Improving Citizen Science Condition Monitoring Reporting:

Condition Monitoring Scale Bar

Master Project Paper

December 7, 2015

David Eckhardt

Contents

Introduction

The National Integrated Drought Information System (NIDIS) initiative to develop a "drought early warning system" (DEWS) has led the Carolinas Integrated Sciences & Assessment (CISA) group to conduct a pilot citizen science study in South Carolina and North Carolina called the Condition Monitoring Project. The pilot project is establishing a baseline understanding of local conditions. Through qualitative coding of citizen science reports, CISA has worked diligently to produce summarized reports on dry and wet indicators. However, qualitative coding of the citizen science data is very labor intensive, creating a delay between the submission of citizen science data and delivery of summaries to decision makers and stakeholders. Condition Monitoring Project design and data collection needs to be reconsidered in order to increase the utility of the citizen data and to support the development of a drought early warning system in the Carolinas.

Background

The National Integrated Drought Information System (NIDIS) is a national program with an interdisciplinary aim to improve drought monitoring and management throughout the United States. In 1998, the National Drought Policy Act lead to the creation of the National Drought Policy Commission, which was tasked with ensuring collaboration between government agencies on drought related problems. The Commission made a series of recommendations to Congress through a report called *Preparing for Drought in the 21st Century*. Following the recommendations from the Commission, the NIDIS act was signed into law in 2006 and reauthorized in 2014 (National Integrated Drought Information System).

After a series of workshops and meetings with federal, state and local agencies, academic researchers, and other stakeholders, a national conference produced the NIDIS Implementation Plan in 2007. Based on the Implementation Plan, NIDIS has moved forward with a series of initiatives. One initiative aims to create a drought early warning system (DEWS) to provide accurate, timely, and integrated information. In theory, a comprehensive DEWS would monitor the early signs of emerging drought in order to trigger early and appropriate responses to emergent drought, reducing and mitigating drought risk and impacts. NIDIS has further expanded upon its goal of creating an early warning system for drought by striving to "develop pilot programs for the design and implementation of an early warning system in selected locations" (National Integrated Drought Information System).

The Carolinas Integrated Sciences & Assessments (CISA) program is working to advance drought preparedness in North and South Carolina. CISA has developed a DEWS pilot project specific to the Carolinas called the Condition Monitoring Project. Much of CISA's research on this project is done in collaboration with NIDIS and the Community Collaborative, Rain, Hail, and Snow (CoCoRaHS) network. CoCoRaHS is a community-based network of local volunteer precipitation reporters. It is a citizen science project that originated at the Colorado Climate Center at Colorado State University in 1998 and now includes thousands of volunteers across the nation (Community Collaborative Rain, Hail and Snow Network). The goal of the Condition Monitoring Project is to assess the quality and utility of citizen science data for drought decision making in the Carolinas.

CISA has utilized tools and resources developed by CoCoRaHS to expand upon the existing drought impact reporting by recruiting citizen scientists to monitor rainfall and submit weekly condition monitoring reports. The Condition Monitoring Project is intended to develop a baseline of local condition information in order to improve understanding of how the level of precipitation impacts local ecosystems and communities. Additionally, CISA is engaging with CoCoRaHS headquarters, National Drought Mitigation Center (NDMC) Drought Impact Reporter staff, State Climate Offices, and National Weather Service coordinators to assess the availability of reporting tools and methods to the citizen scientist participants and meet NIDIS DEWS objectives.

Problem & Solution

As a part of developing local condition monitoring information, CISA has worked diligently to produce summarized reports on dry and wet indicators derived from qualitative coding of condition monitoring reports. Report content is coded and analyzed using NVivo, a software program used for qualitative content analysis. This form of data is very rich in detail and provides a summary of climate conditions over several months and seasons. However, processing all the information from a condition monitoring report into a usable form is an intensive process and requires a high number of work hours per output. Due to the labor-intensive process, there is a significant lag between report submission by citizen scientists and the delivery of report summaries to decision makers. This lag in submission and delivery limits the utility of citizen science reporting for decision makers.

