Scientific Visualization and Reproducibility for "Open" Environmental Science

Judith Bayard Cushing
Computer Science
The Evergrreen State College
Olympia, WA USA
judyc@evergreen.edu

Denise Lach
School of Public Policy
Oregon State University
Corvallis, OR USA
denise.lach@oregonstate.edu

Chad Zanocco
School of Public Policy
Oregon State University
Corvallis, OR USA
zanoccoc@oregonstate.edu

Jonathan Halama
Enivonmental Science
Oregon State University
Corvallis, OR USA
halamaj@oregonstate.edu

***Abstract*—Practicing reproducible scientific research requires access to appropriate reproducibility methodology and software, as well as open data. Strict reproducibility in complex scientific domains such as environmental science, ecology and medicine, however, is difficult if not impossible. Here, we consider replication as a relaxed but bona fide substitution for strict reproducibility and propose using 3D terrain visualization for replication in environmental science studies that propose causal relationships between one or more driver variables and one or more response variables across complex ecosystem landscapes. We base our contention of the usefulness of visualization for replication on more than ten years observing environmental science modelers who use our 3D terrain visualization software to develop, calibrate, validate, and integrate predictive models. To establish the link between replication and model validation and corroboration, we consider replication as proposed by Munafò, i.e., triangulation. We enumerate features of visualization systems that would enable such triangulation and argue that such systems would render feasible domain-specific, open visualization software for use in replicating environmental science studies.**

Keywords— model validation, replication, triangulation, terrain visualization, environmental science

I. Introduction

Reproducibility is presented to college freshmen as a lynchpin of the scientific method. However, a recent survey of 1500 scientists reports that "more than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments" [1]. The term *reproducibility* is often conflated or loosely applied [2]. For some, *reproducibility* is used in a strict sense: using the same data and/or methods to reproduce the results of a prior study. Further, the term *replication* is often used synonymously with reproducibility in formal and common scientific parlance in both this survey of scientists and other recent publications [e.g., 1, 3]. Peng distinguishes between the two terms with **replication** referring to the process of generating "scientific findings using independent investigators, methods, data" *vs.* **reproducibility** "which requires that data sets and computer code be made available to others for verifying published results and conducting alternative analyses" [4]. Throughout this paper, we use Peng's definition of replication, and distinguish replication from strict reproducibility.

When proposing criteria to evaluate whether a given research study can be trusted, others even further relax replication requirements. Munafò and Smith suggest that "an essential protection against flawed ideas is triangulation, i.e., multiple approaches to address one question" [5]. Milcu et al. in an unpublished manuscript go further and suggest that "deliberate introduction of controlled systematic variability (CSV) in experimental designs can increase reproducibility," the idea being that "a robust effect generalizable across many conditions is more likely to stand out" [3, 6]. The Open Science missions—"to increase openness, integrity, and reproducibility of research," and "not to waste time…on results that are not reproducible" [7] would, we believe, be supported by methods and software that makes it easier for environmental scientists and ecologists to produce replicable science and to more easily replicate already published research.

The two questions pursued in this position paper are (1) what role might scientific visualization play in replication for the environmental sciences—to recognize and then confirm the same robust effect over a particular landscape as seen in two different studies, and (2) what software features would facilitate the use of visualization in replication. To that end, we draw on our qualitative studies of environmental scientists using our visualization software to validate, calibrate, integrate, and present results of their predictive models. We argue that those processes are analogous to triangulation—the corroboration and replication approaches recently championed in the literature. Finally, we discuss the visualization features that would facilitate corroboration and replication in environmental science.

Our 3D terrain visualization software Visualization of Terrestrial and Aquatic Systems (aka VISTAS) was launched in 2011 as a collaborative, interdisciplinary project supported by the U.S. National Science Foundation with the purpose of bringing together computer scientists, social scientists, and environmental scientists to address complex problems. The software design process is an extension of Munzner's nested blocks and guidelines model [8], and the use of VISTAS visualizations to present our collaborators' findings are

This research is funded by U.S. National Science Foundation CISE-1637320, DBI-1062572, CISE-0917708.

described elsewhere [9, 10, 11]. For the purposes of this paper, it suffices the reader to know that VISTAS seeks to enable scientists to better understand and communicate information about complex environmental problems. VISTAS can quickly process and display large amounts of geospatial data, and environmental scientists use it to view their modeled or remotely sensed data in conjunction with impacted complex topography and other landscape features. Users can view data interactively over time and space and use simple analytics while viewing graphical images over time (descriptive statistics, linear and multiple regression, and principal component analysis).

