OceanPaths: Visualizing Multivariate Oceanography Data

Carolina Nobre

A Thesis in the Field of Information Technology
for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

June 2016

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. Currents are a key component of ocean dynamics and the current visualization tools do not serve it them well. In an effort to remedy these issues we developed OceanPaths, a browser-based tool that provides powerful interaction and exploration methods for spatial, multivariate oceanography datasets. This tool allows users to define pathways along which the variation of the high-dimensional data can be plotted efficiently. These pathways can be constructed with multiple branches in order to better represent the complex underlying current systems.

Acknowledgements

I would like to thank my supervisor, Dr. Alexander Lex, for helping me find a project that not only took advantage of my background as an oceanographer, but that provided me with such a rewarding foray into the world of information visualization. I am particularly thankful for his support, feedback, and patience as we wrote my first paper. Our meetings, whether in person or through Skype, gave me valuable guidance and support throughout the entire process.

Thank you also Prof. Hanspeter Pfister for teaching the class that first sparked my interest in the topic of Visualization.

I would like to thank my supervisor at the Woods Hole Oceanographic Institution, Dr. Robert Pickart. His patience and support of me as I pursued my master's degree after work were truly admirable.

And finally, I am truly thankful to my mom and dad. Mom, for our conversations, they truly are food for the soul. And Dad, for your unwavering support of my pursuit of something greater than myself.

This thesis is based on the following paper:

Carolina Nobre and Alexander Lex, OceanPaths: Visualizing Multivariate Oceanography Data. Proceedings of the Eurographics Conference on Visualization (EuroVis 15), Short Papers, 2015.

Contents

List of Figures

Chapter 1: Introduction

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. Researchers 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 1.1. The analysts 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.

As noted by Lipsa et al. (2012), attempting to include multiple dimensions in a map-based visualization leads to cluttering and significantly hinders the effectiveness of the visualization (Figure 1.2). 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.

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 1.2 shows the path of the Kuroshio Current as produced by

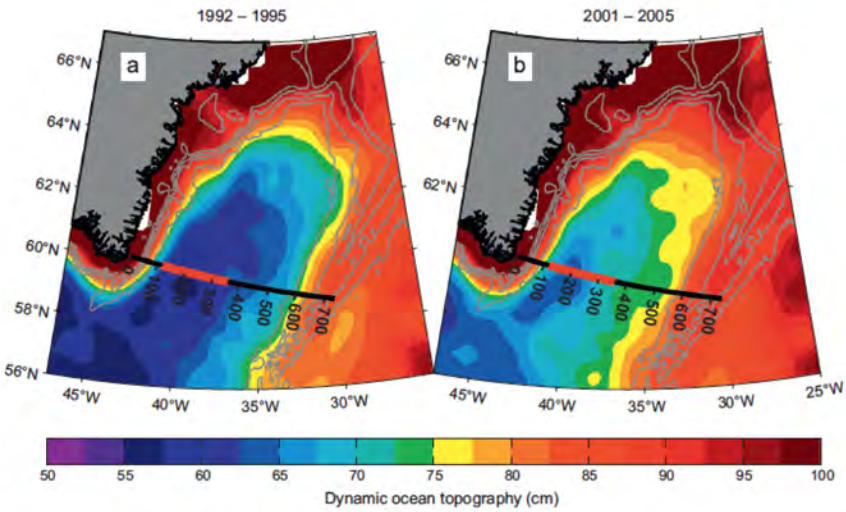

Figure 1.1: Figure from Vage et al. (2011) representing dynamic ocean topography. Example of typical oceanographic visulization.

a model. When assessing current variability along a straight line, whether latitude or longitude, the analyst's distance dimension will clearly not be equivalent to an "along current" dimension due to the sinuous nature of the current's path. Moreover, as discussed by Anderson et al. (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 et al., 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.

In order to address the shortcoming above we present OceanPaths, an interactive visual analytics tool designed to explore multidimensional oceanographic datasets. OceanPaths principal contribution is the support of user defined geographic pathways and plots of variables as a function of distance along these custom pathways. These pathways can be constructed with multiple branches in order to better

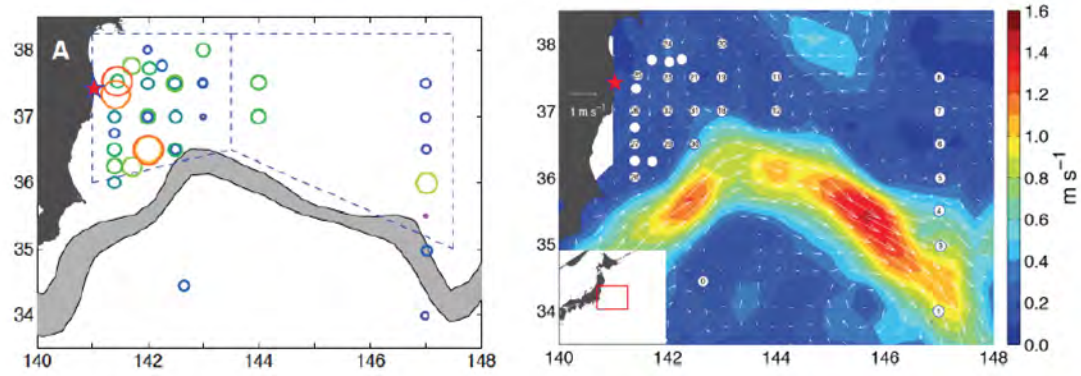

Figure 1.2: From Charette et al. (2013). Left panel: Markers represent 228Ra/226Ra at surface stations. The gray shaded area is the approximate position of the Kuroshio Current during the cruise. Right panel: Mean position of the Kuroshio Current.

represent the complex underlying current systems. The tool also allows users to filter the data by date intervals, as well as visualize horizontal property distribution of properties at specified depth ranges. OceanPaths was developed and validated over the course of several oceanographic expeditions during which the interaction with the lead investigator allowed for continuous refinements to the interaction mechanisms and desired functionality.

The contribution of this work is two-fold. We first present an analysis and abstraction of the domain tasks within oceanography and how they are currently addressed within the field. We then present the design and implementation of OceanPaths. In Chapter 2, we discuss the data analysis tasks that are common to oceanography and that we aim to facilitate with OceanPaths. After reviewing the related work in Chapter 3, we continue with a description of the OceanPaths design and interface in Chapter 4. Chapter 5 covers the implementation details, including challenges encountered and how we chose to address them. Our results are presented in the form of a case study in Chapter 6, followed by a final chapter where we summarize our findings and conclude with a discussion of future work.

Chapter 2: Requirements

We now discuss the data, domain tasks, and associated design requirements for OceanPaths as defined by interviews with domain experts during the design and development phases of this design study.

The development of OceanPaths was carried out in conjunction 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.

2.1. Analysis Tasks

This section describes the analysis tasks that resulted from the interviews with domain specialists.

(1) Determine current trajectories

The relevance of ocean currents for global climate regulation, primary production, carbon dioxide storage, transport of nutrients, and dispersion of radioactive elements is well documented in the literature (Schlosser et al., 1995; Mathis et al., 2007; Brewer et al., 1989). The growing amount of literature on ocean currents and their importance to our ecosystem is evidence that this is a topic of great interest to

researchers across all areas of ocean sciences.

As a result, several researchers focus their efforts on analyzing the transports, seasonality and overall characterization of these currents. The work of Schlosser et al. (1995), for example, discusses how large scale Arctic currents play a role in the transport of contaminants and pollutants brought into the ocean by river runoff. Mathis et al. (2007) analyze how eddies, circular currents of water, influence the transport of organic carbon and nutrients into the western Arctic Ocean.

In analyzing current pathways, researchers must also account for the influence of the ocean floor on the water flow. As noted by Schuur et al. (1998), bottom flowing currents tend to follow lines of constant depth. As a result, researchers often rely on bathymetric charts to aid in their analysis.

(2) Explore the variation of measurements along currents

As mentioned above, currents often dictate how ocean properties are transported and distributed throughout the ocean. As such, oceanographers are interested in quantifying the rate and direction of this dispersion. Prants et al. (2011), for example, analyzed the propagation of radioactive pollutants by an extension of the Kuroshio Current in the Pacific. They discovered that the radionuclides were being trapped by local eddies with potentially harmful effects on marine organisms.

(3) Compare data distribution at several distinct depth and time intervals

The last domain task that emerged was the processes of analyzing the depth and time dependent component of ocean processes. An example of depth dependence are marine organisms that are restricted to certain depths in the ocean due to light and/or food availability, as shown by McCreary et al. (2001). Time dependent processes are also common and include inter-annual events such as El Nino as well as

seasonal variations of established water flows.

2.2. Design Requirements

As a result of the domain tasks described above, we have developed the following design requirements.

Enable sketching of currents/pathways on a map

This design requirement addresses domain task 1 of determining current trajectories. Because of the established importance of bathymetry in ocean currents, this interface should also display bathymetric charts as a background layer.

Determining the pathway of a current is a non-trivial task. OceanPaths relies on the domain expert's knowledge of currents in the area, allied with the bathymetric information, to sketch the current's position. While using measured velocity fields to automatically track the current positions would be ideal, velocity measurements in the ocean are sparse in both time and space and therefore do not provide a comprehensive enough field to automatically detect current pathways. Oceanographers can use these fields as a basis, but must use their knowledge of the area to sketch the entire pathway for the desired study area and time.

