary filter value. These operations initiate the filtering module, which identifies all the events that are within the selected spatial extents and has ancillary data values larger than the selected value. Next, the anomaly detection module finds event anomalies in these events. Finally, the *Map View* and *Event Tree* are updated by the visualization module to display the event anomalies.

When anomaly threshold values are changed another sequence of operations are performed. In this sequence, all the events that previously listed by the filtering module are examined by the event detection module. This is because the event anomalies determined earlier may not be anomalies anymore with the new threshold values (or the other way around). Once the anomalies are detected the visualization module updates the *Map View* and *Event Tree* to display all the anomalies.

4.4 Case Study

This section will show a case study illustrating the benefits of the geovisual analytics components explained in Chapter 3. In this case study, the daily fishing events are extracted from the VMS and MarFis data sets. Then, the event anomalies are detected and represented in the visualizations. Analysts then explore the anomalies for making sense of the data.

The current practice among fisheries experts for analyzing fishing event anomalies is to independently plot these data sets on a geographical map, then manually inspect them for anomalies. This approach is too tedious to apply to data sets that cover a large geographical area and contain data collected from many vessels over a long period of time. Thus, with this approach analysts have limited ability. Currently, a subset of the fishing events are selected for analyzing by using predefined checks, such as vessels that travelled close to protected areas, vessels that reported an exceptionally large catch amount, repeat offenders, or random samples. Although a smaller subset is selected, the analysis process requires focus and attention to manually link the data sets. Even using modern GIS software tools that provides multiple views with interactive layers does little to reduce this cognitive load. Thus, this type of analysis is generally done when there is a clear indication of significant anomalies. This method currently is used more for validating known facts instead of discovering new knowledge or insights from the data sets.

This case study will focus on exploring fishing event anomalies and discovering new knowledge from them. This case study starts with the full spatial and temporal extents of the data sets, and default anomaly threshold values (40 kilometers for distance and 5 hours for time). These threshold values indicate that for a given event, the VMS data should show 5 hours of movement activity within 40 kilometers of the corresponding fishing location reported in MarFis data. Fishing events that do not fulfill these conditions are considered to be anomalies, and will be shown on the visualization for further exploration and investigation. The assumption for defining these threshold values is that they constitute a generous definition of event anomalies in the context of scallop fishing. Using these initial settings, 5,241 events are detected as potential anomalies and represented on the visualizations.

Showing these number of anomalies on the *Map View* results in a significant amount of visual clutter. The tree representation of these anomalies is also too large to analyze. However, the *Map View* provides a high level overview about these anomalies (see Figure 4.3). For example, a significantly high number of events reported their fishing

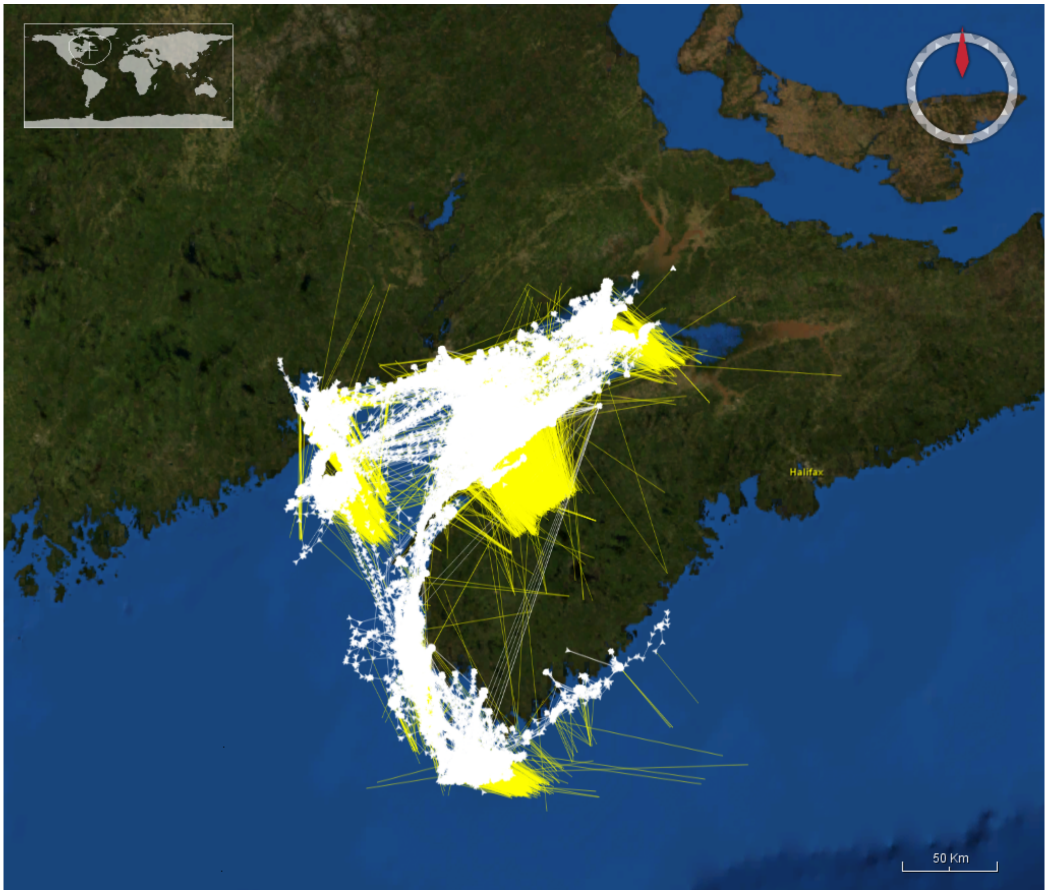


Figure 4.3: All anomalies within a two-year period.

locations on land. While this type of anomaly could easily be found by plotting the MarFis data on a map, it may not be noticed before due to the current practice of inspecting a subset of the data.

Analyzing the fishing events that have fishing locations reported on land may be a result of data entry or processing errors, and can easily be identified and investigated. These events can be hidden to see more interesting event anomalies. A button is provided at the bottom on the event tree, which toggles the visibility of this type of event anomalies. At this point, analysts may hide those event anomalies that reported on land and will proceed to analyze the remaining anomalies. The results of removing such event anomalies is shown in Figure 4.4

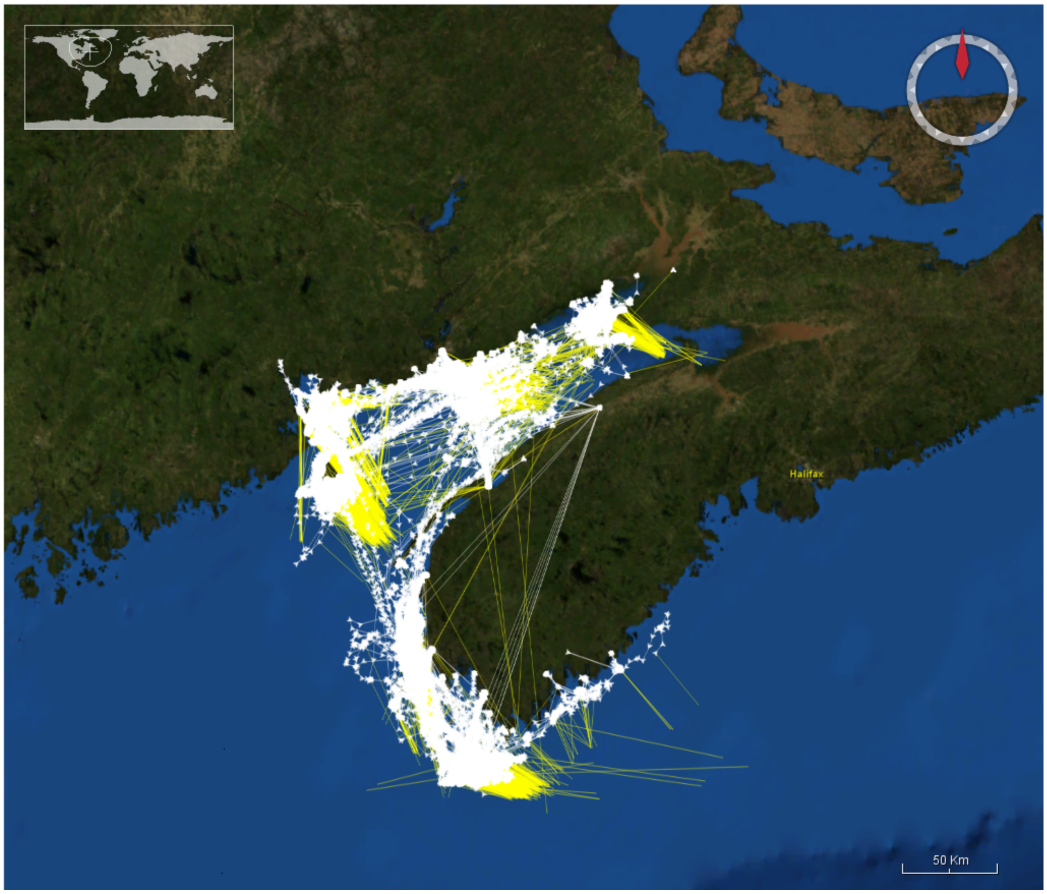


Figure 4.4: All anomalies within a two-year period excluding the events on land.

In order to further explore among the anomalies, the analysts may choose to focus on data from a particular time period, for example the month of September, 2008. After changing the temporal range, 216 fishing events are found, where 46 of them are potential event anomalies. Figure 4.5 shows these anomalies on the *Map View*. To reduce some of the visual clutter, the direction glyphs are kept hidden, as they are not adding much value at this level of detail.

Based on the analysts' knowledge of fishing practices within the selected temporal range, fishing vessels are required to be closer to the reported fishing locations and perform the fishing events for shorter period of time. Thus, the threshold parameters are changed to 25 kilometers for distance and 2 hours for time. The *Ancillary Filter*

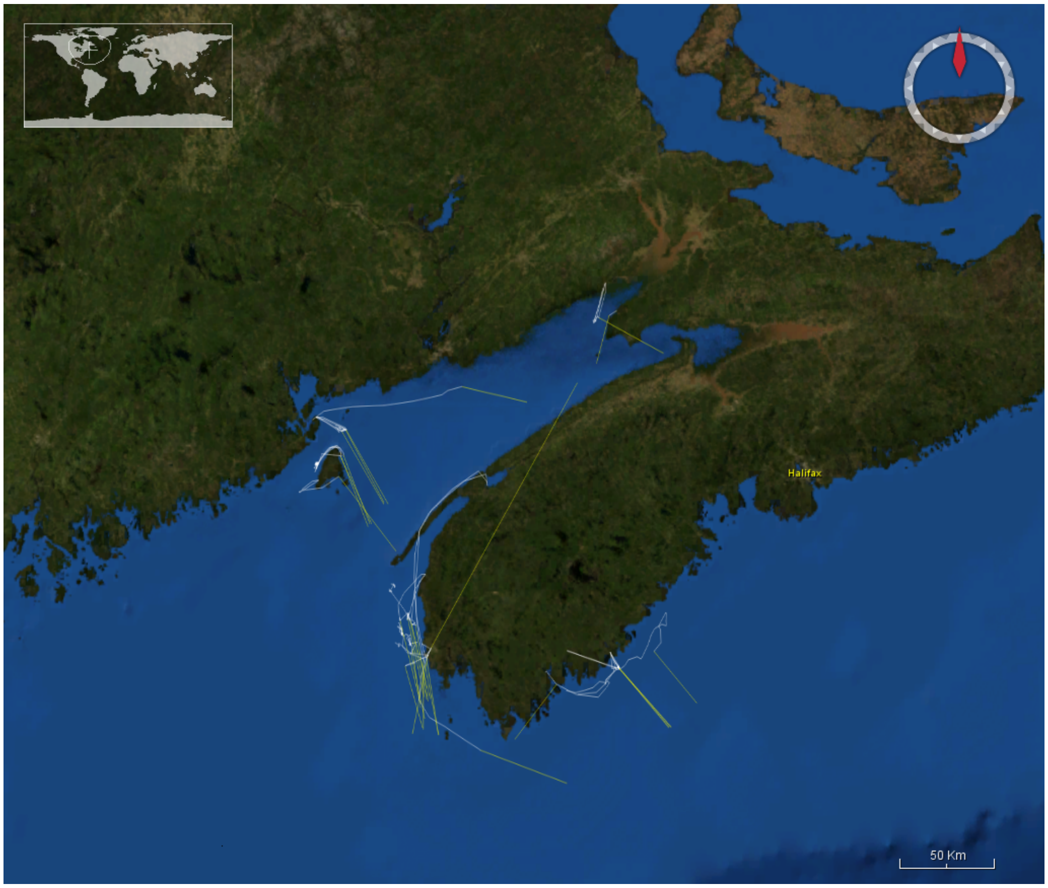


Figure 4.5: All anomalies within a one-month period.

value is also changed to 300 kilograms to remove fishing events with lower catch amounts. These new settings found 141 fishing events, including 32 potential event anomalies. Figure 4.6 shows the map representation of event anomalies based on these settings. From this view three separate geographical regions can be readily identified in which anomalies are present: the northern Bay region, the central Bay region and the southern region.

Noting the larger number of anomalies in the southern region, applying the pan and zoom operations on the *Map View*, analysts can focus on this region. This filter leaves 25 event anomalies for further exploration. The direction glyphs can be turned on at this point, since this direction information is important for the detail analysis.

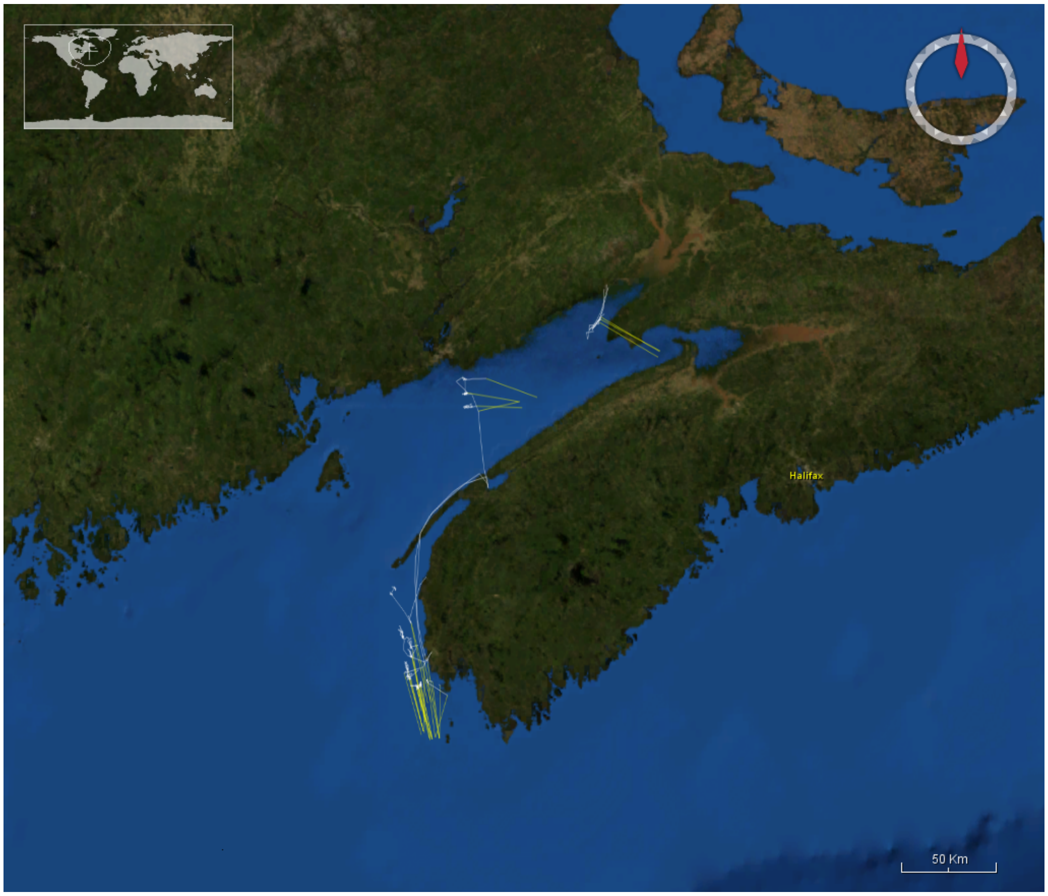


Figure 4.6: Defining anomalies more strictly further reduces the number of anomalies.

