A Survey of Visualization Techniques and Tools for Environmental Data

Likhitha Ravi, Qiping Yan
College of Engineering
University of Nevada
Reno, USA
ravi@cse.unr.edu, qiping@aol.com

Sergiu M. Dascalu, Frederick C. Harris, Jr.
College of Engineering
University of Nevada
Reno, USA
dascalus@cse.unr.edu, fredh@cse.unr.edu

***Abstract*— Visualizing a set of discreet data is often needed in scientific research. In the case of environmental data, it is quite challenging to predict the future variation of a particular variable such as temperature, pressure, and precipitation. Many new visualization techniques and tools are available in the field of software visualization for analyzing large datasets. However, selecting a right tool that will fulfill user requirements for visualizing large datasets is still challenging. This paper reports the findings of a survey we conducted on the techniques and tools currently used for the visualization of environmental data. It overviews several of the more popular visualization tools and briefly assesses their capabilities to support research that involves large datasets of environmental data. A taxonomy of visualization techniques based on the number of variables that can be visualized is also presented. New trends in building related user interfaces and a variety of new visualization techniques with their applicability are also discussed. Finally, several directions of future research in data visualization are outlined.**

Keywords: Data visualization, human-computer interaction, graphical user interfaces; visualization tools; climate change.

I. INTRODUCTION

Nowadays, studies on climate change are gaining significance all around the world. An ever-increasing vast amount of environmental data is measured, stored, and analyzed in order to find the impact of climate change on people and the environment. Many organizations provide environmental data online to help researchers working on climate change all over the globe. Some of such organizations are the National Oceanic and Atmospheric Administration (NOAA) [1], the Western Regional Climate Center (WRCC) [2], Cal-Adapt [3] and Nevada Climate Change Portal (NCCP) [4]. Although most of the climate related data is easily available on the World Wide Web, it is a complex and demanding task to analyze very large datasets without the help of visualization.

As the graphical presentations of data communicate ideas more clearly, the need for visualization is growing in different domains such as business, engineering, education, physical sciences, social sciences, medical sciences, meteorology, hydrology, finances, and genetics. Visualization helps in finding more efficiently the hidden patterns of the raw data. It helps the researchers to focus more on the problem and achieve their goals more quickly. Currently, many industry-standard visualization tools are available, but each of them is developed for a specific purpose. It is a challenge for a researcher to pick a tool from the huge collection of tools available. The survey presented in this paper was conducted while choosing visualization tools for developing a visualization system for the NCCP.

NCCP is a cyber-infrastructure hub and set of valuable resources climate change studies performed in Nevada. An ongoing NSF EPSCoR funded project [4] involves interdisciplinary collaborative work by research teams from UNR, UNLV, DRI, and other Nevada System of Higher Education (NSHE) institutions. The goal of this portal is to efficiently provide the researches, general public, stakeholders, and policy makers with climate change data, and graphically present this data using visualization techniques. One of the tasks before developing the software for the visualization component of the NCCP was to choose a set of visualization tools. The need for selecting the most suitable tools from several existing visualization tools has led to the present survey.

Although there are several taxonomies proposed for visualization tools and techniques, it takes considerable amount of time to choose the right tool that meets the user requirements. The goal of this survey paper is to help the researchers in the field of environmental data to quickly find a visualization tool based on their requirements. Researchers may look at the tools' features, applicability, strengths and limitations while choosing the right tool for their tasks. The proposed taxonomy of visualization techniques and tools makes it is easier for the researchers to find a desired tool.

The rest of the paper is organized as follows: Section II discusses the background and several existing surveys of visualization tools; Section III presents our survey of visualization tools, characterized in terms of capabilities, strengths, and limitations; Section IV proposes a brief taxonomy of visualization techniques; Section V discusses the state of the art in data visualization; finally, Section VI examines several directions of future research in data visualization.

