Department: Graphically Speaking
Editor: Andre Stork, andre.stork@igd.fraunhofer.de
Rahul C. Basole, rahul.basole@gmail.com

Enabling Crosscutting Visualization for Geoscience

Danielle Albers Szafir
University of Colorado at Boulder

Francesca Samsel
Texas Advanced Computing Center
University of Texas at Austin

Stephanie Zeller
Texas Advanced Computing Center
Universary of Texas at Austin

Rick Saltus
Cooperative Institute for Research in the Environmental Sciences
University of Colorado at Boulder

***Abstract*—Our world is a complex ecosystem of interdependent processes. Geoscientists collect individual datasets addressing hyperspecific questions which seek to probe these deeply intertwined processes. Scientists are beginning to explore how investigating relationships *between* disciplines can foster richer and more holistic research, but visualization tools are conventionally designed to address hyperspecific, rather than holistic, analysis. Bridging the vast wealth of available data will require new tools. Visualization has the potential to support holistic crossdisciplinary analysis to understand the complex innerworkings of our world, but doing so requires a paradigm shift to understand how visualization might enable lines of inquiry transcending traditional disciplinary boundaries. We present challenges for visualization in fostering such holistic geoscience analyses.**

Introduction

In 1633, René Descartes completed his ambitious work, *Principia philosophiae*, previously titled *Le Monde* (The World), which set forth his understanding of the principles of nature: a comprehensive, systematic account of the universe. The book represented the then-noble, gargantuan undertaking of uniting the knowns of our universe into a singular *theory of everything*. Many would view such an undertaking today as an overreach, if not ludicrous. The scientific community has since shifted its philosophical approach to empirical research, having discovered that deeper and more rapid progress could be accomplished through hyper-specialization rather than generalization, as was popular in Descartes' time. We may now have reached the apogee of this pendulum swing: as each successive generation of researchers carves out increasingly deeper niches, they are hitting the bedrock of our inextricably connected ecosystem.

This ecosystem is especially central to the geosciences. Earth's systems are complex and interdependent. The study of these systems and the factors that drive their interactions produces a lattice of observed and simulated big data tied to physical properties, processes and spatiotemporal constraints. Our hyper-specialization system, while effective in capturing granularity, has thus far produced few generalized models to accommodate multiple datasets. Geoscientists in their respective subfields hoping to broaden their models and return to a Cartesian-esque approach now face the compounding challenges of uniting diverse datasets at differing scales and resolutions and discipline-specific systems into an *ecosystem* of models mirroring that which they aim to study. As the consequences of rapid climate change come to fruition, a model that can unite domains, characteristics, and scales gains significance beyond knowledge-creation, and must consider not only the study of environmental phenomena and their interrelationships, but also action-oriented cross-community dialog. The field of visualization has the potential to stand at the intersection of these multifarious sub-disciplines and achieve these ends. Here, we set out both the challenges and the potential for such a united ecosystem model through open standards and practices, and how the visualization community can address both the technical and the design aspects of this many-sided problem.

Just as Earth's natural ecosystems operate on both a micro and a macro scale, a data visualization ecosystem should consider how visualizations can not only address geoscience questions in the hyperspecific, but also how they may support scientists in investigating relationships between concepts that transcend disciplinary barriers. To accomplish this goal, we must explore how visualization tools can support (a) Identifying key relationships between diverse datasets to detect productive lines of inquiry, (b) Comparing data from these datasets across subdisciplines to analyze causal and correlative phenomena, and (c) Enabling effective communication of the ensuing multivariate, trans-disciplinary investigation to groups ranging from collaborating scientists to policy makers and the general public.

The Current Situation

Geoscience Systems

On-going efforts have sought to address pieces of these core issues [1]. For example, the EarthCube initiative [2] supports an evolving virtual community engaged in a wide range of hyperspecific projects, each designed to address hyperspecific problems in subsections of the geosciences. The sheer scope of the EarthCube efforts demonstrates the complexity and breadth of data sharing and common data exploration through visualization. Other examples of systems include the Australian Government-sponsored EarthSci [3] and the many and myriad data-specialized packages such as Geosoft's Oasis montaj (potential field data) or IHS's Kingdom Suite (seismic data). These and similar data-specific packages are largely geared toward data processing and analysis. Any output visualization is typically adhoc, limited, and created for a specific dataset.