Given the need to sketch current pathways, OceanPaths must be able to accurately mimic their main properties. An ocean current can be defined as a continuous directed movement of seawater that responds to forces such as gravity, waves, wind, density changes and the topography of the ocean floor. As a result, ocean paths are sinuous, exhibiting bends and branches along their path. Ocean currents also vary greatly in width, depth, and position in the water column.

As a result, the requirement for the map interaction is the ability to properly represent the current properties described above.

Visualize properties along current pathways

As demonstrated in the description of domain task 2, users must be able to visualize data values along the created pathway. As such, each selected data value must be assigned an along path distance.

The main requirements for this functionality are:

(a) Find which current each measurement is associated with.

(b) Visualize selected attribute as a function of along path distance.

Constrict data to user defined depth and time intervals

In order to support domain task 3, OceanPaths must allow the user to restrict the imported dataset to certain depth and time ranges.

In order to do so, the tool must allow the user to interactively filter the data to the desired ranges for both depth and time. The filtered data will be reflected in all viewports within OceanPaths.

Link Views between Map and Scatterplot

As discussed in Chapter 1, oceanographers rely heavily on a combination of maps and scatter plots to analyze and detect ocean processes. These two forms of visualization are complementary as they allow analysts to correlate the geographic position of features to property changes along a chosen spatial dimension. As a result, the requirement to link the map and scatterplot views was established to facilitate the process of correlating geographic position and property distribution for a selected feature. The associated requirements are:

(a) Selecting a measurement highlights the point in all viewports.

(b) Changes in selected data points should be reflected in both the Map and the ScatterPlot.

Details on Demand

This last requirement addresses the need to view and assess the precise value of a given measurements. While color-coding gives the user a sense of the measurements and how it compares to the surrounding points, it is impractical to deduce the exact value. Displaying the data values as a tooltip on hover allows the user to extract the information from either the map or the scatterplot.

Chapter 3: Prior Work

In this chapter we cover related work in the area of multivariate data visualization in the natural sciences. We also discuss how OceanPaths builds on previous work done in order to make a significant contribution to the field.

Multivariate data visualization in the realm of earth science has previously been addressed by several authors (Li et al., 2008; Yuan et al., 2010; Co et al., 2004; Burakiewicz & van Liere, 2006; Jankun-Kelly & Mehta, 2006; Deines et al., 2006). Lipsa et al. (2012) 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. While there have been relatively few oceanography specific multivariate spatial visualization efforts to date (Schlitzer, 2002; Brown, 1998), researchers have employed various methods for handling multivariate spatial visualization in related domains: astrophysical data (Li et al., 2008), satellite-based observational data (Yuan et al., 2010), particle simulation data (Co et al., 2004), ion motions (Burakiewicz & van Liere, 2006), liquid crystal alignment (Jankun-Kelly & Mehta, 2006), and phonon map data (Deines et al., 2006).

What follows is a description of previous contributions to the field and how OceanPaths addresses any issues with each approach.

3.1. Oceanographic Data Visualization Tools

One of the main visualization applications used today by the oceanographic community is Ocean Data View (ODV), a package for the interactive exploration and graphical display of multiparameter profile or sequence oceanographic data (Brown, 1998). When creating visual displays, ODV allows the user to choose between five different visualization modes, all based on display types that are commonly used by the scientific community (Schlitzer, 2002).

The MAP mode generates a map with the measurement locations (Figure 3.1). The map view allows the user to select the projection and includes bathymetry and land topography information. One drawback however is that in order to change the geographic boundaries of the map, the user must access and alter settings in a properties panel, making it cumbersome to zoom in on specific regions.

The map view in OceanPaths is interactive, allowing the user to pan and zoom to their region of interest. The resolution of the bathymetry is also updated accordingly.

In SECTION mode in ODV the user has the option of defining a section interactively on the map by clicking on nodes that define the centerline of the section (Figure 3.2). As stated in the ODV documentation, after specifying the section spine, ODV assigns a default width for the section band and selects distance from the first point entered as the along-section coordinate. All stations inside the section band marked in the map belong to the section and are plotted in subsequent plot operations.

OceanPaths builds on this approach on several fronts. First off, the user can define several section axes, referred to as pathways, creating a complex system of branches that represent the underlying current pathways.Because these pathways are proxies for ocean currents with a horizontal extent, users can interactively define the

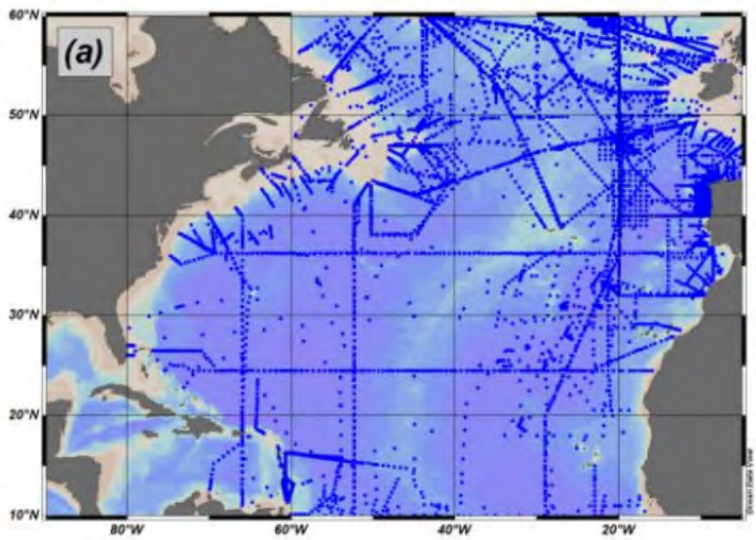

Figure 3.1: ODV MAP mode showing measurement locations.

current width. Only measurements that fall within the current's breadth are selected.

Lastly, each measurement is projected onto the nearest branch and the along track distance assigned is the perpendicular distance to that branch, not the distance to the starting point of the section as is done in ODV. This approach has the advantage of preserving the spatial signature of features and avoiding the distortion that can arise with oblique projections. Figure 3.3 compares the along track distance calculation between ODV and OceanPaths for a given set of points.

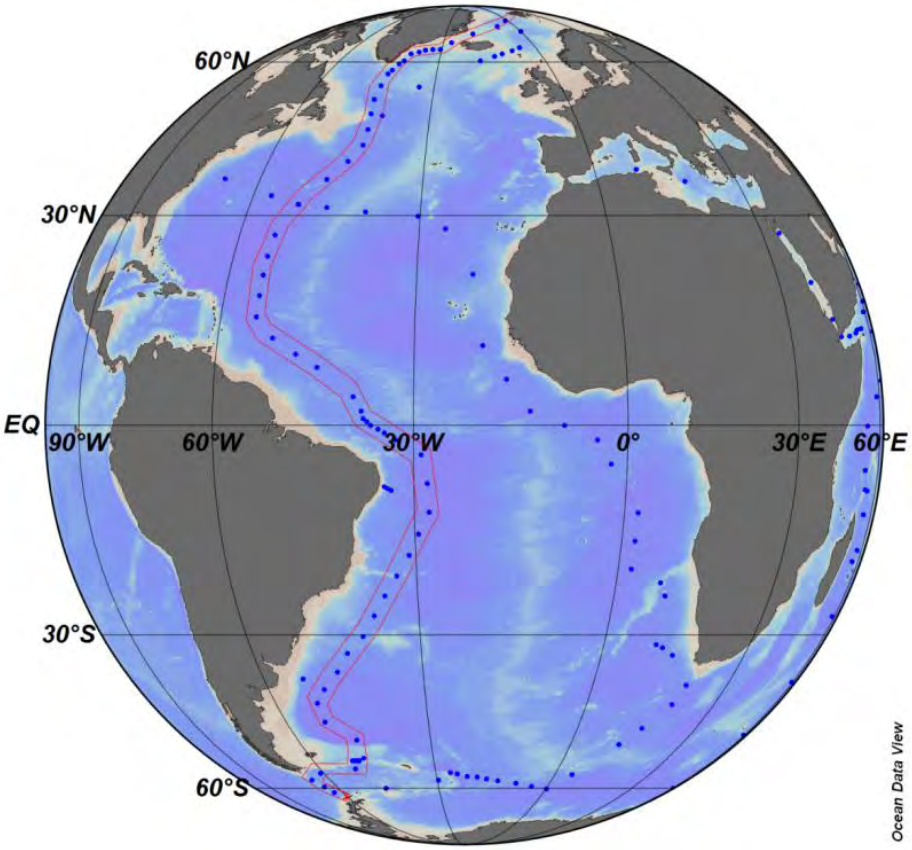

Figure 3.2: ODV map interface for users to select the spine of a section, or the equivalent of a pathway in OceanPaths.

