From Visualization to Visual Analytics for Environmental Science

Taylor Mutch*
Nik Molnar
Conservation Bio. Inst.

Judith Cushing†
Evergreen State Col.

Mike Bailey, Denise Lach, Chad Zanocco
Oregon State University

Genevieve Orr
Willamette Univ.

Peter Drake
Lewis & Clark Col.

Abstract

It is widely accepted that environmental scientists need tools to visualize temporally and spatially complex landscapes on 3D topographies. The interdisciplinary VISTAS (**Vis**ualization of **T**errestrial and **A**quatic **S**ystems) project team (computer, environmental, and social scientists) develops software to rapidly and easily create specialized 3D environmental visualizations without requiring expertise in complex graphics software. VISTAS is currently used to visualize multivariate time series data overlaid on 3D terrain. Current research and development is extending VISTAS to include machine learning and additional analytics to enable scientists to interactively determine not only what is occurring, but why, thus moving from visualization to visual analytics, from qualitative to quantitative analysis.

Keywords: Geospatial visualization, integrating spatial and non-spatial visualization, visualization in environmental sciences, visual analytics.

Index Terms: J.2 [Physical Sciences and Engineering]; I.3.4 [Graphics Utilities], I.3.7 [Three-Dimensional Graphics and Realism]; I.6.4 [Model Validation and Analysis].

1 Introduction

Environmental scientists increasingly use large complex models to discover phenomena and inform decision-making, a critical undertaking in the context of climate change [5]. Prior work demonstrates the importance of 3D scientific visualization to create opportunities both for uncovering relationships within complex spatiotemporal data and for communicating findings to diverse stakeholders and decision-makers [3]. While commercial and open-source tools can produce 3D visualizations (e.g., Paraview), the prerequisite level of expertise required presents a high barrier to entry, and consequently prevents widespread adoption of such software among environmental scientists [1]. The VISTAS project aims to lower barriers associated with producing scientific visualizations, thus empowering scientists to better understand and communicate their research [2].

In this paper, we describe the current state of the VISTAS software and present VISTAS visualizations crafted by collaborators to explore qualitative relationships in their data. We then discuss the need to integrate visual analytics into VISTAS so quantitative relationships in spatiotemporal data can be discerned.

2 VISTAS Software and Applications

The VISTAS project is an interdisciplinary cohort of computer, environmental and social scientists co-producing software that generates visualizations for domain-specific problems. Our focus is the visualization of complex 3D topography over time from sensed and process-model datasets and has resulted in a flexible visualization tool. The VISTAS software is a desktop application that accepts geospatial data and renders complex, spatiotemporal datasets into interactive 3D scenes, allowing users to explore and focus on aspects of their data that are influenced by terrain. Associated social science research (described elsewhere) aims to determine which kinds of visualizations are most effective for which purposes [1, 3]. VISTAS emphasizes a modular, extensible architecture that enables scientists to create data and visualization plugins to ingest geospatial data and produce customized visualizations. VISTAS is open-source and freely available (https://github.com/VISTAS-IVES/pyvistas).

Collaborator Halama uses VISTAS to fine-tune *Penumbra*, his shade-irradiance analysis model [4], to integrate his own research with eco-hydrological models (Figure 1). VISTAS' viewport synchronization enables viewing simultaneous landscape changes across time so that the human recognition system can detect the changes on local or regional scales that are difficult to discern without synchronized 3D perspectives.

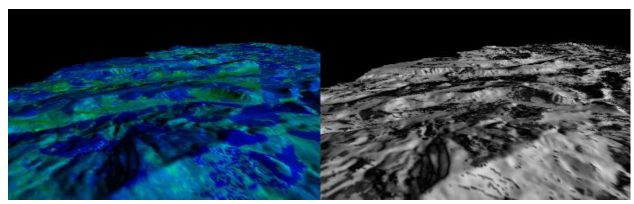

Figure 1: Carbon loss (left), increased irradiance (right).

Encoding multiple attributes into a single terrain visualization has helped scientific users assess relationships among spatial variables. In response to a collaborator's request, watershed flow and accumulation were draped over a rendered terrain as a vector field (Figure 2). While useful as a qualitative measure, the resulting visualization is not enough; the collaborator also wants help determining quantitative relationships within the visualization.

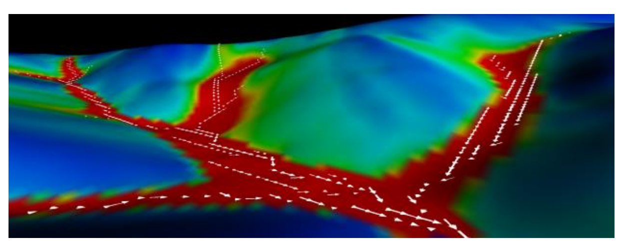

Figure 2: Water flow vectors throughout the Winant watershed.

3 Futures -- Integrating Visual Analytics

Environmental scientists report that VISTAS enables them to perceive qualitative relationships connected to terrain features. However, the need to determine quantitative relationships within spatiotemporal data will require that VISTAS be enhanced with

---

* taylor.mutch@consbio.org † judyc@evergreen.edu

analytical techniques to connect what is visually observed with what is hidden within. Our focus in extending VISTAS to include visual analytics is as before: allow environmental scientists to focus on their domain. Combining VISTAS current qualitative visualization with quantitative analyses as seamlessly as practical will allow our collaborators to better make and support scientific judgments, as Thomas and Cook [7] suggest as the primary goal for visual analytics.

