

Interactive Web-Based Visualization for Lake Monitoring in Community-Based Participatory Research: *A Pilot Study Using a Commercial Vessel to Monitor Lake Nipissing*

Mark P. Wachowiak¹, Renata Wachowiak-Smolíková¹, Brandon T Dobbs^{1,2}, James Abbott² & Daniel Walters²

¹ Department of Computer Science and Mathematics, Nipissing University, Canada

² Department of Geography, Nipissing University, Canada

Correspondence: Mark P. Wachowiak, Department of Computer Science and Mathematics, Nipissing University, Canada. E-mail: markw@nipissingu.ca

Received: December 22, 2014 Accepted: January 15, 2015 Online Published: March 30, 2015

doi:10.5539/ep.v4n2p42

URL: <http://dx.doi.org/10.5539/ep.v4n2p42>

Abstract

Environmental and limnological monitoring is of interest to government agencies, researchers, and the general public. In communities that rely on and are heavily affected by lakes and their watersheds, accessible and intuitive presentation of lake properties influences and aids decision-making, interventions, and the formulation of environmentally sound policies. In this paper, interactive web-based visualizations are employed as a mechanism to communicate environmental information collected from a commercial cruise vessel. A pilot study is presented for monitoring Lake Nipissing, a large culturally and environmentally important lake in northeastern Ontario, Canada. This example of community-based participatory research suggests that: (1) policy makers and researchers can quickly gain insight into what is happening in the lake through visualizations, which helps to direct subsequent, detailed investigations; and (2) through accessible, visual presentation, community members may be encouraged to become involved in contributing to environmental policies that directly affect them, thereby supporting environmental “citizen science”.

Keywords: lake monitoring, visualization, citizen science, community based participatory research

1. Introduction

1.1 Problem

Community-based participatory research (CBPR), including “citizen science”, has gained increasing prominence as a means of addressing challenges in data collection and transparency (Bonney et al., 2009; Newman et al., 2010). CBPR has been defined as research undertaken as a partnership between community members, academic researchers, and other organizations (Israel, Schulz, Parker, & Becker, 1998), whereas “citizen science” primarily refers to avocational research, where research goals are partly obtained by distribution of tasks to non-professionals or amateurs (e.g. the SETI@home or FOLDIT@home initiatives) (Hand, 2010). Both terms imply multiple stakeholders with differing goals and expectations. Furthermore, the research questions themselves can be posed through top-down (scientist-driven) or bottom-up (community-driven) processes (Danielsen et al., 2009). In addition to providing an increased amount of data (Crall et al., 2010; Kéry et al., 2010), stakeholder data collection can better reflect human-environment interactions (Kanjo & Landshoff, 2007; Weckel, Mack, Nagy, Christie, & Wincorn, 2010). Research involving stakeholders can also lead to increased public awareness of environmental issues (Kanjo & Landshoff, 2007) and increased participation in further research (Sullivan et al., 2009). However, the published literature largely focuses on the role that non-scientists can play within the collection of mostly quantitative data within limited parameters [but see (Viegas, Wattenberg, Van Ham, Kriss, & McKeon, 2007) for an example of unstructured collection of qualitative data]. The inclusion of qualitative data collection by non-scientists can also provide a better understanding of complex environmental processes, as well as the perceptions of stakeholders (Abbott & Campbell, 2009; St. Martin, 2001). Moreover, the literature is largely silent about how participatory science can improve policy dialogue beyond improving data quality and transparency.

Among the many research areas covered by citizen science and CBPR, environmental monitoring and ecology (Conrad & Hilchey, 2011; Hochachka et al., 2012; Tulloch, Possingham, Joseph, Szabo, & Martin, 2013; Voinov

& Gaddis, 2008) and the impact of environmental factors on public health (Eghbalnia et al., 2013; Israel et al., 2005; Miller et al., 2013) are prominent. In particular, environmental monitoring is complex, as it relies on multimodal data from sensors, images, and empirical observation. Furthermore, the large quantities of (often multi-dimensional) data acquired from various sensors motivate the need for accessible, interactive visualizations.

The current work focuses on CBPR research and visualization for lake monitoring, where data are collected using a sonde sensor affixed to a recreational cruise vessel, the *Chief Commanda II*. Although the concepts presented in this paper apply to aquatic monitoring in general, the specific target of this research, presented as a pilot study, is Lake Nipissing, the third-largest lake in Ontario, Canada. The lake, which drains into the Georgian Bay of Lake Huron, provides important economic, cultural and ecosystem services to residents and visitors. There is growing concern about declining fisheries, increased blue-green algae blooms, and the appearance of invasive aquatic species (Morgan, 2013; Hutchinson, Karst-Riddoch, & Köster, 2010; Filion, 2011). Because of its importance, the lake is the subject of several ongoing scientific studies of its biophysical properties. These studies involve measuring several parameters in different sections of the lake, generating large quantities of data that must be processed and analyzed. Applying the results of visualization and human-computer interaction studies, these data are subsequently presented through interactive web-based visualizations. The emphasis of the current work is therefore transparent data presentation and interaction.

Major stakeholders include environmental researchers, government policy-makers, First Nations communities, businesses – particularly the fisheries industry – and the general public in North Bay (pop. 54 000) and other communities that surround Lake Nipissing. Researchers, for instance, need to establish baseline environmental indicators, such as oxygen level and temperature. The public can gain an understanding of the data being collected and why they are collected, and how they may contribute potentially important qualitative data, such as unusual environmental events that may be overlooked by conventional surveys. Policy-makers are concerned about fish populations, particularly walleye (Morgan, 2013), as well as the potential effect of three dams on the French River, the primary outflow of the lake.

This research emphasizes the need to enhance data collection, to expand stakeholder participation in the research and policy process, and to disseminate results to the public through intuitive, interactive visualizations. To achieve these goals, an online platform was developed wherein sensor data can be displayed after acquisition. In turn, environmental data can be augmented by qualitative and quantitative input from other stakeholders. This two-way flow of information via the online platform serves several purposes: it allows stakeholders, on a real-time basis, to review the data, and to understand why they are collected; it complements standard data collection by providing additional data input; and it establishes common ground between scientists, policy makers and other stakeholders. The current study encompasses both CBPR, in that data are collected as a collaborative effort from the crew of a commercial cruise vessel, and citizen science, where not only government policy makers and academic researchers can access the data, but where all community members can make use of interactive visualizations and draw conclusions, or relate the information to personal experience.

1.2 The Importance of Lake Monitoring

Many large lakes are of vital significance to local communities, and convey their own important benefits and present their own unique challenges and problems. In the case of Lake Nipissing, the lake experiences considerable variability in biological productivity both within and between years; concerns have been raised regarding apparent declines in fish population and more frequent blue-green algae blooms (Morgan, 2013; Hutchinson et al., 2010). While this may be due in part to the inherent variability of the lake's ecosystem, growing levels of human activity within the region, as well as broader changes in water level and temperature associated with climate change, are also possible factors. In some cases, Lake Nipissing has been affected by human activity, such as modification of tributaries, eutrophication, and the introduction of invasive species (Filion, 2011).

