

Proposal for a  
Thesis in the Field of  
Information Technology  
In Partial Fulfillment of the Requirements  
For a Master of Liberal Arts Degree

Harvard University  
Extension School  
April 20, 2015

Carolina Nobre

Proposed Start Date: July 2014  
Anticipated Date of Graduation: September 2016  
Thesis Director: Alex Lex

# 1 Tentative Thesis Title:

OceanPaths: Visualizing Multivariate Oceanography Data

## 2 Abstract

Geographical datasets are ubiquitous within the realm of oceanography. While map-based visualizations are useful for many different domains, they can suffer from clutter when used for multivariate data sets. As a result, spatial data exploration in oceanography has often been restricted to multiple maps showing various depths or time intervals. This lack of interactive exploration often hinders efforts to expose correlations between properties of oceanographic features, such as currents. In an effort to remedy these issues we plan to develop OceanPaths, a browser-based tool that will provide powerful interaction and exploration methods for spatial, multivariate oceanography datasets. This tool will allow users to define pathways along which the variation of the high-dimensional data can be plotted efficiently.

## 3 Thesis Project Description

This chapter covers the project background, description of the OceanPaths tool, a comparison to previous work in visualizing multivariate data, design considerations, and a prototype of the tool itself.

### 3.1 Background

Spatial visualization of properties in the ocean is important to many oceanography researchers. Oceanographic data can be collected using in situ methods, typically ship expeditions, or remote sensing via satellites. We can also produce fields from computational models. The default method for visualization and analysis for oceanographers are maps overlaid with color-coded marks as shown in Figure 1. The analyst's goal is to better understand property distribution and advection, the transfer of heat or matter by the ocean. The challenge that arises with oceanographic measurements is that they are often multivariate, including temperature, salinity, and oxygen content, and are available for multiple depths and multiple times. This results in a high-dimensional spatial dataset.

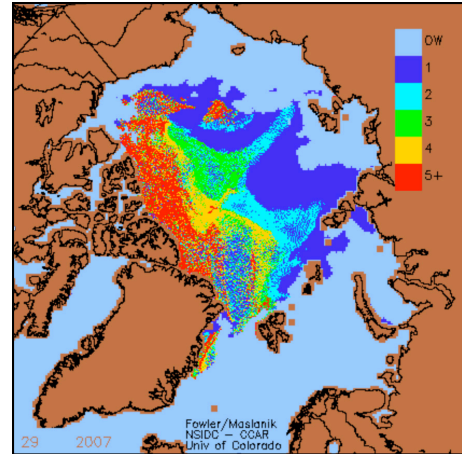


Figure 1. Ice thickness in the arctic. Example of a typical visualization for oceanographers.

As noted by Lipsa (Lipsa, Laramée, Cox, Roberts, & Walker, 2011), attempting to include multiple dimensions in a map-based visualization leads to cluttering and significantly hinders the effectiveness of the visualization (Figure 2a). Exploring trends in these properties, usually as a function of time and geographic location, is a common task in oceanography. In an attempt to reduce the spatial components to a single dimension, researchers often produce a scatter plot with measurements along a fixed latitude or longitude as the independent variable.

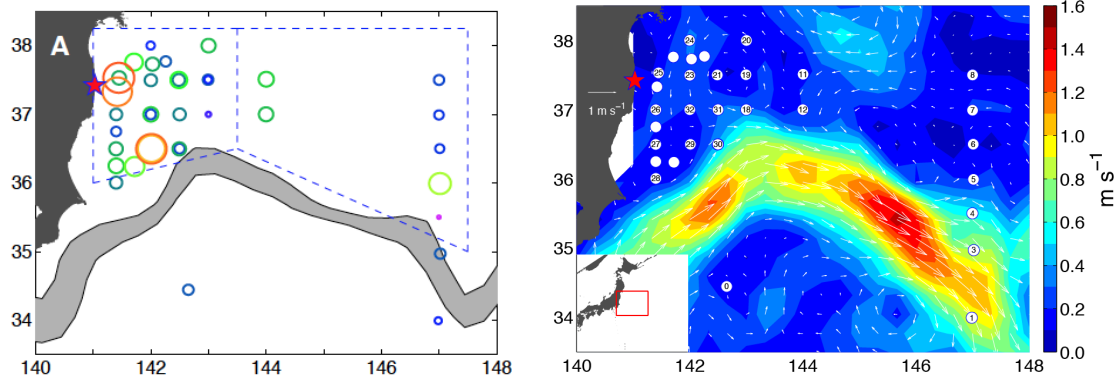


Figure 2. From Charette et al 2013. Left panel: Markers represent  $^{228}\text{Ra}/^{226}\text{Ra}$  at surface stations. The gray shaded area is the approximate position of the Kuroshio Current during the cruise. Right panel: Mean position of the Kuroshio Current.

