Bridging Theory with Practice: An Exploratory Study of Visualization Use and Design for Climate Model Comparison

Aritra Dasgupta, Jorge Poco, Yaxing Wei, Robert Cook, Enrico Bertini, and Cláudio T. Silva, *Fellow, IEEE*

Abstract—Evaluation methodologies in visualization have mostly focused on how well the tools and techniques cater to the analytical needs of the user. While this is important in determining the effectiveness of the tools and advancing the state-of-the-art in visualization research, a key area that has mostly been overlooked is how well established visualization theories and principles are instantiated in practice. This is especially relevant when domain experts, and not visualization researchers, design visualizations for analysis of their data or for broader dissemination of scientific knowledge. There is very little research on exploring the synergistic capabilities of cross-domain collaboration between domain experts and visualization researchers.
To fill this gap, in this paper we describe the results of an exploratory study of climate data visualizations conducted in tight collaboration with a pool of climate scientists. The study analyzes a large set of static climate data visualizations for identifying their shortcomings in terms of visualization design. The outcome of the study is a classification scheme that categorizes the design problems in the form of a descriptive taxonomy. The taxonomy is a first attempt for systematically categorizing the types, causes, and consequences of design problems in visualizations created by domain experts. We demonstrate the use of the taxonomy for a number of purposes, such as, improving the existing climate data visualizations, reflecting on the impact of the problems for enabling domain experts in designing better visualizations, and also learning about the gaps and opportunities for future visualization research. We demonstrate the applicability of our taxonomy through a number of examples and discuss the lessons learnt and implications of our findings.

Index Terms—visualization, design principles, climate model, taxonomy

1 Introduction

The ever-growing data deluge has made visualization an important medium for intuitively portraying and communicating complex information, cutting across various disciplines like life sciences, businesses, journalism, etc. However, creating visualizations demands significant time and effort, which often creates a bottleneck for domain experts [19]; and creating *effective* visualizations requires knowledge about visualization design principles and best practices. However, there has been little work on systematically judging the quality of visualizations used and created by non-experts in the field. While authors like Tufte and Few [17], [48] have critiqued visualization examples and offered guidelines for better design, very few academic attempts exist for classifying types of design problems and judging their consequences, especially when domain experts design visualizations.

To fill this gap, we performed a systematic analysis of how climate scientists use and design visualizations for reflecting upon the causes and effects of design problems. The data that we analyzed comprises of a series of semi-structured interviews with climate scientists, about visualizations collected from research papers and presentations.

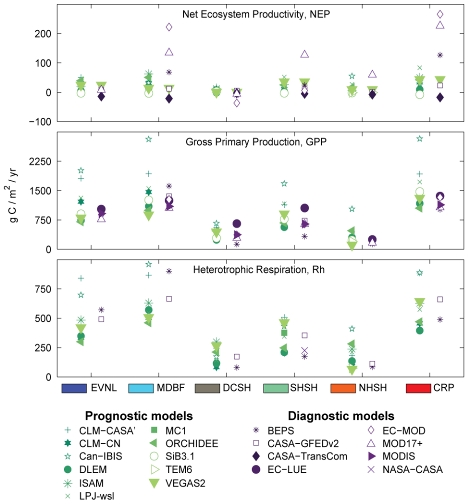
The benefits of such an exploratory study are two-fold. First, it allows domain scientists to better critique their visualization designs and incorporate that knowledge into building more effective visual representations. Second, reflecting on the analysis of visualization design problems is an opportunity for the visualization community to investigate how the state-of-the-art in visualization meets the analysts' needs, and introspect how design principles can be better applied to suit the evolving complexity of data presentation and communication. We know of only two precedents in the visualization community that aimed at examining the cross-domain synergistic capabilities [16], [42]. In our work we take one step farther by judging how well domain experts and visualization researchers agree on design problems, based on which we redesigned some of their existing visualizations and judged the effectiveness of the solutions from their feedback.

In our study, we focus on an important area of climate data modeling, that is comparison of terrestrial biospheric models. Intercomparison of output from the multiple terrestrial biospheric models is a key research area that helps understand the spatial and temporal patterns of carbon sources and sinks (e.g., photosynthesis and respiration) which serve as important feedback into global climate change indicators. Typical visualization usage and design by climate scientists for such comparisons is shown in Fig. 1. Fig. 1a) shows the use of scatter plot for comparing model output variables for multiple models. Fig. 1b) shows the use of multiple maps for analyzing similarity of models over different spatial regions. The challenges for concise visual representation in these cases is non-trivial because of the underlying diversity and complexity of the data. The aim of this exploratory study was to find, for these complex analysis tasks, what are some recurring design problems. While we also found some examples of optimal visualization designs, our goal in this work was not to comment on the general state-of-the-art in visualization practice in climate science, but to focus on the problematic visualization designs and devise a model for describing those problems.

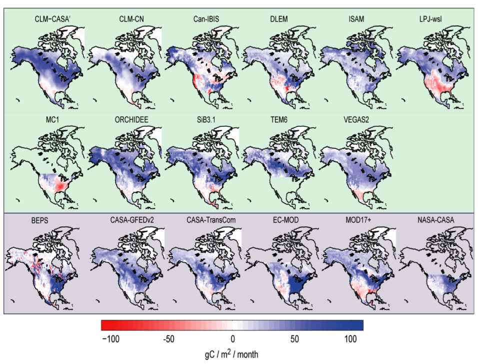
Our high-level analysis questions for understanding visualization design problems were: do the chart types address the goals of visual representation? Are there design flaws specific to those chart types or are there generalizable problems cross-cutting chart types? Does the literature of visualization design offer solutions to those problems? We collected a representative sample of 15 research papers from our collaborators that used visualizations for comparing terrestrial biospheric models. The over-arching goal was to create a taxonomy that systematically answers the aforementioned analysis questions. In summary, the contributions of this paper are three-fold:

1) A systematic classification of visualization design problems in the climate science domain resulting in a descriptive taxonomy

- *A. Dasgupta is with New York University and DataONE. E-mail: adasgupt@nyu.edu*
- *J. Poco, E. Bertini, and C. Silva are with New York University. E-mail: {jpocom, enrico.bertini, csilva}@nyu.edu*
- *Y. Wei and R. Cook are with Oak Ridge National Laboratory. E-mail: {weiy, cookrb}@ornl.gov*



(a) **Design problems in a stacked scatter plot** stemming from over plotting and use of many different symbols.



(b) **Design problems in the multiple maps** stemming from poor encoding of relative similarity.

Fig. 1: **Illustrating two common visualization use case scenarios and their associated visualization design problems**, for comparing terrestrial biospheric models (figures adapted from [25]) In a) stacked scatter plots with multiple visual symbols lead to an ineffective visual search for models and inefficient comparison of spread among their output variables. In b) outliers indicated by red regions are clearly visible but similarity analysis among 17 different maps is difficult without any encoding that reflects relative similarity among the models.

of types, causes, and implications of such problems.

2) Application of the taxonomy for: i) identifying problem consequences and their trade-offs, ii) a detailed analysis of causes of matches and mismatches about design problems between visualization experts and climate scientists, and iii) feedback on redesigned solutions for a representative sample of problem instances.
3) Summary and analysis of the findings for enabling scientists in designing improved visualizations, and for reflecting on the gaps and opportunities for visualization research.

2 Related Work

We discuss the related work with respect to the existing studies on visualization usage and design, and the relevant theoretical models that have been proposed for characterizing the visualization process and design.

2.1 Studies on Usage and Design of Visualization

In recent times, there has been some progress towards studying how people outside the visualization community use, design, and reason about visualizations. This body of research is critical in diversifying the field of visualization by gaining insight into the potential roadblocks that people from different communities face, while designing perceptually effective visualizations and subsequently using the interpretations for their benefit.

To this end, Grammel et al. attempted to understand how people who are unfamiliar with visual data analysis, i.e. InfoVis novices construct visualizations and the potential roadblocks in doing so [22]. They found that constructing effective visual mappings was the most significant roadblock, which was consistent with the findings of Heer et al. [23]. Somewhat related to this, researchers [30] studied the problems low-literacy users face while retrieving information online, and how interactive visualization can help in the process. Researchers have also explored the problem [12] from the point-of-view of existing visual analytics tools: they found that while conducting investigative analysis, several roadblocks exist in understanding, choosing, using, and reading views properly. Some of the studies also focus on collaborative environments. Walny et al. [52] studied how pen and touch interactions on interactive whiteboards facilitate reasoning and understanding of visualizations. Most of these studies focus on the usage patterns of visualizations for novice users. In our work, the focus is on domain experts who have compiled the data to specifically address their research questions but who do not have detailed expertise to design the most effective visualizations. There is a lack of studies that look into the types of problems that arise when domain experts design visualizations.

2.2 Models Characterizing Visualization Process and Design

Among the many theoretical models that exist in visualization, the ones that are relevant for our work fall into two broad categories: i) models which characterize the visualization process, starting from data transformation to human perception and cognition, and ii) models that capture the different aspects of a visualization design and its implications, especially from an end user's perspective. One of the earliest instances of a process model, was the data-state reference model proposed by Card [9], which was later extended by Chi's pipeline model [11] for representing different data transformation stages and the intervening operations. This was further extended by Ware [53] whose model focused more on the visual representation and its perceptual implications. For the visualization design models, we find instances where researchers have studied the use and creation of visualizations from a designer's point of view [38] or as the product of a collaboration between designers and end users [54]. Heer et al. [23] proposed a model for providing guidelines to novice users on the encoding type used. We propose a taxonomy model, which is similar in its characteristics with the visual uncertainty model [14] that combines both visualization process and design in one holistic framework. The functionality of our model is similar in spirit with the work of Walny et al. [51], who generated a taxonomy by studying how visualizations on white boards are typically produced, what their purposes are and how people from outside the visualization community use visual thinking for solving their problems.

3 How Climate Scientists Use Visualization

Our collaborators are climate scientists specializing in Terrestrial Biosphere Models TBMs). As visualization researchers we have worked closely with them in the Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) as part of the DataONE Scientific Exploration, Visualization, and Analysis (EVA) Working Group. In this section we describe briefly what domain-specific problems they aim to address by designing visualizations.

TBMs simulate terrestrial ecosystem processes and the terrestrial-atmosphere carbon exchange in relation to prescribed boundary conditions: vegetation cover, soil properties, climate, etc. They have become

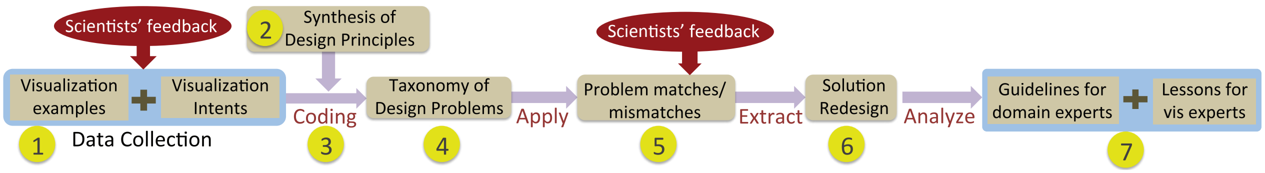


Fig. 2: **Workflow for our qualitative study** comprising of seven different stages as annotated in yellow. The workflow highlights the tight interaction with climate scientists that led to the taxonomy of design problems, its subsequent refinement and application for finding solutions. The problems and solutions were further analyzed and reflected upon for providing general design guidelines to domain experts and highlighting lessons learnt for visualization researchers.

an integral tool for extrapolating local observations and understanding to much larger terrestrial regions, as well as for testing hypotheses about how ecosystems will respond to changes in climate and nutrient availability [25]. TBMs use complex analysis scenarios and generate diverse and large volumes of multidimensional data. Visualization thus constitutes an integral component of most model output analysis processes not only for understanding and representing the data, but also for subsequent dissemination of the scientific knowledge. Visual representations in the form of images and charts used in academic publications and presentations play a critical role in communicating the scientific findings to a broader community.

