

## Using Visualization Science to Improve Expert and Public Understanding of Probabilistic Temperature and Precipitation Outlooks

MICHAEL D. GERST

*Earth System Science Interdisciplinary Center/Cooperative Institute for Satellite Earth System Studies,  
University of Maryland, College Park, College Park, Maryland*

MELISSA A. KENNEY

*Earth System Science Interdisciplinary Center/Cooperative Institute for Satellite Earth System Studies,  
University of Maryland, College Park, College Park, Maryland, and Institute on the Environment,  
University of Minnesota, Saint Paul, Minnesota*

ALLISON E. BAER, AMANDA SPECIALE, AND J. FELIX WOLFINGER

*Earth System Science Interdisciplinary Center/Cooperative Institute for Satellite Earth System Studies,  
University of Maryland, College Park, College Park, Maryland*

JON GOTTSCHALCK, SCOTT HANDEL, MATTHEW ROSENCRANS, AND DAVID DEWITT

*National Oceanic and Atmospheric Administration, Climate Prediction Center, College Park, Maryland*

(Manuscript received 27 August 2018, in final form 5 November 2019)

### ABSTRACT


Visually communicating temperature and precipitation climate outlook graphics is challenging because it requires the viewer to be familiar with probabilities as well as to have the visual literacy to interpret geospatial forecast uncertainty. In addition, the visualization scientific literature has open questions on which visual design choices are the most effective at expressing the multidimensionality of uncertain forecasts, leaving designers with a lack of concrete guidance. Using a two-phase experimental setup, this study shows how recently developed visualization diagnostic guidelines can be used to iteratively diagnose, redesign, and test the understandability the U.S. National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) climate outlooks. In the first phase, visualization diagnostic guidelines were used in conjunction with interviews and focus groups to identify understandability challenges of existing visual conventions in temperature and precipitation outlooks. Next, in a randomized control versus experimental treatment setup, several graphic modifications were produced and tested via an online survey of end users and the general public. Results show that, overall, end users exhibit a better understanding of outlooks, but some types of probabilistic color mapping are misunderstood by both end users and the general public, which was predicted by the diagnostic guidelines. Modifications lead to significant gains in end-user and general public understanding of climate outlooks, providing additional evidence for the utility of using control versus treatment testing informed by visualization diagnostics.

---

### 1. Introduction

Embedding science in tools for supporting decision-making and planning and representing it in public communications has long been a challenge (Dilling and Lemos 2011; Lemos et al. 2012). This is partly due to the time and cost required for aligning user needs and expectations with what the science can provide (Meadow et al. 2015). Even if this engagement or coproduction process is managed well, problems still can emerge in

---

 Denotes content that is immediately available upon publication as open access.

---

 Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WCAS-D-18-0094.s1>.

---

*Corresponding author:* Melissa A. Kenney, [makenney@umn.edu](mailto:makenney@umn.edu)

DOI: 10.1175/WCAS-D-18-0094.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

the design of tools or communications. One particular source of difficulty is that most scientific information is infused with many trends or patterns and often contains significant scientific uncertainty, increasing the dimensions that a tool or communication must try to incorporate (Quinan and Meyer 2016). Given that there only are so many choices available to visually represent the complexity of scientific information, care must be taken to match the most important trends or patterns with the most visually effective design choices (Harold et al. 2016). For public communication, this problem often is compounded by designers not knowing which trend or pattern is of most interest, and in general, the public has a lower scientific skill level than expert users (McMahon et al. 2015).

These problems are particularly salient for maps of extended-range weather (i.e., 6–10 and 8–14 days) to long-lead seasonal forecasts (i.e., 3–4 weeks, 1 month, and 3 months), which add a geospatial component to scientific uncertainty. To varying degrees, weather and seasonal forecasts inform decisions of diverse users ranging from the general public to private and public sector decision-makers. For example, disaster managers have used forecasts for flood and drought management in humanitarian crises (Braman et al. 2013; Tadesse et al. 2016), the wind energy sector has used them for forecasting potential energy production (Roulston et al. 2003; Foley et al. 2012), and the agricultural sector has used them for irrigation decisions and commodity pricing, among other uses (Clements et al. 2013). While these cases show that forecasts are used to some extent, three main factors cause them not to be fully incorporated into decision-making (Changnon and Vonnahme 1986; Hartmann et al. 2002; White et al. 2017).

First, there might be a mismatch between user needs and characteristics of the forecast, such as whether the skill level, lead time, forecast period, weather or climate variable, and spatial resolution is adequate for making decisions. Forecasts also might not be easily accessible or in an understandable format. Additional barriers can occur if users require additional background information to assess forecast credibility (Sonka et al. 1992; Changnon et al. 1995; Pulwarty and Redmond 1997; Callahan et al. 1999; Pagano et al. 2002; Rayner et al. 2005; Lowrey et al. 2009).

Even if there is a reasonable match between forecast characteristics and user needs, user and institutional factors can hinder forecast uptake. For example, both the general public and end users have difficulties reasoning with probabilistic information and often request more forecast accuracy than is needed to make a decision that is better than random guessing (Sonka et al. 1992; Pagano et al. 2002; Steinemann 2006; Wernstedt et al. 2019). Furthermore, users exist within organizations that

set the context for how forecasts are used (Ray and Webb 2016; Simpson et al. 2016). Institutions can be reluctant to change, and their decision-making structure might not be compatible with a forecast's characteristics, especially if processes do not exist to integrate forecasts with other types of information used in decision-making. In addition, the appropriate expertise may not be available at an organization, either in house or through relationships with the forecasters (Changnon et al. 1995; Pulwarty and Redmond 1997; Callahan et al. 1999; Pagano et al. 2001; Rayner et al. 2005; Lowrey et al. 2009).

While much progress has been made in the usable science literature on how to align forecast characteristics with user needs and institutional structure, many open questions remain, including the impact of visualization choices on geospatial forecast understandability (White et al. 2017). In particular, representing geospatial forecast uncertainty is an unsettled topic in practice and in the research literature (Rautenhaus et al. 2018). This manifests in the default settings of many forecast visualization software packages deviating from visualization best practices (Quinan and Meyer 2016). For example, many forecast visualizations use rainbow color maps, which are considered a poor choice by the visualization science community (Borland and Taylor 2007; Stauffer et al. 2015; Dasgupta et al. 2019). In addition, users often prefer extraneous detail that hinders understanding (Hegarty et al. 2009), leading to map conventions that attempt to convey too much information in one image, known as visual clutter (Rosenholtz et al. 2007). These and other deviations from visualization best practice could greatly hinder understandability, as visualization is often the first avenue through which users interpret information embedded in a forecast (Hegarty 2011).

However, making changes to broadly distributed visualizations is not a trivial task. Forecasting and visualization conventions are embedded in the institutions that create them, and therefore require time and effort to change. In addition, users' interpretations of a visualization are in part dependent on the familiarity that has been built over time (Harold et al. 2016). Thus, there is a need to test whether implementing best practices will lead to a change in understandability of forecasts and use by decision-makers. Previous work has touched on separate specific aspects of representing geospatial and forecast uncertainty (MacEachren et al. 2005; Kaye et al. 2012). However, to our knowledge, no study has applied recent advances in diagnosing visualization problems (Dasgupta et al. 2015) to a comprehensive assessment of high-profile forecasts. Such an approach is important for systematically identifying visualization issues in products with long operational histories (Kinkeldey et al. 2014).

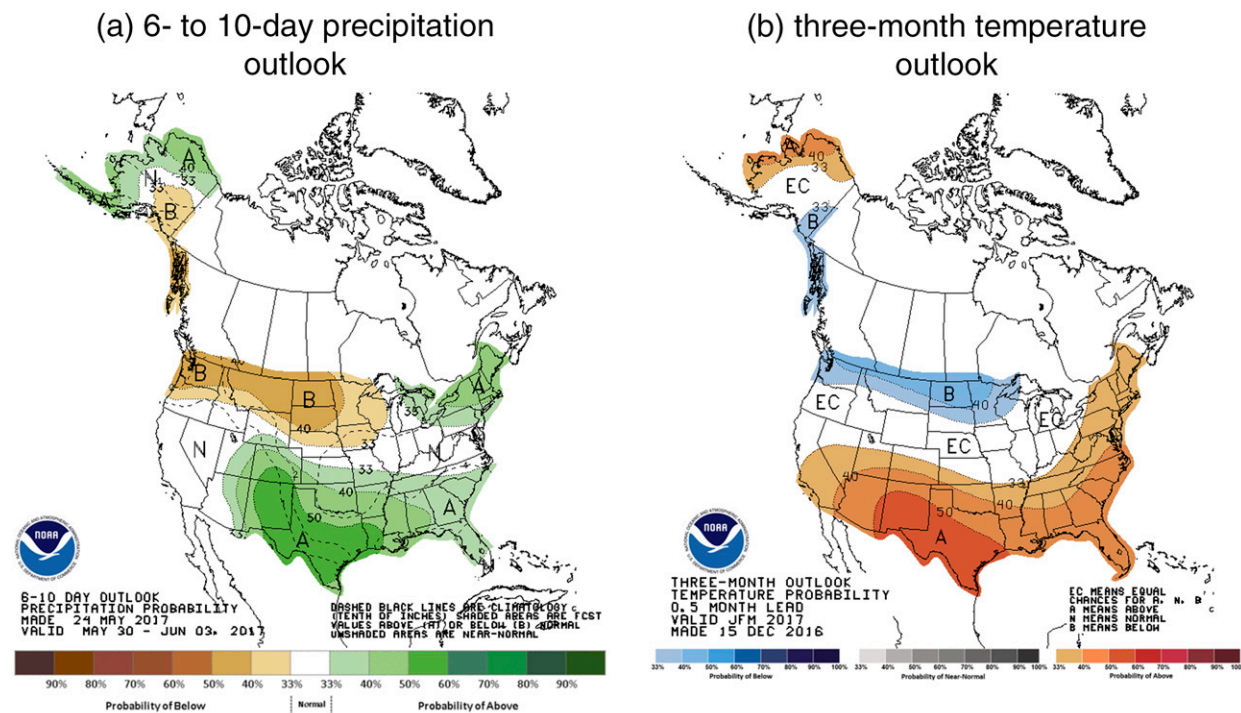


FIG. 1. Example of (a) 6–10-day precipitation (30 May–3 Jun 2017) and (b) 3-month temperature outlooks (January–March 2017). Note that the shorter-time-scale 6–10-day outlook did not include the “equal chances” category. The 6–10- and 8–14-day outlooks (not shown) use similar visual conventions, as do 3- and 1-month outlooks (not shown). Source: NOAA CPC, [cpc.ncep.noaa.gov](http://cpc.ncep.noaa.gov).

