BAMS Article

Users' Cognitive Load

A Key Aspect to Successfully Communicate Visual Climate Information

Luz Calvo, Isadora Christel, Marta Terrado, Fernando Cucchietti, and Mario Pérez-Montoro

ABSTRACT: The visual communication of climate information is one of the cornerstones of climate services. It often requires the translation of multidimensional data to visual channels by combining colors, distances, angles, and glyph sizes. However, visualizations including too many layers of complexity can hinder decision-making processes by limiting the cognitive capacity of users, therefore affecting their attention, recognition, and working memory. Methodologies grounded on the fields of user-centered design, user interaction, and cognitive psychology, which are based on the needs of the users, have a lot to contribute to the climate data visualization field. Here, we apply these methodologies to the redesign of an existing climate service tool tailored to the wind energy sector. We quantify the effect of the redesign on the users' experience performing typical daily tasks, using both quantitative and qualitative indicators that include response time, success ratios, eye-tracking measures, user perceived effort, and comments, among others. Changes in the visual encoding of uncertainty and the use of interactive elements in the redesigned tool reduced the users' response time by half, significantly improved success ratios, and eased decision-making by filtering nonrelevant information. Our results show that the application of user-centered design, interaction, and cognitive aspects to the design of climate information visualizations reduces the cognitive load of users during tasks performance, thus improving user experience. These aspects are key to successfully communicating climate information in a clearer and more accessible way, making it more understandable for both technical and nontechnical audiences.

KEYWORDS: Forecasting; Seasonal forecasting; Decision support; Software; Model interpretation and visualization

https://doi.org/10.1175/BAMS-D-20-0166.1
Corresponding author: Luz Calvo, luz.calvo@bsc.es
In final form 8 July 2021

AFFILIATIONS: Calvo, Christel, Terrado, and Cucchietti—Barcelona Supercomputing Center, Barcelona, Spain; **Pérez-Montoro**—Faculty of Information and Audiovisual Media, Universitat de Barcelona, Barcelona, Spain

The accessibility to climate information has implications for how society makes the best use of scientific knowledge to adapt to climate change (Harold et al. 2016). For years, climate service providers have faced the challenge of how to best communicate climate-related data and information, together with their inherent uncertainty, in an easy and understandable way for both expert and nonexpert users (Kaye et al. 2012; Lorenz et al. 2015). This has resulted in service providers often using visual representations to supply climate knowledge tailored for users' decision-making (Gerst et al. 2020; Taylor et al. 2015).

As a standardized mapping approach to represent climate uncertainty is lacking, different techniques have been applied for this purpose, often resulting in users spending more time trying to understand the mapping approach than focusing on the interpretation of the presented information itself (Kaye et al. 2012). In general, users' familiarity with a type of data visualization has been found to play a significant role in the process of reading and making sense of maps and graphs (Lorenz et al. 2015). Commonly used representations of climate data include choropleths, heat maps, and line charts (Taylor et al. 2015). Although familiar to users, these elements have limitations when used to communicate climate data to decision-makers in a way that is transparent, understandable, and that does not lead to a false sense of certainty. An example of such limitations is seen when communicating climate predictions from the next two weeks up to a few decades into the future. A characteristic of such predictions is that they are probabilistic, meaning they provide information on the probability of a certain climate outcome to occur (e.g., winds below or above a threshold, not optimal for the energy production). In addition, climate predictions are often given in the form of large amounts of data covering the whole globe, and their quality (i.e., level of success of a prediction against observationally based information) depends on the specific location and time (Kaye et al. 2012). Both aspects, probabilities and forecasts quality (referred to as skill by the climate science community), add complexity to the visual communication and can eventually compromise the understanding of climate predictions by users (Bonneau et al. 2014; Terrado et al. 2019).

Visualizations of complex climate data tend to prioritize solutions that take into account the greatest combination of variables and dimensions, e.g., color, size, distance or brightness of the glyphs (the graphic symbols used to represent a value). This complexity calls for formats paying special attention to visual encoding, which encompasses the translation of multidimensional data into visual elements on a chart or map representation. Visual encoding is useful in the sense that it allows to convey a higher amount of information in a single visualization (Grainger et al. 2016; Lloyd 1997). However, it rarely considers if the information needs to be displayed all at a time, with a certain visual aesthetic, or if it will be too complex for its proper interpretation (Cleveland and McGill 1985; Sager et al. 2007). Indeed, aesthetics might be worth considering if striving to create something memorable, that helps to raise awareness about a specific scientific challenge (Borkin et al. 2013). However, an attractive image cannot qualify as effective unless it accurately conveys something meaningful or credible (Holmes 1984; Kosara 2013).

According to Stephens et al. (2012), for a visualization to be effective, it is important to consider a balance between density (the amount of data represented), robustness (the representation of scientific confidence and consensus), and saliency (the relevance of the information to user needs). Although visualizing climate forecast uncertainties and associated probabilities has been thought to increase users' trust (Joslyn and LeClerc 2011;

Roulston et al. 2006), it does not automatically lead to better decisions (Greis et al. 2015). This is especially critical when the visual elements used to represent uncertainty compete with the limited cognitive resources of the users (i.e., their attention, recognition, and working memory) (Antifakos et al. 2004; Riveiro et al. 2014; Davis et al. 2020). Visualizations with a heavy cognitive load, that is, involving a high amount of working memory resources, can have negative effects on users' understanding and learning and can impact their ability to complete a task or make an informed decision (Cairo 2012; Few 2009; McInerny et al. 2014). In such cases, the use of interactive elements can offer mechanisms for progressively dosing the information that is to be shown (Veras Guimarães 2019; Bertin 2010; Ware et al. 2002; Yoghourdjian et al. 2018).

Methodologies grounded on the fields of user-centered design (UCD), user interaction, and cognitive psychology have a lot to contribute to the climate data visualization field (Christel et al. 2018; Bevington et al. 2019). UCD techniques involve users throughout the design process in order to create highly usable visualization tools based on their needs (Davis et al. 2020; Dong et al. 2008; Yucong et al. 2019). User interaction makes use of interactive elements to let users decide what to see, when, or to show only those values that are relevant for a given task. This facilitates decision-making, returns control to the users, and allows them to discard the nonrelevant information at each moment (Lau and Vande Moere 2007; Gerharz and Pebesma 2009; Ware 2012). Within the framework of UCD methodologies, technologies such as the eye-tracker have been used to quantify and analyze visual patterns, attention, and cognitive aspects. Adding cognition and perception (i.e., processes about how humans acquire knowledge, understanding and interpretation) can help detect and solve initial design problems in UCD. Rather than just favoring visual exploration, these disciplines offer increasingly effective methods to develop and evaluate visualization systems that explicitly consider real-world user requirements (Block 2013).

Tools and learnings from the UCD field have already been applied to the visualization of climate information and climate services. For instance, Argyle et al. (2017) showed how incorporating usability evaluation into the design of decision support tools can improve the efficiency, effectiveness, and user experience of a weather forecasting application. Other studies have similarly applied UCD to inform the design of climate information websites, apps, and prototypes (Ling et al. 2015; Oakley and Daudert 2016; Khamaj et al. 2019). Design elements have also been introduced in the development of climate services to increase their usability, in particular for the renewable energy sector (Christel et al. 2018). On the other hand, cognitive and psychological sciences have been applied to the visualization of climate data. Some examples are the use of cognitive psychology methods to help make information provided by IPCC graphs more accessible to expert and nonexpert audiences (Harold et al. 2016) and improve users' task performance (Hegarty et al. 2010). Differences in the interpretation of climate graphs between experienced and nonexperienced users have been explored elsewhere (Atkins and Mcneal 2018; Gerst et al. 2020), both for climate change variables and for temperature and precipitation outlooks.