Unprocessed condition monitoring reports are available online through the CoCoRaHS view data page for drought impacts. This format is efficient at delivering condition monitoring reports in a table view. A user can sort through drought impact reports by searching station id number, state, county, and a specific time range. Querying on the database results in a tabular view of the data, returning the following attributes: start date, end date, station number, state, county, condition monitoring, impact categories, and description. However, this option is time consuming and has limited effectiveness from a decision maker perspective.

Early in Phase I of the Condition Monitoring Project, interviews with decision makers indicated there was a need to provide improved access to data gathered in condition monitoring reports. Hope Mizzell, the South Carolina state climatologist and SC Drought Response Committee Chair, values the precipitation measurement and condition monitoring reports supplied by citizen scientists. Mizzell and other drought decision makers have expressed a need to have data that is delivered in a more processed form, on a more frequent basis, and available online. The lack of efficient access is a shortcoming preventing condition monitoring data from reaching its full potential and utility. A drought metric for reporting conditions could make the assessment of changing conditions more efficient in a short time frame.

Based on the issues and needs outlined above, an overall solution is needed that would enhance the processes of 1) data submission by citizen scientists and 2) the distribution of data to decision makers and stakeholders. First, current data submission should include data that can be more easily processed. Submitting raw text reports produces rich data but there are inherent drawbacks. Currently, Condition Monitoring observers only have the option to report conditions in an open-ended format which produces inconsistency in what is reported in terms of content, as well as spatial and temporal characteristics of the data. It is difficult and time consuming to process Condition Monitoring text reports into a summarized form for end users. Open-ended Condition Monitoring reports are valuable, but data collection needs to implement closed-ended questions in addition to the text reports. Closed-ended questions will create more consistent and structured data that is more easily processed for end users.

Access to the Condition Monitoring report data is limited on the CoCoRaHS website. End users have no easy access to the summarized reports created in NVivo once they are produced by CISA. There are a variety of measures that could be taken to improve the availability of the citizen science data. This study looked at making improvements to the Condition Monitoring pilot project by increasing public availability of Condition Monitoring data and reports. An online mapping application was created to display and visualize condition monitoring data. This online map will give users exploratory capabilities and provide improved spatial context for the data.

The study also looked at revising methods for Condition Monitoring data submission and improving the distribution of data to decision makers and stakeholders. To achieve these goals, research was conducted to assess a Condition Monitoring Scale Bar, technical guidance used with the Scale Bar, and an online mapping application. An iterative process was employed, re-designing aspects of all three solution elements based on feedback from drought decision makers and Condition Monitoring participants.

Developmental work adhered to principles established by the System Development Life Cycle (SDLC). The SDLC is a theoretic framework developed in the 1960's in order to provide structure throughout the creation of large software systems, thus improving project management from beginning to end and producing a high quality end product. The SDLC model defines project steps, which adhere to a schedule for project deliverables (Elliot and Strachan, 2004). The spiral model, a version of the SDLC, is a favored systems development model in information technology and incorporates prototyping into the SDLC (Boehm and Hansen, 2001).

Spiral Model

1) Evaluate current system
2) Identify user/system requirements
3) Design and develop system
4) Evaluate new system
5) Repeat steps 3 – 4 for 1st, 2nd, *n*… prototype iterations
6) Construct final system based on final refined prototype
7) Implement system

To date, work related to the Condition Monitoring Scale Bar has gone through steps 1 and 2 of the Spiral Model, as well as two design iterations (steps 3 & 4). To provide structure, advancements made during the project can be divided into four stages: Stage 1: conceptual development (steps 1 & 2), Stage 2: initial development (1st design iteration), Stage 3: redevelopment (2nd design iteration), and Stage 4: final recommendations. Stage 4 contains final suggestions for the system design as well as future initiatives for Condition Monitoring. Stage 4 recommendations may be used to construct and implement the final system (steps 6 & 7).