To address this gap, we apply visualization diagnostic guidelines to the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) extended-range and long-lead outlooks and test the understandability of modified visualizations on the general public and end users. The results from this study provide evidence for which modifications improve the understandability of NOAA climate outlooks, and more generally, which visualization best practices yield improvements.

## 2. Background

### a. Climate prediction center outlooks

While the U.S. federal government has been producing extended-range and long-lead temperature and precipitation forecasts since the 1940s, the current suite of products generated by NOAA CPC took their modern form in the mid-1990s (Barnston et al. 1994; Barnston et al. 1999). Currently, this includes 6–10- and 8–14-day extended-range outlooks and 3–4-week, 1-month, and 3-month long-lead outlooks (Fig. 1). These outlooks complement shorter-time-scale forecasts produced by other parts of the National Weather Service (NWS), such as the Weather Prediction Center (WPC) and local NWS Weather Forecast Offices

(WFOs). The base set of outlooks from CPC provide information on the most likely range of predicted average temperatures or accumulated precipitation, while other outlook products highlight the potential for extreme events that have impacts to life and property.

Since their introduction in 1994, a key feature of NOAA’s climate outlooks is their characterization and visualization of forecast uncertainty (O’Lenic et al. 2008; Livezey and Timofeyeva 2008). As opposed to mapping probabilistic temperature and precipitation amounts, the outlooks use historical climate data as a reference point for probabilistic-based forecasts. Specifically, the distribution of historical climate over a specified 30-yr period is binned into terciles, which are labeled *below*, *near*, and *above normal*. Using expert judgment and probabilistic model outputs, forecasters designate at any point on the map the category that exceeds 33% probability of occurring. For longer-range outlooks, if it is determined that below-, near-, and above-normal categories are equally likely, then the forecast designation of *equal chances* is given (Fig. 1b) (NWS 2018).

Extended-range and long-lead outlooks, which are less familiar to the general public and many decision-makers, differ from short range forecasts because they show the likelihood of a shift in climate conditions over a specific

timeframe (White et al. 2017). While it accurately represents forecaster judgment, this particular format of characterizing geospatial forecast uncertainty can be confusing to both expert and nonexpert users. Namely, users often confuse the probability of the below-, near-, and above-normal categories with a percentage decrease or increase in temperature or precipitation (Hartmann et al. 2002; Pagano et al. 2002; Steinemann 2006; Wernstedt et al. 2019).

### b. Visualization science

Visualization choices may exacerbate the difficulty users have in interpreting different forecast types that use the same conventions. Shorter-range weather forecasts, which are more familiar to the general public, display likely temperature and precipitation values or the probability of specific values occurring on a particular day. These visualizations follow standard conventions, such as warm colors representing warmer temperatures (White et al. 2017). Despite this substantial difference in forecast characteristics, many of the same short-range visual conventions are used in climate outlooks to indicate whether conditions are forecasted to be below-, near-, or above-normal categories. This may exacerbate observed user confusion over how to interpret extended-range and long-lead forecasts (Pagano et al. 2002; Hartmann et al. 2002; Steinemann 2006; Wernstedt et al. 2019) because red/hot-color and blue/cold-color metaphors are deeply embedded in user expectations (Ho et al. 2014).

How to more effectively represent below-, near-, or above-normal categories is not evident because representing geospatial uncertainty is an open question in visualization science for a few reasons (MacEachren et al. 2005; Rautenhaus et al. 2018). Part of the issue is that the concept of uncertainty in geospatial representations includes many dimensions. Uncertainty in the mapped data can occur during acquisition (e.g., from a satellite signal or weather model), the transformation of the data, and visualization (e.g., from interpolation) (Pang et al. 1997). Also, uncertainty can refer to not knowing the true value of a pixel (as in the CPC outlooks), the location of a feature such as a stream or building, and other properties (Buttenfield and Beard 1994; Thomson et al. 2005; Potter et al. 2012).

Many options exist for how to visualize geospatial uncertainty (Kaye et al. 2012). Separate maps can be used for the variable of interest and a measure of its uncertainty, or both can be shown on the same map as one composite variable or different variables (MacEachren 1992). Once the number of maps and variables are chosen, still more choices exist

for the visual variables: location, size, color value (darkness/lightness), texture/grain, color hue, orientation, shape, color saturation (grayness), and focus (Bertin 1983; Morrison 1984; MacEachren 2004). Choosing the right set for multidimensional data features can be a challenge, as different visual variables draw more visual attention than others and interact with each other in unexpected ways (Wolfe and Horowitz 2004, 2017).

The specific type of geospatial uncertainty represented in CPC outlooks, a two-dimensional scalar, has received significant attention in the literature. Initial studies hypothesized that color saturation would be an effective visual variable for uncertainty (MacEachren et al. 1998). While users in some studies preferred saturation as a representation of uncertainty, it was found to be less effective than other visual variables such as color value, texture, and focus (Schweitzer and Goodchild 1992; MacEachren et al. 1998; Leitner and Buttenfield 2000; Edwards and Nelson 2001; Retchless and Brewer 2016). This lack of effectiveness is partly due to color saturation not efficaciously directing visual attention when other color properties such as hue and value are used at the same time (Retchless and Brewer 2016).

### c. Diagnostic visualization guidelines

Concepts such as directing visual attention have been central to more general diagnostic methods in visualization science, which seek to provide comprehensive visualization guidelines based on the synthesis of evidence and insights from the practitioner and research communities (Hegarty 2011; Dasgupta et al. 2015; Harold et al. 2016). Central to these guidelines is a perception- and cognition-based understanding of how users perceive and interpret images. This account emphasizes the cyclical interaction of *bottom-up* and *top-down* processing (Hegarty et al. 2010; Hegarty 2011). Bottom-up processing describes the direction of visual attention due to the properties of the image, such as color and shape (Wolfe and Horowitz 2004, 2017). In contrast, top-down processing refers to the direction of attention due to users' expectations and prior knowledge (Gilbert and Li 2013). These two types of processing iteratively work to create and refine a mental representation of the image, which is compared to prior knowledge. If problems exist in visualization design, information can be misinterpreted or inefficiently processed. In this study, we focus on the diagnostic method outlined in Dasgupta et al. (2015), which is based on preventing five consequences of visualization design problems: misinterpretation, inaccuracy, lack of expressiveness, inefficiency, and lack of emphasis.

TABLE 1. Diagnostic guideline taxonomy from Dasgupta et al. (2015).

Problem stage	Problem type	Cause of problem
Encoding	Chart appropriateness	Chart mismatch Chart configuration
	Visual variable	Visual variable choice Visual variable ambiguity
	Level of detail	Granularity Jaggedness
	Color map	Quantitative color mapping Qualitative color mapping
Decoding	Clutter	Overlap Color mixing
	Distortion	Scale inconsistency Projection error
	Comparison complexity	Superposition overload Lack of explicit encoding
	Communication gap	Legend Annotation Grid

Misinterpretation is the most severe consequence as the user is drawing an incorrect inference from the visualization. Misinterpretation can occur, for example, if the same visual variable is assigned to more than one data feature, leading to an ambiguous interpretation. A related consequence is an inaccurate inference. While a user might correctly interpret one quantity as greater than another, design problems, such as using the wrong chart type, can estimate that difference as too large or too small. A lack of expressiveness occurs when the design does not clearly indicate the intent of the visualization. For example, if identifying a specific pattern or trend is the intent of the image, then the most effective chart types and visual variables used should direct attention to that specific pattern or trend and not another less important one. A visualization that leads to a correct and accurate interpretation and expresses the underlying data well can still suffer from being inefficient or lack emphasis. Inefficiencies occur when design, for example, creates visual clutter or overly complicated visual comparison tasks. Lack of emphasis occurs when auxiliary elements, such as legends, grids, or annotations, do not highlight essential areas of the image.

Consequences are linked to visual design problems that are grouped into a hierarchical taxonomy (Table 1). At the top level, two design stages are delineated: encoding and decoding. Problems associated with the encoding stage are related to choices made by the image designer to map data features to visual features.

It encompasses problem types of choosing inappropriate chart types, visual variables, levels of detail, and color maps. In contrast, decoding-stage problems are related to how the image interacts with a user’s perceptive and cognitive abilities. It encompasses problem types of too much visual clutter, scale or projection distortion, requiring a visual comparison task that is too complex, and ineffective auxiliary items such as legends. In the Dasgupta et al. (2015) taxonomy, more specific problem causes are delineated under each problem type. For example, inappropriate visual variables can stem from choosing an ineffective visual variable given the intent of the visualization or ambiguity with respect to mapping data features to visual variables. In the following, we describe how this diagnostic method is used in our study of CPC outlook understandability.