The aim of this paper is to provide quantitative and qualitative evidence of how the use of user-centered design, visualization techniques, and interaction elements can reduce the cognitive load of nonexpert users during tasks' performance. We compare two different map visualizations of uncertainty. First, the one used in Project Ukko, a climate service tool prototype that provided climate predictions tailored to the energy sector (http://project-ukko.net/). The second visualization was a redesigned version of Project Ukko, using a simplified visual encoding of uncertainty and the use of interactive elements. We quantified the effects of the redesign on the users' experience performing typical daily tasks, using indicators that include response time, success ratios, eye-tracking measures, and user perceived effort and comments, among others. We show that involving experts from different disciplines [climate experts, user experience (UX) and visualization experts, communicators] in the co-production process and simplifying the visual encoding used in the visualization, has an impact on the users' cognitive load, favoring response times and confidence in decision-making.

Study context: Previous work with experts

Project Ukko introduced design to explore new forms of representation for wind predictions, which was considered a groundbreaking step in the visualization of climate services (Christel et al. 2016). The representation of uncertainty in Project Ukko was achieved through multiple visual resources, namely, glyph thickness representing the intensity (speed) of predicted wind, glyph inclination representing the predicted change in wind conditions (i.e., higher-than-normal, normal, or lower-than-normal wind), and opacity representing the quality of the prediction (skill).

After the development of Project Ukko, a user test was carried out in order to detect functional or usability problems. The test identified a series of conflicting aspects in the tool, mainly related to visual encoding (Makri 2015) and flexibility of use, that needed improvement. These aspects included the use of too many categories for each represented variable, the difficulties to detect color hue associated with narrow glyphs, and the excess of information in nonrelevant areas of the map.

These findings indicated that the development of a new visualization was needed in order to solve the problems detected. For informing this new redesign of Project Ukko, we started by organizing a workshop to gather expert user needs and limitations experienced while using the tool. We also conducted four interviews with target users, including climate experts and operators and managers of wind power plants, to obtain specific behavioral information related to their daily work activities. As a result, some important requirements were identified, e.g., that users often based their decisions on simple metrics or threshold values predefined by their companies. This allowed the redesign of a new version of the tool: the S2S4E Decision Support Tool (S2S4E 2020), an operational climate service which integrates subseasonal to seasonal climate predictions for renewable energy production. In S2S4E, a series of changes were applied to the initial Project Ukko visualization following well-established visual recommendations (Tufte 2001; Few 2009; Yoghourdjian et al. 2018; Veras Guimarães 2019). Some interactive elements were also incorporated to allow users to filter the nonrelevant information, thus reinforcing their cognitive abilities during daily work activities (Gerharz and Pebesma 2009). Specifically, the modifications include (see comparison in Fig. 1): (i) hiding values that do not meet the minimum guarantees of quality (prediction skill); (ii) simplifying the visual encoding for some variables (intensity and predicted change); and (iii) reducing the visual noise by using colors that are similar to the background color for nonrelevant predicted changes, i.e., middle category, showing conditions that can be considered "normal" or close to the average historical observations for a particular region (Chun 2017).

Methodology

To make a comparison and validation from the point of view of perception and cognitive load, we assessed if the changes applied to the original tool (Project Ukko) to create the redesigned one (S2S4E) were fully effective for a general public and not only for an expert audience. To avoid interference in the comparison of cognitive load measurements between both visual representations, not all the improvements included in the S2S4E tools were shown in this study (e.g., color blindness palette, labeling improvement and customizable elements).

The experiment. We conducted an experiment with nonexperts to assess if users' tasks performance when using Project Ukko was improved after redesigning the tool taking into account user-centered design, visualization techniques, and interaction elements. A sample

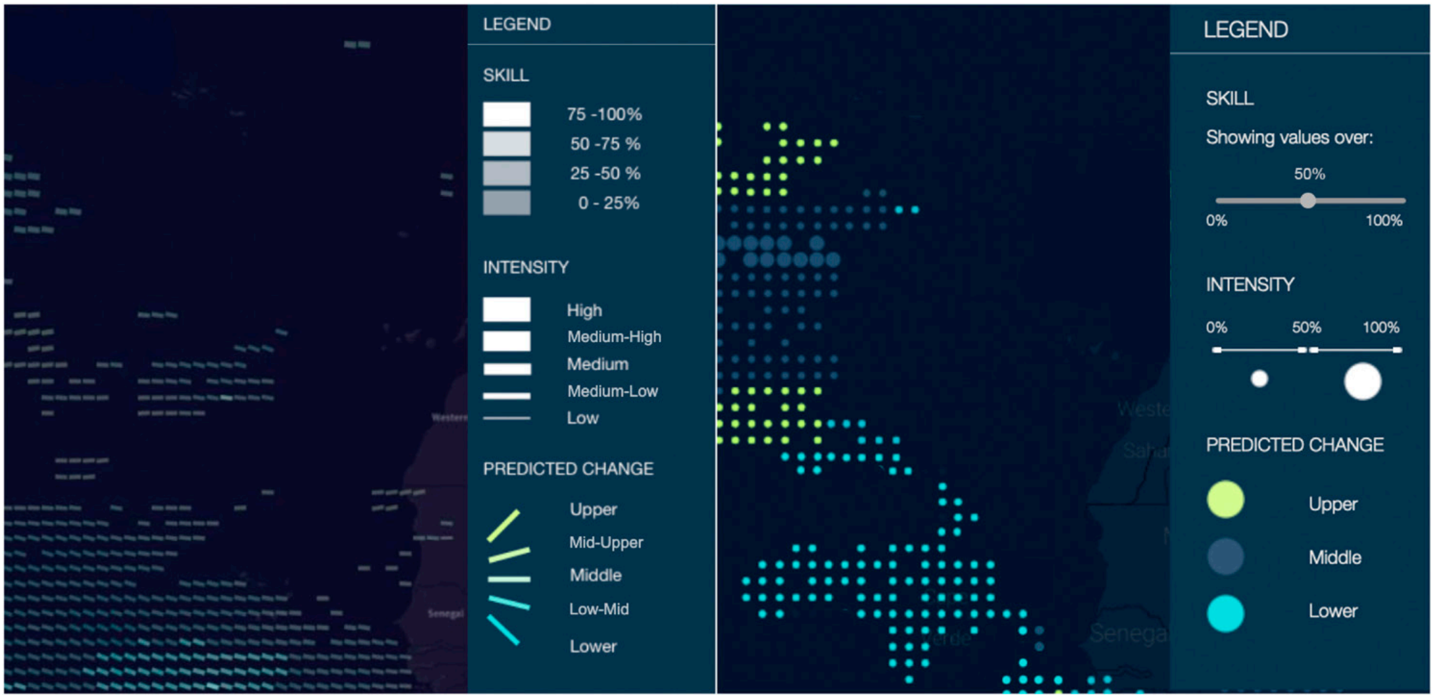

Fig. 1. (left) Project Ukko and (right) the redesigned tool based on user requirements research and visual encoding techniques. Applied changes: filtering out glyphs under a certain prediction skill threshold; simplification of the intensity representation from five to only two size; and simplification of the predicted change color scale from five to three colors, where a similar color to the background one is used for values close to the historical average and more distinguishable colors are used for values higher and lower than the historical average.

of 20 people was tested in order to guarantee 95% of detection of problems (Virzi 1992; Faulkner 2003). The sample was composed by 50% of men and 50% of women aged between 22 and 50, with a low-to-medium knowledge in data visualization, which entailed being familiar with common charts or basic maps but without expertise in visualization of climate or uncertainty information. Recruited participants were students and administrative staff in academia from the human resources, communication, and finance departments. Participants were asked not to have consumed exciting substances that could affect the test results.

The study consisted in two main analyses: (i) a quantitative analysis through the application of a user testing session with two tasks to be completed with both tools (Project Ukko and redesign), and (ii) a qualitative analysis of the positive versus the negative aspects for both tools through the application of a bipolar laddering pocket methodology (Pifarré and Tomico 2007). We also implemented a brief two-question quiz (Schrepp et al. 2017) to determine which visualization needed more mental effort, and which was the tool preferred by participants for decision-making processes.

The development of user testing sessions followed well-established recommendations of planning, moderation, and analysis (Nielsen 1993; Faulkner 2003). User testing sessions were moderated by an expert and recorded to measure time and responses afterward (Holtzblatt et al. 2004). Participants were asked to perform two tasks (task 1 and task 2 below) with each of the tools, based on two typical daily work activities of the intended users (Anderson et al. 2011; Block 2013).

