Exploring Geo-Temporal Differences Using GTdiff

Orland Hoeber*
Department of Computer Science
Memorial University

Garnett Wilson†
Department of Computer Science
Memorial University

Simon Harding‡
Department of Computer Science
Memorial University

René Enguehard§
Department of Geography
Memorial University

Rodolphe Devillers¶
Department of Geography
Memorial University

ABSTRACT

Many data sets exist that contain both geospatial and temporal elements, in addition to the core data that requires analysis. Within such data sets, it can be difficult to determine how the data have changed over spatial and temporal ranges. In this design study we present a system for dynamically exploring geo-temporal changes in the data. GTdiff provides a visual approach to representing differences in the data within user-defined spatial and temporal limits, illustrating when and where increases and/or decreases have occurred. The system makes extensive use of spatial and temporal filtering and binning, geo-visualization, colour encoding, and multiple coordinated views. It is highly interactive, supporting knowledge discovery through exploration and analysis of the data. A case study is presented illustrating the benefits of using GTdiff to analyze the changes in the catch data of the cod fisheries off the coast of Newfoundland, Canada from 1948 to 2006.

Index Terms: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces; J.2 [Computer Applications]: Physical Sciences and Engineering—Earth and Atmospheric Sciences

1 INTRODUCTION

Visualizing and exploring data that include significant geospatial and temporal aspects can be challenging. Often, the complexities in the data lead to visual approaches that are either incomplete or overly complex. The focus of this research is to provide support for a specific type of analysis of such data: examining changes over space and time.

The processes for visually representing one or more data sets on a map are well-known [4]. However, many such approaches make it difficult for users to make meaningful comparisons between multiple data sets. While significant features in the data sets may be readily identified, more subtle elements may be undetectable. Even for complex geo-visualization methods that include a high degree of interactivity or animation, this problem remains.

Our goal in this research is to take advantage of the human vision system and interactive exploration to show when and where the data are changing through geo-temporal differences. By representing the differences (i.e., the regions of increase and/or decrease) in user-defined spatial and temporal bins, changes across space and time may be identified and examined. The goal is not to just provide a single view of the data that can give users the answers they are seeking, but instead to support their knowledge discovery activities through exploration and analysis of the data.

*e-mail:hoeber@cs.mun.ca

†e-mail:gwilson@cs.mun.ca

‡e-mail:simonh@cs.mun.ca

§e-mail: rene@computer.org

¶e-mail: rdeville@mun.ca

To address this problem domain, a geo-visual analytics tool (called GTdiff) was developed to enable users to explore the geo-temporal differences in the data through an intuitive and interactive interface. The data can easily be filtered both spatially and temporally, allowing users to focus their attention on a manageable range. Visual representations of differences in the data allow interesting features to be readily identified and explored further in a knowledge discovery process.

The features of GTdiff are presented in this paper as a design study. We outline aspects of geo-temporal data that make this topic interesting and challenging, and explain our design decisions for visually encoding the data and supporting interactive exploration. A small case study performed with an expert user illustrates the benefits that the approach can have in finding interesting features in the data.

The remainder of this paper is organized as follows. Sections 2 and 3 provide a brief overview of work related to geo-temporal visualization, along with a collection of visualization and interaction principles that guided this research. An outline of the features and implementation details of GTdiff are provided in Section 4. A case study describing how GTdiff can be used in an exploratory process is described in Section 5. The paper concludes with a summary of the contributions and an outline of future work in Section 6.

2 RELATED WORK

While geo-temporal data sets can be very rich, there are often complexities associated with working with them [23, 37]. Nevertheless, their use can add value to application domains including route tracking [35], correlating ocean vessel movements with weather forecasts [24], and identifying behaviours [8]. As a result of the wide-ranging applicability, there is a wealth of research literature that explores the use of geo-temporal data, with a strong focus on practical applications.

Although geo-temporal data can take multiple forms of varying complexity, the type of data we focus on analyzing in this research is point-sample data that includes latitude, longitude, and time elements. Commonly these data contain not only spatial and temporal aspects, but also other meaningful attributes that are of interest to a particular domain. The usefulness of a data set is not that it contains a particular point in space and time, but that it contains some other information that is linked to that point.

Without considering temporal aspects of the data, the visual representation of geospatial data (i.e., geo-visualization) is a very active field of study [4]. The recent increase in interest can be partially attributed to the popularization of geospatial technologies (e.g., GIS and Google Earth), an increase in the amount of data being generated by various organizations, an interest in doing something intelligent with this data, and the democratization of geospatial data [15]. Much work has been devoted to both processing [9] and representing [4] geospatial data, using a multitude of techniques. The use of geo-visualization to enhance the knowledge discovery and decision-making activities of users has also been explored in the development of systems that support geo-visual analytics [3, 20] and exploratory data analysis [2].

Some of the more notable approaches to the visual representation of geospatial data include valuable interactive techniques for manipulating and exploring the data. These include the use of multiple coordinated views [14], filtering [13], highlighting [14], details-on-demand [5], pan and zoom [24], and animation [17].

Considering again the temporal aspects of the data, a classic approach for representing geo-temporal data is the space-time-cube [16, 22]. In this method, time is placed perpendicular to 2D space in a three-dimensional cube. This results in changes in time being represented as lines with varying slopes moving from the origin to the destination of the change. This kind of representation is still used in much of today's research and has remained largely unchanged since its original inception.

