Improving Scientific Data Analysis Through Multi-touch Enabled Interactive Data Visualization with Applications to Neutron Science

Chad A. Steed* Junghoon Chae† John Goodall‡ Steven Hahn§

Oak Ridge National Laboratory
Oak Ridge, TN USA

Abstract

Scientific data analysis is an inherently exploratory process that requires flexible human interaction with data visualization and analysis tools. In this paper, we describe challenges in neutron scattering data analysis and present our approach for allowing more effective human interaction. In addition to multi-touch enabled data visualization techniques, we describe color scale enhancements, layouts, and visual analytics approaches for 3D volume slicing. Since challenges in neutron science are echoed in most scientific fields, we generalize the project objective and present a case for bridging the gap between the latest advances in interactive data visualization and practical scientific scenarios where new critical knowledge is buried in large volumes of complex data.

1 Introduction

Present-day scientific data, whether collected during large-scale experiments, observed by sophisticated sensors, or simulated on supercomputers, hold tremendous potential to solve some of the most pressing scientific challenges. The discovery of new insight, which enables such breakthroughs, hinges upon the accessibility of interactive data visualization and analysis techniques that integrate human direction with machine guidance. However, research and development activities for scientific data analysis disproportionately focus on scalability challenges and rarely consider human interaction techniques. Conversely, fields that focus on human interaction (e.g., information visualization, human-computer interaction) usually apply new techniques to moderately-sized and/or tailored data and rarely involve practical scientific scenarios. Therefore, a chasm exists between interactive data visualization and scientific data analysis applications. By bridging this chasm, we can uncover new depths of knowledge in the most challenging scientific scenarios.

One such scenario exists at Oak Ridge National Laboratory's (ORNL) Spallation Neutron Source (SNS) facility. Researchers from across the globe visit the SNS to execute experiments for scientific research and industrial developments using the most intense pulsed neutron beams in the world. Given the expense and limited time for SNS experiments, interactive data visualization and analysis capabilities that are timely and straightforward are required. Although mainstream scientific data analysis tools are powerful, their interfaces are difficult to learn and use.

*e-mail: csteed@acm.org
†e-mail: chaej@ornl.gov
‡e-mail: goodall@ornl.gov
§e-mail: hahnse@ornl.gov

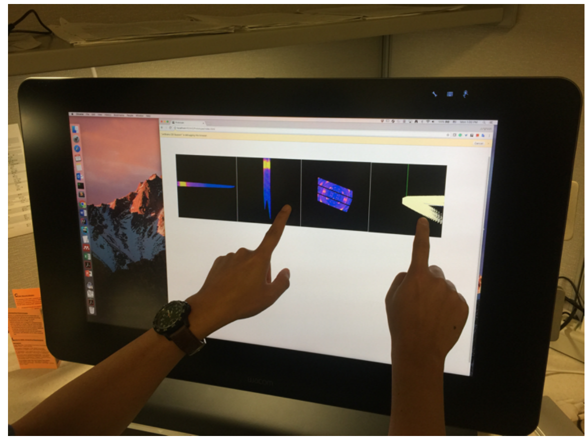
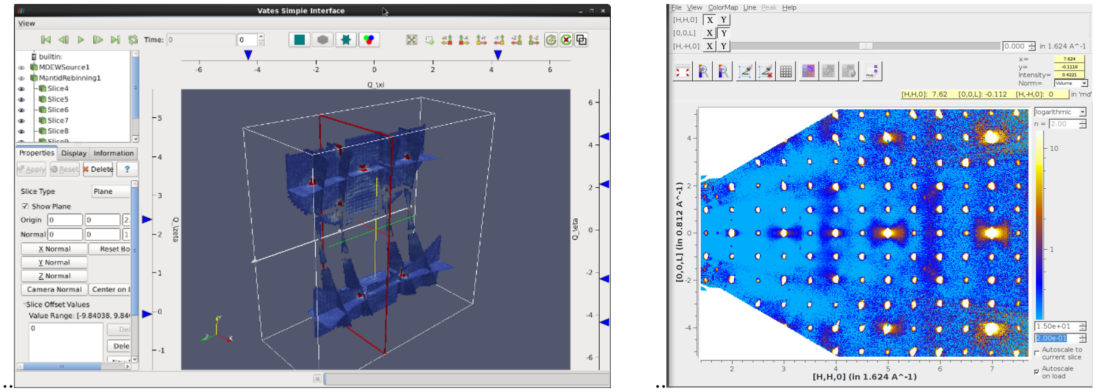


Fig. 1: We are developing new multi-touch interactive data visualization techniques using a web-based prototype to enable more efficient data analysis in neutron scattering science. We are using a Wacom Cintiq 27QHD display and neutron scattering data.