Participants were provided with some context about the real-world conditions in which the tasks would take place. In addition, all the information and conditions necessary to carry out the tasks were presented in the statements, visualization, and captions shared with participants, for them to be able to perform the tasks without having an expert or climate science background. It was simply necessary to identify the requested areas or properties visually (Trivedi 2012). Task 1 was aimed to test if the tool allows good detection of areas with particular conditions while task 2 was aimed to test the differentiability of glyph representation at specific locations.

> **Task 1 statement:** Locate or identify an area on the map that is appropriate to the location of a wind power plant. The area must meet a series of conditions: For Project Ukko, the suitable area should have a prediction skill over 50%, high or medium-high intensity, and an upper or mid-upper predicted change in wind speed (Fig. 1, left panel). For the redesigned version, the area should have a skill over 50%, intensity over 50%, and upper predicted change (Fig. 1, right panel). *[This task required users to be able to identify at least one of the areas that met the conditions specified.]*
> **Task 2 statement:** Identify aloud the conditions that occur in the points included in the highlighted area on the map in terms of skill, intensity, and predicted change. *[This task required users to be able to identify the characteristics of two kinds of glyphs contained in a certain map area]*.

For both task 1 and task 2, we measured the cognitive load and task performance from a quantitative perspective using the following indicators:

1) Success rates when completing a task, including total or partial success (Freitas et al. 2002; Winckler et al. 2004; Ellis and Dix 2006).
2) Response time when completing a task.
3) Number of fixations (i.e., number of gaze points located very close in space, when the eyes are locked toward an object).
4) Fixation duration (i.e., period of time allocated to a fixation (Wang et al. 2014; Majooni et al. 2018).
5) Number of accesses to legend for the completeness of tasks (Pretorius et al. 2005; Klingner et al. 2008).

We also compared qualitative aspects in the use of both tools taking into account the level of user satisfaction and the decision-making facilitation using the following indicators:

6) Subjective effort perception (i.e., user perception on which tool was easier to use while performing the tasks (two-question quiz).
7) Preference between both tools regarding the decision-making process (two-question quiz).
8) Positive and negative aspects (bipolar laddering pocket methodology).

The number of fixations, the fixation duration, and the number of visual paths to legend (quantitative indicators 3, 4, and 5) were monitored using an eye-tracker (GP3 Eye tracker of gazepoint), a sensor-based device that measures eye positions (i.e., point of regard) or eye movement.

To investigate whether there were differences between tools or tasks, a chi-squared test was performed for indicator 1 and a two-way analysis of variance (ANOVA) was performed for indicators 2 to 5 using R software (R Core Team 2018).

During the performance of tasks 1 and 2, we promoted the thinking-aloud technique to identify and detect usability problems (Nielsen 1993; Olmsted-Hawala et al. 2010). Also, the sequence of tasks was randomly presented to the users to avoid biases derived from the learning acquired during the completion of the tasks (Li et al. 2013). We used the bipolar laddering pocket methodology (qualitative indicator 8), which is a reduced version of the bipolar laddering technique. It consists of asking the user about three positive and three negative aspects of both tools (instead of the 10 aspects requested in the extended modality). Then the users rate (from 1 to 10) the aspects mentioned regarding their importance (for the positive ones) or their severity (for the negative ones). By evaluating users' subjective opinions using a scale from 1 to 10 we are able to better quantify input results that are qualitative, which helps identify the key aspects to work with or prioritize the problems that need to be solved (Pifarré and Tomico 2007).

Results

Quantitative analysis of Project Ukko vs redesign. The assessment of success rates (indicator 1) indicated that the number of successfully completed tasks was significantly better when using the redesigned tool than when using Project Ukko, X^2 (9, $N = 20$) = 60.6, $p < 0.001$ (see Figs. 2a,b). In the case of Project Ukko, only 15.8% of the participants successfully completed task 1 and 21.1% task 2. In contrast, with the redesigned tool, task success was much higher, reaching 97.4% and 68.4% for task 1 and task 2, respectively. Although a higher proportion of failures and abandonments occurred during task 1, especially with Project Ukko, completion of task 2 showed various cases of partial success (i.e., identified just one of the two types of glyphs presented) for both tools, reaching 31.6% with the redesigned tool.

The average time to solve a task (indicator 2) was significantly lower for task 1 compared to task 2 ($p = 0.001$) and for the redesigned tool when compared to Project Ukko ($p < 0.001$) (see Figs. 2c,d).

The eye-tracker measurements of the number of fixations (indicator 3) was significantly lower for task 1 compared to task 2 ($p = 0.024$) but showed no significant differences between Project Ukko and the redesigned tool ($p = 0.234$) (see Figs. 2e,f). The fixation duration (indicator 4) did not show a significant difference between tasks ($p = 0.061$) or between both tools ($p = 0.651$) (see Figs. 2g,h). The number of accesses to legend (indicator 5) followed a similar trend as observed for the indicator of response time, with Project Ukko showing higher numbers than the redesigned tool ($p = 0.017$) and significantly lower values for task 1 when compared with task 2 ($p = 0.021$) (see Figs. 2i,j).

Qualitative analysis of project Ukko vs redesign. In terms of perceived effort (indicator 6), a much bigger effort was perceived for Project Ukko (89.9% of participants) than the redesigned

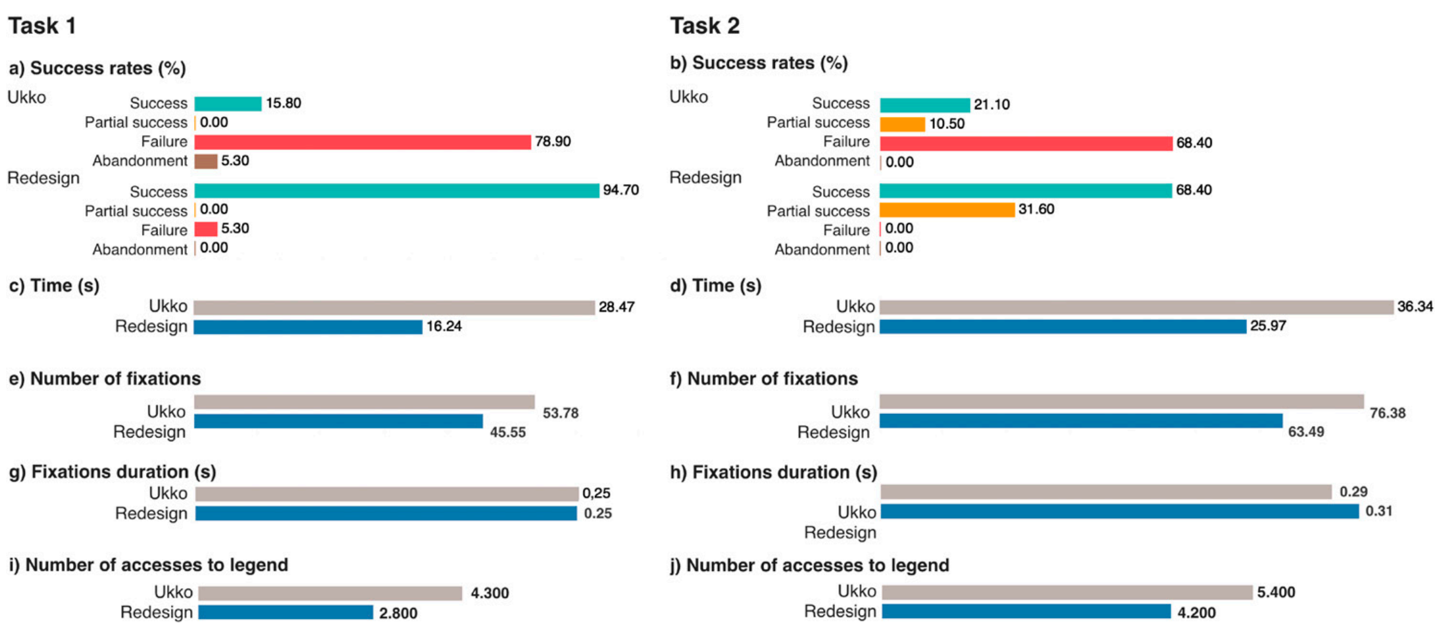

Fig. 2. (a),(b) Comparison of participants' success rates, (c),(d) average time to solve the tasks, (e),(f) number of fixations, (g),(h) fixations duration, and (i),(j) number of accesses to legend when using Project Ukko and the redesigned tool for tasks 1 and 2, respectively.

version (10.5% of participants). Likewise, regarding the tool preferred for decision-making (indicator 7), 89.50% of the participants stated that they would choose the redesigned tool as their working tool for daily tasks decision-making.