We present here the extensive use of VISTAS by two environmental scientist collaborators in validating and calibrating their models, VELMA and Penumbra. The eco-hydrological model VELMA (Visualizing Ecosystem Land Management Assessments) predicts the effectiveness of alternative green infrastructure scenarios for protecting water quality. It is used by scientists and collaborating land manager stakeholders to advise on the establishment of riparian buffers, cover crops, constructed wetlands, and other measures to intercept, store, and transform nutrients, toxics, and other contaminants that might otherwise reach surface and ground waters. The model can handle multiple spatial and temporal scales, from plots to basins to hundreds of square miles and days to centuries [12, 13].

Penumbra is a spatially-distributed irradiance model that provides solar energy reduction from topographic shadowing, forest shadowing, and cloud coverage at multi-year landscape scales. It aims to enhance understanding of how light energy impacts ecological processes within landscapes to answer questions about forest management or restoration of riparian zones or fish habitat [14]. Both VELMA and Penumbra are grid based, and each utilizes precise digital elevation data, coupled with a number of input parameters (aka environmental drivers), to generate model results (aka response variables).

II. Model and Validation Criteria

VELMA and Penumbra underwent standard and stringent validation prior to publication and use; additionally, calibration and further validation to specific locations was conducted prior to subsequent use for predictive purposes in environmental resource management. We distinguish validating model software from software verification; in verification, a developer seeks to assure, often formally, that a given program is correct, i.e., that it correctly implements the program's specifications [15, 16, 17, 18]. Verification says nothing about whether the specifications, and hence the program, match an external physical truth. Validation confirms that the model is accurate with respect to the physical, real-world system it is meant to represent. A model or program can be verifiably correct, but not valid, or valid but not verifiably correct; a model cannot be proven correct for the real world, only validated for a certain use. When scientists validate a model, they make a best effort to confirm that the model represents the real-world situations, per its purpose and intended context (e.g. [19]). Model calibration is a similar process whereby parameters specific to local conditions are set, and the model is then validated for that context. Validation and calibration are therefore related processes. In spatial models that are mechanistic, models can be over-parameterized to generate statistically optimal outputs but not necessarily properly represent the environmental process across the landscape.

To run a model, a user specifies relevant environmental driver variables and any required local parameters. For VELMA these include grid-based digital elevation data and precipitation, temperature, soil porosity, etc.; and for Penumbra Julian start/stop days, a normalized digital surface to represent landscape objects such as tree canopy cover, etc. Calibration for VELMA involves determining and setting hydraulic conductivities and for Penumbra local land cover, landscape, and atmosphere.

Common practice for both validating and calibrating predictive models involves backcasting, a process whereby a model is run for a period of past time for which observed data exists; the predicted (past) model results are then compared to actual (observed) values. See Fig. 1, 2 for backcasting examples with VELMA and Penumbra. When predicted and observed values "agree," the model (or calibration for a particular space/time) is said to be validated.

Backcasting comparisons that show a model is invalid do not, however, tell the modeler much about where the model is going wrong, and what to do to fix it. VELMA and Penumbra modelers used VISTAS to view response variables over time and space, and examined specific topographical and landscape features, observing whether the predicted values made physical sense. If not, then the modelers leveraged intermediate variables to fix errors in the code or revised the science or theory that the code (to be revised) would model.

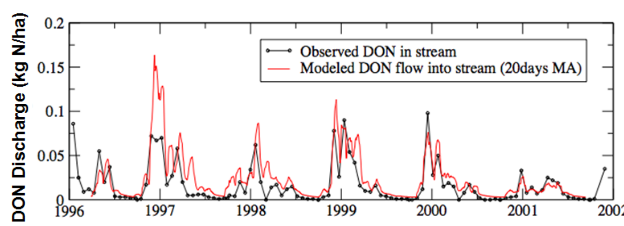

Fig. 1. VELMA backcasting validation example.