Condition Monitoring Scale Bar

Stage 1: Conceptual Development

During conceptual development (Stage 1), literature review of existing citizen science work shows that many citizen science projects have major concerns related to data collection, data validity, and data distribution. Wiggins and Crowston (2011) looked at sociotechnical and macrostructural factors influencing the design and management of participation in citizen science projects by grouping similar projects that share common conditions necessary for success. Wiggins et al. (2011) identified five project types: Action, Conservation, Education, Virtual and Investigation, and analyzed the five project types using scientific, organizational, and technology issues.

Investigation citizen science, such as CoCoRaHS and the Condition Monitoring pilot project, is driven by data collection from the physical environment for the purpose of scientific research. Projects can greatly vary in geographic scope, from a regional level to an international scale. Projects typically provide educational materials to citizen scientists in order to support continued learning. While many conduct research in biology, they can also include projects in meteorology and climatology, such as CISA's CoCoRaHS Condition Monitoring effort.

Projects are often organized by academics with the goal of providing data for formal knowledge production; therefore, a primary scientific concern for investigation citizen science is the validity of data. Well-planned project and task design is paramount, as it directly impacts the generation of reliable results. Data validation plays a role in verifying citizen science contributions. Spatial distribution of citizen science directly impacts the geographic coverage of a project. Unlike virtual projects, the physical location of the volunteers can limit data collection in investigation projects, thus creating bias or gaps in project data collection.

Organization issues originate from the top-down structure common in most investigation projects, making them difficult to manage when there are a large number of project participants. Investigation citizen science projects must contend with long-term sustainability issues due to reliance on funding. Use of web-based technologies greatly varies among investigation citizen science. Projects may employ a fragmented collection of free sites for data collection, like SurveyMonkey and Yahoo SiteBuilder, or have an advanced custom website for data collection. However, data supply issues are common in most investigation citizen science projects. Few projects provide data in readily usable formats for a variety of user types. eBird and the USA National Phenology are notable exceptions.

eBird was launched in 2002 by the Cornell Lab of Ornithology and National Audubon Society. It provides data on bird distribution and abundance throughout the world. eBird participants document the presence or absence of a species, as well as species abundance by using a web-based checklist. By 2015, amateur scientists had reported over 9.5 million bird observations at a variety of spatial and temporal scales. eBird provides observers with a checklist for making bird observations more standardized in the reporting process (eBird).

The USA National Phenology Network (USA - NPN) developed *Nature's Notebook*, which allows volunteers to collect and report standardized ground observations of phenology. Observers monitor phenophases on a regular basis, allowing them to capture the presence and the absence of events (USA National Phenology Network; Sullivan et al., 2014). Similar to the USA - NPN, the Northeast Regional Phenology Network (NE - NPN) coordinates citizen science phenological monitoring for the northeastern United States and eastern Canada. The NE - NPN is an affiliate of the USA - NPN and follows protocol established by NPN in their *Nature's Notebook*, but applies them to specific regional conservation goals.

Drawing from the methodologies used by eBird and the USA-NPN, decision makers were presented the idea of using a standardized question(s) for capturing drought impacts. Response, concerns, and suggestions were noted and used to steer the next stage of development. Decision makers responded by saying consistency is important in considering the credibility of reports. In order to obtain consistent responses from various observers, the drought impact categories would need very clear definitions as to what constitutes each category of drought. Guidance detailing the range of expected conditions for each individual category on the Condition Monitoring Scale Bar will fulfill this need. Feedback suggested reviewing existing drought indices or categories such as the Palmer Drought Severity Index or the U.S. Drought Monitor Categories. Widely used drought indices would give an indication of drought categories and common terms used in drought decision making, providing a foundation for initial development.

Stage 2: Initial Development

Initial development of the Condition Monitoring Scale Bar was based on a select review of well-known drought indices or categories: the Reclamation Drought Index, Palmer Drought Severity Index, and the U.S. Drought Monitor Categories. These three drought metrics are widely used by federal and state agencies as well as scientific researchers. Their use is often pivotal in triggering drought declarations, and consequently the supply of drought disaster relief. Citizen science data should be relatively comparable to existing objective and subjective indicators, allowing the citizen data to be easily integrated into decision maker and stakeholder protocol.