By analyzing the results of the bipolar laddering pocket technique (indicator 8) we found that the number of positive aspects mentioned by participants was just 15 for Project Ukko against 42 mentioned for the redesigned tool (see Fig. 3). Regarding the negative aspects, 32 and 12 aspects were pointed out for Project Ukko and the redesigned tool, respectively. On average, positive aspects were rated with an average score of 7.3 for Project Ukko and 8.75 for the redesign. Conversely, negative aspects were rated with a greater severity for Project Ukko, with an average score of 7.4, than for the redesigned tool, with an average score of 5.8.

Discussion

Considering user requirements when developing climate data visualizations is key to improve decision-making. Moreover, the simplification of a complex visualization through changes in visual encoding and interactivity often increases efficiency. Here we use quantitative and qualitative indicators to assess whether the redesign of the Project Ukko tool, taking into account user requirements, visual encoding and interactivity, enhances communication, users' cognitive capacity, and translates into a better task performance.

When comparing the experience of participants with both Project Ukko and the redesigned tool, the quantitative indicators of success rate and response time when completing a task (indicators 1 and 2) demonstrate that the changes made to the redesigned tool increased the rate of success or partial success and allowed participants to perform the tasks faster. It is necessary to highlight that, in the case of Ukko, the success rate was extremely low for task 1. This was due to an incorrect identification of the area, which was done based on its brightness (high skill values), but that did not meet the minimum requirements of thickness (high intensity values). From a usability point of view, this is commonly defined as a false

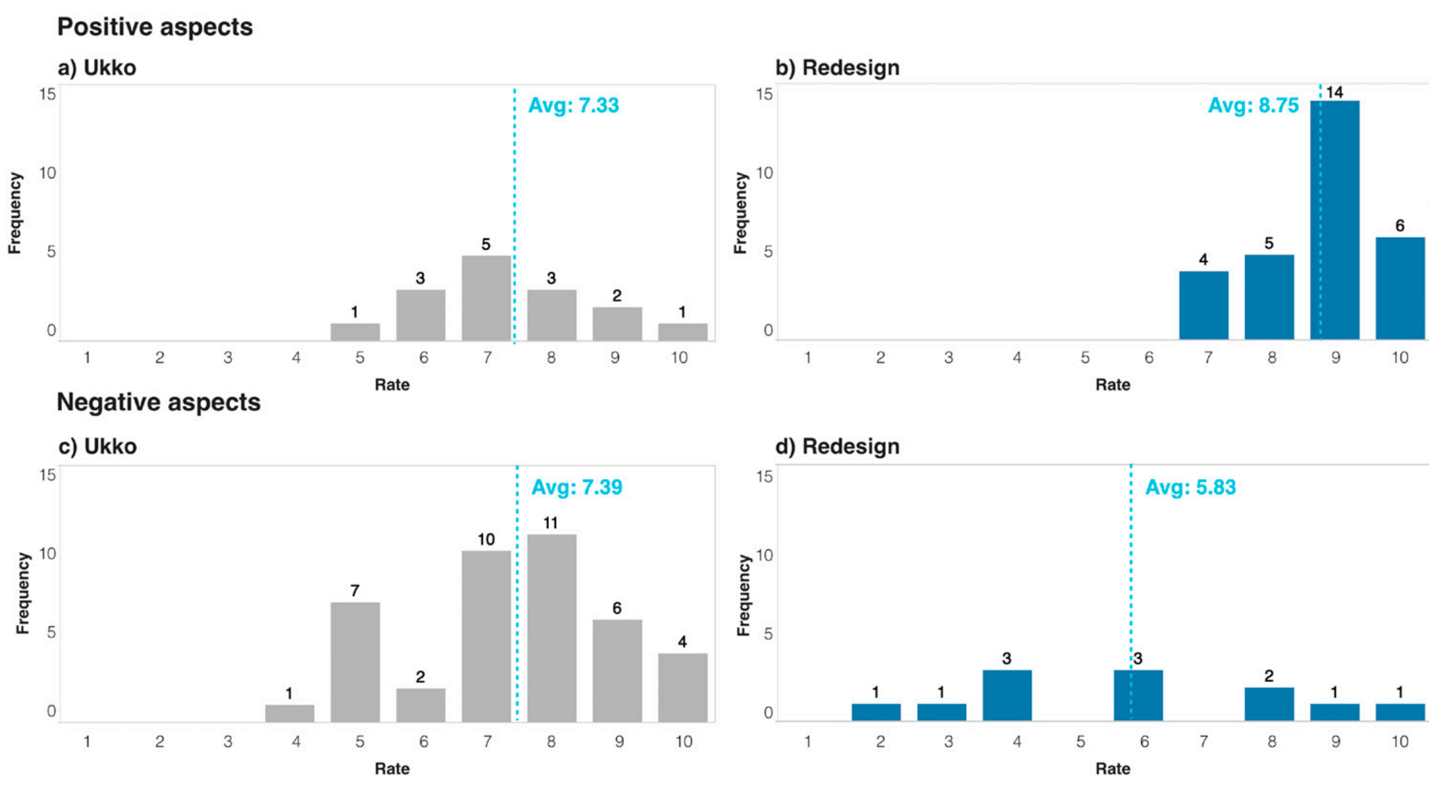

Fig. 3. Frequency histograms of participants' ratings of positive and negative aspects mentioned for (a),(c) Project Ukko and(b),(d) the redesigned tool. All the mentioned aspects (*n* = 97) receive a score from 1 to 10. Positive aspects are rated from 10 (very positive) to 1 (less positive) and negative aspects are rated from 10 (very serious) to 1 (slightly serious).

success, since users believe they have performed the task correctly, when in fact the answer was wrong (Brinck et al. 2002).

Qualitative methods such as the bipolar laddering and two-question quiz applied in this study are based on subjective user assessment (e.g., experience, feelings, intuition, opinions), which is often perceived as less reliable than other quantitative methods (Szafir and Szafir 2016). However, they are well-established practices that support and complement quantitative analysis and have been widely applied (Lim et al. 2019; Navarro et al. 2020). These methods provide useful insights to understand user's preferences and the positive and negative aspects that intervene in the performance and efficiency of users to perform day-to-day tasks.

In general, the changes in shape and size (i.e., the visual encoding of the information) as well as the reduction of categories were highly rated by the participants, enhancing the clarity and ease of use of the redesigned tool. In the case of Ukko, a large number of participants were not able to identify the categories associated with a glyph (skill, intensity, predicted change) nor the exact category, even when they tried to compare them to nearby glyphs. This was because thin lines, which were combined with opacity (a visual encoding that affects visibility), made color detection more difficult. Sometimes the participants were unable to identify the exact thickness of a group of glyphs despite the many accesses to the legend, which led them to randomly select a category to avoid abandoning the task. Overall good practices in data visualization indicate that opacity, combined with color and reduced thickness, can make graphical interpretation worse (Dastani 2002; Jenny and Kelso 2007; Ware 2012). Also, the use of the slope of the glyphs was perceived as a negative aspect by participants. This is confirmed by visual encoding good practices, indicating that changes in slope are more difficult to be interpreted, especially for nominal data (Alexandre and Tavares 2010; Munzner 2014). In addition, in the context of wind data visualization, the use of slopes tends to be related to wind direction (Powers et al. 2017), as was also pointed out by some of the participants during the test. Therefore, using slope to display the wind predicted change can be counterintuitive for users.

Only two sizes and three colors were combined in the redesigned tool. This favors the detection of the areas of interest because the visual encoding does not create competition between variables (in this case, wind intensity and predicted change) (Iliinsky and Steele 2011; Riveiro et al. 2008). Additionally, glyphs below a specific skill threshold (which would be discarded in a decision-making process) can be hidden from the display, allowing the user to focus on feasible options. This has been recognized as an effective means to reduce user memory workload and enhance task performance (Hegarty 2011).