Our collaborations with SNS researchers have revealed that a reliance on mouse-based interaction strategies in scientific data analysis tools hinders the discovery of new knowledge discovery. For example, volume slicing, which involves the extraction of a cross-section from a 3D data volume, requires manual adjustments of several parameters, the number of which exceed the inputs offered by mouse gestures. Application developers resort to a hierarchy of menus and toolbars adjacent to the visualization window, which produces a cumbersome user experience. Nevertheless, SNS users often state that finding new insight in their data hinges on picking the right slice(s).

With this situation in mind, we are developing new interactive data visualization techniques that leverage multi-touch interactions on a high-resolution display in a web-based prototype (see Figure 1). The goal is to increase the efficiency of volume slicing operations on large and multi-dimensional data from neutron scattering experiments. We postulate that such techniques will reduce knowledge discovery time cycles, increase the efficiency of scientific data analysis, and have broad applications to other scientific domains.

This paper highlights the transformative potential of interactive data visualization and analysis for science by focusing a particular challenge in a specific field. Section 2 provides an overview of neutron scattering science at SNS, characteristics of the experiment data, and several data analysis challenges. In Section 3, we present our approach for improving volume slicing in neutron science as a work-in-progress. In Section 4, we discuss additional challenges and relationships with our prior work. Section 5 concludes the paper and provides plans for future work.



(a) Mantid Volume Slice Interface (VSI)

(b) Mantid 2D SliceViewer

Fig. 2: The Mantid data analysis system provides both a 3D (a) and 2D (b) visual interface for volume slice analysis.

2 Background

As advances in computing and experimental facilities feed surges in the size and complexity of scientific data, the transformative potential of interactive data visualization and analysis also rises. Due to the exploratory nature of scientific studies, researchers need flexible tools that allow them to ask questions of their data and generate new hypotheses. In addition to dynamic linkages between automated analytics and data visualizations, human interactions that are embedded within data visualizations are inadequate in mainstream scientific data analysis systems. The current work focuses on addressing this issue for a specific scientific data analysis scenario: volume slicing during neutron scattering data analysis. In the remainder of this section, we provide a domain overview, data characteristics, and specific data analysis challenges at the SNS.

2.1 Neutron Scattering Science at the ORNL SNS

Our project team includes neutron scattering researchers from the ORNL SNS, a unique Department of Energy (DOE) user facility that offers the most intense pulsed neutron beams in the world for scientific research and industrial developments. The SNS produces neutrons with an accelerator-based system that delivers short (microsecond) proton pulses to a target system, where neutrons are produced by a process called spallation. These neutrons are directed to experiment stations that host a variety of sensing capabilities for researchers from diverse disciplines, such as physics, chemistry, materials science, and biology. With its more intense, brighter source of neutrons and instrumentation, SNS allows measurements of greater sensitivity, higher speed, higher resolutions, and more complex sample environments than other existing neutron facilities. Neutron scattering experiments at ORNL facilities have enabled numerous scientific discoveries, such as an enzyme reaction that improve biofuel production [10] and tRNA-nanodiamond combinations that could transform drug delivery design principles [6]

2.2 Neutron Scattering Data

Researchers use SNS neutron scattering instruments to determine complex crystal structures that are described by unit cells with tens to millions of atoms. Experiments generate data sets that are large, multi-dimensional, and involve complex transformations. The size of data files produced in a single experiment may vary from hundreds of megabytes to hundreds of gigabytes, with a typical file being about ten gigabytes. Depending on the experimental configuration, billions of neutron events are measured and transformed into scientifically relevant dimensions, three of which are related to the spatial location (x, y, and z) [9, 11].