Design Systems

Traditional processes for visualization design tend to focus on solving a single complex problem, such as developing a specific statistical model [4] or analyzing ensembles of meteorological predictions [5]. While definitions of rigor within design studies call for reflections on how individual systems inform visualization practice, such reflections seldom actively develop solutions for neighboring questions. For example, SimilarityExplorer [4] allows climate modelers to develop terrestrial biosphere models across an array of variables, but focus only on prespecified datasets and sets of methods. This problem-driven approach necessitates hyperspecific design thinking: designers focus on the immediate question motivating the system design. However, a lack of visualization standards between domains means that this specificity, at best, forces scientists to develop and translate findings across a series of separate tools and, at worst, can introduce fundamental incompatibilities that inhibit translating insight between geoscience datasets.

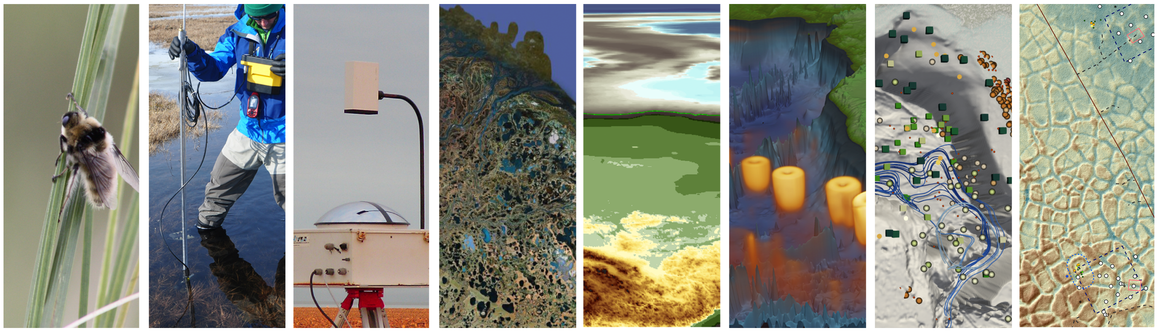

Figure 1. Geoscience data is collected in many ways, including direct observation (left) and simulation (center). Datasets often combine multiple collection methods to provide more holistic insight into a question (right).

Visualization across the geosciences needs a unifying visual and data language to facilitate crosscutting, holistic analytical solutions and to develop an ecosystem of visual and statistical techniques that encourage both deep problem-driven analysis and broad cross-disciplinary investigation of complex environmental systems. Establishing such standards requires conceptualizing visualization as not simply a solution to a single problem or as a shared set of techniques, but rather as a set of open standards and practices that *evolve and adapt* to new knowledge, tools, and data. We motivate our discussion of this vision through a representative use case drawn from one of the co-authors experiences as a geophysicist.

Motivating Use Case

Co-Author Rick Saltus is a Senior Research Scientist at the Cooperative Institute for Research in the Environmental Sciences at the University of Colorado, Boulder. During his 30-year career as a research geophysicist at the U.S. Geological Survey, he worked with several multi-disciplinary teams on tectonic framework and resource assessment studies. These experiences offer a broad and pertinent perspective on the challenges of cross-disciplinary geoscience visualization.

Saltus's field work is often interdisciplinary, conducted alongside geoscientists from a range of subdomains. Each scientists deals with different data and communicates in similar, but distinct, languages. Saltus, a geophysicist, has worked to acquire knowledge about his colleagues' fields to better communicate with them. He notes that even this approach is complicated by the myriad challenges created by varying data collection practices: Geologists often make connections between an outcrop of current study and a previous outcrop miles away from memory, but aren't able to make empirical comparisons because those datasets are disconnected. While Saltus's approach prioritizes uniform spatial coverage, it trades high-resolution for a wider area of intake. Geochemists may collect soil samples from specific locations to acquire surface-level knowledge of the soil's chemical breakdown on a regular grid, while seismologists collect broad, 2D data on specific geological features.