The most frequently mentioned negative aspect of the redesigned tool referred to the color chosen for the middle prediction category ("Small dark glyphs, showing average values, have a similar color to the background"). Users found the color too similar to the background, therefore lacking sufficient contrast. However, this was decided on purpose, since target users are more interested in situations that depart from normal (i.e., upper and lower than average values), since these are the ones in which they need to take action (Kohlhammer and Zeltzer 2004). Therefore, by choosing a color similar to the background for nonrelevant values, we reduced visual noise in the representation. On the other hand, values of upper and lower predicted change (indicating wind conditions above and below normal) use a green-yellow color hue, which stimulates more photoreceptors in the human eye and hence are easily detected by users (UNSW 2015).

Despite color being a crucial element of visual encoding, we did not include further changes in the color scale used in the test to compare Project Ukko and its redesigned visualization. This was decided in order to focus the analysis in the visual encoding and interactivity aspects and to avoid a major change between both tools that could bias the results of the test.

However, in order to improve the accessibility of any climate service, color-blindness should be taken into account in color choices for visual representations (Light and Bartlein 2004). In the case of Project Ukko, even using color-blind-friendly scales, the combination of color, opacity, and certain widths reduced the effective perception by color-blind people. In this sense, the S2S4E Decision Support tool (S2S4E 2020) already included changes in the color palette to improve its accessibility by taking into account color-blindness aspects.

Some changes applied in the redesign of Project Ukko were also related to the use of interactive filters to dose or personalize information. The skill filter in the redesigned version of the tool allows one to explore the uncertainty associated with the predicted change focusing just on the relevant data for the user (i.e., values below a preferred skill threshold are hidden). In the same way, the intensity slider allows the user to establish a preferred threshold, giving more visual presence (larger glyph size) to the values above this threshold. This capacity of filtering and personalizing was highlighted as a strong positive aspect of the redesigned tool. One of the users even suggested adding interactivity to the color legend, to be able to filter by the predicted change. Indeed, interactivity allows users to consume information step by step, explore particular aspects of complex datasets, and display relevant information in their own world view (Beddington 2011; McInerny et al. 2014). This is thus a highly recommended feature for online climate services taking into account that there may be limits to how useful interactive visualizations are if the viewers do not have the required skills to interact with the presented information (diSessa 2004).

The second task proposed to participants was more challenging than the first task regardless of the tool used. The classification of two types of glyphs proposed in task 2 took more time to complete than the identification of an area of interest in task 1. This was especially remarkable for Ukko, with a higher number of categories competing at the same time for visual attention (Alhadad 2018; Munzner 2014). Also, the number of times that participants needed to check the legend (indicator 5) was larger for Project Ukko. The difference between Project Ukko and the redesigned tool is probably related to Project Ukko's negative aspects linked to problems for understanding the legend, which had a more complex visual encoding. The combination of different categories in Project Ukko was also considered as a negative aspect (e.g., "overwhelming representation," "mixing too many categories increases complexity") in contrast to positive aspects of the redesigned tool linked with the simplicity of the representation (e.g., "the representation is very clear," "easy to distinguish between glyphs"). This would also explain the low success rate of task 1 for Ukko where, despite having more accesses to the legend, the area selected by participants did not meet the requirements of the statement in terms of skill, intensity and predicted change.

The difference in the purpose of task 1 and task 2 (identification versus classification, respectively), might also be the reason why participants needed to check the legend more often at the second task. Overall, the obtained number of accesses to the legend (combined with a simpler visualization) suggests a reduction in the cognitive load of the participants during the completion of the tasks with the redesigned tool as they could retain the legend better and therefore reduce the number of times they had to check it.

Regarding the fixation duration, the quantitative indicator behaved almost equally between tasks and between tools, with durations ranging between 0.25 and 0.31 s. The number of fixations are the number of times a user pays attention to a certain point or area of interest on the screen. According to available bibliography, a longer fixation duration may indicate a bigger cognitive load during task performance (Duchowski 2007; Ooms et al. 2014; Andrzejewska and Skawińska 2020; Klingner et al. 2008; Krejtz et al. 2018) or that users have found more interesting elements to fix their attention for a longer time, without necessarily implying a greater difficulty or cognitive load (Henderson and Ferreira 2004; Klingner et al. 2008; Ooms et al. 2012; Krejtz et al. 2018; Andrzejewska and Skawińska 2020).

Advanced brain monitoring tools, such as electroencephalograms and eye-tracker measures of pupillometry, can be useful to further study cognitive load (Anderson et al. 2011; Jiang et al. 2014; Keskin et al. 2020). It would be interesting to test them with other eye-tracker models to explore if this could be due to the accuracy of the model used.

The perceived effort during the task performance (indicator 6) identifies Project Ukko as being more complex to use than the redesigned tool, which was indicated as the tool preferred by 90% of the participants tested in this work (indicator 7). This contributes to the hypothesis that by eliminating or simplifying visual encodings nonrelevant to a target action or task and increasing interactivity, we favor decision-making.

The results and opinions of the bipolar laddering (indicator 8) clearly supported the previous indicators. A total of 29 positive comments were obtained for the redesigned tool, compared to 15 for the original Ukko tool (Table 1). This was also confirmed by the higher average score obtained for the redesign (8.75) when compared to Project Ukko (7.33). In the same way, the redesigned tool received fewer negative comments (12 compared to 41 for Ukko) and they were less serious (obtained scores of 5.83 for the redesign against 7.33 for Ukko). The most serious aspects associated with Ukko referred to the difficulty to differentiate the categorization of glyphs due to the combination of encoding through color, intensity, and thickness, often making participants unable to identify the corresponding category. These aspects had a very high frequency, 25 comments with notable severity and an obtained score of 8.

Table 1. Participants' positive and negative comments for Project Ukko and the redesigned tools with the number of participants that mentioned a particular aspect (freq.) and average rate of its importance/severity (avg.).

Project Ukko					
Positive	**Freq.**	**Avg.**	**Negative**	**Freq.**	**Avg.**
Easy to distinguish extreme areas	7	7.8	Difficult to distinguish opacity	9	8
A very detailed version with lots of information	3	6.6	Mixing width and brightness is too complex	9	8.22
Static legends are more traditional	3	6.6	Thin glyphs with low visibility are impossible to distinguish	7	7.57
Visually attractive	2	7.5	Using the combination of two variables (color and slope) for prediction change is too complex	4	6.25
			Slopes are confusing, they usually are used to show wind direction	3	7.33
			Mixing too many categories increases complexity	3	6.33
			Overwhelming representation	2	7
			Terrain not visible enough	1	6
			Slope value (of glyphs) is difficult to measure	1	5
			Legend is confusing	1	5
Total	15	7.33		41	7.39
Redesign					
Positive	**Freq.**	**Avg.**	**Negative**	**Freq.**	**Avg.**
Shapes and sizes are easier to identify	8	8.25	Small dark glyphs (showing average values) have a color similar to the background	5	5.6
The representation is very clear	6	9.3	The descriptive labels of the skill slider (using mathematical terms) may not be clear to all audiences.	3	6.66
Easy to distinguish between glyphs	5	8.4	Too basic to represent predicted change	1	6
Skill filtering is very useful	4	9.25	By simplifying some of the categories, we lose information	1	6
High contrast	2	9.5	Terrain not visible enough	1	6
Easy location of an area	2	9	All the three filters or descriptive labels could have been interactive (such as color category)	1	4
Simple categorization	2	8			
Total	29	8.75		12	5.83

Glossary

Our work is based on the collaboration of multiple disciplines that, when combined, improve the user experience and favor decision-making. **User-centered design (UCD)** allows us to meet the needs of the users. **Interaction design** allows us to dose and customize the way the data are displayed. **Design** and **data visualization** rules improve the way of visually encoding information and highlight what is really important. **Cognitive psychology** helps us measure the impact of complex visualizations on users and, with the help of other disciplines, favors their understanding.