A key approach for environmental modeling is to use multiple models as a way to gain confidence in the results and bound understanding. Therefore, inter-comparison of a suite of terrestrial biospheric models over space, time, and different land cover types is an important research area. But the volume and complexity of model outputs present many challenges for analysis and visualization.

To gain additional confidence in model output, researchers compare observations with model simulations in a benchmarking activity. Furthermore, modelers want to know which models are similar, and why, when, and where they are similar. Linking model structures with model output can be used to understand why models are different from benchmarks and each other. Visualization plays a crucial role in all of these steps for understanding model characteristics and visually representing the scientific findings.

4 Methodology

The goal of our study was to diagnose design problems in visualizations created by climate scientists. While in course of our research we also discovered visualization examples which adhered to the best practices (Fig. 3), our aim here was to focus exclusively on the causes and consequences of the problems. We followed a descriptive approach where we could provide useful guidelines for climate scientists and discover challenges for visualization researchers. To achieve this purpose we adopted a qualitative methodology featuring in-depth analysis of climate visualization examples, generation of descriptive classifications schemes, as well as multiple interviews, workshops, surveys, and focus groups. In the following section, we describe our methodology in details.

4.1 Participants

In the course of our project we collaborated with 20 climate scientists, with 5 of them being direct collaborators from the Multi-Scale Synthesis and Terrestrial Model Inter-comparison Project (MsTMIP). Most of them have over ten years of experience in climate modeling. The overall goal of MsTMIP is to provide feedback to the terrestrial biospheric modeling community to improve the diagnosis and attribution of carbon sources and sinks across regional and global scales. Our group of collaborators were mostly climate modelers working as part of the Exploration, Visualization, and Analysis (EVA) working group under the DataONE initiative; and they spanned across different national labs and universities within the United States.

With our direct collaborators, we interacted over a six month period via semi-structured interviews, which were both in-person and through teleconferences, and three workshops where we exchanged knowledge about our respective domains and conducted interviews. With the indirect collaborators, we attended their presentations at workshops, took note of their visualization design and received their feedback on our findings through teleconferences.

4.2 Evaluators

The group of evaluators (henceforth referred to as *we*) who were involved in data collection, analysis, and synthesis; comprised of four data visualization experts (all co-authors of the paper): one doctoral student, one research scientist and two faculty members. All the evaluators have at least four (and for two of them more than ten) years of research and practical experience in visualization. The coding part of our work loosely follows the tradition of expert-based evaluation of user interfaces like *heuristic evaluation*, where it has been demonstrated that a small number of experts can reliably detect most of the problems [36].

Also, following-up on the same tradition, rather than relying exclusively on the personal judgment of the evaluators, we created guidelines and support material to inform and guide their work. Since a single established set of visualization heuristics does not exist yet, we decided to: i) review the few initial attempts to create visualization heuristics we found in the literature ([18], [55]). and ii) create our own synthesis of visualization principles drawn from the visualization design and research literature. We provide more details about the synthesis of visualization principles in Section 4.5.

4.3 Workflow and Goals

The workflow we adopted for our exploratory study is outlined in Fig. 2 and is characterized by seven distinct stages we performed to gather the necessary data and perform our research.

In (1) **Data collection** we interacted with our collaborators, through in-person meetings and teleconferences to collect *visualization examples* and *intents* that are representative of the typical tasks performed by the climate scientists. In (2) **Synthesis of design principles** we reviewed the existing literature on visualization design principles and organized them into a reference list we used to inform and guide our critique of visualization examples. In (3) **Coding** we used the reference list to manually annotate the collected examples and generate descriptive codes that captured potential design problems. In (4) **Taxonomy generation** we systematically and iteratively refined and organized the codes to generate a design problem taxonomy. In (5) **Problem matches and mismatches** we discussed representative examples of the collected design problems with the climate scientists to gather instances of diverging opinion between visualization experts and the domain experts. This phase allowed us to refine the taxonomy and to build a much richer view on how visualization design principles can and should be instantiated in practice. In (6) **Solution Redesign**, based on the suggestion of our group of collaborators, we extended our analysis to include discussions of solutions. We redesigned some

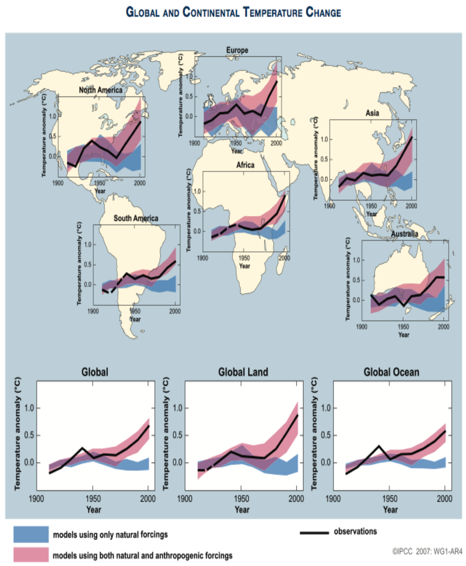


Fig. 3: **One of the few examples of optimal visualization design from our collected examples [45].** where the intent was effectively captured and communicated. Here small multiple line charts are used effectively in conjunction with maps for showing region-wise temperature variation of two classes of climate models.

selected examples and gathered additional feedback from the scientists. Finally, in (7) **Guidelines and lessons learned** we reflected on the output generated by our analysis and interactions and came up with a set of general guidelines, pitfalls and lessons learned.

For data collection, the qualitative analysis of the examples, and the following derivation of a design problems taxonomy; we followed the grounded theory methodology [10]. Grounded theory is a systematic methodology used in social sciences to derive classification schemes from the analysis of large quantities of qualitative data and guides the researchers through iterative phases of data collection, code generation, and their organization into descriptive categories. This approach recently gained some momentum in the visualization research community and it has been successfully adopted in a growing number of studies to analyze visualization artifacts and their use in real-world scenarios [24], [26], [43], [47], [51]. This methodology allowed us to critique and analyze visualizations created by scientists without any pre-formed hypothesis, thereby allowing the data to dictate the taxonomy that emerged. The aspects we imbibed from grounded theory were alternating between data collection and analysis, refining the conceptual relationships within the data, and subsequent generation of a theoretical paradigm for structuring our findings.

4.4 Data Collection

During our data collection phase we first interacted with the climate scientists to generate a representative sample of visualizations to be used for our analysis.

Visualization Examples. The visualization examples were collected in consultation with our collaborators, from a set of 20 presentations in two workshops, 15 research papers from the climate science domain and four interview sessions. Our effort was to ensure that the collected sample represents the state-of-the-art in visualizations used for comparing climate model data. From these sources we generated a total of 105 images which we used as the basis for our study. Given the high experience level of our collaborators we were confident from our interactions, that these images constituted a representative sample for our study. Among the examples we collected, 80% of the visualizations comprised of geographical maps, scatter plots and variants of line charts. The remaining 20% was a heterogeneous set of examples which could not be organized into any consistent group or description. For this reason we decided to exclude them from the analysis and focused on 80% consistent group of images, which comprised of 40 line charts, 30 geographical maps and 15 scatter plots.

Visualization Intents. In a preliminary coding pass we realized it was hard to judge the collected examples without first knowing the visualization intents. Rather then evaluating the collected example exclusively under an abstract set of principles, we preferred to ground our analysis on the following main questions: **Q1**)"*Does the chart represent the intent correctly?*" **Q2**)"*Does the chart convey the main message efficiently and effectively [34]?*" These high-level questions guided the latter stages of the evaluation pipeline, such as the synthesis of design principles and coding.

As we felt we could not derive the intent and message of every single example without the help of our domain experts, we conducted surveys and teleconferences to come up with a more reliable set of intents. To achieve this purpose we performed the following steps: i) two investigators formed an initial idea of the intents from the descriptions in the collected material and collected them in a document containing pairs of images and intents; ii) we distributed a survey with the intents to a group of 2 scientists asking them to mark whether they agreed or disagreed with the stated intent and to add their own version of the intent where necessary; iii) for those cases where the scientists disagreed (with us or between them) we performed an interview session for further clarification and collected the right intents.

This phase was crucial in understanding the motivation and context behind the creation of the visualizations. In some cases, we initially misread the intent, which was subsequently clarified in these interview sessions. For example, as shown in Fig. 1(a), we initially deduced that the initial intent was to identify the models which showed high or low values for the carbon flux output variables NEP, GPP, Rh (organized vertically in the three stacked panels). But in an interview session it was clarified that the primary intent is to show the overall variability of prognostic (green) and diagnostic (purple) models and also to show the which models belonged which eco-regions for the different variables.

An analysis of the collected intents showed that in the context of the climate model inter comparison, the general intents of scientists were mainly to identify similar models and compare their spatial or temporal variability over different granularity of space and time. In previous work, we had designed an interactive interface that helped scientists realize these intents through a rich exploratory visualization tool [37]. In this work, we judged the design problems of the static visualizations with respect to these intents.

4.5 Synthesis of Design principles

The basis for the judgment of design problems in the examples we collected was the rich literature of visualization design principles that have evolved over time. While we observed that there is no single conceptual framework which can be applied for such analysis, various theoretical principles, starting from Bertin's seminal work, to the most recent research on information graphics, have guided visualization design over the years. These design principles come from different areas of visualization research, and address different but often complementary issues such as: principles for optimal visual design, criteria for design based on data type, matching the visual properties to best support human perception and cognition, and also more recently, how best to communicate the message by properly structuring the information. All these categories subsume the two general questions **Q1** and **Q2** on which we eventually based our judgment of the visualizations. Before starting the coding phase, we worked on collecting design principles and synthesizing them in a way

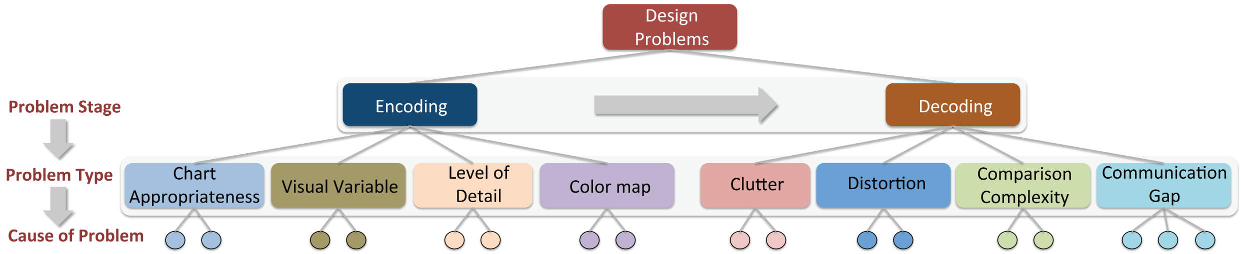


Fig. 4: **Different levels in the design problem taxonomy**. Problems are categorized according to the stage in the visualization pipeline, the type and the cause. The leaf nodes for problem cause are shown in Figures 5 and 7. The most frequently occurring problems were the visual variable problem (37% cases), communication gap (30% cases), and clutter (29% cases); followed by color map choice (28% cases) and distortion (20% cases). Some of the less frequent problems were level of detail (17% cases), comparison complexity (15% cases), and chart appropriateness (13% cases).

that codes based on those principles could be used for the study. Note that these principles were applied only in the context of the scientists' intents for any visualization. Here, we provide a brief summary of the main sources we used to inform our research.

