GEMS202 - 0102

SL: Jill Smith

Citation style: AMS

DIVA - Data Imaging and Visualization Analysis

Methodology

We pledge that we have not given or received any unauthorized assistance on this assignment.

Table of Contents

Introduction

Current climate data visualization tools are non-interactive; they typically display only a few variables, cannot render real-time results (with the exception of World Wind), and do not allow users to see correlations or other interesting trends in the data (Alder et al. 2013; Du et al. 2015; Liu et al. 2015; Potter et al. 2009; Teuling et al. 2011; Wickham et al. 2012; Zhang et al. 2016). To address these issues, we will first develop a virtual reality (VR) tool to visualize climate data using techniques derived from previous research. Once we complete a prototype of our tool, we will collect some initial data to guide our design and inform us of areas that could be improved. Finally, once the final product is complete, we will collect data on the performance of our tool. Thus, our research can be broken up into three main phases: product development, product improvement, and product evaluation.

Product development entails constructing a Graphical User Interface (GUI), integrating VR to visualize climate data, and comparing methods of visualization. Therefore, we have divided the team into three subteams: a front-end design team, a back-end design team, and a data analysis and visualization team. The front-end design team determines how a visualization appears in the user's Heads-Up Display (HUD) and develops a control method for the user to interact with the tool. The back-end team specializes in implementing climate data and visualization methods into the VR framework. Finally, the data analysis team determines the optimal methods to visualize climate data. During the product development phase, we will focus on obtaining numerical data on computation time, feature effectiveness, and storage space, which will quantify the effectiveness of our tool.

After completing the product development stage, we will proceed to the product improvement stage. During this phase we will hold focus groups, first consisting primarily of graphics or visualization experts, then with the general public, and finally with experts in the climate field. The purpose of these focus groups is to receive constructive feedback on our prototypes as we develop our product through an iterative process. Once we develop our final product, we will administer individual surveys to both the general public and the climate experts in order to evaluate our visualization tool. Subteams for this phase will remain virtually the same, but the data analysis team will have added responsibilities to process and analyze the survey data. These data will answer how well the tool is received, as we will measure how informative, comprehensive, and user-friendly our product is. Using our proposed methodology, we will answer the following research questions:

1. In terms of computation time, feature selection, and storage, how can we most effectively design and create a Virtual Reality climate data visualization tool?
2. What are the most user-friendly, aesthetically pleasing and informative ways for scientists and the general public to visualize climate data through VR?
3. How can our revolutionary tool help scientists gain insights they would not have been able to gain otherwise?

Section 1 - Product Development Plan

Current methods of climate data visualization, such as 2D maps and 3D globes, are generally restricted to univariate data sets, and options to view multivariate data sets are quite limited (Liu et al. 2015; Teuling et al. 2011). This causes difficulties for climate scientists to identify areas of interest where there may be meaningful correlations that are hidden between multiple variables. The older methods of data visualization are also typically not interactive via an intuitive interface (Alder et al. 2013; Liu et al. 2015; Potter et al. 2009; Wickham et al. 2012; Zhang et al., 2015). For two-dimensional maps, users cannot zoom in, change their location, or adjust elevation on the map to focus on different parts of the data without coding significant control scripts. Due to the lack of interactivity and sometimes complicated nature of current maps and globes, years of experience are often required for users to use these tools proficiently. We plan to address some of these shortcomings by integrating NASA's open source World Wind Globe API (Application Programming Interface) with Oculus Rift VR technology so that technical users can observe multiple climate data sets simultaneously, focus in on areas of interest, and exercise greater control over the display of their data.