### 3. Methods

This study is a performance assessment of CPC temperature and precipitation outlooks. As such, it measures how well users understand the outlooks through task-oriented controlled experiments administered via online surveys, which is a common study design in the visualization literature (Lam et al. 2012; Kinkeldey et al. 2014). The Dasgupta et al. (2015) taxonomy-based diagnostic guidelines are used in two ways. First, they are used to generate hypotheses about how CPC outlook visualizations might be misinterpreted, and second, as a way to guide visualization redesign.

To check the reasonableness of the hypothesized design problems, we first met with eight federal government experts that were identified by CPC staff. These participants were chosen based on their extensive experience with the outlooks and knowledge of end users. The semistructured interviews ranged in length from 40 to 90 min; after the interview, participants were sent a transcript and allowed to review and remark on their comments. The content of the interviews focused on (i) user communities and their needs and (ii) challenges in interpreting the outlooks. The questions were open ended, asking interviewees to relate what they knew about how end-user communities utilized the outlooks. In addition, interviewees were asked to describe extended-range and long-lead outlooks along with what they liked and did not like about them and what they would do to improve them. The results of these initial interviews were used to refine the hypothesized design problems.

Next, a set of questions was developed to test whether the hypothesized design problems led to misinterpretation of the outlooks. These questions were



incorporated into a broader instrument that was used in individual interviews, focus groups, and online surveys. Additional questions covered users' decision context and familiarity and use of extended-range and long-lead outlooks. In consultation with CPC, four user group sectors were targeted: agriculture, emergency management, water resources, and energy. Participants were identified by CPC and through web searches, referrals from colleagues, and snowball sampling (i.e., a participant suggested a colleague to participate) (Atkinson and Flint 2004), leading to a sample size of 32 focus group and interview participants and 131 survey respondents. Focus groups varied in size but were structured so that participants were given time to consider their responses individually before sharing with the group, which is a common focus group structure (Krueger and Casey 2015). The understandability problems identified in the interviews, focus groups, and surveys were then used to inform final control versus treatment testing.

Final testing included two rounds of online survey testing using a randomized control versus treatment design, where redesigned outlooks were informed by diagnostic guidelines used in first assessing the outlook images. The first round tested differences in how well end-user and general public populations understood two redesigned outlooks (treatments) versus the original (control) visualization. End users ( $n = 427$ ) were identified based on consultation with CPC, snowball sampling, web searches, and a survey link on CPC's website. The general public sample ( $n = 658$ ) was taken from a random U.S.-based pool provided by survey company ROI Rocket ([roirocket.com](http://roirocket.com)) that was balanced by gender and restricted to having a college education. Based on these results, a final redesigned image was tested with the general public ( $n = 223$ ), for a total general public sample size of 881. The division of respondents by image is shown in Table A1 in appendix A.

Each respondent was randomly assigned either the control (Fig. 2a) or one of the modified outlooks (Figs. 2b–d). The control was taken from an archival forecast, as there was a desire for CPC to have a baseline measurement of how users understood an outlook as is, even though these outlooks had not issued near-normal forecasts since 2006. This design choice necessitated constructing hypothetical outlooks for the treatments because there was a need for CPC to test the effectiveness of different near-normal visualizations. Even though this constraint created two separate forecasts, forecasts across the treatments were the same and the interpretive tasks for the control and treatments were the same. In addition, survey questions were designed to

minimize the effect of the specific forecast on user understandability.

Two types of multiple-choice understandability questions were posed. Both were designed to be answered independent of the forecast shown. One type asks about specific visual variables used on a map, such as the meaning of white color mapping in the United States or the meaning of the warm color scale. The other type of question is more task specific. For each category, respondents were asked to give the probability of the category in a specific state, such as the probability of above normal in Maine (for Fig. 2a) or Washington (Figs. 2b–d). Although the location of the state is different, the question and potential answers are the same. Any effect on user responses would be introduced if a user is more familiar with one particular state over another. Given the location of a state should be familiar to end users, and is easily referenced with an internet search, we believe this effect is minimal.

Additional questions were asked about familiarity, background on how the outlooks were produced, and demographic information. The background questions were used to test whether user knowledge of the outlooks was predictive of understanding their content. Such objective measures of background knowledge, as opposed to more subjective measures such as confidence or familiarity, are thought to be better predictors of understandability (Lam et al. 2012; Kinkeldey et al. 2014).

The effect of background knowledge on the odds of answering a question correctly can be measured by examining the slope parameters from a logit regression (Wooldridge 2016). Here the independent variables are the number of questions answered correctly (BK) and the number of questions answered "I don't know" (UNC) and the dependent variable  $y$  is 1 or 0 depending on whether the question was answered correctly or not, respectively. As detailed in appendix B, we test the robustness of the effect of background knowledge by specifying a variety of equation structures, with each one embodying a hypothesized relationship between the dependent and independent variables. The main equation we report here controls for the type of modification  $i$  (MOD) and whether or not the respondent is an end user or part of the general public (EX):

$$y = \beta_0 + \sum_{i=1}^3 \beta_i \text{MOD}_i + \beta_4 \text{EX} + \sum_{i=5}^6 \beta_i \text{EX} \times \text{MOD}_i + \beta_7 \text{BK} + \beta_8 \text{EX} \times \text{BK} + \beta_9 \text{UNC} + \beta_{10} \text{EX} \times \text{UNC} + \beta_{11} \text{BK} \times \text{UNC} + \beta_{12} \text{EX} \times \text{BK} \times \text{UNC}. \quad (1)$$

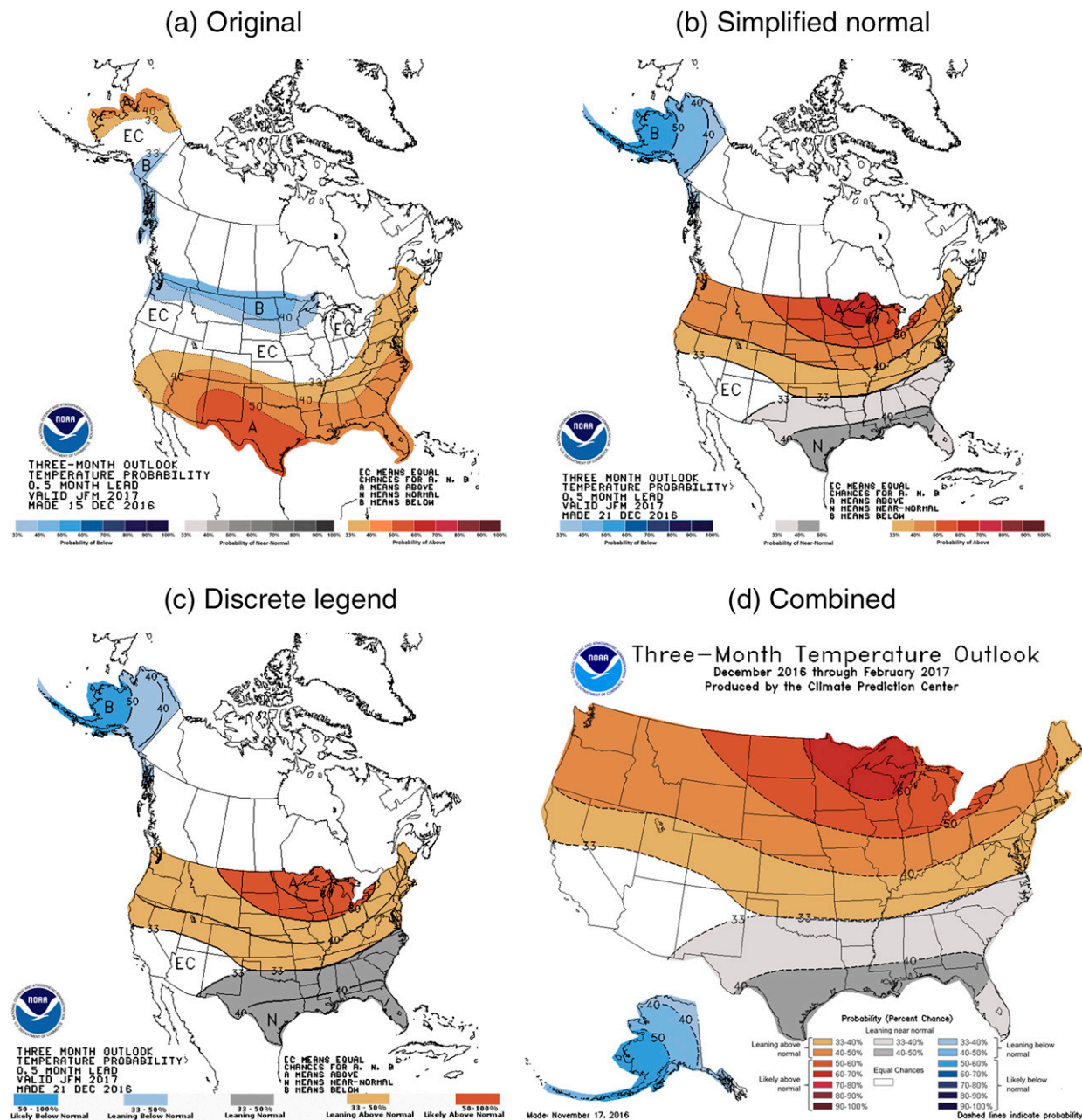


FIG. 2. NOAA CPC climate temperature outlooks for (a) original graphical approach presented by NOAA; (b) simplified representation of near-normal conditions using grayscale and updated legend; (c) discrete legend that represents normal using grayscale, aggregated probability ranges, and qualitative probability descriptors in the legend; and (d) combined approach addressing all five understandability diagnoses. Final images tested with end users are in (a)–(c) and with the general public are in (a)–(d). Note that the maximum probability value for above- or below-normal temperatures is 94%. Sources are (a) NOAA CPC ([cpc.ncep.noaa.gov](http://cpc.ncep.noaa.gov)) and (b)–(d) hypothetical forecasts developed by authors.