Optimality of design: From Tufte's seminal work [48] and Tukey's research on statistical graphics [49] we adopted the principles of graphical excellence and integrity. Graphical excellence encompasses a number of criteria for design by maximizing the *the data/ink ratio* (i.e., information and data density) and avoiding accessory elements and embellishments. Graphical integrity refers to a truthful representation of the data (*related to Q1*) for avoiding potential misinterpretation due to scaling issues or distortion.

Criteria for design based on data type: From the early work of Mackinlay [34] we adopted the criteria of *expressiveness* and *effectiveness* to qualify visualizations in terms of the encoding parameters and degree of salience of visual attributes. From Bertin's seminal work [3] we derived principles of effective visual encoding of data features into visual variables. We also considered the work of Card et al. [8] and MacEachren et al. [33] which revise and extend the early of work of Bertin respectively in information visualization and geo-visualization. These enabled us to judge the appropriateness of the visualization parameters (*related to Q1*) based on the data type they represented.

Perceptual Implications of Design: Bertin's seminal work on visual variables such as position, size, shape, color, orientation, and texture; formed the most important basis of our judgment of the appropriateness of the design parameters. Bertin focused on defining the possible visual variables and reflected on their perceptual implications for visualization design. The work of Bertin was extended by Cleveland and McGill [13] and Ware [53], who provided much needed empirical evidence of the perceptual effectiveness of visual variables through controlled user studies. Together with Bertin's work, such experimental research form the core of the science of visualization. *For addressing Q2*, we utilized the following concepts inspired from these threads of research: the importance of visual encoding keeping pre-attentive processing in mind, ranking of visual variables based on different tasks, perceptual effects and properties of color, importance of spatial organization of visualization design, etc.

Design for more effective visual communication: Finally we also considered recent approaches from data visualization practitioners. These mainly address the concern of how visualization should not only support exploration and analysis, but should also be able to visually communicate the data. *For addressing Q2*, we utilized design guidelines from Stephen Few's book "Show Me the Numbers" [17] and from "The Functional Art" [7] a data narrative-oriented book written by data journalism expert Alberto Cairo.

4.6 Coding

In the coding phase we analyzed all the image instances for potential problems with respect to the visualization intents that were collected in the initial phase. The codes we used for describing the problems were based on our synthesis of design principles. For each example we collected codes describing design problems and relevant issues. For instance, the scatter plot example shown in Fig. 1(a), was coded with: *clutter, chart selection, and color map*. Wherever more clarity was needed, we resolved our doubts by asking further questions to the scientists. We met at regular intervals to share, compare, merge and refine the set of codes and after several iterations we reached a stable set. Halfway into this process, during our discussions we realized design problems sometime have non-trivial solutions. For this reason we started collecting, together with problems, descriptions of design consequences and their trade-offs, which are presented in Sections 5.3 and 8.

4.7 Generating a Taxonomy of Design Problems

After collecting the codes we moved to the *axial and selective coding* [10] phase where we merged, grouped and structured the codes into a full taxonomy. During this phase we went through several refinements by having one of the investigators mainly working on the classification scheme and another investigator testing the scheme with the library of examples, while discussing inconsistencies collaboratively. We stopped the process when we felt that we reached a stable and satisfactory description of all the problems.

One of the issues during this phase was to choose an agreed upon level of abstraction for categorizing the design problems. For this we used a bottom-up approach by analyzing which problems are similar in terms of: which stage of the visualization process they were introduced and what effect they had on the visual representation. Accordingly we came up with a three-level taxonomy, that helped us categorize the causes and implications of the problems.

5 Taxonomy of Design Problems

The taxonomy we have derived is a classification scheme where a visualization example can be associated with multiple design problems that are described by different nodes of the taxonomy tree. For deciding a classification that captures the causes and effects of design problems, we took inspiration from the taxonomy of visual uncertainty [14] based on the traditional information visualization pipeline [11]. The latter can be regarded as a communication channel [44] and thought of being composed of two distinct phases: encoding, that is associated with mapping the data on to the screen-space; and decoding, that is associated with perceptual and cognitive processes on the user's side. The classification scheme, based on encoding and decoding stages as the first level, enables us to systematically analyze different dimensions of the design problem. The levels are described below:

i) Problem Stage: The first level decides whether it is in the encoding or the decoding stage that a design problem is found. Encoding deals with problems that mostly depend on the choices the designer makes when deciding how to transform data features to visual features.

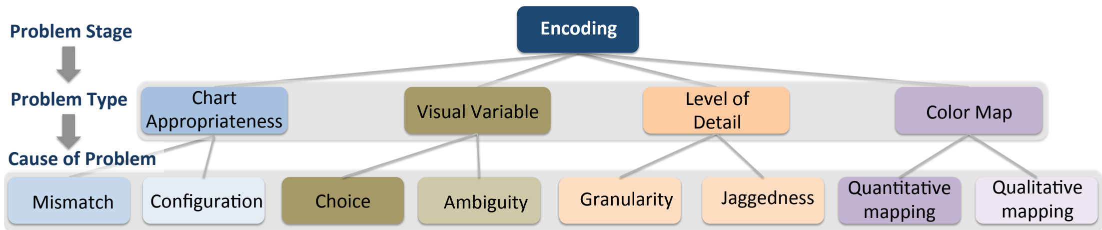


Fig. 5: **Design problems found at the encoding stage of the visualization pipeline** reflected how well the input parameters such as chart type, visual variables, level-of-detail and color map were chosen by the scientists. Darker color at the lowest level indicates higher frequency of a particular cause (e.g. Choice) within a problem type (e.g. Visual Variable problem).

Problems on the encoding side enable us to address if the encoding strategy (e.g., chart type, visual variables, color map) is based on visualization best practices. Decoding captures design problems that go beyond the specific scope of visual encoding and the deliberate encoding choices a designer makes. These problems may be due to the effect of the limited screen resolution, perceptual implications of the visual parameters and auxiliary elements (grids, legends, annotations). They mainly affect the effectiveness with which the user can decode the presented information.

ii) Problem Type: An encoding problem or a decoding problem can be classified into multiple types, depending on the characteristic of the problem. The type level separates the encoding and decoding problems into different classes, which encapsulate the low-level causes of these problems. The classes belonging to the encoding side reveal the gaps in fulfilling the necessary conditions for a good encoding. For example, it could reveal if the chart type and visual mapping were appropriate. Fulfilling the necessary conditions for a good encoding might not always be sufficient for a visualization to be useful. The classes belonging to the decoding side reveal if the necessary conditions were also sufficient, by revealing if there was too much clutter, or if the there was distortion of information or the visual complexity was too high.

iii) Cause of problem: This is the level at which leaf nodes of the taxonomy tell us the precise cause of the problem. These causes reveal the low-level details of the problem types. For example, from this level we know the causes of a color map problem or the causes of a distortion problem.

5.1 Encoding Problems

The encoding side of the taxonomy helps us ask questions such as: "*Do the encoding parameters such as chart selection reflect the intent?*", "*Can the visual mapping and color map choice be improved?*", "*Given the intent, is the data shown at an appropriate level of detail?*" In this section we describe the causes of design problems during the encoding stages (Fig. 5).

Chart appropriateness: The first design decision that the scientists have to make for reflecting their intent, is which chart type to use. The chart appropriateness issue deals with whether the charts selected by the scientists appropriately reflected their intent. For judging this problem, we analyzed if any inherent limitation of a chart type, or the resulting configuration of the visual representation interfered with the intent. The two causes for the appropriateness problem were as follows:

Mismatch Mismatch captured cases in which the chosen visual representation was not the best option for conveying the intent due to its inherent limitations. This issue was observed mainly in scatter plots.

For example, one of the intents in the scatter plot in Fig. 1(a) was to find which models belonged to which eco-regions. The author attempted to convey this intents through a variant of a traditional scatter plot, where the *X*-axis represents a categorical attribute (the eco-regions) rather then a numeric value, as is usually expected. This unexpected configuration created confusion among the scientists and made the chart difficult to interpret. A scatter plot is unable to clearly show the models that belong to a particular eco-region due to over plotting. An additional problem is the use of the many different symbols for representing each model, which leads to an inefficient conjunctive visual search. In Section 7 we discuss a solution to this problem.

Configuration: Configuration problem deals with the arrangement of multiple charts in a common canvas. We found several examples where scatter plots, line charts and maps were stacked together for comparing climate model behavior. This can be observed in Fig. 6. The intent in this case was to compare the temporal variation of annual cycles. But the horizontal stacking of the line charts, where time was represented on the X-axes, made it difficult to compare the Y-axis values. This problem was also observed in Fig. 1(b) where multiple maps where stacked together without any ordering based on similarity.

Visual Variable: Visual variable problem captures cases in which the designers made poor choices in the mapping data attributes to visual variables. This is one of the most important design decisions in the visualization process [3], [8]. Since the number of data attributes generally outnumbers the number of visual variables (position, shape, size, color, orientation, etc.) by far, effective utilization of the latter is crucial in designing effective visualizations. The two causes for the visual variable problem were as follows:

Choice: The *choice* of visual variables was one of the leading causes of problems we found in our collection. While the choice affects all the subsequent visualization stages of human perception and cognition, here we focus on how visual mapping can "*above all show the data*" as suggested by Tufte's principle of graphical excellence [48]. The different classes of problems due to choice of visual variables were: representation of discrete data attributes in scatter plots and line charts using a combination of visual variables, and use of color as a quantitative channel for comparing averages and differences on geographical maps.

For example in the scatter plot in Fig. 1(a), one of the main problem was the representation of discrete data attributes, i.e. climate models, using visual variables such as shape, texture, and orientation concurrently. The use of multiple symbols causes conjunctive visual search [53] which is inefficient and not a good use of the pre-attentive capabilities of the human vision system. Moreover, combination of shape (different symbols), texture (filled and unfilled shapes), and orientation (triangles pointing in different directions), do not adhere to the rule of integral and separable visual dimensions [53]. In line charts, a recurring problem was the use of dots, solid lines, and dotted lines which would create difficulty in recognizing and tracing the different items.

Ambiguity: Another category of problems with visual variables, was ambiguity, where the use of visual variables was inconsistent: either different visual variables were used to represent the same data attribute, or the same visual variable was used to represent different data attributes. While the choice of visual variables reflects how well the latter reflects the different data properties, ambiguity reflects if even after a correct choice was made, there were additional inconsis-

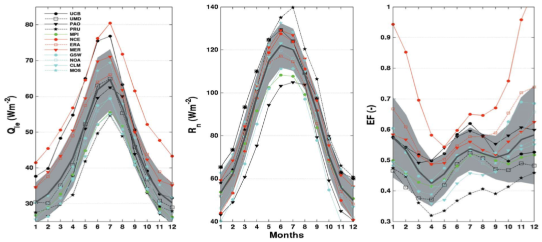


Fig. 6: **Problems due to clutter: color mixing, visual variable problem: ambiguity, and chart appropriateness: configuration** The intent behind this multiple line chart figure [27] is to enable readers to analyze the variation of annual cycles over time in terms of the ensemble mean, the standard deviation and the individual values. It is difficult to compare temporal trends due to the side-by-side placement. Color mixing among the lines causes clutter.

tencies. For example, in the line chart example in Fig. 6, there are only a few different colors used for representing the different categories. It almost seems there is a relationship among them, although none is explicitly mentioned in the text. The same problem was observed in maps, where a white or grey was used to represent two different factors: absence of data and lack of correlation among values that are represented. As evident ambiguity can lead to misinterpretation of the data where a relationship can be deduced even if there is none and if there are multiple relationships, only one of them might have been conveyed.