Figure 4.7 shows these event anomalies in the *Map View*, where four clusters can be easily identified. A pattern of reporting the fishing locations to further southern regions is also noticed from this view.

In order to support analysts' understanding of these anomalies, detail inspection and evaluation of individual vessels are required. This can be done by interactively selecting one of these event anomalies. Two event anomalies are selected from the *Map View* that are of particular interest. Upon making such selections, the anomalies from the other vessels in the region are dimmed, the selected anomalies and their vessel paths are highlighted, and contextual information regarding the paths of the vessels before and after the anomalous event are included in the display. More insight can also be

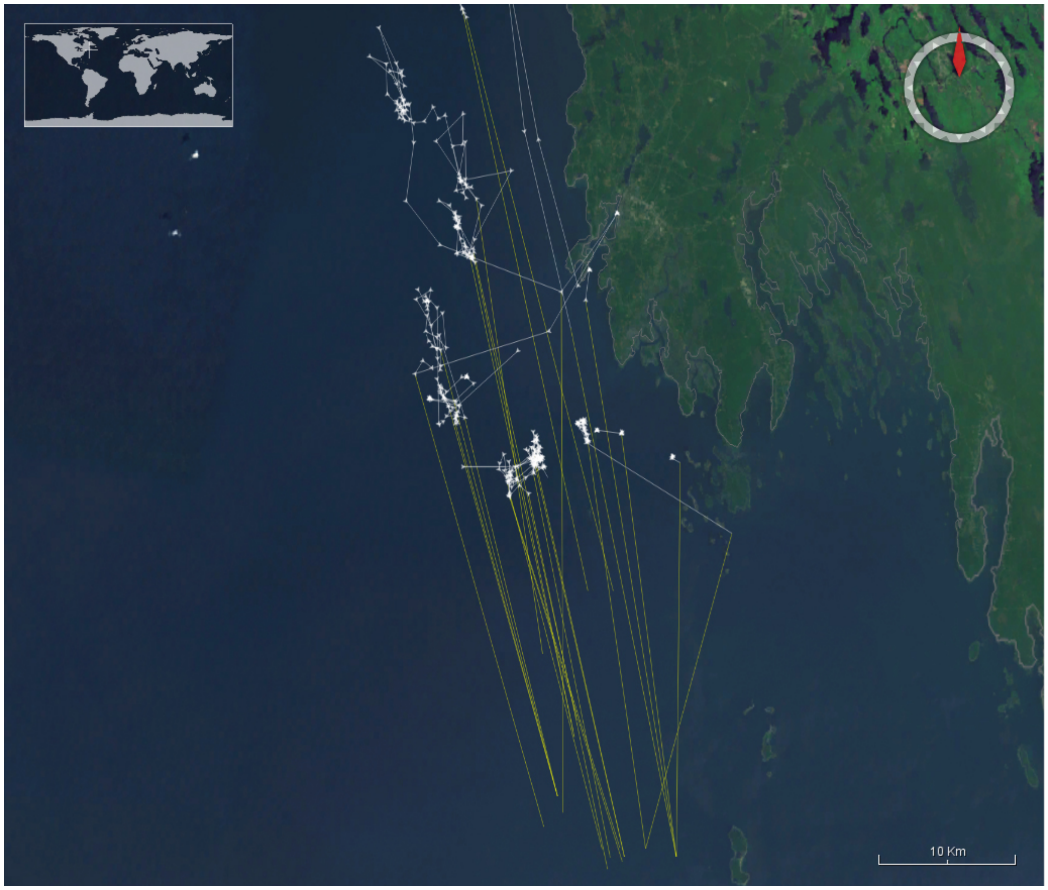


Figure 4.7: Zooming into a spatial region allows for more detailed analysis of the vessels.

provided by turning on the highlighting of missing data points (see Figure 4.8).

At this view analysts can infer specific activities of the highlighted vessels. For example, the highlighted vessel on the right reported the same location for four of its fishing events, but the vessel was not close to that reported location anytime during this month. Moreover, this vessel's VMS system was not responding for 11 consecutive hours within this period, thus no data are available regarding the location of the vessel for this period of time. An equipment failure can explain these missing data points, however, it also could be an intentional equipment sabotage for disguising illegal activities.

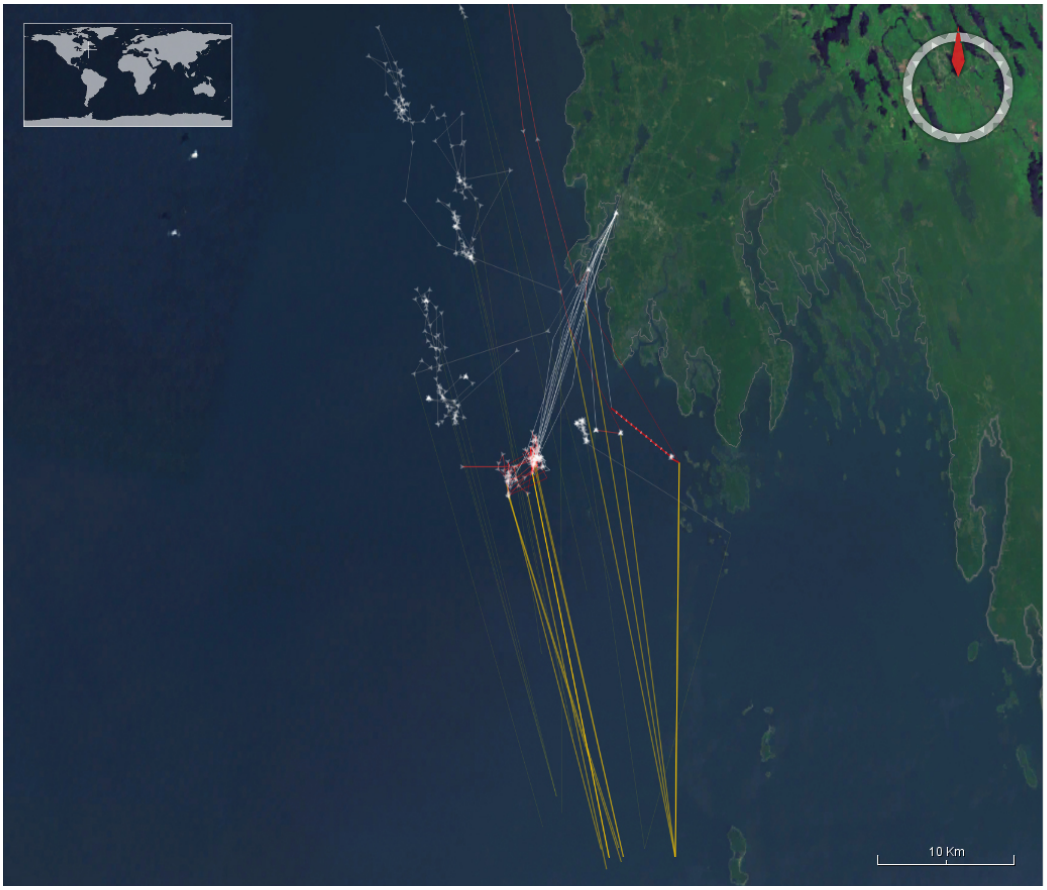


Figure 4.8: Highlighting vessels allows for the comparison of anomaly patterns.

The other selected vessel (on the left) performed fishing activities close to the port and made several back and forth trips between the fishing locations and the port. The movement pattern indicates fishing activities, however, these events were never reported correctly, The reported locations are apparently double in distance from the port than the actual distance. The reported locations also slightly vary from the other selected vessel's reported location.

The analysis of these two vessels allows analysts to discover some of the insights of their activities and provides evidence for further investigation. The context regarding these vessels' fishing activities within a wider temporal range may be analyzed. Thus, the analyst may wish to broaden the temporal range to six months, zoom out to explore where else the vessels have been travelling, and show even those event anomalies that are on land. In this view, analysts may also hide all other vessels' activities from the map (see Figure 4.9), which reduces the visual clutter and avoid misinterpretations. From this setting, these highlighted vessels are found reporting their activities correctly in the bay regions, however, they did not do the same when they fished in the southern region. This may be another indication to introduce more monitoring of the southern region.

This case study shows that the interactive features of this prototype system support exploration of the anomalies and allow analysts to perform analytic reasoning about the underlying behaviour that caused the anomalies. The highly interactive features

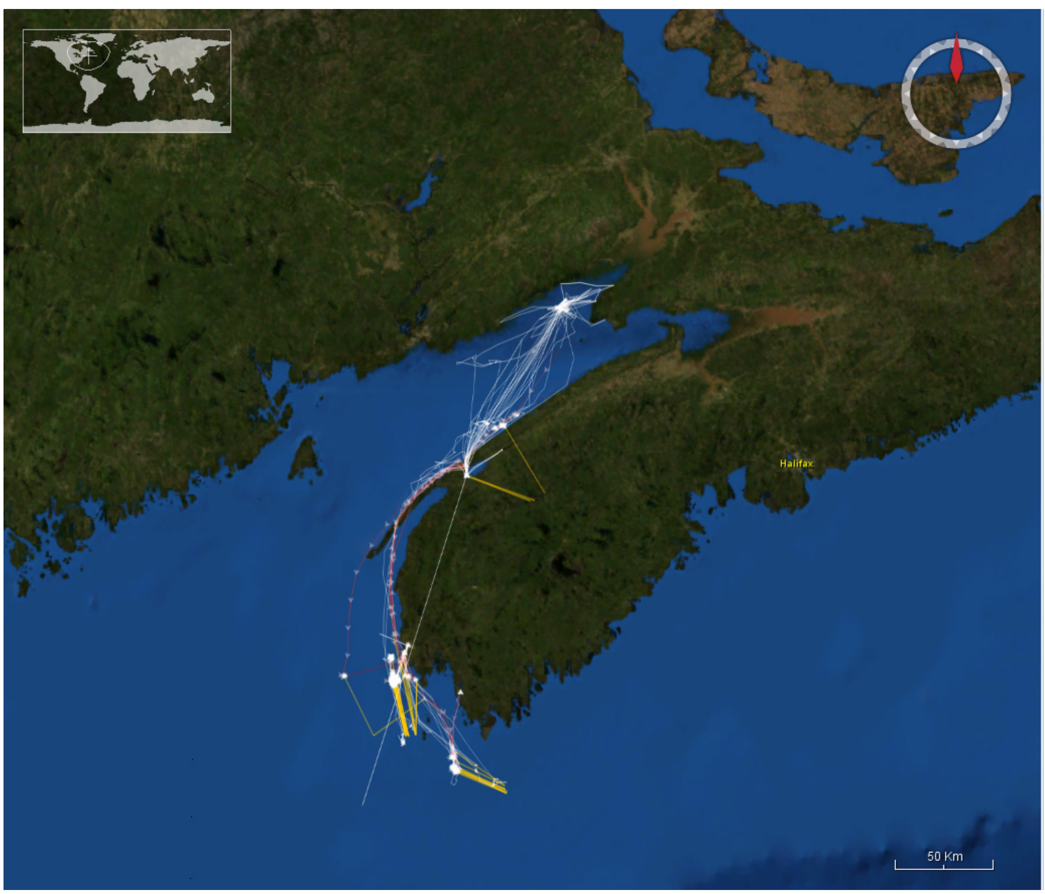


Figure 4.9: Zooming out geographically and temporally provides more context to the anomalies.

of this prototype allow analysts to easily focus on the spatial and temporal extents of interest. Furthermore, anomalies from the same or different vessels can be compared within their geographic context. This reveals the patterns of the anomalies which are readily interpreted by the analysts based on their existing knowledge about the domain. In addition to verifying known facts of the data, new knowledge can also be discovered from the data sets using the geovisual analytics approach presented in this prototype system.

4.5 Discussion

This chapter has shown an example of how the proposed geovisual analytics components presented in Chapter 3 can be implemented for real world data analysis tasks. The real world data sets associated with daily fishing events were taken as an example for this implementation. The implementation details were described. A case study was also presented to show example data analysis. The entire chapter was aimed to provide details of implementation so that the geovisual analytics components can be implemented for data sets from other domains.

This work was intended to identify and analyze event anomalies from movement data and geographical point data where both of these data sets explain the same events. In this regard, data sets related to daily fishing events were an ideal candidate for the case study. The types of anomalies found in the data sets are caused by many different reasons, ranging from systematic data quality problems to intentional misreporting. Thus, the case study validates the benefits of the approach.

The guideline for choosing colours of the visual variables was discussed in the previous chapter. The specific colours for representing fishing events explained in this chapter show an example of implementing the guideline. Since choosing the colours is strictly dependent on the data set, the colour scheme used in this chapter may work on other oceanic data sets. For other types of data, different colour schemes may need to be chosen.

Finally, the case study provided in this chapter shows how this geovisual analytic system helps analysts to explore, analyze, and make sense of the data. The system was designed inspired by Keim et al. proposed mantra [40] for visual analytics (an extension of Shneiderman's information seeking mantra [66]). Following the mantra, analysts started with viewing all of data and then used a series of interactive actions to reveal insights of it. During the case study, a large number of events were noticed that reported their fishing location on land, which may be known to the analysts. Further, analysts were able to iteratively filter uninteresting events and identify two interesting vessels, which are investigated further with their domain specific knowledge. This also reveals a weakness in the monitoring practices. Thus, this case study shows the benefits of using the system. While this is an example for showing the benefits of the prototype system, a comprehensive user evaluation using field trials is shown in next chapter, which confirms the value of this system in a real world setting.