Section 1.1 - World Wind Globe API

Currently, the most advanced climate data visualization methods involve 3-D rendering displayed on virtual globes. The most prominent of this type of visualization is NASA's open source World Wind Globe API (Liu et al. 2015; Zhang et al., 2015). Researchers from the Chinese Academy of Sciences and Wuhan University have used World Wind to display vectors representing cyclones and dust storms, respectively (Liu et al. 2015; Zhang et al., 2015). However, these tools are neither versatile nor interactive. Our goal is to generalize World Wind's applications to encompass multiple kinds of data so that climate scientists can observe correlations among multiple variables and domains. For example, we would like to display not only wind direction but also other variables, such as temperature, atmospheric humidity, and oceanic salinity. Although World Wind has only been used to display a few climate patterns, it is well suited to multivariate visualization as it can display both vectors and volume rendering (Zhang et al., 2015). Since the World Wind API is written in Java, it supports any rendering from the Java Open Graphics Library (JOGL), which is Java's versatile standard 3-D graphics library (Zhang et al., 2015). With this framework, we could display, for instance, land and ocean temperature with a color gradient, wind speed and direction with vectors, and precipitation with height fields. Climatologists will then be able to observe these values change over time, thereby determining if there are any consistent relationships between the variables or any anomalies which may indicate a significant regional change in climate. We can then generalize this visualization process to any variable of interest and add customizability for color and texture, giving researchers the opportunity to simultaneously view as many data sets as needed and choose how they would like to view them.

Section 1.2 - Integration of Oculus Rift

As literature has indicated, existing forms of climate data visualization are not interactive, which often limits climate scientists' understanding of the data (Alder et al. 2013; Liu et al. 2015; Potter et al. 2009; Wickham et al. 2012; Zhang et al., 2015). For instance, the temperature variation with altitude may affect climate patterns, but this is essentially impossible to visualize on a two-dimensional map. Another example of how an interactive display may be helpful is if a researcher observes a localized trend in the data and would like to zoom in on that particular region for further investigation.

The implementation of a climate data visualization system within a VR framework will allow for greater interactivity between the user and the data. This provides users with the control to choose which parts of data sets to focus on in order to observe important relationships.

Specifically, we will use the Oculus Rift to develop our visualization system. The Rift headset has an accelerometer, gyroscope, and magnetometer, from which head orientation (yaw, pitch, roll) can be inferred. In addition, the Oculus Rift comes with an infrared camera, which tracks the position of an array of infrared micro-LEDs on the headset, allowing developers to track head position (x, y, z) of the user (Desai et al. 2014). The Oculus Rift software development kit (SDK) includes a head model code, which we will use to access the position and orientation measurements (Desai et al. 2014). We will take advantage of head orientation tracking to allow the data display to rotate with the user's head, creating a truly immersive experience. Furthermore, we will utilize head position tracking to allow users to zoom in on and zoom out of fields of interest by leaning in and leaning out, making our product even more user friendly and intuitive. To reduce latency, the time between head movements and updated display, we will use the predictive tracking code included in the Oculus SDK. Oculus will soon release the Oculus Touch, which are two controllers, one for each hand, that allow users to interact with VR. Oculus Rift can track wrist movement using these controllers in addition to tracking conventional button pressing (Oculus n.d.). Using the Oculus Touch, the user will be able to interface with select menus and other aspects of our GUI. Overall, we will take advantage of these systems to allow users to control their movements throughout the World Wind environment, observe changes in multiple sets of data, and focus in on areas where correlations exist.

Another benefit of using Oculus Rift is that it supports web-based applications through a Javascript framework called ReactVR (Developer, 2016). Hosting the application remotely and serving users through a web interface has two main benefits: a wider audience and a consistent architecture. Not everyone has an Oculus-ready PC since running Oculus Rift requires certain minimum specifications for the processor, graphics card, and operating system. By developing a web-based application, we will allow users without the required hardware to use our visualization tool, eliminating one boundary between product and user base. This format will also allow for taxing calculations and rendering tasks to be performed on a remote server with a known computing power, so that the speed and efficiency of the application is consistent from user to user. The Rift's ability to interface with the web allows us to ensure that our VR application will not need to rely on any users' hardware and can deliver consistent, reliable results (Developer, 2016).

Section 1.3 - Data Processing and Computational Efficiency

Oculus Rift's web compatibility allows us to outsource large computations and graphical visualizations to the cloud, reducing the load on client machines. Climate data are typically stored in large files on the order of gigabytes as either Network Common Data Form (NetCDF) or Gridded Binary (GRIB) files. NetCDF is a self-describing data format, which means the file includes a description of how the data is formatted throughout the file. This description allows NetCDF parsers to access meaningful values from millions of lines of text. GRIB is an older, less flexible type of climate data file which simply attributes values to a series of grid squares, and can be converted to NetCDF. However, due to the sheer magnitude of global data, these files can take substantial amounts of time to process on an ordinary desktop computer (Zhang et al. 2016). This is why we plan to use Oculus's cloud connectivity to outsource computations to Graphical Processing Units (GPUs). We will then utilize Oculus Rift's ReactVR interface to send the rendered frames directly to the Oculus Rift display. In this way, the client machine hosting the Oculus Rift hardware will only need to display the visualization and not process huge sets of data on a local machine (Fig. 1). Therefore, we will be able to display very smooth renderings of our visualization for the user.