1.3 The Importance of Interactive Visualization in Environmental Monitoring and Citizen Science

Accessible, web-based visualization technologies can stimulate scientific innovation for both researchers and members of the community who participate in the research and in environmental monitoring. Such visualizations encourage non-scientists to develop new questions, and scientists to address questions in new ways, given the availability of citizen participation (Goodchild, 2007). Specifically, interactive browser-based visualization and analysis allow participants and scientists to explore the data with fewer constraints, thereby enhancing innovation (Newman et al., 2012). For instance, visualization, visual communication tools, and immersive virtual experiences are encouraging community involvement in sustainability efforts aimed at assessing climate change

effects (e.g. <http://cirs.ubc.ca/research/research-area/visualization-tools-and-community-engagement>).

Such highly interactive visualization techniques are increasingly used in geographical visualization (Dykes, MacEachren, & Kraak, 2005). Utilizing a variety of interactive methods allows data to be explored more thoroughly, thereby aiding understanding. For instance, multiple graphs and maps showing related data can be linked together such that an operation performed on one graph or map will be reflected on the others, creating a dynamic connection between the information and the visualizations (Theus, 2005). Especially for representation of geospatial data, maps and other visualizations often require simplification so that a high level understanding of fundamental concepts (e.g. datums, projection, resolution) is not necessary for participants to successfully use the site (Newman et al., 2010). This principle was employed in the present study, where the data collected from the *Chief Commanda II* are presented using multiple visualizations linked through an interactive map. Users can interact with a variety of different visualizations (described below) in such a way that they can explore different spatial, temporal, properties and other aspects using multiple visualization mechanisms. Such a “linked approach” is conducive to understanding and successful use by people who are not trained in geographic information systems or in spatial representation methods (Newman et al., 2010).

1.4 Interactive Visualization of Multidimensional Transect Data

Data were collected at regularly-sampled intervals, usually every minute, every three minutes, or every four minutes depending on the transect (a *Chief Commanda II* cruise where data were collected), and can therefore be considered as continuous. As cruise durations are generally less than three hours, and there may be hours or even days between each transect, the temporal dimension is described as discrete. Finally, the spatial coverage of each transect is described as continuous because the GPS collects positions at regular intervals. Due to the hybrid nature of the data, alternative graphing approaches are used: (1) An interactive overlay map for visualizing properties with high spatial resolution; (2) Parallel coordinate plots for communicating high dimensionality (i.e., the different properties collected by the sensors attached to the cruise vessel), providing both high temporal and spatial resolution; (3) Cloud lines, which are new, sophisticated scatter plots for displaying long-term trends in properties over time, allowing users to probe how properties change as a function of distance from a specific reference point; (4) Hovmöller (space-time) plots that provide a broad representation of changes in properties over time and space, targeted primarily to the research community and environmental organizations; and (5) Visualization of descriptive statistics.

1.4.1 Interactive Overlay Map

Easily understood maps can be overlaid with data to greatly aid understanding and ease of interpretation (Hochachka et al., 2012). The primary component of the linked visualizations is a map of Lake Nipissing overlaid with paths representing each transect taken by the *Chief Commanda II*. The cruise paths are by default coloured such that the colour at any particular coordinate reflects the value for a chosen attribute at that point along the path. The data can be coloured based on any chosen property, using one of several different colour maps. These include specialized colour maps, such as a greyscale colour map, colour maps for users with colour-deficient vision, and the standard pH scale, which can be adjusted to show small differences which normally occur in lake monitoring (e.g. setting 5 to red/acidic, 8 to purple/alkaline). Groups of points can be selected on the map, and these spatial selections will be reflected throughout the rest of the visualization.

1.4.2 Parallel Coordinate Plots

Another facet of the linked visualization is a parallel coordinates plot (PCP), where n parallel axes, each corresponding to a dimension of the data, are drawn. A single n -dimensional tuple is represented by a polyline (a piecewise connected line) that intersects each axis at the location corresponding to the tuple's value for that axis' dimension (Streit et al., 2006), allowing all dimensions of multidimensional data to be displayed simultaneously. The PCP used in the linked visualization assigns a parallel axis to each of the measurements collected in the study. PCP can be used for an arbitrary number of dimensions. It can also indicate when data fall outside the range of the sensor by showing these values below their corresponding coordinate axis (Siirtola & Räihäm, 2006).

Many enhancements have been made to the standard parallel coordinates technique. In the current study, a 3D plot was implemented so that every parallel axis was extended along the z -axis into a plane. The different *Chief Commanda II* transects are separated along the z -axis, clearly indicating the date of each collection (Johansson et al., 2014). Opacity controls were provided to reveal the density distribution of the graph, as polylines in dense areas are more visible than in less dense areas (Wegman & Luo, 1996). The order of the parallel axes can be changed to better observe relationships between two arbitrarily juxtaposed axes (Blaas et al., 2008).

The polylines can also be grouped into k categories (Johansson et al., 2005), with each polyline coloured based on the cluster to which it belongs. The clustering can be performed based on any combination of m properties, using some distance measure between the m -D data points (in the current study, Euclidean distances). Although many sophisticated clustering algorithms exist, the standard iterative k -means clustering approach was used in this work.

Groups of polylines can be selected directly on the plot by “brushing” (Edsall, 2002), wherein users draw a “brush line” over the plot to highlight polylines. Consistent with the goal of linked visualization, the selection (“brushed” polylines) is reflected throughout the other visualization components (i.e. the map and the statistics table). The polylines in the PCP are also coloured based on the property chosen for the map, with each polyline’s colour matching that of its corresponding point on the map.

1.4.3 Cloud Lines

To represent time, a discrete measure due to the temporal distance between transects, a visualization based on cloud lines was employed (Krstajic et al., 2011). Cloud lines are essentially scatter plots where dependent variables (sonde properties) are each shown as single lines, as a function of the independent variable (time). Events (the sonde readings) along the time axis are represented as circular markers that are coloured according to their value, or whose radius and/or opacity reflects the value, resulting in lines of varying width and colour. The advantage of this technique is that it allows a large amount of data to be intuitively displayed with a relatively small amount of screen space. Initially proposed for discrete, episodic data (Krstajic et al., 2011), cloud lines were adapted to continuous data in the current study to convey properties represented in a straight line, whose colours vary over time according to their value. To represent the spatial dimension, the radius of each point in the cloud line represents the proximity (inverse distance) of that spatial coordinate from a user-specified point, where readings closer to the selected point appear larger. Cloud line plots for multiple properties can be shown simultaneously, and can be juxtaposed arbitrarily using a dragging mechanism.