While this approach accomplishes the goal of reducing dimensionality, it also confines the user to assessing movement in pathways along equal longitude or latitude. As an example, Figure 2b shows the path of the Kuroshio Current as produced by a model. Assessing current variability along a straight line, whether latitude or longitude clearly will not follow the current. Moreover, as discussed by John C. Anderson (Anderson, Andres, Davis, Fujiwara, Fang, & Nedbal, 2010), oceanographers are still confined to using predominantly static maps, which are often restricted to a single depth and time.

The authors of a study on the deep circulation in the Arctic Ocean (Spence, Fyfe, Montenegro, & Weaver, 2010) acknowledge the shortcomings associated with limiting cross sections to either latitude or longitude. Their paper points out that complex features in the local bathymetry, the topography of the ocean floor, can have a significant impact on local dynamics. This cannot be effectively quantified when plotting along a straight line. Our proposed tool, OceanPaths, represents a novel visual tool for exploring multidimensional oceanographic datasets, addressing the shortcomings discussed above, by enabling users to define pathways along which to plot the multivariate data. These pathways can be constructed with multiple branches in order to better represent the complex underlying current systems. Ocean currents play a fundamental role in shaping the earth's climate and heat distribution. As such, a better understanding of this data can inform policies on ocean warming, ice melt, and overall environmental awareness.

### 3.2 Prior Work

Multivariate data visualization in the realm of earth science has previously been addressed by several authors (Li, Fu, & Hanson, 2008; Yuan, Xiao, Guo, Guo, Kendall, Huang, & Zhang, 2010; Co, Friedman, Grote, Vay, Bethel, & Joy, 2005; Burakiewicz & Van Liere, 2006; Jankun-Kelly & Mehta, 2006; Deines, Michel, Bertram, Hagen, & Nielson, 2006). Lipsa (Lipsa et al., 2011) presents a comprehensive overview of data visualizations used in the physical sciences and notes that there is no shortage of challenges when producing effective data visualizations in this field. Our work focuses on visualization for multivariate geospatial oceanography datasets. Researchers have employed various methods for handling multivariate spatial visualization in other domains.

### *Li et al., 2008*

Hongwei Li (Li et al., 2008) visualized multivariate 4D astronomy data by mapping the galactic coordinates to the X-Y (image) dimension, the wavelength values to a third spatial dimension perpendicular to the image plane, and the flux density values to color values (Figure 3). This is similar to our approach in that the spatial dimensions are conserved in a map like encoding while the remaining two dimensions are encoded separately. In their work, the authors allow the user to draw a mask over the map in order to outline a particular feature that the user would like to explore. This will then highlight that feature in all subsequent visualizations of the data.

While we also support interaction and selection of points in our map based visualization, we further improve on this interaction by projecting each of the selected points onto the user defined pathway, calculating the along track distance and using this value as the x dimension in the scatterplot visualization. This metric is particularly useful for oceanographers, since it is more physically meaningful for analyzing spatial variability than latitude or longitude alone.

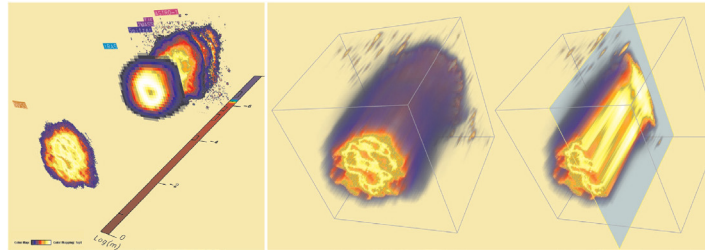


Figure 3. From Li et al., 2008. A visualization example of a multi-wavelength crab nebula.

### *Yuan et al., 2010*

Yuan (Yuan et al., 2010) studied 4D seismic and satellite-based observational data. Similar to Li et al, 2008 and the work proposed here, the authors map latitude and longitude to a 2D map. The two remaining dimensions are time and intensity of the earthquake, which are encoded along the Z and color/size dimensions respectively (Figure 4). While encoding time on an axis is a common approach for 2 dimensional plots, we found that doing so in a 3D plot led to visual clutter.

