Visualizing Multivariate Ocean Data with Multiple Views

Lawrence Sun, Gabor Csapó

Abstract—As global temperatures rise, the affects glacier warming are becoming more and more important to understand. Thorough data collection has been done in regions of glacier melting to study its effects. The three main pieces of data being tracked is water's temperature, depth, and salinity. Water that melts from the glacier and flows into the ocean has both a significantly lower temperature that ocean water and also a lower salt content. To observe glacier melting's affect on the ocean, scientists observe how temperature and salinity changes as depth in water changes. However, there exists no meaningful visualization to accomplish such a task. Scientists are currently using line graphs that observe this desired change for only one specific location, not for an entire region. Furthermore there exists no tool to compare how salinity and temperature changes over time, as glaciers continue to melt. This paper provides a meaningful tool to address these issues through use linked multiple views.

1. Introduction

Glacier research has been at the forefront of environmental research for the past few decades. The shrinking size of glaciers worldwide are an accurate indicator of global warming. By observing the melting ice, scientists can reveal fascinating information about future sea level predictions.

David Holland, an NYU researcher and his team leads cutting edge research in several parts of the earth, measuring temperature, ice thickness, salinity and running numerical models to better understand the interactions between glacier, ocean and fjord. In 2011 the team came up with a novel idea to measure different properties of the water by tagging ctd trackers onto seals. Seals are capable of swimming into places that are normally not reachable by scientists and probes.

Holland's team tagged 4 seals that year, which collected information for a year until they shed their skin. The tags would constantly make measurements of location, salinity, temperature, depth and send off the data to a database through a satellite connection whenever the seal emerged from the water. The previous visualizations of the data, however, failed to provide a meaningful overview of the results.

Our analysis tool will aid the scientist to find insights into how the ice melting affects the ocean. The significance of the project lies in the fact that the seals reach places that have never been measured before and therefore could help make predictions more accurate. The main obstacle in the visualization is the number of attributes that need to be plotted at the same time. Through the visualization humans are capable of drawing conclusions about the relation of these properties, which is hard for a computer to do as the seals are swimming into open oceans that create outliers that need to be discarded and the seasonal change is also hard to account for mathematically, but humans can easily estimate and see the overall picture. Therefore this paper will provide three main contributions:

- **System:** We introduce a novel technique to visualize spatial, temporal, and quantitative data that still allows for meaningful comparison. This technique highlights the important changes and presents them in a way that is easily understood.
- **Scientific Contribution:** We create a tool for scientists that makes use of already gathered data that has no tool for evaluation. This tool may allow them to make full use of their data and make new discoveries in oceanography and the affects of glacier melting.
- **Expansion on current visualizations:** This paper expands on current spiral graph visualizations scatter plot visualizations and uses them in novel ways to convey more meaning.

2. Related Work

As this task visualizes multivariate data and looks for patterns in time and geographic location, previous work in visualizing time-oriented data and spacial data is examined. Furthermore, methods of breaking down multivariate data into multiple types of visualizations is also examined.

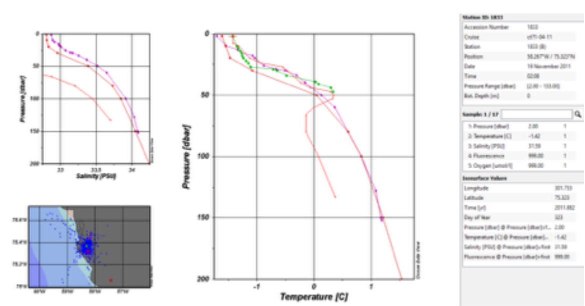

Fig. 1. Current visualization that's graph is too congested and lacks any means of obtaining an overall view.