1.4.4 Space-Time Plots

Hovmöller (space-time plots) present a standard image of how properties change over space (y -axis) and time (x -axis). The image colour indicates the value of the property currently being displayed (Buzzelli, Ramus, & Paerl, 2003). To make these diagrams more useful, space is represented as distance from a user-selected point on the map, allowing users to view changing properties based on specific geographic locations.

The time sparsity of the transects in this pilot study posed challenges for space-time interpolation, as the data are sparse in time but not in space (Wright & Goodchild, 1997). Therefore, the natural neighbour method based on Voronoi methods was employed (Gold & Condal, 1995; Wright & Goodchild, 1997).

1.4.5 Statistics

Two statistics tables and corresponding graphs are also available. The statistics tables show the mean, standard deviation, median, interquartile range, minimum, and maximum values of every property. The first table shows these statistics for the data points that are currently selected in the linked visualizations, and is updated dynamically whenever the selection changes on either the interactive map or the PCP. The second table displays statistics on a transect-by-transect basis, and is not dynamic. The available graphs are a multi-bar graph of the mean \pm standard deviation for all selected properties, and a box-whisker plot of the minimum, first quartile, median, third quartile, and maximum for a selected property.

1.4.6 Colour Maps

For the coloured properties described in 1.4.1 to 1.4.4, the standard rainbow (blue = low, red = high) and grayscale (black = low, off-white = high) colour maps that vary colour across the visual spectrum are included, but the former are sometimes considered to be among the least useful. Recent visualization research has focused on diverging colour maps, where a transition between two chosen colours passes through an unsaturated colour (Moreland, 2009). Several of these diverging colour maps are implemented in the current system. These colour maps may also be more easily understood by users with colour-deficient vision (Moreland, 2009).

2. Methods

2.1 Study Area

The study area covered a large NE to SW sector of Lake Nipissing (see the map on Figure 1). Most data were collected in an area within approximately 20 km of North Bay, Ontario, but many routes also included the environmentally important Callander Bay area in the southeastern part of the lake. A limited amount of data is also available for the French River, which is the primary outflow for the lake, located in its southwestern sector.

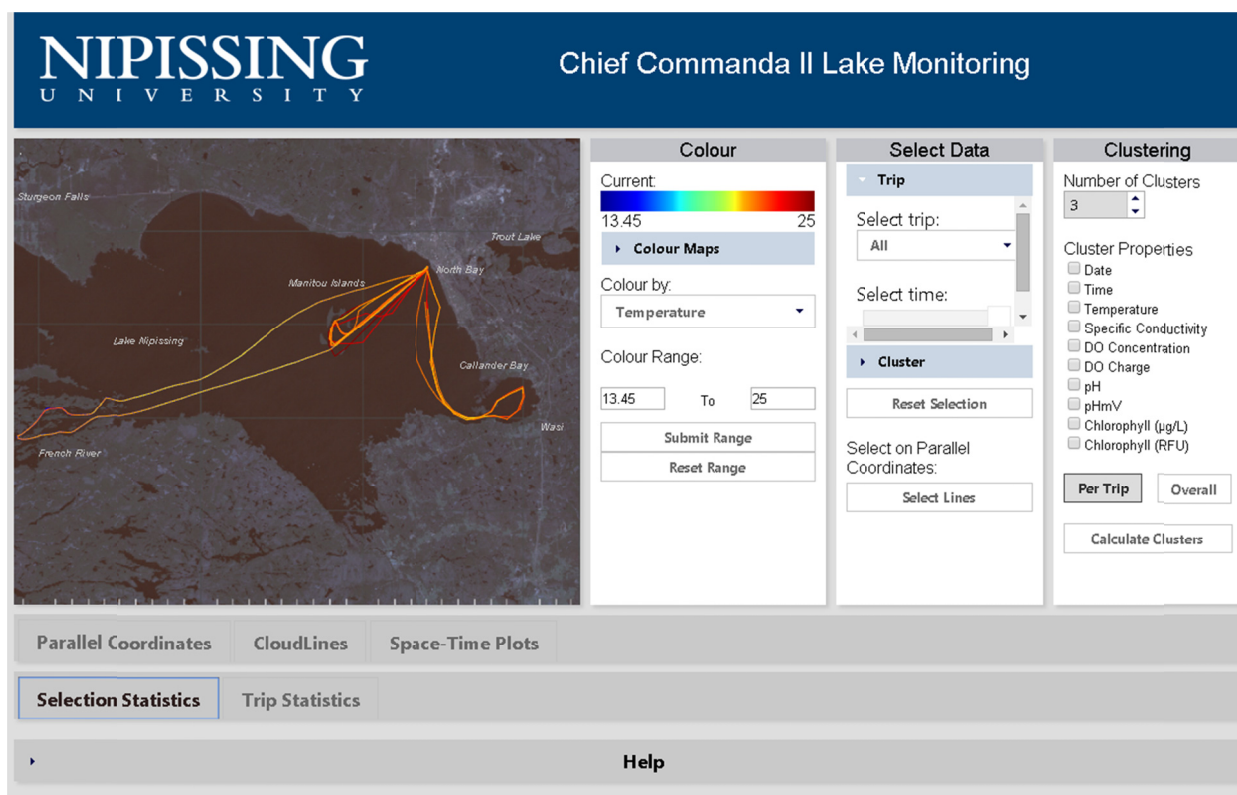


Figure 1. The main web-based graphical user interface, with map coloured by temperatures for all transects. The interactive map of Lake Nipissing is on the left. The expandable visualization and statistics panes are located below the main view pane

2.2 Data Collection

The data were collected as part of the *Chief Commanda II* Water Quality Monitoring Pilot Study. This project is one of several studies to gauge the health of the Lake Nipissing ecosystem. The *Chief Commanda II*, a cruise ship operating daily on Lake Nipissing between May and October, takes several different cruise routes through the lake. Due to the regularity and consistency of its routes, as well as the coverage of large parts of the lake, the cruise ship is an excellent tool for surveying the conditions of the lake over time and at multiple points. Examples of other community-based research from commercial vessels contributing to lake monitoring have been presented in the literature (Buzzelli et al., 2003). In the summer months between 2010 and 2012, the research team attached a YSI (YSI Incorporated, Yellow Springs, Ohio) 6600 V2 sonde to the ship to acquire and transmit readings from the lake at regular intervals during cruises. The sonde, simultaneously used for other studies, was calibrated biweekly. The measurements taken were: (1) temperature ($^{\circ}\text{C}$); (2) specific conductivity ($\mu\text{S}/\text{cm}$); (3) dissolved oxygen (DO) concentration (mg/L); (4) DO charge; (5) pH; (6) pH charge (mV); and (7) chlorophyll ($\mu\text{g}/\text{L}$ and RFU). Additionally, all transects from 2011 measured conductivity (mS/cm), all transects from August 2011 measured turbidity (NTU), and all transects from September 2011 measured total algae (cells/mL and RFU). These attributes were recorded along with the time and GPS coordinates.