OceanPaths deals with the time dimension in two ways. First, the user can constrain the time interval for which the data is plotted by means of a slider, which updates all visualizations. The user can further explore temporal variability by selecting time as the desired variable for the y-axis or the color dimensions in the scatterplot display.

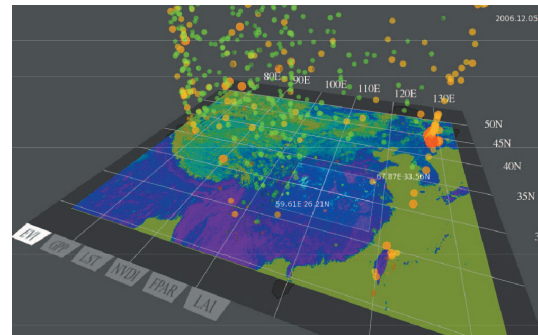


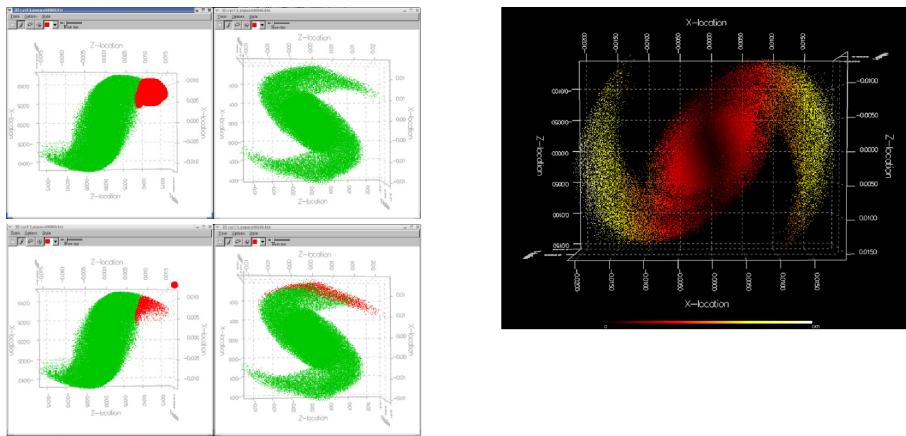
Figure 4. From Yuan et al., 2010. Seismic catalog data set is displayed as 3D points, together with satellite images in a common temporal geographical coordinates system.

Yuan chose to encode earthquake intensity as both size and color. While this approach works well for spatially sparse data, it can lead to cluttering when the data points are closer together, a common occurrence in oceanographic data.

As such, we chose to plot all points in a constant size, and used only color to encode the property value. We also expanded on the approach used by Yuan by allowing the user to select which field is encoded as color on both the map and the associated scatterplots.

*Co et al., 2005*

Multiple linked views (Co, Friedman, Grote, Vay, Bethel, & Joy 2005; Kehrer, Ladstadter, Muigg, Doleisch, Steiner, & Hauser, 2008) has been employed within physics to handle multidimensional data sets. In his paper, Christopher Co (Co et al., 2005) employs linked views, as well as color, to highlight a given set of points and their properties in separate visualizations. The authors first employ color to indicate which subset of points the user has selected (Figure 5, left panel). A second visualization uses color to encode the values of the selected points (Figure 5, right panel). While this approach accomplishes the task of analyzing the values of only a subset of the data, it has the drawback of only allowing the user to visualize the measurements once a selection is made. This can be particularly onerous if the user would like to use the value of the measurements to guide his selection.



*Figure 5. From Co et al., 2005. Left panel: Demonstration of manual partial selection through the use of a painting tool. Right panel: Corresponding points now colorized according to the magnitude of their momentum vector.*

OceanPaths addresses this drawback by highlighting selected points with opacity and size of the markers, while keeping the point's measurement encoded in color (Figure 6). This allows the user to clearly see which points are selected, while still being able to use the values of the points to inform their selection.

In summary, OceanPaths takes a unique approach in handling multivariate, spatial datasets by meaningfully condensing the spatial dimensions of a dataset while still preserving the spatial variability. Our tool allows the user to define a meaningful set of branching pathways along which currents flow. As such, it is related to previous work done for analyzing multivariate networks (Partl, Lex, Streit, Kalkofen, Dashofer, & Schmalstieg, 2013), yet uses significantly different types of data and visual encodings. We also expand on the user of color to encode a property by allowing the user to select which property to encode in color. The use of marker size and opacity to indicate selected points further allow the user to clearly distinguish selected points while still visualizing the selected property in color.