II. BACKGROUND AND RELATED WORK

To the best of our knowledge, there are no comprehensive surveys performed on visualization tools for environmental data. However, a survey similar to ours was conducted for visualization tools of biological network analysis [5]. Visualization tools such as Medusa, Cytoscape, BioLayout Express3D, ProViz, and Pajek were reviewed in this survey. These tools were assessed based on their visualization capabilities, efficiency, compatibility with other tools, supported input data formats, user-friendliness, and applicability. The authors summarize the survey by suggesting tools for each outlined challenge such as integration of heterogeneous data, scalability, inter-connectivity, and pattern recognition. Also, several of the more specialized tools that are used for visualizing the biological networks are mentioned, but are not reviewed in the survey. Furthermore, standard network file formats used for representing biological networks such as BioPAX, SBML, CellML and RDF are discussed and evaluated. The paper also discusses several challenges faced by visualization tools representing biological networks. For example, one of such challenges is user-friendliness in the case when thousands of nodes need to be visualized (because most tools become less efficient and slower when operating with large datasets). The authors suggest outsourcing the computational load by using web services or rewriting the existing algorithms. Another challenge faced by the visualization tools is that of representing heterogeneous data.

A comparison of information visualization tools was conducted by Aigner *et al* [6]. Tools such as Tableau, Xmdv Tool, Spotfire, and ILOG Discovery were evaluated and compared, basic user tasks with one or two variables and advanced tasks dealing with complex problems were tested, the features and functions offered by the tools were categorized, and a matrix stating the features offered by each tool was presented. The suggested improvements to the surveyed visualization tools include incorporating standard input file formats and supporting large datasets.

Existing visualization tools for environmental data such as IBM Data Explorer, OceanShare, ImmerseDesk, CAVE, and Infinity Wall were briefly reviewed by Mozzafari and Seffah [7]. These tools were found to be useful for integrating heterogeneous data but less intuitive and unable to link multivariate data. The paper mainly presents a toolkit that offers interaction and visualization services. The services include gaining insight into singular data with the help of visualizations, using interaction techniques to link visualized data, and discovering hidden patterns and relationships in large datasets.

A survey of hundred and one visualization techniques that are used in general for representing time and time-oriented data was performed by Aigner *et al* [8]. The techniques were categorized based on data, time, mapping, and dimensionality. Based on this categorization a table showing the features supported by technique was provided. From the survey, it was found that none of the visualizations can handle data that is based on multiple perspectives, and has cyclic time evolution. Also, most of techniques are not reusable for another visualization problem even if the problem is similar to the original one. Thus, visualization techniques need to have broader applicability.

From the past research it is apparent that there are many challenges facing the visualization tools and techniques. In the following section we review several existing visualization tools.

III. A SURVEY OF VISUALIZATION TOOLS

Some of the tools that are currently used for visualizing environment data were selected for our review. Usually, these tools are used for specialized tasks. For each tool surveyed we provide a concise summary of its capabilities, applicability, strengths, and limitations.

A. ArcGIS

ArcGIS is a software package that allows users to quickly create maps by using geographic knowledge [9]. It is a proprietary software offered by the Environmental Systems Research Institute (ESRI), the world leader in Geographic Information Systems (GIS).

Features

ArcGIS is available in many versions. End users can access the maps, data, and applications from their desktop computers, mobile phones, tablets, and the web. Web developers can build applications on multiple platforms by using ArcGIS API for JavaScript, Flex, Silverlight and SharePoint. The tool's web mapping feature allows the sharing of data, maps, and applications from other users. The desktop version of ArcGIS has predefined templates that automate the creation of maps, which can be shared without any limitation. The ArcGIS runtime SDK is available for Windows mobile, smartphone and tablet devices. The ArcGIS's online facility offers to the users a ready to use software environment with rich content.

Strengths

- ArcGIS is easy to use.
- High quality maps can be created quickly by using automation for desktop users and no installation is needed for ArcGIS online users.
- Several input data formats are supported.

Drawbacks

- ArcGIS is expensive and there are additional costs for upgrades.
- Clip geoprocessing tool is slow and often produces inaccurate results [10].

B. AVS/Express

AVS/Express supports object oriented development and is mainly used for visualization purposes by programmers and non-programmers [11].

Features

AVS/Express is available for several platforms and supports several visualization techniques for 2D and 3D environments. Moreover, visualization of multi-gigabyte datasets is possible. The cross-platform GUI system and cross-platform compatibility of the tool allows rapid application development. AVS/Express has a library with 900 modules for performing various visualization and data management tasks. The parallel processing capabilities of the tool allow distributed computing across shared memory, which results in very fast processing. The tool's multi-channel output makes possible to display visualizations over multiple screens.