2.3 Web Service Implementation

The interactive visualization web service was programmed in JavaScript, a lightweight language that runs on virtually all current web browsers, and employs the JavaScript-based WebGL application programming interface, which facilitates 3D graphics by utilizing the client's graphics hardware. Customized free Dygraphs libraries (<http://dygraphs.com>) were specially adapted for the cloud lines and box-whisker visualizations. Database connectivity and data retrieval were effected through the PHP language.

2.4 Visualization Implementation

Users have the ability to change colour maps for the map overlay, the 3D parallel coordinate plots, cloud lines, and Hovmöller plots. However, pH property colour defaults are set to the standard pH colour scale. Furthermore, the user can select a rectangular spatial region on the interactive map. The property values for this selection are

then updated on the PCP, cloud lines, and statistics visualization. The user may also select individual transects for analysis. Although transects generally run less than three hours, a time slider is provided to allow users to track the progress of a single selected transect, and to assess how properties change during this transect. For the cloud lines, an interactive range selector (adapted from Dygraphs) located below the graphs allows zooming in on smaller date ranges, and hovering over a point on the graph displays the value of all properties at that point in time.

Due to the computational complexity of computing the Hovmöller plots and the limited system resources generally available for web-based visualization, Hovmöller plots were pre-computed for thirty-nine points close to meaningful observation sites across the lake. The site closest to the user-selected point was used for display. Up to four diagrams can be displayed simultaneously.

2.5 Human-Computer Interaction

To facilitate the community-based research aspect of this work, the web-based interface was designed to be as intuitive and interactive as possible. The diverse visualizations are linked through the interactive map “hub”. This map allows users to select specific transects (either by clicking them on the display or by selecting them from a drop-down list), dates, and geographic locations.

The interface is divided into three components to reduce scrolling. Each component can be minimized to save screen space. The interactive map, colour map selection, and transect selection are located in separate panels in the top component. The second component houses the PCP, cloud lines, and space-time plots. A single visualization can be displayed in this panel, allowing the user to rapidly cycle through the visualization with tabs, resulting in a cleaner, easier-to-navigate interface. Users, especially non-researchers, may find it easier to concentrate on one representation at a time. Options that are unique to each visualization are displayed on the right side of the panel.

The bottom component features the statistical tables and plots. A tab interface separates statistics over all data collections with per transect representations. Finally, although the interface, visualization, and components were designed to be intuitive and self-teaching, “tool tips” (text that explains the function of a button, menu, or feature that appears when the pointing device/mouse hovers over it) and text in the interface are liberally employed to aid navigation. Because this is a citizen science initiative, a tutorial video is included that provides an overview of all of the features of the service with special attention given to the more advanced visualizations.

3. Results

The website is hosted on Nipissing University’s server system, and is found at the URL <http://visual.nipissingu.ca/Commanda2>. All the features previously described in Section 2 can be found on this site (Figure 1). Three examples illustrating the use of the visualizations for exploratory data analysis follow.

3.1 Temperature and Dissolved Oxygen

Using the North Bay municipal dock as the reference point from which spatial distances were calculated, a cursory analysis of the Hovmöller plots for temperature and DO may lead to the interpretation of an inverse relationship between the two properties (Figure 2). Because of the low temporal resolution of these plots, the relationship is further examined with cloud lines visualizations, using the same reference point (Figure 3). Here, four properties of interest (temperature, DO concentration, specific conductivity, and pH) are displayed, with the temperature and DO concentration cloud lines moved to the top to facilitate interpretation. The larger radii denote closer proximity to the reference location. The apparent inverse relationship seen in the Hovmöller plots was also observed in the cloud lines, with possible hypoxic (or even anoxic) conditions in parts of the lake.

However, the PCP plot shows that some polylines fall below the DO coordinate axis; consequently, those DO values are suspect (Siirtola & Räihäm, 2006). On the specific sonde used in this study, when the DO charge falls below 25, DO readings are suspect. When the cloud lines zoomed to show the last five transects (all with valid DO values), a positive relationship is seen on the August 22, 2011 transect, and a negative relationship on the September 4, 2011 trip. The visualization led to a closer statistical analysis being performed, the results of which are shown in Table 1. The correlation coefficients (r) confirm the visual results.

To take the analysis further, the PCP was clustered based on temperature and DO concentration with five clusters (Figure 4). Not only is an inverse relationship generally observed between these properties, but upon further examination with rotating the PCP in 3D and by brushing the time period covering the three clusters, it was seen that three of the five clusters were generated for the longest cruise towards the distant French River, suggesting wide spatial variability in DO concentration. By reducing the number of clusters to two, the French River cruise was clearly separated from the other cruises.

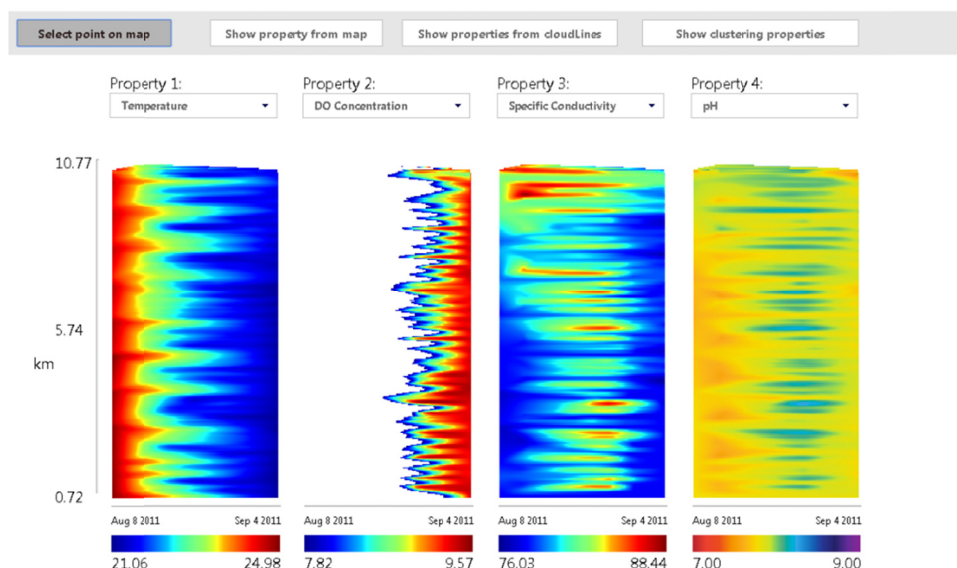


Figure 2. Hovmöller (space-time) plots of four properties. Invalid DO values are not displayed in the second panel