Other recent work of note is that of Maciejewski et al. [27, 26]. They created a comprehensive system designed to find data aberrations (or "hotspots"). They examined the application of the system to a number of domains, including crime and syndromic surveillance. Their system implements a geo-temporal view with pan and zoom controls. Time series are shown as line plots, with regions and points of the time series used to interactively control the main geospatial temporal window. Interactive temporal tools in a separate menu include filter and aggregation control. Users can identify hotspots using heat maps and temporal contours. The exploration of the data is enhanced with interactive control of the colour scale and binning functions.

In the work of Lundblad et al. [24], ship routes and weather forecast data are represented in a single unified interface. Users can investigate the relationships in the data using linked interactions, highlighting, smooth zooming capabilities, and filtering of data. The goal of this work was to allow for the easy identification of situations where ships may become exposed to severe weather conditions.

In other work by Maciejewski et al. [28], the LAHVA system examined the pet and human health data sets for disease surveillance. The system is similar to the more recent work in that it also combined spatial exploration with history controls and time sliders. Pan and Mitra's FemaRepViz system [31] allowed users to examine the spatial placement of emergency reports as they chose particular time periods. The reports were dynamically placed at specific locations based on their textual contents.

Work by Aeschliman et al. [1] used colour to visually encode the collective movement patterns of shoppers within a retail setting using DVisRFID. The intensity of the colour represents the relative popularity of the location, and the hue of the colour represents the direction in which the people move through the location. Although their usability tests were positive, this use of colour may be difficult to visually decode when the movement is not in one of the cardinal directions (which are mapped to the colours red, green, yellow, and blue).

Although these representations address various issues associated with geo-temporal data (e.g., representing position using spatial data, showing high-dimensional data simultaneously with spatial position, visually identifying outliers), they do not directly address the issue of representing such data to show the differences or changes over time. It is this aspect of geo-visual analytics that we address in this paper. In particular, GTdiff provides a visual representation that highlights the temporal differences of numeric data at the same spatial locations, allowing users to discover where and when changes have occurred.

3 VISUAL REPRESENTATION OF DATA

While the set of methods for visually representing data is broad, it is well understood that the effectiveness of a method for encoding a particular attribute of a data set is influenced by its type [10, 12, 29]. For example, the use of colour can be very effective to represent categorical information, but is less effective for representing numeri-

cal data. To complicate matters further, once a particular method is chosen to visually encode an attribute of the data, it can be difficult to re-use it without causing confusion or misinterpretation. As a result, while there are many different data encoding choices available from which the visualization designer may choose, the resulting visualization systems often make use of both optimal and sub-optimal choices.

When the data sets include spatial attributes that are fundamentally important to understanding the data, the choices for visually encoding the remaining attributes are further constrained. That is, since it is logical to represent the spatial attributes of the data spatially, this visual feature can no longer be used to represent other aspects of the data. For example, in geo-visualization, the data attributes are visually represented at their appropriate locations on a map. Since the spatial location is being used, how the data are represented at that location must be chosen from the other visual encoding methods.

In addition to this restriction of choice for visual encoding, geo-visualization introduces further complexities that must be addressed [25]. The task of understanding geospatial data requires that users assimilate two different sets of data attributes: that which is being visually represented, and the representation of the geographic space. Providing the geographic context in a way that does not overpower the visual representation of the data can be challenging, and is often dependant on the tasks and analyses the users need to perform.

While the use of visual encoding features such as position, length, angle, area, and shape are well understood, colour hue and intensity are often misused [34]. Evidence of this can be found not only within the public domain, but also commercial products, and even the academic literature. As such, colour theory warrants further discussion within the context of the visual representation of data.

The opponent process theory of colour [18, 36] suggests that the human perception of colour is best described by arranging six elementary colours on three channels: black-white, red-green, and yellow-blue. This theory provides guidance in the selection of colour to visually encode data, including the labelling of categorical information (e.g., choosing colours that are at the extremes of the colour channels) the use of perceptually ordered colour scales (e.g., varying the colour monotonically on one or more colour channels), and the representation of data that has a true zero value (e.g., using a neutral colour for zero, and diverging colours on one or more colour channels to represent positive and negative values). Specific advice regarding the use of colour scales in geo-visualization is provided by ColorBrewer [11].

Interaction is an important aspect of modern visualization systems [32, 36]. The ability to manipulate the visual representation through interactive features such as filtering, brushing, and focusing are fundamentally important. Such interactive features enable users to explore the data and manage the visual complexity, as they seek to gain insight and knowledge from the data.

When data sets are highly complex and multi-variate, a common technique for dealing with the resulting visual complexity is to use multiple coordinated views of the data. Providing multiple views of a single conceptual entity can not only enhance the understanding of the data, but can also reduce the cognitive overhead associated with interpreting complex data [7]. The keys to using multiple views effectively are to ensure that selection, manipulations, and changes made in one view are made apparent in the other views.