Chapter 5

Evaluation

5.1 Purpose

The evaluation was performed for the developed prototype system for understanding the extents of the intended tasks the system supports as seen from analysts' eyes, and to probe for requirements and needs [45]. The evaluation also focuses on testing the design decisions and evaluating the analysts' performance on the prototype system. Therefore, the evaluation process shown in this Chapter will address the research questions related to the usefulness, ease of use, and enhancement of analysts' ability for analyzing event anomalies. This evaluation process was guided by a set of hypotheses related to the system's usability and usefulness. A Field trial method was used for this evaluation where the data were collected using questionnaires, interviews, and investigator's observations.

5.2 Hypotheses

Due to the absence of any baseline system, a valid measurement of the benefits and limitations of geovisual analytics system for analyzing event anomalies was challenging. To understand the user acceptance of a system like this, different theories are available, such as Theory of Planned Behaviour (TPB) [3], Technology Acceptance Model (TAM) [20], and Unified Theory of Acceptance and Use of Technology (UTAUT) [78]. Among them, TAM is a motivational model that traces the effects of system design characteristics on users' intentions to use the system through perceived usefulness and perceived ease of use. TAM also provides the basis for a practical and effective user acceptance testing methodology for predicting the degree of user acceptance of a new system [21]. TAM2 [77] is an extension of TAM, which is also can be used for predicting the degree of user acceptance of a new system. In addition, provides additional elements of user acceptance of a system, such as the determinants of perceived usefulness. The objective of this evaluation was to understand the degree of user acceptance of this system. Thus, TAM was used to collect structured data about participants' acceptance of the system. Beside TAM, interviews and observations were also used in this evaluation procedure. The entire evaluation procedure was guided by the following hypotheses:

Hypothesis 1: *Using yellow lines to connect the movement paths and event locations are useful for visually representing the event anomalies.*

The visual variable for representing event anomalies uses lines, flow lines, and dots. Flow lines are used to represent the movement paths and dots are used to represent the event locations. Yellow lines are used to connect these two objects. This design decision was made based on the Gestalt Law [42] and Opponent Process Theory of Colour [53]. The yellow colour for these lines was chosen because of the chromatic contrast with blue (the colour used on the map for representing the ocean). The lengths of these lines are longer for larger positional discrepancies. Thus, the severe anomalies and clusters of anomalies are expected to be pre-attentively identified because of the length of the line and chromatic difference of colours. Therefore, it was expected that analysts will find the yellow lines useful for representing event anomalies.

Hypothesis 2: *The use of red empty circles for visually representing missing movement data points is useful when analyzing event anomalies.*

The proposed visual variable uses a different symbol shown in separate colour for representing the missing data points: red empty circle for missing data points and white dot for others. This approach visually separates the missing data points from the others. Since position and colours are the visual features that are pre-attentively processed [81], so the expectation is that these missing data points bring attention to the analysts instantly and effortlessly [65]. This quick identification and pre-attentive processing may be an important aspect of identifying interesting event anomalies. Thus, the expectation was that the use of these different symbols and colours for representing missing data points will be useful for the analysts.

Hypothesis 3: *The adjustable threshold values are useful in extracting interesting event anomalies.*

Complexities are associated in defining anomalies for the computational method to detect potential event anomalies because of the absence of the information regarding the exact locations and sizes of the fishing regions. The adjustable distance and time threshold values allow analysts to set a maximum allowable distance between movement data points and event locations, and the length of valid fishing sessions. This feature allows analysts to filter out event anomalies that are the result of mismatches in temporal scale or are within a reasonable distance from one another. With this feature analysts can adjust the threshold values at any time throughout their analysis session, which interactively extract anomalies and display them in the visualizations. The expectation was that this feature will be useful for the analysts.

Hypothesis 4: *The adjustable threshold values for extracting interesting event anomalies are easy to use.*

The threshold values are adjusted within the prototype system using standard slider controls. This feature provides an interactive visual query system for analysts to extract event anomalies based on different threshold values and presents them in the visualizations. The slider control is used with proper labeling to facilitate this feature. Since sliders are commonly used for increasing or decreasing values, with the proper labeling in it, the expectation was that this feature will be easy for the analysts to use.

Hypothesis 5: *Filtering event anomalies using the ancillary data is useful for analysts to focus on the event anomalies having certain ancillary data values.*

This ancillary information within spatio-temporal data sets plays important roles in the data analysis process. The ancillary data used in the prototype system is the estimated catch amount. Analysts may be interested in the fishing events that report higher catch amounts. Hence, filtering information on the basis of this ancillary data allows analysts to remove visual clutter by filtering out events that may be of less interest. Therefore, the design decision was made to provide an ancillary data filtering feature. The expectation was that the analysts will find this filtering feature useful for analyzing event anomalies that have higher interest to them.

Hypothesis 6: *Filtering event anomalies using the ancillary data is easy to use for analysts.*

The prototype considers the estimated catch amounts in kilograms, which are integer values, as the ancillary value associated with each event. By adjusting the filtering parameter for this value, all the events that do not exceed the chosen value are interactively removed from the visualizations. A standard spin control is used to implement this feature with appropriate labeling. This control allows analysts to type the desired value directly and then adjust the value with small steps (+/- 10 kilograms). This spin control is commonly used when values are varied within a large range. Thus, the expectation was that this feature will be found easier for the analysts to use.

Hypothesis 7: *Filtering event anomalies with the zoom and pan map operations is useful for analyzing the event anomalies from a specific geographical location.*

Spatio-temporal data sets may contain events from a wide spatial extent. However, in each step of an analysis session, the analysts may focus on event anomalies from a certain geographic region by applying zoom and pan operations on the map. Representing event anomalies only from the focused geographical region reduces information overload. Therefore, a spatial filtering feature was provided in the prototype system for filtering out event anomalies from outside of the focused region. Clutter removal is the basis of this hypothesis and the expectation was that the analysts will find these filtering options useful for their tasks.

Hypothesis 8: *Filtering event anomalies with the zoom and pan map operations for analyzing event anomalies from a specific geographical locations are easy to use.*

Zoom and pan map operations are the common features of geovisual analytics applications and commercial GIS tools [26] [27] [57] [73]. Professional data analysts are familiar with these zoom and pan map operations using mouse scroll and drag options. These basic zoom and pan map operations of NASA World Wind SDK were extended within the prototype system to perform the filtering of event anomalies interactively on the change of spatial extent of the map. Thus, the filtering operations work with analysts' native zooming and panning tasks, which they are already familiar with. This basis of these filtering features gave an expectation that the analysts will find the filtering of event anomalies on zoom and pan map operations easy to use.

Hypothesis 9: *The interactive highlighting of event anomalies from the Event Explorer is useful for detail investigations of event anomalies.*

The *Event Explorer* lists the events that are within the selected temporal and spatial range in a tree view (group by entity identifier) and displays ancillary data, event data, entity identifier, and comparison of event and event anomalies. Analysts can choose one or more events from the *Event Explorer* based on one of more of these additional information. This selection also highlights the selected events on the *Map View*, which shows their geographical contexts of the selected events. This gave an expectation that analysts will find the feature of highlighting event anomalies from the *Event Explorer* useful.

Hypothesis 10: *The interactive highlighting of event anomalies from Event Explorer is easy to use.*

In this prototype system, event anomalies are grouped by the vessel identification numbers in the *Event Explorer* within a tree structure in ascending order. Event anomalies associated with each vessel are also ordered by the event date and displayed in the lower level of the tree. This ordered representation may make searching for a specific vessel or event easy. Multiple event selection or deselection operations also follow the user interaction techniques of operating systems' file and directory selection operations. Analysts can select the first event by clicking on the events. For selecting additional events, analysts can hold control key and click on the events. The deselection can also be done in the same way. The expectation was that the analysts will find the event selection and highlighting task in *Event Explorer* easy.

Hypothesis 11: *The interactive highlighting of the event anomalies using the Map View is useful for detailed investigations of event anomalies.*

The *Map View* provides the spatial context of event anomalies. This context may instigate more detailed investigation. The prototype provides an option for analysts to select one or more event anomalies by clicking on anywhere of the visual variables that is representing those event anomalies in the *Map View*. This shows the entire path taken by that vessel within the selected temporal range and highlights all the event anomalies within that entire path. It also highlights the events and the entity identifiers on the *Event Explorer* to provide their ancillary information. Thus, analysts can see information about the entity identifier, their statistical information, and their movement before, after, and in between event anomalies. These may support analysts to understand the event anomalies and their behaviour. Therefore, the expectation was that analysts will find this feature useful.

Hypothesis 12: *The interactive highlighting of the event anomalies using Map View for detailed investigation is easy to use.*

The event anomaly selection process is developed by following the conventional methods of selecting items in a computer system. The analysts' mouse hovers on event anomalies and clicks to select them. Using the control keys in the keyboard they can also select multiple event anomalies. These event anomalies interactively highlight on both of the visualizations, which is a visual feedback for the analysts. All the events can also be unselected by right clicking on anywhere in the map other than events. Since the operations are aligned with the other selection procedure of the prototype system and following the conventional method of selecting items in a computer system, the expectation was that the analysts will find this feature easy to use.

Hypothesis 13: *Showing contextual data related to the highlighted event anomalies is useful for detail analysis of the entities of select event anomalies.*

This prototype allows analysts to extract interesting facts about event anomalies by analyzing them within their contexts. Once an event anomaly from a vessel is highlighted, the entire activities performed by that vessel within the selected temporal extent become visible to the analysts. Selecting events from multiple vessels also displays details for all of them. This extra information supports analysts to understand the vessels' activities. Thus, the expectation was that this feature will be useful for the analysts to extract the details of event anomalies and the correlations between them. Table 5.1 shows the list of all the hypotheses.

#	Hypothesis
1	Using yellow lines to connect the movement paths and event locations are useful for visually representing the event anomalies.
2	The use of red empty circles for visually representing missing movement data points is useful when analyzing event anomalies.
3	The adjustable threshold values are useful in extracting interesting event anomalies.
4	The adjustable threshold values for extracting interesting event anomalies are easy to use.
5	Filtering event anomalies using the ancillary data is useful for analysts to focus on the event anomalies having certain ancillary data values.
6	Filtering event anomalies using the ancillary data is easy to use for analysts.
7	Filtering event anomalies with the zoom and pan map operations is useful for analyzing the event anomalies from a specific geographical location.
8	Filtering event anomalies with the zoom and pan map operations for analyzing event anomalies from a specific geographical locations are easy to use.
9	The interactive highlighting of event anomalies from the Event Explorer is useful for detail investigations of event anomalies.
10	The interactive highlighting of event anomalies from Event Explorer is easy to use.
11	The interactive highlighting of the event anomalies using the Map View is useful for detailed investigations of event anomalies.
12	The interactive highlighting of the event anomalies using Map View for detailed investigation is easy to use.
13	Showing contextual data related to the highlighted event anomalies is useful for detail analysis of the entities of select event anomalies.

Table 5.1: List of all hypotheses.

5.3 Field Trial Methodology

The data analysis tasks that are supported by this prototype system are complex in nature. The availability of professional data analysts for this domain is also limited. In addition to these issues, no baseline data analysis system is available against which to compare. However, the evaluation of a geovisual analytics system conducted with the real world data sets, domain specific tasks, and domain experts as participants provide concrete and realistic evidence of effectiveness [17]. Thus, to evaluate a system like this, a field trial methodology is often used in the visual analytics research [17] [62], which evaluates the system under real world conditions.

In the field trials, professional data analysts directly use the system with real world data sets to conduct analysis tasks of their own choosing and provide informed opinions about the system. The data used in this field trial was the same fisheries data sets described in Section 4.1. The field trials were performed at Fisheries and Oceans Canada, Halifax, NS. During a field trial, three types of data are collected: investigator observation, survey, and interviews. Analyzing these data provides the understanding of the system's usefulness, ease of use, and effectiveness.

Because of the limited number of participants and the open ended nature of tasks, field trials do not provide comparable quantitative data. However, field trials do provide insightful qualitative feedback from the participants. This feedback is more valuable than the statistical analysis of quantitative data collected over constrained tasks [17].

5.4 Study Design

Each field trial started with obtaining informed consent from the participant. The field trial was then divided into five phases: demographic investigation about the participant, training, system use, post-study questionnaire, and interview session.

In the demographic investigation phase, each participant was asked to complete a pre-study questionnaire. This questionnaire was designed to obtain the participant's knowledge about the domain, such as experience level with VMS and MarFis data, their familiarity with virtual globe and geovisual analytics systems, and the number of years they worked with VMS and MarFis data.

After completing the pre-study questionnaire the training session was conducted where each participant was instructed by the investigator about how to use the prototype system. The investigator showed different features of the system and explained how these features work. The investigator further asked the participants to perform some test tasks as a part of their training.

After the training, each participant was asked to explore the event anomalies with the prototype system. Using the same data sets each participant performed undirected and open-ended data analysis tasks based on their interests and experience. The investigator helped the participants to operate the software when required. This helped participants to perform their tasks with the software at a level beyond novice users. While participants were performing the analysis tasks there activities was video recorded, which were further investigated by the investigator.

After the system was evaluated sufficiently by the participants, a post-study questionnaire was administered using a survey instrument adapted from the TAM. In this survey, each participant was asked a set of questions regarding the perceived usefulness (PU) and perceived ease of use (PEU) for the features. Six questions were asked to measure the PU of a given feature. Six other questions also asked for measuring the PEU for a given feature. The questionnaire are given in Appendix A. A five-point Likert scale was used to measure their responses. Next, semi-structured interviews were conducted to gather the detailed opinions about the system from the participants. In the interview sessions, the same set of questions was asked to all the participants to measure their satisfaction, and understanding of the different features of this system.

5.5 Participants

The participants of these field trials were fisheries data analysts who had experience with the data generated by VMS units, the MarFis data, and the patterns and behaviours of fishing vessels conducting both legal and illegal activities. An invitation was sent to all the potential analysts and five participants (P1, P2, P3, P4, and P5) were voluntarily participated.

The pre-study questionnaire provided some insight into the composition of the participant pool for this study. Table 5.2 shows the participants demographics collected in the pre-study questionnaire. Four participants were very familiar with the anomaly

	P1	**P2**	**P3**	**P4**	**P5**
MarFis data analysis experience	4 years	6 years	6 years	8 years	2 years
VMS data analysis experience	4 years	6 years	6 years	4 years	1 year
Anomaly analysis experience	Very High	Very High	Very High	High	Medium
Familiarity with geovisualization system	High	Low	Very High	Medium	Medium
MarFis data visualization experience	High	Very High	Very High	Very High	High
VMS data visualization experience	Very High	Very High	Very High	High	Medium
Number of software used for MarFis data analysis	3	0	3	4	3
Number of software used for VMS data analysis	4	1	4	4	3

Table 5.2: Demographics of participants in field trials.

analysis among MarFis and VMS data and worked with these data sets from 4 to 6 years. The one participant who was not familiar with analysis of fishing event anomalies for a longer period of time, had medium level of experience in analyzing anomalies within these data sets.