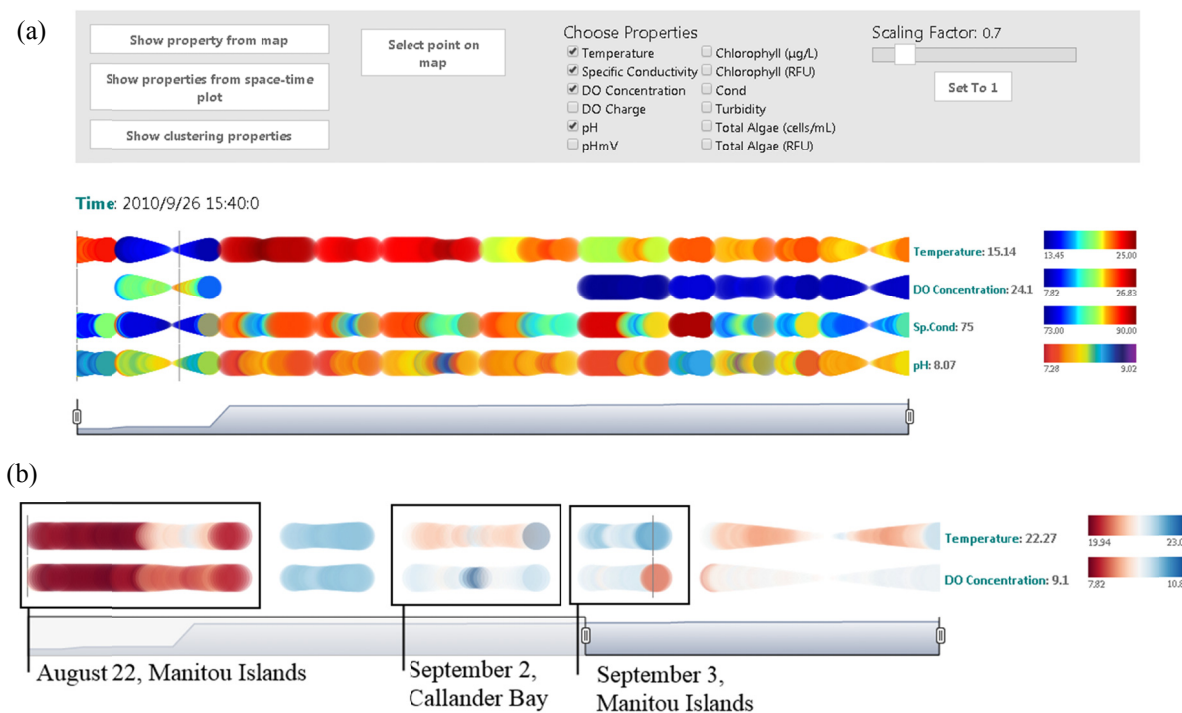


Figure 3. Cloud lines of: (a) four properties; the contrast of Temperature and DO concentration is seen by the two cloud line plots placed at the top of the panel and (b) Temperature and DO concentration zoomed by selecting transects in late August and early September 2011 to more clearly show relationships

Table 1. Correlation coefficients (r) and p -values for three pairs of properties. The Temperature/DO and Chlorophyll/Total Algae relationships were investigated after analyzing the visualizations of these properties

Date	Extent	Temperature and DO		Temperature and pH		Chlorophyll and Total Algae	
		r	p	r	p	r	p
Aug 30 2010	Manitou Islands	N/A	N/A	0.09	0.24	N/A	N/A
Sep 26 2010	French River	0.49	0	-0.28	0.00	N/A	N/A
Aug 8 2011	Manitou Islands	N/A	N/A	0.81	0.00	N/A	N/A
Aug 9 2011	Manitou Islands	N/A	N/A	0.59	0.00	N/A	N/A
Aug 9 2011	Callander Bay	N/A	N/A	0.72	0.00	N/A	N/A
Aug 11 2011	Callander Bay	N/A	N/A	0.12	0.36	N/A	N/A
Aug 22 2011	Manitou Islands	0.73	0	0.77	0.00	N/A	N/A
Sep 2 2011	Manitou Islands	0.07	0.57	0.10	0.43	0.07	0.55
Sep 2 2011	Callander Bay	0.59	0.00	0.72	0.00	0.57	0.00
Sep 3 2011	Manitou Islands	-0.46	0.00	-0.22	0.07	0.34	0.00
Sep 4 2011	French River	-0.01	0.86	0.43	0.00	0.95	0.00

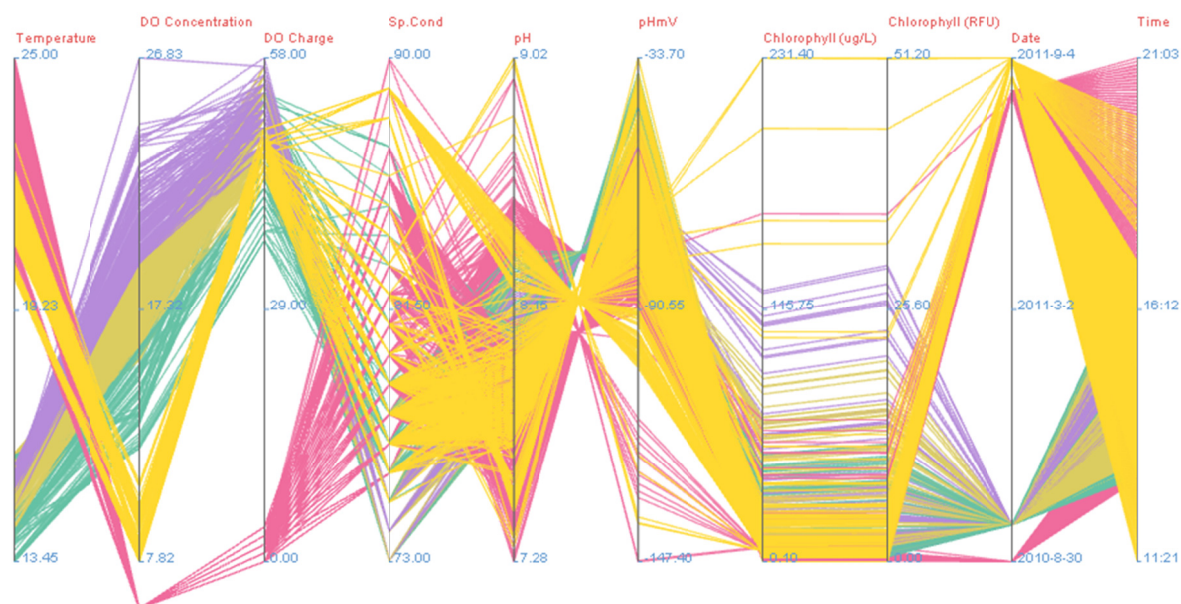


Figure 4. PCP of five clusters based on temperature and DO concentration. Invalid DO concentration readings are displayed below the DO axis

As expected, water temperature and dissolved oxygen increased and decreased, respectively, as the vessel approached shallower water near the Manitou Islands (10 km SW of North Bay) and North Bay waterfront. Spatial variation in these values showed that the sampling method was sensitive enough to detect even slight changes in temperature and dissolved oxygen.