These disciplines have specific terminology used in this paper:

Accessibility: Discipline and rules that guarantee that websites and technologies are designed and developed so that people with disabilities can use them independently from their capability limitations: auditory, visual, cognitive, physical, or neurological.

Cognition: Mental process of acquiring knowledge and understanding through senses, thought, and experience.

Cognitive load: Refers to the used amount of human working memory resources, which is limited in both capacity and duration.

Cognitive psychology: The scientific study of mental processes such as attention, memory, perception, problem solving, and understanding.

Eye-tracker: Sensor-based device which measures where the participant is looking at (the point of gaze) and the motion of the eyes.

Fixation: A period of time during which the eyes are locked toward an object or visual element.

Color hue: The attribute of color defined by wavelength (red, blue, etc.).

Glyph: A hieroglyphic character or symbol used in visualization as a part of a chart or graph.

Interaction design: Design of interactive products focusing on the way users interact with them, including visual representation, terminology, devices, and behavior.

Multidimensional visualization: Graph or visualization showing more than one variable through visual encoding (color, size, etc.).

Opacity: Property of a visual element that determines how transparent (or less visible) it will be. The lower the opacity value, the more transparent the element is.

Perception: The way in which something is regarded, understood, or interpreted.

Usability testing: Evaluation of a product or a service in order to detect problems or evaluate how easy it is to use it.

User-centered design: Iterative design process in which designers focus on the user needs and involve them in each phase of the design process.

Visual encoding: Translating the data into a visual element on a chart/map or graph using visual properties as length, position, size, color, slope, opacity, etc.

Working memory: The cognitive system with a limited capacity that can hold information temporarily.

Regarding the negative comments obtained for the redesigned tool, five participants referred to the color similarity of some glyphs to the background of the tool (values close to the mean in terms of predictive change). However, after explaining the reasons for this design decision (reduce visual noise by attenuating non relevant points), all users found the change appropriate. Another negative aspect was related to the label terminology of the skill slider, which included mathematical terms that may not be clear to all audiences. When delving into the reasons for the negative assessment, the participant clarified that the control seems useful, but that the texts used in the labels could be clearer or more intuitive.

Paradoxically, some positive comments received for Ukko, referred to its higher density of information, greater detail in the representation of predicted change (five categories instead of the three categories in the redesigned tool) and visual appeal. However, although users who mentioned these aspects believed that these characteristics could be valuable in a context where exploration was the objective of visualization, they did not favor clarity or decision-making.

Conclusions

The behavioral decision-making literature shows how people often struggle to understand particular climate terminology. There is often a mismatch between the understanding of concepts such as probabilities or uncertainty between experts and nonexperts. Although visualizing forecast uncertainties and associated probabilities is thought to increase users' trust, it does not automatically lead to better decisions.

Our results identify relevant aspects that can improve user experience and reduce cognitive load and that are worth considering when designing climate data visualizations. These include choosing representations and categories tailored to specific decisions, avoiding visual encoding that interferes with users' perception of the represented forms, and offering interactive elements that allow users to filter nonrelevant information or highlight relevant information for the decision at hand. In the redesign of Project Ukko we included all these changes at the same time. Hence, we cannot empirically establish the relative influence of each of these individual aspects in the overall reduction of the users' cognitive load. Nevertheless, we demonstrate that all these aspects can help reduce cognitive load, favor decision-making, and thus improve the overall user experience with a climate service.

In future works, it would be interesting to delve into the weight of each of the implemented actions (simplifying the number of categories, avoiding redundant visual encoding, customizing the visualizations based on user needs through interactive controls) in the total reduction of the cognitive load.

Likewise, analyzing the implemented changes in the context of the final tool, in combination with other improvements not assessed in the framework of this study (redesign of the navigation, color-blind aspects, customization, and levels of detail available), can highlight additional benefits. This would delve further into the visual communication of climate information.

Our study highlights that, when combining techniques and knowledge from different disciplines [climate science, design, user-centered design (UCD), user interaction, and cognitive psychology], we are able to find better solutions for the visualization of climate data, especially when aimed at supporting decision-making. In addition, we identify a clear need for co-design and increased empirical testing of the resulting products. We recommend information providers and tool designers in the field of climate services to collaborate more with end users throughout the whole design process to identify what is effective and to leverage the knowledge and well-established techniques from nonclimate related disciplines that have a lot to offer.

Acknowledgments. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreements 776787 (S2S4E), 776613 (EUCP), and (ClimatEurope). This work was also supported by the MEDSCOPE project. MEDSCOPE is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by AEMET (ES), ANR (FR), BSC (ES), CMCC (IT), CNR (IT), IMR (BE), and Météo-France (FR), with co-funding by the European Union (Grant 690462). The research team would like to thank the participants of the test who generously shared their time and opinions for the purposes of this research. This study is a part of the PhD of the corresponding author, Luz Calvo.

References

Alexandre, D., and J. Tavares, 2010: Introduction of human perception in visualization. *Int. J. Imaging*, **4** (10), 60–70.

Alhadad, S. S. J., 2018: Visualizing data to support judgement, inference, and decision making in learning analytics: Insights from cognitive psychology and visualization science. *J. Learn. Anal.*, **5**, 60–85, https://doi.org/10.18608/jla.2018.52.5.

Anderson, E. W., K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva, 2011: A user study of visualization effectiveness using EEG and cognitive load. *Comput. Graph. Forum*, **30**, 791–800, https://doi.org/10.1111/j.1467-8659.2011.01928.x.

Andrzejewska, M., and A. Skawińska, 2020: Examining students' intrinsic cognitive load during program comprehension – An eye tracking approach. *AIED 2020: Artificial Intelligence in Education*, I. I. Bittencourt et al., Eds., Springer, 25–30, https://doi.org/10.1007/978-3-030-52240-7_5.

Antifakos, S., A. Schwaninger, and B. Schiele, 2004: Evaluating the effects of displaying uncertainty in context-aware applications. *UbiComp* 2004*: Ubiquitous Computing*, N. Davies, E. D. Mynatt, and I. Siio, Eds., Springer, 54–69, https://doi.org/10.1007/978-3-540-30119-6_4.

Argyle, E. M., J. J. Gourley, Z. L. Flamig, T. Hansen, and K. Manross, 2017: Toward a user-centered design of a weather forecasting decision-support tool. *Bull. Amer. Meteor. Soc.*, **98**, 373–382, https://doi.org/10.1175/BAMS-D-16-0031.1.

Atkins, R., and K. Mcneal, 2018: Exploring differences among student populations during climate graph reading tasks: An eye tracking study. *J. Astron. Earth Sci. Educ.*, **5**, 85–114, https://doi.org/10.19030/jaese.v5i2.10219.

Beddington, J., 2011: Blackett Review of high impact low probability events. Government Office for Science, 46 pp.

Bertin, J., 2010: *Semiology of Graphics: Diagrams, Networks, Maps*. 1st ed. Esri Press, 456 pp.

Bevington, K., S. Ansari, R. Bilotta, A. B. Courtright, and A. Lang, 2019: Improving climate data visualization and information dissemination on the U.S. drought portal through user experience (UX) design. *2019 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract PA33C-1103.

Block, G., 2013: Reducing cognitive load using adaptive uncertainty visualization. Ph.D. thesis, Nova Southeastern University, 132 pp. https://nsuworks.nova.edu/gscis_etd/95/.

Bonneau, G.-P., H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz, 2014: Overview and state-of-the-art of uncertainty visualization. *Scientific Visualization*, C. D. Hansen et al., Eds., Springer, 3–27, https://doi.org/10.1007/978-1-4471-6497-5_1.

Borkin, M. A., A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister, 2013: What makes a visualization memorable? *IEEE Trans. Vis. Comput. Graph.*, **19**, 2306–2315, https://doi.org/10.1109/TVCG.2013.234.

Brinck, T., D. Gergle, and S. D. Wood, 2002: *Usability for the Web*. Elsevier, 497 pp.

Cairo, A., 2012: *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders Publishing, 384 pp.

Christel, I., and Coauthors, 2016: Project Ukko: Seasonal wind speed predictions at a glance. *WindEurope Summit 2016*, Hamburg, Germany, Wind Europe, https://discovery.dundee.ac.uk/en/publications/project-ukko-seasonal-wind-speed-predicitions-at-a-glance.