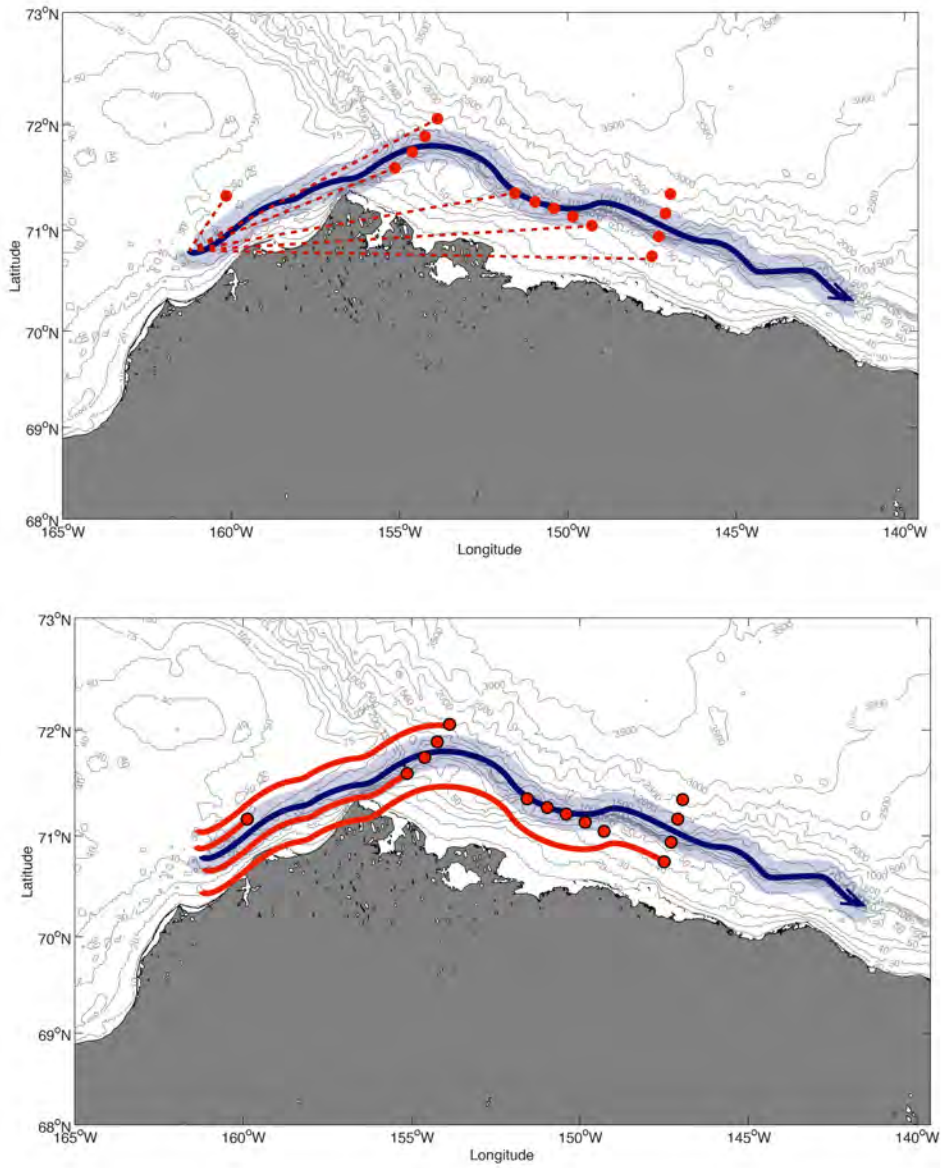

Figure 3.3: Comparison between ODV and OceanPath methods of calculating distance along track for a set of measurements. Blue track represent the user defined path, red markers represent the imported data points. Top panel: ODV method where the distance of each measurement is the distance to the start of the path. Bottom panel: OceanPaths method where each measurement is projected onto the path and the distance of the path line up to that point is assigned to the measurement.

Overall, ODV and OceanPaths exhibit overlapping capabilities but essentially different purposes. ODV is an application that must be installed on the user's device. Also, it is designed to quickly generate scatter plots, sections, and maps from a potentially large oceanographic database. The generated figures are static and aside from the section definition mode, offers no interactive exploratory features such as linked axis or station selection.

OceanPaths on the other hand was developed to maximize user exploration of smaller datasets through interactive features. The tool builds on oceanographer's need to understand spatial variability by allowing the user to interact with the map view and define a distance axis along features of interest. Additionally, OceanPaths is a browser based tool thus bypassing any need for installation and platform dependencies which are characteristic of installed software.

Aside from ODV, oceanographers often use scripting languages such as Matlab, Octave, Python, and R to import data files and create customized graphics. While this approach has the advantage of allowing for a higher level of customization for the figures produced, these scripts are often time consuming to produce, are restricted to scientists with some programming experience, and often are not made widely available to the scientific community.

OceanPaths aims to remedy this by facilitating the process of data analysis for researchers regardless of their programming skills and specific fields.

3.2. Multivariate Encoding

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 3.4).

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.

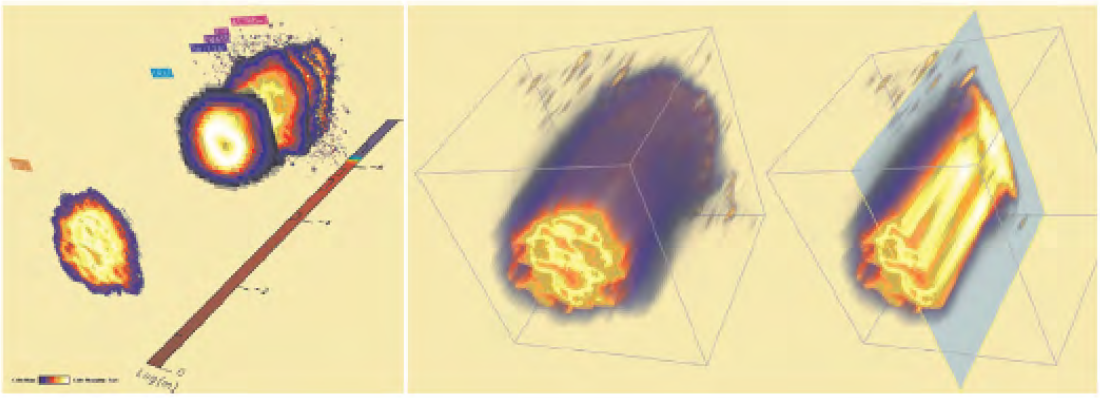

Figure 3.4: From Li et al. (2008). A visualization example of a multiwavelength crab nebula.

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.

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 (Figure3.5). While encoding time on an axis is a common approach for 2 dimensional

plots, doing so in a 3D plot leads to visual clutter.

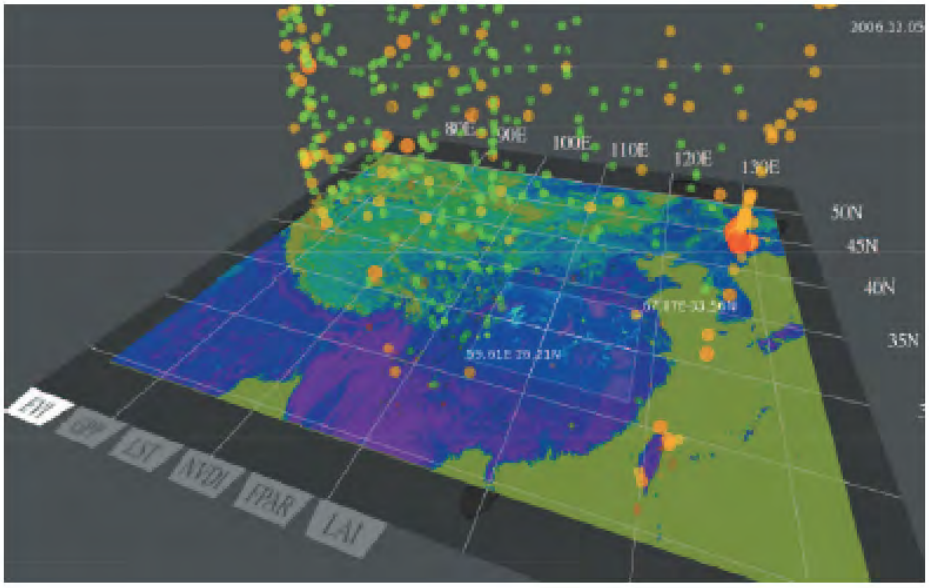

Figure 3.5: From Yuan et al. (2010). Seismic catalog data set is displayed as 3D points, together with satellite images in a common temporal geographical coordinates system.

Yuan et al. (2010) chose to encode earthquake intensity in both the size and color of markers. 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 et al. (2010) by allowing the user to select which field is encoded as color on both the map and the associated scatterplots.

3.3. Linked Views

Multiple linked views (Co et al., 2004; Kehrer et al., 2008) have been employed within physics to handle multidimensional data sets. In their paper, Co et al. (2004)

employ 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 3.6, left panel). A second visualization uses color to encode the values of the selected points (Figure 3.6, 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.