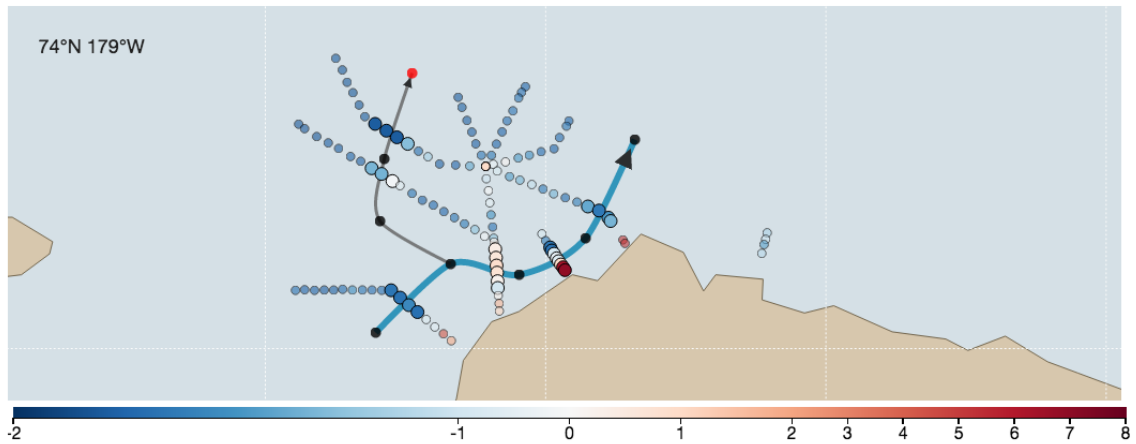


Figure 6. OceanPaths map visualization which makes use of marker size and opacity to indicate selected nodes.

### 3.3 Application Design

OceanPaths will be developed in collaboration with a team of oceanographers at the Woods Hole Oceanographic Institution. The team consists of researchers from the physical, chemical, and biological oceanography departments and includes experts in arctic circulation, dispersion of radioactive elements in the ocean, and permanence of radioisotopes on marine organisms. The design of our tool will be guided by the following domain tasks, which we elicited through interviews with these domain experts:

1. Determine current trajectories.
2. Explore the variation of various measurements along currents.
3. Compare data distribution at several distinct depth and time intervals.

In order to support the elicited domain tasks, we propose the use of two coordinated views (see Fig 7): a map, where selected attributes can be directly plotted and pathways can be interactively defined, and a “Scatterplot Network”, showing the multivariate data along the pathways.

#### Map

In the map we plot points to show where measurements are available. On demand, these points can be color coded to show a selected attribute at a chosen time and depth, or an aggregate thereof. The background layer shows the depth of the ocean. Since currents tend to follow lines of constant depth, experts can create and edit pathways in the map interface, addressing domain task 1. The pathways can have multiple incoming and outgoing branches, typically resulting in a crossing free, directed, acyclic graph. We use the distance along these pathways as a one-dimensional representation of the spatial variation of data points. The map in Figure 1 shows measurement points that are color-coded by the statistical mean temperature from the Arctic Regional Climatology dataset, provided by the National Oceanographic Data Center (NODC) (Boyer, Baranova, Biddle, Johnson, Mishonov, Paver, Seidov, & Zwent, 2012). Overlaid on the map is a user-created path network representing the main Norwegian Atlantic Current, an extension of the Gulf Stream, as it travels north and branches off into the Greenland and Barents Seas. Pathways are created by



clicking on the map to define nodes, which are automatically connected to a previously selected node. Users can easily add or remove nodes and segments, thus creating a system of branches that represents the underlying currents.