2.1. Conductivity, temperature, depth (CTD) analysis and visualization

The ultimate goal of the paper is to find the optimal visualization for CTD data that has been collected with tags put on the back of seals. The visualization poses many challenges as we need to visually encode CTD data, which are 3 quantitative attributes already, and on top of that we need to add the location of the data points and show how the data changes over time. We could not find any paper that focuses on the most efficient visualization techniques of CTD data, but many oceanography papers [11,12, 13] implicitly created sections for that. These papers usually disregard the basic design principles (i.e.encoding an interval with the rainbow spectrum). They all struggle to encode the 3 quantitative attributes in one graph. As Nobre [30] points out, spatial data exploration in oceanography has often been restricted to multiple maps showing various depths or time intervals. However, Herlemann [11] in one his diagrams tries to encode salinity and oxygen content on the axis while depth by the size of the dots on the scatterplot, which is a nice try, but is not very intuitive. We'll have to find a better way since we also have to take into account the timeliness and spatiality of our data. What makes Herlemann [11] stand out from other authors is his use of different channels. In another graph he uses position, color and size to create an efficient representation of salinity, depth, distance from starting point, bacterial abundance and categorical data.

From reading through the literature on oceanography it is clear that not many scientists had time to consciously think about efficient visualization techniques of multidimensional data.

2.2. Spatial Data Visualization

Spatial data visualization has been one of the oldest and most looked at method, however, still in areas like multivariate spatial data, progress has not been great. Yixian [15] summed up the types of location-based visualizations. Point represented locations are the most intuitive and direct representation. However, connecting them to show movement makes the graphics cluttered. As Ropinski [22] points out using continuous lines or line sets helps pushing the limits of the number of visualizable points. Region based visualizations come in handy when the number of data points becomes too large as they are used to show aggregated information. Examples of such are flow maps and heatmaps. Line based visualizations can also come in handy and many techniques such as line bundling exist to make the visuals less cluttered. When it comes to multivariate data, many scientists use 3D maps, but that creates many issues of clarity and bias. Other techniques include pixel-based and geometric projection. The most promising technique is the glyph based visualization.

While Winnie Wing-Yi Chan [18] says that these visualizations get very cluttered as the size of the data increases, I still think it is a viable solution. As Ropinski[20] said glyphs can both support preattentive and attentive processing, which is exactly what we need. The design of glyphs can make use of many different visual channels such as shape, colour, texture, size, orientation, aspect ratio or curvature, enabling the depiction of multi-dimensional data attributes as Borgo[21] also realized. Meanwhile, glyphs are also capable of representing clusters of data. He lists many aspects of creating efficient glyphs (attention balance, visual orderability, channel capacity, separability, etc.).

Brownian bridges [25 18, 19] may be used to the expected movement path of an animal and map an estimate of the animal's probability of occurrence in a given area. A Brownian bridge is a continuous-time stochastic model of movement. It is reliant on an animal’s starting and ending location, elapsed time between those points, and the speed of movement. The brownian bridge excels at modeling movement and occurrence probability, but is suited for short time intervals. This would be very unsuited for this project, as it tracks movement through the year and seeks to find patterns between years.

2.3. Time-oriented, Multivariate Visualization and Interactivity

The challenge of visualizing multivariate data against time is consolidating all the data into a visualization that is informative and meaningful yet clean and digestible. Healey [1], in his study on real-time multivariate visualization, defines a well designed real-time multivariate visualization. In addition to being able to visualize data in real-time, which for this paper we do not concentrate on, visualizations are to be visualized on a two-dimensional screen, be rapidly intelligible to users, often catering to pre-attentive processing, and be accurate in representing the investigated relationships.

Ainger [2] claims that to visualize time-oriented multivariate data, a single visualization does not suffice. A simple and singular framework, he claims, cannot justly encapsulate all the data needed to be visualized. Therefore visualizations must either include multiple views or interaction for the visualization to be effective. The goal of multiple views is to provide certain views that primary focus on certain aspects. The goal of interaction is to provide intuitive exploration of the data, not navigation of time. Designers must also be wary not to overload users with too much functionality, thus making the visualization more difficult to perceive. Buja [6, 26] uses linking and focusing in addition to multiple views to solve the issue of visual overload found in singular visualizations. Linking multiple views through a common variable such as time provides context for the views. In visualizing tidal levels, Buja links a time series tidal level plot with a lag plot through the common variable time. Focusing is achieved by panning and zooming on the graphs. These interactions allow viewers to isolate their focus and maintain context at the same time. However these multiple view visualizations do not cater much to pre-attentive processing and require a thorough amount of attention and searching to understand.