Level-of-Detail: For the visual encoding of data attributes, it is important to choose an appropriate level-of-detail that would not only preserve the fidelity of the data as much as possible, but also effectively communicate the intent. The two causes of the level-of-detail problem were as follows:

Granularity: This problem was observed in cases where either a coarser or finer granularity level could better reflect the intent. For example, a recurring issue with the maps was that pixel-based representation was used for mapping quantitative variables and enabling model comparison. While this led to high fidelity, comparison across multiple maps was difficult because of the low-level details that readers had to classify and compare. As shown in Fig. 1(b), this would be inefficient as the lowest level of granularity would not facilitate a high-level overview of the salient patterns. It would instead cause a sequential search for finding similarities and dissimilarities among the maps.

Jaggedness: A special case of the level-of-detail problem was the jaggedness of the lines in case of line charts representing time series. The salient peaks and crests in the time series were occluded because of the jaggedness. The main source of the problem was the tendency of the scientists to plot daily or monthly data, even when the intent was to show the annual variability of any given entity. In those cases smoothing could be used by computing an average and that would highlight the main trends. In the redesign section (Section 7), we present a line chart example (Fig. 12) that shows these jagged patterns.

Color Map: Choosing an appropriate color map is essential for the effective display and analysis of data. Based on fundamental human perceptual principles and the type of data being displayed (sequential, diverging, or categorical), there are formal and systematic ways to make an appropriate color choice based on the task at hand. Color maps for quantitative attributes are important for making an accurate judgment, while those for qualitative attributes are important for distinguishing among different categories efficiently. Since the implications of these two types of color maps are different, we treated these two problems separately:

Quantitative mapping: For quantitative color maps, the rainbow color map was used in most cases. As extensively documented in the visualization literature [5], [40], the lack of perceptual ordering

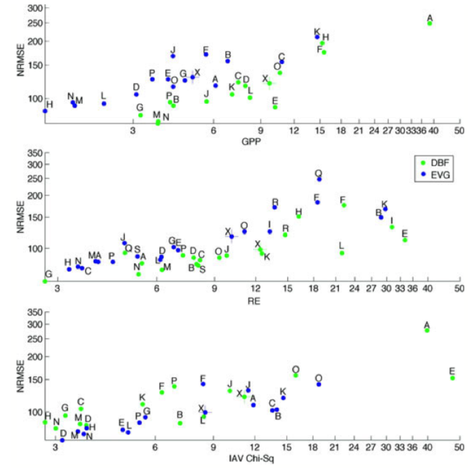


Fig. 8: **Problems due to chart appropriateness: mismatch, distortion: scale inconsistency, comparison complexity: superposition overload, communication gap: legend, annotation.** The intent in this figure [29] was to compare the errors of two plant functional types (DBF, EVG), color-coded in green and blue, and quantified in two ways: NRMSE and Chi-Sq. The X-axis in all three scatter plots represent the Ch-Sq statistic. It is hard to separate the patterns between the two functional types: a regression line and annotation of key trends would more clearly communicate the message.

and isoluminance in case of rainbow color map can cause inaccurate interpretation of the data. It has also been shown in case of scientific data, the crucial role that a perceptually motivated color map plays, for example in case of diagnosing heart conditions [4]. From the examples we collected, geographical maps suffered from this problem the most. In many cases we found that scientists are only interested in recognizing the extreme values, and the colors red and blue are associated with the semantics of temperature: red signifying hot regions and blue signifying cooler regions. But in many of those cases, all the hues of the rainbow are used for encoding the data. A divergent color map with only a luminance variation [6] would be appropriate in that case.

Qualitative mapping: For qualitative color maps, the problem was when representing discrete variables with color (Fig. 12). If the hues are not separable enough, visual search for the variables would be inefficient. We found that it is a common requirement for the climate scientists to represent more than 10 discrete variables (in the form of regions or climate models) in a single chart. If color is the chosen visual variable, the choice of hue then becomes critical. The Tableau 20 color palette can be used in that case. ColorBrewer [6] only offers about 11 distinct colors.

5.2 Decoding Problems

Once the encoding parameters are chosen in the design process, to the judge the quality of the visualization, we have to judge its perceptual implications. Analysis of problems at the decoding stages of the visualization, that is the perception and cognition stages, enables us to evaluate a visual representation by asking questions such as: *"Is it perceptually confusing?" "Does it represent the patterns without distorting it or being too complex?" "Does it emphasize the intended message clearly enough?"* In this section we describe the problems caused during these stages of the pipeline (Fig. 7).

Clutter: We adopt the definition of clutter which relates the degradation of a display with the number, representation, and organization of items [41]. Many of the visualization examples, across maps, line charts, and scatter plots were cluttered and there were different reasons for clutter. The two causes of clutter were as follows:

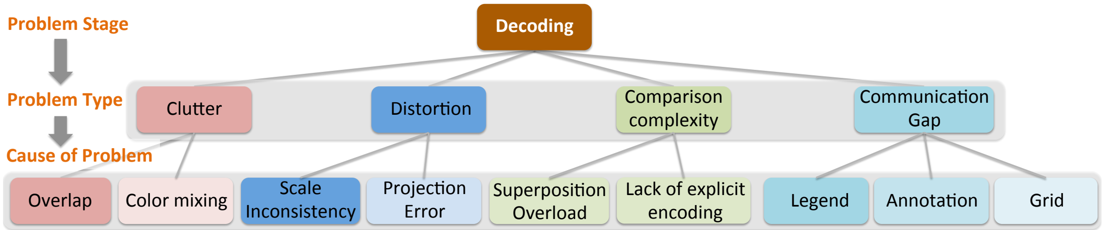


Fig. 7: **Design problems found at the decoding stage of the visualization pipeline** reflected how the communication of the intended message was affected by the design choices. Darker color at the lowest level indicates higher frequency of a particular cause (e.g. Overlap) within a problem type (e.g. Clutter).

Color mixing: Color mixing (Fig. 6) was one of the causes for clutter. This is different from the color map problem, because color mixing mainly occurred between the the chart elements and the background or among the different sets of symbols. For this case, the color map could have been appropriate, but there needed to be another extra degree of caution for avoiding color mixing. Color mixing was observed mainly in maps and line charts. For example in the line chart in Fig. 6, the color mixing between the grey band, the grey mean line and the chosen colors for the other lines cause clutter.

Overlap: Overlap encompasses over-plotting of data items in scatter plots and maps, and large number of crossings in line charts. In some instances of line charts, the thickness of individual lines made it difficult to identify and trace the paths of individual lines. Over plotting of different visual variables on a scatter plot (Fig. 1(a)) made it difficult to recognize and visually search for the individual data points. While over plotting and overlap are artifacts of the representation, and are often unavoidable, the key question here is whether these artifacts interfered with the intent. For example, in case of the scatter plot example, over plotting interferes with the intent, as identifying each model is one of the intents of the author.

Distortion: Distortion of the data in a visual representation is a serious problem that can either lead to potential misinterpretation or an inaccurate perception of the data, especially when quantitative attributes are involved. The causes for distortion are as follows:

Scale inconsistency: Choosing different scales for the same variable leads to inconsistent representation of the patterns. This is a decoding problem, because the chosen scale is appropriate for a given variable, but when multiple variables are involved, lack of attention to consistency can mislead readers.

We found this to be a problem in some geographical maps, where a single rainbow color map was used to represent data which were at different scales, leading to misrepresentation of the patterns. In some scatter plots and line charts (Fig. 8) we found that the scale of one or more of the attributes are different from others. In this case the X axis represents the chi-squared statistic and in the topmost line chart, the tick placement is different from the others, signifying a different log-scale for encoding the data than others.

Projection error: This problem is typical of maps, and due to the inherent mapping from a 3D sphere to a 2D surface, we observed projection error in some of the map examples. We observed that while some error is unavoidable, use of better projection techniques could reduce the amount of the error. For example, an equal-area projection will be more appropriate in displaying area-sensitive data like fire-burnt area.

Comparison complexity: The main goal in TBM domain being model inter-comparison, the primary intent behind most visualizations was to facilitate comparison at different levels. We found that the comparison complexity in terms of number of data points per chart, or number of charts per views, or their placement, led to some design problems. We take inspiration from Gleckler et al.'s taxonomy of visual comparison methods [20] for categorizing these problems. We classified this as a decoding problem because even a correct encoding choice could lead to comparison complexity and influence the communication of the main message. The two causes of comparison complexity were as follows:

Superposition overload: This category deals with the case where the number of entities in a chart are far too many for facilitating an effective comparison. This issue was observed mainly in line charts and scatter plots. As opposed to a small multiple display, a large single [50] was often required by climate scientists for comparing models to observations, or comparing individual values to ensemble mean. In some of those cases, superposition overload of too many elements led to clutter (Fig. 12) and in some case, although clutter was not caused, superposition overload interfered with the intent (Fig. 8). In the first case, the drawbacks of the superposition are obvious and we discuss a small multiple solution to this problem in Section 7.3. We found small multiples being used by scientists in case of maps, but we found only one line chart example where a small multiple was used for reducing overloading.

Lack of explicit encoding: This case with the issue where explicit encoding of relationship among the compared entities would have led to better design. It has been observed that small multiples are important while visualizing multiple variables [48] but care should be taken to position and sequence the individual charts appropriately so that visual search is optimized [17]. For example, in case of the multiple maps (Figure 1(b)), the intent here is to deduce the degree of similarity among the different models. However a random arrangement does not immediately show how much similar, the maps are. It requires visual inspection almost on a pixel-by-pixel basis for judging similarity. In that case, extracting some summary statistic about the degree of similarity and using that for positioning the maps seemed to be a good solution. We will discuss this solution and its evaluation in the case study section.

Communication Gap: In many of the visualization examples we collected, we found problems with factors which do not directly interfere with the intent, but might create problems with communication of patterns. These are auxiliary information about charts, which were categorized as follows:

Grids: Grids can be used for chunking the important pieces of information, which might not be intuitive immediately. Human brains are good at picking out patterns. Often, fairly small changes to a graphic layout that strengthen the appearance of grouping or other types of patterns will add to the ability of the graphic to deliver an instant impression or overview of the message being communicated. While unnecessary grid lines must be avoided in keeping with the idea of minimizing non-data ink proposed by Tufte [48], judicious use of grids help in capturing the reader's attention to the salient portions of the chart. For example, use of column-wise grid lines in Fig. 1(a) could separate the ecoregion-wise patterns for the different models. In our collected examples, we found two scatter plots were grids were used to chunk the information space, for denoting groups of data points belonging to a model or a year.

<table>
<tr><th>Causes of Problem</th><th>Consequence</th></tr>
<tr><td>Visual variable problem: ambiguity
Distortion: scale inconsistency</td><td>Misinterpretation</td></tr>
<tr><td>Distortion: projection error
Distortion: scale inconsistency
Color map: quantitative mapping</td><td>Inaccuracy</td></tr>
<tr><td>Chart appropriateness: mismatch
Chart appropriateness: configuration
Visual variable problem: choice
Level-of-detail: jaggedness</td><td>Lack of expressiveness</td></tr>
<tr><td>Visual variable problem: choice
Level-of-detail: granularity
Color map choice: qualitative mapping
Clutter: color mixing
Clutter: overlap
Comparison complexity: superposition overload
Comparison complexity: lack of explicit encoding
Communication gap: grids</td><td>Inefficiency</td></tr>
<tr><td>Comparison complexity: lack of explicit encoding
Communication gap: grids, legend
Communication gap: annotation</td><td>Lack of emphasis</td></tr>
</table>

TABLE 1: **Connecting design problems to problem consequences sorted by severity**. *Misinterpretation* has the highest degree of severity owing to the misrepresentation of the intent. *Lack of emphasis* is least severe as the problems are dependent on the inefficiency of the visual communication process, and not the incorrectness of the representation.