2.3 Limitations of Scientific Data Analysis Tools

The unique characteristics of neutron scattering data often limit the utility of commercial systems (e.g., Tableau, Excel) that offer basic data visualization and analysis capabilities or tailor to more abstract and smaller scale data. Although several scientific data visualization and analysis tools are available (e.g., ParaView, VisIt, Mantid [2], Horace [8], DAVE [3]), challenges remain that can only be addressed through more interactive techniques. Developed by the SNS staff, Mantid is a customization of ParaView with specific neutron science capabilities.

Visiting SNS researchers are experts in their respective fields (e.g. materials science, chemistry, biology) but often inexperienced with advanced scientific data visualization and analysis tools. This scenario, which is partially due to complex user interface designs (see Figure 2), handicaps the researcher's ability to quickly and accurately explore experiment data. Indeed, researchers state that they "don't look at their data enough" due to the complexity associated with using the tools. Perhaps with more intuitive interactions, immersive experiences, and analytical assistance, researchers would spend less time with each individual inquiry during an analysis session and use the extra time to consider more of the data. That is, we expect the application of more efficient interactive data visualization and analysis techniques to yield more thorough and accurate investigations in the same way that physical interactions with data visualizations for large scale displays improved user performance during document analysis tasks in studies by Ball et al. [4] and Andrews et al. [1].

SNS users also state that although they are good at finding patterns that are already known, unexpected discoveries are slow to occur. Known patterns may include established theories or the original ideas that influenced the experimental design. Although confirmation of such patterns is important, researchers are most interested in discovering new, and often unanticipated, insights that lift the fog of uncertainty clouding some realms of modern science. Indeed, historical reflections show that the most impactful discoveries are often of an unexpected nature [5]. Again, this observation hints at the promise of more efficient interactive data visualizations as well as automated data analytics that are directly linked to data visualizations to highlight potentially significant patterns.

If researchers have access to tools that offer more intuitive interactions embedded within the visualization, they will enjoy flexibility to explore their data, consider more patterns in greater depth, and presumably uncover profound insights. Moreover, the judicious combination of such human interactive capabilities with advanced scientific processes holds the potential to transform all scientific domains, let alone neutron science. But without such tools, researchers will fall further behind the curve as both data sizes and complexities escalate. If we are unable to exploit the vast data sets, these size and complexities increases will hinder progress and waste the power of user facilities and supercomputing centers. Given the operating expenses and limited time to conduct experiments, it is critical that researchers have timely access to advanced analysis techniques. The development of new visual data analysis tools, especially those that provide human interaction and/or immersive experiences, will remove this barrier to scientific knowledge discovery by increasing the user's ability to quickly and accurately find new insights and solve the most pressing scientific challenges.

3 APPROACH

In light of these challenges, we are collaborating with SNS neutron researchers to develop and evaluate new interactive data visualization techniques that leverage multi-touch enabled, high-resolution display technology. In particular, we are focusing on improving the efficacy of a specific visual analysis technique called volume slicing, which SNS users state is critical to understanding neutron scattering data. Volume slicing is applied to a 3D volume of data. The user extracts a slice by specifying a 2D plane, which is intersected with the 3D volume. Usually, the 2D plane is orthogonal to either the *x*, *y*, or *z* axis and multiple slices are sometimes extracted and viewed together. The slice extraction yields a 2D array of values that are visually encoded as colored pixels in an image. The discovery of important patterns in neutron scattering experiments hinges on selecting the right slice and applying a color scale that maps the widest range of colors to the most important range of values. In both cases, the background knowledge and intuition of the researchers are the primary drivers for finding interesting slices.

Finding interesting slices and an optimal color scale is a complicated process, especially for novice users. As shown in Figure 2a, Mantid's Volume Slice Interface (VSI) requires users to specify several parameters. In fact, the number of parameters exceeds the inputs available with mouse-based gestures. To allow parameter adjustments, a series of menus and toolbars are displayed and linked to the 3D view. In addition to occupying precious screen space that could be devoted to the data visualization, these menus and toolbars require the user to manually specify slice positions, color scale mappings, and other related parameters that control the slice extraction using text fields, sliders, and other user interface components. The current VSI interactions yield an ineffective user experience that inhibits direct human interactions with the data visualization and delays visual perception of key patterns. The VSI primarily uses a 3D view, which helps maintain spatial awareness of the slice. However, the 3D view subjects the user to occlusion effects, perspective distortions, and navigational issues. As a result, some users reject the VSI and opt for a more basic 2D slice view (see Figure 2b).