Many visualizations [4, 5, 15] such as the Time Wheel, Theme River, and Parallel Coordinate accomplish the task of visualizing in a two dimensional space and being rapidly intelligible, but do not focus on the relationships this paper is seeking. The nature of studying oceanic data points this paper to seek a visualization for cyclic data, whereas the mentioned visualizations best cater to linear-time data. Weber [7] proposes the spiral graph to support the need for comparison of cyclic data. Not only can spiral graphs support a very large data set, they can support multivariate data by means of hue, thickness, use of icons, and can support informative scales for absolute comparison. The downfall of Weber’s visualization, however, is that he proposes a 3D model of a spiral graph to support interactive browsing. Once brought into the third dimension, the spiral graph’s power of comparison is greatly diminished and also may not be as easily digestible as the 2D counterpart. It’s only benefit is its ease of exploration. Very similar visualizations such as Kaleidomaps [28] or Circle Segments [10] maintain the same overall structure as spiral graphs but divide the spiral into regions to function as separate graphs. These method still maintains a good level of comparison between regions.

From reading through the literature on oceanography it is clear that not many scientists had time to consciously think about efficient visualization techniques of multidimensional data.

3. Visualization Design

The goal of the visualization is to visualize both spatial and temporal data along with two other quantitative attributes: temperature and salinity. The visualization also seeks to give the option of filtering for certain geographical regions or periods of time so that specific patterns can be more easily identified and outliers can be easily discarded. To do this, the visualization incorporates four different views that are all connected.

This visualization choses to use multiple views as opposed to an a multivariate visualization so that the ability to perceive relationships is not lost in the visualization. It is also to provide an intuitive visualization of the data.

The map view maps the data points collected from seals onto a geographic map of the ice fjord. It does not connect data points to show the seal’s path because doing so congests the map and makes the user unable to gather any information. The map clusters data points the more zoomed out the map is, and gradually un-clusters the data points as the user zooms in. Users may select an area of the map to examine data only for that specific area. Users also select two points to serve as a distance axis when examining the data selected within the area.

The individual data view will display the traditional and currently used method for visualizing salinity, temperature and depth. Depth is plotted along the y-axis, salinity is plotted along the top x-axis, and temperature is plotted along the bottom x-axis. Two lines are plotted, one to show the Temperature vs. Depth relationship and the other to show the Salinity vs Depth relationship.

The scatter plot view provides an aggregated cross-section view of the data selected on the map, plotting the distance between the two selected points on the map on the x-axis and depth on the y-axis. Each data point’s coordinates are passed through a function that returns the point's distance from the two selected points on the map. This new distance value is plotted on the scatter plot's x-axis. The data points are plotted as dots that show each point's temperature and salinity. The dot's size encodes the point's salinity value and the dot's hue and saturation will encode the point's temperature. Points with high salinity will have larger sized dots while points with low salinity are visualized as smaller dots. This is a very intuitive design because the user can easily assimilate a larger dot with a greater mass and higher salinity. Likewise, a smaller dot visually aligns with the point's low mass and low salinity. Color is also very intuitive and universally understood encoding for temperature, so it is easily understood that the more blue hue colors are cooler than the more orange hue colors.

Because the scatter plot shows a cross section of the ice fjord, the user is able to visually see abnormalities in the water due to the melting fjord. Users can see how the basic physics principle that warm water rises and cold water sinks is disrupted by the water's salinity. Water with high salinity will have a greater mass and thus sink, while water with low salinity will rise. The decision to visualize a cross section of the water easily and intuitively displays this relationship.

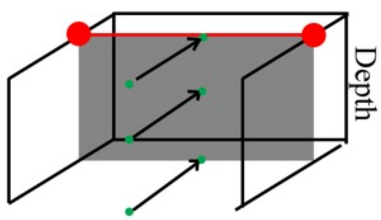

Fig. 2. Portrayal of the aggregated cross section. The red line and dots show the user selected points the distance value is calculated from. The green dots show how all data points in the volume are aggregated onto the plane.