The pre-study questionnaire also showed that participants had a moderate level of familiarity with geovisualization and virtual globes. All the participants were also familiar with different tools to accomplish their daily activities. Such tools included non geovisualization tools, such as SQL manager and Microsoft Excel; generic GIS tools, such as ArgGIS, Google Earth, and Map Info; and customized tools, such as Virtual Data Center Mapping Application (VDCMA) and R.

The prototype was developed for analyzing fishing event anomalies and the field trials were performed in the office of Fisheries and Oceans Canada, Halifax. The participants were professional fisheries data analysts. Therefore, the number of professional data analysts was limited for these field trials. In addition, the user evaluation was subjective in nature and the participants performed open ended data analysis tasks during the field trials. Since all of the participants were expert data analysts within the domain in which these field trials were conducted, their use of software is considered to be representative of the real world data analysis practices and their responses are considered to be reliable. Therefore, this sample size was considered large enough to ensure that the investigator would gain a full understanding of the variability in responses between expert participants [16] [84].

5.6 Data Analysis Methods

Three types of data were collected regarding the participants use of the software: post-study questionnaire, interviews, and investigator's observations. Different types of data analysis methods are used for analyzing these data.

The post-study questionnaire provides participants' perceptions from multiple perspectives about common underlying features, such as representing positional discrepancies using yellow lines and the use of threshold values for detecting event anomalies. Due to the participants' free and open ended use of the system, individuals' responses are analyzed separately. The percentages of each types of answer on the 5-point Likert scale from a given participants for a given feature are calculated. All such results for a given feature from all the participants are shown within a divergent stacked bar chart for visually depicting this result. Agreement is represented to the right, in light green for agreement and dark green for strong agreement; neutral is in the middle in grey; disagreement is to the left in red, noting that there were no responses that were of strong disagreement. The same approach is also taken for analyzing the responses of features' ease of use. Section 5.7.1 describes the results of features that were evaluated during the field trials.

Although, statistical analysis has been applied in many user evaluations for validating the hypotheses, this is not done in this thesis. Statistical analysis is useful for analyzing quantitative data. The field trials for this user evaluation were open ended in nature, where all the participants chose their own data analysis objectives and used the different features of the prototype with their own choices. Thus, the field trials were primarily subjective. The validations of the hypotheses (accepted or refuted) were also subjective in nature. Moreover, the number of participants was insufficient for applying statistical analysis. Thus, statistical analysis was not applied for analyzing responses of the post-study questionnaire.

Beside the post-study questionnaire, broad opinions about the system were collected with interviews. Their responses were audio recorded during the interview and transcribed afterwards. These responses are then coded based on the themes of their responses. Three main themes were found in their responses: positive responses, negative responses, and new features or improvements of the system, which are described in Section 5.7.2.

The field trials were also video recorded, which are further analyzed by the investigator. The post-study questionnaire and interview responses show the participants' understanding and the degree of acceptance of the system, assessing their activities during the system use session (from the video recording) provides further understanding of uses of the features for analyzing event anomalies. The investigator observed the activities which are grouped to identify common themes, will be discussed in Section 5.7.3.

5.7 Results

5.7.1 Post-Study Questionnaire

Within the post-study questionnaire eight sets of questions were provided to evaluate eight features of the prototype software. Among them, both perceived usefulness (PU) and perceived ease of use (PEU) were measured for five features: anomaly thresholds, ancillary data filter, spatial filter, event anomalies highlighting from map representation, and event anomalies highlighting from event explorer. For other three features (anomaly representation, missing data point representation, and showing contextual data on event highlighting) only the perceived usefulness (PU) was measured, since these are purely visual features, thus asking to evaluate the ease of use of these features was meaningless.

5.7.1.1 Anomaly Representation

The visual representation of the event anomalies (positional discrepancies) is one of the core features of this geovisual analytics system. This feature was designed to visually encode the severity of discrepancies and support the pre-attentive identification of the events with higher discrepancies or geographical regions having a higher number of event anomalies. Hypothesis 1 was about the perceived usefulness of this feature and the expectation was that this visual representation will be useful for the analysts.

Figure 5.1 represents the participants' responses on the statements regarding the perceived usefulness of this feature. In this figure, each horizontal bar is 100% wider and represents an individual participant's responses. Each of these Individual bars is partitioned by the respective participant's percentage of the agreeing and disagreeing level with the six statements regarding the perceived usefulness of this feature, and are marked with different colours as described earlier. The bars are also split on the neutral positions. The portions of the bars representing the percentage of disagreement or strongly disagreement are presented on the left of this neutral position and the portions representing the percentage of agreement or strongly agreement are presented on the right. The portion representing neutral responses are split equally on left and right of the neutral positions. In this figure, the chart shows that the participant P2 provided neutral responses to all the six statements (since the entire bar is gray) regarding the perceived usefulness of anomaly representation, and participant P3 agreed with all of those statements. Participant P1 agreed with one (16.67%) and strongly agreed with the rest of the statements (83.33%). Participant P4 agreed with four statements (66.67%) and strongly agreed with the other two (33.33%). Participant P5 provided a neutral response on one statement (16.67%), agreed with four (66.67%), and strongly agreed with the other one (16.67%).

Most of the participants saw the value of this representation for analyzing event

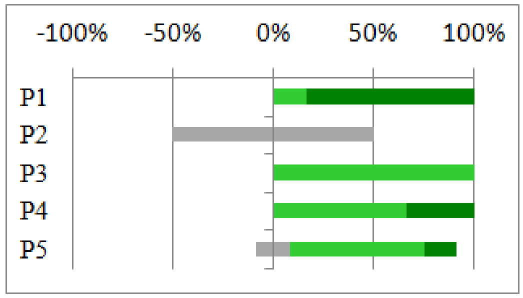


Figure 5.1: Usefulness of the visual representation of the event anomalies.

anomalies and no participant indicated that this feature was not useful, which can be considered as a positive finding. Thus, Hypothesis 1 was supported by the data.

5.7.1.2 Missing Data Point Representation

The missing movement data points of the fishing paths were interpolated to make the data sets consistent. These interpolated data points were encoded differently in the visual variable to avoid misinterpretation. In Hypothesis 2, the expectation was that this feature will be useful for analyzing event anomalies.

The participants' responses with the statements regarding the usefulness of representation of missing data points are shown in Figure 5.2. Mix responses were found for this feature. There were not many instances of the missing data points in the data sets. All of the participants were also not aware of specific events that have missing data points. Due to the open ended nature of their data analysis activities, some of

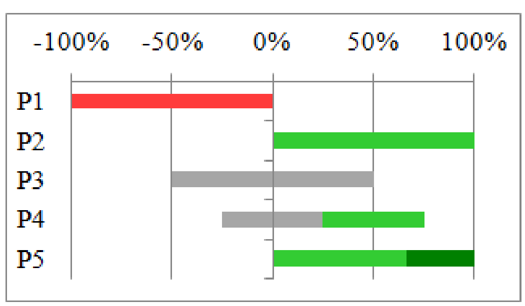


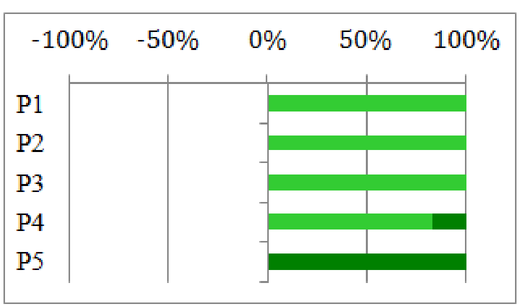
Figure 5.2: Usefulness of the visual representation of the missing data points.

the participants were able to find such events and they used this feature for analyzing them and others were not able to find them, which influenced these responses. Thus, some participants were (P2 and P5) agreeing and some other participant (p1) was disagreeing with the statements regarding the perceived usefulness of this feature. One of the participants (p3) showed natural responses. Therefore, Hypothesis 2 was neither supported nor rejected by the data.

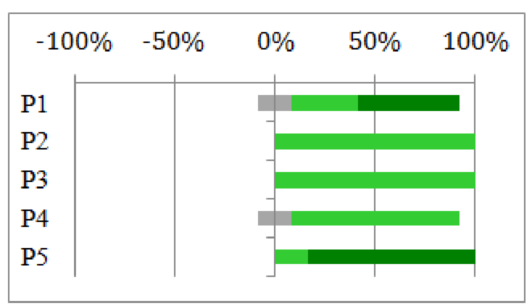
5.7.1.3 Anomaly Thresholds

Detecting event anomalies based on the analysts' experience, spatial context, and temporal context of the data sets was another key feature of this system. Adjustable threshold values were introduced for this purpose which allow analysts to guide the anomaly detection process in different settings of the system. This interactive feature was expected to be useful and easy to use for analysts in Hypothesis 3 and Hypothesis 4, respectively.

Figure 5.3a shows the responses with the statements regarding the perceived usefulness of this feature. All the participants found this feature useful. Their responses were leaned towards agreeing to strongly agreeing with the statements regarding the perceived usefulness of this feature. No participant responded with neutral or disagreeing in this regards. Therefore, Hypothesis 3 was supported by the data.



(a) Usefulness of anomaly threshold values.



(b) Ease of use of anomaly threshold values.

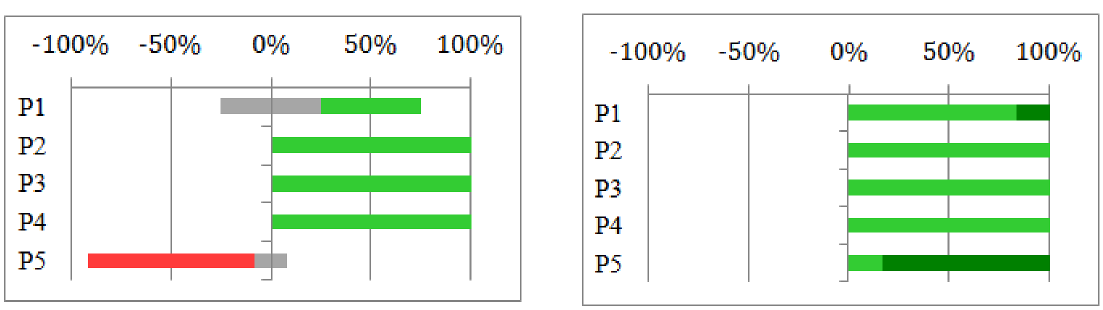
Figure 5.3: User acceptance of the anomaly threshold for detecting event anomalies.

Figure 5.3b shows the responses with the statements regarding the perceived ease of use of this feature. Participants also found the feature easy to use. Their responses in this case were also leaned towards agreeing to strongly agreeing with the statements regarding the perceived ease of use of this feature. No response of disagreement was provided by any of the participants. Thus, Hypothesis 4 was also supported by the data.

5.7.1.4 Ancillary Data Filter

To remove some of the visual clutter from the visualizations, *Ancillary Data* filter was provided to filter out event anomalies with lower catch amount. This was assumed that event anomalies with a lower catch amount were of lower interest to the analysts. The perceived usefulness and ease of use of this feature reported by the participants are represented in Figure 5.4.

The participants' responses with the statements of perceived usefulness regarding this filtering feature are represented in Figure 5.4a. Three participants (P2, P3, and P4) were agreeing, while one participant (P1) provided responses which leaned towards agreeing with the statements regarding the perceived usefulness of this feature. Participant P5 was pessimistic or reserved in his/her opinions. Therefore, this usefulness of this feature was found promising. As a result, the responses support Hypothesis 5.



(a) Usefulness of ancillary filter.

(b) Ease of use of ancillary filter.

Figure 5.4: User acceptance of the ancillary filter for analyzing event anomalies.

The participants' responses with the statements regarding perceived ease of use of this filtering feature are represented in Figure 5.4b. Although participant P5 was pessimistic in his/her opinion with the statements regarding usefulness of this feature, this participant was strongly agreeing about the ease of use of this feature. The other four participants were also agreeing with the same statements. Hypothesis 6 anticipated that the participants would find this filtering feature easy to use and no participant provided a neutral or disagreeing response. Therefore, the hypothesis was supported by the data.

5.7.1.5 Spatial Filter

The system filters out event anomalies that are outside the area of focus on the map. This spatial filter was added to remove some of the visual clutters from the visualizations. The participants' perceived usefulness and ease of use of this feature are shown in Figure 5.5.

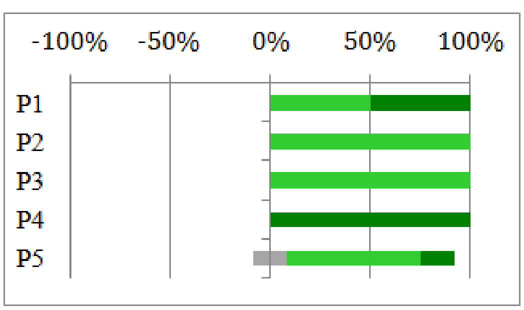
The participants' responses with the statements regarding the perceived usefulness of this feature are shown in Figure 5.5a. All the participants either agreed or strongly agreed with these statements. Hypothesis 7 anticipated that the filtering event anomalies with zoom and pan operation (change of spatial focus) show event anomalies only from the area of focus, which can be useful for the participants. Their responses in this case supported the hypothesis.

The participants' responses with the statements regarding the perceived ease of use of this spatial filter are shown in Figure 5.5b. The responses are distributed from disagreeing to strongly agreeing. The filtering was performed based on the reported fishing locations. Due to the open ended nature of the participants' data analysis activities, some participants were analyzing the reported fishing locations and found this feature easy to use. Some other participants were more interested with the movement path of the events and few of these events had larger degree of discrepancies than others. For such cases, zooming in to the movement paths were taking the reported fishing locations outside of spatial focus, hence the events were removed from the visualizations. Which negatively influenced the participants' responses with the statements regarding ease of use of this feature. Three participants (P2, P4, and P5) were agreeing or strongly agreeing with the statements regarding ease of use of this feature, while one participant (P3) was provided natural responses and the other participant(P1) was disagreeing with these statements. Therefore, Hypothesis 8 was neither supported nor rejected by the data.

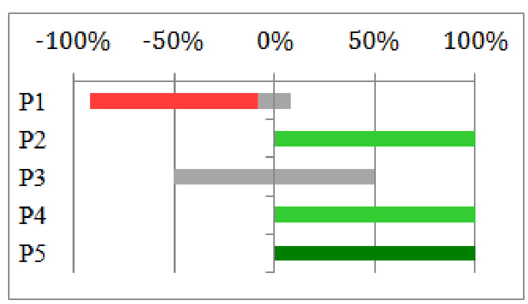
5.7.1.6 Highlighting Event Anomalies from the Tree Representation

Event anomalies can be highlighted based on their entity or ancillary data during the analysis sessions. In this system, event anomalies are represented in a tree view where analysts can interactively highlight them based on their vessel identifier, event date, and reported catch amount. Hypothesis 9 and Hypothesis 10 were formed regarding the perceived usefulness and ease of use of this feature, respectively.

