Visualization as a Tool for Ecological Analysis

S Kyle McKay, U.S. Army Engineer Research and Development Center, New York, NY, United States

Glossary

Information visualization "The processes of producing visual representations of data and the outputs of that work. Information visualisation aims to enhance one's ability to carry out a task by encoding often highly abstract information into a visual form. Visualisations can be static, or interactive and dynamic, and hosted in a variety of media (e.g., journal poster, website, or software)" (McInerny *et al.*, 2014).

Visual analytics "The science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2005 in Keim *et al.*, 2008) or "combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision-making on the basis of very large and complex data sets" (Keim *et al.*, 2008).

Visualization "A method of computing, [which] transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights" (McCormick *et al.*, 1987).

Introduction

Visual exploration of empirical, experimental, or model data is a powerful tool for increasing understanding of complex, long-term, and variable data sets common in ecology (Keim *et al.*, 2008; Fox and Hendler, 2011; McInerny *et al.*, 2014). Data visualization methods are well-studied in fields of visual analytics, information visualization, computer graphics, and scientific communication (e.g., McCormick *et al.*, 1987; Tufte, 2001; Keim *et al.*, 2008; Aigner *et al.*, 2011). Ecologists informally use visualization to parse data sets, guide analyses, and explore new ideas, but the field rarely acknowledges formally the role of visualization in ecological analysis and synthesis.

Large data sets are increasingly available in ecology (e.g., stream gage networks, high resolution sensor networks, large-scale remote sensing), and effective visualization techniques will be crucial to rapidly and efficiently understand and communicate these observations (Michener and Jones, 2012). Visualization cannot substitute for more rigorous quantitative and statistical methods (Garbrecht and Fernandez, 1994). However, visual exploration takes advantage of the capacity of the human eye to rapidly detect and discern visual patterns (e.g., color, shape, grouping), when presented effectively (McCormick *et al.*, 1987; Keim *et al.*, 2008; Fox and Hendler, 2011; Healey and Enns, 2012).

Given the breadth of ecological data types, formats, volumes, and analytical needs, innumerable data visualization approaches are potentially pertinent to the ecological community of practice. Rather than undertake a foolhardy review of these methods, the objective of this article is to highlight the value of visualization as a component of ecological analysis and synthesis and to present a variety of key issues that must be addressed in the selection and application of a visualization approach. The fields of visual analytics, information visualization, computer graphics, and scientific communication provide a rich body of literature on the subject, and this article serves only as an entry point for uncovering the seemingly endless body of data visualization approaches. To this end, data visualization examples are presented relative to four common ecological applications: data exploration, experimental analysis, numerical model output and evaluation, and ecological decision-making. The article concludes with a set of questions to guide ecologists in the selection and application of a visualization approach.

Reviewing Data Visualization Via Case Study

Ecological data visualization is inherently specific to a problem, purpose, or question. For instance, three questions about the management of an invasive riparian plant would drive an analyst to explore vastly different visual media: What is the plant's current extent (may lead to a map)? What environmental conditions influence the current distribution (may lead to a scatterplot between variable-x and plant density)? Does chemical-y effectively control the invasive plant (may lead to a barplot of mortality relative to treatment and control groups)? This pedestrian example is merely intended to suggest that visualizations are akin to other ecological analysis tools; the method must befit the need. Because of this intimate connection to applications, case studies are used herein to review common issues in visualization of complex ecological data sets. These examples often omit ecologically and analytically relevant details in the interest of focusing of key aspects of the visual approach. Case studies were selected to present a diversity of ecological applications and highlight crucial considerations for the visual presentation. Many potentially interesting visualization approaches were not considered (e.g., interactive graphics, animations) due to the constraints of the two-dimensional, print medium (See section Selecting a Visualization Method).

 doi:10.1016/B978-0-12-409548-9.10566-4

Data Exploration

Often ecological data are collected over long time scales to understand trends, patterns, and variability associated an ongoing process or system. Additionally, the resolution of these data streams is increasing as sensor networks improve and computational power increases. This often leads to the need to analyze patterns in large, complex data sets, and visualization provides an ideal tool for exploration of large data sets. For example, the Luquillo Long-Term Ecological Research Program in eastern Puerto Rico has trapped freshwater shrimp in tropical streams weekly for more than 20 years (Crowl *et al.*, 2017). Here, catch per unit effort from a single pool (Pool 0 in Quebrada Prieta) and species (*Atya lanipes*) from 1993 to 2014 are examined through the lens of multiple time series visualization methods (**Fig. 1**).