Because EX is a binary variable that takes the values 0 and 1 when a respondent is part of the general public or an end user, respectively, the coefficients on terms with EX can be interpreted as a test of whether being an end user matters or not. For example, if  $B_8$  is statistically significant, then background knowledge affects the

ability to interpret an outlook differently for an end user relative to the general public.

In addition to testing the statistical significance of the slope coefficients, the joint significance of the interaction terms are tested using a likelihood ratio test (Wooldridge 2016). In the results that follow, we report

TABLE 2. Summary of design problems and their diagnoses from the literature.

Design problem	Diagnoses <sup>a</sup>	Possible remedies
White color mapping in Canada and U.S.	Visual variable ambiguity	Choose different colors to represent different information types; remove Canada from map
Color mapping and conceptual representation of near normal	Quantitative mapping, communication gap (legend)	Choose different color; redesign legend
Misinterpreting probability ranges and precise probabilities	Communication gap (legend and annotation)	Redesign legend
Climatology contour lines <sup>b</sup>	Superposition overload	Remove climatology contour lines
Difficulty finding and reading map annotations	Communication gap (annotation)	Add notations for all information represented in graphic, change font size/type; move text or remove visual clutter, such as locations where there is no outlook presented (e.g., islands)

<sup>a</sup> Diagnoses use terms from Dasgupta et al. (2015).

<sup>b</sup> Not all outlooks used climatology contour lines.

the slope coefficients for Eq. (1) only if the interaction terms are jointly significant.

## 4. Results

### a. Diagnosis and redesign

As described in the methods, diagnosis of outlook visualization problems relied on the convergence of guidance from the Dasgupta et al. (2015) taxonomy and preliminary interviews, focus groups, and online surveys. Similar themes emerged from diagnostic guidance, experts, and end users, which may be summarized into five diagnoses (Table 2). The importance of these visual problems to correctly interpreting the outlooks was then tested through online surveys of end users and general public (section 4b). Almost two-thirds of the focus group participants were from the emergency management or agriculture sectors. Similarly, the online surveys of end users were composed of about two-thirds emergency management and one-quarter agriculture, with the remaining spread across the water resources and energy sectors.

Among focus group participants, the meaning of the color maps across the United States and Canada was identified as a potential source of confusion for non-expert users. This was corroborated by initial end-user survey results. Depending on the specific outlook, 20%–30% of end users misidentified white color mapping in Canada to mean near-normal or equal chances. This finding is not surprising given that the visualization diagnosis literature predicts that visual variable ambiguity will cause misinterpretation (Dasgupta et al. 2015).

Understanding white color mapping in the United States is dependent on whether white is assigned to mean equal chances or near normal. For outlooks that

use white to denote equal chances, 89%–98% (depending on the outlook) of end users correctly identified it as such. In contrast, extended-range outlooks, which use white to denote near normal, only 71%–77% of end users correctly identified the meaning of the color maps. Since the extended-range outlooks do not include equal chances as an outlook category, confusion for these outlooks may be more associated with misunderstanding of what “near normal” means and how it is visualized. In terms of visualization, white could be interpreted as the lowest certainty level of the below- or above-normal category, as opposed to an entirely separate near-normal category. To avoid this problem, Kaye et al. (2012) suggest using a pale yellow color. Independent of this study and after initial survey work was completed, CPC replaced white with gray shading for near normal on extended-range outlooks.

For long-lead outlooks, about half of survey respondents correctly interpreted gray shading as the probability of the near-normal category, with the majority of other respondents incorrectly identifying gray as a certainty of near normal, indicating a broader problem with the understanding of near normal. At the time of the study, using a grayscale for the near-normal category was a fairly new change to the suite of outlooks, which could contribute to lower understanding. Visually, this misunderstanding could originate in a few areas. One explanation is that there is a communication gap in the legend, as the placement of the different scales could imply a continuous range when in fact there are three separate scales, separated by white blocks the same size as scale values. A user could misinterpret these white blocks as being part of the



blue, gray, or red scales, a similar problem with the extended-range outlooks.

Understandability of below and above normal categories were in the range of 65%–77%, with no discernable trend across outlook types. The primary source of misunderstanding was interpreting probability ranges (e.g., 30%–40%) as a precise probability (e.g., 30%). Two potential sources of confusion are the legends and contour/color combinations. Legends, which are used in the extended-range and long-lead outlooks, are drawn in a way could be interpreted as continuous: color blocks have no gaps among them, and edges are labeled with precise probability values as opposed to ranges. In addition, contours are also not explicitly drawn or labeled as ranges. Focus group participants also thought that it would be easy for the unfamiliar user to interpret the outlooks as being proportional to the magnitude (or change in magnitude) of temperature and precipitation as opposed to the likelihood of temperature or precipitation being below, near, or above normal, which has occurred in other studies of forecast products (Hartmann et al. 2002; Pagano et al. 2002; Steinemann 2006; Wernstedt et al. 2019).

Survey and focus group results identified a group of problems associated with visual elements that contributed to difficulties in visually processing the outlooks. Climatology lines in the extended-range forecasts were frequently overlooked by respondents, and when noticed, their use was found to be difficult as they were superimposed with contour lines and state boundaries. This problem, referred to as superposition overload (Dasgupta et al. 2015), can be remedied by reducing the number of overlapping visual variables and visual tasks required of users. Respondents also identified problems with being able to read explanatory text and labels for probability contour lines and descriptive text toward the bottom of all the graphics. These are displayed in small, blocky font and are superimposed over Caribbean islands that are not part of the official outlook.

Given the identified design problems (Table 2), three modified images, tested in two rounds (round 1: Figs. 2b,c; round 2: Fig. 2d), were created to test the effectiveness of the diagnoses. The *simplified normal* image (Fig. 2b) focused on confusion of what white and gray color mapping meant by reducing the range of the grayscale to only include what is empirically feasible in the legend and adding space among the legend scales. The *discrete legend* image (Fig. 2c) introduces more noticeable modifications by decreasing the precision of the color-mapped scales to two colors each for below and above normal and one for near normal. In addition, scale color maps are labeled with ranges (e.g., 33%–50%) and qualitative probability language (e.g., “leaning below normal”).

Using test results from these two images, a final image was created using modifications from the *simplified normal* and *discrete legend* images. This visual, labeled *combined* (Fig. 2d), merged the discretized nature of the *discrete legend* image with the more precise color map of the *simplified normal* image. In addition, the legend is stacked vertically in three columns to emphasize that the scales are not separate and not continuous. All images included a more simplified background map that did not include climatology contour lines. A summary of the modifications by image and design problem is provided in Table 3.

### b. Efficacy of redesigned images

Of the 427 end users surveyed, about half were working in the agriculture, forestry, and land-management sectors. About 20% and 10% worked in emergency management and water resources, respectively. The remaining 30% of respondents were spread across the energy, government, education, and weather forecasting sectors. Of the 881 general public respondents surveyed, about half had an associate’s or bachelor’s degree as their highest level of education. About 20% had a graduate degree, and the remaining attended but did not finish college.

In the original outlook images, white color mapping in Canada and United States are assigned two different meanings: no outlook in Canada and equal chances in the United States. By a large margin (difference of 45%), end users are more adept than the public at making the distinction that white color mapping in Canada means no outlook (Fig. 3a). In the original image, about two-thirds of the public misidentifies white color mapping in Canada as meaning equal chances or near-normal outlook. The *simplified normal* and *discrete legend* modifications lead to large improvements in understandability for the public, but the increases were not significant for the end users. Both modifications provide less cluttered legends, which might serve to make clearer the meaning of white color mapping in Canada. Since Canada was removed from the *combined* modification, no question was asked about white color mapping outside the United States.

Compared to the interpretation of white color mapping in Canada, the public has a much higher understanding of white color mapping in the United States, albeit still less so than end users (Fig. 3b). The *simplified normal* and *discrete legend* modifications yield statistically significant decreases in understandability for end users and marginal decreases for the public. The *combined* modification, which removed Canada from the map and explicitly labeled white color mapping in the legend, showed statistically significant improvement in public understanding. While it is clear that removing Canada has a positive impact due to

TABLE 3. Summary of outlook modifications.

Design problem	Modifications		
	Simplified normal	Discrete legend	Combined
White color mapping in Canada and U.S.	Add more space between color maps so that white does not appear as part of the scales	Add more space between color maps so that white does not appear as part of the scales	White is explicitly labeled as equal chances and remove Canada from map
Color mapping and conceptual representation of near normal	Reduce range of near-normal color map	Reduce precision of near-normal color map and label with qualitative language and probability range	Reduce range of near-normal color map and label with qualitative language and probability range
Misinterpreting probability ranges and precise probabilities	Provide more space between near-normal and below-/above-normal color maps	Reduce precision of color maps and label with qualitative language and probability range	Rearrange direction of color maps and label with qualitative language and probability range
Climatology contour lines	Remove climatology contour lines	Remove climatology contour lines	Remove climatology contour lines
Difficulty finding and reading map annotations	—	—	Remove Caribbean islands from map, change font type/size, and reposition title

removing the ambiguity of white color mapping, there is not a clear explanation for the slight decrease in understanding for the *simplified normal* and *discrete legend* modifications in Fig. 3b. Further experiments with higher sample sizes and modified experimental design could be done to probe whether the effect is robust.