Figure 5.6a shows responses with the statements regarding the usefulness of this interactive highlighting feature. All the participants were acknowledging the value of this feature. Thus, Hypothesis 9 was supported.



(a) Usefulness of spatial filter.



(b) Ease of use of spatial filter.

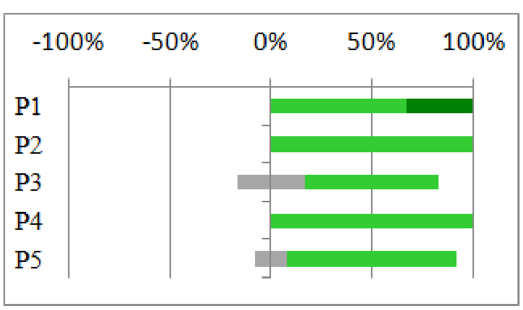
Figure 5.5: User acceptance of the spatial filter for analyzing event anomalies.

While this feature was perceived as useful by all the analysts, Figure 5.6b shows that the analysts had a slightly different opinion regarding the perceived ease of use. One participant (P1) strongly agreed and two other participants (P2 and P4) agreed, whereas, another participant (P5) showed more neutral responses with the statements regarding the perceived ease of use of this feature. Participant P3 was pessimistic about his/her answer about the statements regarding perceived ease of use of this feature. According to these participants responses (P3 and P5), this features was not easy for them to learn. Thus, they were not able to use this feature properly within their analysis tasks, which influenced their responses. Thus, Hypothesis 10 was neither supported nor rejected by the data.

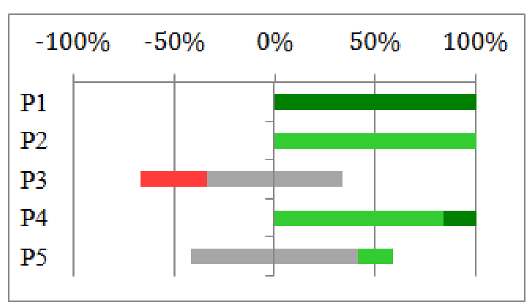
5.7.1.7 Highlighting Event Anomalies from the Map Representation

Event anomalies can be highlighted based on the location or surrounding event anomalies during the analysis session. Such an option was added to the map representation where analysts can click on the event anomalies to highlight them. Hypothesis 11 and Hypothesis 12 involved the perceived usefulness and ease of use of this feature, respectively.

Figure 5.7a shows the responses with the statements regarding the usefulness of high-



(a) Usefulness of anomaly highlighting from tree representation.



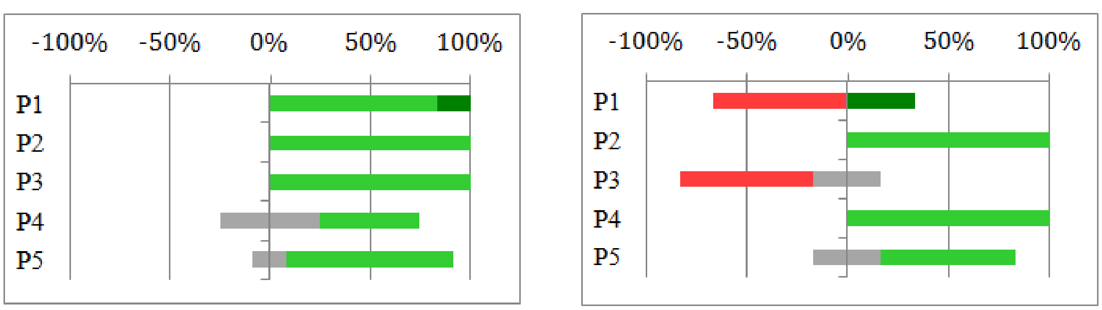
(b) Ease of use of anomaly highlighting from tree representation.

Figure 5.6: User acceptance of the anomaly highlighting from the tree representation.

lighting event anomalies from the map representation. Four participants were found agreeing, whereas the other participant's (P4) responses were leaned towards agreeing with the statements regarding the perceived usefulness of this feature. Since all the participants find the value of this feature, Hypothesis 11 was supported by the data. Figure 5.7b shows the responses with the statements regarding the ease of use of this feature. Mixed responses were found for this features. Analyzing the individuals' responses on the six questions, it is found that this feature is easy to understand and learn, but due to the lower thickness of the line and the spatial filter on zoom and pan map operations, some participants did not find this feature flexible and becoming skillful to use this feature was not easy for them. At the same time three participants (P2, P4, and P5) were agreeing or leaning toward agreeing with the responses regarding the ease of use of this feature. Therefore, this hypothesis was neither supported not rejected by the data.

5.7.1.8 Showing Contextual Data on Event Highlight

To understand the underlying phenomenon of the event anomalies, the entire movement path of the corresponding vessels are required to be analyzed. Thus, for the selected event anomalies, the details of the vessels' movement paths are shown on the



(a) Usefulness of anomaly highlighting from map representation.

(b) Ease of use of anomaly highlighting from map representation.

Figure 5.7: User acceptance of the anomaly highlighting from the map representation.

map. In Hypothesis 13, it was expected that the analysts will find this feature useful for their analysis tasks.

Figure 5.8 shows the participants' responses with the statements regarding the usefulness of this feature. Participants provided positive feedback about this feature. Their responses were leaned toward agreeing to strongly agreeing. Thus, the hypothesis was supported by the data.

5.7.2 Interview Questions

At the end of the post-study questionnaire, interviews were conducted. Each participant was asked the same six questions (see Appendix A) regarding their experience of using the system. While most participants provided positive feedback about the prototype system, some commented on their difficulties to use the system, and suggested specific improvements. The key findings from the interview questions are outlined below.

5.7.2.1 Positive Feedback

The analysis of the interview responses revealed a number of common themes in the participants' qualitative feedback. All the participants found the representation of the degree of discrepancies of event anomalies using yellow lines useful for their analysis

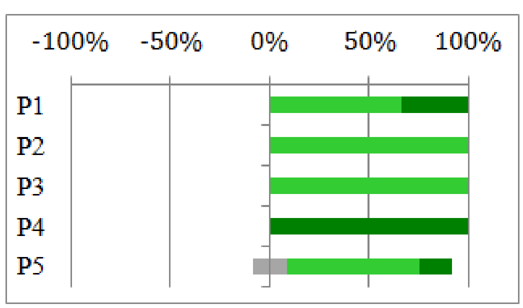


Figure 5.8: Usefulness of the contextual data when highlighting event anomalies.

tasks. The participants used multiple software including commercial GIS tools to gather the same information, however, the relations between these two data sets was not explicitly provided by any of their existing systems. Participant P3 stated that "*In terms of being able to quickly visualize the differences in data and the magnitude of discrepancies, this is huge*". Participant P2 echoed, "*You know, I am involved but I did not realize that there are some cases where there are quite a discrepancy*". Participant P4 provided a detailed comparison of this feature with their existing work practices, "*It would be incorporated almost as a daily tool for me. It could replace probably three pieces of software I am currently using. Literally what happens for me, I use SQL script to draw these data from the database, then a series of checks with different software such as Excel, but the visualization comes after that, and this method is much more integrated. Thus, one piece of software gives me my anomalies instantaneously, as opposed to having series of added checks and then doing analysis of subsequent uses of software*".

All the participants agreed that their existing work practices are too tedious and they always analyzed a subset of the data. This system provides a view of the event anomalies over a broad temporal and geographical scale. Participant P5 stated that "*I usually would not plot out all the record for a month or year, at the same time I would do the case by case basis, so this kind of whole picture is very different than how I look at the data right now on my work. [This new approach] is definitely useful*". Participant P1 noted that "*It certainly highlights the erroneous data from the log records, and I see the application of it for selecting out the anomalous log entry and further investigating them*".

While this prototype system has a clear application in the area of fisheries enforcement, participants also find that this system can also be applied for correcting the underwater fish stock assessment data, which is important for the analysts from ocean science division of Fisheries and Ocean Canada. Participant P4 commented that *"This is becoming more and more important in this business, in terms of science and in terms of working with enforcement. It looks like an efficient package to make my job easier"*. Participant P5 added that *"It is really important for us to allocate the catch amount in the right location. This would be an application for our stock assessment"*.

5.7.2.2 Negative Feedback

During the interview sessions, two issues were reported by the participants about the prototype system. The first issue was about the difficulty of selecting event anomalies from the *Map View*. Since the straight lines and flow lines were thin to keep the visual weight low, it was difficult for the analysts to click on them when the map was not zoomed in enough. In such cases, they had to zoom in to click on them, and sometimes that caused lose of spatial context. This is also the root cause of the mixed response on the Hypothesis 11.

The other difficulty the participants noted was the usability of spatial filter. Currently the spatial filter is performed based on the event locations. Some of the events had very large discrepancies where zooming in to the movement paths left the event locations outside of the viewing window and the event filtered out. The participants found difficulties analyzing such events closely. They suggested filtering out event anomalies only when no portion of those events are visible in the map view, which will increase the ease of use of this feature and the Hypothesis 8 may be supported.

5.7.2.3 New Features or Improvements

During the interview sessions, participants also made comments about improvements and additions of new features to the system as follows:

First, the missing VMS data point is a visual feature in the prototype systems. The analysts were more keen to analyze events with missing data points separately. Thus, they suggested providing a filtering feature so that they can separate such event anomalies and analyze them. According to their statements, they analyze this kind of anomalies differently than others.

Second, the anomaly detection procedure currently use the length of time that the vessels spend around the event locations which are calculated by finding the VMS data points around those locations. The consecutiveness of these VMS data points was not considered within the procedure. For example, if a vessel spent an hour of time around the reported fishing location, moved to some other location, then came back to the previous location and spent another hour, this procedure considers two hours of time span around the event location. Analysts commented that, this procedure may not be able to correctly identify all the event anomalies. They suggested for an option that allows them to calculate these hours from the consecutive data points, when required.

Third, the VMS data used in this prototype has only one temporal resolution (hourly data). VMS data may also consist of multiple resolutions (15 minutes and hourly data). Thus, the participants suggested that the system should support extracting and analyzing event anomalies from data sets having multiple temporal resolutions.

Fourth, the ocean is divided into several regions for administrative purpose. Each of these regions is administrated separately. Licensing conditions for each region is also different. Thus, one type of event anomalies analysts were interested in where the movement paths were in one region and the event locations were reported in another. A filtering option for extracting such events was suggest by the analysts.

5.7.3 Observations

While the participants were analyzing the data sets with their own choices of tasks, a pattern of using the system was noticed among the participants. All the participants started with setting a temporal range and configuring the anomaly threshold filter. Next, they zoomed to a geographical region. After that, they undertook further filter refinement steps, which included the ancillary data filter, adjusting the spatial extent, and changing the temporal window. All the participants were found using the event anomaly highlighting feature after they brought down the number of event anomalies to a manageable size with their filter refinement actions.

While analyzing the event anomalies, all the participants except P4 were analyzing event anomalies in the bay regions. Participant P4 was analyzing the anomalies in the southern peninsula and further southern regions. When participant P4 was analyzing events from offshore, because of the absence of any land mark it was difficult for this participant to understand the spatial extent of the viewing window of that map.

During the use of the system, all of the participants were able to find and evaluate vessels that misreported their fishing activities, and conducted detailed investigations of the vessels' activities over a wider temporal range. That gives them the understanding of these vessels' activities before, during, and after the event in question. The temporal range was varied from one month to three months for all of the participants, one participant (P3) tried to use a range of one year for a particular vessel. All the participants also compared the event anomalies with their surrounding event anomalies. One of the participants (P4) compared event anomalies performed by a certain group of vessels from different times of the year to understand how the pattern of incorrect reporting varies over the time.

All the participants were enthusiastic about this prototype software. Although they use this software for the analysis tasks of their own choices, all the features of this software were used by all the participants. They were found adjusting deferent filtering parameters and highlighting events iteratively, which indicated their reasoning activities about the data. Their replies on interview questions reveals that they were able to identify their known facts about the data, and also able to discover new information. The vessel identifier was coded within the prototype for privacy policy. Thus, all the analysts except P4 was finding difficulties to validate some of their know facts about the data. All the participants were also found facing difficulties using the spatial filter and selecting events from the map view. Some of them were able to learn this feature quickly within their analysis session and provided positive response in the post-study questionnaire.

5.8 Discussion

In this evaluation, a set of field trials were conducted among a group of fisheries data analysts to validate the potential value of this geovisual analytics system. A summary of the field trials results regarding the usefulness and ease of use of the features of this system is shown in Table 5.3. All the features of this system, except representation of mission data points, were found useful for the analysts and two features were found easy to use. The field trials also reveal some improvements of some of the features. One of the possible reasons for finding the features of this prototype system useful

Hypothesis	**Feature**	**Usefulness**	**Ease of Use**
1	Event representation	Supported	
2	Representation of missing data points	Mixed	
3,4	Anomaly threshold	Supported	Supported
5,6	Ancillary data filter	Supported	Supported
7,8	Filtering event with zoom and pan operations	Supported	Mixed
9,10	Event highlighting from event explorer	Supported	Mixed
11,12	Event highlighting from map view	Supported	Mixed
13	Showing contextual information	Supported	

Table 5.3: Usefulness and ease of use of the features.

by the analysts is the underlying theories (mantra of visual analytics, Gestalt Laws of Pattern Perception, Opponent Process Theory of Colour) used for designing the system. Using this system, the analysts were able to analyze large data sets; filter out event anomalies that are not of interest based on spatial, temporal, and ancillary data; visualize the degree of discrepancies within event anomalies; and investigate and compare event anomalies. The mantra for the visual analytics process proposed by Kiem et al. is: analyze first, show the important, zoom, filter, analyze further, and details on demand (see Section 2.2). The observations of the field trials show that the analysts were following this visual analytics mantra during their data analysis session and were able to perform analytic reasoning about the data by validating the known facts and discovering new knowledge.

The other possible reason for finding the system useful by the participants was the absence of any existing system. Although the participants are analyzing fishing event anomalies for a significant period of time, due to the absence of proper tools they were not able to analyze the full data sets. With the use of modern GIS tool they could only analyze the data sets separately. While the data sets are representing the same events which are collected separately, it is important to analyze them together, explicitly showing the discrepancies and relationships among them, which is not possible by analyzing them separately. Thus, analysts were never able to analyze all the aspects of event anomalies. Using this prototype, they were able to visualize the event anomalies on the map that includes the spatial context, surrounding event anomalies, and the degree of discrepancies. From the interviews it was found that all the participants were able to discover new knowledge from the data sets, and also were able to separate event anomalies that relevant to their analysis tasks. Thus, this system provided the analysts with what they were looking for.