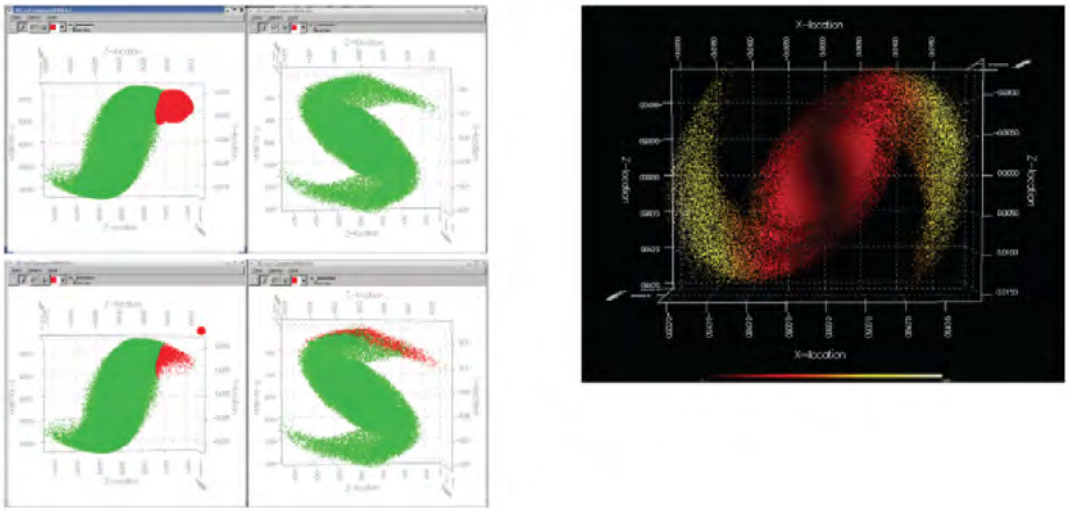

Figure 3.6: From Co et al. (2004). Left panel: Demonstration of manual partical selection through the use of a paiting tool. Right panel: Corresponding points now colorized according to the magnitude of their momentum vector.

OceanPaths addresses this drawback by highlighting selected points with a white outline while keeping the points measurement encoded in color, as shown in Figure 4.2. 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 et al., 2013), yet uses significantly different types of data and visual encodings. We also expand on the use 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.

While existing work in the realm of multivariate data visualization provides a range of valid approaches, it does not sufficiently address the tasks associated to analyzing multivariate spatial oceanographic data. The work done with multivariate data sets tends to encode spatial dimensions in color or size, but does not allow the combination of spatial coordinates in a single, meaningful, axis as is done in OceanPaths. Within the domain of multiple linked views, existing literature allows for selection of points but in doing so, overrides the color encoding of the points. OceanPaths improves on this method by outlining selected data points in all linked views, while maintaing the color encoding of the original measurements.

Chapter 4: OceanPaths Design

In this chapter we explain the design decisions that went into the creation of OceanPaths. In each subsection we highlight a component, giving a detailed description of the user interaction features as well as the reasoning behind the chosen visual encodings.

In order to support the elicited domain tasks, we use two coordinated views, as shown in Figure 4.2: 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.

Figure 4.1 contains a schematic of the two viewports in OceanPaths. For each path drawn in the map (left panel) a corresponding scatterplot is created in the scatterplot network (right panel). The positioning of the scatterplots is defined so as to mirror their position in the map. More detail in how these is implemented is described in section 4.2

4.1. Map

4.1.1 Measurements

We plot the imported data as points on the map to show where measurements are available. By using the drop down menus and range sliders, these points can be color-coded to visualize a selected attribute at a chosen time and depth, or an

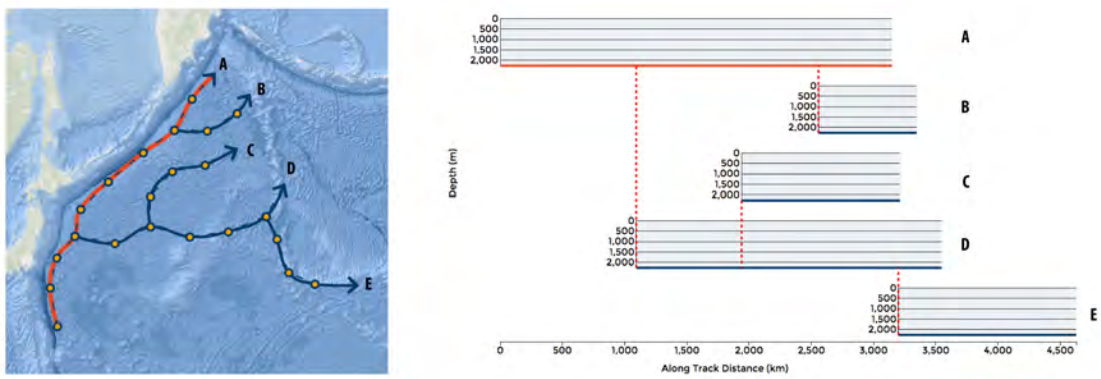

Figure 4.1: Schematic of the two viewports in OceanPaths and how pathways drawn in the map (left panel) correspond to the scatterplot network (right panel). The main path is shown in red and branch paths are drawn in blue.

aggregate thereof. The drop down menu is populated with all the attributes detected in the imported dataset. In order to allow the user to restrict the active range of the depth and time attributes, the two range sliders below the map have moveable handles for quick positioning of the desired ranges. Imposing depth constraints on the data allows the user to focus on features and processes that are constrained to a given ocean layer. This addresses domain task 3 discussed in Chapter 2. The time constraint allows for analysis to be performed on events that are confined to a given period, such as El Nio (Giese & Ray, 2011) or other seasonal and interannual phenomena.

The background layer of the map shows the bathymetry, i.e. the depth of the ocean. Because bottom currents often follow lines of constant depth (Shetye, 1982; Schuur et al., 1998)), domain experts find it useful to consult bathymetric charts when analyzing current movement.

The map in Figure 4.2 shows measurements acquired by floating oceanographic devices, known as drifters, in the North Atlantic. Drifters are often employed in oceanographic research to investigate ocean currents as they tend to follow the flow

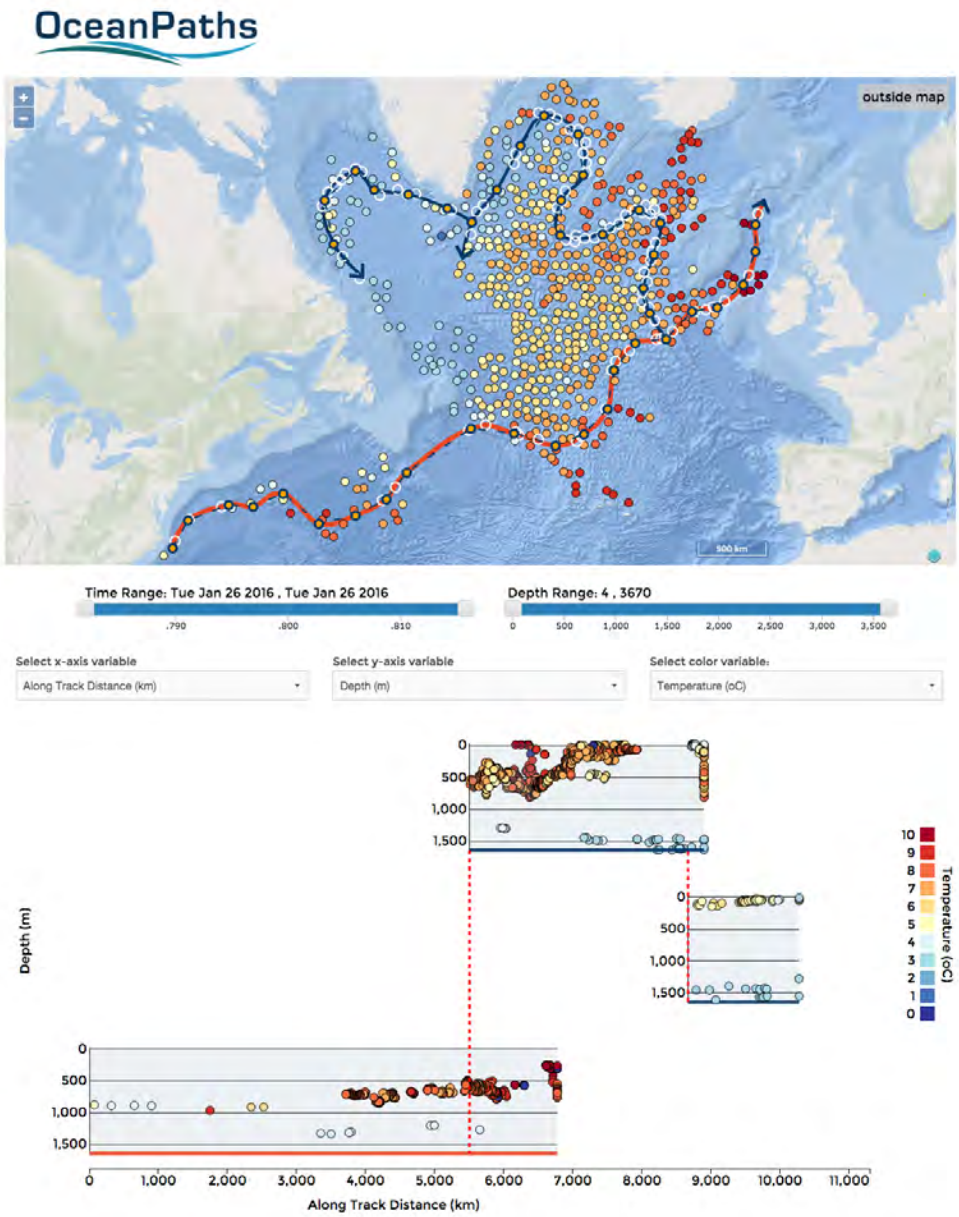

Figure 4.2: Screenshot of OceanPaths with the map as the top view and the scatterplot network as the bottom view. The map shows points of measurement color-coded by temperature averaged across the depths, as well as a user-defined path based on currents. The scatterplot network shows one scatterplot 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 red and branch paths are drawn in blue.

of major currents. The data set used here is the WOCE Subsurface Float Data Assembly Center (WFDAC) and was provided courtesy of the Woods Hole Oceanographic Instituion. Overlaid on the map is a user-created path network representing the main North Altantic Current, an extension of the Gulf Stream, as it travels north and branches off towards the Nordic and Labrador Seas.