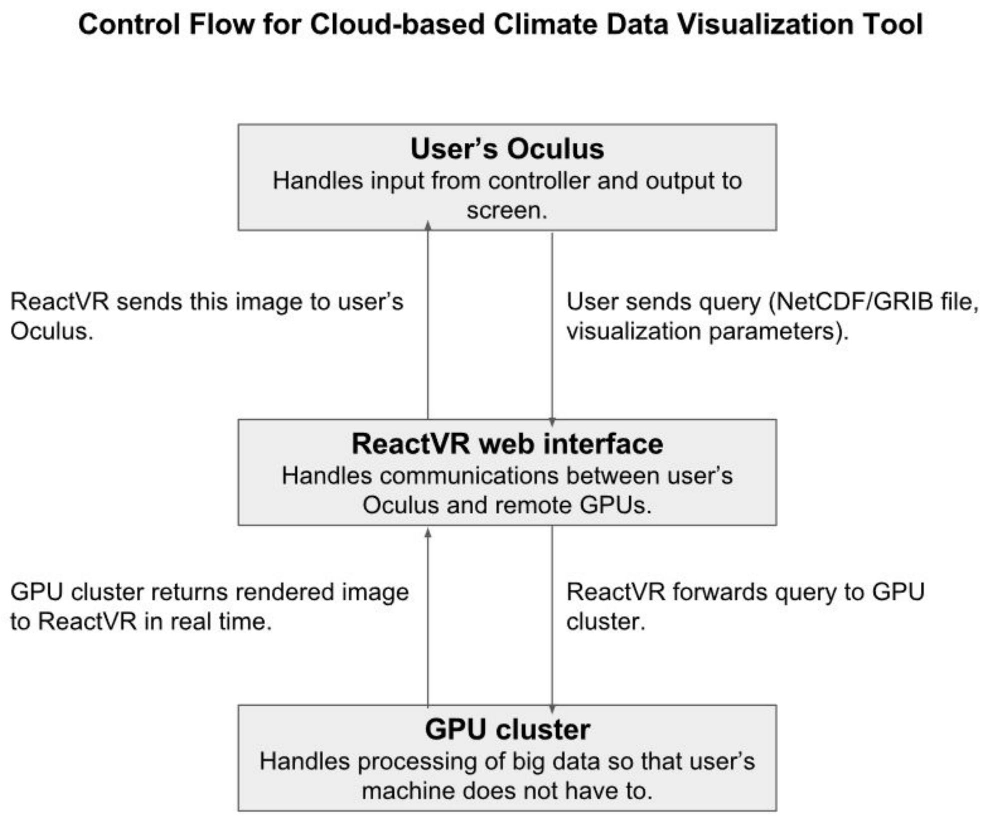

Figure 1: Control flow diagram of our data visualization tool.

Section 1.4 - Application of Product: El Niño

In order to gather survey and focus group information regarding our project, we plan to display and analyze a climate phenomenon to showcase the potential of our tool's ability to render multivariate interactions in full 3D over a time domain. We believe that El Niño is an ideal choice to display the potential of our visualization system. El Niño is a climate pattern during which warmer ocean water from the western Pacific flows eastward, affecting global weather patterns by increasing rainfall in the eastern Pacific and causing drought in countries bordering the western Pacific (NOAA n.d.). This is a great example of ocean conditions influencing atmospheric conditions, and vice versa, which will allow our visualization to demonstrate its ability to highlight the correlations among these data sets.

Since El Niño is a such a complex phenomenon, researchers in the fields of both visualization and analysis have used this phenomenon to evaluate the success of their software. Marwan and Kurths (2002) modified a method of analyzing nonlinear data through cross recurrence plots and chose to apply this technique to El Niño data in order to test their ability to identify relations in multivariate data. By comparing the results of their new method and the traditional linear method, they deemed their method a success because it could better identify a trend of increased local rainfall in Argentina than the traditional model (Marwan and Kurths 2002). In another study, McCormick et al. (2004) chose El Niño as a case study to test the effectiveness of multivariate visualization rendering on GPUs as opposed to CPUs (Central Processing Units). We believe that displaying a visualization of El Niño to focus group and survey participants will adequately exhibit the capabilities of our software and allow the participants to evaluate our VR tool against traditional visualization methods.