Drought Indices

The Reclamation Drought Index (RDI) was developed by the Bureau of Reclamation and originates from initiatives outlined in the Reclamation States Drought Assistance Act of 1988. The RDI is used as a tool for characterizing drought duration and severity, as well as a trigger in releasing drought emergency relief funds. Temperature, precipitation, snowpack, streamflow, and reservoir levels are inputs for the RDI. The RDI was originally calculated at the river basin level but is adaptable to specific regions. The RDI has six categories: extreme drought, moderate drought, normal to mild drought, normal to mild wetness, moderately wet, and extremely wet (Hayes, n.d.).

The Palmer Drought Severity Index was first developed by W.C. Palmer in 1965 and measures the departure of moisture supply (Palmer, 1965). It was the first comprehensive drought index developed in the United States, leading to its wide use among many states and federal government agencies in drought relief programs (Hayes, n.d.). In order to enable comparisons between locations Palmer strived to make the index a standardized measurement of moisture supply (Palmer, 1965).

There are considerable limitations to the Palmer Index, however it continues to be integral in drought mitigation plans. Three positive characteristics contribute to its continued popularity: it provides a regional measure of weather abnormality for decision makers, it provides a historical perspective to current conditions, and it presents a spatial and temporal description of past drought events (Alley, 1984). The Palmer Drought Severity Index includes eleven categories: extreme drought, severe drought, moderate drought, mild drought, incipient dry spell, near normal, incipient wet spell, slight wet, moderately wet, very wet, and extremely wet.

The U.S Drought Monitor (USDM) provides a high level view of drought conditions in the United States and was established in 1999. NOAA, the U.S. Department of Agriculture, and the National Drought Mitigation Center (NDMC) work collaboratively to produce a weekly map displaying drought conditions. The USDM is a composite index and is derived from climatic, hydrological, and soil condition measurements, as well as observations and impacts reported by more than 350 contributors. The drought map produced from the index is used extensively by government policy makers and the media drought discussion. From 2008 to 2011, the USDA Farm Service Agency used the USDM drought designations to distribute an estimated $1.64 billion in Livestock Forage Disaster Program and $50 million through the Livestock Assistance Grant Program in 2007. In 2012, the USDA altered the process for secretarial drought disaster declarations, allowing for near automatic drought disaster declarations for counties experiencing eight consecutive weeks of severe drought according to the USDM drought intensity categories (National Drought Mitigation Center).

The drought intensity categories classify drought intensity into five categories, derived from the USDM composite index. The list of U.S. Drought categories includes (D0) Abnormally Dry, (D1) Moderate Drought, (D2) Severe Drought, (D3) Extreme Drought, and (D4) Exceptional Drought. The D0 category purposefully uses the adjective “dry” and is considered a drought watch area, meaning that an area may be going into drought or recovering from drought. The drought severity classification is a table that shows the corresponding range for each drought category to five key indicators used in the composite index: Palmer Drought Severity Index, CPC Soil Moisture Model, USGS Weekly Streamflow, Standardized Precipitation Index, and Objective Drought Indicator Blends (National Drought Mitigation Center). The complete summary of the USDM drought categories and the drought severity classification is available in the Appendix (Figure F).

Initial Prototype Development

Initial prototypes for the Condition Monitoring Scale Bar were designed in Google Drawings, a free, web-based vector graphics editing software. It can be used for creating basic website wireframe, flowcharts, diagrams, and logos. The Likert Scale question format was chosen as the format for initial development because of its ubiquitous use in surveys and its well established protocol. Likert Scale question response options are displayed on a horizontal axis, divided in the middle by a neutral response option. Negative response options are usually listed on the left and positive response options are listed on the right. The individual index categories were used in place of the response options found on a Likert Scale question, making them easily transferable to a Likert Scale question format. Dry conditions were listed on the left in place of the negative response options and wet conditions were listed on the right in place of the positive response options. That is typically the arrangement of the three drought indices when they are displayed on a horizontal axis.