Many visualization analyses begin with traditional plotting methods such as line, scatter, or bar plots, and these simple figures can prove extremely valuable, particularly if a visual benchmark is used to call attention to a specific aspect of the data (e.g., the mean and standard deviation are shown in the upper left figure). As with quantitative methods, a variable may be transformed to highlight important nonlinearity in the distribution of a variable, but for visualization transformations should be conducted by way of altering the axis not the data to maintain natural units for the analyst. For instance, shrimp catch data are logarithmically

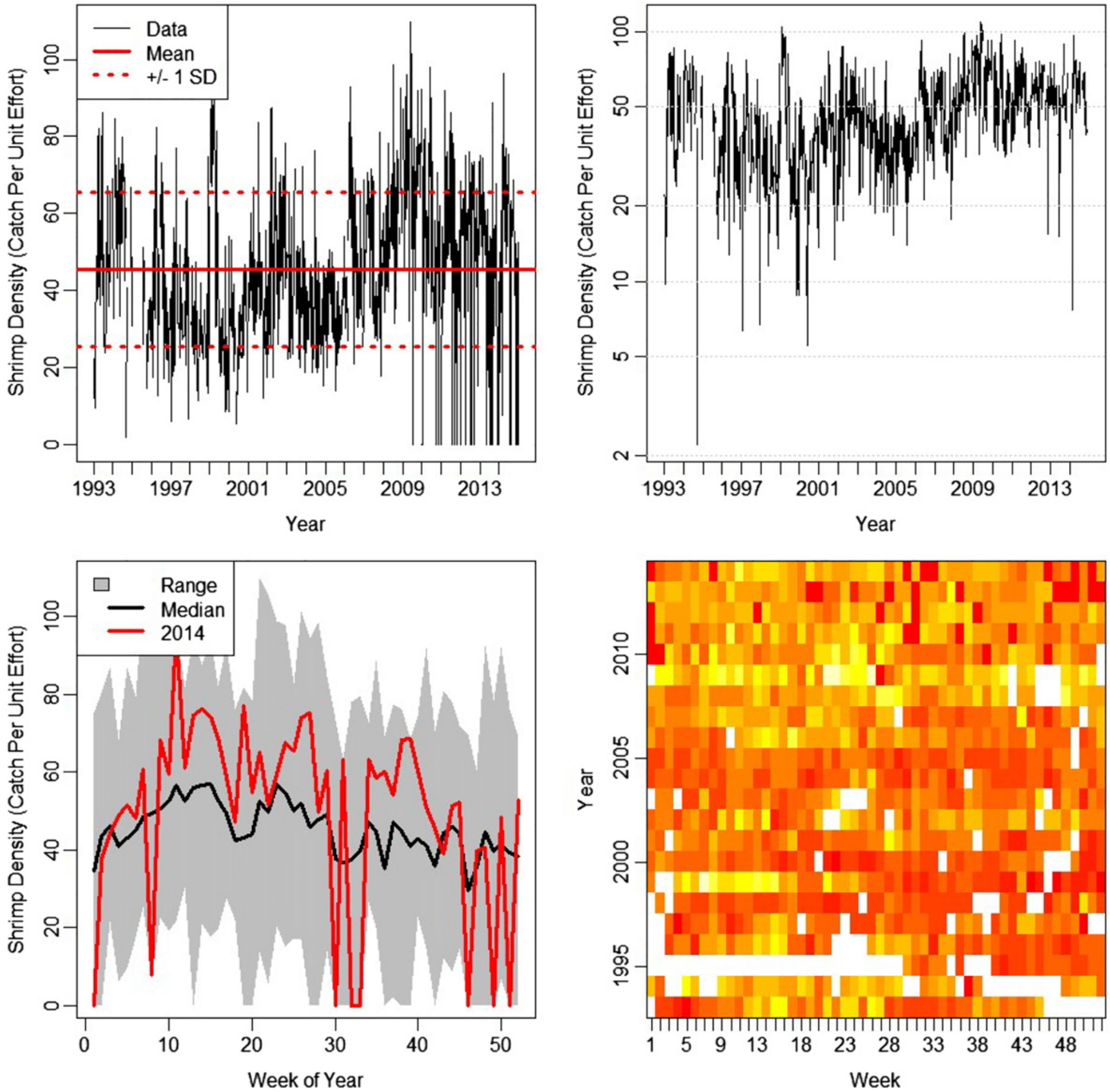

Fig. 1 Multiple visualization methods applied to a long-term, ecological time series. Data represent weekly freshwater shrimp (*Atya lanipes*) trapping densities collected at pool-0 of the Luquillo Long-Term Ecological Research (LTER-LUQ) site in Puerto Rico: (top-left) line plot with linear scale and reference points, (top-right) line plot with logarithmic scale, (bottom-left) envelope plot summarizing 2014 data relative to long-term observations, (bottom-right) a raster plot using color rather than location to represent density (red = high, yellow = low, white = missing).