The spiral graphs view shows two spiral graphs: one for temperature, one for salinity. These graphs aggregate data for the region selected by the user. The spiral graphs encode the data for each year onto one ring of the graph in order to make comparisons between years. Hue and saturation are encoded to show whether the values are high or low.

We chose to create a spiral graph for each piece of quantitative data: temperature and salinity, as opposed to a spiral graph for each year to more easily show changes in specific data between years. For example, temperature change can be directly seen as each year's data is next to each other. If we chose to create an overall spiral graph for each year that plots both temperature and salinity, this comparison would be much more difficult.

There is both a depth slider and a time slider for users to concentrate the data on a specific depth range or time. Users may want to specifically see the change in data for only a specific depth and not have the rest of the data offset the visualization. Also a time slider is implemented so that users can focus on a specific time frame.

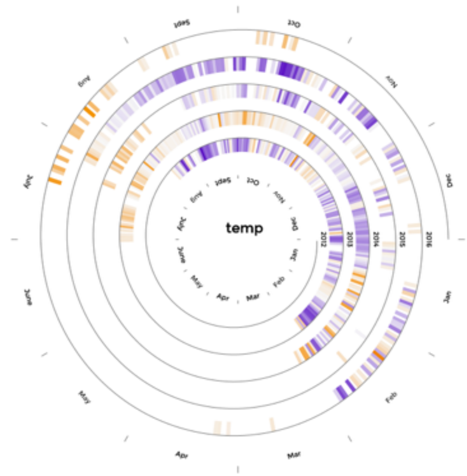

Fig. 3. Spiral graph of temperature for years 2012-2016

Acknowledgments

We wish to thank David Holland and his research team for giving us the opportunity to work on this project and use the data they collected. We would also like to thank Alexander Mackay for connecting us to Dr. Holland and guiding us through the team's missions and goals.

References

1. Healey, Christopher G., Kellogg S. Booth, and James T. Enns. "Visualizing Real-time Multivariate Data Using Preattentive Processing." *ACM Transactions on Modeling and Computer Simulation* 5.3 (1995): 190-221. Web.
2. Aigner, Wolfgang, Silvia Miksch, Wolfgang MÃ¼ller, Heidrun Schumann, and Christian Tominski. "Visualizing Time-oriented DataâA Systematic View." *Computers & Graphics* 31.3 (2007): 401-09. Web.
3. Muller, W., and H. Schumann. "Visualization Methods for Time-dependent Data - an Overview." Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693) (n.d.): n. pag. Web.
4. Guo, Diansheng, Jin Chen, A.m. Maceachren, and Ke Liao. "A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP)." *IEEE Transactions on Visualization and Computer Graphics* 12.6 (2006): 1461-474. Web.
5. Aigner, W., S. Miksch, W. Muller, H. Schumann, and C. Tominski. "Visual Methods for Analyzing Time-Oriented Data." *IEEE Transactions on Visualization and Computer Graphics* 14.1 (2008): 47-60. Web.
6. Buja, A., J.a. Mcdonald, J. Michalak, and W. Stuetzle. "Interactive Data Visualization Using Focusing and Linking." *Proceeding Visualization '91* (n.d.): n. pag. Web.
7. Weber, M., M. Alexa, and W. Muller. "Visualizing Time-series on Spirals." *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.* (n.d.): n. pag. Web.
8. A. Pang. Visualizing Uncertainty in Geo-spatial Data. In Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology, Arlington, 2001.
9. Evans, Beverley J. "Dynamic Display of Spatial Data-reliability: Does It Benefit the Map User?" *Computers & Geosciences* 23.4 (1997): 409-22. Web.
10. Ankerst M., Keim D. A., Krienel H.-P.: "'*Circle Segments': A Technique for Visually Exploring Large Multidimensional Data Sets*", Proc. Visualizatoin'96, Hot Topic Session, San Francisco, CA, 1996.