4 GTDIFF PROTOTYPE DESIGN

The primary goal in the creation of GTdiff was to support data analysts in exploring and understanding how geospatial data sets change over time. The system was implemented as a Java application, using a virtual globe generated by World Wind [30] for the
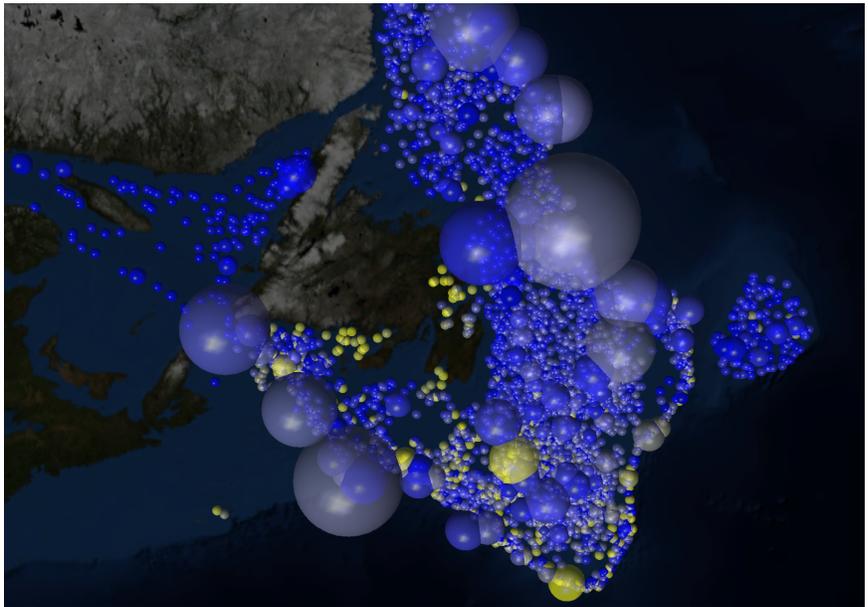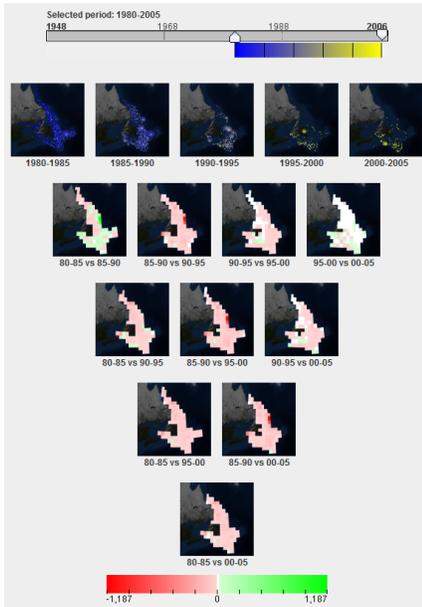
Figure 1: The main visual components of GTdiff include a temporal view (top portion of the left screenshot), a difference view (bottom portion of the left screenshot) and a geospatial view (right screenshot). The data shown is from the cod fisheries off the coast of Newfoundland, Canada. The data is filtered to a 25-year timeframe, divided into 5-year temporal bins. The yellow-blue colour encoding represents the data in the temporal bins; the sizes of the spheres are proportional to the mass of the catch at each location; the red-green colour encoding represents the changes in the difference graphs.

framework upon which the core geo-visual representations are layered. The design of the visual and interactive features were strongly influenced by the aspects of visually representing data described in the previous section. The interface of GTdiff includes three primary visual elements: the temporal view, the difference view, and the geospatial view (see Figure 1). Although the methods used to visually represent where and when the data are changing have been previously presented [19], here we focus on how these views and their associated features are tightly integrated, providing support for the interactive exploration of the data.

4.1 Temporal View

The key interactive feature that the temporal view provides is the ability for users to filter the data temporally, aggregating what remains into a user-specified number of equal length temporal bins. For example, a user may wish to filter the data set to only include a five-year timeframe, and then group the data into five one-year bins. Alternately, a user may be interested in twelve years of data, grouped into six two-year bins. Since the goal of GTdiff is to allow users to understand how the data are changing over time, it is necessary to aggregate the data into manageable units of time. As will be shown in the description of the difference view, doing so allows GTdiff to clearly illustrate changes that have occurred between the temporal bins.

A perceptually ordered colour scale of mid-level brightness on the yellow-blue colour channel is provided within the temporal filter. This colour scale is used to label the data in the temporal bins, and provides a persistent visual legend to support users in the decoding process. For the set of temporal bins within the selected temporal range, equally spaced colours are selected and assigned to the bins in order. While the use of colour is not an optimal method for visually differentiating the data contained in one temporal bin from the data contained in another, few options remain for visually differentiating multiple data sets once the data is represented spatially.

Each temporal bin is displayed side-by-side under the temporal filter, using the colour encoding described above. These representations show a zoomed-out geospatial view of the data. Their purpose is to support users in visual scanning and comparison activities, and to allow users to select one or more temporal bins to investigate further within the geospatial view. The value of a specific attribute of the data under investigation (e.g., product sales, biomass of fisheries catch data, green house gas production) is encoded as either the area of a circle or the volume of a sphere, placed at the specific location of the data point, using the colour assigned to the temporal bin.

The choice as to whether the data is represented as 2D circles or 3D spheres is user-configurable depending on the source of the data. Matching the visual encoding as closely as possible to the end-user's understanding of the meaning of the data [33] can increase the ability of the user to understand and make sense of what they are seeing. Since the data set used within this paper represents the mass of fish caught at specific locations, representing this data as 3D spheres is more logical from the user's perspective than the alternative of 2D circles.

Whether the data is represented as spheres or circles, the visual encoding allows users to readily make comparisons between the numeric data at different locations and in different temporal bins. The objects are rendered semi-transparently in order to address the occlusion problems that occur when a large object covers one or more smaller objects at nearby locations. When the data is encoded using spheres, a simple shading model enables the proper perception of their 3D shapes.