The Reclamation Drought Index and the Palmer Drought Severity Index required no additions to their respective categories, as they already included classifications for wet conditions. The U.S Drought Monitor categories, however, only contain classification for dry or drought conditions. In order to create a condition scale that captured wet conditions as well, wet condition categories were created that mirrored the dry classifications in the U.S. Drought Monitor categories. Once the prototypes based on the three indices were created, they were saved as image files for use during decision maker feedback interviews. Examples of the three prototypes are listed below. An additional version of each prototype was created in order to show a modified version that contained fewer categories or used slightly different wording.

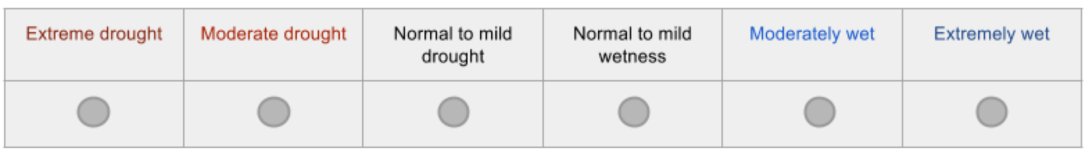


Figure 1: Condition Monitoring Scale Bar prototype based on the Reclamation Drought Index

Figure 2: Condition Monitoring Scale Bar prototype based on the Palmer Drought Severity Index

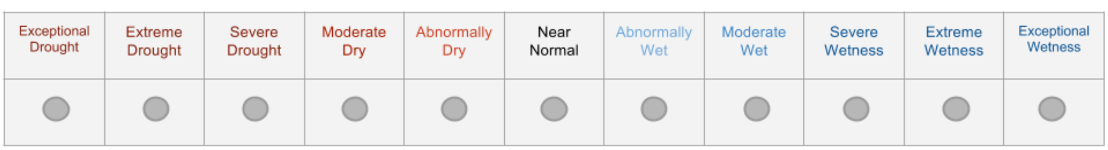


Figure 3: Condition Monitoring Scale Bar prototype based on the US Drought Monitor Categories

Decision Maker Feedback

These three Condition Monitoring Scale Bar prototypes and their respective modified versions were presented to drought decision makers during feedback interviews (Appendix: Figure A). Responses, concerns, and suggestions were noted. The Condition Monitoring Scale Bar based on the U.S Drought Monitor was favored among most interviewees. A division began to emerge related to using all USDM categories or a condensed version for the Scale Bar. The prototype using all USDM categories has eleven choices and the condensed version has nine. Decision makers favored using all USDM categories because of their familiarity with those categories. They also expressed concern about the transferability of the citizen scientist data if the Condition Monitoring Scale Bar categories did not directly match the USDM categories.

Observer Feedback

Observer feedback was solicited during Stage 2. CISA was concerned with the length of the survey and how this might limit participation. Therefore survey participants were only asked two close-ended questions and one open-ended question. The closed-end questions used simple response options (i.e., yes, no, and not sure). Observers were presented with only one of the three different scale bar options, the modified USDM prototype, using nine categories (Appendix: Figure A). A short article explained the logic behind the use of the Condition Monitoring Scale Bar and Web Map. Feedback was collected via a Google Forms questionnaire link. After reading the article and clicking on the link, observers were prompted with an image of the Condition Monitoring Scale Bar and asked questions to gauge their perception and confidence in using the Scale Bar. Instructions for using the Condition Monitoring Scale Bar were not available at that time.

Eleven people answered the survey. No population details for the survey were recorded due to privacy restrictions and concerns about survey length. Participants generally responded well to the idea of a condition monitoring scale bar. The majority of people indicated they would feel confident in using the Condition Monitoring Scale Bar and felt that the Likert scale question would help them in describing conditions in their area. Responses to the question "Would this scale bar help you to describe or report conditions in your area?" consisted of eight answers of **yes** (72.7%), one **no** (9.1%), and two **not sure** (18.2%). Responses to the question "Would you feel confident selecting from these categories?" consisted of eight answers of **yes** (72.7%), and three **not sure** (27.3%). It is interesting to note that the respondents who selected "Not sure" or "No" in the first question are the same respondents that selected "Not sure" in the second question, collectively comprising 27.3% of the responses.