Saltus notes that someone familiar with ARCGIS often acts as a point-person for these interdisciplinary campaigns, pulling the group's diverse datasets together into one map or model. However, this approach removes scientists from the analysis pipeline: Saltus needed to give up agency over his data and his ability to interactively and iteratively drive data exploration. The point-person also likely lacked deep expertise in each subdiscipline present, limiting their ability to tie patterns to knowledge.

The underlying difficulty here, according to Saltus, is twofold: (1) geoscientists lack the tools and agency to enable them to explore and analyze multiple different datasets across disciplines, and (2) geoscientists lack a common language to facilitate this cross-disciplinary data-sharing. In Saltus's view, a visualization tool or set of tools that could bridge these gaps would be invaluable to his community, and would likely enable cross-disciplinary scientific breakthroughs by quantifying, verifying, and supplementing geospatial and temporal models from other disciplines.

Key Cross-Data Relationships

Field scientists like those Saltus describes are often working with isolated datasets, unable to expand the scope of their research beyond their own science questions in order to derive a big-picture, systemic understanding of their data in context.

This holistic approach to exploratory data analysis in the geosciences should enable scientists to rapidly identify meaningful avenues for deeper investigation from a library of datasets and to combine and interrogate those datasets within the same space. Achieving this vision requires techniques that allow scientists to intuitively sift through massive quantities of heterogeneous information and to establish a common visual language across disparate datasets to facilitate communication between datasets and disciplines.

Technical Challenges

Allowing scientists to navigate intra-datum relationships requires visual analytics tools that can compute, extract, and visually represent meaningful patterns across disparate datasets. Developing these tools raises interesting technical challenges for visualization.

First, visual analytics systems must create innovative *techniques for data fusion* that align large collections of data from disparate investigations to facilitate comparison. Data fusion techniques must wrangle the vast diversity of data types (e.g., ice density, magnetic fields, etc.) and sources (e.g., observations, simulations, sensor readings, etc.) found across the geosciences at scale and over multiple dimensions (Fig. 1). While systems might fuse data across time or location, alternative approaches may help scientists find relations across more esoteric yet critical attributes, such as specific patterns (e.g., global regions with similar frequency variations) or common analytical techniques (e.g., approximation techniques or modeling approaches).

While some of this fusion may be managed through sophisticated statistical approaches, the geosciences often rely on a broad variety of techniques both during data wrangling (e.g., interpolation or simulation to fill in missed values for sparse data and assimilation or thinning when data are too abundant) and collection (e.g., varying sensor precision) that can introduce *compounding uncertainties* or otherwise complicate automatically identifying relationships. For example, scientists studying magnetic fields may couple hand-collected measurements across a field-site with big-picture assays from remote sensing data. Manual measurements are subject to sensor error and interpolated across samples to estimate a continuous set of measures, introducing computational uncertainties. Remote sensing data contains its own measurement uncertainties. Pairing manual and remote sensing data aligns the data across space and time and over varying levels of resolution, introducing yet more uncertainty. This problem becomes compounded when we scale up this comparison to datasets across regions, times, or data types. These layered uncertainties create a challenging problem for comparison: how might systems enable people to reason across a range of datasets in light of complex, disparate uncertainties?

These same challenges also complicate system interaction: people need to be able to use data, metadata, and patterns as a means to *intuitively query* a large library of datasets. While strategies such as searching over time, topic, or geographic location lend themselves to traditional querying techniques like sliders or searchboxes, other correlations require more complex reasoning. Systems must allow scientists to fluidly and intuitively formulate and apply potential complex query criteria *through* data (e.g., sketching, query-by-example, etc.) rather than simply *about* data (e.g., SQL queries). Scientists will need to adjust these selections on the fly to accommodate evolving knowledge and insight.