transformed in the upper right figure, but the original units for catch per unit effort (CPUE) are preserved as a mechanism for maintaining understandability in the plot rather than *log* (*CPUE*). As a result of transformation, a more pronounced pattern of long-term catch declines and variability from 1995 to 2005 is shown, which is followed by a period of relative stability and high catch from 2005 to 2014. Ecological time series often exhibit periodic or cyclical patterns (e.g., diel movement, seasonal rainfall), and periodicity can provide a mechanism for compressing data for visualization. For instance, an envelope plot is used to provide historical context of the data (e.g., the range of observations and median for each sample week) for understanding a particular component of the data set (e.g., 2014 highlighted here; lower left figure). Although these line-based methods can lead to important observations, large data sets are challenging to communicate through this media because data are compressed into a finite plotting space. A variety of time series methods exist for overcoming these weaknesses, and the reader is encouraged to explore these approaches (Aigner *et al.*, 2011). As one example, a raster-based summary (lower right figure) presents the shrimp catch data set, and new observations may be made that were unapparent in other visuals such as data gaps (white cells), interannual variability (e.g., large-scale catch reduction across 2009), and massive intraannual variability (e.g., sequential high and low catch values in 2012).

Experimental Data

Experiments often help distinguish causal mechanisms in ecological systems, and visualization can help identify potential patterns, guide quantitative analyses, and facilitate the presentation of findings. More detailed guidance on the appropriate use of visualization in experimental analysis is addressed elsewhere (e.g., Weissgerber *et al.*, 2015), and only a few common issues are illustrated here. For instance, McKay *et al.* (2017a) used a tethered flight mill to study immune trade-offs with flight effort in North American monarch butterflies (*Danaus plexippus*). The experiments collected flight distance and duration data for individual monarchs in reproductively active summer conditions and reproductively inactive fall conditions for a maximum of four, 60-min flight trials conducted over sequential days. While additional analyses were originally pursued, the simplest form of the raw flight data are used here to address key aspects of experimental data visualization (**Fig. 2**).

The type of data generally serves as the first distinguishing feature driving a particular plotting approach with continuous and categorical data types providing the primary points of differentiation (e.g., 0.1, 0.8, and 1.2 and treatment-A and treatment-B, respectively; Weissgerber *et al.*, 2015). Typically, continuous variables lend themselves well to scatter and line plots, while categorical variables lend themselves to bar and box plots (top and bottom figures, respectively). Data grouping can also lead to alternative observations within a continuous variable. For instance, the top figures address the same data with all trials and individuals lumped into a single data set (left) and data parsed by individual monarch across four trials (right), and the two plots give the reader an alternative understanding of the amount of uncertainty and individual variation within the data. Generally speaking, a set of visualizations that effectively embed more data and information are preferred to those that present a single view of data. For instance, in the bottom figures, the bar plot conveys the general effect of the two seasonal treatments based on the mean observation, but the box plot provides additional information about the distribution of outcomes between the treatments (e.g., the lower tail of the duration data are affected more by season, nonnormality of the distribution is more apparent, the complete range of the data are shown).

Ecological Model Development

In addition to empirical approaches, visualization can inform many of the steps in an ecological model development process of conceptualization, quantification, evaluation, and application. **Fig. 3** presents a diversity of examples of visual approaches used to inform multiple types of ecological models. While not explicitly data visualization, conceptual representations of ecological models can guide the development of a model, structure data input–output, and serve as a mechanism for communication in an interdisciplinary team. For example, a conceptual model helped structure the development of an ecological model for quantifying the benefits of restoration actions in the Proctor Creek Watershed in Atlanta, Georgia (upper left, McKay *et al.*, 2017b), and the model was subsequently used by the restoration team to catalog potential restoration actions (orange boxes), input variables (white boxes), summary variables (gray boxes), and categorical outputs (yellow boxes). Visualization can also inform the evaluation of ecological models during calibration, verification, or testing phases. For instance, Shrestha (2016) developed a watershed-scale hydrologic model and applied separate portions of the streamflow gaging record to calibrate and verify model outcomes (upper right). Joint visualization of these phases, along with reference points of perfect prediction and $\pm 20\%$ error, allowed simultaneous consideration of the relative value of predictions and avoided systematic bias or error (i.e., calibration and verification have similar distributions of observed and predicted values). As ecological models become more sophisticated, so too must the visualization approaches for verification. For instance, Swannack *et al.* (2009) developed a model of Houston toad (*Bufo houstonensis*) movement through a complex, patchy landscape, and model outcomes were verified by a "Turing test," where a local subject matter expert familiar with toad movement was asked to select between observed and modeled landscape utilization patterns. Model results may also be summarized with visualizations in the application phase of ecological model development, but the model type and audience will play a large role in the selection of the visual approach. For instance, a population model of an oyster reef network was used to develop a network representation of the connectivity between reefs as well as a chord diagram summarizing key ecological aspects of connectivity such as the proportion of larvae moving to and from a given reef (middle row;