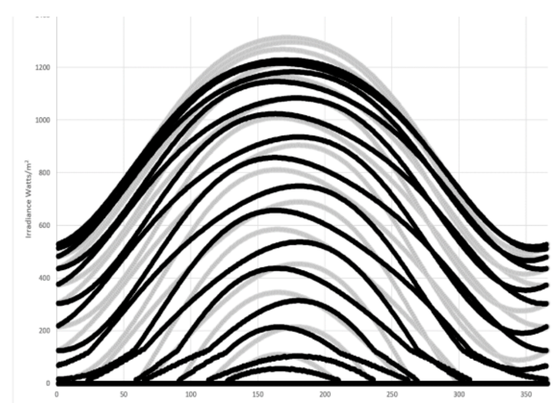

Fig. 2. Penumbra backcasting validation example.

The VELMA model was validated by using VISTAS to display model outputs that were difficult to interpret using prior 2D runtime displays or for which no runtime displays were available. VISTAS 3D still images and animations displayed source areas and downslope transport of nitrate within coastal watersheds; these verified that VELMA was correctly representing measured data describing the timing and location of nitrate hotspots within the watershed. When calibrating VELMA for watersheds where data for backcasting were not available, VISTAS 3D animations were used to examine whether the modeled data were consistent with general principles and patterns seen in similar well-validated areas.

To validate Penumbra, Halama et al. used the Environmental Protection Agency (EPA) Crest-to-Coast dataset, a transect of field monitored locations with paired open- versus forested-sites [20]. Each location has an array of sensors, including LICOR photosynthetically active radiation (PAR) sensors that measure irradiance in micromoles/m^2s^{-1}. They set spatial data inputs for Penumbra and captured modeled irradiance data at the location of the open- and forested-sites and then compared model results to the PAR sensor data. VISTAS was used both to provide a qualitative understanding that Penumbra's shading and irradiance made logical sense, and to visualize the modeled shade and irradiance results. The modelers created videos with four windows to visualize the topographic shade, object shade, total shade (topographic and object combined), and solar energy, and then viewed them simultaneously as part of the model validation process. This process involved synchronizing the views and controlling the speed of the video so any anomalies or unexpected outputs could be identified (e.g. by interactively selecting cells or areas in the visualization) and further examined. During model development, multiplicative interactions among variables can easily be inversed or over calibrated. Visualizing variables independently in VISTAS allowed model developers to identify such issues. Additionally, modelers might modify model runs so that additional variables can be viewed; for example, one of the modelers (Halama) wrote into Penumbra the ability to output the processed biomass to tree height results (even though they were intermediate data) so he could visualize the conversions and assess whether they were spatially correct with appropriate minimum and maximum values (Fig. 5).

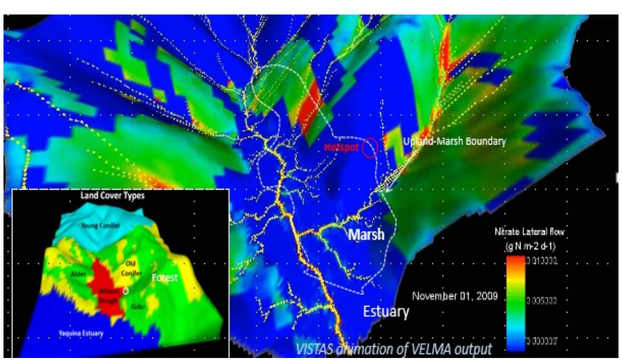

Fig. 3. Single frame from VISTAS VELMA animation.

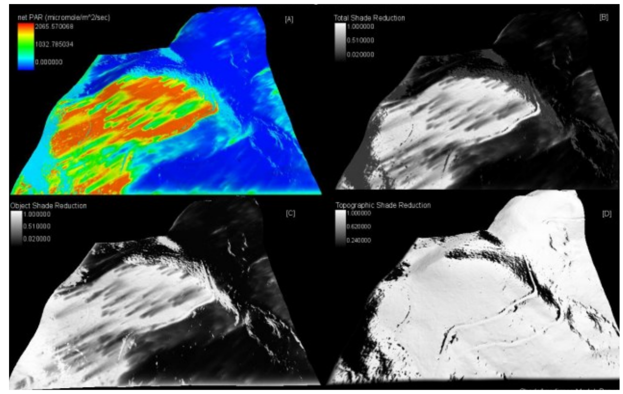

Fig. 4. VISTAS visualization of Penumbra and VELMA simulations.