There is significant confusion among both end users and the public on the meaning of gray color mapping (Fig. 4a). The most common misinterpretation is that it corresponds to intensity: a region with gray shading *will* have near-normal temperatures as opposed to a chance of near-normal temperatures. Part of this could come from difficulty in conceptualizing near normal as

probabilistic, whereas the probability of below or above normal might have a more intuitive interpretation. In addition, previous versions of the publicly available outlooks did not map out differing probabilities of near normal. As a result, end users would be unfamiliar with interpreting gray shading, putting on them a similar level of familiarity as the general public.

The modifications, which use the legend to clarify the probabilistic meaning of near normal in relation to below and above normal, provide some improvement. Especially significant are the increases in understanding for the *discrete legend* and *combined* modifications. Both of these modifications use qualitative uncertainty

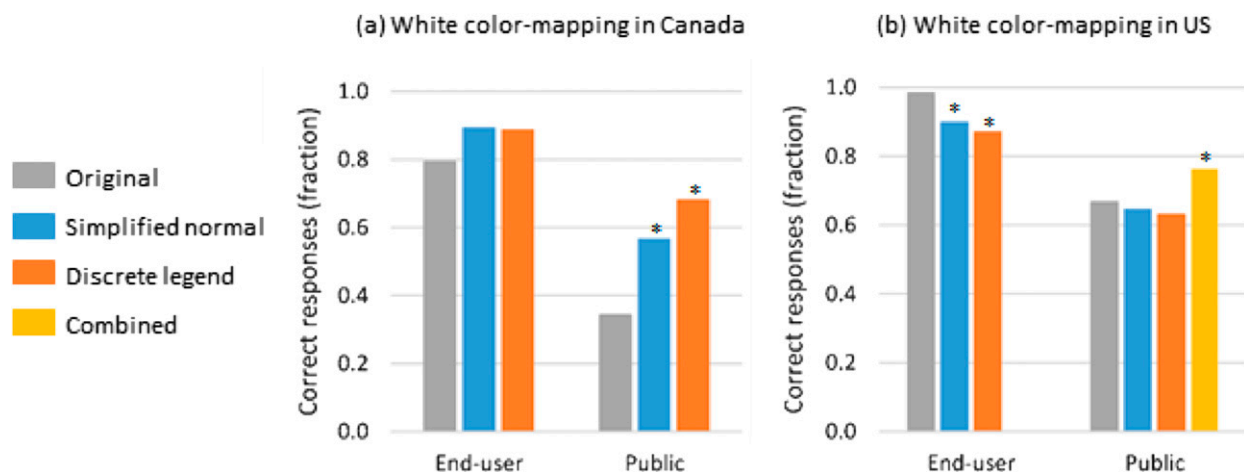


FIG. 3. Fraction of respondents to correctly interpret white color mapping in (a) Canada and (b) U.S. An asterisk (\*) indicates a significant ( $p < 0.05$ ) difference in understanding from the original.

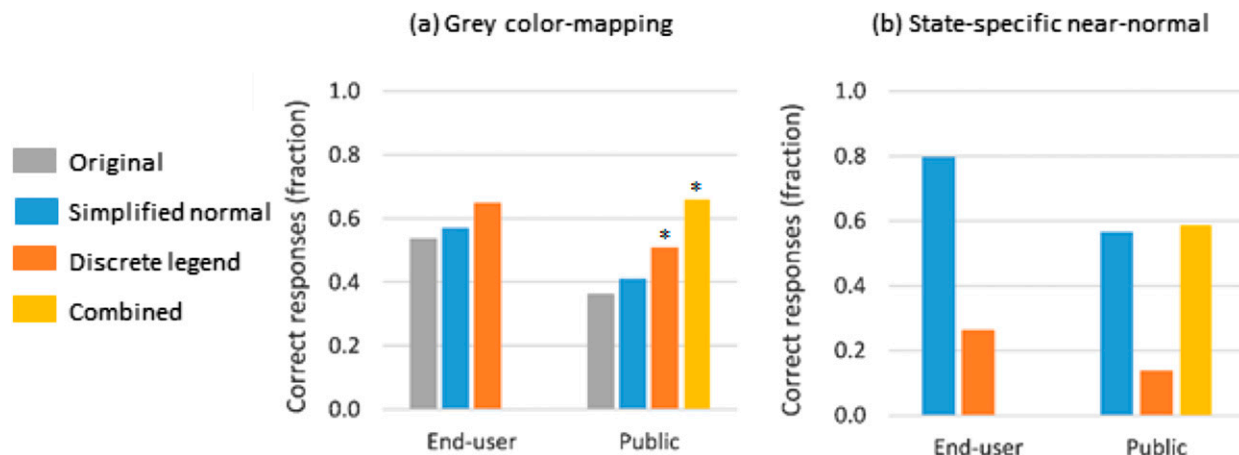


FIG. 4. Fraction of respondents to correctly interpret (a) gray color mapping and (b) near-normal category in a specific state. An asterisk (\*) indicates a significant ( $p < 0.05$ ) difference in understanding from the original.

language that might be helpful in reinforcing the probabilistic nature of the outlook.

These levels of comprehension are largely corroborated by how well respondents perform at interpreting a state-specific near-normal outlook, which was only performed on the modifications (Fig. 4b). For this task, respondents were asked to provide the outlook for the state of South Carolina. Those who saw the *simplified normal* and *combined* modifications performed fairly well at this task. However, those who saw the *discrete legend* modification (Fig. 2c) consistently misinterpreted the outlook. For this modification, fewer colors were used on its scale, requiring users to rely on contour lines to make more precise outlook readings. These results show that a vast majority of respondents used color as their primary cue for outlook interpretation, which corresponds to what the visualization literature would predict (Wolfe and Horowitz 2004, 2017).

Compared to gray shading of near normal, warm-cool shading is much better understood by both end users and the public (Ho et al. 2014) (Fig. 5). Although, as with other aspects of understandability, a gap remains between the two groups. Modifications appear to have marginal effects, with the exception of the *combined* modification, which yields statistically significant less understandability than the original. As discussed earlier, the overall higher levels of understandability of warm-cool shading than gray near-normal shading might be due to the inherent ease of conceptualizing above-below as probabilistic.

When asked to correctly interpret a below- or above-normal category in a specific state (Figs. 5b,d), a similar pattern emerges for the *discrete legend* modification. Respondents are primarily using color to guide interpretation as opposed to contour lines. The *simplified normal* modification provides improvements in

understandability, which are almost significant for above-normal outlooks and significant for below-normal outlooks.

### c. Understandability factors

In addition to questions gauging end-user and general public understanding of the outlook visualizations, the online survey asked respondents three background questions about how the temperature and precipitation probabilities were created. Unsurprisingly, end users answered the questions correctly at a greater rate: 64% of end users answered at least one question correctly compared to only 43% of the general public (Fig. 6a). Large differences are shown in understanding how many years of data make up the climate baseline, with 44% of end users answering correctly versus 9% of the general public (Fig. 6b). Similar levels, 35% and 29% respectively, of end users and the general public understood that the below-, near-, and above-normal categories were calculated by dividing the climate baseline into terciles. However, only 19% of the general public understood how locations on the outlook map are designated below, near, or above normal, versus 35% of the end-user respondents. Despite large differences in how well end user and the public answered background questions, the average number of “I don’t know” answers were remarkably similar at 1.41 and 1.49 out of 3, respectively.

A few main patterns suggest that end users and the general public interact differently with the outlook images (Table 4). Given that top-down processing of an image involves preexisting user expectations of a visualization, this is not surprising (Hegarty 2011). Foremost, background knowledge figures much more prominently for end users than the public. All of the BK