——, D. Hemment, D. Bojovic, F. Cucchietti, L. Calvo, M. Stefaner, and C. Buontempo, 2018: Introducing design in the development of effective climate services. *Climate Serv.*, **9**, 111–121, https://doi.org/10.1016/j.cliser.2017.06.002.

Chun, R., 2017: Redundant encoding in data visualizations: Assessing perceptual accuracy and speed. *Vis. Commun. Quart.*, **24**, 135–148, https://doi.org/10.1080/15551393.2017.1343153.

Cleveland, W. S., and R. McGill, 1985: Graphical perception and graphical methods for analyzing scientific data. *Science*, **229**, 828–833, https://doi.org/10.1126/science.229.4716.828.

Dastani, M., 2002: The role of visual perception in data visualization. *J. Vis. Lang. Comput.*, **13**, 601–622, https://doi.org/10.1006/jvlc.2002.0235.

Davis, B. M., K. Dickerson, and S. C. Gillmore, 2020: User centered design strategies for improving visualization of sensor data in rotorcraft cockpit displays for degraded visual environment operations. *Advances in Usability and User Experience*, T. Ahram and C. Falcão, Eds., Springer, 131–141, https://doi.org/10.1007/978-3-030-19135-1_13.

diSessa, A. A., 2004: Metarepresentation: Native competence and targets for instruction. *Cognit. Instr.*, **22**, 293–331, https://doi.org/10.1207/s1532690xci2203_2.

Dong, F., G. Ghinea, and S. Y. Chen, 2008: *User Centered Design for Medical Visualization*. Medical Information Science Reference Global, 436 pp.

Duchowski, A., 2007: *Eye Tracking Methodology*. Springer, 334 pp.

Ellis, G., and A. Dix, 2006: An explorative analysis of user evaluation studies in information visualisation. *Proc. of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, Venice, Italy, ACM, 1–7, https://doi.org/10.1145/1168149.1168152.

Faulkner, L., 2003: Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behav. Res. Methods*, **35**, 379–383, https://doi.org/10.3758/BF03195514.

Few, S., 2009: *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. 1st ed. Analytics Press, 329 pp.

Freitas, C. D. S., P. R. G. Luzzardi, R. Cava, M. Winckler, M. Pimenta, and L. Nedel, 2002: On evaluating information visualization techniques. *Proc. of the Working Conf. on Advanced Visual Interfaces*, Trento, Italy, ACM, 373–374, https://doi.org/10.1145/1556262.1556326.

Gerharz, L. E., and E. J. Pebesma, 2009: Usability of interactive and non-interactive visualisation of uncertain geospatial. *Geoinformatik 2009 Konferenzband*, W. Reinhardt, A. Krüger, and M. Ehlers, Eds., Institut für Geoinformatik, 223–230.

Gerst, M. D., and Coauthors, 2020: Using visualization science to improve expert and public understanding of probabilistic temperature and precipitation outlooks. *Wea. Climate Soc.*, **12**, 117–133, https://doi.org/10.1175/WCAS-D-18-0094.1.

Grainger, S., F. Mao, and W. Buytaert, 2016: Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environ. Modell. Software*, **85**, 299–318, https://doi.org/10.1016/j.envsoft.2016.09.004.

Greis, M., T. Ohler, N. Henze, and A. Schmidt, 2015: Investigating representation alternatives for communicating uncertainty to non-experts. *INTERACT 2015: Human-Computer Interaction*, J. Abascal et al., Eds., Springer, 256–263, https://doi.org/10.1007/978-3-319-22723-8_21.

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.

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.

——, M. Canham, and S. Fabrikant, 2010: Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *J. Exp. Psychol. Learn. Mem. Cognit.*, **36**, 37–53, https://doi.org/10.1037/a0017683.

Henderson, J. M., and F. Ferreira, Eds., 2004: *The Interface of language, Vision, and action: Eye Movements and the Visual World*. Psychology Press, 414 pp.

Holmes, N., 1984: *Designer's Guide to Creating Charts and Diagrams.* 2nd ed. Watson-Guptill, 192 pp.

Holtzblatt, K., J. B. Wendell, and S. Wood, 2004: *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Elsevier, 320 pp.

Iliinsky, N., and J. Steele, 2011: *Designing Data Visualizations*. O'Reilly, 114 pp.

Jenny, B., and N. V. Kelso, 2007: Color design for the color vision impaired. *Cartogr. Perspect.*, **58**, 61–67, https://doi.org/10.14714/CP58.270.

Jiang, X., M. Atkins, G. Tien, R. Bednarik, and B. Zheng, 2014: Pupil responses during discrete goal-directed movements. *Proc. SIGCHI Conf. on Human Factors in Computing System*, Toronto, ON, ACM, 2075–2084, https://doi.org/10.1145/2556288.2557086.

Joslyn, S., and J. LeClerc, 2011: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exp. Psychol. Appl.*, **18**, 126–140, https://doi.org/10.1037/a0025185.

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**, 243–256, https://doi.org/10.5194/gmd-5-245-2012.

Keskin, M., K. Ooms, A. O. Dogru, and P. De Maeyer, 2020: Exploring the cognitive load of expert and novice map users using EEG and eye tracking. ISPRS Int. Geo-Inf., **9**, 429, https://doi.org/10.3390/ijgi9070429.

Khamaj, A., Z. Kang, and E. Argyle, 2019: Users' perceptions of smartphone weather applications' usability. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **63**, 2216–2220, https://doi.org/10.1177/1071181319631098.

Klingner, J., R. Kumar, and P. Hanrahan, 2008: Measuring the task-evoked pupillary response with a remote eye tracker. *ETRA'08: Proc. of the 2008 Symp. on Eye Tracking Research & Applications*, Savannah, GA, ACM, 69–72, https://doi.org/10.1145/1344471.1344489.

Kohlhammer, J., and D. Zeltzer, 2004: Towards a visualization architecture for time-critical applications. *Proc. of the 9th Int. Conf. on Intelligent User Interfaces*, Funchal, Madeira, Portugal, ACM, 271–273, https://doi.org/10.1145/964442.964502.

Kosara, R., 2013: InfoVis is so much more: A comment on Gelman and Unwin and an invitation to consider the opportunities. *J. Comput. Graph. Stat.*, **22**, 29–32, https://doi.org/10.1080/10618600.2012.755465.

Krejtz, K., A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, 2018: Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, **13**, e0203629, https://doi.org/10.1371/journal.pone.0203629.

Lau, A., and A. Vande Moere, 2007: Towards a model of information aesthetics in information visualization. *2007 11th Int. Conf. Information Visualization (IV'07)*, Zurich, Switzerland, Institute of Electrical and Electronics Engineers, 87–92, https://doi.org/10.1109/IV.2007.114.

Li, Y., N. Hollender, and T. Held, 2013: Task sequence effects in usability tests. *IADIS Int. Conf. Interfaces and Human Computer Interaction and Game and Entertainment Technologies 2013*, Prague, Czech Republic, University of Economics in Prague, 141–148.

Light, A., and P. Bartlein, 2004: The end of the rainbow? Color schemes for improved data graphics. *Eos, Trans. Amer. Geophys. Union*, **85**, 385, https://doi.org/10.1029/2004EO400002.

Lim, K., A. Selamat, R. Alias, O. Krejcar, and H. Fujita, 2019: Usability measures in mobile-based augmented reality learning applications: A systematic review. *Appl. Sci.*, **9**, 2718, https://doi.org/10.3390/app9132718.

Ling, C., L. Hua, C. D. Karstens, G. J. Stumpf, T. M. Smith, K. M. Kuhlman, and L. Rothfusz, 2015: A comparison between WarnGen system and probabilistic hazard information system for severe weather forecasting. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **59**, 1791–1795, https://doi.org/10.1177/1541931215591387.

Lloyd, R., 1997: *Spatial Cognition: Geographic Environments*. Springer, 118 pp.

Lorenz, S., S. Dessai, P. M. Forster, and J. Paavola, 2015: Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. *Philos. Trans. Roy. Soc.*, **373A**, 20140457, https://doi.org/10.1098/rsta.2014.0457.