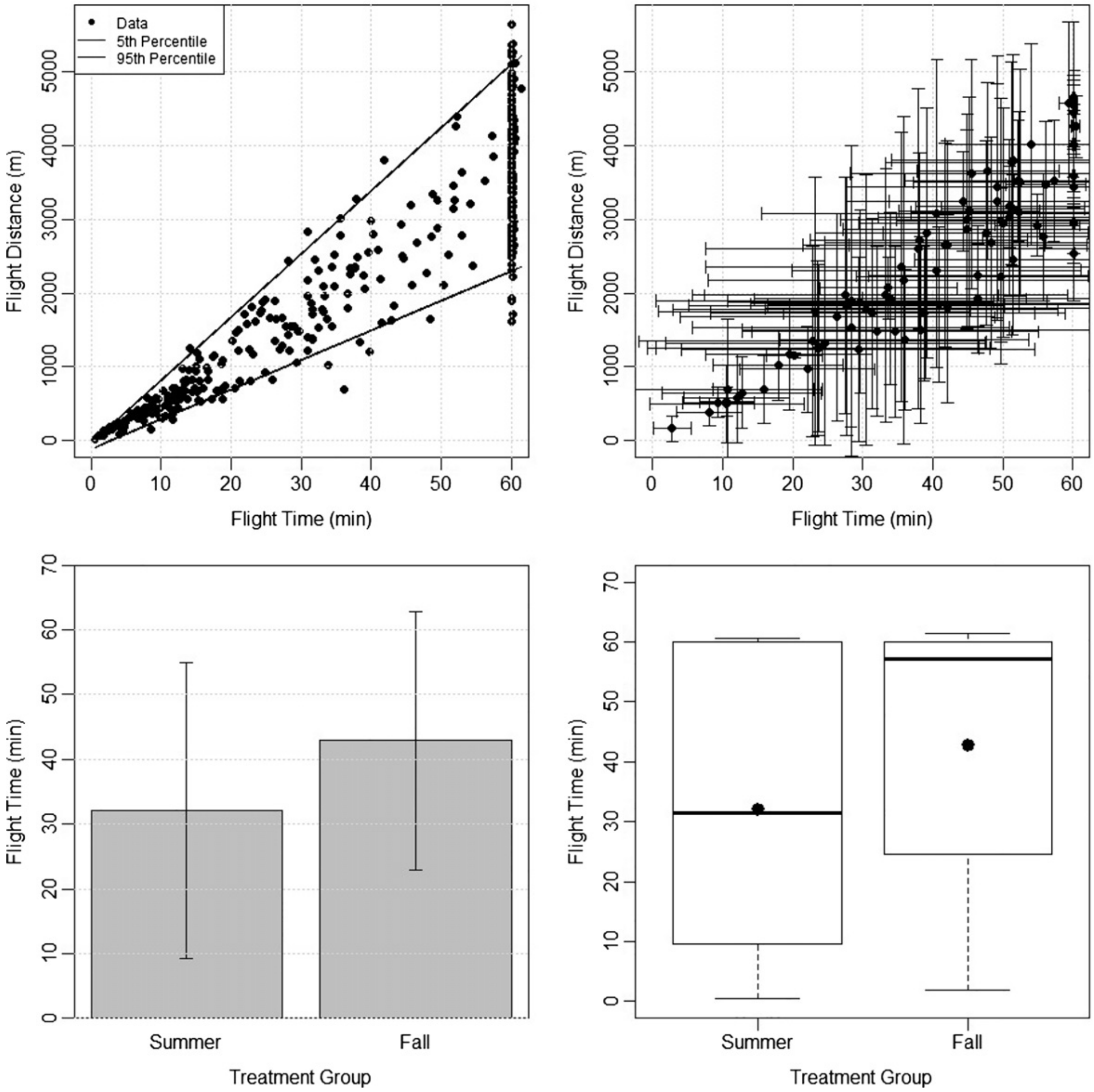

Fig. 2 Visualization of experimental results for monarch butterfly (*Danaus plexippus*) flight trials (McKay *et al.*, 2017a): (top-left) flight data from individual trials with a quantile regression bracketing outcomes, (top-right) flight data summarized by individual butterfly across multiple trials, (bottom-left) bar plot summarizing mean flight duration $\pm$ one standard deviation by treatment, and (bottom-right) a boxplot summarizing the effect of an experimental treatment across the entire sample population where the black point is the mean, black line is the median, box extent is the interquartile range, and whiskers represent maximum and minimum values.