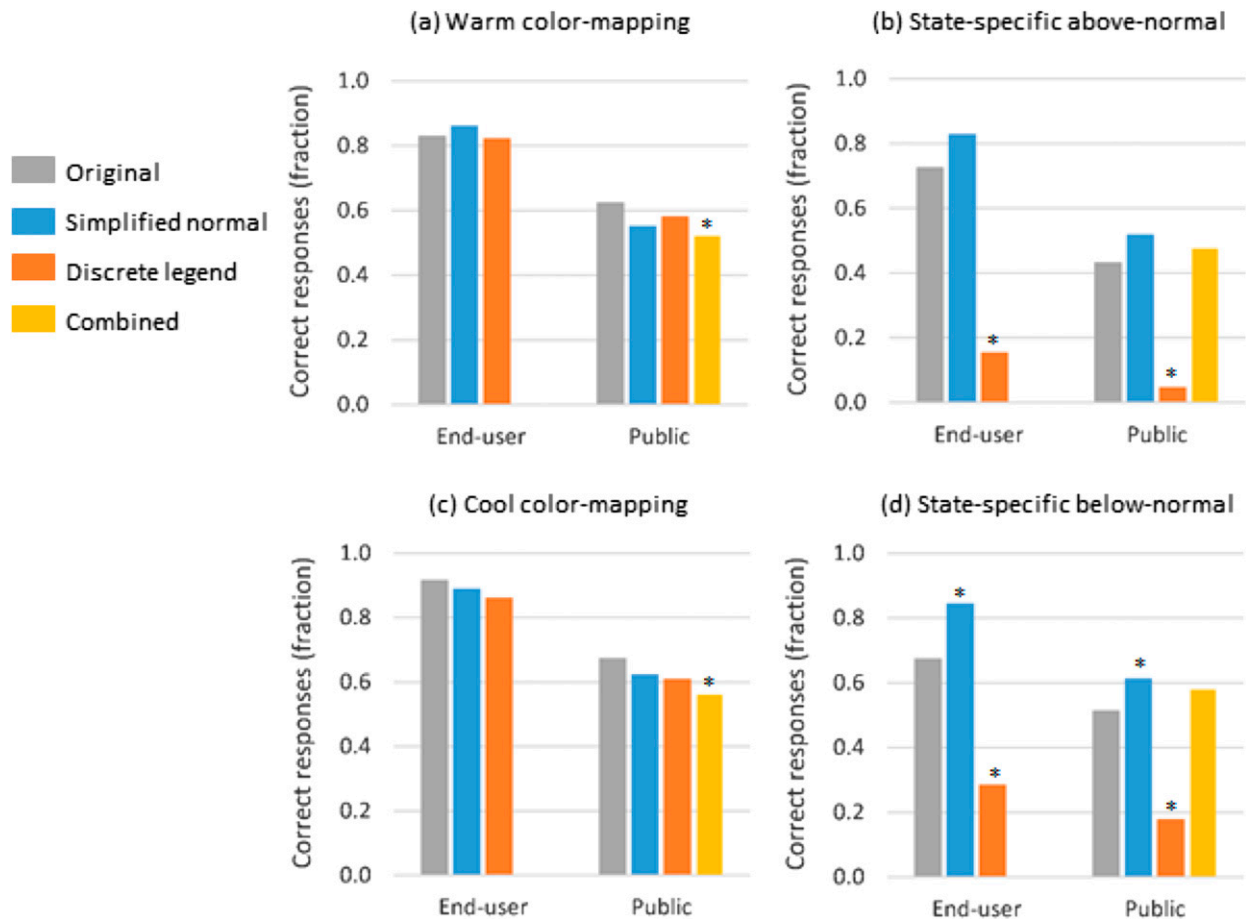


FIG. 5. Fraction of respondents to correctly interpret (a) warm color mapping, (b) above-normal category in a specific state, (c) cool color mapping, and (d) below normal category in a specific state. An asterisk (\*) indicates a significant ( $p < 0.05$ ) difference in understanding from the original.

slope coefficients ( $B_7 + B_8$ ) are positive, and all but one are statistically significant ( $p < 0.05$ ). In comparison, only one BK slope coefficient is statistically significant for the public. A positive slope coefficient indicates that more background knowledge increases the odds of understanding the outlook.

Second, uncertainty is much more of a factor for the public than for end users, and for some questions it appears to interact with background knowledge individually and through the interaction term. Even though many of the public slope coefficients are not individually significant, the ones included in the table are jointly significant, meaning that in total their interaction is statistically significant. This is reinforced by a large difference in the Spearman correlation coefficients  $\rho$  between BK and UNC, with  $\rho$  for end users and the public of  $-0.83$  and  $-0.60$ , respectively. This indicates a weaker relationship between self-assessed and actual ability for the general public; a correlation of  $-1$  would indicate perfect self-assessment. It is noteworthy

that this entanglement of BK and UNC also shows up for end users when answering a question about gray color mapping. This is the question with the lowest understandability, presumably because of lack of familiarity with using a scale for the near-normal category.

## 5. Discussion

Our results show the efficacy of image modifications based on the visualization diagnosis literature (Dasgupta et al. 2015). Of the diagnoses tested (first three rows in Table 2), all were shown to affect understandability. For example, visual variable ambiguity is one of the most serious visualization problems because users might misinterpret or confuse one variable for another. Our results show this occurs for white color mapping in Canada and the United States, and that removing ambiguity has a large effect on improving understandability (Fig. 3) for the general public. Similarly, the literature predicts that confusion might occur in



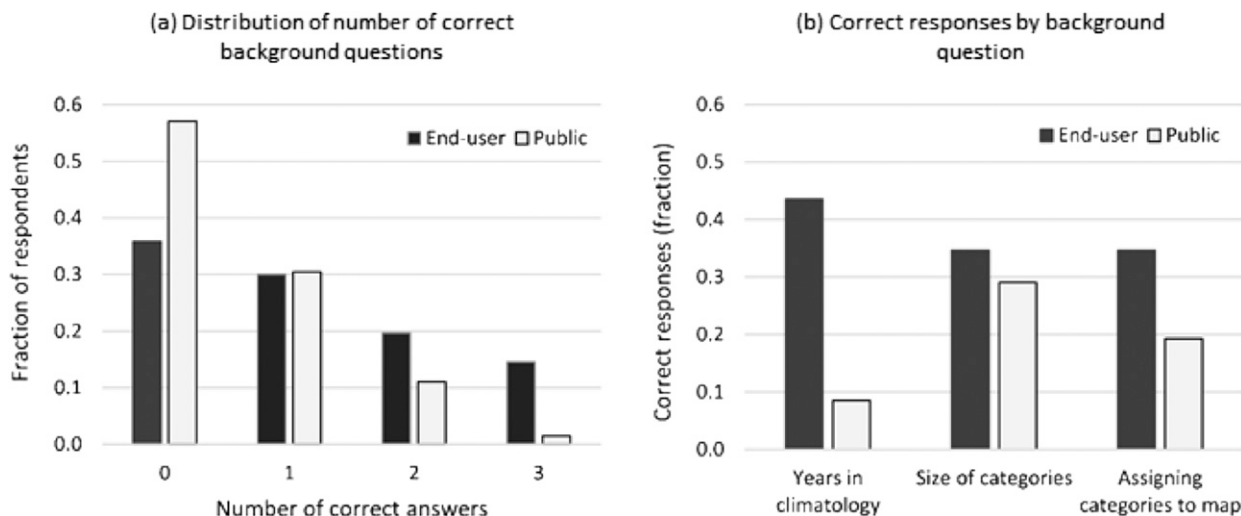


FIG. 6. Comparison of end-user and general public responses to outlook background questions. (a) Distribution of number of correct responses. (b) Fraction of correct answers by specific background question.

interpreting the below-, near-, and above-normal color maps because there is a communication gap resulting from them being aligned in a way that might imply a single linear scale and precise increments in probability instead of probability ranges. Redesigning the legends to emphasize that near normal and below/above normal are separate scales of probability ranges instead of a single continuous probability scale improved end-user and general public understandability (Figs. 4a and 5b,d). Both of these modifications, as well as all others in Table 3, were collectively prototyped and tested in the combined image (Fig. 2d), which resulted in the most comprehensive improvement in understandability of the modifications tested (Fig. 2).

The difference in how end users and the general public understand the outlook images highlights the dual nature of how users comprehend an image. The visualization science literature emphasizes that images are easier (harder) to understand if viewer expectations

of the visualization converge (diverge) (Hegarty 2011; Harold et al. 2016). This phenomenon occurs because viewing and interpreting an image is an iterative process that involves (i) attraction of visual attention and (ii) comprehension of viewed information. The significance of these coupled processes is that many avenues may exist for improving a user’s scientific visual understanding and that the most efficient and/or effective choice may depend on the characteristics of the user. In this study, we have primarily focused on modifying images, as opposed to improving visiospatial ability or user background knowledge. Nevertheless, our results intersect with these user-centric properties.

For example, improvement in public understanding of white color mapping in the United States is slightly offset by a decrease in end-user understanding, which could be tied to end users’ more entrenched expectations of how an outlook legend is structured. Some user attributes, however, are more general as both end

TABLE 4. Slope coefficients for background knowledge (BK) and uncertainty (UNC). Methods for populating table are listed in appendix B. Blank cells indicate that slope coefficient was not necessary for interpreting results. Asterisks indicate statistical significance: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

Modifications	Public			End user		
	BK	UNC	BK × UNC	BK	UNC	BK × UNC
White color mapping in Canada		0.28***		0.82***		
White color mapping in U.S.	0.18		0.35*	0.84*		0.14
Gray color mapping				0.67***	0.34*	
Warm color mapping	0.14	0.11	0.23	1.0***		-0.32
Cool color mapping		0.18*	0.32*	0.48		0.10
Above normal	0.15	0.14	0.34*	0.96***		-0.10
Below normal	0.38***	0.20*		0.43*		
Equal chances	0.15	0.15	0.24	1.1*		-0.48

users and the general public had poor performance interpreting below, near, and above normal questions for the discretized legend modification. This outcome is linked to so-called Gestalt laws, where users perceive differences in color as primary to other differences denoted by contour lines or other grouping types such as shape or size (Wolfe and Horowitz 2004, 2017).

While this research reinforces common advice from the literature, such as not trying to explain too many patterns in one image and matching the number visual and data variables, there is a need for more experimental studies. One direction more directly addresses the issue of whether blue, gray, and red are the most effective color choices for below, near, and above normal. There are reasons to believe a different hue such as yellow might better convey the three category nature of the outlooks given that gray and color saturation is frequently associated with overall uncertainty (Kaye et al. 2012). The number of discrete categories to represent uncertainty is also an important issue as it affects the complexity of the visualization (Kinkeldey et al. 2014). CPC has recently created an experimental two-category 3–4-week outlook that does not include near normal in the visualization. It is an open question as to whether a change in understandability would occur from applying this format to other forecast products.

In addition, studies that focus on nonvisual elements are needed, including accompanying captions, key points, and text. Research on these factors can be accomplished by isolating and testing (i) specific modifications based on the visualization literature and (ii) modifications based on visiospatial ability and user expectations. The latter can be especially crucial for legacy scientific visualizations, such as the NOAA CPC outlooks and other government data products, where users have long-standing expectations and experience that would need to be considered when redesigning such high-profile graphics.

Ultimately, continued research along these lines can serve to produce more comprehensive diagnostic guidance for combinations of visual and associated text-based changes. This guidance can be important in cases, such as gray color mapping and near normal, where visual modifications lead to improvements, but additional gains in understanding are potentially left unrealized due to unchanged nonvisual elements. Additionally, expanding visualization research to dynamic or user-controlled graphics or decision support systems would address improved graphics communication across a diversity of platforms. This work demonstrates the importance of experimental coproduced decision support research to improve understandability and better support the use of evidence in decision-making.