While the sheer complexity of these challenges appears to call for fully automated solutions, holistic inquiry in the geosciences relies so heavily on scientists' evolving intuitions and knowledge about a given question that automated solutions are not only infeasible but undesirable. Automated solutions lack the intimate knowledge of the local context and science associated with a question while reducing scientists' agency. However, automated processes may help optimize scientists' time, "learning" from the scientist to help sift through vast quantities of complex data and enabling meaningful alignment between datasets. Enabling expertise to drive these investigations at scale introduces core design challenges to

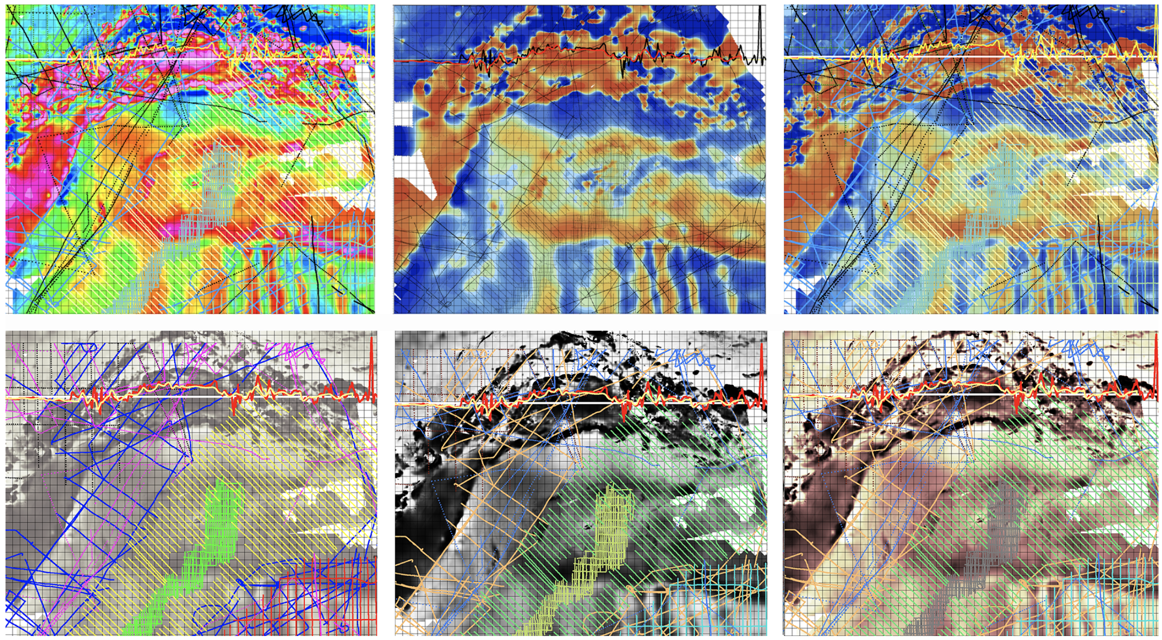

Figure 2. Rick Saltus's data is shown here as an example of the design challenges present in layering just one series of variables within the same image. Saltus needs to see the characteristics of the underlying bathymetry, each set of lines which represent six separate regions, and the horizontal line graph near the top of each image. The top left image shows the data with a default rainbow colormap applied. The top middle and right images are Saltus's attempts to remedy the rainbow's shortcomings. The bottom row shows more successful attempts, moving from grayscale topology overlaid with default discrete colors on the left, to a tinted grayscale and carefully chosen discrete hues in the middle, to Matlab's pink colormap on the right, which improves discrimmination power between hues. Even in two dimensions, scientists like Saltus face arduous challenges of both perception and design.

establish a visual language to facilitate scientific interchange between scientists and disciplines.

Design Challenges

To facilitate the preliminary exploratory analysis required by geoscientists like Saltus, we must consider not only the technical challenges of identifying points of overlap between diverse datasets from varying subdisciplines, but also the basic design challenges of visualizing these datasets in concert with the explicit goal of drawing out initial and useful intersections before moving forward. This kind of problem appears ready-made for an automated solution, but each dataset often requires unique visual maneuvering in order to reveal properties informative to the underlying science questions. Further, over-automating precludes the possibility of enabling scientists to adjust and explore novel intravariate and intra-datum relationships themselves—a process which can act as a catalyst for discovery.