3.2 Similarities across All Properties

Clustering enables many interesting features to be seen. For example, the PCP was clustered into three groups across eight (8) properties (all properties except date and time) to assess overall relationships (Figure 5). The 8D data were clearly clustered on date, with all data from transects in early August of 2011 (August 8 to August 11) categorized into the same cluster, all data from late August to early September of 2011 (August 22 to September 4) being placed into another cluster, and the data from September 26, 2010 comprising the third cluster. If the summers of 2010 and 2011 had similar conditions, it is expected that the readings from late August 2010 would be grouped with that from late August of 2011. However, the data from August 30, 2010 was placed in the same

cluster as the 2011 early August readings, suggesting that there were substantial difference between the summers of 2010 and 2011, as is more clearly seen in the 3D view of the parallel coordinate plot (Figure 6).

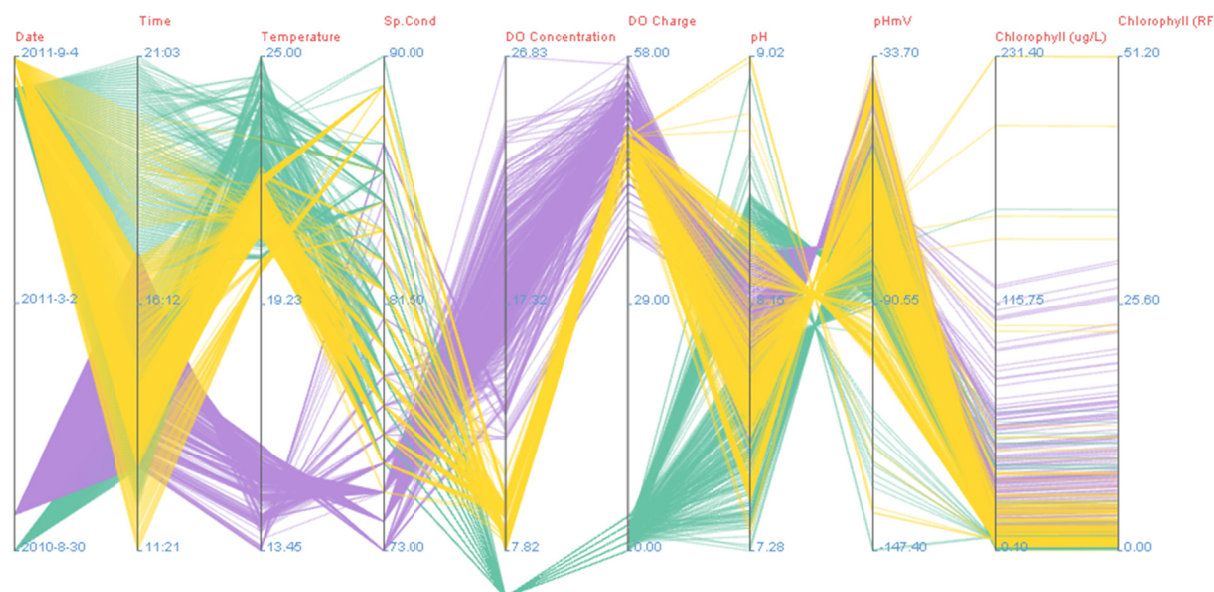


Figure 5. PCP of three clusters across all properties except date and time, indicating that the cruise from August 30, 2010 is in the same cluster as the cruises from early August 2011

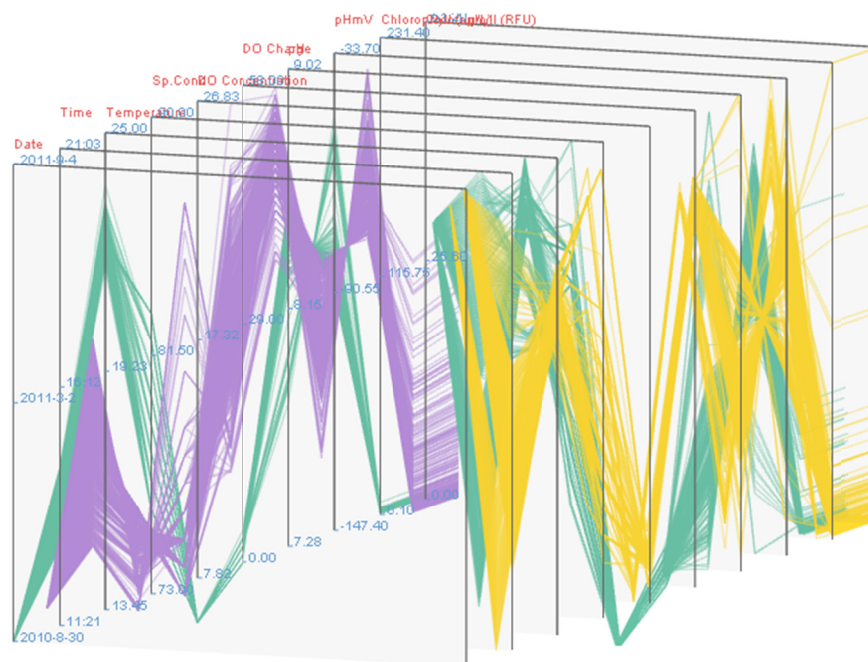


Figure 6. 3D representation of PCP for three clusters, showing the temporal separation of the clusters

3.3 Total Algae

Algae analysis is another example of how the visualization can be used. Densities of algae (as estimated by fluorescence) were recorded for four sampling transects (September 2 [two trips], September 3 and September 4, 2011). For most of the entire dataset, algae readings stayed below 10.00 Relative Fluorescence Units (RFUs). RFU values “spiked” at several points of the September 4, 2011 sampling transect. However, results explaining these fluctuations have been inconclusive. On the one hand, chlorophyll readings rose at the same time (Figure

7), suggesting that there were potentially high levels of primary production. However, the high values were only seen on one leg of the transect, despite the fact that the return leg sampled water that was sometimes 0.25 km away or less. Table 1 shows the detailed correlation analysis of chlorophyll and total algae, and suggests a strong positive relationship in regions close to the shore (Callander Bay and the French River), and not in more interior regions of the lake (Manitou Islands). Furthermore, pH and DO did not change significantly with high RFU readings, although such correlations are associated with post-bloom decomposition of algal material. High turbidity has been shown to also be correlated with high RFU values (Ahn et al., 2007). Unfortunately, turbidity was not recorded for these transects.