4.2 Difference View

The difference view provides a visual representation of the difference between each pair of temporal bins in the form of a set of difference graphs. The difference graphs are organized in an inverted pyramid, where the top layer shows the difference graphs for neighbouring pairs of temporal bins, the second layer shows the difference graphs for pairs of temporal bins with a one-bin gap, and so
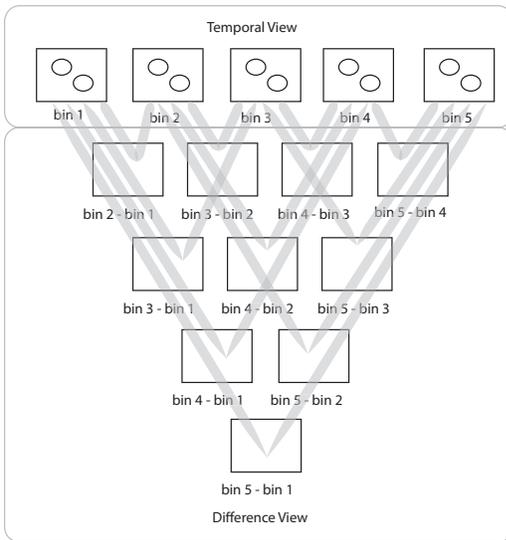
Figure 2: The difference view provides a visual comparison of every pair of temporal bins, arranged in an inverted pyramid.

(a) The raw approach adds the data in the spatial bins.

(b) The normalized approach averages the data in the spatial bins.

Figure 3: Examples of the spatial binning and associated colour encoding of the differences employed in GTdiff. Note that the numbers listed within the figures represents the raw data at the associated locations.

on until the final layer shows the difference graph between the pair of temporal bins at the extremes of the temporal range. As such, every possible pair-wise comparison of temporal bins is shown simultaneously within the difference view (see Figure 2). This organization encodes the source data for each difference graph within its spatial location within the inverted pyramid. In addition to being able to see where and when the differences are occurring, the user can also visually inspect and compare the raw data shown in the corresponding temporal bins at the top of the inverted pyramid, a method which is common in exploratory data analysis domains [2].

Since the goal of the difference graphs is to allow users to perceive how the data are changing between the temporal bins, spatial binning is necessary. Without spatial binning, the only situation in which showing the differences would have meaning is when data points are at the exact same spatial location. For example, if there are two data points at the same location with the same value, but in two different temporal bins, then these will balance out and show no change. However, if the two points are at slightly different locations, then it will look like there is a decrease at one location and an increase at the other. Spatial binning groups data points that are near one another to avoid this situation.

The current GTdiff prototype performs spatial binning by dividing the space over the geographic coordinate system (e.g., using latitude/longitude). For each spatial bin, the difference between the values in each pair of temporal bins is calculated. The maximum of the absolute value of all of the differences is used as the extreme value in a positive/negative scale. A divergent colour scale is used to visually encode these differences within the visual representations of each spatial bin, following a method similar to that in [21]. White represents a value of zero (no change), the degree of saturation of green is used to represent positive values, and the degree of saturation of red is used to represent negative values. In order to assist with decoding, a labelled legend is provided at the bottom of the difference view.

When aggregating the data within the spatial bins, users may choose to either combine the raw data, or first normalize the data. The choice of whether to normalize the data depends on the domain in which the system is being used. If the data represents complete and accurate information (e.g., the value of coffee sales within a region) then calculating the total in the spatial bins using the raw data is appropriate. However, if the data represents point
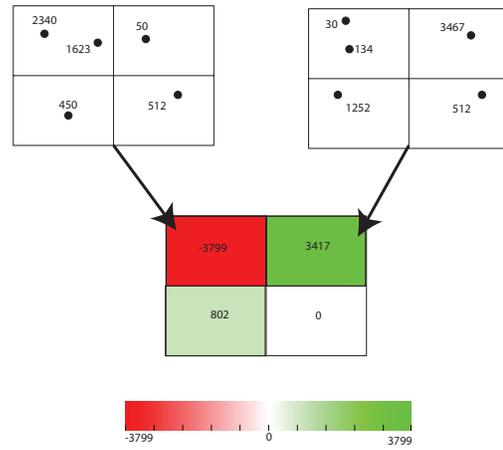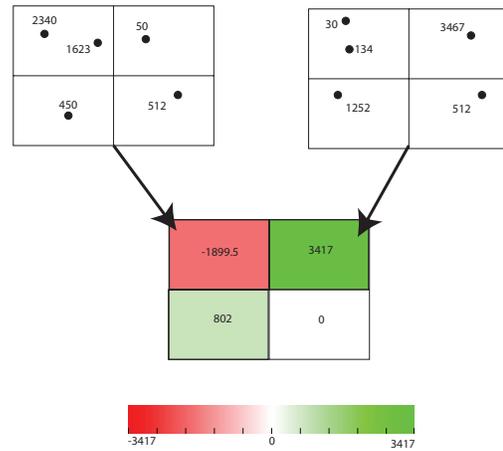
samples of physical phenomena (e.g., fisheries catch data, or population samples), then normalizing the data first is more accurate. A simple example illustrating the spatial binning approach, and the corresponding encoding of the differences using colour, is provided in Figure 3. One can readily identify both the spatial regions and magnitude of the changes.

Spatial binning introduces two potential issues. The first of these is when there are significant data points near the boundaries of the spatial bins. In this case, the data points may be split between the spatial bins, making comparisons between pairs of temporal ranges difficult. The second problem is when a spatial bin overlaps a location where it is not possible for data to exist (i.e., fisheries data over land, population data over the ocean, etc.). In order to alleviate these problems, GTdiff provides users with control over the resolution of the spatial binning. Examples of different resolutions can be seen in Figure 4.