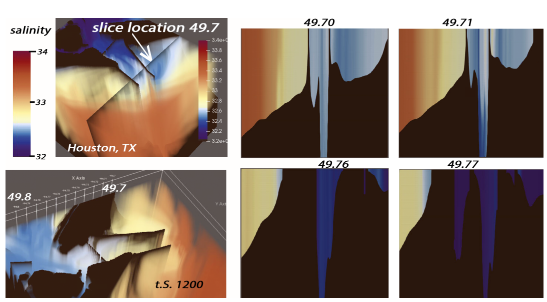

Figure 3. Volume rendering of salinity flow in Houston Harbor. Segmenting the opacity levels within the colormap exposes flow features hidden in slices alone.

Open visualization systems should include visual assets that support this interactive discovery phase. Because such a system would fundamentally require cross-community collaboration, usability that translates across disciplines is critical. Designing open systems and tools for broad

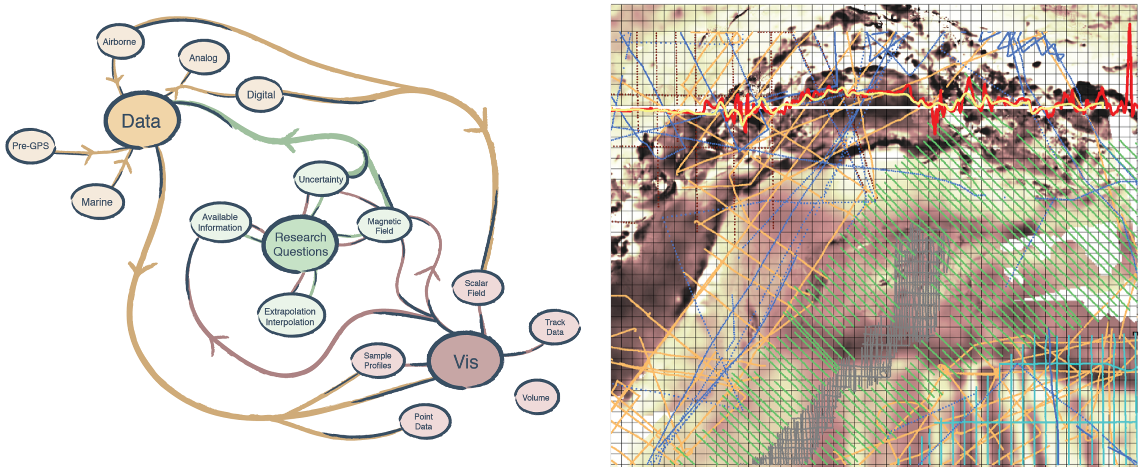

Figure 4. Holistic geoscience visualization introduces new complexity into the interplay between data, questions, and the corresponding visualizations. The above diagram summarizes current relationships between these variables for a single investigation from Saltus' work (Fig. 2, right) with connections indicating relations underlying the visualization. Integrating new data can further expand these relations, offering scientists broader perspective on each question, but requires addressing core visualization challenges.

use would necessitate deep investment in user interface and user experience research anchored in scientists' workflows. In an era of purely digital collaboration, we've witnessed this sort of innovation in the private sector to promote productive team dynamics, collaborative work and ideation spaces, and ease of communication and file sharing. A similar approach could theoretically be taken in cross-disciplinary visualization.

Further considerations include problems of holistic perceptual design. Generating clear layered and multidimensional data is a challenge rife with false artifacts, multiple transparencies, conflicting grid systems and resolutions, limited topologies, temporal variables and conflicting visualization standards across subdisciplines. Designing to maximize intuitive visual processing and minimize confusing discordance will present formidable design problems which may benefit from the inclusion of more diverse visualization professionals, such as career visual artists and designers, whose work in visual communication theory would expand the possibilities of cross-community collaboration within the geosciences.

Causal and Correlative Phenomena

Once scientists have identified potentially interesting datasets, they need to be able to dig into them, exploring the data in detail to assess how it may inform their core questions. Such investigation requires carefully inspecting relationships between their data and new datasets to verify and validate potential causal or explanatory hypotheses emerging from the data.