Kjelland *et al.*, 2015). Conversely, the time series visualization methods described above were used to present a set of water management simulations to inform local officials of the magnitude of hydrologic change associated with a municipal water supply (bottom row). While ecological models are often system- or question-specific, visualization methods are agnostic to application, but selecting an appropriate visualization requires careful consideration of the step of the modeling process, goals of the visualization, and structure of the ecological output.

Informing Management Decisions

Visualization methods inform not only the ecological analyses themselves, but also the communication and condensation of those data for management and policy decision-makers (McInerny *et al.*, 2014). For example, an interagency team of federal, state, local, and nonprofit groups is partnering to restore the highly urbanized Proctor Creek watershed in Atlanta, Georgia. Visualization methods have served a vital role in facilitating knowledge transfer between the partners, and here two particular applications are highlighted that benefited from the visualization tools discussed in the article (**Fig. 4**). Urban watersheds represent diverse landscapes with many competing social, ecological, and economic objectives, and stakeholders often hold different relative value across those objectives. In this project, structure decision-making methods are being applied to examine the relative difference in

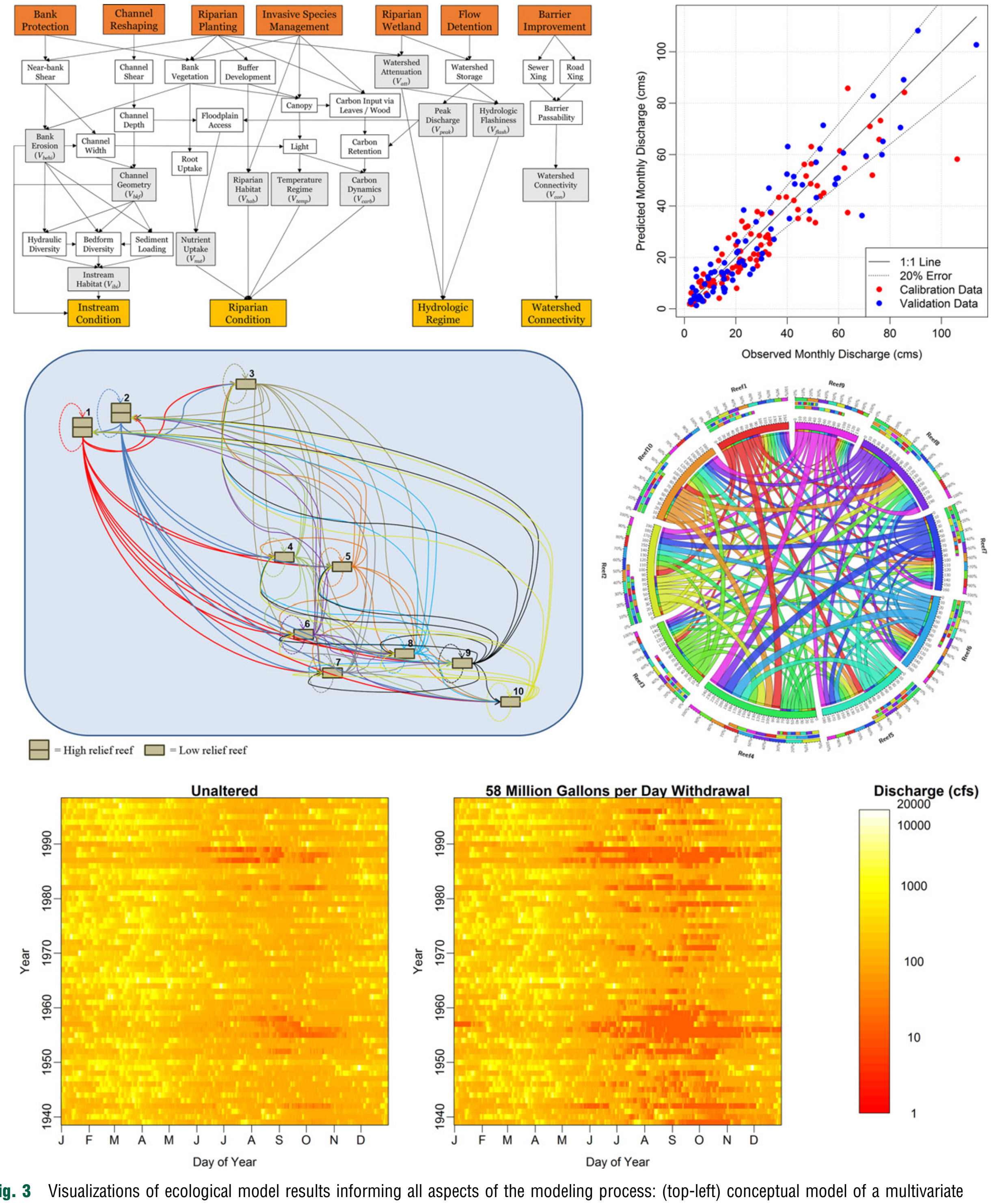

Fig. 3 Visualizations of ecological model results informing all aspects of the modeling process: (top-left) conceptual model of a multivariate model for stream restoration used in Proctor Creek, Atlanta, Georgia, (top-right) hydrologic model outcomes for both calibration and verification analyses (Shrestha, 2016), (middle) Oyster reef larvae dispersal network for a given year in a metapopulation dynamics model illustrating proportion of larvae originating from a given reef and going to other reefs, as well as the proportion of larvae from other reefs going to a given reef (Kjelland *et al.*, 2015), and (bottom row) raster-based time series of river discharge with and without water withdrawal.

stakeholder values across eight primary objectives. While not explicitly ecological data, the family of visualization methods presented in this article were used to summarize these data for communication among these divergent groups. Over 85 stakeholders were asked to distribute 100 "points" across each of the eight objectives, which were subsequently summarized in visualizations which show each stakeholder's values individually (top-left) as well as the aggregate value of all stakeholders (top-