By visually scanning the collection of difference graphs within the inverted pyramid, users can quickly and easily perceive significant changes in the data, both spatially and temporally. If users wish to investigate a specific difference graph in more detail, it may be selected and explored further in the geospatial view.
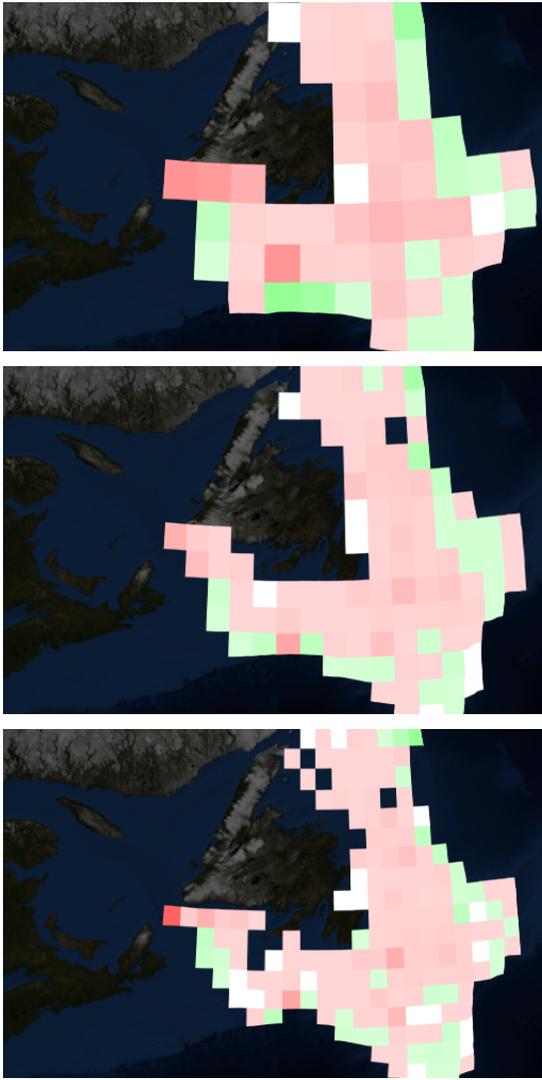
Figure 4: Adjusting the resolution for the spatial binning results in difference graphs that range from coarse-grained to fine-grained comparisons of the data.

4.3 Geospatial View

The primary purpose of the geospatial view is to provide a detailed visual representation of selected aspects of the data in the context of their spatial location. As users explore the data in the temporal and difference views, they may wish to investigate specific elements further. One or more temporal bins may be selected in order to conduct a detailed analysis of the data. Or a specific difference graph may be selected for detailed inspection. In either case, the data is layered over satellite imagery in order to provide users with spatial awareness of the data.

In order to enhance the ability of users to perceive the foreground data from the background contextual information, the system was designed to ensure sufficient luminance contrast between the two. The satellite imagery that makes up the background is darkened by placing a semi-transparent layer over top of it. Bright colours were chosen to represent the foreground information (i.e., the spheres in the temporal bins or the squares in the difference graphs). As such, the foreground information can be readily perceived as being a separate layer placed over the background information.

When selecting two or more temporal bins to be shown simultaneously in the geospatial view, the spheres are rendered using the colour encoding associated with their source temporal bin. While the semi-transparency of the spheres allows the user to see smaller spheres that are embedded within larger ones, the side effect of using transparency is that the colour of a sphere that is encapsulated within a larger sphere is affected, making the decoding of the colour to the associated temporal bin difficult in some situations. The ambiguity of the temporal bin to which a particular sphere belongs can be eliminated by interactively showing or hiding specific temporal bins. Further, the underlying data can be inspected and examined through focusing and brushing operations as needed.

The geospatial view only supports the selection of a single difference graph at one time, since the meaning of displaying multiple difference graphs is not obvious. As with the viewing of the temporal bins, the goal is to allow users to analyze and explore the data in greater spatial detail. As such, focusing and brushing operations are also available when viewing a difference graph, allowing the raw data to be inspected as needed.

The geospatial view also supports the spatial filtering of data through pan and zoom operations. As the users manipulate the location and scale of the map, all other geospatial representations are updated in both the temporal view (i.e., temporal bins) and the difference view (i.e., difference graphs). In this way, users can manipulate the map in the geospatial view, explore the features of the data within the other views at the desired location and level of detail, and then make further selections of temporal bins and difference graphs in support of their knowledge discovery tasks.

5 CASE STUDY

In order to enable a discussion of the geo-visual analytic capabilities of GTdiff, a case study using historical data for the cod fisheries off the coast of Newfoundland, Canada is presented. This case illustrates the value of providing data analysts a visual approach to exploring how the data are changing over space and time.

5.1 Cod Fisheries in Newfoundland

Fisheries ecology and fisheries management require that domain experts know not only where fish are located and in what quantity, but also understand complex biological processes that have a strong spatial and temporal component. Part of this understanding is reached by mapping fish species distribution and abundance, and comparing it with the same data available for the previous years in order to identify potential changes. This process is normally done using static maps generated for separate years. Such a process limits the ability of fisheries experts to understand and build hypotheses about complex phenomena as it does not allow an exploration of the data that could help the expert to compare regions and years.