Section 2 - Data Collection

In the following section, we describe how we will collect data and what types of methods we will use to analyze the data. We take into account the limitations of our methods as well as other factors that may affect our data collection. We will have three focus groups during our development phase: graphics experts, general public, and climate researchers. Once we complete our development phase, we will conduct surveys in order to evaluate our tool. All of these focus groups and surveys will require IRB approval.

Section 2.1 - Focus Groups

Our first focus group will consist of about five graphics experts from the University of Maryland and will last about an hour. The purpose of this focus group is to refine the aesthetics and user interface of our visualization. This will be a facilitated discussion guided by questions concerning our current state of visualization techniques and headed by two teammates. These questions will not directly concern climate data, but instead will concern the aesthetics of the visualization. We will give a demographic survey to each of the graphics experts before the discussion. We will inform them that the discussion will be recorded for research purposes. Based on the feedback from these experts, we will make improvements to the graphics and user interface.

Next, our second focus group will consist of members of the general public. This focus group will have the same structure as the group of graphics experts, except the questions will focus more on the general feel and usability of our tool. To obtain participants, we will advertise throughout the University of Maryland, targeting a diverse group of students not only majoring in STEM fields, but also in the arts. This is because we want to make our user control scheme as intuitive as possible. In addition, we do not want poor aesthetics to hamper our later focus group of climate researchers from giving feedback on our visualization techniques. To do this, we will advertise at the Computer Science Instructional Center, Tawes Hall, William E. Kirwan Hall, and the Clarice Smith Performing Arts Center. We will obtain approximately 30 students and plan to separate them into five subgroups of six students. We can then make further improvements to the usability and aesthetics of our VR tool based on the general comments of the public.

Lastly, we will hold a climate researcher focus group, which will consist of 10 climate data experts from NOAA, NASA, and/or within the University of Maryland. This focus group will have the same structure as our previous two focus groups. However, the main subject of this focus group will be what the experts can interpret and learn from our visualization such as making general comments about correlations and viewing specific data.

Section 2.2 - Individual Surveys

After the focus group data has been obtained and consequent improvements have been made to our VR tool, we will test our prototype against earlier visualization tools, such as older implementations using World Wind, by administering an extensive survey to the general public and experts.

The variables we will measure are usability, aesthetic appeal, comprehensibility, motion sickness caused by the visualization tool, and the time to complete given tasks. For this data, we will ask users to rate these variables on a scale from 1-10 and will time how long it will take to complete the tasks assigned. This data will help us focus on where our product shines and fails compared to current visualization tools. Usability can be measured by both the time spent and accuracy of the tasks as well as the participant's direct rating on a scale of 1-10. Aesthetic appeal is also an important data set to consider. Not only does it increase user interest and engagement for the data, but it also improves effectiveness of our tool (Lau and Moere 2007). The comprehensibility of the data is also important, especially for the public, because it defines how much information the data visualization is able to convey to the user (McCormick et al. 2004). Measuring motion sickness is important because ideally our visualization tool should not cause any motion sickness for the user.

To account for confounding variables, we will collect demographic information of our participants including age, ethnicity, gender, major, and socioeconomic status. For the general public, it will be important to measure the user's familiarity with both technology and climate data as these are potential confounding variables. For the climate experts, we will collect data on familiarity and experience with technology, place of work, position in the research community (such as seniority and authority level), and research interests (theoretical or applied). For both groups, we will run correlation analyses, such as multiple logistic regression, between test results and demographic information to determine if there is any relationship between them. This way, we can determine whether or not our results are statistically significant.