Legend: In some visualization examples, we found that the charts are not self-contained: lack of legends for different symbols or relationships meant one has to either browse through the captions or the textual description for making sense of what the symbols mean. This was especially difficult where lots of different symbols are used on a chart, for example, the scatter plot shown in Fig. 8.

Annotation: In different visualization examples, we observed that annotation of salient patterns or data points on the chart could communicate the intent or some other critical aspects more effectively. For example, in case of the multiple maps (Figure 1(b)), the white color on the maps denotes a lack pf spatial extent, but that is not documented within the image itself. An annotation would clearly communicate this important aspect of the chart.

5.3 Problem consequences

After we created the final version of the taxonomy, we realized that while the taxonomy enabled us to categorize the problems and their causes, it did not capture their severity, and most importantly, their impact. In light of the numerous trade-offs a visualization practitioner has to face when creating a visualization, it would be useful to have: guidance on how severe a visualization problem could be, and a categorization of consequences it may lead to. To solve this problem we went through our list of problems again and consulted our synthesis of design principles which were based on the two high-level questions: Q1, about correctness; and Q2, about effectiveness and efficiency of visual representations. Based on these questions and inspired by the seminal work on graphical integrity [48] and the criteria of expressiveness and efficiency by Mackinlay [34], we created a list of potential problem consequences. These consequences bridged the low-level causes of design problems to high-level effects, which were more comprehensible from a domain scientists' point-of-view. The association of design consequences with design problems is shown in Table 1, sorted by their level of severity. The level of severity is defined by the graphical integrity principle [48], according to which the most important criteria for a visual representation is to represent the data correctly and accurately. The different problem consequences are described below.

Misinterpretation: Certain design problems could lead to misinterpretation of the data. Since this consequence directly interfered with the correctness of the interpretation, and violated the principles of graphical integrity [48], it had the highest level of severity. As shown in Table 1, ambiguity of visual variables and distortion due to scale inconsistency (Fig. 8) could lead to the misinterpretation of the data.

Inaccuracy: In scientific data analysis, an important requirement for visual representations is to allow scientists to deduce accurate estimates from the display. When certain design problems could lead to an inaccurate interpretation of the data with respect to the original intent, they would cause inaccuracy. The most prevalent design problems that caused this issue were distortion due to projection error and quantitative color maps in the form of rainbow color maps. Problems like chart mismatch could also cause an inaccuracy problem. For example in case of the scatter plot example (Fig. 1(a)), one has to mentally compute the spread of the different output variables, and therefore inaccurately perceive the differences in the spread.

Lack of expressiveness: The expressiveness [34] criteria dictated whether the visual representation matched with the properties of the data attributes. A lack of expressiveness condition would not clearly convey the intent as the certain aspects of the representation would not match the intent. The problems leading to lack of expressiveness from our taxonomy were chart mismatch, chart configuration, visual variable problem due to choice, level-of-detail due to jaggedness, and lack of explicit encoding.

Inefficiency: Efficiency of algorithms are measured in terms of speed. Inefficiency in visualization design could also be traced to the slowness of the interpretation on the part of the reader. When certain design problems did not directly interfere with the interpretation of the data with respect to the original intent in terms of its correctness or accuracy, but affected the speed and efficiency, they led to inefficiency. This category encompasses the principles of effectiveness [34], use of pre-attentive features [53] and visual variables for efficient search for patterns [13]. The problems leading to this consequence based on our taxonomy were level-of-detail due to granularity, qualitative color maps, clutter due to color mixing, superposition overload, and communication gap due to lack of grids and legend.

Lack of emphasis: In static charts it is often important to draw the reader's attention to salient portions of patterns that have higher priority than the rest. This can be done by highlighting different aspects and organizing the information in a structured way [17]. While these do not directly correspond to the data being shown, the emphasis on the key aspects of a chart affects that message that readers decode from a chart. The problems leading to a lack of emphasis consequence based on our taxonomy were lack of explicit encoding, grids and annotations. Since this consequence does not directly interfere with the intent, it has the lowest level of severity.

6 Matches and mismatches

At this point of our study, we realized we had the opportunity to get back to our group of climate scientists and get feedback on our categorization of the design problems. We realized this would not only be a useful way of validating our work, but it would also be interesting to observe how visualization problems compiled by a group of visualization expert would be received by a group of domain scientists. We realized that while extensive research exists on reporting design problems when evaluating visualization and, as we have seen above, on providing visualization guidelines, there's very little understanding or even exploration of how criticism and guidance is received and used by domain experts. We were interested in spotting cases where visualization experts and domain scientists disagree and dig deeper into why and how this kind of disagreement happens.

6.1 Interview Procedure

Before conducting the interview, for avoiding redundancy, we made a pass through all the problem categories, in an attempt to filter out

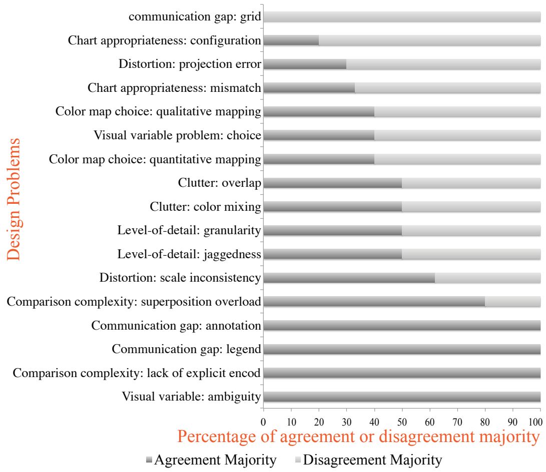


Fig. 9: **Causes of problems sorted by high percentage of majority disagreement** from top to bottom along the Y-axis. Scientists were mostly in agreement with the most severe problems. They were generally in disagreement with problems which required significant knowledge about visualization best practices and those which were in conflict with the domain conventions.

the images which are very similar to one another. For example, in the case of a rainbow color map problem, we only showed a few examples which expressed the problem.

We arranged for a face-to-face interaction with our direct collaborators, as part of a workshop, and the entire interaction lasted for about *four* hours. We described the taxonomy along with problems from the collected examples. Since some of the scientists did not have a background in visualization, we first gave examples of best practices in the choice of visual variables, color maps, chart selection, etc. We exercised caution in not using too many technical terms, but explained the problems as illustratively and simply as possible. We asked the scientists to fill up a spreadsheet where they had to write if they agreed or disagreed with a design problem. They also added in a comments section, the reasons for their disagreement.

We realized that there could be disagreement among the climate scientists themselves, about design problems. We did not want an apriori settlement of their disagreement, but instead wanted to collect raw data about the same and see if there is a majority disagreement. Therefore we requested our collaborators to record their feedback independently. After collecting all the responses, we separated cases which had majority agreement (more than half the people agree about a design problem) and majority disagreement (more than half the people disagreed among themselves in acknowledging a design problem).

6.2 Cases with Majority *Agreement*

Fig. 9 shows the distribution of the percentage of majority agreement and disagreement, sorted by high percentage of disagreement majority from top to bottom. We can observe that the scientists were generally in agreement with problems that lead to the most severe consequence, i.e., misinterpretation: scale inconsistency and ambiguity of visual variables. Also, there was a high percentage of agreement for comparison complexity problems and communication gap problems caused by lack of legend and annotation. As observed from Table 1, these categories lead to lack of emphasis consequence whose degree of severity was low and did not directly interfere with the communication of the intent.

In course of our interaction with the scientists we could reason with this apparent dichotomy, that is, they tended to agree with problems with lead to consequences with both highest and lowest degrees of severity. One of the reasons was that, the scientists could immediately recognize why certain problems led to misinterpretation as the visual representation in those cases misrepresented the data. There were other cases, like the comparison complexity problem, where majority of the scientists agreed with the problem, but they were not aware of the solution. It took a while for us to illustrate how lack of explicit encoding or superposition overload hindered their main intended task, which was comparison of models, and which solutions could work better. We showed them sketches and examples of how these problems could be solved by making the decoding process more efficient and emphasizing the salient patterns; after which they agreed with the problem. Understanding the effect of the communication gap problem caused by lack of legend or annotation did not require much visualization expertise. In several cases the scientists commented *"This figure desperately needs a legend, it is so difficult to flip back and forth to know what the symbols mean"*. For the annotation problem, the scientists sometimes acknowledged that annotation of the main trends would help them to focus directly on the main message instead of searching for it.

6.3 Cases with Majority *Disagreement*

During our interview, we found several instances where it was difficult for the visualization experts to convince the climate scientists about design problems, including well-known pitfalls like rainbow color map and $3D$ views. The cases with majority disagreement are shown in Fig. 9. These were, especially cases where the problems did not directly interfere with the intended tasks. In other words, the problems did not have any misleading consequence, however they interfered with accuracy, expressiveness, efficiency, and emphasis. Some of the comments we got for problematic charts based on these consequences were: *"improvements are subjective"*, *"minor problem"*, *"This might be a problem but I am ok with this plot"*.

Next we describe categories of major disagreements we have found during our interactions with the scientists. They are categorized in three main classes:

Domain Conventions: We found that the source of some of the design problems were existing conventions that the climate scientists followed. In some examples of line charts and scatter plots, we found them to be too cluttered or superposed with too many details to make sense of. For example, in many cases we found the use of dots along with lines on line charts, which cluttered the display and which we coded for both clutter due to overlap and choice of visual variables. However scientists explained that the observed data is by convention encoded by black dots, and the simulated data is encoded otherwise to distinguish them, and enable comparison between the two categories, i.e. observed and simulated.

For the superposition overload problem we found a group of line charts, that were similar to the one in Fig. 12, but with the additional complexity of multiple dots in addition to the lines, representing observation data. This was an obvious candidate for overload. However, some of the scientists asserted that this was more of a convention in the climate science community for representing observation data on top of lines for comparing simulation data, and they were comfortable with such a representation.

Loss Aversion: Similar to the tendency to avoid losses rather than acquire gains, which is popularly known as a *loss aversion* problem ([28]), climate scientists tended to focus more on avoiding loss of data in their visualizations, than on tuning the chart parameters for gaining insight from them. The recurring level-of-detail problem with line charts and maps exemplified this tendency. In examples where multiple maps were compared for understanding similarity of models, we suggested that a coarser granularity would facilitate more effective comparison. This was because if the number of maps being compared is more than three or four, it becomes difficult to perform a pixel-by-pixel comparison, where the data is encoded at the finest level of granularity.

Fig. 10: **Solution redesign for improving the scatter plot shown in Fig. 1(a)**. A box plot conveys the intent of showing spread among the output variables of prognostic and diagnostic models, and the accompanying heatmap alleviates the problem of having to visually search multiple symbols for knowing which models belong to which eco-regions.

For the granularity problem, were multiple maps were compared using a pixel-by-pixel mapping (Fig. 1(b)), although there was largely a consensus, some scientists said *"I would have to see the coarser version to really know if it is better, though"*. While one of them said *"Difficult to get a widespread trend. Personally, My eye tends to be drawn to red dots so I may be missing much of the information presented. A seldom problem is that I have trouble mentally overlaying the different plots. Each plot is a different variable/model. From this visualization, I have trouble comparing the locations and extremism of values at the same geography."* This showed that the comparison mechanism is not effective due to the low granularity but they were not convinces unless alternate solutions showed them the real benefit.