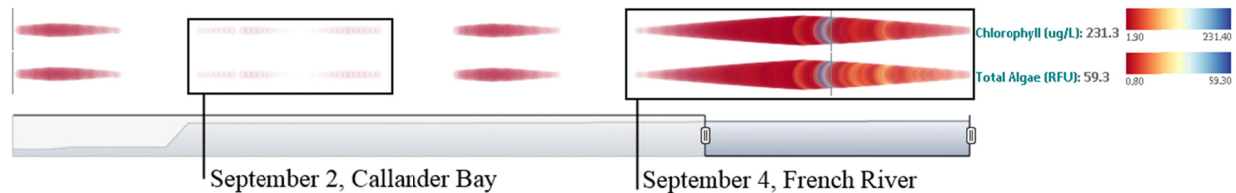


Figure 7. Cloud line plots of total algae Relative Fluorescence Units and chlorophyll readings for September 2 and September 4, 2011

4. Discussion

From the results presented above, the interactivity of the linked visualizations is expected to encourage the community to examine different visualizations to draw conclusions and to formulate hypotheses. It is again noted that the focus of the current study is to demonstrate the potential for visualizing and interacting with limnological data and the transparent presentation of these data. The specific data discussed here should not be treated as representative. However, although the data used in this pilot study may lead to incongruous results, the visualizations provide new impetus for demonstrating outliers instead of eliminating them, as well as for querying sensors (see the discussion of the temperature/DO relationship in Section 3.1).

As shown in Section 3.1, a combination of Hovmöller plots, cloud lines, and PCP were used to assess a relationship that, with a cursory look, could be misinterpreted (e.g. by viewing only the space-time plots). The ability to identify outliers or invalid sensor readings through PCP, and the time selection/zooming property of the cloud lines allows observers to see both positive and negative correlations between the properties, which can then be subsequently statistically analyzed (e.g. Table 1).

The clustering example in Section 3.2 demonstrates spatial and temporal separation, where the $k = 3$ clusters (k was chosen empirically) indicates a division between sampling years. The choice of k , as well as which and how many properties on which to cluster, is dependent upon whether the investigator is researching specific phenomena, or exploring the data for unforeseen relationships. In the latter case, clustering parameter selections are generally empirical. In both cases, an intuitive interface and easy human-computer interaction allows efficient exploration of multiple aspects of the situation in the lake.

Finally, the cloud line visualizations suggest spatially and temporally varying relationships between chlorophyll and total algae, which were subsequently investigated with statistical methods.

Even the more complex visualizations, such as PCP, are supplied with features that allow the general public to make sense of high dimensional data. The interactive 3D aspect of the PCP, while not common for scientific investigations (Johansson et al., 2014), can aid discovery and insights that can be tested with other investigations. For instance, in the example above, PCP was used to further investigate the insights obtained through space-time plots and cloud lines. Brushing dense areas in the PCP, which are linked to the interactive map showing those cruise routes, identified specific geographic regions of interest.

The visualizations draw attention to patterns in the data (e.g. a negative correlation in the case of temperature and dissolved oxygen) that may be subsequently explored with more rigorous comparative statistical methods (e.g. computing the correlation coefficient).

Such web-based visualizations allow the community and other stakeholders to perform in-depth explorations of high-dimensional spatial-temporal data, or of single properties of interest.

5. Conclusion

The data collected for the present study are an important first step in gathering baseline data about Lake Nipissing's environment. Baseline data are especially crucial given the variability of the lake ecosystem.

Further investigation with the visualizations also shows that this variability was particularly apparent in 2010, when lake levels were several meters lower than average. As such, it was expected that the recorded temperature and oxygen values also depart from the norm. However, historical data with which to compare them is not available.

This pilot project indicates that remote monitoring of aquatic parameters in Lake Nipissing is feasible. The method will be adapted to allow the YSI 6600 V2 sonde readings and spatial data to be integrated on a data logger, which will simplify the data collection procedure and make detection of anomalous readings easier and quicker.

The community-based research pilot project presented in this paper is part of a larger initiative for monitoring Lake Nipissing, which presently includes collecting data from sensors at different depths on buoys placed at various locations around the lake. It is also anticipated that the visualizations presented in this web service will be of interest to other community, government, and academic organizations that focus on lake monitoring, such as the Global Lake Ecological Observatory Network (www.gleon.org).

As the applications for real-time aquatic monitoring expand, the potential of using multi-parameter data to better understand and anticipate adverse events in Lake Nipissing – such as anoxia or blue-green algae blooms – is also seen. For example, recent work has demonstrated how evolutionary algorithms can be used to model and forecast algal blooms in a dam reservoir in China (Ye, Cai, Zhang, & Tan, 2014). The visualization of these analyses could potentially aid in identifying correlations and increasing model accuracy.

With the rapid collection and analysis of data, and the implementation of a publicly-available web service, it is hoped that the flow of information will be two-way, where interested parties can comment on the data, as well as contribute their own, including location-tagged images, thereby fulfilling the mandate of citizen science and community-based participatory research.

Acknowledgments

This research was made possible by funding from Nipissing University. M.P.W. is supported by NSERC Grant #386586-2011. We also thank Capt. R. Stirvins and the crew of the *Chief Commanda II* for their cooperation and assistance. Jacquelyne Forgette, Christopher Boasman, Kyle Jaynes, and Bryan Sarlo provided valuable technical assistance. Marlena Pearson helped to proofread and improve the text. Finally, we thank the anonymous reviewers for their constructive suggestions and criticisms.

References

- Abbott, J. G., & Campbell, L. M. (2009). Environmental histories and emerging fisheries management of the upper Zambezi river floodplains. *Conservation and Society*, 7(2), 83.
- Ahn, C., Joung, S., Yoon, S., & Oh, H. (2007). Alternative alert system for cyanobacterial bloom, using phycocyanin as a level determinant. *Journal of Microbiology*, 45(2), 98.
- Blaas, J., Botha, C. P., & Post, F. H. (2008). Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1436-1451. <http://dx.doi.org/10.1109/TVCG.2008.131>
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy; improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions*, 59; 12(11; 10), 977; 3419-984; 3428. <http://dx.doi.org/10.1525/bio.2009.59.11.9>
- Buzzelli, C. P., Ramus, J., & Paerl, H. W. (2003). Ferry-based monitoring of surface water quality in North Carolina estuaries. *Estuaries*, 26(4), 975-984. <http://dx.doi.org/10.1007/BF02803356>
- Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment*, 176(1-4), 273-291. <http://dx.doi.org/10.1007/s10661-010-1582-5>
- Crall, A. W., Newman, G. J., Jarnevich, C. S., Stohlgren, T. J., Waller, D. M., & Graham, J. (2010). Improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions*, 12(10), 3419-3428. <http://dx.doi.org/10.1007/s10530-010-9740-9>