If an error is found in the model or calibration—uncovered through visualization or by some other method—our modelers employ an iterative process to refine the model, correcting the code (adjusting model theory or its implementation) or changing an input or calibration parameter, and then visualizing and backcasting. See Fig. 3, 4 for example VISTAS visualizations used in validating VELMA and Penumbra.

Once Penumbra and VELMA were validated, the modelers used VISTAS primarily for sharing information with other scientists or decision makers. However, when those models were later extended, as when calibrated for a new ecosystem, modelers again visualized results to ensure that models were consistent with their understanding of the ecosystem context; if not, they began again the tedious process of adjusting the theory (code), visualizing, parameterization, backcasting, calibration, etc.

Our experience working with scientist-collaborators on model visualizations suggest that scientists who integrate models representing components of complex environmental systems typically go through a process of continual model refinement. While we observed modelers using visualization for model validation in our collaborative work, we posit that as models are iteratively refined and improved with increasing complexity, visualization will become more important as a validation tool.

III. From Validation to Replication for Complex Systems

In this section we first distinguish replication from reproducibility and establish why reproducibility for studies of complex systems is an oxymoron since each complex system is unique by definition [21, 22]. We then explain how the process of replicating studies of complex ecological systems is similar to the process of validating models of such systems and that, just as visualization has been shown to be helpful in validating complex models, visualization could be helpful in validating and replicating ecological studies.

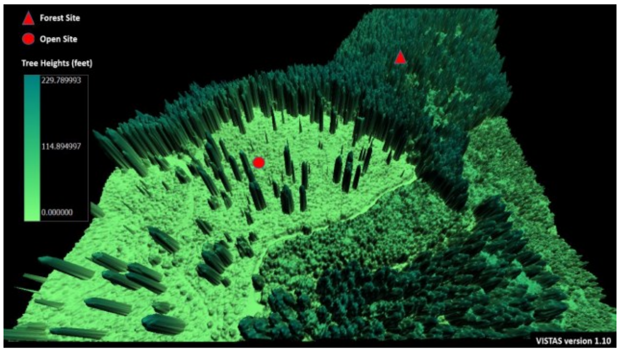

Fig. 5. VISTAS visualization of Penumbra-processed biomass to leaf transmittance.

A. *Replication of Ecological Studies*

Recent discussions of a crisis in science regarding reproducibility across the disciplines have led to refined criteria for "trustworthiness." Problems of reproducibility have been reported in fields that examine complex systems such as medicine, psychology, and ecology [1, 3, 23]. Clearly some published studies contain obvious errors and can be shown to be untrustworthy without reproducing them [24], but some reproducibility studies show significant issues. For example, the Open Science Collaboration (OSC) replicated 100 (psychology) studies, comparing the percentage of statistically significant results of the original (97%) to that of the replicated (68%) and reporting that only 36-47% of the original studies were successfully replicated [23]. This report has been countered, however: others found that when the OSC results were corrected for error, power and bias, the replication study provided no support for a crisis, and was "consistent with the opposite conclusion" [25].

Similar debates about reproducibility and replication, and the trustworthiness of science occur among ecologists. Where no reproducible errors are found in a study, there often remains doubt in the validity of the conclusions, which is especially problematic where the cost of a type II error (falsely inferring the absence of some phenomenon) is high in areas such as climate change impacts. In such cases, scientists seek to replicate study results and increase the trustworthiness of the findings so that stakeholders are more likely to take appropriate action.

For such systems, trustworthiness might be established via "the replication of scientific findings using independent investigators, methods, and data" [4]. Milcu et al. have experimented with an approach to increase reproducibility/replicability by deliberately introducing systematic variability in experimental design, varying experiments, increasing sample size, and hence noise in complex experiments [6]. A recent *Nature* editorial suggests that these methods "sow the seeds of trust for multi-lab replication efforts" and shore up "the reliability of field studies" [2]. For Milcu's group of researchers, subsequent data analysis, while non-trivial, is attainable because the 14 studies were designed for cross analysis (Fig. 6).

For replication studies that cannot be tightly controlled in a laboratory as Milcu's, or rigorously designed prior to execution of the study under replication, Munafò suggests using triangulation, where different methods are used to confirm the same result, rather than "producing statistically significant, definitive studies centered on an endpoint that supports a hypothesis." He suggests triangulation as one way to carry out Lipton and E. O. Wilson's ideas of "inference to the best or 'loveliest' explanation," abductive over deductive reasoning, likely explanations, and consilience" [5]. The question then arises for scientists: how to demonstrate that these different studies corroborate each other.