We are currently developing new human interaction capabilities for adjusting the volume slice operation parameters using multi-touch capabilities on a Wacom Cintiq 27QHD display (see Figure 1). In addition to multi-touch interactions, the Cintiq 27QHD provides pressure sensitive pen inputs, programmable buttons, and it supports display resolutions up to 2560x1440 pixels. Our prototype currently focuses on multi-touch interactions in both 2D and 3D views of neutron scattering data to control the location and viewing parameters of volume slices.

In addition to positional, transparency, and thickness parameters, we are designing interaction schemes for adjusting the color scale settings. For example, we have developed touch interactions that allow the user to adjust the color scale mappings and dedicate a wider range of colors to the data values. Furthermore, multi-touch interactions with the volume slice visualization will allow the user to create and adjust volume slice planes dynamically. All interaction techniques are embedded within the visualization, meaning the user touches the graphical objects that are rendered in the data visualization instead of using adjacent dialogs and toolbars. The prototype is implemented as a web-based application that uses both three.js and D3.js JavaScript libraries.

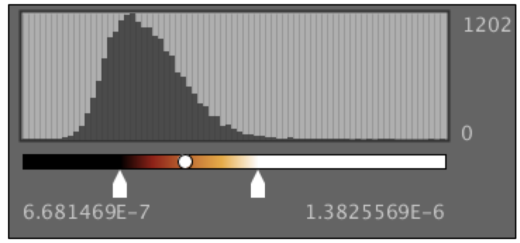
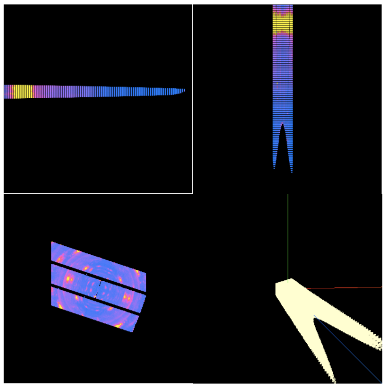


Fig. 3: We have developed a statistical technique to automatically set the the high, middle, and low color values of a color scale to the 25th, 50th, and 75th percentiles of values, respectively. The adjustable value mappings and a histogram of values are shown in the lower left corner of this volume slicing prototype.

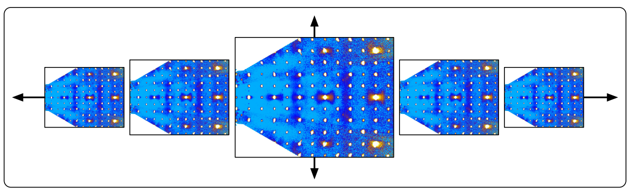
To evaluate the effectiveness of embedded multi-touch interactions, the prototype operates in either a mouse-based or a touch-based interaction mode. The mouse-based mode operates like the original VSI interface using a combination of mouse, keyboard, and menus to control view parameters. When complete, this prototype will establish a baseline to enable evaluations of the touch-based interactions with SNS researchers in real data analysis scenarios. We expect to utilize both empirical user studies and more informal feedback from domain experts in these evaluations. Anticipated measures of usability include scientific accuracy, performance time, and display space utilization.

In addition to evaluating the embedded touch-based interactions, the prototype enables studies of additional ideas for scientific data analysis. For example, we have developed a statistical technique that automatically sets the initial color scale mappings using the interquartile range and median values for the overall data distribution (see Figure 3). The approach maps the 25th and 75th percentiles to the high and low color values and the 50th percentile, or median, to a middle color value. The objective of this process is to improve the initial view of the data, which is often skewed by large outliers or peaks, by assigning the maximum range of colors to the bulk of the data value space. Through direct interactions with the color scale, the user can adjust these settings. Additionally, we are designing new view layouts and browsing modes to improve the user's ability to scan more slices in the data. As shown in Figure 4a, our initial prototype used a volume slicing view with a grid of four views, one of which is a 3D view. We are exploring alternative views that remove the reliance on a 3D view due to the perception issues. We believe a 3D view may still be used for spatial awareness, but it can play more of a supplement role. As shown in Figure 4b, our concept would provide a carousel of slice images, that the user can swipe through. A distortion function is applied to make the current slice larger and the neighboring slices progressively smaller based on the distance from the current slice. Adjacent slices can be shown either as flat images or slightly rotated to convey the 3D spatial arrangement. Our goal is to provide spatial awareness in the 2D views through linking and interaction and thereby reduce, or perhaps even remove, the need for a 3D view. Multiple slice directions can be stacked using juxtaposition.