4.1.2 Pathways

Based on the imported data, the depth information, and their background knowledge about currents in the ocean, experts can create and edit pathways in the map interface. The pathways can have multiple incoming and outgoing branches, typically resulting in a crossing-free, directed, acyclic graph. The importance of pathways being crossing-free stems from the fact that currents in the ocean do not intersect. Instead, they merge and branch off from one another. We use the distance along these pathways as a one-dimensional representation of the spatial variation of data points.

Pathways are created by clicking on the map to define nodes, which are automatically connected to a previously selected node. Users can easily add (click) or remove (shift click) nodes and segments, thus creating a system of branches that represents the underlying currents. Dragging a node will change its position as well as all the pathway segments that connect to it. Users can also create separate segments and later merge them into a single path or system of paths by dragging one node on top of another. Nodes exhibit a snap behavior that allows for easy merging of path segments.

After the path is constructed, data points located within the width of the current are automatically selected and plotted in the Scatterplot Network, described below. The selection of data points occurs in the following manner. Each data point

is associated to its closest path. If the distance of a measurement to its closest path is within the user defined width of the current, it is selected.

What follows is a description of the steps taken to construct a pathway in the map interface.

User Input

As described in the Libraries section, the map interface in OceanPaths was implemented with the OpenLayers javascript library. As such, we were able to take advantage of the provided draw feature, where users can click on the map and create a series of nodes that define the vertexes of a path. This interaction also allows user to delete, move, and merge existing vertexes. The release of OpenLayers v3.8.0 during the development of OceanPaths provided the added functionality of "snapping vertexes". This was a valuable addition to the OceanPaths user interface as users can now draw branches separately and later join them simply by dragging the nodes close to each other allowing them to "snap" together. Figure 4.3 contains an example of a simple user defined path in the NorthWest Atlantic. The yellow circles indicate the locations where the user clicked, generating vertexes. The dashed blue line indicates the resulting line generated by OpenLayers. The red line is an approximation we implemented in OceanPaths as described in the subsection below.

Bezier Approximation

In order to better represent the path of actual ocean currents, we approximated the line segment provided by OpenLayers (dashed blue line) into a bezier curve, represented in Figure 4.3 as the red curve.

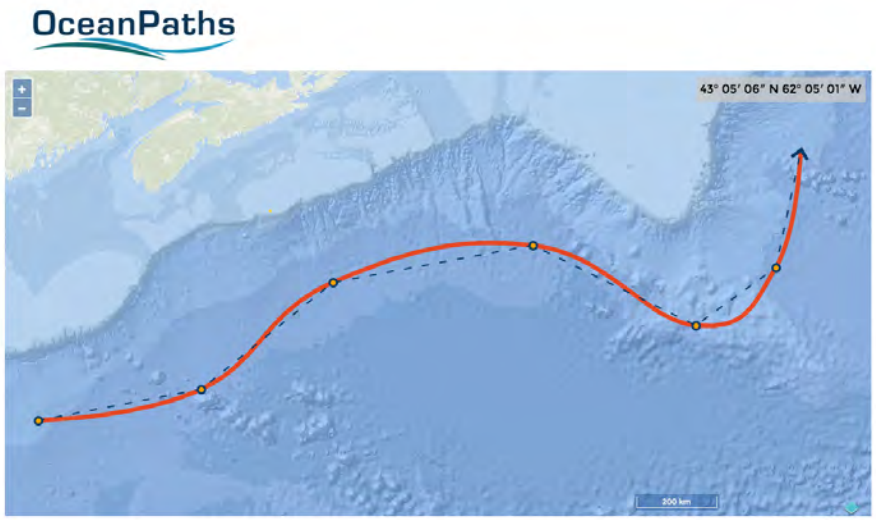

Figure 4.3: Screenshot of user created path in OceanPaths map interface. Yellow circles indicates vertexes as created by the user. The dashed blue line represents the OpenLayers path connecting these vertexes. The red line is a bezier approximation of the blue line to better represent ocean currents.

Branches

One of the major innovations provided by OceanPaths is the ability to create a network of pathways, represented by multiple branches. Figure 4.4 exemplifies a network of three branches in the Pacific Arctic. The main branch is represented by the red path and the branches are drawn in blue. Users can click on any node to create a branch that splits from an existing pathway.

Once a network of paths has been drawn, each branch is represented by a scatter plot in the scatterplot viewport. The positioning of the scatterplots was designed to best reflect their geographic position in the map above. In order to achieve this we used an in order tree traversal to determine the relative ordering of the branches. As a result, a path to the right of another path is shown below it and a path to the left is shown above. Figure 4.5 shows the scatterplot network that resulted from the paths shown in Figure 4.4.

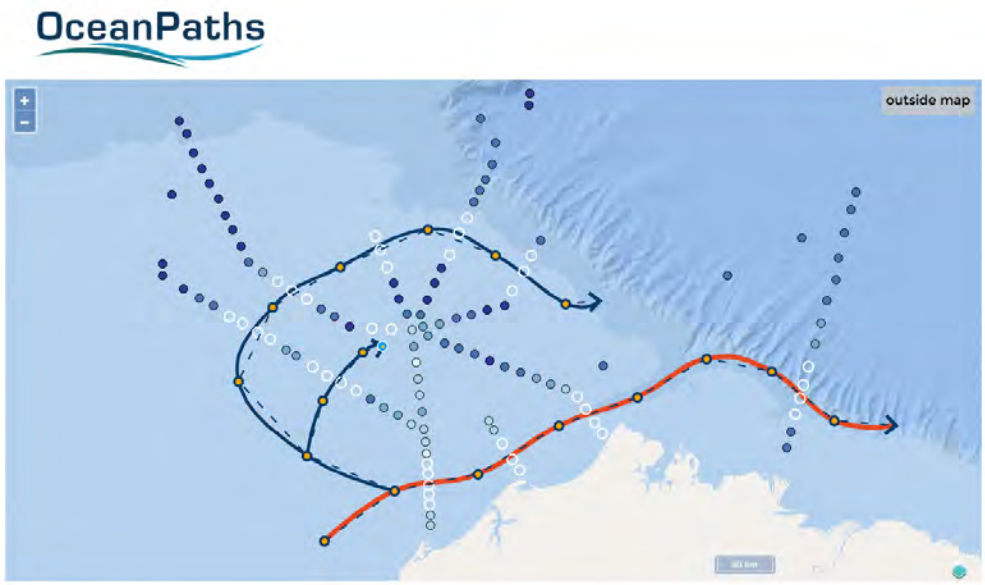

Figure 4.4: Screenshot of a network of paths composed of three branches. The main branch is highlighted in red and the secondary branches are blue.

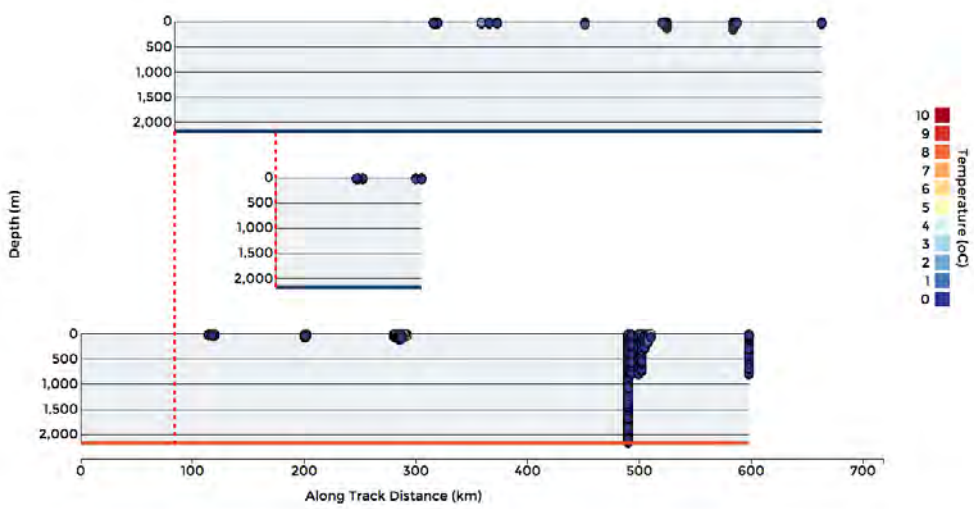

Figure 4.5: Screenshot of scatterplot network associated to the map in Figure 4.4.