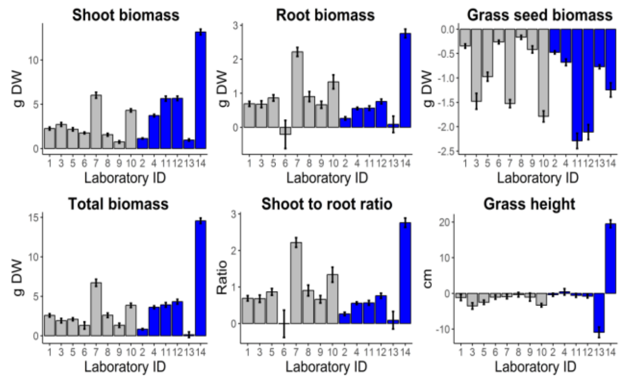

Fig. 6. Example analysis of Milcu's replicaiton study, 6 of 12 response variables from 14 laboratories.

B. *A Role for Validation in Replication*

In prior sections, we established that environmental systems research requires approaches different from classic reproducibility to achieve replication goals. Munafò proposed triangulation as a method to increase trustworthiness of studies of such complex systems as environmental sciences—using multiple studies to independently establish a causal relationship between driver and response variables.

We have observed that environmental scientists use visualization to validate and calibrate models; in relating drivers to response variables in a model, an ecologist is in effect attempting to establish a causal inference. We anticipate that, with the correct tools, scientists could similarly use visualization to compare multiple results from different studies on the same (or even verifiably similar) landscapes to use one study to replicate causal inferences made in another. The research question in such a replication study, using a method analogous to triangulation, is whether the several studies similarly characterize the ecosystem under consideration, and are converging (or not) on a commonly understood phenomenon or causal inference.

If ecologists work together in this way—different studies producing independent evidence in support of some central ideas—those of us who produce software need to think about how we can provide ways of viewing results of different studies in new ways. Munafò points to Wegener's initial observations that the shape of the west coast of Africa matched that of the east coast of South America. That visual intuition led him to look for evidence from many fields to support a new theory.

Scientists seeking to use, for example, two independent studies to corroborate a third could intuitively explore data from the three studies side-by-side. Visualization allows for these comparisons without significant additional data preparation that would otherwise be required to perform statistical replication. If the visualization shows that independent studies are mutually supporting, scientists could then devise statistical or other tests that compare results more formally. Visualizing multiple results on a flexible topographical background, as for our modelers, might give the intuition of whether the studies being compared corroborate, and hence whether a detailed data analysis is warranted.

C. Visualization Software and Open Science

VISTAS, coded in Python and developed specifically for a certain kind of environmental science, is a domain-specific application that is (according to our users) “very easy” to learn with only a brief training video [26]. As it stands, however, VISTAS would only be useful in replication for a relatively narrow range of environmental scientists: those dealing with grid-based data for topographically complex landscapes and using spatiotemporal data. To generalize beyond the user class, domain and functions for which VISTAS was developed would require accepting and wrangling more kinds of input data, creating new visualization types, making the system more easily usable for a wider range of users, and including more analytical capability.

While environmental scientists can use general scientific visualization systems for replication, these systems often require significant learning time beyond what most scientists can invest, and do not usually provide users with the level of analysis required for conducting replication studies. Heer has used design methods for domain-specific languages that are now widely deployed in data transformation and information visualization software (with reported order-of-magnitude productivity gains) [27-30]. His strategy has been to model user interface actions in a domain-specific language, and then leverage the language to predict potential actions and decouple the user interface from the underlying runtime.

If triangulation using scientific visualization is indeed, as we have argued, a valid method for replicating environmental science studies, and if the datasets for such replication studies are made available in open data repositories, then more effort should be devoted to making the necessary visualization software openly available and usable.