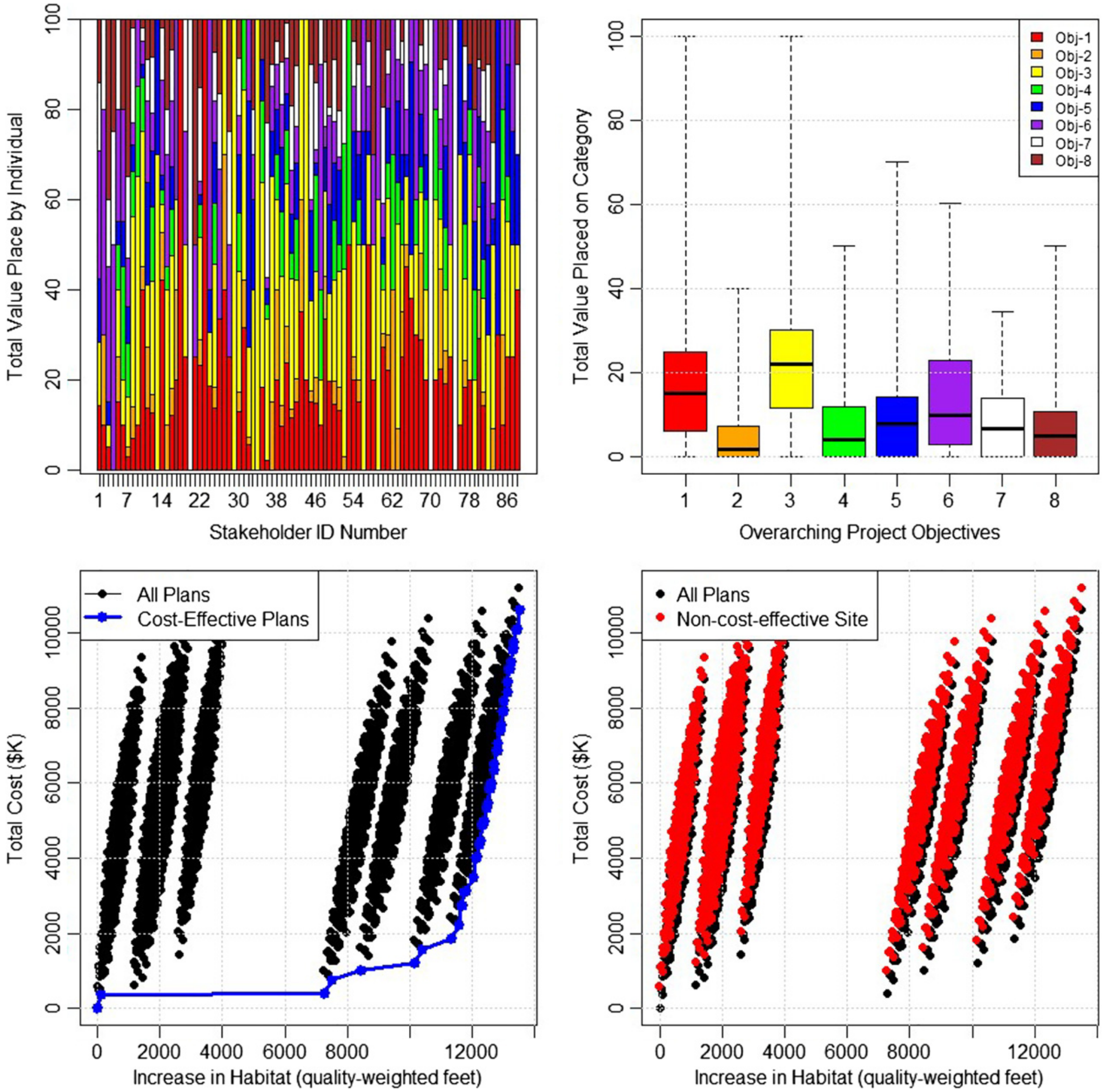

Fig. 4 Sample visualizations used to inform stream restoration decision-making in Proctor Creek, Atlanta, Georgia, United States: (top-left) distribution of values across 8 management objectives for 88 staekholders, (top-right) summary of the distribution of stakeholder values across objectives, (bottom-left) trade-off diagram highlighting sets of restoration actions that provide the most cost-effective restoration schemes, and (bottom-right) trade-off diagram highlighting plans including a particular site which shows that all sets incorporating this location are suboptimal (i.e., there are black points below and to the right of all red points).

right). In these figures, color is used to call information to the eye of the user and create a visual summary of the data (Healey and Enns, 2012), while still showing the range of values. For one of these objectives, an ecological model was developed to quantify the expected change in stream condition for a variety of potential restoration actions at 13 locations (8192 combinations of restoration actions watershed-wide). Model outputs were condensed into trade-off diagrams for decision-makers (bottom row), but the figures were then used to highlight particular aspects of the decision. For instance, suites of restoration actions were identified that were cost-effective (i.e., maximum habitat gain per cost and minimum cost per habitat gain), and color was used to call attention to these for decision-makers (bottom-left). Conversely, restoration sites were identified that were never included in cost-effective actions, which were also summarized with visualization (bottom-right). In communicating for decision and management, visualization must be careful to accurately represent information without imposing a judgment (McInerny *et al.*, 2014). For instance, color selection often implies preference, which can be used appropriately to guide decision-makers away from ineffective actions based on ecological evidence, but analysts must be careful not use visualization to impose their judgment on complex environmental decisions.

Selecting a Visualization Method

The vast array of options for data visualization can make choosing a technique challenging. The following questions are proposed to help ecologists navigate these methods. These questions are merely intended to structure thinking on method selection and are presented in an approximate priority of importance based on the key issues visualization (Aigner *et al.*, 2007, 2011; Kelleher and Wagener, 2011). Importantly, multiple methods can (and should) be applied simultaneously to explore a data set, emphasize specific components of those data, highlight different elements of variability, and guide quantitative analyses.