For this reason, the GTdiff prototype has been tested using an extensive fisheries data set that has both large spatial (about 1,000,000 km^2) and temporal (1948-2006) extents. The data was compiled from annual multi-species bottom trawl surveys conducted by Fisheries and Oceans Canada for the region of Newfoundland and Labrador, Canada. This data set records information about a number of different species, but only one was used to test the prototype. Atlantic cod (*Gadus morhua*) is a species that offers the most data and the longest temporal extent within the data set. Cod is also a species that played a critical role in fisheries management in Eastern Canada as it used to be very abundant but collapsed and led to a moratorium on cod fisheries in 1992-1993 that severely impacted the local economy. The years preceding the collapse have been documented by fisheries experts and showed interesting spatial and temporal characteristics that could benefit from examination using a system like GTdiff. The data set has a number of attributes, including the mass of cod caught by the survey vessels at various locations and times.

Fisheries data also have inherent complexities that make them interesting for this case study. Data collection followed a randomly stratified sampling scheme which ensured a representative coverage of the study area, allowing robust statistical analyses. However, point data are samples which are not always representative of the phenomenon being observed, their number being limited by the resources that can be spent to collect the data (e.g. number of days at sea to cover a given region). Specific point locations also change from one year to the next, as the random sample of sites is done separately each year. In addition, a number of problems can occur during the data collection (e.g. ship engine failure, large number of days with bad weather) that can lead to significant gaps in the spatial coverage of the data that should not be interpreted as a lack of fish.

5.2 Fisheries Analyst Tasks

Bottom trawl surveys such as the ones used for this case study are mainly performed by the government in order to define quotas to be allocated for commercial fisheries in the next year (i.e. the amount of fish from each species that can be fished in certain zones). Geospatial tools are still marginally used in day-to-day operations of fisheries analysts which mostly rely on statistics and mathematical population models, usually aggregating data in large spatial management units (e.g. NAFO Zones in the northwest Atlantic) and comparing the average catches with previous years. However, geospatial tools are penetrating this field as space is increasingly being recognized as a neglected component when it comes to understanding processes regulating changes in fisheries resources. Static maps representing the location and abundance of catches (e.g. proportional symbols maps) are produced routinely in order to look at spatial variations within management units. Comparing a map for one year and one for another year can provide an insight into changes at given locations and sometimes into movement of fish populations or overexploitation of fish stocks. Analysts also tend to aggregate and bin different years in order to compare groups of years which can highlight long-term trends in the data. The ability to analyze data both in space and time is important due to the dynamic nature of the fisheries.

5.3 Using GTdiff for Data Exploration

GTdiff has been developed and tested in collaboration with an expert in fisheries data mapping and analysis. An attempt was made to visualize the cod catch data before, during, and after the major collapse of the cod stocks (1988 to 1993), as this period had significant changes over space and time that have been well documented by biologists [6]. Different temporal blocks and grid sizes were selected, allowing for the identification of visual patterns in the difference graphs that have been described in the fisheries literature. The expert found the temporal-based analysis of the data valuable in identifying this well-known pattern. However, further study is needed to determine the benefits that GTdiff can provide for an analyst to discover new knowledge from within a geo-temporal data set.

Figure 5 shows the temporal and difference views of the data using five temporal bins in the 1988-1993 timeframe. Three difference graphs were quickly identified by the expert that show the main steps that led to the cod moratorium. The difference graph comparing 88-89 vs. 89-90 (see Figure 6(a)) shows a decrease in the cod catches in some regions, and increases in others. In addition, there were significant decreases in some northern regions, and significant increases in some northeastern regions. These decreases continued to spread to the southeast, which can be identified in the 89-90 vs. 90-91 difference graph (see Figure 6(b)). At the same time, a number of much higher catches were observed in the east of the region, due to what has been described by biologists as the hyper-aggregation of cod. This increased catch gave the impression
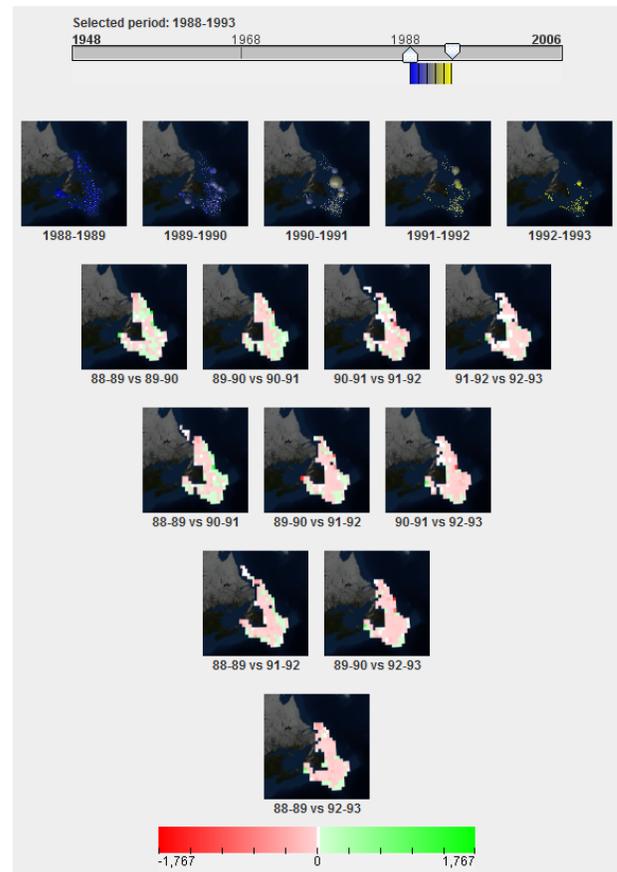
Figure 5: A view of the cod fisheries data over a five-year timeframe. The catch data in each of the years is shown in the temporal bins across the top. The expansion of the fisheries can be identified by the green regions in the difference graphs; the subsequent depletion of stocks can be identified by the red regions.

to the industry and some biologists that cod stocks were healthy and led to an intense fishery in this region. The 90-91 vs. 91-92 difference graph (see Figure 6(c)) shows not only a depletion of the stocks in the regions that had high catch rates in the previous years, but also widespread reductions in most other regions as well. The result of these reductions led to a decision to enact a moratorium on cod fisheries in the summer of 1992. Viewing the data from the extremes of the temporal range (see the bottom difference graph in Figure 5), the effects of the depletion of the cod stocks can be seen.