Strengths

- Easily integrates modules from programming languages such as C, C++, and FORTRAN.
- Scales to very complex and large datasets.

Drawbacks

- It depends on virtual memory for sending results to users, which at times may result in low performance.
- As it supports multiple platforms, the number of available standard widgets and controls is relatively limited.

C. GrADS

Grid Analysis and Display System (GrADS) is a visualization tool used for data manipulation and visualization of Earth science data in a 5-dimensional space [12].

Features

GrADS uses FORTRAN-like command-line expressions to execute the operations and provides rich built-in functions. Furthermore, the users can add external functions written in any programming language and output graphics can be saved as images or PostScript.

Strengths

- This toolset is open source.
- It supports several input data formats.

Drawbacks

- Only a desktop version is available.
- The users need to learn commands to draw and control the graphics, which could make the tool less intuitive.

D. Integrated Data Viewer (IDV)

IDV is a Java-based software framework for analyzing and visualizing geosciences data [13].

Features

IDV is available on multiple platforms. Java and Java3D are needed to work with the IDV framework. Its user interface can be customized to perform specific tasks apart from geoscience applications. IDV provides labels for longitude and latitude axes and it can display different data types at the same time.

Strengths

- It is free.
- Provides high quality 3D visualizations.
- It can plot data from remote servers.
- Supports several data types.

Drawbacks

- It requires a lot of RAM, which can make it slow for large databases.

E. UV-CDAT

Ultrascale Visualization Climate Data Analysis Tools is a set of tools that support data analysis and visualization for large climate datasets [14].

Features

This is a python-based software toolset that can be integrated with other tools such as CDAT, VisTrails, and Paraview to allow researchers to design advanced scientific visualizations. Many climate-related data functions such as mean, standard deviation, and linear regression are provided by this tool that supports multiple platforms. It also supports provenance functionality and allows data analysis via VisTrails.

Strengths

- Provides a suitable, useful suite of tools for climate related data visualizations.
- It is open source.

Drawbacks

- Supports a limited number of operating systems.
- As it is a new product, it may still have bugs and need more testing.

F. VisTrails

VisTrails is an open source system that provides support for scientific data workflow and visualization [15].

Features

The provenance feature of VisTrails allows the comparison of different workflows visually. The results of the workflows can be displayed on a broad range of displays, from large display walls to small iPod screens. VisTrails is a python-based tool to which existing code or new modules can be added. Specialized plugins allow for other tools such as VisIt and ParaView to be integrated. Among its educational purposes, VisTrails is used for teaching Scientific Visualization and Digital Media.

Strengths

- Has a simple and easy to use user interface.
- Benefits from a broad user community.
- Contains good comparative visualization capabilities.

Drawbacks

- Sometimes it may hang up while updating large amounts of data from a remote site.
- Has limited parallel computing capabilities.

G. VisIt

VisIt is an open source tool that provides capabilities for visualization of complex scientific data [16].

Features

VisIt supports C++, Python and Java interfaces and can operate on multiple platforms. Terascale data can be visualized through parallel and distributed architecture and rich 2D and 3D visualizations are possible. Extensibility is achieved by dynamically allowing plugins.

Strengths

- Provides a good framework for customization.
- Has capabilities for interactive parallel visualizations.
- Accepts input data of different formats.
- Supports terascale data sets.

Drawbacks

- Data movement could be challenging in future machines.

H. Visualization Toolkit (VTK)

VTK is an open source software that supports an object-oriented environment and consist of libraries written in C++. It is mainly used for image processing and visualization [17].

Features

VTK provides surface and volume rendering support. It enables the visualization of large complex datasets through parallel processing. Also, it supports multiple platforms and includes an extensive set of 3D widgets. Most of the wrapper code for binding Tcl, Java and Python is automated.

Strengths

- Manages and represent complex scientific data.
- Supports many visualization techniques.
- Has a large user base.

Drawbacks

- Contains a limited number of modeling tools.