The field trials also revealed three of the features that were not easy for some of the participants to use: spatial filtering, event highlighting from map view, and event highlighting from event explorer. Two out of these three features was related to the visual variable for representing event anomalies. Since the visual variable is introduced in this research, participants did not have any prior experience to interact with it. In addition, the field trials were conducted within a limited period of time. Therefore, the participants had little chance to get familiar with the entire prototype system, which was limiting their ability to become more skilful and might have influenced their opinion.

While the features of the prototype was found useful for the data sets used in it, the field trials also revealed a few improvements that are required to make it robust and more useful. Anomaly detection is one of them that needs improvement in order to detect anomalies from other types of fisheries data and detect more refined and true positive events (as discussed earlier in Section 5.7.2.3). While a few limitations were found related to user interaction with visual variable for representing event anomalies, one specific direction was also found to improve the spatial filtering.

Although the features of the prototype system were designed with the guidance from the literature and by following the best practices of geovisual analytics system, the field trials revealed limitations of the prototype system. The limited ease of use of the visual variable for representing event anomalies in the map view was one of the important findings among them. Since this is a new visual variable introduced in this research, the outcome related to its interaction provides a guideline for further improvement. Interviews and investigator's observation revealed more insights in this direction. The future works related to the visual variable and other features of this prototype system revealed from the field trials, interviews and investigator's observation are discussed in next chapter.

Chapter 6

Conclusions and Future Work

6.1 Summary

The goal of this thesis has been to address fundamental issues related to the current geovisual analytics systems for analyzing event anomalies among multiple data sets. To fulfill this goal, methods for detecting events and anomalies within these data sets were introduced; a visual variable was developed to represent these event anomalies in the map-based visualization, which adds spatial context to these event anomalies; and the user interactions provided within the system, which enable analysts to control what it means for an event to be considered as an anomaly and explore them among the data sets by filtering and highlighting. This system allows analysts to incorporate their domain specific knowledge within their data analysis tasks and provides support for their reasoning about the data. This approach takes advantage of multiple coordinate views and interactive data filtering.

A set of field trials were conducted to evaluate this event anomaly analysis system. With this approach, fisheries data analysts analyzed commercial fishing vessel movement (VMS) data and catch report (MarFis) data, exploring interesting anomalies and patterns of anomalies among the data and performing analytic reasoning about them. The remainder of this chapter summarizes the contributions of this research work, limitations, generalizability of this work, and potential future research directions.

6.2 Research Contributions

The geovisual analytics system for analyzing event anomalies from movement and geographical point data was an unexplored area of study. Thus, the fundamental research question was: *Is it possible to design a geovisual analytics system for analyzing event anomalies?* This work is presented as a design study in this thesis and it outlines the aspects of event anomalies that make this work interesting and challenging. The mismatch in temporal granularity among the data sets makes this work more complicated by adding uncertainty into the anomaly detection process, which leads towards a human-centred analysis for separating the meaningful anomalies from those that are a result of the mismatch between the temporal scales. To address this problem, controllable threshold values were introduced where analysts can define the parameters for filtering event anomalies. These threshold values consider both temporal and positional differences within the data. A new visual variable was introduced to represent the event anomalies on map. User interactions were provided to perform exploration and analytic reasoning about the data. Chapter 3 and 4 explain the details of design and implementation of this system and conform the answer of the first research question.

The other research questions were related to the usefulness, ease of use, and enhancement of analysts' ability of this system. The field trials conducted within this research were designed to address these questions. The answers of these questions are discusses in the following.

6.2.1 Usefulness

The fundamental research question raised regarding the usefulness of this approach was: *Is the geovisual analytics system useful for the analysts to identify and analyze potential event anomalies?* To identify the answer to this question, eight hypotheses were formulated, one for each of the features of the system, and validated by the field trials. The feature anomaly threshold and representation of event anomalies on the geographical map are the novel contribution of this research. The anomaly threshold is designed to address the difference in temporal granularity of the data sets and the visual representation of event anomalies is designed to represent the differences in the data sets when multiple data sets are representing the same entities performing the same events. Both of these features were found useful. The field trials also reveals that the representation of missing data points was not very useful, but have a moderate usefulness. The other features were also found useful in the field trials. Thus, overall, it can be concluded that the system was useful for the analysts.

6.2.2 Ease of Use

The next fundamental question raised about this approach was, *Does the geovisual analytics system make it easy for the analysts to explore the data sets for extracting knowledge about event anomalies and entities' activities?* During the field trials participants were asked to identify the ease of use of five of the eight features of the system. Anomaly threshold was found easy to use, however, mixed opinion was found about the user interaction with the event representations.

Anomaly threshold was implemented using common controls (sliders). Thus, the participants were comfortable with these features. However, interacting with the representation of event anomalies on the map was not easy. The lines of the visual representation were kept thin for keeping the visual weight low, but that was causing problem on selecting the anomalies. The spatial filter also considered the event location for filtering, which was confusing for the analysts. The filtering should remove the event when both the event location and the movement path are outside of the focus of the map view.

Despite the medium and mixed opinions on user interactions of anomaly representation, all the participants were able to gain new knowledge within their limited system use session. Thus, it can be concluded that with some improvement and more system use, the features which were found less easy to use may become easier for the analysts. Therefore, the system is also promising in terms of their ease of use.

6.2.3 Enhancement of Analysts' Ability

The final fundamental question raised regarding this approach was: *Does the system presented in this research enhance the analysts' abilities to make sense of the data and discover anomalies that are both known and previously unknown among the event anomalies within the data sets?* To answer this question, participants of the field trials were interviewed about their experience with the system, the new knowledge they gained from their system use session, and the differences from their existing work practices. Their answers shows that the system enhanced their ability to analyze the data sets. Their present work practice allows them to analyze data only for the chosen events, while this system allowed them to analyze the entire data sets, separate the event anomalies that were caused by systematic errors, and analyze those events that were of interest to them. One participant also mentioned that this system will increase their efficiency and productivity. Although they pointed out some limitations of this system and suggested a few improvements, but all of them agreed with the fact that this system enhances their ability to explore and analyze the event anomalies.

6.3 Limitations

Although the field trials showed that this anomaly detection and analysis methods were useful for the analysts and enhanced their ability to analyze the data sets, this work has several limitations. This work using movement data and velocity of the moving entity was not considered for identifying and analyzing the event anomalies. This velocity information could be an important aspect of the analysis [25]. Thus, this may limit the knowledge that can be discovered by using this system.

In this work, the movement data has higher level of temporal granularity than the geospatial point data. Each event has only one geospatial point data with a set of data points from movement data. The anomaly detection method was designed considering this fact. Thus, this method will not work for the data sets where the geospatial point data has a higher level of temporal granularity than the movement data.

The fishing events considered in this work were daily fishing events, where the calendar date was used to define the temporal boundary of individual events. A fishing event may start in the evening and ends in the next morning. Such event will be considered as two events by the current event detection method, which will also limit the use of this system.

The geographic visualization used in this work has limited flexibility and shows only the movement paths and event locations. The colour for representing the landscape information on map can not be changed in this system, thus its effect can not be studied. As a result, the colours used in the visual variables for representing the events were influenced by the colour used on the map. Additional information about the event anomalies was also not shown on the map view. As a result, analysts were required to use interactive brushing and selection facilities for retrieving this information while analyzing the data.

The visual variable used in this work focused on representing event anomalies where the root cause of the anomalies is positional discrepancies. Event anomalies that have different root causes such as sudden changes of speed or direction may not be easily represented with this technique. Further research is required to adapt this system for representing such anomalies.

6.4 Generalizability

Although the prototype system was implemented and evaluated by analyzing fishing event anomalies in commercial fisheries data, which may be considered a specialized type of data analysis, this system was designed for and can be applied more generally to cases where geospatial point data and movement data representing events are collected independently, but need to be verified as correct. For example, individuals' movement paths measured from their smart phones and locations where they used their credit cards at point-of-sale machines can be analyzed by using this system for identifying fraudulent charges. The charge amount and/or merchant type may be used for ancillary data filter in this analysis. A different colour scheme is required to be chosen for representing this event anomalies, since the underlying landscape of these events are different than that of the fishing events.

This system can also be used to analyze the events of distractions of ship voyages from their original courses. Ships often change their planned courses due to weather forecast or other unwanted circumstances. For this type of analysis, the event detection method is required to be changed. The segment of the trajectories that were distracted from their original course can be defined by the threshold values: the distance between the original course and the planned course, and the time the ship was away from the original course. Then, the distraction events are represented by linking these trajectory segments with the nearest locations where weather conditions were changed. The weather conditions and the ship voyage information can be visualized on a geographical map to understand the spatial contexts of these events. Analyzing these events can help designing safe route, and also can identify the events where distractions were not made because of the change of weather conditions. In the latter case analysts can further investigate such events with the contextual data, such as nearest port and another ship close to that distracted ship.

6.5 Future Work

Addressing the limitations of this work can be a good future work for this system. Currently the system considers movement data points collected at regular intervals (one data points in one hour). The intervals between these data points are one of the key parts of this system. Thus, this system can be improved for using movement data collected at irregular intervals (i.e., some tracking devices are recording data at every hour interval time and some others are recording data at every 15 minutes interval time). Incorporating the velocity for analyzing event anomalies will also be an important improvement of this system. A graph within the time line showing when anomalies are occurring with the ratio of event and event anomalies can also be added to guide the analysts to choose the time span for detailed analysis.

Another interesting future work may be analyzing other types of anomalies, such as moving entities that are traveling in one geographical region and reporting their location in another region. This type of events anomaly are not detected by the current system if the distance between the movement paths and reported event locations are within the distance threshold. Thus, this system is required to be adjusted for analyzing this type of anomaly.

This work was designed for analyzing the positional discrepancies in events, however, it can also be used for analyzing events based on similarities by applying an inverted logic in the event anomaly detection algorithm. For example, movement data of animals can be collected using special collars with devices that measure the positions at chosen time intervals and transmit the measurements via radio networks. This data can be analyzed to identify and understand the events of animals' preference to visit locations with specific properties, such as previously hunted locations or open area. In this case, the anomaly thresholds can be used to define the length of the trajectory segments of these animals. If these trajectory segments are found within the selected distance thresholds from these locations, it will be considered as a visit to this area. Then, these visits can be considered as the unit of analysis for the rest of the components of this system and represented in the visualizations with suitable colour schemes. This analysis system may become a useful tool for the wild animal researchers for understanding the animals' behaviour, such as preferred locations of animals for hunting, and daytime resting.

The prototype used NASA World Wind mapping technology, which limits the flexibility of choosing the colour used in the map for representing different landscape information. Thus, the usability of the visualization cannot be studied with different settings, such as using a dark blue ocean background vs. light blue. A different mapping technology can be used in future work allowing changes of the base map, which may identify other optimal visualization settings.

The map view currently shows only the movement paths and the event locations using colour encoding. No other information, such as vessel ID, event identifier, and ancillary data are shown on the map. While such information can be found using interactive brushing and selection tools, the effect on the usability for showing such information on the map can also be studied in the future.

The system currently can identify the obvious event anomalies (i.e., fishing locations on land) or having missing data points in movement data. Exporting these types of event anomalies and analyzing them separately may reveal interesting information about the data sets, such as instrumental or communication error. Therefore, such data export features can also be added to the system in future.

The participants of the field trials did not have any existing system that could do a similar analysis, and for this reason they always analyze a part of the data. This fact may influenced by the field trials results. Therefore, field trials can be conducted within a domain where the system can be compared with a baseline system.

This anomaly analysis approach is presented in general, however, the case study and field trials were focused on a specific domain. The system found useful for this domain, however, to understand the generalizability of this system, user evaluations are also required to be conducted on the data from other domains.

Bibliography

[1] B. Aeschliman, B. Kim, and M. Burton. A visual analysis of spatio-temporal data associated with human movement. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 400–403, 2009.

[2] C. Ahlberg, C. Williamson, and B. Shneiderman. Dynamic queries for information exploration: an implementation and evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 619–626, 1992.

[3] I. Ajzen. From intentions to actions: a theory of planned behavior. In Kuhl, editor, *Action Control: From Cognition to Behavior*, chapter 2, pages 11–39. Springer-Verlag, 1985.

[4] G. Andrienko, N. Andrienko, and M. Heurich. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, 25(9):1347–1370, 2011.

[5] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M.-J. Kraak, A. MacEachren, and S. Wrobel. Geovisual analytics for spatial decision support: setting the research agenda. *International Journal of Geographical Information Science*, 21(8):839–857, 2007.

[6] G. Andrienko, N. Andrienko, I. Kopanikis, and A. Ligtenberg. Visual analyics methods for movement data. In F. Giannotti and D. Pedreschi, editors, *Mobility, Data Mining and Privacy*, chapter 13, pages 375–408. Springer Berlin Heidelberg, 2008.

[7] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46, 2007.

[8] N. Andrienko and G. Andrienko. Designing visual analytics methods for massive collections of movement data. *Cartographica*, 42:117–138, 2007.

[9] N. Andrienko, G. Andrienko, and P. Gatalsky. Supporting visual exploration of object movement. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 217–220, 2000.

[10] N. Andrienko, G. Andrienko, and P. Gatalsky. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 14(6):503–541, 2003.

[11] N. Andrienko, G. Andrienko, and P. Gatalsky. Impact of data and task characteristics on design of spatio-temporal data visualization tools. In J. Dykes, A. M. MacEachren, and M.-J. Kraak, editors, *Exploring Geovisualization*, chapter 10, pages 201 – 222. Elsevier, 2005.

[12] R. Arsenault, C. Ware, M. Plumlee, S. Martin, L. Whitcomb, D. Wiley, T. Gross, and A. Bilgili. A system for visualizing time varying oceanographic 3D data. In *Proceedings of the IEEE Techno-Ocean Conference*, volume 2, pages 743 – 747, 2004.

[13] R. A. Becker and W. S. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[14] J. Bertin. *Semiology of Graphics.* Translated by W. J. Berg. University of Wisconsin Press, 1983.

[15] C. Brewer and M. Harrower. ColorBrewer 2.0: color advice for cartography. `http://colorbrewer2.org/`, 2013.

[16] B. Brown, S. Reeves, and S. Sherwood. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1657–1666, 2011.

[17] S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, chapter 2, pages 19–45. Springer-Verlag, 2008.

[18] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms.* McGraw-Hill Higher Education, second edition, 2001.

[19] T. Crnovrsanin, C. Muelder, C. D. Correa, and K.-L. Ma. Proximity-based visualization of movement trace data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 11–18, 2009.

[20] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989.

[21] F. D. Davis. On the relationship between HCI and technology acceptance research. In P. Zhang and D. Galletta, editors, *Human-Computer Interaction and Management Information Systems: Foundations*, chapter 4, pages 395–421. Advances in Management Information Systems, 2006.

[22] A. Dix and G. Ellis. Starting simple: Adding value to static visualisation through simple interaction. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 124–134, 1998.

[23] J. Duncan and G. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96:433–458, 1989.

[24] J. A. Dykes and D. M. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, 43(4):581–603, 2003.