(1) What is the best method for communicating a message? Purpose and objectives should always drive the selection of a visualization method (Kelleher and Wagener, 2011). For instance, a simple line plot may be sufficient for communicating the time of sampling relative to recent environmental disturbances. However, a pixel-based approach might be more effective for communicating long-term trends (e.g., drought periodicity; **Fig. 1**). In particular, methods might change relative to the components of a data set that are of interest (e.g., extremes vs. central tendencies; **Fig. 2**).
(2) What are the relevant scales? Relevant ecological temporal and spatial scales range from minutes to decades and centimeters to thousands of kilometers in both length and interval. For many ecological processes, valuable understanding can be gained from the presentation of data in the historical context of a long-term record or large spatial domain. The resolution of the data relevant to a particular problem also influences the efficacy of a particular visualization method.
(3) How are data distributed? Ecological data often occur over many orders of magnitude (e.g., changes in streamflow, boom-and-bust of a population). Visualization can often be made more effective by rescaling figures, but care should be taken to maintain understandability in units (e.g., natural units of a process, not transformed units; **Fig. 1**).
(4) What are the constraints of the visualization environment? Selection of methods is influenced by not only the need to present data effectively, but also the limits of the presentation medium (Aigner *et al.*, 2011). The spatial extent, resolution, and use of color in a figure are often limited by screen or page size and the medium of interest (e.g., journal article vs. presentation vs. interactive on-line infographic). This article has focused on static presentation of a single time series, but the capacity to animate or interact with figures can increase visualization options significantly.
(5) What tools and expertise are available? A large variety of software is available to visualize data (e.g., Microsoft Excel, MATLAB, R Statistical Software, Geographic Information Systems). While all of these tools can present data in multiple formats, not all visualizations are easily conducted in all programs. In addition to availability and capability of the software, there may be personnel limitations in terms of expertise or time availability, which should not be overlooked. Herein, the freely-downloadable, R statistical software package was used to develop all figures (R Development Core Team 2016 Version 3.3.2), and code is available upon request from the author.