I. Other data visualization tools

Several other tools from different application domains such as Ferret, Ggobi, Mapobjects, Mathematica, Matlab, OpenDX, Prefuse, R, S-PLUS, SPSS, and Tableau were also reviewed. Details of these tools are available in a comparison table that is available upon request from the authors. For each tool, the table contains information about the operating systems supported, the visualization techniques employed, the programming and scripting languages used, and the number of variables used in representing data (e.g., 2D, 3D, multi-dimensional).

IV. A SIMPLE TAXONOMY OF VISUALIZATION TOOLS

The existing visualization methods for environmental data can be classified based on several factors. Many researchers have introduced a taxonomy for visualization techniques [18]. For example, Shneiderman classified visualization techniques based on data types and user tasks [19]. Specific data types include *1D, 2D, 3D, multidimensional, temporal, tree,* and *network*. The user tasks considered by Schneiderman to classify the methods were *overview, zoom, filter, details-on-demand, relate, history,* and *extract*. Keim classified the methods based on data types and interaction/distortion techniques [20]. His data types were similar to those used by Schneiderman, except for *algorithms/software*. Interaction methods such as *standard, projection, filtering, zoom, distortion,* and *link brush* were considered for classification. Silva *et al* [21] classified the methods based on visualization and interaction features, Muller *et* al [22] structured them as static and dynamic, Chi [23] classified them based on visualization processes and Tory [24] grouped them based on the characteristics of the models of data employed. While these classifications are more generic, in this paper we propose a taxonomy of visualization methods based on the data types used to represent environmental data.

Specifically, the usual types of environmental data are: *one-dimensional,* such as atmospheric pressure and wind velocity; *two-dimensional,* which result from a combination of two variables, for example temperature and humidity; *three-dimensional,* which involve a combination of three variables; *multi-dimensional,* which are a combination of more than three variables; and finally, *climate-related text data* that can be found in the documents or news.

A. One-Dimensional

For a one dimensional data set, the data values correspond to one variable and there is only a value per each data item. Some of the data visualizations of one-dimensional data are histograms as and normal distributions as shown in [25].

B. Two-Dimensional

Two-Dimensional data corresponds to two variables. The relationship between two variables can be easily found through visualization. The 2D visualizations of climate data are line graphs [26], comparison of variables using plotting [27], bar charts [28], area charts, pie charts, maps, scatterplots [29] , and stream line and arrow visualizations [30].

C. Three-Dimensional

Data values in three-dimensional space have three attributes. The graphical representation of the three attributes shows the depth and rotation in addition to the two dimensional data. The methods for representing three-dimensional data are Isosurface techniques such as presented in [31] and [32], direct volume rendering [33], slicing techniques [34], 3D bar charts and realistic renderings [35].

D. Multi-Dimensional

Data attributes in Multi-Dimensional space ranges from four to hundreds. To understand the relations between multiple variables several techniques are available. The methods for the visualization of multivariate data are based on using scatterplot matrices [36], parallel coordinates [37], star coordinates [38], maps [39], and autoglyphs [40].

V. Discussion

After surveying about 20 visualization tools, it is apparent that none of the tools fulfill the needs of the users completely. Users need to switch among the tools to get the desired results for the climate researchers. Even though most of the data (especially in the field of meteorology, environmental sciences and climatology) is easily and openly accessible over the Internet for the researchers [41, 42, 43], handling large streams of data is complex when the data is in different formats. Also, only few visualization tools have their source code available, enabling the researchers in the field of visualization to implement their ideas quickly by not having to start everything from scratch.

With the availability of high quality computer display devices, visualization is currently mainly focused on 3D/4D techniques. The benefits of these techniques, which replace the traditional 1D/2D techniques that have been in use for many years, need to be investigated further. Furthermore, the user interaction with the visualizations is becoming increasingly more complex. For example, users are enabled to discover many details available in the visualization by rolling the mouse over the visualization, or are able to change the visualization by using functions such as zoom in, zoom out, moving left/right/up/down or maps by selecting several views –Google, for instance, provides all these features in its visualizations [44].