- Danielsen, F., Burgess, N. D., Balmford, A., Donald, P. F., Funder, M., Jones, J. P. G., . . . Yonten, D. (2009). Local participation in natural resource monitoring: A characterization of approaches; participación local en el monitoreo de recursos naturales: Una caracterización de métodos. *Conservation Biology*, 23(1), 31-42. <http://dx.doi.org/10.1111/j.1523-1739.2008.01063.x>
- Dykes, J., MacEachren, A. M., & Kraak, M. (2005). Exploring geovisualization. In J. Dykes, A. M. MacEachren & M. Kraak (Eds.), *Exploring geovisualization* (pp. 3-19). Elsevier.
- Edsall, R. M. (2003). The parallel coordinate plot in action: Design and use for geographic visualization. *Computational Statistics and Data Analysis*, 43, 605-619.
- Eghbalnia, C., Sharkey, K., Garland-Porter, D., Alam, M., Crumpton, M., Jones, C., & Ryan, P. H. (2013). A community-based participatory research partnership to reduce vehicle idling near public schools. *Journal of Environmental Health*, 75(9), 14-19. Retrieved from <http://europepmc.org/abstract/MED/23734527>
- Filion, J. (2011). *Summer 2010 collapse of the Lake Nipissing zooplankton community subsequent to the introduction of the invasive zooplankter bythotrephes longimanus*. (Environmental report). North Bay, ON: Lake Nipissing Partners in Conservation.
- Gold, C. M., & Condal, A. R. (1995). A spatial data structure integrating GIS and simulation in a marine environment. *Marine Geodesy*, 18, 213-228.
- Goodchild, M. F. (2007). Citizens as sensors: Web 2.0 and the volunteering of geographic information. *GeoFocus*, 7, 8-10.
- Hand, E. (2010). Citizen science: People power. *Nature*, 466(7307), 685-687.
- Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W., & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution*, 27(2), 130. <http://dx.doi.org/10.1016/j.tree.2011.11.006>
- Hutchinson, N., Karst-Riddoch, T., & Köster, D. (2010). *Callander Bay subwatershed phosphorus budget*. Bracebridge, ON: Hutchinson Environmental Sciences Ltd.
- Israel, B. A., Parker, E. A., Rowe, Z., Salvatore, A., Minkler, M., López, J., . . . Halstead, S. (2005). Community-based participatory research: Lessons learned from the centers for children's environmental health and disease prevention research. *Environmental Health Perspectives*, 113(10), pp. 1463-1471. Retrieved from <http://www.jstor.org/stable/3436118>
- Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). REVIEW OF COMMUNITY-BASED RESEARCH: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, 19(1), 173-202. <http://dx.doi.org/10.1146/annurev.publhealth.19.1.173>
- Johansson, J., Forsell, C., & Cooper, M. D. (2014). On the usability of three-dimensional display in parallel coordinates: Evaluating the efficiency of identifying two-dimensional relationships. *Information Visualization*, 13(1), 29-41.
- Johansson, J., Ljung, P., Jern, M., & Cooper, M. (2005). Revealing structure within clustered parallel coordinates displays. Paper presented at the *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on Information Visualization*, 125-132. <http://dx.doi.org/10.1109/INFVIS.2005.1532138>
- Kanjo, E., & Landshoff, P. (2007). Mobile phones to monitor pollution. *IEEE Distributed Systems Online*, 8(7).
- Kéry, M., Royle, J. A., Schmid, H., Schaub, M., Volet, B., Haeffliger, G., & Zbinden, N. (2010). Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, 24(5), 1388-1397.
- Krstajic, M., Bertini, E., & Keim, D. A. (2011). CloudLines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2432-2439. <http://dx.doi.org/10.1109/TVCG.2011.179>
- Miller, P., Waghiyi, V., Welfinger-Smith, G., Byrne, S. C., Kava, J., Gologergen, J., . . . Seguinot-Medina, S. (2013). Community-based participatory research projects and policy engagement to protect environmental health on St Lawrence Island, Alaska. *International Journal of Circumpolar Health*, 72(0). Retrieved from <http://www.circumpolarhealthjournal.net/index.php/ijch/article/view/21656>
- Moreland, K. (2009). Diverging color maps for scientific visualization. Paper presented at the *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part II*, Las Vegas, Nevada. 92-103. http://dx.doi.org/10.1007/978-3-642-10520-3_9

- Morgan, G. E. (2013). *Lake Nipissing data review 1967 to 2011* (technical report). North Bay: Ontario Ministry of Natural Resources.
- Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012). The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6), 298-304. <http://dx.doi.org/10.1890/110294>
- Newman, G., Zimmerman, D., Crall, A., Laituri, M., Graham, J., & Stapel, L. (2010). User-friendly web mapping: Lessons from a citizen science website. *International Journal of Geographical Information Science*, 24(12), 1851-1869. <http://dx.doi.org/10.1080/13658816.2010.490532>
- Siirtola, H., & R  ih  m, K. (2006). Interacting with parallel coordinates. *Interacting with Computers*, 18(6), 1278-1309.
- St. Martin, K. (2001). Making space for community resource management in fisheries. *Annals of the Association of American Geographers*, 91(1), 122-142. <http://dx.doi.org/10.1111/0004-5608.00236>
- Streit, M., Ecker, R. C.,   sterreich, K., Steiner, G. E., Bischof, H., Bangert, C., . . . Radu, R. (2006). 3D parallel coordinate systems--A new data visualization method in the context of microscopy-based multicolor tissue cytometry. *Cytometry Part A*, 69A(7), 601-611. <http://dx.doi.org/10.1002/cyto.a.20288>
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282. <http://dx.doi.org/10.1016/j.biocon.2009.05.006>
- Theus, M. (2005). Statistical data exploration and geographical information visualization. In J. Dykes, A. M. MacEachren & M. Kraak (Eds.), *Exploring geovisualization* (pp. 127-142). Elsevier.
- Tulloch, A. I. T., Possingham, H. P., Joseph, L. N., Szabo, J., & Martin, T. G. (2013). Realising the full potential of citizen science monitoring programs. *Biological Conservation*, 165(0), 128. <http://dx.doi.org/10.1016/j.biocon.2013.05.025>
- Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., & McKeon, M. (2007). Manyeyes: A site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1121-1128.
- Voinov, A., & Gaddis, E. J. B. (2008). Lessons for successful participatory watershed modeling: A perspective from modeling practitioners; A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment*, 216; 176(2; 1-4), 197; 273-291. <http://dx.doi.org/10.1016/j.ecolmodel.2008.03.010>
- Weckel, M. E., Mack, D., Nagy, C., Christie, R., & Wincorn, A. (2010). Using citizen science to map Human--Coyote interaction in suburban New York, USA. *The Journal of Wildlife Management*, 74(5), 1163-1171. <http://dx.doi.org/10.2193/2008-512>
- Wegman, E. J., & Luo, Q. (1996). High dimensional clustering using parallel coordinates and the grand tour. *Computing Science and Statistics*, 28, 361-368.
- Wright, D. J., & Goodchild, M. F. (1997). Data from the deep: Implications for the GIS community. *International Journal of Geographical Information Science*, 11(6), 523-528. <http://dx.doi.org/10.1080/136588197242176>
- Ye, L., Cai, Q., Zhang, M., & Tan, L. (2014). Real-time observation, early warning and forecasting phytoplankton blooms by integrating in situ automated online sondes and hybrid evolutionary algorithms. *Ecological Informatics*, 22, 44-51.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).