Another more quantitative question that is of interest to SNS tool developers is related to the scale limits of a web-based visual analysis approach. Using three.js, graphics are rendered using WebGL for more scalable performance. However, we have encountered file size limits and we plan to document the rendering performance as the volume of data increases. In the process, we will identify bottlenecks in the rendering performance and devise new strategies



(a) Grid Layout Prototype



(b) Proposed Horizontal Slice Layout Using a Carousel Metaphor

Fig. 4: The normal slice view uses a grid layout (a), but we are exploring alternative layouts such as a horizontal layout for each slice plane (b) that lets the user see effects of changing the slice location without interacting with a 3D view.

for enabling larger scale analysis on the web. Existing scientific data analysis tools are rarely provided as web based applications, but the accessibility and ease of deployment make this an important goal in the SNS's mission to support users.

4 Discussion

There are many opportunities to improve visual data analysis in scientific domains. In particular, domains that study data from experimental or observational systems need more effective human interaction techniques that enable exploratory data analysis. In earlier work with our SNS collaborators, we developed an immersive visual analysis prototype using a head-mounted virtual reality device that gave researchers the ability to fly-through and interact with with 3D models of diffuse scattering data [7, 12]. In several demonstrations of the immersive system to the SNS community, we observed that researchers welcome new visual analysis ideas and they are eager to evaluate new approaches, especially if they are given the opportunity to share their ideas and feedback on the system design. One of the main issues we encountered in our virtual reality experiments related to the practicality of using the equipment at a scientific user facility. In this project, we focus on large, touch-based displays that can be easily connected to existing workstation or conference room scenarios, which can be used in conjunction with virtual or augmented reality equipment.

By building bridges of collaboration with experts in scientific domains, we can translate the latest advances in interactive data visualization into practice in data-rich fields that desperately need help. In doing so, data visualization researchers will encounter challenges related to the scalability of their techniques, especially with respect to maintaining responsive interaction. Although it is not always the case, scientific data is often much larger than the data used in information visualization or visual analytics research. As the scientific visualization community has primarily focused on scale issues, the coalesce of scientific and information visualization expertise, as well as visual analytics approaches, promises to provide a tremendous boost to the data analysis endeavors in most fields.

In particular, a visual analytics approach, which is often a foreign concept to researchers, is particularly promising. In neutron scattering and similar scientific domains, the questions asked of the data are too exploratory for a completely automated solution. While researchers may have some specific patterns in mind, they often don't know all the questions that can be asked of the data beforehand and true breakthroughs are often unexpected. Thus, analysis techniques must allow humans to direct the analytical discourse and flexibly ask questions of the data at variable scales. On the other hand, the size and complexity of the data preclude an entirely manual process as a human cannot feasibly look at all of the data. Automated analytics are necessary, which reduce the search space and guide the human to potentially important patterns in the data. Therefore, a visual analytics approach that judiciously combines human exploration strengths with computation power is a necessity rather than an option for realizing the full potential of extreme scale scientific data. In the data visualization systems that researchers presently use, the analysis process is either completely manual or automated. Researchers are eager to adopt such systems, but data visualization experts are needed to realize a new class of tools. In the second phase of this project, we will explore this idea more thoroughly. We plan to extend our prior work [13, 14], which leverages human interactions with data visualizations to feed semi-supervised machine learning algorithms that infer user intent and highlight patterns.

5 Conclusion

In this project, we are developing new interactive data visualization and analysis techniques that allow SNS researchers to unlock the full potential of the large and complex data they analyze for neutron scattering experiments. The need for such tools is felt in most scientific domains. In the future, we will explore new visual representations and view layouts that are designed for the larger pixel resolutions of modern workstation displays. We will also incorporate machine learning techniques to infer intent and automatically find patterns of interest by monitoring the researcher's interaction with the tool. By making such tools a reality, we will allow researchers to achieve breakthroughs and deliver on the central promise of data visualization–amplified cognition.