VI. Future Work and Conclusion

Currently, the challenges facing the designers of the data visualization include creating applications that : run on a broad range devices, including desktop computers, mobile phones, display walls, and touch pads; support diverse operating systems such as Windows, Mac, and UNIX; provide several visualization options and capabilities so that the users do not need additional tools; support various interaction and visualization techniques; provide high quality graphics with no loss of useful information; support a large variety of input data formats; enable processing large datasets without performance losses; and allow easy integration with other tools.

As the developed visualizations play a crucial role in decision making it is also important to check if there are any missing data values from defective equipment, and if the received data is accurate. Also, a standard input format would be useful for the visualization tools. In the future, we aim to develop a comprehensive visualization system for the NCCP that includes the above capabilities.

Acknowledgments

This work was made possible through the support provided by the National Science Foundation under Cooperative Agreements No. EPS-0814372 and No. EPS-0919123.

References

[1] "Graphical Forecasts," Nation Oceanic and Atmospheric Administration, Accessed December 7, 2012, available at: http://graphical.weather.gov/.

[2] "UNR Valley Road Weather Station," Western Regional Climate Center, Accessed December 7, 2012, available at: http://www.wrcc.dri.edu/weather/unr.html.

[3] "Snow Pack: Decadal Averages Map," Cal-adapt Exploring California's Climate Change Research, Accessed December 7, 2012, available at: http://caladapt.org/snowpack/decadal/.

[4] Nevada Climate Change Portal, Accessed December 7, 2012, available at <http://www.sensor.nevada.edu/NCCP/>.

[5] Pavlopoulos G. A., Wegener A., and Schneider R., "A survey of visualization tools for biological network analysis", BioData Mining, November 2008.

[6] Aigner W., Bertone A., and Miksch S., "Comparing Information Visualization Tools Focusing on the Temporal Dimensions," 12th International Conference on Information Visualisation, pp. 69 - 74, July 2008.

[7] Mozzafari E. and Seffah A., "From Visualization to Visual Mining: Application to Environmental Data", IEEE Confererence on Advances in Computer-Human Interaction, pp. 143-148, February 2008.

[8] Aigner W., Miksch S., Schumann H., and Tominski C., Visualization of Time-Oriented Data, Springer, May 2011.

[9] "ArcGIS - Mapping and Spatial Analysis for Understanding Our World", ESRI, Accessed December 3 2012, available at: <http://www.esri.com/software/arcgis>.

[10] "ArcGIS QGIS Faceoff", blog.donmeltz.com, Accessed: December 3 2012, available at: <http://donmeltz.com/ blog/index.php/2011/06/10/arcgisqgis-faceoff/>.

[11] "AVS/Express Data Visualization Software", AVS/Express, accessed December 4 2012, <http://www.avs.com/products/avs-express/index.html>.

[12] "GrADS Home Page", Grid Analysis and Display System, Accessed December 4 2012, available at <http:// www.iges.org/grads/grads.html>.

[13] "Unidata | IDV", Unidata, Accessed December 5 2012, available at <http://www.unidata.ucar.edu/software/idv/>.

[14] "UV-CDAT", UV-CDAT, Accessed December 5 2012, accessed <http://uv-cdat.llnl.gov/>.

[15] "VisTrailsWiki", VisTrailsWiki, Accessed December 3 2012, available at: <http://www.vistrails.org/index.php/Main_Page# VisTrails_Overview>.

[16] "VisIt Visualization Tool", visIt, Accessed December 7 2012, available at <https://wci.llnl.gov/codes/visit/home.html>.

[17] "VTK - The Visualization Toolkit", Visualization Toolkit, Accessed December 7 2012, available at <http://www.vtk.org/

[18] Wenzel S. B. J. and Jessen U., "A taxonomy of visualization techniques for simulation in production and logistics," Proceedings of the 2003 Winter Simulation Conference, pp. 729 - 736, December 2003.

[19] Shneiderman B., "The eyes have it: a task by data type taxonomy for information visualizations," Proceedings of IEEE Symposium on Visual Languages, pp. 336-343, 1996.

[20] Keim D.A., "Information visualization and visual data mining," IEEE Transactions on Visualization and Computer Graphics, pp. 1-8, January/March 2002.

[21] Silva S.F., Catarci T., “Visualization of Linear Time-Oriented Data: a Survey,” First International Conference on Web Information Systems Engineering (WISE), pp. 310-319, Hong Kong, Chine, IEEE Computer Society, 2000.