11.Hu, Yue Oo, Bengt Karlson, Sophie Charvet, and Anders F. Andersson. "Diversity of Pico- to Mesoplankton Along the 2000 Km Salinity Gradient of the Baltic Sea." (2015): n. pag. Web.
12.Straneo, Fiammetta, Gordon S. Hamilton, David A. Sutherland, Leigh A. Stearns, Fraser Davidson, Mike O. Hammill, Garry B. Stenson, and Aqqalu Rosing-Asvid. "Rapid Circulation of Warm Subtropical Waters in a Major Glacial Fjord in East Greenland." *Nature Geoscience* 3.3 (2010): 182-86. Web.
13.Holland, David M., et al. "Acceleration of Jakobshavn Isbrae triggered by warm subsurface ocean waters." Nature geoscience 1.10 (2008): 659-664.
14.Sheng, Yafang Su Yongwei. "Visualizing Upwelling at Monterey Bay in an Integrated Environment of GIS and Scientific Visualization." *Marine Geodesy* 22.2 (1999): 93-103. Web.
15.Y. Zheng, W. Wu, Y. Chen, H. Qu and L. M. Ni, "Visual Analytics in Urban Computing: An Overview," in IEEE Transactions on Big Data, vol. 2, no. 3, pp. 276-296, Sept. 1 2016.
16.Menemenlis, Dimitris, et al. "ECCO2: High resolution global ocean and sea ice data synthesis." Mercator Ocean Quarterly Newsletter 31 (2008): 13-21.
17.Kamachi, Masafumi, et al. "Ocean Data Assimilation and Prediction system in JM and MRI." Quarterly Newsletter: 31.
18.Chan, Winnie Wing-Yi. "A survey on multivariate data visualization." Department of Computer Science and Engineering. Hong Kong University of Science and Technology 8.6 (2006): 1-29.
19.Horne, Jon S., et al. "Analyzing animal movements using Brownian bridges." Ecology 88.9 (2007): 2354-2363.
20.Buchin, Kevin, et al. "Detecting movement patterns using brownian bridges." Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, 2012.
21.Ropinski, Timo, Steffen Oeltze, and Bernhard Preim. "Survey of glyph-based visualization techniques for spatial multivariate medical data." Computers & Graphics 35.2 (2011): 392-401.
22.Borgo, Rita, et al. "Glyph-based visualization: Foundations, design guidelines, techniques and applications." Eurographics State of the Art Reports (2013): 39-63.
23.Alper, Basak, et al. "Design study of linesets, a novel set visualization technique." IEEE Transactions on Visualization and Computer Graphics 17.12 (2011): 2259-2267.
24.Downs, Joni A., and Mark W. Horner. "Analysing infrequently sampled animal tracking data by incorporating generalized movement trajectories with kernel density estimation." Computers, Environment and Urban Systems 36.4 (2012): 302-310.
25.Guo, Diansheng. "Flow mapping and multivariate visualization of large spatial interaction data." IEEE Transactions on Visualization and Computer Graphics 15.6 (2009): 1041-1048.
26. Shamoun-Baranes, Judy, et al. "Analysis and visualization of animal movement." Biology letters (2011): rsbl20110764.
27.Buja, Andreas, Dianne Cook, and Deborah F. Swayne. "Interactive high-dimensional data visualization." Journal of computational and graphical statistics 5.1 (1996): 78-99.
28.Ward, Matthew O. "A taxonomy of glyph placement strategies for multidimensional data visualization." Information Visualization 1.3-4 (2002): 194-210.
29.Bale, Kim, et al. "Kaleidomaps: a new technique for the visualization of multivariate time-series data." Information Visualization 6.2 (2007): 155-167.
30.Zhao, Jian, Fanny Chevalier, and Ravin Balakrishnan. "Kronominer: using multi-foci navigation for the visual exploration of time-series data." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011.
31.Nobre, Carolina, and Alexander Lex. "OceanPaths: Visualizing Multivariate Oceanography Data." (2015).