Acknowledgments

This work is sponsored by Oak Ridge National Laboratory LDRD project No. 8711. A portion of this research at ORNL's Spallation Neutron Source was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, US Department of Energy. We thank Songxue Chi, Feng Ye and Wei Bao for providing inelastic neutron scattering data of (Tl,Rb)2Fe4Se5 measured on the SEQUOIA instrument at SNS and Feng Ye for providing single crystal diffraction data measured on the CORELLI instrument at SNS.

References

[1] C. Andrews, A. Endert, and C. North. Space to think: Large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 55–64, 2010.

[2] O. Arnold, J. Bilheux, J. Borreguero, A. Buts, S. Campbell, L. Chapon, M. Doucet, N. Draper, R. F. Leal, M. Gigg, V. Lynch, A. Markvardsen, D. Mikkelson, R. Mikkelson, R. Miller, K. Palmen, P. Parker, G. Passos, T. Perring, P. Peterson, S. Ren, M. Reuter, A. Savici, J. Taylor, R. Taylor, R. Tolchenov, W. Zhou, and J. Zikovsky. Mantid—data analysis and visualization package for neutron scattering and sr experiments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 764:156–166, 2014. doi: 10.1016/j.nima.2014.07.029

[3] R. T. Azuah, L. R. Kneller, Y. Qiu, P. L. Tregenna-Piggott, C. M. Brown, J. R. Copley, and R. M. Dimeo. Dave: a comprehensive software suite for the reduction, visualization, and analysis of low energy neutron spectroscopic data. *Journal of Research of the National Institute of Standards and Technology*, 114(6):341, 2009.

[4] R. Ball, C. North, and D. A. Bowman. Move to improve: Promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 191–200, 2007.

[5] W. I. B. Beaveridge. *The Art of Scientific Investigation*. The Blackburn Press, Caldwell, NJ, 1957.

[6] G. K. Dhindsa, D. Bhowmik, M. Goswami, H. O'Neill, E. Mamontov, B. G. Sumpter, L. Hong, P. Ganesh, and X.-q. Chu. Enhanced dynamics of hydrated tRNA on nanodiamond surfaces: A combined neutron scattering and md simulation study. *The Journal of Physical Chemistry B*, 120(38):10059–10068, 2016.

[7] M. Drouhard, C. Steed, S. Hahn, T. Proffen, J. Daniel, and M. Matheson. Immersive visualization for materials science data analysis using the oculus rift. In *Proceedings of the IEEE International Conference on Big Data*, pp. 2453–2461, 2015.

[8] R. Ewings, A. Buts, M. Le, J. van Duijn, I. Bustinduy, and T. Perring. Horace: Software for the analysis of data from single crystal spectroscopy experiments at time-of-flight neutron instruments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 834:132–142, 2016. doi: 10.1016/j.nima.2016.07.036

[9] M. D. Lumsden, J. Robertson, and M. Yethiraj. Ub matrix implementation for inelastic neutron scattering experiments. *Journal of applied crystallography*, 38(3):405–411, 2005.

[10] W. B. O'Dell, P. K. Agarwal, and F. Meilleur. Oxygen activation at the active site of a fungal lytic polysaccharide monooxygenase. *Angewandte Chemie International Edition*, 56(3):767–770, 2017.

[11] P. F. Peterson, S. I. Campbell, M. A. Reuter, R. J. Taylor, and J. Zikovsky. Event-based processing of neutron scattering data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 803:24–28, 2015.

[12] C. A. Steed, J. Daniel, M. Drouhard, S. Hahn, and T. Proffen. Immersive visual analytics for transformative neutron scattering science. In *IEEE Workshop on Immersive Analytics (IA)*, pp. 38–43, 2016.

[13] C. A. Steed, J. E. Swan, T. J. Jankun-Kelly, and P. J. Fitzpatrick. Guided analysis of hurricane trends using statistical processes integrated with interactive parallel coordinates. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pp. 19–26, 2009.

[14] C. A. Steed, C. T. Symons, F. A. DeNap, and T. E. Potok. Guided text analysis using adaptive visual analytics, 2012.