In many line charts, daily data were plotted where monthly or annual were being compared. Scientists observed that there is a need or tendency to show all the data because the time spent in extracting the data is significant. Also they believe that there might be some anomalies that might be missed by aggregating the data. Though all of them did not agree, there was general consensus about this fact. When shown alternatives with monthly averages computed, one of the scientists commented: *"I agree with you on this. But the situation is that people, especially scientists, they tends to show data as raw as possible. I think if this figure is used in scientific publications, it's fine."*

Awareness about visualization best practices: We found that in many cases scientists were not aware of what the visualization best practices are and why they should adhere to them. The categories which led to a lot of discussion between the scientists and visualization experts, involved use of 3*D* plots (coded under chart mismatch), use of color maps and choice of visual variables. It has been well documented in the literature that 3*D* plots lead to distortion and ineffective reading of the data [39]. However, since climate scientists are already oriented towards reading 3*D* volume visualizations, they did not think that a 3*D* layout for abstract data could create a problem. The same applies to the color map problems. In most cases they felt that the since they are already used to reading data from rainbow color maps, a more perceptually motivated color map would not make a difference to the goal of the intent. In a few cases they commented that: *"I agree that the color map can be better but that would be a cosmetic changes and won't affect the intent"*.

The effective use of pre-attentive features was also another category where climate scientists did not agree with most of the problems.. For example, the scatter plot in Fig. 1(a) encodes all the models by using different symbols. Even in absence of over plotting, the different symbols would cause an inefficient, conjunctive visual search. We discuss later in Section 7 how we could avoid this problem.

7 Solution Redesign

During our interactions with the climate scientists suggested that they needed to look at some solutions for better understanding the consequences of problems and how to avoid those. We agreed while it was useful to directly see why some problems should be avoided, and it was also important to see if the redesigned visualizations solved their problems better. We believed this phase would useful for visualization experts, because we got additional inputs which were not explicitly revealed in the previous phases. However, it is worth noting here that we follow a descriptive approach rather than a prescriptive one [32] and the final decision to judge the merit of a solution is left to the scientists.

For selecting images for our solution redesign, it was necessary to select a sampling of cases where scientists disagreed with the problem, or they agreed with the problem but were unaware of the solution. This would potentially demonstrate the effectiveness of optimal visualization designs to the climate scientists, if they found the solutions to be beneficial. As described in Section 6.3, there was a high level of agreement for problems that led to misinterpretation, inaccuracy, and lack of emphasis. Therefore, we selected examples for which the problems mainly led to lack of expressiveness and inefficiency, given by Table 1. We also selected examples where there was a high degree of agreement about the problem, but they were unaware of the solution, like, the problem due to comparison complexity caused by superposition overload and lack of explicit encoding.
To get feedback on the solutions, for each of the images, we specifically asked them if the redesigned solution conveyed the original intent better, and also if there were additional information they could gather which was not possible in the original visualization.

7.1 Scatterplot

Intent(s): The context of analysis here was comparison of models with respect to their output variables and respective eco-regions (Fig. 1(a)). The primary intent here was to look at the spread of the prognostic and diagnostic models with respect to the different output variables, and a secondary task was to identify, for each variable, which models belonged to which eco-region.

Design Problems and Consequences: The design problems were chart mismatch, choice of visual variables, granularity, and lack of grid lines. Chart mismatch happens because for the scatter plot to convey the first intent, readers have to mentally compute the spread, which is avoidable with a different representation; and more seriously, the second intent is very difficult to convey on a scatter plot due to over plotting. The granularity problem is caused by plotting every data point in a scatter plot, whereas the intent was to look at the spread of the models. Visual variables with different colors, shapes and orientation cause users to perform an inefficient, conjunctive visual search for the models. Absence of grid lines leads to a lack of emphasis of where the users should focus their attention: the chart should be read column-wise, which can be emphasized by use of grid lines. Grid lines were used by the climate scientists in two other examples, but as we observed from Fig. 9, most of them did not agree that adding grids could be beneficial in several examples.
Solution: A box plot is a more appropriate solution for conveying the first intent, i.e., allowing the users to readily understand the different patterns of spread. This is shown in Fig. 10. Since there

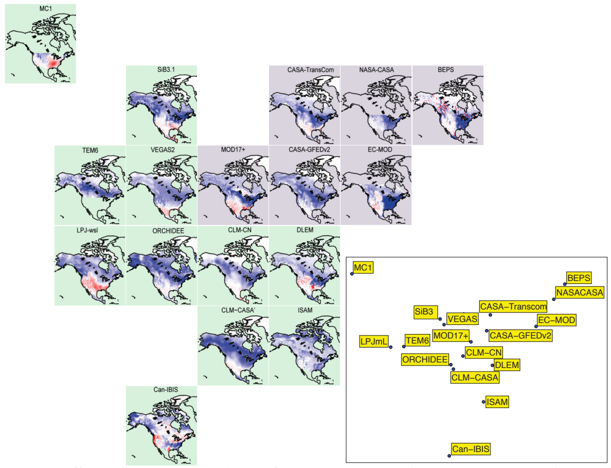


Fig. 11: **Solution redesign for the multiple maps (Fig. 1(b))**. Explicit encoding of similarity through multidimensional projection, where higher proximity signifies greater similarity among the models. The layout of the maps is based on the projection shown at the inset. Subsequent layout optimization for ensuring no overlap enables efficient comparison among multiple models.

are only a few outliers, we label them directly on the plot. For showing the membership of models in an eco-region, we use a tabular representation of the models in the form of a heat map, eliminating the need of additional visual variables that might lead to confusion. The heat map is basically a presence-absence chart where a cell is colored if a model belongs to an eco-region, and left white, if the model is absent for that eco-region.

Using the box plot, one can immediately recognize the much higher than average spread for diagnostic models, for the CRP eco-region for the NEP variable. One can also compare across different variables and models; for example the less average spread for both classes of models for GPP and NEP variable, where there are a few outlier models. This trend is however absent for the Rh variable. From the heatmap one can also immediately detect that most of the outlier models are diagnostic models, and find that the Rh variable is contained by much less number of diagnostic models as opposed to NEP and GPP.

Scientists' Feedback: Initially scientists were not convinced that the box plots are an improvement as they thought the scatter plot showed more information, like the ability to spot the models directly on the visualization. However when we added the heat map, showing their eco-region membership much more efficiently, they were convinced that the combination of box-plot and heat map eliminated the design problems. This is evident from the following comment by one of the scientists: "*Initially, I was inclined to reply that I liked seeing the model scatter on the first graph, i.e., I can think that I can see skew, bi-modality, etc from the scatter plot and if we just slightly offset the points horizontally and with only* 20 *points, then the overlap would not be too bad and I could glean more information. But upon examining the plot, well, you convinced me. What you provided does as good of a job as what I had imagined I would have preferred. In particular, showing the box and stem AND the outliers gave a good bit of information, as did the heat map*." They were also convinced about the utility of the dual view: *"one can immediately detect that most of the outlier models are diagnostic models. This was very difficult to achieve using the original scatter plots. One can also see that most of those models are for NEP and GPP and those are not present for the Rh variable"*.

7.2 Map

Intent(s): The intent here was to identify similarities and dissimilarities among models for summer months during the period 2000-2005, based on the spatial distribution of the NEP variable.

Design Problems and Consequences: The design problems was mainly a lack of explicit encoding as the positioning of the maps do not represent the degree of similarity among the models. Thus, the scientists have to sequentially search and compare models to get insight into their relative similarity. It is thus hard to find pairwise similarity between maps and find groups of similar maps.

Solution: We aimed to improve the visualization by deriving a summary statistic about similarity that climate scientists use, and manipulate the layout for encoding similarity. In discussions with the climate scientists, we used statistical information about the models, that is, root mean squared difference (RMSD) which is widely used in the climate science domain. Using the pairwise computation of RMSD between models, we applied multidimensional projection for displaying the maps in a two-dimensional space (Fig. 11a) using the ISOMAP [46] technique. RMSD was used as the distance function and in the two-dimensional space the proximity of the models denote similarity.

For representing the maps directly based on the projection view, we adjusted the layout using an optimization algorithm [21], so that maps did not overlap and spatial information was retained. Displaying the maps directly was important as the spatial extents of the models were different and the scientists wanted to see them on the geographical map. The projection view shows clear patterns. Note the point representing the MC1 model is far away from the rest of points, it means that its corresponding map is very different than the others. Another example are the maps SibB3 and VEGAS, their points are near meaning that they are similar (confirmed by the looking at the maps). Another similar group is formed by CLM-CASA and ORCHIDEE maps.

Scientists' Feedback: There was consensus among the scientists that the resultant figure not only conveyed the original intent but also showed additional information, like quantifying the degree of similarity or dissimilarity of the models based on a metric they were familiar with. They observed that this is a new visualization approach than what they are used to, and one of them expressed caution about the abstraction being used: *"I have to be cautious about the MDS method used. I agree that placing maps in different locations will be beneficial. But the MDS method is only one way to represent the similarity among those maps from one certain specific aspect"*. However, they were convinced about the utility of the approach and its benefit in expressing model similarity: *"Shows the outliers, and their degree of outlying, more clearly than the original. This is a great solution to a very commonplace visualization in climate modeling"*.

7.3 Line Chart

Intent(s): The intent here was to compare the temporal variability of multiple models with respect to each other and also with respect to the ensemble mean.

Design Problems and Consequences: The problems with the spaghetti plot ([1]) as shown in Fig. 12 were comparison complexity due to superposition overload, level-of-detail due to jaggedness of lines; and clutter due to overlap. Superposition overload and overlap led to an inefficient comparison of the temporal patterns. Jaggedness was caused by plotting of monthly data and this lead to a lack of expressiveness as the salient annual peaks and crests were occluded.

Solution: We aimed to solve this problem by converting the large single [50] or the spaghetti plot, into a series of small multiples. As shown in Fig. 13, we converted the individual lines into a band for showing the range of variation, and plotted the ensemble mean in each of the plots, shown by the black line, similar to the approach taken by Andrienko and Andrienko [2]. Each line plot now belongs to a model, and it is highlighted in red. We avoid using different colors for each

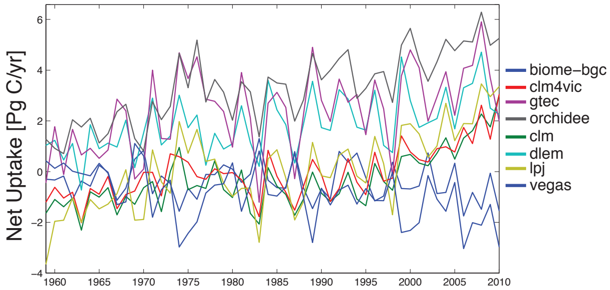


Fig. 12: **Original representation of the spaghetti plot** for comparing temporal variation of net uptake of models.

model, as the labels are sufficient for identification of a model. We found the small multiple approach being used the scientists mostly in case of maps, and in one more example, where maps and line charts were used for linking the spatio-temporal trends (Figure 3). One of the problems we did not address was the jaggedness of lines. Since we are losing resolution by using small multiples, one option was to compute an average by combining several years, and smooth out the time series. However due to information loss, scientists were not comfortable with the idea of smoothing by computation of average.

Scientists' Feedback: The scientists unanimously felt that the resulting small multiple display overcomes the problems that are traditionally present with a spaghetti plot. One of them commented: *"The new plots are definitely better than the original one. It's difficult to identify each model line in the original plot due to over-plotting"*. They appreciated the minimalist design by using few colors and also the fact that temporal variation could be compared quickly and intuitively with both the multi-model range, and the lines representing different models. One of the scientists also observed that: *"If the goal is to visualize model similarity then we can apply the same layout optimization as applied to the multiple maps example to rearrange similarly behaving models together"*. They were confident that this would be an exemplary visualization which will be emulated in model comparison scenarios and preferred over the traditional spaghetti plot.