IV. Summary and Future Work

In this position paper, we reported that VISTAS, software for 3D terrain visualization, is useful to modelers in conducting model validation. We argued that model validation is analogous to corroborating multiple studies using triangulation, and that since visualization is useful in validating ecological models of ecosystem processes across landscapes, it is also likely useful in replicating studies of ecosystem processes across landscapes—where triangulation could be used in lieu of more traditional methods of replication. To demonstrate our claim, however, we recognize that one would need to observe ecologists using visualization in triangulation-replication. We envision two kinds of studies that would be needed to establish our claim and determine under what circumstances terrain visualization, at least as powerful as VISTAS, would be helpful as a first step in replication: where two (or more studies) seek to establish a relationship, or a causal inference between/among variables (1) over the same landscape, or (2) over different terrain but similar topographic or ecosystems.

Clearly (1) is most similar to the model validation and calibration where we have already established the usefulness of terrain visualization. (2) would be more difficult to establish and would require that the visualization system allow for such exploration as an interactive shifting of the landscape to align similar topographic features, modifying the scale of one or all scenes independently.

Where two or more studies seek to establish a relationship or causal inference among variables ranging across different topographic and ecosystem types, the value of visualization is more tenuous. However, in our work we have observed that researchers are indeed linking multiple models across different ecosystems, looking for links and causal inferences across time and space. These scientists plan to use visualization of outputs from different models to validate, calibrate, and revise multiple integrated models in a way similar to the way VELMA and Penumbra used VISTAS to validate model output. This lends some credence to the possibility that visualization would be useful in such replication efforts.

With any replication/corroboration, visualization would be only the first step towards replication; researcher(s) would need to follow up with further analyses prior to claiming that one study replicates another. For these next steps, having the visualization system anticipate and seamlessly feed data into an appropriate data analysis would be ideal. Outside the scope of this paper are recommendations about what kinds of statistics to integrate with visualization: changes in analysis methodology may be needed for the new ways of thinking about what constitutes robust findings, and these analyses might be more similar to “big data” explorations than traditional statistical analysis.

One desirable outcome of this paper would be for ecologists or environmental scientists in the Open Science community to test the hypotheses presented above and conduct replication experiments using Open Science Data and terrain visualization software such as VISTAS. Since VISTAS is open-source, implemented in Python, and runs on either Windows or Macintosh machines, and accepts a number of data input formats, such an experiment would be feasible; other terrain visualization software (e.g., ParaView) might also be appropriate for such a study.

We also call for research and development of Open Science scientific visualization analogous to those that Jeffrey Heer promotes for information visualization and data transformation [27-30]. While a visualization system like VISTAS is useful to, and actively used by, a small group of scientists, it is financially infeasible to develop such specialized applications for a great number of users in other even similar domains. To have an impact on environmental science, visualization systems need to facilitate interactive data exploration, with recommendations for such systems summarized as follows: (1) be scalable from meters to thousands of miles and from temporal scales of days or weeks to months, seasons, and years; (2) be intuitive to use; (3) be able to anticipate user needs; (4) be open source; and (5) be rendered domain-specific for a greater number of input data types and domains.

Acknowledgment

We thank collaborators Bob McKane and Allen Brookes (Environmental Protection Agency, Office of Research and Development, Western Ecology Division), Kirsten Winters (Ph.D., Oregon State University), Susan Stafford (formerly of University of Minnesota), John Bolte (Department Head, Biological and Ecological Engineering, Oregon State University), and members of the VISTAS development team Mike Bailey (Computer Science Department, Oregon State University) and, from The Conservation Biology Institute in Corvallis Oregon: Taylor Mutch and Ken Ferschweiler and, in particular, chief software architect Nik Molnar.

References

[1] M. Baker, "REPRODUCIBILITY CRISIS?," *Nature*, vol. 533, p. 26, 2016.

[2] "Reality check on reproducibility," Nature News, vol. 533, no. 7604, p. 437, May 2016.

[3] E. Callaway, "REPRODUCIBILITY Ecologists sow seeds of trust," *Nature*, vol. 548, no. 7667, pp. 271–271, 2017.

[4] R. D. Peng, "Reproducible Research in Computational Science," *Science*, vol. 334, no. 6060, pp. 1226–1227, Dec. 2011.

[5] M. R. Munafò and G. D. Smith, "Robust research needs many lines of evidence," Nature, vol. 553, no. 7689, p. 399, Jan. 2018.

[6] A. Milcu *et al.*, "Systematic variability enhances the reproducibility of an ecological study," *bioRxiv*, p. 080119, 2016.

[7] Center for Open Science, https://cos.io/.