Technical Challenges

Many geoscience simulations and models are extrapolations based on sample-size observable phenomena: for example, Saltus often works with region-specific geological maps derived from a single method of measurement and data collection over several areas then amalgamated using statistical aggregation and sampling to fill in the gaps. The results serve as models to guide further research and as references against observed data to verify the validity of ongoing work.

This approach is widely used and extremely effective, but only provides a confidence interval of how well the predictions reflect reality within singular domains and regions. Cartographers often combine different datasets within the same region to generate layered maps, but this is a post-hoc process, occurring after scientists have refined their data. If visualization could instead support integration early in the analysis process, scientists could iteratively and collaboratively layer their respective *knowns*, finding points of overlap, verifying estimations, analyzing causal

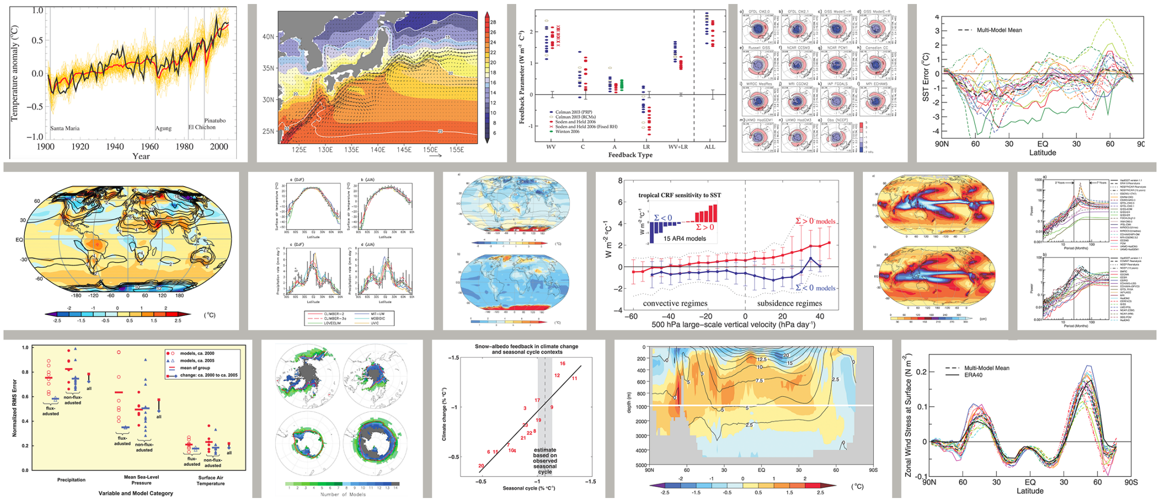

Figure 5. Graphics: IPCC Report, AR4 Climate Change 2007:The Physical Science Basis, Chapter 8, [6] showing a range of methods used for displaying climate data. The diversity of colors, forms, and projections highlights the difficulties in relating data across data representations.

and correlative phenomena, and bringing each subdiscipline closer to a collective ground truth.

Visualization can support either manual and automated integration. Manual processes are directly driven by the scientists. Visualizations add transparency to the statistical approximations scientists make, which can either introduce or resolve uncertainties in the data or other artifacts. Doing so, however, would mean finding ways to align these artifacts so a scientist can focus on the "new" data as well as new visual encodings representing that data. Automated processes can scaffold layered analyses either through machine learning techniques drawing from past analyses to apply manual procedures *en masse* or intelligently predicting data uncertainty and alignment using scientifically grounded "rules" (e.g., the amalgamated regions in Saltus' data) or by learning from prior examples [7].

Design Considerations

An image is like a symphony with many individual parts, each complementing the other and working in harmony. Harmony emerges when all of the notes, chords and tempo are in tune with each other. If one instrument is out of sync or off key, the harmony is disturbed and an audience's attention shifts to the discord. The same is true with visual imagery, including visualization.

Communicating through imagery requires a sophisticated understanding of visual cues and the ability to implement them in both the micro and the macro in order to transmit specific concepts.

For centuries, artists have employed Artistic Design Theory [9] to convey an intentional and directed plexus of meaning, hierarchy, motion, dominance, and intra-subject relationships in visual images.