4.2. ScatterPlot Network

The Scatterplot Network, shown below the map, displays the selected data as a function of distance along the defined pathways. It consists of individual scatterplots

that correspond to the branches of the network defined in the map, with the (user-defined) main path shown highlighted in red. The positioning of the subplots is defined by the paths 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, as shown in Figure 4.1. 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 vertical axis and by the color of the marks can be defined by the user. In Figure 4.2 the vertical axis represents depth while we show temperature using color. The data displayed in the scatterplot shows a clear cooling of the water column starting at the point where the northernmost branch spawns another branch into the Labrador sea. These cooler waters, represented in the scatterplot in Figure 4.2 by the blue markers appear to be of deeper origin as they are only present at waters deeper than 1000 m.

We determine which measurement point is associated with which exact position along a path using two complementary methods. By default, we automatically select all data points that fall within the width of the pathway. This value can be adjusted. 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. A complementary 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. This allows domain experts to quickly identify the geographic location of a given feature in the ScatterPlot, as well as isolate the ScatterPlot marker for a given measurement on the map.

Chapter 5: Implementation

This chapter covers the implementation details of OceanPaths. We start by discussing the system setup which includes a client side web interface and a server to store and serve the data. We then discuss the choice of using JavaScript/HTML/CSS, followed by a description of the specific libraries used. The remainder of the chapter includes a detailed description of how the pathways were implemented and a discussion on the challenges encountered and our chosen solutions.

5.1. System Setup

As noted by Gasston (2013), web technologies have evolved significantly over the past years. As such, browser based tools have gained traction in several domains, including information visualization. Developing OceanPaths as a browser based tool also obviated the need to install any software and the tool can be easily used by anyone with a computer and an internet connection.

5.2. HTML5

The reasons we chose to use JavaScript as the primary language are three-fold. First, JavaScript has become a ubiquitous language in the realm of web programming (Gasston, 2013). As such, there is a large online community which provides numerous graphic libraries as well as extensive documentation. Next, D3 and OpenLayers, the

two main libraries used in the development of OceanPaths, are JavaScript libraries. The choice of using D3 and OpenLayers is discussed in more detail below.

HTML/CSS was a natural selection for the web interface as they have become the de facto choice for web development. We used the bootstrap template for the HTML/CSS component and made changes to incorporate extra elements such as sliders and menus.

5.3. Libraries

OceanPaths makes use of two main libraries: D3, and OpenLayers. Both are JavaScript libraries and thus integrate seamlessly with OceanPaths.

What follows is a description of each library and how the features are used to achieve the domain tasks described in Chapter 4.

D3

D3 is a JavaScript library for manipulating documents based on data (Bostock et al., 2011). Since its initial release in 2011 it has grown rapidly in usage and is currently a very popular choice for data driven visualizations. We used the D3 library in OceanPaths for (1) loading and parsing out the data and (2) constructing and updating the ScatterPlot Network.

In order to allow users to upload their own data to OceanPaths, we use D3s xhr module to asynchronously load the data into OceanPaths. Once loaded, D3 functions such as d3.json and d3.csv are used to parse, filter, and store the incoming data. When the user updates the pathway in the Map view or selects an attribute from the drop down menus, D3 functions are used to update the plot axis, marker colors, and marker positions in the ScatterPlot Network.

OpenLayers

OpenLayers is an open source JavaScript library for displaying map data in web browsers (Gratier et al., 2015). It is one of several mapping libraries available today and was chosen based on its intuitive API and available functionality and support.

As described in Chapter 2, one of OceanPath's principal features is the ability for users to define current pathways through an interactive interface. OpenLayers provides a suite of "interactions", such as draw, select, and modify, which provided key functionalities with which we were able to implement the OceanPaths map interactivity described in Chapter 4. We present a more detailed description of user interaction for creating pathways below.

5.4. Challenges

The following subsections describe the two major challenges we encountered throughout the development of OceanPaths and our approach to solving them.

Large Datasets

The initial prototype of OceanPaths was developed and tested on a relatively small dataset (1000 data points). In those initial stages we were still developing and optimizing the algorithms that processed and filtered the imported data. However, once we tested the tool on a much larger dataset (10,000 points), we detected a noticeable delay in the tool's response rate as the user's path was drawn. After some investigation, we found that the bulk of the execution time was inside the function that calculated the distance of each measurement to the user-defined path.

We first attempted to solve the problem by loading the data into a SQLite

database and making queries as needed when user interaction occurred. While this did improve the speed of tasks, it did not completely solve the problem. We then refined the algorithm to only calculate the exact distance for measurements within a bounding box drawn around the user path. This approach improved the performance significantly but still showed a reduced response rate for very large datasets.

Our final solution involved the approach above in conjunction with an OpenLayers functionality. The user drawn paths on the map are saved as "features" within the OpenLayers map instance. As such, we were able to take advantage of an OpenLayers function that finds the nearest feature to a given point. Because the execution time of this function was faster than the algorithm we wrote to compute the exact distance of a measurement to each path, we were able to use it to filter out only points that were within the pathway breadth, and then run our distance algorithm on that select subset of points.

Data Sparsity

A common drawback to observational oceanographic data is the potential for geographic sparsity of the measurements. These data points are collected aboard research vessels and when compared to the output of computational models or satellite imagery, are particularly scarce.

As a result, the ScatterPlot Network can suffer from gaps in the "along-path" dimension. In order to address this issue we proposed a collapsing mechanism for the X axis that would get rid of empty regions, thus taking better advantage of the available plotting area. The downside to this approach is that data points that in reality are far from each other are brought closer together on the X dimension. The collapsed regions would have to be particularly noticeable to ensure the analyst did not inadvertently misinterpret the scatterplot.

Ultimately we did not have enough time to implement this proposed solution and have added this challenge to the chapter on possible future work.

Chapter 6: Case Study

In this chapter we present a case study that illustrates the usefulness of OceanPaths for oceanographers to explore the variability and correlation of selected properties in both time and space. We report on an analysis session conducted by an oceanographer, Dr. Bob Pickart, from WHOI. The study was done with hydrographic data collected during two oceanographic expeditions in the Pacific Arctic and aimed to elucidate the circulation on the northeast Chukchi shelf. Figure 6.1 shows a schematic of the current understanding of the circulation in the region.

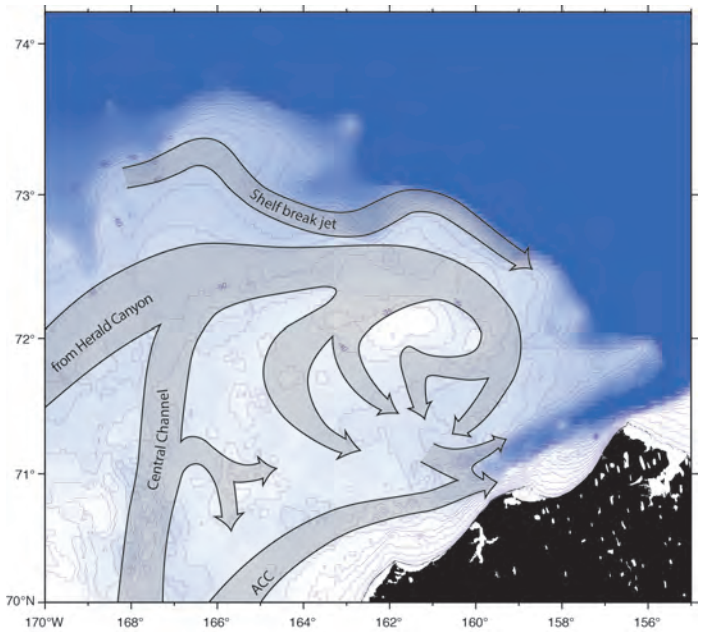

Figure 6.1: A schematic of the circulation of Pacific water in the Chukchi Sea.

Dr. Pickart started his analysis by drawing a path across each of the hydro-

graphic sections occupied. Figure 6.2 shows the resulting configuration in OceanPaths for the Chukchi North section. This interaction exemplifies how OceanPaths can be used to view data along a section, independent of the oceanographic features it represents. It can be particularly useful for oceanographic expeditions where data is acquired as a series of sequential stations along transects.

The chosen path allowed Dr. Pickart to detect two lenses of warmer water flows. The first is located near the coast, and is confined to the top 20m of the water column. The second is further offshore and has a smaller horizontal signature.

The current understanding of the circulation in the region (Figure 6.1) implies that the flow crossing through the western portion of the chukchi north section would progress northwestward, and therefore cross the section taken to the north. In order to confirm this, Dr. Pickart added a pathway along that section, resulting in the configuration shown in Figure 6.3. The resulting ScatterPlot Network confirms the continued presence of the warmer lens of water as the water progresses northwestward through the two sections.

After analyzing the individual transects, Dr. Pickart relied on a combination of the temperature distribution, the bathymetry information, and his pre-existing knowledge of the area to construct a flow schematic of what he believed to be representative of the progression of water in the region. Once the pathway was determined, he used the automatic selection of data points to choose the points that fell within the pathways and were therefore relevant for the ensuing analysis of property evolution along the central pathway. He took advantage of the single point data selection toggle to exclude the portion of the flow that emanated from the west since that water has a different source than the rest of the sampled data points in the pathway. When generating the accompanying scatter plot, Dr. Pickart selected depth as the vertical axis and temperature as color. The resulting displays are shown in Figure 6.4.