Due to convenience and proximity, we only plan to conduct our general public surveys on students from the University of Maryland. To gather participants, we will post advertisements and flyers across campus at the most popular locations such as McKeldin Library, Adele H. Stamp Student Union, the Computer Science Instructional Center, and Glenn L. Martin Hall. We will obtain approximately 50 participants for these surveys. Since we expect our participants will not have much prior experience dealing with climate data, we will set up a pre-made data visualization on both our platform and an existing visualization tool for all participants. We will allow the participants to get familiar with the visualizations for approximately five minutes. We will begin by asking the participants to fill out a survey in which they will rate both tools on the variables we mentioned previously on a scale of 1 to 10, with 1 being "Poor" and 10 being "Near perfect". Afterwards, we will set up questionnaires for each participant, asking them to analyze the data displayed on our visualization tool. For example, we will ask them to zoom in on a particular region and read some data points. The participants will be given a score based on the accuracy of their answers.

The surveys for the researchers will be held at a different time but will consist mostly of the same process. We will have a group of 10 experts to take our survey. For this group, we will provide a bare data set, and ask them to visualize it given a set of instructions instead of giving them a pre-made data visualization. Due to their extensive knowledge of climate data visualization, these experts will be able to give more insight into how easy it is to load data into our device and visualize it. We will measure how much time it takes for the experts to visualize the data on our tool as well as an older tool. For this older tool, the same survey will be given to the experts to complete.

Section 2.3 - Data Analysis

Our surveys will supply us with data of the participants' evaluations of the old visualization tool and our virtual reality tool. The two objective data sets we will analyze are the accuracy of their analysis and the time taken to complete the analysis. In addition, we will have multiple subjective data sets that gauge the participants' attitudes towards the features of the tool. We can tabulate the results by tool and factor. Our null hypothesis is that the scores and the evaluations of the features are not significantly different between the old tool and the VR tool. Our alternate hypothesis is the opposite - that there exists a significant difference between the scores of the old tool and our VR tool. We will use a One-Way Analysis of Variance test (ANOVA) to calculate how likely it is that the two sets of results occurred simply due to chance. We will run this test on each feature as well as the score variable to see how much our VR tool improved or did not improve. A P-value will generated from our statistical tests. This P-value states the probability that our results occurred due to chance. If this value is below a certain threshold such as 0.05, then we can reject the null hypothesis and state that there is a significant difference between the usage results of the old tool and those of the virtual reality tool. The analysis is similar for the set of experts, except that we will also analyze how much time it takes for the experts to visualize the data. It is important to note that we will analyze both groups separately. In addition, for the expert group, we will have a greater degree of uncertainty because of the small sample size.

Conclusion

Overall, our methodology will involve three main phases: product development, product improvement and product evaluation. Our product development phase will entail creating a working prototype that uses preliminary visualization techniques. We plan to utilize the World Wind API with Oculus Rift to produce our visualizations. In order to reach a wider audience, we will use the web-based ReactVR interface and outsource our computations to a cloud computing service so that any computer connected to the web will be able to use our tool. Once we create our prototype, we will transition to the product improvement phase. This will involve a series of focus groups: first, with graphics experts to improve the interface, then with the general public to assess usability, and finally with climate data experts to evaluate our final product. This process will be iterative, during which we will improve our tool based on each focus group's feedback. This ensures that our final product will be tailored to our target audience of climate data researchers. Finally, in our product evaluation phase, we will collect both quantitative and qualitative data to determine how users interpret our program's features and how users feel about existing visualization tools. We will use ANOVA to compare these results to determine whether a statistically significant difference exists between the usage of our tool and an existing tool. This way, we will determine whether integrating virtual reality with a globe environment is indeed superior to existing methods of climate data visualization.

Because we are focusing heavily on interactivity, are employing a multilayered iterative design, and are testing our product during each phase, we believe that our tool will demonstrate a statistically significant improvement when compared to existing tools. This is largely because existing methods of climate data visualization have not been developed based on feedback from experts and the public, but have instead been based on mere intuition and a general dissatisfaction with previous tools (Alder et al. 2013; Liu et al. 2015; Potter et al. 2009; Teuling et al. 2011; Wickham et al. 2012; Zhang et al. 2016). Furthermore, no older method has employed VR technology as a potential solution to these problems of interactivity (Alder et al. 2013; Liu et al. 2015; Potter et al. 2009; Teuling et al. 2011; Wickham et al. 2012; Zhang et al. 2016). By utilizing feedback from a series of focus groups to optimize our product, we therefore anticipate that VR technology can be used to improve climate data visualization methods.

Glossary

Application Programming Interface (API) - The set of pre-made programming libraries, documentation, and tools used by developers to write their code.

Central Processing Unit (CPU) - A hardware component in a computer which handles basic arithmetic, input/output, and control of other components.