Artistic Design Theory contains a set of Design Elements (the individual building blocks of image): points, lines, shapes, forms, texture and color, which act as the building blocks of an image. They align with the building blocks of visualizations: points, lines and streamlines, shapes, forms and glyphs, textures, hues and colormaps.

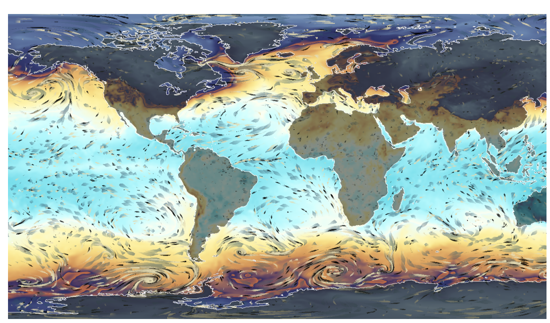

Figure 6. Global oceanographic data layering five variables plus uncertainty. Layering datasets allows scientists to deeply investigate interrelations within diverse datasets but requires significant support [8].

Design *Principles*, a subset of elements within Design Theory, are a set of compositional guidelines that organize and distill imagery to support the conceptual themes within an image or a work of art. These guiding principles include harmony, balance, repetition, emphasis, proportion, rhythm, hierarchy, pattern, variety and unity.

Artists use these theories and principles to distill, organize, and harmonize complex, disjointed forms within an image. However, unlike symphonies or works of art, visualizations do not have conductors to keep the elements working in harmony, nor are they compositionally dependent on an artist. Everything within the visualization composition impacts everything else and is determined by the data. As the number and variety of forms within the composition grows, so does the likelihood of cacophony and visual distortion, hindering the visualization's intended purpose. Each successive variable layer creates a new visual situation that can change over time with the data values. In three dimensional visualization, the composition can also evolve depending on a viewer's perspective within the model.

These challenges are compounded by limited colormap and glyph selections in most visualization software. Without a full range of encoding variety, generating visualizations that are both perceptually sound and aligned with relevant science questions is extremely challenging. Figure 2 illustrates the distracting interaction of variables produced from one scalar field and six track lines categories differentiated by discrete hues. Figure **??** is a visualization created with Vis-By-Sketching [8], a visualization tool designed to enable the fidelity of control needed to exercise artistic color principles.

The perceptual community has conducted significant work to tackle the individual elements which provide visual distinction, but less work has been done on identifying methods to combine the visual encodings. [10] When reaching across disciplines to pull *multiple* datasets together into the same visualization, these challenges will increase tenfold. In order to maintain cohesion and promote high-level exploration and sense-making for scientists, designer methodologies need to be seamlessly integrated into visualization software. These integrations must address not only individual encodings, but also the *holistic* design challenges that emerge in multivariate, multidiscipline data visualizations. We argue that if Design Principles and guidance can be integrated into the software, scientists and visualization professionals would be able to handle the increased complexity of multidisciplinary layering. Harmoniously layered data will enable discussion, exploration, and communication across disciplines and fields.

Figure 3, a volume rendering of salinity in Houston Harbor, demonstrates the challenges of representing 3D data as compared with slices, visible on the right. Here, a segmented opacity transfer function provides some insight into single variable volumetric data and illustrates the difficulty of incorporating multiple 3D variables. On the cover of this issue is a visualization of a 3D multivariate rendering from Keefe's team that uses a sampling algorithms to address this challenge.

Visualization as Dialog

Creating visual imagery is fundamentally an act of translation. From antiquity to modernity, visual imagery has always provided us with a means to transcend the abstractions of language and mutually consider an object on whose properties we could all agree, discuss, and form conclusions. This characteristic fortifies visualization as not merely a personal tool but rather a vehicle for iterative data interrogation and communication across individuals, domains, and communities. If we are to pursue a unified system as described here, we must consider the scope of communication opportunities enabled by transdisciplinary visualization design practices and open software. As we've learned from the humanities, each individual brings their own experiences to a visualization, interpreting objects, textures, and colors through personal associations built over a lifetime. Rather than seeking to minimize these varied interpretations by creating one rigid, standardized system, visualization professionals should instead allow for and encourage customization of visual assets through open software, standards, and practices to enable points (1)-(4).