5.4 Discussion

In the case study, the expert user was provided with a short training session and description of the meanings of each of the visual representations. He was able to quickly filter the data spatially and temporally using GTdiff, producing views of the collapse of the cod fishery around Newfoundland. Even though the spatial and temporal parameters pertaining to the collapse were known by the expert, users unfamiliar with the system could readily identify that a drop in cod stocks had taken place during the time over which all data were available (e.g., the 25 year timeframe in Figure 1 clearly shows a severe decline in the data beginning in the early 1990s).

The expert selected a known five-year timeframe and chose to produce five temporal bins, one corresponding to each year. Different granularities of spatial binning were then explored. While an increased resolution of the difference graphs produced an indica-
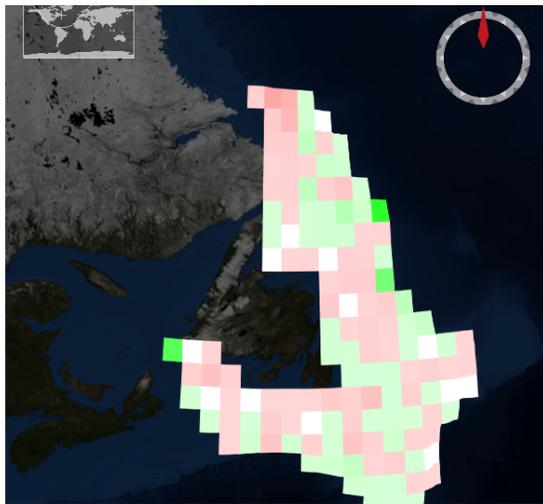
(a) Difference graph for 88-89 vs. 89-90.
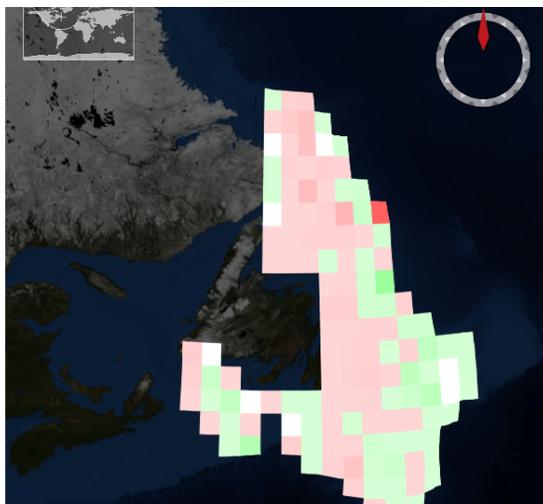
(b) Difference graph for 89-90 vs. 90-91.
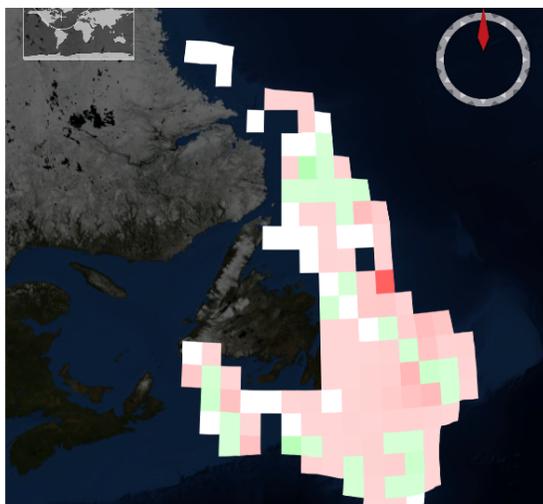
(c) Difference graph for 90-91 vs. 91-92.

Figure 6: The important difference graphs discovered in the case study, shown in detail within the geospatial view.

tion of the differences in catch data over a more precise region, the expert opted to use lower to mid-range spatial bin resolutions. The benefit in choosing the lower resolution was that broader, overarching changes became visible that were not evident with finer resolutions. He also found the ability to investigate specific difference graphs using the geospatial view to be very useful in examining the ecological events leading up to the cod moratorium. In particular, the higher catches associated with the hyper-aggregation of cod were examined in great detail, as described in the previous section.

There is an important difference between how this data has been explored in the past and how it was explored by the expert using GT-diff. As noted previously, traditional approaches to analysis would be for the expert to generate multiple maps that illustrate the location and abundance of catches, and then make comparisons of this data using these snapshots of the data. GTdiff provides support for a more interactive and dynamic exploration of the data. Analysts can adjust the spatial and temporal filtering and binning, identify features of interest within the difference pyramid, and investigate these in detail as necessary. As such, GTdiff does not just provide snapshots of the data, but tools to analyze and explore the changing aspects of the data over space and time.