[25] R. A. Enguehard, O. Hoeber, and R. Devillers. Interactive exploration of movement data: a case study of geovisual analytics for fishing vessel analysis. *Information Visualization*, 12(1):64–84, 2013.

[26] Esri. ArcGIS. `http://www.esri.com/software/arcgis`, 2012.

[27] Google. Google Earth. `http://www.google.com/earth/`, 2013.

[28] D. Guo, J. Chen, A. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1461–1474, 2006.

[29] C. G. Healey, K. S. Booth, and J. T. Enns. High-speed visual estimation using preattentive processing. *ACM Transection of Computer Human Interaction*, 3(2):107–135, 1996.

[30] E. Hering. *Outlines of a Theory of Light Sense*. Harvard University Press, 1964.

[31] Q. Ho and M. Jern. Exploratory 3D geovisual analytics. In *Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future*, pages 276–283, 2008.

[32] O. Hoeber, G. Wilson, S. Harding, R. Enguehard, and R. Devillers. Exploring geo-temporal differences using GTdiff. In *Proceedings of the Pacific Visualization Symposium*, pages 139–146, 2011.

[33] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the IEEE Conference on Visualization*, pages 361–378, 1990.

[34] S. Jänicke, C. Heine, R. Stockmann, and G. Scheuermann. Comparative visualization of geospatial-temporal data. In *Proceedings of the International Conference on Information Visualization Theory and Applications*, pages 613–625, 2012.

[35] M. Jern, T. Astrom, and S. Johansson. GeoAnalytics tools applied to large geospatial datasets. In *Proceedings of the International Conference on Information Visualisation*, pages 362–372, 2008.

[36] M. Jern and J. Franzen. GeoAnalytics - exploring spatio-temporal and multivariate data. In *Proceedings of the International Conference on Information Visualization*, pages 25–31, 2006.

[37] S. Johansson and M. Jern. GeoAnalytics visual inquiry and filtering tools in parallel coordinates plots. In *Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, pages 33:1–33:8, 2007.

[38] T. Kapler and W. Wright. GeoTime information visualization. *Information Visualization*, 4(2):136–146, 2005.

[39] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: definition, process, and challenges. In A. Kerren, J. T. Stasko,

J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, volume 4950, pages 154–175. Springer Berlin Heidelberg, 2008.

[40] D. A. Keim, F. Mansmann, D. Oelke, and H. Ziegler. Visual analytics: combining automated discovery with interactive visualizations. In *Proceedings of the International Conference on Discovery Science*, pages 2–14, 2008.

[41] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. Challenges in visual data analysis. In *Proceedings of the Conference on Information Visualization*, pages 9–16, 2006.

[42] K. Koffka. *Principles of Gestalt Psychology*. Harcourt Brace and Company, 1935.

[43] M. J. Kraak. The space-time cube revisited from a geovisualization perspective. *Proceedings of the International Cartographic Conference*, pages 1988–1995, 2003.

[44] M.-J. Kraak and D. E. V. de Vlag. Understanding spatiotemporal patterns: visual ordering of space and time. *Cartographica*, 42(2):153–161, 2007.

[45] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.

[46] P. Lundblad, O. Eurenius, and T. Heldring. Interactive visualization of weather and ship data. In *Proceedings of the International Conference on Information Visualisation*, pages 379–386, 2009.

[47] P. Lundblad, M. Jern, and C. Forsell. Voyage analysis applied to geovisual analytics. In *Proceedings of the International Conference on Information Visualisation*, pages 381–388, 2008.

[48] A. M. MacEachren. *How Maps Work: Representation, Visualization, and Design*. Guilford Press, 1995.

[49] A. M. MacEachren and M.-J. Kraak. Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28(1):3–12, 2001.

[50] R. Maciejewski, B. Tyner, Y. Jang, C. Zheng, R. V. Nehme, D. S. Ebert, W. S. Cleveland, M. Ouzzani, S. J. Grannis, and L. T. Glickman. LAHVA: linked animal-human health visual analytics. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 27–34, 2007.

[51] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1986.

[52] D. Mountain. Visualizing, querying and summarizing individual spatio-temporal behaviour. In J. Dykes, A. M. MacEachren, and M.-J. Kraak, editors, *Exploring Geovisualization*, chapter 9, pages 181 – 200. Elsevier, 2005.

[53] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1):381–400, 1985.

[54] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[55] M. Nanni, R. Trasarti, C. Renso, F. Giannotti, and D. Pedreschi. Advanced knowledge discovery on movement data with the GeoPKDD system. In *Proceedings of International Conference on Extending Database Technology*, pages 693–696, 2010.

[56] NASA. NASA World Wind. `http://worldwind.arc.nasa.gov/java/`, 2004.

[57] Open Source Geospatial Foundation. Quantum GIS. `http://www.qgis.org/`, 2012.

[58] Oracle Corporation. Java development kit 7. `http://www.oracle.com/technetwork/java/javase/overview/index.html`, 2013.

[59] D. Orellana, M. Wachowicz, N. Andrienko, and G. Andrienko. Uncovering interaction patterns in mobile outdoor gaming. In *Proceedings of the International Conference on Advanced Geographic Information Systems & Web Services*, pages 177–182, 2009.

[60] S. Palmer and I. Rock. Rethinking perceptual organization: the role of uniform connectedness. *Psychonomic Bulletin & Review*, 1(1):29–55, 1994.

[61] N. Pelekis, I. Kopanakis, G. Marketos, I. Ntoutsi, G. Andrienko, and Y. Theodoridis. Similarity search in trajectory databases. In *Proceedings of the International Symposium on Temporal Representation and Reasoning*, pages 129–140, 2007.

[62] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 109–116, 2004.

[63] P. Quinlan and G. Humphreys. Visual search for targets defined by combinations of color, shape, and size: an examination of the task constraints on feature and conjunction searches. *Perception & Psychophysics*, 41(5):455–472, 1987.

[64] C. Rinner. A geographic visualization approach to multi-criteria evaluation of urban quality of life. *International Journal of Geographical Information Science*, 21(8):907–919, 2007.

[65] J. F. Rodrigues, Jr., A. J. M. Traina, M. C. F. de Oliveira, and C. Traina, Jr. The spatial-perceptual design space: a new comprehension for data visualization. *Information Visualization*, 6(4):261–279, 2007.

[66] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[67] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1):126–146, 2008.

[68] G. Strong and M. Gong. Similarity-based image organization and browsing using multi-resolution self-organizing map. *Image Vision Computing*, 29(11):774–786, 2011.

[69] The government of United States. National Aeronautics and Space Administration . `http://www.nasa.gov/`, 2014.

[70] J. J. Thomas and K. A. Cook, editors. *Illuminating the path: the research and development agenda for visual analytics*. IEEE Computer Society, 2005.

[71] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006.

[72] L. A. Treinish. Visual data fusion for applications of high-resolution numerical weather prediction. In *Proceedings of the Conference on Visualization*, pages 477–480, 2000.

[73] USA-CERL. Geographic resources analysis support system (GRASS GIS). `http://grass.fbk.eu/`, 2012.

[74] W. R. van Hage and D. Ceolin. *The Simple Event Model*. Springer, 2013.

[75] J. van Wijk. The value of visualization. In *Proceedings of the IEEE Visualization*, pages 79–86, 2005.

[76] A. Vandecasteele, R. Devillers, and N. Aldo. From movement data to objects behavior using semantic trajectory and semantic events. *Marine Geodesy*, 37(2):126–144, 2014.

[77] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science*, 46(2):186–204, 2000.

[78] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27(3):425–478, 2003.

[79] M. Vlachos, D. Gunopoulos, and G. Kollios. Discovering similar multidimensional trajectories. In *Proceedings of the International Conference on Data Engineering*, pages 673–684, 2002.

[80] Y. Wakabayashi and T. Ishikawa. Spatial thinking in geographic information science: a review of past studies and prospects for the future. *Procedia - Social and Behavioral Sciences*, 21:304 – 313, 2011.

[81] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, second edition, 2004.

[82] J. Wood, J. Dykes, A. Slingsby, and K. Clarke. Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1176–1183, 2007.

[83] J. S. Yi, Y. a. Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007.

[84] R. K. Yin. *Case Study Research: Design and Methods*. Sage Publications, fourth edition, 2008.

[85] H. Yu. Spatio-temporal GIS design for exploring interactions of human activities. *Cartography and Geographic Information Science*, 33(1):3–19, 2006.

Appendix A

User Evaluation

This appendix includes all the documents related to the field trials.

A.1 Approval of Field Trials

MEMORIAL UNIVERSITY

Interdisciplinary Committee on Ethics in Human Research (ICEHR)

Office of Research Services
St. John's, NL Canada A1C 5S7
Tel: 709 864 2561 Fax: 709 864 4612
www.mun.ca/research

ICEHR Number:	**20140910-SC**
Approval Period:	December 6, 2013 – December 31, 2014
Funding Source:	Supervisor's NSERC grant [title: *Geo-Visual Analytics of Capture Fisheries Statistical Data*]
Responsible Faculty:	Dr. Orland Hoeber Department of Computer Science, University of Regina (also adjunct at Memorial University)
Title of Project:	*Field Trials with Geovisual Analytics Software for Exploring Anomalies in Fisheries Data*

December 6, 2013

Md. Monjur Hasan
Department of Computer Science, Faculty of Science
Memorial University of Newfoundland

Dear Md. Hasan:

Thank you for your submission to the Interdisciplinary Committee on Ethics in Human Research (ICEHR) seeking ethical clearance for the above-named research project.

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* (TCPS2). *Full ethics clearance* is granted for one year from the date of this letter.

If you need to make changes during the course of the project which may give rise to ethical concerns, please forward an amendment request with a description of these changes to Theresa Heath at icehr@mun.ca for the Committee's consideration.

The *TCPS2* requires that you submit an annual status report on your project to the ICEHR before December 31, 2014. If you plan to continue the project, you need to request renewal of your ethics clearance, including a brief summary on the progress of your research. When the project no longer requires contact with human participants, is completed and/or terminated, you need to provide the final report with a brief summary, and your file will be closed. The annual update form is on the ICEHR website at http://www.mun.ca/research/ethics/humans/icehr/applications/.

We wish you success with your research.

Yours sincerely,

M Shute

Michael Shute, Th.D.
Chair, Interdisciplinary Committee on
Ethics in Human Research

MS/th

copy: Supervisor – Dr. Orland Hoeber, Department of Computer Science, University of Regina
Director, Office of Research Services

A.2 Participants Recruitment Letter

Subject: participant recruitment for field trials

Hello,

My name is Md. Monjur Ul Hasan and I am a M.Sc. student in the Department of Computer Science at Memorial University. As part of my thesis, I am conducting research under the supervision of Dr. Orland Hoeber and Dr. Wolfgang Banzhaf in the domain of geovisual analytics.

In the course of this research, we have developed prototype software with the purpose of assisting analysts with their tasks of exploring anomalies between multiple related fisheries data sets. The software is developed to analyze VMS and MarFis datasets to explore the discrepancies between them, and consists of interactive filtering tools and a visualization method to show the anomalies on a virtual globe.

The primary objective of this study is to gain insight into how the prototype system that has been developed can be used in real-world exploration of the discrepancies between VMS data and MarFis data. We will conduct field trials with individual domain experts such as yourself at the Fisheries and Oceans Canada offices within the Bedford Institute of Oceanography, in Dartmouth, NS on [date to be determined]. The study is expected to take approximately two hours, depending on your level of engagement in the analysis activity. Your participation in this study is entirely voluntary. If you choose not to take part in this research or if you decide to withdraw from the research once it has started, there will be no negative consequences for you, for that time or in the future.

Although we will communicate with you via email to coordinate your participation, your identity is not required during the actual study. You will not be required to write your name or any identifying information on the research questionnaires that will assure your anonymity.

If you are interested in participating in this study, please send us an email at mmuhasan@mun.ca so that we can coordinate a schedule for the field trials. Once the schedule is arranged, we will inform you of the details via email and will provide you with a copy of the consent form so that you are aware of the details of the study in advance.

The proposal for this research has been reviewed 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 709-864-2861.

If you would like more information about this study, you may contact either of the primary investigators:

Md Monjur Ul Hasan
M.Sc. Student
Department of Computer Science
Memorial University of Newfoundland
Email: mmhuasan@mun.ca

Or

Dr. Orland Hoeber
Assistant Professor
Department of Computer Science
University of Regina
Email: orland.hoeber@uregina.ca

A.3 Consent Form

Informed Consent Form

Field Trials with Geovisual Analytics Software for Exploring Anomalies in Fisheries Data

Researchers: MdMonjurUlHasan,
Department of Computer Science
Memorial University of Newfoundland
Email: mmuhasan@mun.ca

Dr. Orland Hoeber
Department of Computer Science
University of Regina
Email: orland.hoeber@uregina.ca

Dr. Wolfgang Banzhaf
Department of Computer Science
Memorial University of Newfoundland
Email: banzhaf@mun.ca

You are invited to take part in a research project entitled *"Field Trials with Geovisual Analytics Software for Exploring Anomalies in Fisheries Data"*.

This form is part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. It also describes your right to withdraw from the study at any time. In order to decide whether you wish to participate in this research study, you should understand enough about its risks and benefits to be able to make an informed decision. This is the informed consent process. Take time to read this carefully and to understand the information given to you. Please contact the researcher, Md Monjur Ul Hasan, if you have any questions about the study or for more information not included here before you consent.

It is entirely up to you to decide whether to take part in this research. If you choose not to take part in this research 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
My name is Md. Monjur Ul Hasan and I am a M.Sc. student in the Department of Computer Science. As part of my thesis, I am conducting research under the supervision of Dr. Orland Hoeber and Dr. Wolfgang Banzhaf in the domain of geovisual analytics.

In the course of this research, we have developed a research prototype with the purpose of assisting analysts with their tasks of exploring anomalies between multiple related fisheries data sets. The prototype is developed to analyze VMS and MarFis data sets to explore the discrepancies among them. Our prototype software consists of interactive filtering tools and a visualization to show the data on a virtual globe.

You have been selected to participate in this field trial due to your experience in analyzing such data sets.

Purpose of study:
The primary objective of this study is for the researchers to gain insight into how the prototype system that has 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 anomaly analysis activities that are being undertaken by the participants.

What you will do in this study:
In this study, you will be asked to use our system to analyze the anomalies within a subset of the VMS and MarFis data, identify how these relate to specific fishing events, and exploring interesting features that might emerge from the identification of anomalous data. The datasets have been provided by Fisheries and Oceans Canada. After using our system to explore the anomalies in a loosely-directed manner, you will be asked to complete a questionnaire. A short interview will also be conducted in which we will ask your opinion on various aspects of our system and the types of anomaly analysis tasks you normally perform.

Your use of our system will be video recorded so that we can analyze your activities at a later date, and so that we can focus our attention on helping you to perform your data analysis tasks using the software. 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 approximately2 hours, depending on your level of engagement in the analysis activity.