General Regularly-distributed Information in Binary form (GRIB) - A self-describing data format which is often used for big weather data. The files contain a description of a grid space and the values of data points in each square of the grid space. GRIB files can be converted to NetCDF files.

Graphical User Interface (GUI) - What the user sees on the screen; an interactive visual display which the user can interact with to operate the software.

Graphics Processing Unit (GPU) - A hardware component in many computers which specifically handles generation of graphics.

Heads-up display (HUD) - A program which displays information on the screen without requiring the user to move his or her head.

Network Common Data Form (NetCDF) - A self-describing data format which is often used for big weather data. The files contain a description of the data, followed by the data itself, allowing for greater flexibility in the actual data format.

Java Open Graphics Library (JOGL) - An API which allows the user to interact with GPUs of computers to produce computer graphics. It is cross-platform, which means that it can be used on different operating systems, like Windows and macOS, and with different programming languages, like Java and C.

Three-dimensional (3D) data visualizations - Globe APIs which render data, usually represented as vectors or color gradients, in such a way that users are able to move throughout the 3D environment to see the data at different points.

Two-dimensional (2D) data visualizations - Two-dimensional maps which use colors and/or symbols distributed across a geographic area to indicate values of variables at different points in space.

Virtual Reality (VR) - Technology utilizing goggles equipped with LEDs and motion detection to make users feel as though they are in a different environment than their actual surroundings.

References

Alder, J., Hostetler, S., and Williams, D., 2013: An Interactive Web Application for Visualizing Climate Data. *Eos, Transactions American Geophysical Union*, **94,** 197–198, doi:10.1002/2013eo220001.

Desai, P. R., Desai, P. N., Ajmera, K. D., and Mehta, K., 2014: A Review Paper on Oculus Rift-A Virtual Reality Headset. *International Journal of Engineering Trends and Technology*, **13**, 175–179, doi:10.14445/22315381/ijett-v13p237.

Du, Z., Fang, L., Bai, Y., Zhang, F., and Liu, R., 2015: Spatio-temporal visualization of air–sea CO2 flux and carbon budget using volume rendering. *Computers & Geosciences*, **77**, 77–86.,doi:10.1016/j.cageo.2015.01.004.

Lau, A., and Moere, A. V., 2007: Towards a Model of Information Aesthetics in Information Visualization. *2007 11th International Conference Information Visualization (IV '07)*, doi:10.1109/iv.2007.114.

Liu, P., Gong, J., and Yu, M, 2015: Visualizing and analyzing dynamic meteorological data with virtual globes: A case study of tropical cyclones. *Environmental Modelling & Software*, **64**, 80–93, doi:10.1016/j.envsoft.2014.11.014.

Marwan, N., and Kurths, J., 2002: Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, **302**, 299–307, doi:10.1016/s0375-9601(02)01170-2.

Mccormick, P., Inman, J., Ahrens, J., Hansen, C., and Roth, G., 2004: Scout: a hardware-accelerated system for quantitatively driven visualization and analysis. *IEEE Visualization 2004*, doi:10.1109/visual.2004.95.

NOAA Pacific Marine Environmental Laboratory: What is El Niño?. Accessed 14 November

2016. [Available online at http://www.pmel.noaa.gov/elnino/what-is-el-nino.]

Oculus: Developer Center. Accessed 14 November 2016. [Available online at https://developer.oculus.com/.]

Potter, K., Wilson, A., Bremer, P.-T., Williams, D., Doutriaux, C., Pascucci, V., and Johhson, C.,2009: Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. *Journal of Physics: Conference Series*, **180**, 012089, doi:10.1088/1742-6596/180/1/012089.

Teuling, A. J., Stöckli, R., and Seneviratne, S. I., 2010: Bivariate colour maps for visualizing climate data. *International Journal of Climatology*, **31**, 1408–1412, doi:10.1002/joc.2153.

Wickham, H., Hofmann, H., Wickham, C., and Cook, D., 2012: Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics*, **23**, 382–393, doi:10.1002/env.2152.

Zhang, T., Li, J., Liu, Q., and Huang, Q., 2016: A cloud-enabled remote visualization tool for time-varying climate data analytics. *Environmental Modelling & Software*, **75**, 513–518, doi:10.1016/j.envsoft.2015.10.033.