[22] Muller W., Schumann H., “Visualization Methods for Time-Dependent Data - An Overview,” Proceedings of the Winter Simulation Conference WSC, pp. 737-745, New Orleans, USA, ACM Press, 2003.

[23] Chi E.H., “A Taxonomy of Visualization Techniques Using the Data State Reference Model,” Proceedings of the IEEE Symposium on Information Visualization InfoVis,” pp. 69-76, Salt Lake City, USA, IEE E Computer Society, 2000.

[24] Tory, M., “Rethinking Visualization: A High-Level Taxonomy, IEEE Symposium on Information Visualization,” pp. 151-158, 2004

[25] Edward Linacre, Climate Data and Resources: A Reference and Guide, New York: Routledge, 1992.

[26] Stolte C., Tang D., Hanrahan P., "Polaris: a system for query, analysis, and visualization of multidimensional relational databases," IEEE Transactions on Visualization and Computer Graphics, pp. 52-65, Jan/Mar 2002.

[27] Daniel K. R., Dmitry B., Joseph V., David J. S., Donald S., "A simple 1-dimensional, climate based dissolved oxygen model for the central basin of Lake Erie," Journal of Great Lakes Research, pp. 465-476, March 2010.

[28] Matthew W., Georges G., Daniel K., Interactive data visualization: foundations, techniques, and applications, Mass.: A K Peters, 2010.

[29] “Climate Modeling”, OSU Website, Accessed June 27, 2012, <http://mgg.coas.oregonstate.edu/~andreas/OC599/climate_modeling_11/Script.pdf>.

[30] Nocke T., Flechsig M., Bohm U., "Visual Exploration and Evaluation of climate-related simulation data," Simulation Conference, pp. 703-711, December 2007.

[31] Lloyd A. T., “Case study: severe rainfall events in northwestern Peru (visualization of scattered meteorological data),” Proceedings of the Conference on Visualization '94, pp. 350-354, October 1994.

[32] Riley K., Ebert D., Hansen C., Levit J., "Visually accurate multi-field weather visualization," IEEE Visualization, pp. 279-286, October 2003.

[33] Bajaj C., Yagel R., “Efficient Techniques for Volume Rendering of Scalar Fields,” 1998.

[34] Chaoli W., Kwan-Liu M., Wittenberg A.T., "Correlation study of time-varying multivariate climate data sets," IEEE Pacific Visualization Symposium, pp. 161-168, April 2009.

[35] Baker M.P., "After the storm: considerations for information visualization," IEEE Conference on Computer Graphics and Applications, pp. 12-15, May 1995.

[36] Richard A.B., William S.C., “Brushing scatterplots,” Techno-metrics, pp. 127–142, 1987.

[37] Robert M. E., "The parallel coordinate plot in action: design and use for geographic visualization," Computational Statistics & Data Analysis, pp. 605–659, November 2002.

[38] Kandogan E., “Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions,” Proceedings of IEEE Information Visualization, Hot Topics, pp. 4-8, 2000.

[39] Bottinger M, Scheuermann G., "Brushing of Attribute Clouds for the Visualization of Multivariate Data," IEEE Transactions on Visualization and Computer Graphics, pp. 1459-1466, December 2008.

[40] Wong P.C., Bergeron R.D., “30 Years of Multidimensional Multivariate Visualization,” Scientific Visualization-Overviews, Methodologies, and Techniques, IEEE Computer Society Press, pp. 3-33, 1997..

[41] “S.E.N.S.O.R”, NCCP Website, Accessed July 19 2012, available at <http://sensor.nevada.edu/NCCP/DataSearch/ Silverlight Data/Client.aspx>.

[42] “Environmental Reports| EPA Response to BP Spill in the Gulf of Mexico | US EPA”, EPA website, Accessed July 19 2012, available at <http://www.epa.gov/bpspill/download.html>.

[43] “Environment | Data”, The World Bank Website, Accessed July 19 2012, <http://data.worldbank.org/topic/environment>.

[44] “Google Maps,” Google Website, Accessed July 19 2012, available at < https://maps.google.com/maps>.