Towards this effort, VISTAS architecture has been enhanced to interface with commonly used analytic libraries (Figure 3). Our goal is to map analyses of visualized data into VISTAS' visualizations, rendering visual markers (e.g., hot spots, heat maps, clusters) to indicate significant findings.

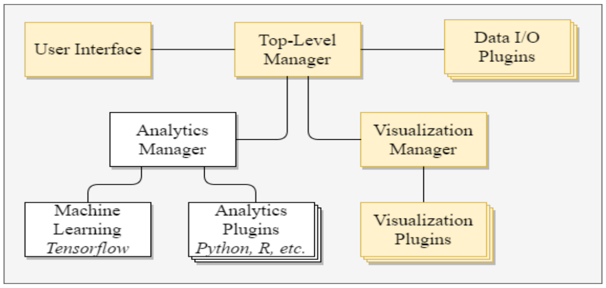

Figure 3: VISTAS architecture with Analytics Management interface. White boxes indicate future functionality.

To accomplish the visual analytics integration, developers have migrated VISTAS to the Python programming language. Many within the scientific community have encouraged adopting Python for data analysis and visualization. Community-backed numerical and statistical packages (e.g., NumPy, SciPy, TensorFlow) alleviate the need to self-develop libraries, allowing the team to focus on domain-specific requirements and rapidly incorporate new techniques as they become available.

3.1 Deciphering Images with Machine Learning

Manually deciphering and encoding images of the physical environment for research is time-consuming and error-prone. VISTAS team's ongoing experiments with deep convolutional neural networks (DCNNs) have had some success using machine learning classification for *cloud segmentation*, determining which pixels of a sky image are clouds (Figure 4). We plan to extend this work into the more challenging task of *cloud classification*, identifying type(s) of clouds within an image. Our eventual goal is to apply lessons learned in machine learning classification to other spatiotemporal data, such as the National Land Cover Database, to identify landscape types within image-based, remote sensing datasets.

Figure 4: A portion of a Total Sky Imager (TSI) image and segmentation into cloud and non-cloud pixels.

3.2 *Getting to Why* with Time Extrusions

Environmental scientists want to understand not only which landscape types are where, but where and why landscape change occurs. However, change over time is difficult to discern by looking at animations where maps are portrayed linearly over time. Schultz and Bailey have prototyped visualizations of spatiotemporal land use within a single image frame as a time-extrusion [6], where slices of a 2D spatial domain at consecutive time intervals are layered into a 3D volume (Figure 5). With a time extrusion, using time as the vertical dimension, scientists could interactively select and query individual markers. Our eventual goal is to combine time-extrusion with machine learning classification, so scientists can better determine where and when significant changes occur as well as what features lead to those changes and which are engendered by the changes.

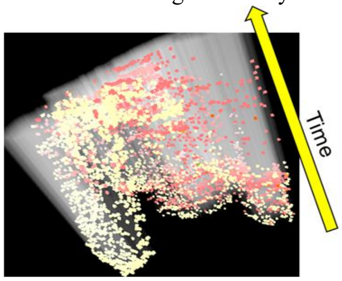

Figure 5: A Time-Extrusion Visualization

Acknowledgments

This project is supported by NSF grants: 1) BIO-ABI 10-62572: From Data to Knowledge in Grand Challenge Environmental Science Research and 2) CISE-CNS 16-37334: Connecting Communities Through Data, Visualizations, and Decisions. We also acknowledge our environmental science collaborators and all VISTAS developers at http://blogs.evergreen.edu/vistas.

References

[1] J. B. Cushing, E. Hayduk, J. Walley, K. Winters, D. Lach, M. Bailey, C. Thomas, and S.G. Stafford. Which visualizations work, for what purpose, for whom: evaluating visualizations of terrestrial and aquatic systems. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors—Novel Evaluation Methods for Visualization*, 2012.

[2] J. B. Cushing, N. Molnar, V. Ratanasangpunth, M. Bailey, J. Bolte, A. Brookes, D. Lach, B. McKane, G. Orr, et al. Visualizing Terrestrial and Aquatic Systems in 3D. VisWeek, 2014.

[3] J. B. Cushing, K. Winters, D. Lach. Software for Scientists facing Wicked Problems – Lessons from the VISTAS Project. 16th Annual International Conference on Digital Government Research, Phoenix, AZ, May 27-30, 2015.

[4] J. J. Halama. Penumbra: A Spatiotemporal Shade-Irradiance Analysis Tool with External Model Integration for Landscape Assessment, Habitat Enhancement, and Water Quality Improvement. Doctoral dissertation, Oregon State University, 2017.

[5] J. Harold, I. Lorenzoni, T. F. Shipley, and K. R. Coventry. Cognitive and psychological science insights to improve climate change data visualization, *Nature Climate Change* 6, 1080–1089 (2016). doi:10.1038/nclimate3162

[6] N. Schultz and M. Bailey. Using extruded volumes to visualize time-series datasets. In: *Expanding the Frontiers of Visual Analytics and Visualization*, Springer, 127–148, 2012.

[7] J. J. Thomas and K. A. Cook (eds). *Illuminating the path – The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society, 2005.