8 Design Problem Trade-Offs

In this section we present one of the the key findings of our study, which is a reflection on the trade-offs among the different problem consequences. Many design problems and consequences cannot be simultaneously avoided. An awareness of the trade-offs is necessary for the scientists to judge how best to configure a visualization. The first decision that scientists have to take, is to weigh which design consequence more, and accordingly decide which potential design problems to avoid the most. A perfect visualization is hard to achieve and there is no one-size-fits-all formula for generating one. Visualization design is heavily parameterized by the scientists' intent, which needs to take into account the different trade-offs. Following were the different trade-offs we found in our analysis:

Lack of expressiveness vs Inaccuracy: This trade-off was observed in cases where scientists' intent for visualizing the data at the finest level of detail, led to a lack of expressiveness of the salient patterns. For example, in case of jagged lines, an average could be computed to reduce the number of steps, and that would lead to higher expressiveness, but at the cost of inaccuracy. Another similar example of this trade-off was choice of visual variable and granularity problem in maps due to pixel-by-pixel representation on maps. As acknowledged by some of the scientists, a coarser representation would have lead to better expressiveness of the data.

Inefficiency vs Inaccuracy: This trade-off was observed in cases where an accurate representation was achieved at the cost of an efficient one. Superposition of multiple lines and points for comparison with observation data is a common practice with climate

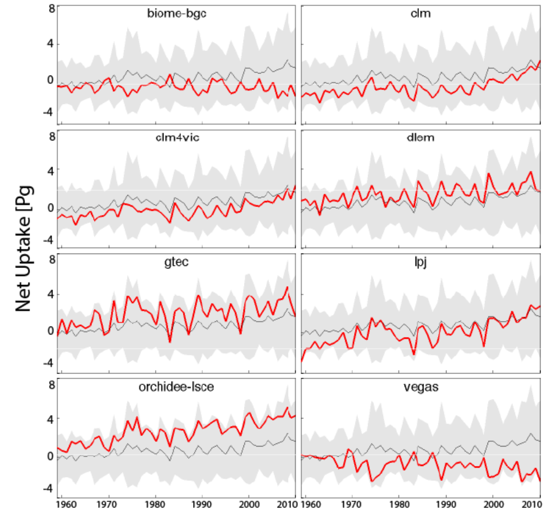


Fig. 13: **Solution redesign for improving the spaghetti plot.** Separating the model representations into small multiples of line charts enables efficient comparison of each model trend with the ensemble mean and range of variance.

scientists. While in many cases superposition facilitates accurate multiway comparison, fulfilling the expressiveness criteria, in some cases this also leads to clutter leading to inefficiency. Especially during in publications and broader dissemination, these criteria are important. In these cases, small multiples and use of explicit encoding of relationships should be considered.

Lack of Emphasis Vs Inefficiency: This trade-off was observed in cases where scientists' intent of keeping charts free of clutter, for achieving more efficiency, came at the cost of a lack of emphasis. Charts should be self-contained by use of proper labelling, grids and annotations if necessary, which help emphasize the intended message. Improper use of these auxiliary information however can clutter charts and make the decoding process inefficient. This trade-off is also echoed my Few's mantra of minimizing non-data ink [17].

9 Guidelines for Avoiding Design Problems

The design problems and consequences enable a visualization expert to reflect on design-trade-offs and formulate solutions based on the intents that consider those trade-offs. Climate scientists are not familiar with all visualization best practices, which we demonstrated in Section 6.3. It was thus necessary to abstract the problems, consequences and their solutions in the form of guidelines, that are more comprehensible from a scientist's perspective. Our objective was to distill the general problem trends and provide guidelines that can enable climate scientists to avoid those problems. The following guidelines should be understood in the context of the classes of visualizations we collected, which were maps, scatter plots, and line charts; and the scientific intents, which mainly centered around understanding and expressing similarity of climate models in those visualizations. The following guidelines are discussed with the context of the related design problems, so that scientists are able to bridge the gap between the design recommendations and current practices, and embrace the best practices in visualization.

G1. Keep audience in mind. A recurring issue cutting across different design problems was the tendency of scientists to use the visualizations designed for their own analysis, for publication and dissemination of their results as well. This was triggered by an implicit assumption about the familiarity of the audience with what to look for in the data. On the other hand, to cater to a broad audience, whether internal or external to the climate science community, the visualization itself should be expressive enough to convey the intent, without overwhelming the audience with the details.

In those cases, rather than representing all the data (level-of-detail problem), it is more important to show and highlight the trends by abstracting or aggregating some of the data. As we had pointed out in Section 6.3, the problem of loss aversion was a leading cause of problems in maps and line charts. There were many cases where the old adage of 'less is more' held true for the visualization designs. For example, if a line chart is too jagged because all time-steps are represented, it can obscure the message. In cases where the intent is to visually express similarity of multiple models, scientists can choose to represent the data at a coarser granularity by choosing a visual variable (choice of visual variable problem) other than color, like orientation of lines or glyphs which have been successfully adopted in the geographical data visualization domain [33].

G2. Guide users attention to salient visual objects. A critical requirement of any visualization design is to explicitly guide user's attention to the salient patterns. Enabling visual comparison of similar and dissimilar models was the underlying intent of the images we collected. To facilitate such comparison, key elements of the visual representations should serve as indexes for visual search for finding models that are similar or dissimilar.

We did not find effective use of Gestalt laws of grouping, which can be an effective visual cue in these cases. This led to the comparison complexity problem (Figure 7. For example, as we had shown in Fig. 11 absence of explicit encoding the message about similar or dissimilar models is not fully expressed. Another examples is the problem of superposition overload (Fig. 12). This causes clutter which can either disinterest the audience or cause trouble in finding the patterns. Scientists should avoid relying on the audience's mental operations to make those visual comparisons, which can be both inefficient and ineffective in absence of any visual cue. As we had elaborated in Section 5.3, explicit encoding of relationships and emphasis of the key message can alleviate these problems.

G3. Focus on the message and make it self-contained: In many cases, the scientists' intents were not fully conveyed as the message of a chart was incomplete due to either lack of emphasis of the take-away message or a lack of synergy between the auxiliary information and visual representation. In complex visualizations, it is often necessary not just showing the data, but also explaining what the visualization conveys through highlights and texts. To make the message clear, scientists can use size, color or orientation, that is substantially different from that of the other objects in the visualization, than the one which is most important. This is especially true of outlier objects.

The design problems related to this category were mostly those associated with a lack of emphasis. Although auxiliary information about charts only help when a chart is effective in the first place, they can help focus human attention very quickly to the salient portions of a chart. In cases of complex charts with multiple messages, this aids the user in decoding the intended message very efficiently. Charts should also be self-contained, without the audience having to search for the meaning of the legends in some other table or graph, which was true of some examples we collected from the research papers.

G4. Tie color selection with data semantics. We observed in Fig. 9 that choice of color was one of the problem categories where there was a lot of disagreement between visualization experts and climate scientists. We found that our collaborators generally considered the use of color as more of an aesthetic issue than it being tied to the data semantics. While there were some good examples of choice of color maps, in majority of the cases we found that the choice of color map was not appropriate.

Apart from color maps, we also found some other inappropriate choices of colors. For example, in some cases, we found the most important visual object, such as the mean or the trend line being encoded in gray, in which case it would be hidden in clutter and not be emphasized. Apart from different color scales for different data types, we also recommend the effective use of color in emphasizing certain objects (e.g., red can create a pop-out effect) or muting certain aspects of the data, like using gray to de-emphasize points in a scatter plot that create noise and use color only to encode certain salient points.

G5. Be mindful of defaults: For several design problems we had assessed, one of the precursors for the problems was the defaults of the tool that the scientists were using for creating those visualizations. One of the infamous defaults in many tools is the rainbow color map. The other one is the selection of random symbols for showing discrete data. In our taxonomy, these led to the color map choice and visual variable problems (Figure 5). The consequences of these problems can be as severe as misinterpretation, or lead to lack of emphasis for salient patterns (Table 1). In these cases, it is necessary for the scientists to look beyond the defaults and introspect if the defaults impede data analysis and visual communication. Such introspection might ultimately require scientists to manually configure visualizations for overcoming the problems with the default settings.

10 Scope and Impact

In this section we discuss the scope of our work and the impact in terms of the generalizability and utility of mapping design principles to domain-specific, static visualization designs.

Limitations: Our work has some important limitations to take into account; first of all is its subjective nature due to the qualitative methodology, While use of grounded theory and bottom-up approach to building visualization usage models are gaining ground [26], [51], it is an acknowledged fact that subsequent research needs to be done to develop prescriptive solutions to the problems. Accordingly, our focus in this work has been to reflect on the design problems through the descriptive taxonomy, which can be expanded in scope through further research to build prescriptive, broadly applicable solutions. Moreover, there are multiple ways of describing the design problems we found. We are not claiming that this is the only way to classify visualization design problems we found. However, we are confident that through our collaboration with a broad group of climate scientists and our understanding of the state-of-the-art in visualization practices, our classification provides a good starting point for bridging the gap between visualization best practices and existing climate data visualizations. The guidelines should be understood in the context of the sample of images we collected. These guidelines still need to be validated through empirical evaluations.

Generalizability: Although the sample of visualizations we collected was limited by their type (maps, line charts, and scatter plots) for us to build prescriptive solutions, we believe many aspects of our study are generalizable. First, although we used only three types of visualizations, they represent a broad set of usage scenarios in climate science: understanding spatial patterns, temporal patterns and looking at bivariate relationships among variables. The tasks mainly involved visual comparison of distributions, correlations, and variability, which are common analysis tasks cutting across climate science and even other domains. From that perspective, we are confident that the problem classification will be applicable to different domains and usage scenarios.

Second, the problem classification itself follows a mapping between general design principles and visualization examples from a domain. Even if some of the problems we found in the climate science domain do not exist in other domains, the same principles and classification scheme based on encoding and decoding *problem stages* would still apply. The same would apply for the *problem type* level, only the causes of the problems might be different. For example, there can be different causes for a level of detail problem, or the problem or clutter or distortion, but these problem types are still applicable for judging the quality of visualizations. As mentioned before, to the best of our knowledge, our work is a first step towards bridging the gap between general design principles and how they are realized in practice.

Third, from a visualization perspective, some problems we found are symptomatic of general gaps in research involving static visualizations. First, while much research has focus on judging effectiveness of interactive visualizations, many mediums such as publications and presentations are constrained by their static nature. We found that representing multivariate or multi-model relationships and effectively visually communicating their relationships have non-trivial challenges. Second, use of bad defaults has been widely talked about, but rarely addressed in the tools available today, with a few exceptions, like the Paraview tool where the rainbow color map was changed to the perceptually effective divergent color map [35]. Our findings should encourage such changes in the visualization tools, which will ensure better designs by domain experts.

Utility: A large body of research focuses on interactive visualization and it is a general assumption that good interactive visualization design can be easily and directly turned into good visual presentation design. But our work points to the fact there are different challenges and gaps and we need to better understand and research this difference. Visual presentation is not just taking pictures from our interactive tools and placing them into our papers and presentations. The design has to tell a compelling story about the findings of the scientists to the non-technical stakeholders, and in visualization, the presentation and story-telling aspect has received much less attention till date [31]. Some well-defined best practices like harmfulness of rainbow color maps [5] need more empirical validation, especially in the science community [4] for establishing the objective reasons behind recommendation of perceptually motivated color maps. A survey of existing visualization tools, investigating the quality of the defaults, will be helpful in identifying these issues [15] and will enable visualization non-experts like domain scientists design visualizations more efficiently.