*Acknowledgments.* This research was supported by NOAA Climate Prediction Center via Grant NA14NES4320003 [Cooperative Institute for Climate and Satellites (CICS)] at the University of Maryland/ESSIC. Data collection support was provided by Riley Cassidy, Dean Sproul, Jason Winik, Candela Cerpa, Natalia Jaffee, Samantha Ammons, and Greta Easthom.

## APPENDIX A

### Sample Size Determination

Table A1 lists sample sizes used in phase 2 online survey testing. Using power analysis, sample sizes were chosen to be able to detect changes in understandability greater than 0.15 with a  $p < 0.05$ . Therefore, small changes in understandability might be significant if a larger sample size is used.

## APPENDIX B

### Statistical Regression Analysis

Tables B1 and B2 show the slope coefficients for the logit regression in Eq. (1). Table B1 shows the version of the equation with interaction terms estimated. The last column indicates whether the interaction terms are jointly significant ( $p < 0.05$ ). Table B2 shows slope coefficient estimates of a version of Eq. (1) without interaction terms.

Table 4 in the main text, which summarizes the effect of background understanding and uncertainty on the probability of answering a visualization question correctly, is populated with a mix of coefficients from Tables B1 and B2. If the interaction terms are not jointly significant or one of the terms is not individually significant, then slope coefficients from Table B2 are used. Coefficients with  $p > 0.05$  are left blank. Otherwise, the slope coefficients are from Table B1. Both interaction terms are included if they are jointly significant, even if they are individual not significant. For rows where interaction terms are shown, non-interaction terms are shown if they are significant in Table B1, or if they are significant in Table B2. We include

TABLE A1. Sample size by modification.

	Public	End user
Original	225	58
Discrete legend	222	184
Simplified normal	211	185
Combined	223	—
Total	881	427

TABLE B1. Slope coefficients for Eq. (1) with interaction terms. Asterisks indicate statistical significance: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

Modifications	Understanding		Uncertainty		Interaction		Joint $p$ value
	Public	End user	Public	End user	Public	End user	
White color mapping in Canada	-0.12	0.84***	0.26***	0.24	0.091	-0.037	
White color mapping in U.S.	0.18	0.84*	0.060	-0.0080	0.35*	0.14	*
Gray color mapping	0.081	0.67***	0.083	0.35*	0.17	-0.0080	
Warm color mapping	0.14	1.0***	0.11	0.24	0.23	-0.32	*
Cool color mapping	0.088	0.48	0.18*	-0.22	0.32*	0.10	*
Above normal	0.15	0.96***	0.14	0.27	0.34*	-0.10	*
Below normal	0.32*	0.53*	0.17*	0.14	0.098	-0.25	
Equal chances	0.15	1.1*	0.15	0.071	0.24	-0.48	*

TABLE B2. Slope coefficients for Eq. (1) without interaction terms. Asterisks indicate statistical significance: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

Modifications	Understanding		Uncertainty	
	Public	End user	Public	End user
White color mapping in Canada	-0.055	0.82***	0.28***	0.23
White color mapping in U.S.	0.39***	0.94***	0.13	0.024
Gray color mapping	0.19	0.67***	0.12	0.34*
Warm color mapping	0.28*	0.77***	0.16*	0.14
Cool color mapping	0.28*	0.55	0.25***	-0.20
Above normal	0.36***	0.91***	0.22***	0.24
Below normal	0.38***	0.43*	0.20*	0.047
Equal chances	0.29*	0.65*	0.20***	-0.074

terms that are significant in Table B2 to account for joint significance among noninteraction and interaction terms.

REFERENCES

Atkinson, R., and J. Flint, 2004: Snowball sampling. *The SAGE Encyclopedia of Social Science Research Methods*, M. S. Lewis-Black, A. Bryman, and T. F. Liao, Eds., SAGE Publications, <https://doi.org/10.4135/9781412950589.n931>.

Barnston, A. G., and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.*, **75**, 2097–2114, [https://doi.org/10.1175/1520-0477\(1994\)075<2097:LLSFDW>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<2097:LLSFDW>2.0.CO;2).

—, A. Leetmaa, V. E. Kousky, R. E. Livezey, E. A. O’Lenic, H. Van den Dool, A. J. Wagner, and D. A. Unger, 1999: NCEP forecasts of the El Niño of 1997–98 and its U.S. impacts. *Bull. Amer. Meteor. Soc.*, **80**, 1829–1852, [https://doi.org/10.1175/1520-0477\(1999\)080<1829:NFOTEN>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<1829:NFOTEN>2.0.CO;2).

Bertin, J., 1983: *Semiology of Graphics: Diagrams, Networks, Maps*. Esri Press, 456 pp.

Borland, D., and R. M. Taylor, 2007: Rainbow color map (still) considered harmful. *IEEE Comput. Graph. Appl.*, **27**, 14–17, <https://doi.org/10.1109/MCG.2007.323435>.

Braman, L. M., M. K. van Aalst, S. J. Mason, P. Suarez, Y. Ait-Chellouche, and A. Tall, 2013: Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008. *Disasters*, **37**, 144–164, <https://doi.org/10.1111/j.1467-7717.2012.01297.x>.

Buttenfield, B. P., and M. K. Beard, 1994: Graphical and geographical components of data quality. *Visualization in Geographic Information Systems*, D. Unwin and H. Hearnsaw, Eds., John Wiley and Sons, 150–157.

Callahan, B., E. Miles, and D. Fluharty, 1999: Policy implications of climate forecasts for water resources management in the Pacific Northwest. *Policy Sci.*, **32**, 269–293, <https://doi.org/10.1023/A:1004604805647>.

Changnon, S. A., and D. Vonnahme, 1986: Use of climate predictions to decide water management problems. *J. Amer. Water Resour. Assoc.*, **22**, 649–652, <https://doi.org/10.1111/j.1752-1688.1986.tb01919.x>.

—, J. M. Changnon, and D. Changnon, 1995: Uses and applications of climate forecasts for power utilities. *Bull. Amer. Meteor. Soc.*, **76**, 711–720, [https://doi.org/10.1175/1520-0477\(1995\)076<0711:UAAOCF>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0711:UAAOCF>2.0.CO;2).

Clements, J., A. Ray, and G. Anderson, 2013: The value of climate services across economic and public sectors: A review of relevant literature. USAID, 43 pp., [http://www.climate-services.org/wp-content/uploads/2015/09/CCRD-Climate-Services-Value-Report\\_FINAL.pdf](http://www.climate-services.org/wp-content/uploads/2015/09/CCRD-Climate-Services-Value-Report_FINAL.pdf).

Dasgupta, A., J. Poco, Y. Wei, R. Cook, E. Bertini, and C. T. Silva, 2015: Bridging theory with practice: An exploratory study of visualization use and design for climate model comparison. *IEEE Trans. Visualization Comput. Graphics*, **21**, 996–1014, <https://doi.org/10.1109/TVCG.2015.2413774>.

—, —, B. Rogowitz, K. Han, E. Bertini, and C. T. Silva, 2019: Effect of color scales on climate scientists’ objective and subjective performance in spatial data analysis tasks. *IEEE Trans. Visualization Comput. Graphics*, <https://doi.org/10.1109/TVCG.2018.2876539>, in press.

Dilling, L., and M. C. Lemos, 2011: Creating usable science. *Global Environ. Change*, **21**, 680–689, <https://doi.org/10.1016/j.gloenvcha.2010.11.006>.

Edwards, L. D., and E. S. Nelson, 2001: Visualizing data certainty: A case study using graduated circle maps. *Cartogr. Perspect.*, **38**, 19–36, <https://doi.org/10.14714/CP38.793>.