Majooni, A., M. Masood, and A. Akhavan, 2018: An eye-tracking study on the effect of infographic structures on viewer's comprehension and cognitive load. *Inf. Vis.*, **17**, 257–266, https://doi.org/10.1177/1473871617701971.

Makri, S., 2015: EUPORIAS seasonal wind prediction prototype: User evaluation. City University London, 36 pp., https://earth.bsc.es/wiki/lib/exe/fetch.php?media=library:external:services:euporias_seasonal_wind_prediction_prototype_-_user_evaluation_report_nov_2015.pdf.

McInerny, G. J., and Coauthors, 2014: Information visualisation for science and policy: Engaging users and avoiding bias. *Trends Ecol. Evol.*, **29**, 148–157, https://doi.org/10.1016/j.tree.2014.01.003.

Munzner, T., 2014: *Visualization Analysis and Design*. 1st ed. CRC Press, 428 pp.

Navarro, I., O. de Reina, D. Fonseca, M. Gomez, and Á. Ferrer, 2020: Virtual reality using smart-devices in educational frameworks: Case study Museum Casa Batlló. *Int. J. Technol. Hum. Interact.*, **13**, 50–61, https://doi.org/10.4018/IJTHI.2017100104.

Nielsen, J., 1993: Evaluating the thinking-aloud technique for use by computer scientists. *Advances in Human-Computer Interaction*, H. R. Hartson and D. Hix, Eds., Ablex Publishing, 69–82.

Oakley, N. S., and B. Daudert, 2016: Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. *Bull. Amer. Meteor. Soc.*, **97**, 263–274, https://doi.org/10.1175/BAMS-D-14-00121.1.

Olmsted-Hawala, E. L., E. D. Murphy, S. Hawala, and K. T. Ashenfelter, 2010: Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination websites for usability. *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, Atlanta, GA, ACM, 2381–2390, https://doi.org/10.1145/1753326.1753685.

Ooms, K., P. De Maeyer, V. Fack, E. Van Assche, and F. Witlox, 2012: Interpreting maps through the eyes of expert and novice users. *Int. J. Geogr. Inf. Sci.*, **26**, 1773–1788, https://doi.org/10.1080/13658816.2011.642801.

——, A. Coltekin, P. De Maeyer, L. Dupont, S. Fabrikant, A. Incoul, and L. Van der Haegen, 2014: Combining user logging with eye tracking for interactive and dynamic applications. *Behav. Res. Methods*, **47**, 977–993, https://doi.org/10.3758/s13428-014-0542-3.

Pifarré, M., and O. Tomico, 2007: Bipolar laddering (BLA): A participatory subjective exploration method on user experience. *Proc. of the 2007 Conf. on Designing for User EXperiences, DUX 2007*, Chicago, IL, ACM, 2–13, https://doi.org/10.1145/1389908.1389911.

Powers, J. G., and Coauthors, 2017: The weather research and forecasting model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Pretorius, M., A. P. Calitz, and D. van Greunen, 2005: The added value of eye tracking in the usability evaluation of a network management tool. *SAICSIT'05: Proc. of the 2005 Annual Research Conf. of the South African Institute of Computer Scientists and Information Technologists*, Cape Town, South Africa, South African Institute for Computer Scientists and Information Technologists, 1–10, https://doi.org /10.5555/1145675.1145676.

R Core Team, 2018: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.

Riveiro, M., G. Falkman, and T. Ziemke, 2008: Improving maritime anomaly detection and situation awareness through interactive visualization. *Proc. of the 11th Int. Conf. on Information Fusion*, Cologne, Germany, Institute of Electrical and Electronics Engineers, 1–8.

——, T. Helldin, G. Falkman, and M. Lebram, 2014: Effects of visualizing uncertainty on decision-making in a target identification scenario. *Comput. Graph.*, **41**, 84–98, https://doi.org/10.1016/j.cag.2014.02.006.

Roulston, M. S., G. E. Bolton, A. N. Kleit, and A. L. Sears-Collins, 2006: A laboratory study of the benefits of including uncertainty information in weather forecasts. *Wea. Forecasting*, **21**, 116–122, https://doi.org/10.1175/WAF887.1.

S2S4E, 2020: Decision support tool. Accessed 16 December 2020, https://s2s4e.eu/dst.

Sager, L., R. Grier, C. Jackson, Y. Levchuk, and E. Stelzer, 2007: Visualization techniques for revealing uncertainty to decision makers. *Human Systems Integration Symp. 2007*, Annapolis, MD, American Society of Naval Engineers.

Schrepp, M., A. Hinderks, and J. Thomaschewski, 2017: Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *IJIMAI*, **4**, 103–108, https://doi.org/10.9781/ijimai.2017.09.001.

Stephens, E. M., T. L. Edwards, and D. Demeritt, 2012: Communicating probabilistic information from climate model ensembles—Lessons from numerical weather prediction. *Wiley Interdiscip. Rev.: Climate Change*, **3**, 409–426, https://doi.org/10.1002/wcc.187.

Szafir, D. A., and D. Szafir, 2016: Cognitive load in visualization: Myths & misconceptions. *Workshop on Creation, Curation, Critique and Conditioning of Principles and Guidelines in Visualization (C4PGV)*, Baltimore, MD, IEEE, https://cmci.colorado.edu/visualab/papers/16-c4pgvcogload.pdf.

Taylor, A. L., S. Dessai, and W. B. de Bruin, 2015: Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philos. Trans. Roy. Soc.*, **373A**, 20140454, https://doi.org/10.1098/rsta.2014.0454.

Terrado, M., and Coauthors, 2019: The weather roulette: A game to communicate the usefulness of probabilistic climate predictions. *Bull. Amer. Meteor. Soc.*, **100**, 1909–1921, https://doi.org/10.1175/BAMS-D-18-0214.1.

Trivedi, M. C., 2012: Role of context in usability evaluations: A review. *J. Adv. Comput. Sci. Appl.*, **3**, 69–78, https://doi.org/10.5121/acij.2012.3208.

Tufte, E. R., 2001: *The Visual Display of Quantitative Information*. 2nd ed. Graphics Press, 200 pp.

UNSW, 2015: Colour mixing and colour vision. Accessed 16 December 2020, www.animations.physics.unsw.edu.au/jw/light/colour-mixing.htm.

Veras Guimarães, R., 2019: Visual encoding quality and scalability in information visualization. Ph.D. thesis, Faculty of Science, University of Ontario Institute of Technology, 151 pp., http://ir.library.dc-uoit.ca/handle/10155/1024.

Virzi, R. A., 1992: Refining the test phase of usability evaluation: How many subjects is enough? *Hum. Factors*, **34**, 457–468, https://doi.org/10.1177/001872089203400407.

Wang, Q., S. Yang, M. Liu, Z. Cao, and Q. Ma, 2014: An eye-tracking study of website complexity from cognitive load perspective. *Decis. Support Syst.*, **62**, 1–10, https://doi.org/10.1016/j.dss.2014.02.007.

Ware, C., 2012: *Information Visualization: Perception for Design*. Elsevier, 536 pp.

——, H. Purchase, L. Colpoys, and M. McGill, 2002: Cognitive measurements of graph aesthetics. *Inf. Vis.*, **1**, 103–110, https://doi.org/10.1057/palgrave.ivs.9500013.

Winckler, M. A., P. Palanque, and C. M. D. S. Freitas, 2004: Tasks and scenario-based evaluation of information visualization techniques. *Proc. of the 3rd Annual Conf. on Task Models and Diagrams*, Prague, Czech Republic, ACM, 165–172, https://doi.org/10.1145/1045446.1045475.

Yoghourdjian, V., D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu, 2018: Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Vis. Inf.*, **2**, 264–282, https://doi.org/10.1016/j.visinf.2018.12.006.

Yucong, Y., F. Sauer, K.-L. Ma, K. Aditya, and J. Chen, 2019: A user-centered design study in scientific visualization targeting domain experts. *IEEE Trans. Vis. Comput. Graph.*, **26**, 2192–2203, https://doi.org/10.1109/TVCG.2020.2970525.