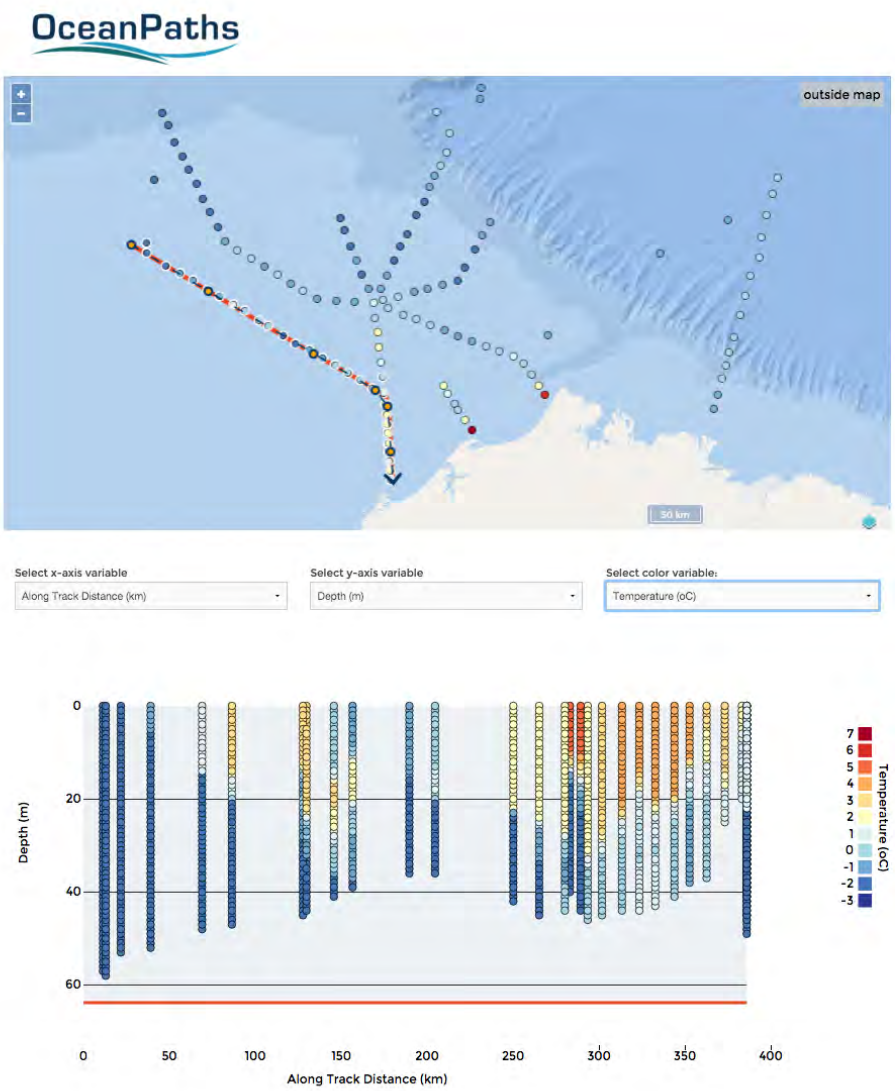

Figure 6.2: OceanPath configuration for the highlighted transect.

A particularly notable feature was that the warmer waters present in the two sections analyzed earlier (highlighted branch A in the map of Figure 6.4) were not present farther along that pathway. Instead, those warmer waters appear to have branched off and can be seen in the highlighted portions of branches B and C.

In order to pinpoint the spatial location of the warm water lenses, he used his mouse to hover over the area in the scatter plot revealing the appropriate stations

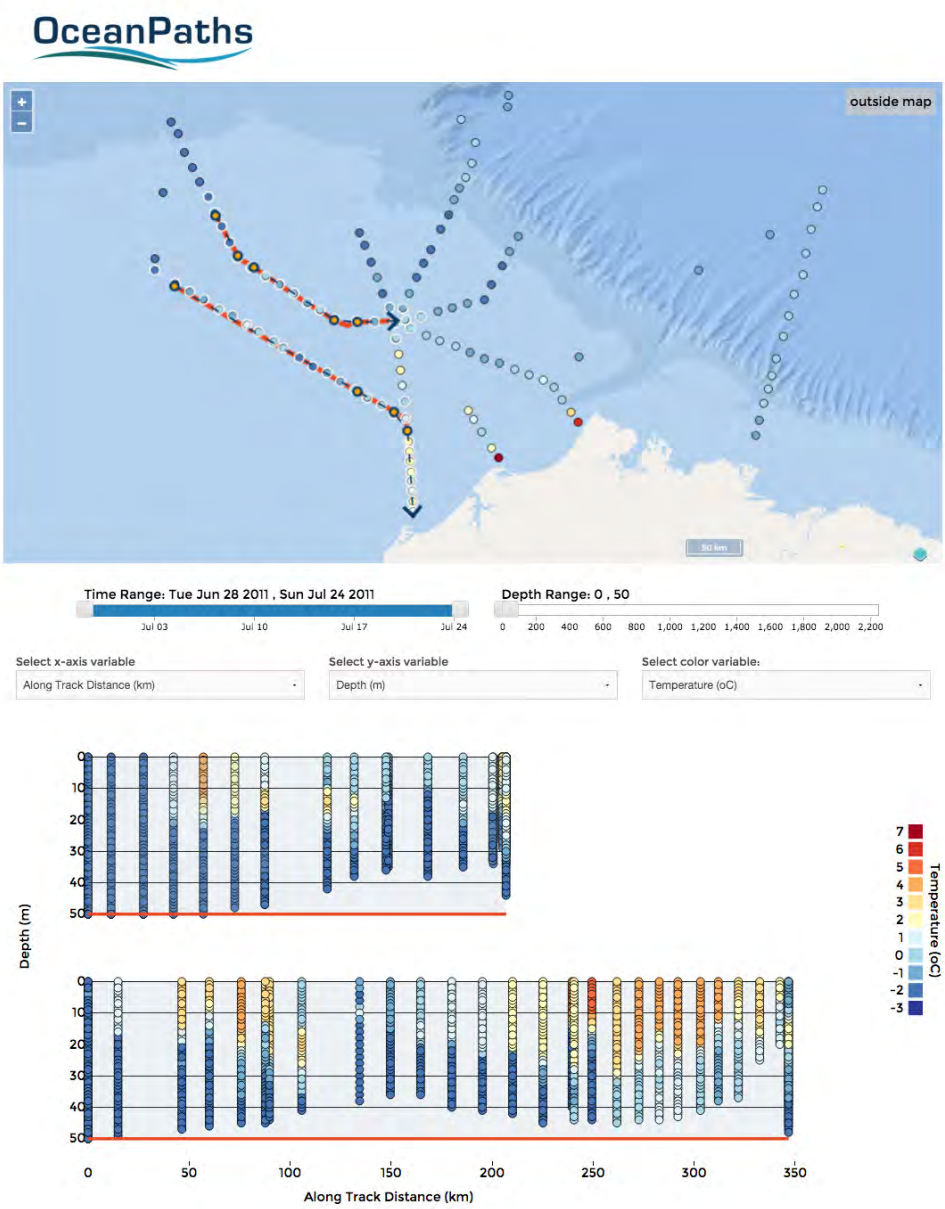

Figure 6.3: OceanPath configuration for the highlighted transects.

in the map display. Dr.Pickart used this information to confirm existing hypothesis that water masses traveling north through the main channel can be diverted by smaller branches, thus affecting the northward transport of heat. The existence of these branches is most likely due to the presence of a shoal to the northeast of the