- Foley, A. M., P. G. Leahy, A. Marvulgia, and E. J. McKeogh, 2012: Current methods and advances in forecasting of wind power generation. *Renewable Energy*, **37**, 1–8, <https://doi.org/10.1016/j.renene.2011.05.033>.
- Gilbert, C. D., and W. Li, 2013: Top-down influences on visual processing. *Nat. Rev. Neurosci.*, **14**, 350–363, <https://doi.org/10.1038/nrn3476>.
- Harold, J., I. Lorenzoni, T. F. Shipley, and K. R. Coventry, 2016: Cognitive and psychological science insights to improve climate change data visualization. *Nat. Climate Change*, **6**, 1080–1089, <https://doi.org/10.1038/nclimate3162>.
- Hartmann, H. C., T. C. Pagano, S. Sorooshian, and R. Bales, 2002: Confidence builders: Evaluating seasonal climate forecasts from user perspectives. *Bull. Amer. Meteor. Soc.*, **83**, 683–698, [https://doi.org/10.1175/1520-0477\(2002\)083<0683:CBESCF>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2).
- Hegarty, M., 2011: The cognitive science of visual-spatial displays: Implications for design. *Top. Cognit. Sci.*, **3**, 446–474, <https://doi.org/10.1111/j.1756-8765.2011.01150.x>.
- , H. S. Smallman, A. T. Stull, and M. S. Canham, 2009: Naïve cartography: How intuitions about display configuration can hurt performance. *Cartographica*, **44**, 171–186, <https://doi.org/10.3138/carto.44.3.171>.
- , M. S. Canham, and S. I. Fabrikant, 2010: Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *J. Exp. Psychol. Learn. Mem. Cogn.*, **36**, 37–53, <https://doi.org/10.1037/a0017683>.
- Ho, H. N., G. H. Van Doorn, T. Kawabe, J. Watanabe, and C. Spence, 2014: Colour-temperature correspondences: when reactions to thermal stimuli are influenced by colour. *PLOS ONE*, **9**, e91854, <https://doi.org/10.1371/journal.pone.0091854>.
- Kaye, N. R., A. Hartley, and D. Hemming, 2012: Mapping the climate: Guidance on appropriate techniques to map climate variables and their uncertainty. *Geosci. Model Dev.*, **5**, 245–256, <https://doi.org/10.5194/GMD-5-245-2012>.
- Kinkeldey, C., A. M. MacEachren, and J. Schiewe, 2014: How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualization user studies. *Cartogr. J.*, **51**, 372–386, <https://doi.org/10.1179/1743277414Y.0000000099>.
- Krueger, R. A., and M. A. Casey, 2015: *Focus Groups: A Practical Guide for Applied Research*. Sage Publications, 280 pp.
- Lam, H., E. Bertini, P. Isenber, C. Plaisant, and S. Carpendale, 2012: Empirical studies in information visualization: Seven scenarios. *IEEE Trans. Visualization Comput. Graphics*, **18**, 1520–1536, <https://doi.org/10.1109/TVCG.2011.279>.
- Leitner, M., and B. P. Battenfield, 2000: Guidelines for the display of attribute certainty. *Cartogr. Geogr. Inf. Sci.*, **27**, 3–14, <https://doi.org/10.1559/152304000783548037>.
- Lemos, M. C., C. J. Kirchoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nat. Climate Change*, **2**, 789–794, <https://doi.org/10.1038/nclimate1614>.
- Livezey, R. E., and M. M. Timofeyeva, 2008: The first decade of long-lead U.S. seasonal forecasts. *Bull. Amer. Meteor. Soc.*, **89**, 843–854, <https://doi.org/10.1175/2008BAMS2488.1>.
- Lowrey, J. L., A. J. Ray, and R. S. Webb, 2009: Factors influencing the use of climate information by Colorado municipal water managers. *Climate Res.*, **40**, 103–119, <https://doi.org/10.3354/cr00827>.
- MacEachren, A. M., 1992: Visualizing uncertain information. *Cartogr. Perspect.*, **13**, 10–19, <https://doi.org/10.14714/CP13.1000>.
- , 2004: *How Maps Work: Representation, Visualization, and Design*. The Guilford Press, 513 pp.
- , C. A. Brewer, and L. W. Pickle, 1998: Visualizing georeferenced data: Representing reliability of health statistics. *Environ. Plann.*, **30A**, 1547–1561, <https://doi.org/10.1068/a301547>.
- , A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, 2005: Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartogr. Geogr. Inf. Sci.*, **32**, 139–160, <https://doi.org/10.1559/1523040054738936>.
- McMahon, R., M. Stauffacher, and R. Knutti, 2015: The unseen uncertainties in climate change: Reviewing comprehension of an IPCC scenario graph. *Climatic Change*, **133**, 141–154, <https://doi.org/10.1007/s10584-015-1473-4>.
- Meadow, A. M., D. B. Ferguson, Z. Guido, A. Horangic, G. Owen, and T. Wall, 2015: Moving toward the deliberate coproduction of climate science knowledge. *Wea. Climate Soc.*, **7**, 179–191, <https://doi.org/10.1175/WCAS-D-14-00050.1>.
- Morrison, J. L., 1984: Applied cartographic communication: Map symbolization for atlases. *Cartographica*, **21**, 44–84, <https://doi.org/10.3138/X43X-4479-4G34-J674>.
- NWS, 2018: Climate outlooks. NOAA/National Weather Service Instruction 10-1001, 53 pp., <http://www.nws.noaa.gov/directives/sym/pd01010001curr.pdf>.
- O’Lenic, E. A., D. A. Unger, M. S. Halpert, and K. S. Pelman, 2008: Developments in operational long-range climate prediction at CPC. *Wea. Forecasting*, **23**, 496–515, <https://doi.org/10.1175/2007WAF2007042.1>.
- Pagano, T. C., H. C. Hartmann, and S. Sorooshian, 2001: Using climate forecasts for water management. *J. Amer. Water Resour. Assoc.*, **37**, 1139–1153, <https://doi.org/10.1111/j.1752-1688.2001.tb03628.x>.
- , —, and —, 2002: Factors affecting seasonal forecast use in Arizona water management. *Climate Res.*, **21**, 259–269, <https://doi.org/10.3354/CR021259>.
- Pang, A. T., C. M. Wittenbrink, and S. K. Lodha, 1997: Approaches to uncertainty visualization. *Visual Comput.*, **13**, 370–390, <https://doi.org/10.1007/s003710050111>.
- Potter, K., P. Rosen, and C. R. Johnson, 2012: From quantification to visualization: a taxonomy of uncertainty visualization approaches. *Working Conf. on Uncertainty Quantification in Scientific Computing 2011*, Boulder, CO, International Federation for Information Processing, 226–249.
- Pulwarty, R. S., and K. T. Redmond, 1997: Climate and salmon restoration in the Columbia River Basin: The role and usability of seasonal forecasts. *Bull. Amer. Meteor. Soc.*, **78**, 381–398, [https://doi.org/10.1175/1520-0477\(1997\)078<0381: CASRIT>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0381: CASRIT>2.0.CO;2).
- Quinan, P. S., and M. Meyer, 2016: Visually comparing weather features in forecasts. *IEEE Trans. Visualization Comput. Graphics*, **22**, 389–398, <https://doi.org/10.1109/TVCG.2015.2467754>.
- Rautenhaus, M., M. Böttinger, S. Siemen, R. Hoffman, R. M. Kirby, M. Mirzargar, N. Röber, and R. Westermann, 2018: Visualization in meteorology—A survey of techniques and tools for data analysis tasks. *IEEE Trans. Visualization Comput. Graphics*, **24**, 3268–3292, <https://doi.org/10.1109/TVCG.2017.2779501>.
- Ray, A. J., and R. S. Webb, 2016: Understanding the user context: Decision calendars as frameworks for linking climate to policy, planning, and decision-making. *Climate in Context: Science and Society Partnering for Adaptation*, A. S. Parris et al., Eds., John Wiley and Sons, 27–50.
- Rayner, S., D. Lach, and H. Ingram, 2005: Weather forecasts are for wimps. *Climatic Change*, **69**, 197–227, <https://doi.org/10.1007/s10584-005-3148-z>.



- Retchless, D. P., and C. A. Brewer, 2016: Guidance for representing uncertainty on global temperature change maps. *Int. J. Climatol.*, **36**, 1143–1159, <https://doi.org/10.1002/JOC.4408>.
- Rosenholtz, R., Y. Li, and L. Nakano, 2007: Measuring visual clutter. *J. Vision*, **7**, 17, <https://doi.org/10.1167/7.2.17>.
- Roulston, M. S., D. T. Kaplan, J. Hardenberg, and L. A. Smith, 2003: Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy*, **28**, 585–602, [https://doi.org/10.1016/S0960-1481\(02\)00054-x](https://doi.org/10.1016/S0960-1481(02)00054-x).
- Schweitzer, D. M., and M. F. Goodchild, 1992: Data quality and choropleth maps: An experiment with the use of color. *Proc. GIS/LIS '92*, San Jose, CA, ACSM and ASPRS, 686–699.
- Simpson, C. F., L. Dilling, K. Dow, K. J. Lackstrom, M. C. Lemos, and R. E. Riley, 2016: Assessing needs and decision contexts: RISA approaches to engagement research. *Climate in Context: Science and Society Partnering for Adaptation*, A. S. Parris et al., Eds., John Wiley and Sons, 3–26.
- Sonka, S. T., S. Changnon, and S. L. Hofing, 1992: How agribusiness uses climate predictions: Implications for climate research and provision of predictions. *Bull. Amer. Meteor. Soc.*, **73**, 1999–2009, [https://doi.org/10.1175/1520-0477\(1992\)073<1999:HAUCPI>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<1999:HAUCPI>2.0.CO;2).
- Stauffer, R., G. J. Mayr, M. Dabernig, and A. Zeileis, 2015: Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations. *Bull. Amer. Meteor. Soc.*, **96**, 203–216, <https://doi.org/10.1175/BAMS-D-13-00155.1>.
- Steinemann, A. C., 2006: Using climate forecasts for drought management. *J. Appl. Meteor. Climatol.*, **45**, 1353–1361, <https://doi.org/10.1175/JAM2401.1>.
- Tadesse, T., T. Haigh, N. Wall, A. Shiferaw, B. Zaitchik, S. Beyene, G. Berhan, and J. Petr, 2016: Linking seasonal predictions to decision-making and disaster management in the Greater Horn of Africa. *Bull. Amer. Meteor. Soc.*, **97**, ES89–ES92, <https://doi.org/10.1175/BAMS-D-15-00269.1>.
- Thomson, J., B. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, 2005: Typology for visualizing uncertainty. *Proc. Conf. Visualization Data Analysis 2005*, San Jose, CA, Institute of Electrical and Electronics Engineers, 16–20.
- Wernstedt, K., P. S. Roberts, J. Arvai, and K. Redmond, 2019: How emergency managers (mis?)interpret forecasts. *Disasters*, **43**, 88–109, <https://doi.org/10.1111/disa.12293>.
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- Wolfe, J. M., and T. S. Horowitz, 2004: What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.*, **5**, 495–501, <https://doi.org/10.1038/nrn1411>.
- , and —, 2017: Five factors that guide attention in visual search. *Nat. Hum. Behav.*, **1**, 0058, <https://doi.org/10.1038/s41562-017-0058>.
- Wooldridge, J. M., 2016: *Introductory Econometrics: A Modern Approach*. Cengage Learning, 912 pp.