Conclusions

As ecological analysis has increased in sophistication and rigor, so too have visualization approaches (Kelleher and Wagener, 2011). However, broad adoption of many visualization methods has lagged. Large data, complex, and long-term data sets are becoming increasingly common in ecology (Michener and Jones, 2012). Using a variety of case studies, this article has highlighted a few (of many) techniques for visualizing ecological data and provided a set of criteria for guiding analysts to an appropriate technique. Visualization cannot substitute for rigorous quantitative analyses, but it can inform the analyst, guide the analyses, and facilitate communication (McCormick *et al.*, 1987).

See also: Ecological Data Analysis and Modelling: Mediated Modeling and Participatory Modeling. Statistical Inference. General Ecology: Communication. Principal Components Analysis. Human Ecology and Sustainability: Ecological Systems Thinking

References

Aigner, W., Miksch, S., Muller, W., Schumann, H., Tominski, C., 2007. Visualizing time-oriented data: A systematic view. Computers & Graphics 31, 401–409.
Aigner, W., Miksch, S., Schumann, H., Tominski, C., 2011. Visualization of time-oriented data. London: Springer-Verlag.
Crowl, T., Covich, A.P., Melendez-Colom, E., Perez-Reyes, O., 2017. Shrimp populations in Quebrada Prieta (pools 0, 8, 9, 15) (El Verde). Luquillo Long-term Ecological Research Site. http://luq.lternet.edu/data/luqmetadata54 Accessed 13 March 2017.
Fox, P., Hendler, J., 2011. Changing the equation on scientific data visualization. Science 331, 705–708.
Garbrecht, J., Fernandez, G.P., 1994. Visualization of trends and fluctuations in climatic records. Water Resources Bulletin 30 (2), 297–306.
Healey, C.G., Enns, J.T., 2012. Attention and visual memory in visualization and computer graphics. IEEE Transactions on Visualization and Computer Graphics 18 (7), 1170–1188.
Keim, D., Andrienko, G., Fekete, J.D., Gorg, C., Kohlhammer, J., Melancon, G., 2008. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (Eds.), Visual analytics: Definition, process, and challenges. Information visualization: Human-centered issues and perspectives. Berlin: Springer-Verlag, pp. 154–175.
Kelleher, C., Wagener, T., 2011. Ten guidelines for effective data visualization in scientific publications. Environmental Modelling & Software 26, 822–827.
Kjelland, M.E., Piercy, C.D., Lackey, T., Swannack, T.M., 2015. An integrated modeling approach for elucidating the effects of different management strategies on Chesapeake Bay oyster metapopulation dynamics. Ecological Modelling 308, 45–62.

McCormick, B.H., DeFanti, T.A., Brown, M.D., 1987. Visualization in scientific computing. Computers and Graphics 21 (6), 1–14.
McInerny, G.J., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F., Spiegelhalter, D.J., Stefanr, M., Tessarolo, G., Hortal, J., 2014. Information visualisation for science and policy: Engaging users and avoiding bias. Trends in Ecology & Evolution 29 (3), 148–157.
McKay, A.F., Ezenwa, V.O., Altizer, S., 2017a. Unravelling the costs of flight for immune defenses in the migratory monarch butterfly. Integrative and Comparative Biology 56 (2), 278–289.
McKay, S.K., Pruitt, B.A., Zettle, B., *et al*., 2017b. Proctor Creek ecological model (PCEM) phase 2: Benefits analysis. ERDC EL-TR. Vicksburg, Mississippi: U.S. Army Engineer Research and Development Center.
Michener, W.K., Jones, M.B., 2012. Ecoinformatics: Supporting ecology as a data-intensive science. Trends in Ecology & Evolution 27 (2), 85–93.
Shrestha, S., 2016. Impact of wood pellet production on water availability: A case study from Northeast Oconee River Basin in Georgia. Athens, Georgia: Master's Thesis, University of Georgia.
Swannack, T.M., Grant, W.E., Forstner, M.R.J., 2009. Projecting population trends of endangered amphibian species in the face of uncertainty: A pattern-oriented approach. Ecological Modelling 220, 148–159.
Tufte, E.R., 2001. The visual display of quantitative information. Cheshire, Connecticut: Graphics Press.
Weissgerber, T.L., Milic, N.M., Winham, S.J., Garovic, V.D., 2015. Beyond bar and line graphs: Time for a new data presentation paradigm. PLoS Biology 13 (4), doi:10.1371/journal.pbio.1002128.

Relevant Websites

www.visual-literacy.org/periodic_table/periodic_table.html—A Periodic Table of Visualization.
http://www.informationisbeautiful.net/—Information is beautiful.
http://survey.timeviz.net/—Survey of methods for time series visualization.
http://www.creativebloq.com/design-tools/data-visualization-712402—"The best 38 tools for data visualization".