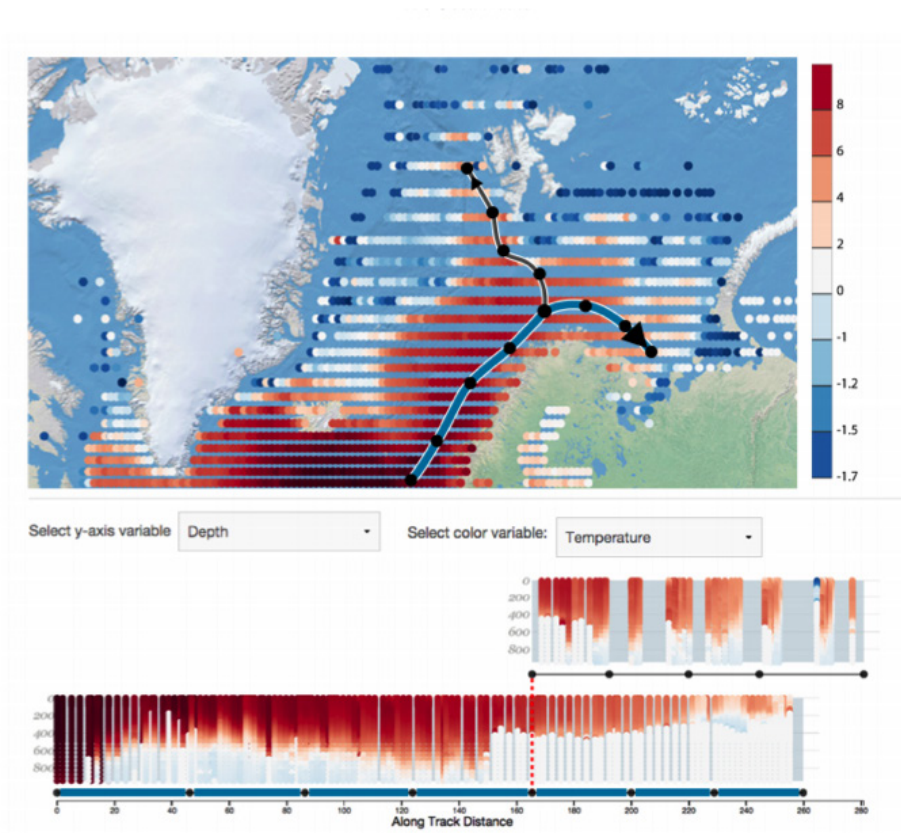


Figure 7. Layout of OceanPaths with the map as the top view and the scatterplot network as the bottom view. The map shows points of measurement color-coded by temperature averaged across the depths, as well as a user-defined path based on currents. The color scale used is non-linear in order to bring out the variability in sub-zero temperatures. The scatterplot network shows one scatterplot per path segment where depth is plotted on the vertical axis and temperature is encoded as color. The user-defined “main” path is shown in blue without branches at the bottom.

## Scatterplot Network

The Scatterplot Network, shown below the map, visualizes the selected data as a function of distance along the defined pathways. The Scatterplot Network consists of individual scatterplots that correspond to the branches of the network defined in the map, with the (user-defined) main path shown straight and highlighted in blue. The support points of the paths are shown along the X-axis. The positioning of the subplots is defined by the path’s relative positions to each other: a path to the right of another path is shown below it, a path to the left is shown above. Branching points are indicated with red, stippled lines. We devised this method of subplot ordering and positioning to clearly visualize the properties along contiguous branches. The properties encoded by both the y-axis and by the color of the marks can be defined by the user. In Fig 7 the y-axis represents depth and the color property encodes temperature values. The data displayed in

the scatterplot shows a clear cooling of the water column starting at the point where the current spawns a branch into the Greenland sea. The branch that veers east and contours the east coast of Norway experiences greater cooling than the Greenland sea branch. We determine which measurement point is associated with which exact position along a path using two alternative methods. By default, we automatically select all data points that fall within a user-editable distance of the pathway. Editing this distance can account for varying width of currents. Selected data points are then projected onto the closest position on the closest branch, i.e., they are assigned an along-track distance and a branch. An alternate method of selecting and/or refining the selection of data points is manual selection, which makes fine-tuning of the measurement assignments possible. The map and the scatterplot network are fully coordinated views. Selecting a measurement point or a path in one plot will highlight the corresponding point in the other view.

## 4 Work Plan

### 4.1 Assumptions, Risks and Alternatives

The development environment for OceanPaths is fairly straightforward since it is a browser based tool and does not require access to any databases or proprietary software. A potential risk associated to the development of OceanPaths relates to the scope of the tool. While we have created the proposed timeline leaving plenty of time for iterations with the domain experts, there is always the possibility that domain experts will suggest additional features, and we may or may not have the time to implement these. In order to successfully complete the thesis by spring/summer of 2015, we may need to restrict ourselves to the proposed features.