11 Conclusion and Future Work

In this paper, we have presented a comprehensive study of visualizations designed by climate scientists and classified their shortcomings by categorizing the causes and consequences of design problems in the form of a taxonomy. In the process, we have investigated the cross-domain agreement and disagreement about design problems and highlighted their reasons. Further, we have demonstrated the utility of our taxonomy by getting feedback on redesigned solutions, which the scientists found to be beneficial for their practical use. In this work, we collaborated directly and indirectly with a large group of domain experts. This gave us a unique opportunity to understand visualization practices in the climate science community. We found that our collaborators are open to embracing new ideas about visualizing their data and the best practices that exist in our community. Currently, we are planning to work towards extracting the problem classifications defined by our taxonomy and building an automated system that will be able to detect design problems and provide recommendations for task-based solutions.

12 Acknowledgement

This work was supported by: the DataONE project (NSF Grant number OCI-0830944), NSF CNS-1229185, NASA ROSES 10-BIOCLIM10-0067, and DOE Office of Science Biological and Environmental Research (BER). The data was acquired through the MAST-DC (NASA Grant NNH10AN68I) and MsTMIP (NASA Grant NNH10AN68I) projects funded by NASA's Terrestrial Ecology Program. We extend our gratitude to members of the Scientific Exploration, Visualization, and Analysis working group (EVA) for their participation in the study and their continuous feedback and support in course of the project.

References

[1] T. T. Allen. *Introduction to engineering statistics and six sigma: statistical quality control and design of experiments and systems*. Springer, 2005.

[2] G. Andrienko and N. Andrienko. Visual exploration of the spatial distribution of temporal behaviors. In *Information Visualisation*, pages 799–806. IEEE, 2005.

[3] J. Bertin. Semiology of graphics: diagrams, networks, maps. 1983.

[4] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman, and H. Pfister. Evaluation of artery visualizations for heart disease diagnosis. *TVCG*, 17(12):2479–2488, 2011.

[5] D. Borland and R. M. Taylor. Rainbow color map (still) considered harmful. *Computer Graphics and Applications*, 27(2):14–17, 2007.

[6] C. A. Brewer. Color use guidelines for data representation. In *Proceedings of the Section on Statistical Graphics, American Statistical Association*, pages 55–60, 1999.

[7] A. Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.

[8] S. K. Card and J. Mackinlay. The structure of the information visualization design space. In *Symposium on Information Visualization*, pages 92–99. IEEE, 1997.

[9] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[10] K. Charmaz. *Constructing grounded theory: A practical guide through qualitative analysis*. Sage Publications., 2006.

[11] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Symposium on Information Visualization*, pages 69–75. IEEE, 2000.

[12] B. chul Kwon, B. Fisher, and J. S. Yi. Visual analytic roadblocks for novice investigators. In *Visual Analytics Science and Technology (VAST)*, pages 3–11. IEEE, 2011.

[13] W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society.*, pages 192–229, 1987.

[14] A. Dasgupta, M. Chen, and R. Kosara. Conceptualizing visual uncertainty in parallel coordinates. In *Computer Graphics Forum*, volume 31, pages 1015–1024. Wiley Online Library, 2012.

[15] C. Demiralp, C. Scheidegger, G. Kindlmann, D. Laidlaw, and J. Heer. Visual embedding: A model for visualization. *Computer Graphics and Applications*, 2014.

[16] T. Etiene, H. Nguyen, R. M. Kirby, and C. T. Silva. flow visualization juxtaposed with visualization of flow: Synergistic opportunities between two communities. pages 1–13, 2013.

[17] S. Few. Show me the numbers. *Designing Tables and Graphs to Enlighten*, 2012.

[18] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 199–206. ACM, 2010.

[19] P. Fox and J. Hendler. Changing the equation on scientific data visualization. *Science(Washington)*, 331(6018):705–708, 2011.

[20] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[21] E. Gomez-Nieto, W. Casaca, L. G. Nonato, and G. Taubin. Mixed integer optimization for layout arrangement. In *Proceedings,Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2013.

[22] L. Grammel, M. Tory, and M. Storey. How information visualization novices construct visualizations. *TVCG*, 16(6):943–952, 2010.

[23] J. Heer, F. van Ham, S. Carpendale, C. Weaver, and P. Isenberg. Creation and collaboration: Engaging new audiences for information visualization. In *Information Visualization*, pages 92–133. Springer, 2008.

[24] J. Hullman, S. Drucker, N. Henry Riche, B. Lee, D. Fisher, and E. Adar. A deeper understanding of sequence in narrative visualization. *TVCG*, 19(12):2406–2415, 2013.

[25] D. Huntzinger, W. M. Post, Y. Wei, A. Michalak, et al. North american carbon program (nacp) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, 232:144–157, 2012.

[26] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proceedings of the 2008 conference on BEyond time and errors: novel evaLuation methods for Information Visualization*, page 6. ACM, 2008.

[27] C. Jiménez, C. Prigent, B. Mueller, S. Seneviratne, et al. Global intercomparison of 12 land surface heat flux estimates. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 116(D2), 2011.

[28] D. Kahneman and A. Tversky. Choices, values, and frames. *American psychologist*, 39(4):341, 1984.

[29] T. Keenan, I. Baker, A. Barr, P. Ciais, et al. Terrestrial biosphere model performance for inter-annual variability of land-atmosphere co2 exchange. *Global Change Biology*, 18(6):1971–1987, 2012.

[30] N. Kodagoda, B. Wong, C. Rooney, and N. Khan. Interactive visualization for low literacy users: from lessons learnt to design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1159–1168. ACM, 2012.

[31] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[32] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *TVCG*, 18(9):1520–1536, 2012.

[33] A. M. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 2004.

[34] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

[35] K. Moreland. Diverging color maps for scientific visualization. In *Advances in Visual Computing*, pages 92–103. Springer, 2009.

[36] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM, 1990.

[37] J. Poco, A. Dasgupta, Y. Wei, W. Hargrove, C. Schwalm, R. Cook, E. Bertini, and C. Silva. SimilarityExplorer: A visual inter-comparison tool for multifaceted climate data. In *Computer Graphics Forum*, volume 33, pages 341–350. Wiley Online Library, 2014.

[38] A. J. Pretorius and J. J. Van Wijk. What does the user want to see? what do the data want to be? *Information Visualization*, 8(3):153–166, 2009.

[39] K. Risden, M. P. Czerwinski, T. Munzner, and D. B. Cook. An initial examination of ease of use for 2d and 3d information visualizations of web content. *International Journal of Human-Computer Studies*, 53(5):695–714, 2000.

[40] B. E. Rogowitz, L. A. Treinish, and S. Bryson. How not to lie with visualization. *Computers in Physics*, 10(3):268–273, 1996.

[41] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 761–770. ACM, 2005.

[42] M. Sedlmair, M. Bremer, S. Ingram, and T. Munzner. Dimensionality reduction in the wild: Gaps and guidance. Technical report, 2012.

[43] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. In *Computer Graphics Forum*, volume 31, pages 1335–1344. Wiley Online Library, 2012.

[44] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[45] S. Solomon. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge University Press, 2007.

[46] J. B. Tenenbaum. Mapping a manifold of perceptual observations. In *Advances in Neural Information Processing Systems 10*, pages 682–688. MIT Press, 1998.

[47] M. Tory and S. Staub-French. Qualitative analysis of visualization: a building design field study. In *Proceedings of BELIV Workshop*, page 7. ACM, 2008.

[48] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 31. Graphics press, 1983.

[49] J. W. Tukey. Graphic comparisons of several linked aspects: Alternatives and suggested principles. *Journal of Computational and Graphical Statistics*, 2(1):1–33, 1993.

[50] S. Van Den Elzen and J. J. Van Wijk. Small multiples, large singles: A new approach for visual data exploration. *Computer Graphics Forum*, 32(3pt2):191–200, 2013.

[51] J. Walny, S. Carpendale, N. Henry Riche, G. Venolia, and P. Fawcett. Visual thinking in action: Visualizations as used on whiteboards. *TVCG*, 17(12):2508–2517, 2011.

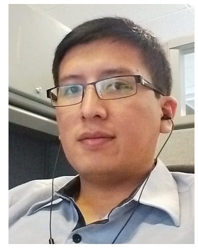
[52] J. Walny, B. Lee, P. Johns, N. Henry Riche, and S. Carpendale. Understanding pen and touch interaction for data exploration on interactive whiteboards. *TVCG*, 18(12):2779–2788, 2012.

[53] C. Ware. *Information visualization: perception for design*, volume 5. 2000.

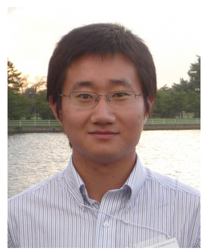
[54] C. Weaver, D. Fyfe, A. Robinson, D. Holdsworth, D. Peuquet, and A. M. MacEachren. Visual analysis of historic hotel visitation patterns. In *Visual Analytics Science And Technology (VAST)*, pages 35–42. IEEE, 2006.

[55] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and S. Carpendale. Heuristics for information visualization evaluation. In *Proceedings of the BELIV Workshop*, pages 1–6. ACM, 2006.

Aritra Dasgupta Aritra Dasgupta is a Research Assistant Professor at the School of Engineering, New York University. Over the past two years, he has collaborated extensively with climate scientists as part of the DataONE project, for applying visualization based approaches to climate science problems. He received his Ph.D. in Computing and Information Systems in December 2012 from the University of North Carolina at Charlotte. His main research interests are information visualization and visual analytics.



Jorge Poco received his BE degree in system engineering in 2008 from Universidad Nacional de San Augustin, Peru, and MS degree in Computer Science in 2010 from Universidade de São Paulo, Brazil. Currently, he is working towards his PhD degree at the School of Engineering, New York University. His research interests include Data Visualization and Computer Graphics.



Yaxing Wei Yaxing Wei is currently a scientist at the ORNL Distributed Active Archive Center (DAAC) in the Environmental Sciences Division. At ORNL, Wei has been working on several research projects to provide geospatial information management, analysis, visualization, and sharing. Those projects include ORNL DAAC, Modeling and Synthesis Thematic Data Center (MAST-DC), and the National Hydropower Assets and Resources Assessment Project. Wei received his PhD in safety engineering from the University of Science and Technology of China (USTC) in 2007, while his doctoral research was conducted at George Mason University in Fairfax County, Virginia.

Robert Cook (Environmental Sciences Division, ORNL) is a biogeochemist who works on large interdisciplinary projects. He is currently involved in several model-observation and model-model intercomparison projects, which involve developing a cyberinfrastructure framework for discovering, accessing, integrating, and visualizing diverse observation data and model output. Cook is the Chief Scientist for NASAs ORNL Distributed Active Archive Center. In addition to his work at ORNL, Cook has acted as an associate editor for the distinguished journal Biogeochemistry since 1996.

Enrico Bertini is Assistant Professor at the NYU Polytechnic School of Engineering in the Department of Computer Science and Engineering. His research focuses on the study of effective data visualization methods and techniques to explore and make sense of large and high-dimensional data. His research has been applied to several application domains including: biochemistry, cybersecurity, development, healthcare. Professor Bertini earned his PhD degree in Computer Engineering at Sapienza University of Rome in Italy.

Cláudio T. Silva is a Professor of Computer Science & Engineering at New York University. From 2003 to 2011, he was with the School of Computing and the Scientific Computing and Imaging Institute at the University of Utah. He received the BS degree in mathematics from the Federal University of Ceara, Brazil, in 1990, and the PhD degree in computer science from the State University of New York at Stony Brook in 1996. He coauthored more than 200 technical papers and 12 U.S. patents, primarily in visualization, geometry processing, computer graphics, scientific data management, HPC, and related areas.