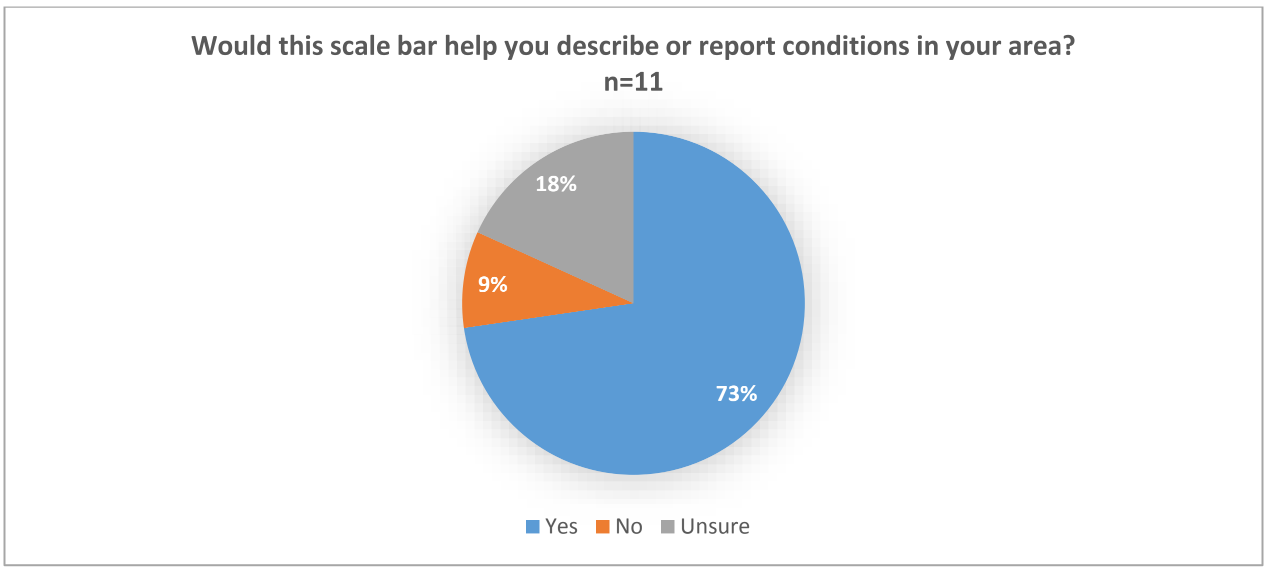


Figure 4: Observer Feedback Survey 1, Question 1

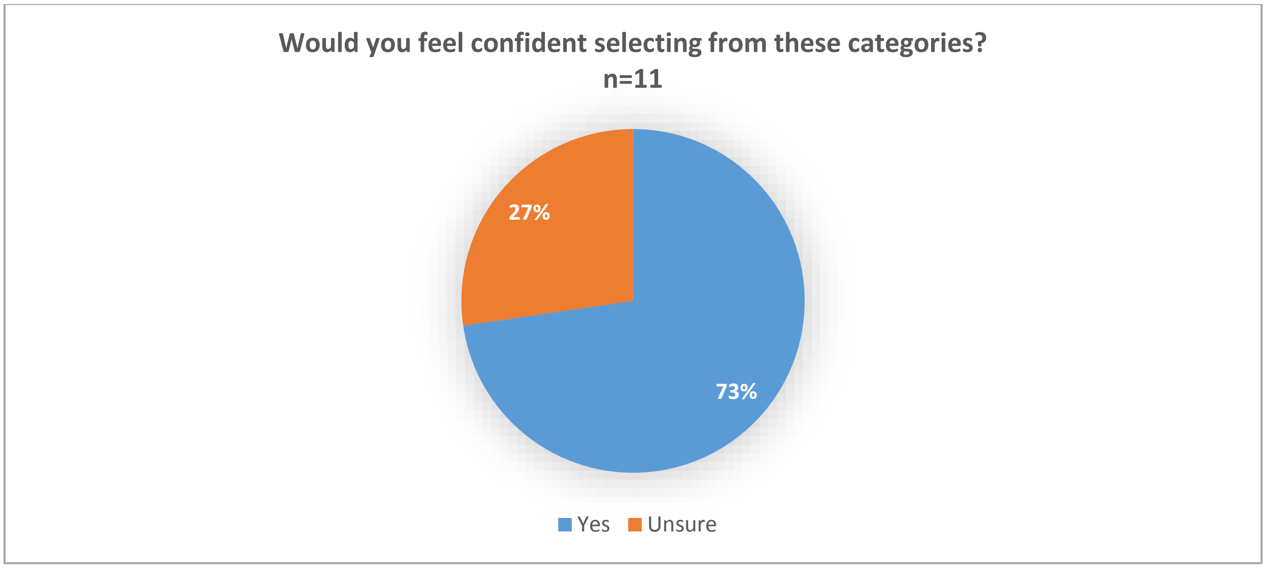


Figure 5: Observer Feedback Survey 1, Question 2

Responses to the open-ended response question revealed that there were too many choices, making it difficult to distinguish between categories like “severe wet” and “extreme wet.” Observers also indicated they wanted additional information or guidance to be able to use the Condition Monitoring Scale Bar. Lastly, some observers found it odd that “abnormally wet/dry” was placed near “normal”, suggesting that it is more logical to place “moderate” closer to normal conditions and “abnormally” closer to the tail-end of the scale. The citizen scientists’ response to the number and ordering of the categories in the USDM-based scale bar raised substantial concern in terms of usability for the observer. Finding a balance between decision maker and stakeholder utility and observer usability is important for the success of the Condition Monitoring Scale Bar (Appendix: Figure B).

Stage 3: Redevelopment

During stage 3, development of the Condition Monitoring Scale Bar sought to find a balance between decision maker utility and observer usability and investigate the possibility of using a sliding scale bar as opposed to the Likert scale buttons. The Condition Monitoring Scale Bar, based on all USDM categories, would likely be too intimidating for citizen scientists. User feedback showed that nine categories may be too many choices for an observer. If so, how can a citizen scientist distinguish between the eleven options using all USDM categories? Arrangement of adjectives can be confusing as well; observers indicated confusion over the position of “abnormal.” Distinguishing between dry and drought may be a difficult decision for citizen scientists. Citizen science research states that one should avoid asking for more than an observer is capable of answering.