Location:
The field trials will be conducted at the Fisheries and Oceans Canada offices within the Bedford Institute of Oceanography, in Dartmouth, NS.

Withdrawal from the study:
Your participation in this field trial is entirely your decision. You are free to withdraw from the field trial anytime before or during the activities. 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. Any collected data, both paper and electronic, will be destroyed immediately if you decide to withdraw from this study. Your decision of whether or not to participate in this study will not be shared with any of our associates and supporters at Fisheries and Oceans Canada. The raw data will only be made available to the principal investigators in this project, and not shared with any of our partners or external collaborators.

Possible benefits:
The primary benefit that you may find when participating in this study is the identification of specific anomalies with the data of which you may not have been previously aware. It may also highlight the potential for semi-automatic data analysis approaches that could be useful in other aspects of your work. Further, your participation will provide us with valuable information regarding how you are able to perform anomaly analysis tasks using our system, leading to further refinement and perhaps integration of the approach within the Fisheries and Oceans Canada information systems.

Possible risks:
There are no risks or harms associated with this study beyond the normal use of a computer system.

Confidentiality and Storage of Data:
In order to maintain the privacy of your participation in this study, the data collected will be held strictly confidential by the researchers. 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, as per Memorial University policy on Integrity in Scholarly Research. 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.

Anonymity:
Although we will communicate with you via email to coordinate your participation, your identity is not required during the actual study. You will not be required to write your name or any identifying information on the research questionnaires. Any identifying information will be kept separate from the details of your participation in the study. Any reporting of the outcomes of this research will exclude identifying information of the participants. The data itself will only be used by the researchers indicated in this consent form, and will not be shared in raw format with anyone.

Recording of Data:
Your use of the prototype system will be video recorded. However, the focus of the video recording will be on what you are doing with the system. As such, the video camera will be pointed at the computer screens, keyboard, and mouse. The audio portion of the recording will capture the discussions between yourself and the researcher during the anomaly analysis process. This video and audio 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:
The result from the study may be published in a student thesis, conference proceedings, and scientific journals. These publications will also be shared with Fisheries and Oceans Canada. All data will be anonymous and the raw data will not be included.

Sharing of Results with Participants:
Once the results of this study are published, we will inform you of this via email. If you do not wish to be included in such communications, you may indicate this at any time before, during, or after the study.

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, you may contact any of the researchers listed at the beginning of this document.

ICEHR Compliance:
The proposal for this research has been reviewed 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 709-864-2861.

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 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.
- You understand that your use of the software will be video recorded and your responses to the interview questions will be audio recorded.
- You understand that any data collected from you up to the point of your withdrawal will be destroyed.

If you sign this form, you do not give up your legal rights and do not release the researchers from their professional responsibilities.

Your signature:
I have read and understood what this study is about and appreciate the risks and benefits. I have had adequate time to think about this and had the opportunity to ask questions and my questions have been answered.

☐ I agree to participate in the research project understanding the risks and contributions of my participation, that my participation is voluntary, and that I may end my participation at any time.

A copy of this Informed 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 Principal Investigator Date

A.4 Pre-study Questionnaire

Pre-Study Questionnaire **Participant:_____**

Please answer the following questions with regard to your background.

1. How many years have you been involved in the VMS data analysis?

__________ Years.

2. How many years have you been involved in the MarFis data analysis?

__________ Years.

3. Please list any data analysis software systems or approaches you have used for analyzing VMS data.

1)
2)
3)
4)

4. Please list any data analysis software systems or approaches you have used for analyzing MarFis data.

1)
2)
3)
4)

5. What is your level of experience in visualizing VMS data?

1	2	3	4	5
Not at all				Very familiar

6. What is your level of experience in visualizing mafias data?

1	2	3	4	5
Not at all				Very familiar

7. How familiar are you with GIS or geovisualization systems such as Google Earth or ArcGIS?

1	2	3	4	5
Not at all				Very familiar

8. How familiar are you with analyzing anomalous fishing events between VMS and MarFis data?

1	2	3	4	5
Not at all				Very familiar

A.5 Post study Questionnaire

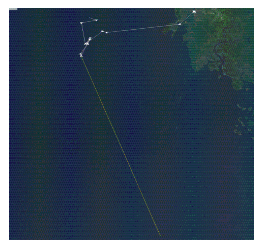
Post Study Questionnaire   Participant:_______

The following questions relate to your experience using the prototype system for exploring geospatial representations of anomalous fishing events. In the following you will be asked about different features you just used in the prototype system. A set of statements are provided about each of the features. We would like to know how strongly you agree or disagree with these statements. Your answers will allow us to understand your impressions of the software and to make future improvements.

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

Feature 1:

While working with the prototype software, you were shown the anomalous fishing events between vessel paths and the reported fishing locations. The vessel paths and reported locationswere connected using yellow lines (see an example in the picture at right). Please indicate how strongly do you agree or disagree with the following statements about the usefulness of this feature.



	Strongly Disagree	Disagre e	Neutral	Agree	Strongly Agree
Using the he **visual representation** enabled me to accomplish my anomaly analysis tasks more quickly.	1	2	3	4	5
Using the **visual representation** improved my anomaly analysis performance.	1	2	3	4	5
Using the **visual representation for analyzing anomalies** increased my productivity.	1	2	3	4	5
Using the **visual representation** enhanced my effectiveness in analyzing anomalies.	1	2	3	4	5
Using the **visual representation** made it easier for me to analyze the anomalies within the data.	1	2	3	4	5
I found the **visual representation** useful for analyzing the anomalies.	1	2	3	4	5

Feature 2:

In the prototype software,the missing VMS data points are interpolated and represented using whitecircles (see the picture at right). Please indicate how strongly do you agree or disagree with the following statements about the usefulness of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using the **visual representation of missing VMS data points** enabled me to accomplish my anomaly analysis tasks more quickly.	1	2	3	4	5
Using the **visual representation of missing VMS data points** improved my anomaly analysis performance.	1	2	3	4	5
Using the v**isual representation of missing VMS data points** increased my productivity.	1	2	3	4	5
Using the v**isual representation of missing VMS data points** enhanced my effectiveness in analyzing anomalies.	1	2	3	4	5
Using the **visual representation of missing VMS data points** made it easier for me to analyze anomalies within the data.	1	2	3	4	5
I found the **visual representation of missing VMS data points** was useful for analyzing the anomalies.	1	2	3	4	5

Feature 3:

In the prototype software, interactive controls are provide to adjust the anomaly threshold based on the distance between the VMS and MarFis data points, and the amount of time spent within this distance (see the picture at right). This feature allows you to filter out anomalies that are not of interest, and highlight those that are. Please indicate how strongly do you agree or disagree with the following statements about the usefulness and ease of use of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using the **Anomaly Threshold** enabled me to accomplish my anomaly analysis task more quickly.	1	2	3	4	5
Using the **Anomaly Threshold** improved my anomaly analysis performance.	1	2	3	4	5
Using the **Anomaly Threshold for analyzing anomaly** increased my productivity.	1	2	3	4	5
Using the **Anomaly Threshold** enhanced my effectiveness in analyzing anomalies.	1	2	3	4	5
Using the **Anomaly Threshold** made it easier for me to analyze the anomalies within the data.	1	2	3	4	5
I found the **Anomaly Threshold** was useful for analyzing the anomalies.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Learning to operate the **Anomaly Threshold** was easy for me.	1	2	3	4	5
I found it easy to get the **Anomaly Threshold** to do what I wanted it to do.	1	2	3	4	5
My interaction with the **Anomaly Threshold** was clear and understandable.	1	2	3	4	5
I found the **Anomaly Threshold** to be flexible to interact with.	1	2	3	4	5
It was easy for me to become skillful at using the **Anomaly Threshold**.	1	2	3	4	5
I found the **Anomaly Threshold** was easy to use.	1	2	3	4	5

Feature 4:

In the prototype, you were presented an option to filter out anomalies based on catch amount within the MarFis data (see the picture at right). This feature allows you to filter the data and to focus on aspects that are of interest based on the catch amount. Please indicate how strongly do you agree or disagree with the following statements about the usefulness and ease of use of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using the **Ancillary Filter** enabled me to accomplish my anomaly analysis tasks more quickly.	1	2	3	4	5
Using the **Ancillary Filter** improved my anomaly analysis performance.	1	2	3	4	5
Using the **Ancillary Filter** increased my productivity.	1	2	3	4	5
Using the **Ancillary Filter** enhanced my effectiveness in analyzing anomalies.	1	2	3	4	5
Using the **Ancillary Filter** made it easier for me to analyze the anomalies within the data.	1	2	3	4	5
I found the **Ancillary Filter** was useful for analyzing the anomalies.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Learning to operate the **Ancillary Filter** was easy for me.	1	2	3	4	5
I found it easy to get the **Ancillary Filter** to do what I wanted it to do.	1	2	3	4	5
My interaction with the **Ancillary Filter** was clear and understandable.	1	2	3	4	5
I found the **Ancillary Filter** to be flexible to interact with.	1	2	3	4	5
It was easy for me to become skillful at using the **Ancillary Filter**.	1	2	3	4	5
I found the **Ancillary Filter** was easy to use.	1	2	3	4	5

Feature 5:

In the prototype, the anomalies are shown at their corresponding locations within the Map View. Panning and zooming features allow you to change the location and spatial extent of your area of focus. The software always presents events from the focused area and hides the rest. Please indicate how strongly do you agree or disagree with the following statements about the usefulness and ease of use of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using **panning and zooming** to focus on specific geographic regions enabled me to accomplish my anomaly analysis tasks more quickly.	1	2	3	4	5
Using the **panning and zooming** to focus on specific geographic regions improved my anomaly analysis performance.	1	2	3	4	5
Using the **panning and zooming** to focus on specific geographic regions increased my productivity.	1	2	3	4	5
Using the **panning and zooming** to focus on specific geographic regions enhanced my effectiveness in analyzing anomalies.	1	2	3	4	5
Using the **panning and zooming** to focus on specific geographic regions made it easier for me to analyze the anomalies.	1	2	3	4	5
I found **panning and zooming** to focus on specific geographic regions useful for analyzing the anomalies.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Learning to operate the **panning and zooming** was easy for me.	1	2	3	4	5
I found it easy to get **the zoom and pan operations** to do what I wanted it to do.	1	2	3	4	5
My interaction with the anomalous **the zoom and pan operations** on the virtual globe were clear and understandable.	1	2	3	4	5
I found **the zoom and pan operations** on the virtual globe were flexible to interact with.	1	2	3	4	5
It was easy for me to become skillful at using **the zoom and pan operations** on the virtual globe.	1	2	3	4	5
I found **the zoom and pan operations** on virtual globe were easy to use.	1	2	3	4	5

Feature 6:

In the prototype, you were able to highlight anomalous fishing event from the Event Explorer based on the vessel identification number and the event data. With the highlighting of events, additional information was presented to you. Please indicate how strongly do you agree or disagree with the following statements about the usefulness and ease of use of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using events highlighting from the Event Explorer enabled me to accomplish my anomaly analysis task more quickly.	1	2	3	4	5
Using events highlighting from the Event Explorer improved my anomaly analysis performance.	1	2	3	4	5
Using events highlighting from the Event Explorer increased my productivity.	1	2	3	4	5
Using events highlighting from the Event Explorer enhanced my effectiveness in analyzing anomalies.	1	2	3	4	5
Using events highlighting from the Event Explorer made it easier for me to analyze the anomalies within the data.	1	2	3	4	5
I found events highlighting from the Event Explorer useful for analyzing the anomalies.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Learning to operate the feature of **highlighting events from the Event Explorer** was easy for me.	1	2	3	4	5
I found it easy to get **highlighting events from the Event Explorer** to do what I wanted it to do.	1	2	3	4	5
My interaction with the feature of **highlighting events from the Event Explorer** was clear and understandable.	1	2	3	4	5
I found the feature of **highlighting events from Event Explorer** to be flexible to interact with.	1	2	3	4	5
It was easy for me to become skillful at using the feature of **highlighting events from the Event Explorer**.	1	2	3	4	5
I found the feature of **highlighting events from the Event Explorer** was easy to use.	1	2	3	4	5

Feature 7:

In the prototype, you were able to highlight fishing event anomalies from the Map View by clicking on the anomaly. With the highlighting of events additional information was presented to you. Please indicate how strongly do you agree or disagree with the following statements about the usefulness and ease of use of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using events highlighting from the Map View enabled me to accomplish my anomaly analysis tasks more quickly.	1	2	3	4	5
Using events highlighting from the Map View improved my anomaly analysis performance.	1	2	3	4	5
Using events highlighting from the Map View improved my productivity.	1	2	3	4	5
Using events highlighting from the Map View enhanced my effectiveness in analyzing the anomalies.	1	2	3	4	5
Using events highlighting from the Map View made it easier for me to analyze the anomalies within the data.	1	2	3	4	5
I found events highlighting from the Map View useful for analyzing the anomalies.	1	2	3	4	5

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Learning to operate the feature of **highlighting events from the Map View** was easy for me.	1	2	3	4	5
I found it easy to get the feature of **highlighting events from the Map Viewer** to do what I wanted it to do.	1	2	3	4	5
My interaction with the feature of **highlighting events from the Map View** was clear and understandable.	1	2	3	4	5
I found the feature of **highlighting events from the Map View to be** flexible to interact with.	1	2	3	4	5
It was easy for me to become skillful at using the feature of **highlighting events from the Map View**.	1	2	3	4	5
I found the feature of **highlighting events from the Map View** was easy to use.	1	2	3	4	5

Feature 8:

In the prototype, when you highlighted a fishing event anomaly (either from the Event Explorer or theMap View), contextual information about the vessel movement path before and after the anomalous event and other fishing events were shown. Please indicate how strongly do you agree or disagree with the following statements about the usefulness of this feature.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using the feature of **showing contextual data about the highlighted events** enabled me to accomplish my anomaly analysis task more quickly.	1	2	3	4	5
Using the feature of **showing contextual data about the highlighted events** improved my anomaly analysis performance.	1	2	3	4	5
Using the feature of **showing contextual data about the highlighted events** increased my productivity.	1	2	3	4	5
Using the feature of **showing contextual data about the highlighted events** enhanced my effectiveness in analyzing the anomalies.	1	2	3	4	5
Using the feature of **showing contextual data about the highlighted events** made it easier for me to analyze the anomalies within the data.	1	2	3	4	5
I found showing contextual data about the highlighted events useful for analyzing the anomalies.	1	2	3	4	5

A.6 Interview Questions

Interview Questions:

1. What was your experience with using the prototype software for analyzing anomalies between VMS data and MarFis data?
2. Using this software can you explain what kind of new information you were able to gather from the data sets?
3. Can you explain how using this software differs from how you currently analyze the anomalies between the VMS data and the landing reports in MarFis.
4. How do you think this system could be integrated within your current data analysis activities?
5. Did you experience any problems, difficulties, or confusion while using the prototype software?
6. What is your overall impression of the prototype software?
7. Do you have any comments or suggestions about how we can improve the prototype software?