Researcher-data dialog: Enabling clear dialog between a researcher and their data is an iterative process of data exploration. This facilitates what cognitive scientists call ”self-talk”:

internal dialog that allows each of us to logically walk ourselves through problems. By creating dynamic tools tuned to the scientists needs which allow for customization within a collaborative system, visualization professionals can revitalize the researcher-data dialog.

Intra-team dialog: The next layer of dialog is that which occurs between team members. Visualizations spur discussion by acting as a group commonality with agreed-upon properties. Building interactive, customizable tools that incorporate design practice guidance will enable research team members to effectively translate the results of their "self-talk" process into a shared visual language congruent with established group norms for specific subdomains, carving a path for collective data interrogation and collaborative analysis.

Intra-domain data dialog: A key piece of this proposed paradigm shift in visualization is that of intra-domain dialog. While the technical and design challenges of bringing together disparate datasets and enabling scientists to rapidly extract and emphasize key takeaways are formidable, they would usher in a new era of cross-disciplinary communication—a dearth of which has begun to stifle progress in academia and beyond. In order to understand our vast and complex ecosystem, the visualization community must both encourage and facilitate this bridging of disciplines by providing tools that allow us to create a shared language to investigate what are ultimately shared science questions.

Scientist-policy maker + public data dialog: If we can accomplish points (1)-(3), productive scientist-policy maker dialog will follow. Using a shared visual language, scientists may be far more effective in communicating to a lay public. An engaged, informed public begets attentive government officials and actionable science policy, at which point important insights about our world would begin to directly impact the lives of millions of people.

Conclusion

Addressing these challenges will enable us re-envision geoscience analysis. They do not address every micro- and macro-dimension of geoscience data but *will* move us closer to a holistic, shared *ecosystem* of models, visual languages, and communication strategies. This approach should enable geoscientists to more thoroughly explore their own data and the underlying phenomena informed by other relevant research *in context*. Visualization's contribution to addressing climate change and other key environmental issues depend on its ability (1) engineer tools and guidelines tuned to the needs of the scientists and their data, and (2) enable cross-cutting collaboration that transcends disciplines.

REFERENCES

1. A. Karpatne and S. Liess, "A guide to earth science data: Summary and research challenges," *Computing in Science Engineering*, vol. 17, no. 6, 2015.
2. "Earthcube." [Online]. Available: https://www.earthcube.org/
3. "Earthsci." [Online]. Available: https://www.ga.gov.au/data-pubs/interactive-3d-models/EarthSci-3D-Data-visualisation-tool
4. J. Poco, A. Dasgupta, and Y. e. a. Wei, "Similarityexplorer: A visual inter-comparison tool for multifaceted climate data," in *Computer Graphics Forum*, vol. 33. Wiley Online Library, 2014.
5. P. S. Quinan and M. Meyer, "Visually comparing weather features in forecasts," *IEEE TVCG*, vol. 22, no. 1, 2015.
6. D. Randall, R. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. Stouffer, A. Sumi, K. Taylor, S. Solomon, Z. C. M. M. K. A. M. D. Qin, M. Manning, and e. H.L. Miller, "2007: Climate models and their evaluation. in:climate change 2007: The physical science basis. faq 8.1, figure 1; 8.1; 8.14; 8.12; 8.8; 8.17; 8.3; 8.2; 8.15; 8.5; 8.13; 8.11; 8.10; 8.16; 8.9 and 8.6," Cambridge, United Kingdom and New York, NY, USA.
7. D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geoscience Frontiers*, vol. 7, no. 1, pp. 3–10, 2016.
8. D. Schroeder and D. F. Keefe, "Visualization-by-sketching: An artist's interface for creating multivariate time-varying data visualizations," vol. 22, 2015.
9. L. F. Rathus, *Foundations of Art and Design*. Stamford, CT: Cengage Learning, 2015.
10. C. Ware, *Information Visualization: Perception for Design, Edition 4*. Morgan Kaufmmann, 2020.