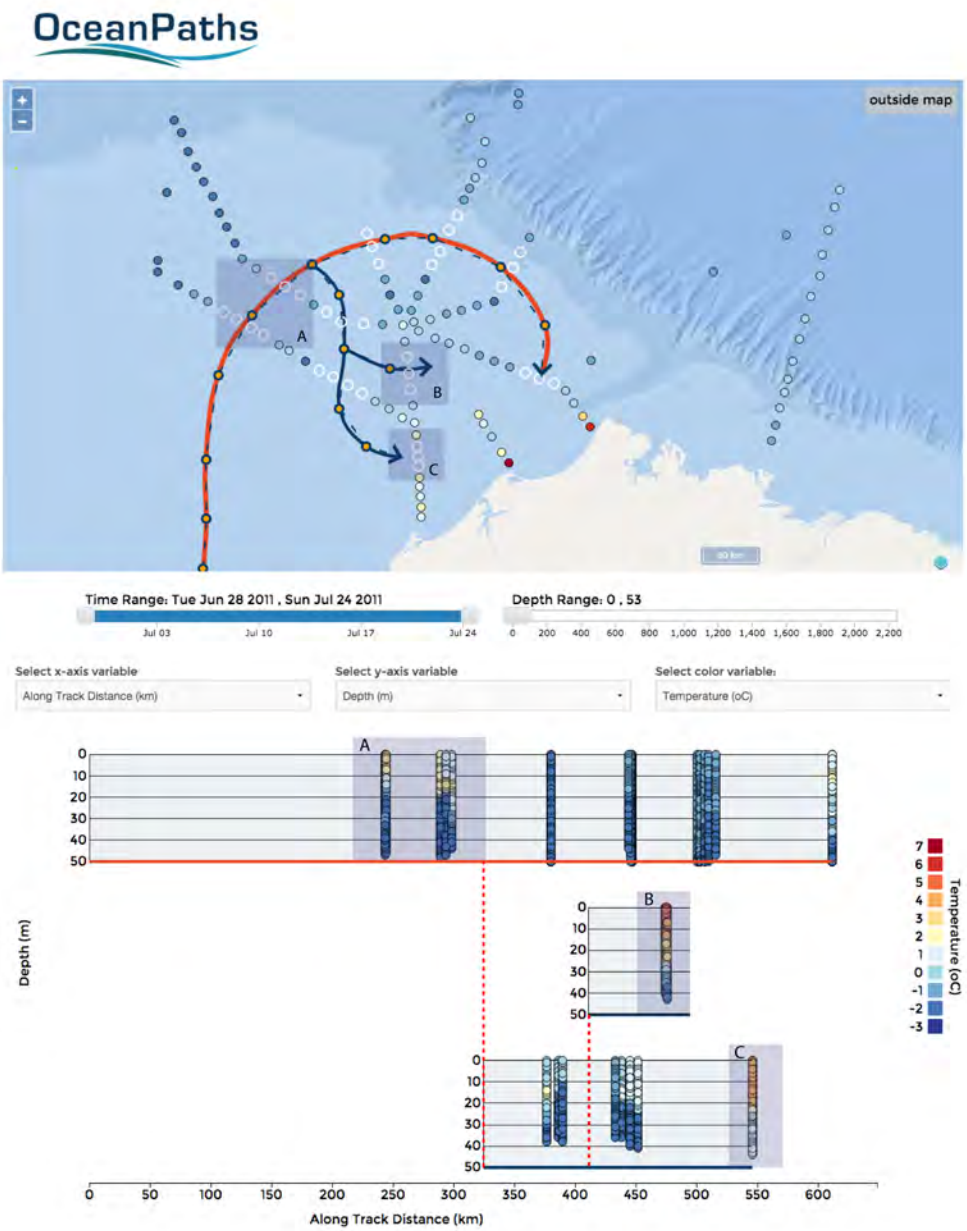

Figure 6.4: OceanPath configuration for the current system in the western Chukchi Sea. Highlighted areas in the map correspond to the highlighted areas in the scatterplot network below.

branching point. As noted in the literature, bathymetry can often play an important role in determining current paths. Dr. Pickart found that the bathymetric infor-

mation often informed his decisions when generating pathways and interpreting the corresponding data points. He concluded the study by noting that this close inspection of properties along the determined pathway was a great facilitator to interpreting property changes and both corroborating existing theories and potentially generating new ones in regards to the circulation in the area.

Chapter 7: Conclusions and Future Work

7.1. Summary and Conclusions

During the course of this project we worked closely with oceanographers to develop an exploratory visual analysis tool designed to assist in understanding how water-properties vary in space and time along currents. We initiated the study with a series of interviews with domain specialists, which allowed us to articulate the domain tasks and system requirements for the tool. The resulting OceanPaths tool facilitates the analysis of multidimensional oceanographic data by allowing users to define geographic pathways and plot available data as a function of distance along these custom pathways.

Despite the previous existence of studies which address multi-variable datasets in earth sciences, OceanPaths is a unique and novel solution for reducing the dimensionality of spatial oceanographic data by condensing geographic coordinates into a network of meaningful paths. Moreover, the application of well-documented visualization techniques such as brushing, linked plots, as well as customizable axes, applied specifically to the domain of oceanographic data, proved to be a valuable tool enabling a more efficient process of data exploration and a better understanding of the spatial distribution of prominent features

OceanPaths was used by a physical oceanographer for the case study presented in Chapter 6. The response of the domain expert was highly positive, verifying that

our visual framework has the potential to complement traditional methods used by oceanographers in exploring multidimensional data.

7.2. Future Work

Including methods to increase scalability of the Scatterplot Network would be interesting future work to augment OceanPaths current capabilities. Introducing more compact representations for the network view and only showing the full scatterplots on-demand would be one way to do so. Another potential improvement would address the issue of data sparsity, a common attribute of observational oceanographic data, by exploring different interpolation techniques and by enabling analysts to collapse empty regions, so that more space is used for denser parts.

We plan on making OceanPaths available to interested parties in early 2016, shortly after the publication of this thesis.

References

Anderson, J. C., Andres, J. M., Davis, M., Fujiwara, K., Fang, T., & Nedbal, M. (2010). Voyager: An Interactive Software for Visualizing Large, Geospatial Data Sets. *Marine Technology Society Journal*, 44(4), 8–19.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301–2309.

Brewer, P. G., Goyet, C., & Dyrssen, D. (1989). Carbon Dioxide Transport by Ocean Currents at 25 N Latitude in the Atlantic Ocean. *Science*, 246(4929), 477–479.

Brown, M. (1998). Ocean Data View 4.0. *Oceanography*, 11(2), 19–21.

Burakiewicz, W. & 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), 1037–1043.

Charette, M. A., Breier, C. F., Henderson, P. B., Pike, S. M., Rypina, I. I., Jayne, S. R., & Buesseler, K. O. (2013). Radium-based estimates of cesium isotope transport and total direct ocean discharges from the Fukushima Nuclear Power Plant accident. *Biogeosciences*, 10(3), 2159–2167.

Co, C. S., Friedman, A., Grote, D. P., Vay, J.-L., Bethel, E. W., & Joy, K. I. (2004). Interactive methods for exploring particle simulation data. *Lawrence Berkeley National Laboratory*.

Deines, E., Michel, F., Bertram, M., Hagen, H., & Nielson, G. M. (2006). Visualizing the Phonon Map. In *Proceedings of the Eighth Joint Eurographics / IEEE VGTC Conference on Visualization*, EUROVIS'06 (pp. 291–298). Aire-la-Ville, Switzerland, Switzerland: Eurographics Association.

Gasston, P. (2013). *The Modern Web: Multi-device Web Development with HTML5, CSS3, and JavaScript*. No Starch Press.

Giese, B. S. & Ray, S. (2011). El Nio variability in simple ocean data assimilation (SODA), 18712008. *Journal of Geophysical Research: Oceans*, 116(C2), C02024.

Gratier, T., Spencer, P., & Hazzard, E. (2015). *OpenLayers 3 : Beginner's Guide*. Packt Publishing.

Jankun-Kelly, T. J. & 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., & 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., & Hanson, A. (2008). Visualizing Multiwavelength Astrophysical Data. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1555–1562.

Lipsa, D. R., Laramee, R. S., Cox, S. J., Roberts, J. C., Walker, R., Borkin, M. A., & Pfister, H. (2012). Visualization for the Physical Sciences. *Computer Graphics Forum*, 31(8), 2317–2347.

Mathis, J. T., Pickart, R. S., Hansell, D. A., Kadko, D., & Bates, N. R. (2007). Eddy transport of organic carbon and nutrients from the Chukchi Shelf: Impact on the upper halocline of the western Arctic Ocean. *Journal of Geophysical Research: Oceans*, 112(C5), C05011.

McCreary, J. P., Kohler, K. E., Hood, R. R., Smith, S., Kindle, J., Fischer, A. S., & Weller, R. A. (2001). Influences of diurnal and intraseasonal forcing on mixed-layer and biological variability in the central Arabian Sea. *Journal of Geophysical Research: Oceans*, 106(C4), 7139–7155.

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 Suppl 19, S3.

Prants, S. V., Uleysky, M. Y., & Budyansky, M. V. (2011). Numerical simulation of propagation of radioactive pollution in the ocean from the Fukushima Dai-ichi nuclear power plant. *Doklady Earth Sciences*, 439(2), 1179–1182.

Schlitzer, R. (2002). Interactive analysis and visualization of geoscience data with Ocean Data View. *Computers & Geosciences*, 28(10), 1211–1218.

Schlosser, P., Swift, J. H., Lewis, D., & Pfirman, S. L. (1995). The role of the large-scale Arctic Ocean circulation in the transport of contaminants. *Deep Sea Research Part II: Topical Studies in Oceanography*, 42(6), 1341–1367.

Schuur, C. L., Coffin, M. F., Frohlich, C., Massell, C. G., Karner, G. D., Ramsay, D., & Caress, D. W. (1998). Sedimentary regimes at the Macquarie Ridge Complex: Interaction of Southern Ocean circulation and plate boundary bathymetry. *Paleoceanography*, 13(6), 646–670.

Shetye, S. R. (1982). *Effect of Bathymetry on the Wind- and Thermohaline-Driven Ocean Circulation*. Ph.D., University of Washington, United States – Washington.

Spence, P., Fyfe, J. C., Montenegro, A., & Weaver, A. J. (2010). Southern Ocean Response to Strengthening Winds in an Eddy-Permitting Global Climate Model. *Journal of Climate*, 23(19), 5332–5343.

Vage, K., Pickart, R. S., Sarafanov, A., Knutsen, ., Mercier, H., Lherminier, P., van Aken, H. M., Meincke, J., Quadfasel, D., & Bacon, S. (2011). The Irminger Gyre: Circulation, convection, and interannual variability. *Deep Sea Research Part I: Oceanographic Research Papers*, 58(5), 590–614.

Yuan, X., Xiao, H., Guo, H., Guo, P., Kendall, W., Huang, J., & Zhang, Y. (2010). Scalable Multi-variate Analytics of Seismic and Satellite-based Observational Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1413–1420.