### 4.2 Preliminary Schedule

Milestone	Duration(weeks)	Description	Expected Completion Date
Background research	4	Becoming acquainted with what work has been done in visualizing multivariate data	Aug 2014
Determine domain tasks	4	Meetings with Domain Experts to determine domain tasks for OceanPaths	Sep 2014
UI design and preliminary implementation	4-8	Discuss possible layouts for the tool. Implement select features to test out user interaction.	Dec 2014



First prototype	4	With a layout determined and the design tasks clearly outlined, a prototype will allow for user testing to begin.	Jan 2015
Continued development of OceanPaths	3-5	Testing by domain experts will allow for continued improvement of OceanPaths.	Feb 2015
Paper Submission to EuroVis 2015	4	Opportunity to write up and present OceanPaths as a novel contribution to the DataVis community.	Feb 2015
Improvements based on EuroVis Feedback	8	The feedback received from the reviewers will allow for further refinements to OceanPaths as both a research tool and a contribution to the data visualization community.	Apr 2015
Write Up Thesis	8	Expand on paper to produce thesis document	May 2015

## Bibliography

- Anderson, J., Andres, J., Davis, M., Fujiwara, K., Fang, T., and Nedbal, M. (2010). Voyager: An interactive software for visualizing large, geospatial data sets. *Marine Technology Society Journal*, 44(4), 8-19.
- Boyer, T. P., Baranova, O. K., Biddle, M. Johnson, D. R., Mishonov, A. V., Paver, C. Seidov, D., and Zwent, M. (2012). Arctic Regional Climatology, Regional Climatology Team NOAA/NODC ([www.nodc.noaa.gov/OC5/regional\\_climate/arctic](http://www.nodc.noaa.gov/OC5/regional_climate/arctic)).
- Burakiewicz, W., and Van Liere R. (2006). Analyzing Complex FTMS Simulations: a Case Study in High-Level Visualization of Ion Motions. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 1037-1044.
- Co, C., Friedman, A., Grote, D., Vay, J., Bethel, E., and Joy, K. (2005). Interactive Methods for Exploring Particle Simulation Data. *EG/IEEE VGTC Symposium on Visualization (EuroVis)*, 279- 286.
- Deines, E., Michel, F., Bertram, M., Hagen, H., and Nielson, G. (2006). Visualizing the phonon map. *EG/IEEE VGTC Symposium on Visualization (EuroVis)* 291-298.
- Jankun-Kelly, T., and Mehta, K. (2006). Superellipsoid-based, real symmetric traceless tensor glyphs motivated by nematic liquid crystal alignment visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 1197-1204.
- Kehrer, J., Ladstadter, F., Muigg, P., Doleisch, H., Steiner, A., and Hauser, H. (2008). Hypothesis Generation in Climate Research with Interactive Visual Data Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1579-1586.
- Li, H., Fu, C.-W., and Hanson, A. (2008) Visualizing Multiwavelength Astrophysical Data. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1555-1562.
- Lipsa, D. R., Laramée, R. S., Cox, S. J., Roberts, J. C., and Walker, R. (2011). Visualization for the Physical Sciences. In *EG 2011 - State of the Art Reports*, Eurographics Association, 49-73.
- Martine, A. R., and Ward, M. O. (1995). High Dimensional Brushing for Interactive Exploration of Multivariate Data. *IEEE Conference on Visualization, IEEE Computer Society*, 271-278.
- Partl, C., Lex, A., Streit, M., Kalkofen, D., Kashofer, K., Schmalstieg, D. (2013). enRoute:

Dynamic Path Extraction from Biological Pathway Maps for Exploring Heterogeneous Experimental Datasets. BMC Bioinformatics, 14(19), S3

Pickart, R. S., Moore, K. W. G., Mao, C., Bahr, F., Nobre, C., Weingartner, T. (2015) Circulation of Winter Water on the Chukchi Shelf in Early Summer. To be submitted to Deep-Sea Research.

Qu, H., Chan, W.-Y., Xu, A., Chung, K.-L., Lau, K.-H., and Guo P (2007). Visual analysis of the air pollution problem in Hong Kong. IEEE Transactions on Visualization and Computer Graphics, 13(6), 1408-1415.

Spence, P., Fyfe, J. C., Montenegro, A., and Weaver, A. J. (2010). Southern Ocean response to strengthening winds in an eddy-permitting global climate model. Journal of Climate, 23(19), 5332-5343.

Yuan, X., Xiao, H., Guo, H., Guo, P., Kendall, W., Huang, J., and Zhang, Y. (2010). Scalable multivariate analytics of seismic and satellite-based observational data. IEEE Transactions on Visualization and Computer Graphics, 16(6), 1413-1420.