While we examined the case study of the Newfoundland cod fishery in particular, GTdiff may be a beneficial tool for exploring any data set involving a strong connection between geospatial and temporal elements. From a general usage perspective, the system allows data analysts to focus on particular time periods in a data set for a geospatial region by manipulating the geospatial view and associated temporal view. They can quickly assess the changes that have taken place over that period from longer to shorter time intervals (difference view), and further examine these at an appropriate level of detail (geospatial view). Natural choices for GTdiff application domains include population statistics (e.g., examining the growth and decline of urban and rural areas), and business intelligence (e.g., studying consumer purchasing patterns or product adoption rates), among many others.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we present a geo-visual analytics system that supports an interactive exploration of changes in a data set through geo-temporal differences. Spatial and temporal filtering allow data analysts to easily focus on specific locations and timeframes of interest. Visual representations of the data in both temporal bins and difference graphs simultaneously show the raw data and how it is growing or shrinking within spatial bins. Multiple coordinated views support the analyst in examining the data from multiple different perspectives, focusing on interesting aspects of the data as needed. At any time, the underlying data can be inspected in detail through simple focusing and brushing operations.

A case study using data from the cod fisheries off the coast of Newfoundland between 1948 - 2006 was conducted with an expert user. The system was found to be useful for both exploring the data, and for showing and explaining known phenomena. The expert was able to quickly grasp the meaning of the visual representations, the value of the specific features, and the methods for interactively exploring the data.

Future work includes further refinement and enhancements of the prototype implementation, and exploring methods for supporting the visual identification of more complex trends in the data. User evaluations in a controlled laboratory setting are currently in the planning stage, the goal of which are to measure the benefits of the specific design choices in the creation of GTdiff. Field trials with expert and novice users are also being planned, in order to gain an understanding of how GTdiff can be used to explore, discover, and explain both known and unknown phenomena.

This work was conducted as part of a larger project that focuses on discovering and evaluating techniques to support geo-visual an-

alytics of capture fisheries statistical data. As such, portions of GT-diff may be integrated into future research prototypes developed within this broader research domain.

REFERENCES

[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] G. Andrienko and N. Andrienko. *Exploratory Analysis of Spatial and Temporal Data - A Systematic Approach.* Springer, 2005.

[3] 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.

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

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

[6] D. B. Atkinson, G. A. Rose, E. F. Murphy, and C. A. Bishop. Distribution changes and abundance of northern cod (gadus morhua), 1981-1993. *Canadian Journal of Fisheries and Aquatic Science*, 54(Suppl. 1):132–138, 1997.

[7] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the ACM Advanced Visual Interfaces*, pages 110–119, 2000.

[8] F. Bartumeus and J. Catalan. Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, 43:1–12, 2009.

[9] M. A. Bayir, M. Demirbas, and N. Eagle. Discovering spatiotemporal mobility profiles of cellphone users. In *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile, and Multimedia Networks*, pages 1–9, 2009.

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

[11] C. Brewer and M. Harrower. ColorBrewer 2.0: color advice for cartography. http://colorbrewer2.org/, 2010.

[12] W. S. Cleavland and R. McGill. Graphical perception: Theory, experimentation, and application of the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[13] U. Demšar, S. Fotheringham, and M. Charlton. Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics. *Information Visualization*, 7(3):181–197, 2008.

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

[15] M. F. Goodchild. Towards a geography of geographic information in a digital world. *Computers, Environment and Urban Systems*, 21(6):377–391, 1999.

[16] T. Hägerstrand. What about people in regional science? *Papers of the Regional Science Association*, 24:7–21, 1970.

[17] M. Harrower and S. Fabrikant. The role of map animation in geographic visualization. In M. Dodge, M. Turner, and M. McDerby, editors, *Geographic Visualization*. Wiley and Sons, 2008.

[18] E. Hering. *Outlines of a Theory of Light Sense (Grundzüge der Lehre vom Lichtsinn, 1920)*. Harvard University Press, 1964.

[19] O. Hoeber, G. Wilson, S. Harding, R. Enguehard, and R. Devillers. Visually representing geo-temporal differences. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 229–230, 2010.

[20] D. A. 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*, LNCS 4950, pages 154–175. Springer, 2008.

[21] D. A. Keim, T. Nietzschmann, N. Schelwies, J. Schneidewind, T. Schreck, and H. Ziegler. A spectral visualization system for analyzing financial time series data. In *Proceedings of Eurographics/IEEE-VGTC Symposium on Visualization*, pages 195–202, 2006.

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

[23] M. P. Kwan. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8(1-6):185–203, 2000.

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

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

[26] R. Maciejewski, S. Rudolph, R. Hafen, A. M. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, M. Wade, and D. S. Ebert. Understanding syndromic hotspots–a visual analytics approach. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 35–42, Oct. 2008.

[27] R. Maciejewski, S. Rudolph, R. Hafen, A. M. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, M. Wade, and D. S. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):205–220, 2009.

[28] R. Maciejewski, B. Tyner, Y. Jang, C. Zheng, R. 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, Oct. 2008.

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

[30] NASA. World Wind Java SDK. http://worldwind.arc.nasa.gov/java/, 2010.

[31] C.-C. Pan and P. Mitra. FemaRepViz: Automatic extraction and geo-temporal visualization of fema national situation updates. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 11–18, Oct. 2007.

[32] R. Spence. *Information Visualization*. Prentice Hall, 2nd edition, 2007.

[33] M. Tory and T. Möller. Rethinking visualization: A high-level taxonomy. In *Proceedings of IEEE Symposium on Information Visualization*, pages 151–158, 2004.

[34] E. Tufte. *Envisioning Information*. Graphics Press, 1990.

[35] U. D. Turdukulov, M. J. Kraak, and C. A. Blok. Designing a visual environment for exploration of time series of remote sensing data: In search of convective clouds. *Computers & Graphics*, 31(2):370–379, 2007.

[36] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition, 2004.

[37] 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.