[8] T. Munzner, "A nested model for visualization design and validation," *IEEE transactions on visualization and computer graphics*, vol. 15, no. 6, 2009.

[9] J. B. Cushing, K. M. Winters, and D. Lach, "Software for scientists facing wicked problems lessons from the VISTAS project," in *Proceedings of the 16th Annual International Conference on Digital Government Research*, 2015, pp. 61–70.

[10] K. M. Winters, J. B. Cushing, and D. Lach, "Designing visualization software for super-wicked problems," *Information Polity*, vol. 21, no. 4, pp. 399–409, 2016.

[11] K. M. Winters, D. Lach, and J. B. Cushing, "A conceptual model for characterizing the problem domain," *Information Visualization*, vol. 15, no. 4, pp. 301–311, 2016.

[12] A. Abdelnour, M. Stieglitz, F. Pan, and R. McKane, "Catchment hydrological responses to forest harvest amount and spatial pattern," *Water Resources Research*, vol. 47, no. 9.

[13] Environmental Protection Agency, "VELMA Eco-hydrological Model, Version 2.0." [Online]. **https://www.epa.gov/sites/production/files/2016-09/documents/508_velma_fact_sheet_revised_6-21-16.pdf**. [Accessed: 29-Jun-2018].

[14] J. J. Halama, "Penumbra: A Spatiotemporal Shade-Irradiance Analysis Tool with External Model Integration for Landscape Assessment, Habitat Enhancement, and Water Quality Improvement," 2017.

[15] D. Beyer, "Status report on software verification," in *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, 2014, pp. 373–388.

[16] B. Cook, E. Koskinen, and M. Vardi, "Temporal property verification as a program analysis task," *Formal Methods in System Design*, vol. 41, no. 1, pp. 66–82, 2012.

[17] B. Cook, D. Kroening, and N. Sharygina, "Accurate theorem proving for program verification," in *International Symposium On Leveraging Applications of Formal Methods, Verification and Validation*, pp. 96–114, 2004.

[18] N. Heath, "Why the blue screen of death no longer plagues Windows users," *ZDNet*. [Online]. Available: https://www.zdnet.com/article/why-the-blue-screen-of-death-no-longer-plagues-windows-users/. [Accessed: 29-Jun-2018].

[19] E. J. Rykiel Jr, "Testing ecological models: the meaning of validation," *Ecological modelling*, vol. 90, no. 3, pp. 229–244, 1996.

[20] US EPA, LI-COR PAR sensor dataset from the Pacific Coastal Ecology Branch, Science Inventory » US EPA, Newport, https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NHEERL&dirEntryId=341837.

[21] M. Mitchell, *Complexity: A guided tour,* Oxford University Press, 2009.

[22] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-Lite: A Grammar of Interactive Graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 341–350, Jan. 2017.

[23] Open Science Collaboration, "Estimating the reproducibility of psychological science," *Science,* vol. 349. no. 6251, p. aac4716, Aug. 2015.

[24] D. B. Allison, A. W. Brown, B. J. George, and K. A. Kaiser, "Reproducibility: A tragedy of errors," *Nature*, vol. 530, no. 7588, pp. 27–29, Feb. 2016.

[25] D. T. Gilbert, G. King, S. Pettigrew, and T. D. Wilson, "Comment on 'Estimating the reproducibility of psychological science,'" Science, vol. 351, no. 6277, pp. 1037–1037, Mar. 2016.

[26] C.M. Zanocco, "Co-producing visualization software for environmental science applications" Ph.D. dissertation draft, October 2018, (unpublished).

[27] J. Heer, J. M. Hellerstein, and S. Kandel, "Predictive Interaction for Data Transformation.," in *CIDR*, 2015.

[28] J. Heer, ACM Grace Murray Hopper Award. *https://awards.acm.org/award_winners/heer_1520709*. [Accessed: 29-Jun-2018].

[29] J. Heer, "Predictive Interaction." CMU HCII Seminar, March 2017. [Online]. *https://hcii.cmu.edu/news/seminar/event/2017/03/predictive-interaction*. [Accessed: 29-Jun-2018].

[30] J. Heer, "Predictive Interaction." NWDS Annual Meeting, January 2018. [Online]. *http://db.cs.washington.edu/events/database_day/2018/slides/heer.pdf*. [Accessed: 29-Jun-2018].