Citizen science research shows that projects should avoid being overly ambitious when assigning tasks to amateur scientists. The level of expertise of volunteers must be considered when determining what task a volunteer is capable of completing (Jennett, Blandford, Brohan, and Cox, 2014). Standard protocol dictates that a Likert scale have no more than five to seven categories. Research shows that adding granularity to a Likert scale question does not improve the effectiveness of a Likert scale question (Mattel and Jacob, 1972). Measures of reliability, validity and discriminating power improve with an increasing number of response categories up to roughly seven choices, with reliability decreasing with additional categories beyond seven (Preston and Colman, 2000).

Following standard protocol, a new version of the USDM-based Condition Monitoring Scale Bar was created. The second version had three categories on each side, with neutral in the middle, giving a total of seven categories: severely dry, moderately dry, mildly dry, near normal, mildly wet, moderately wet, severely wet. USDM categories were used as the basis for the new design because it has the most potential for producing citizen science data that is useful to decision makers. The U.S. Drought Monitor is already a mixed methods composite index, thus allowing for greater potential integration of subjective citizen science data into the USDM composite index. Additionally, feedback interviews with drought decision makers has shown using the USDM categories is the preferred option.

Three USDM category descriptors were used to build the Condition Monitoring Scale Bar, with "abnormal" changed to "mildly." The Scale Bar used the same adjectives to describe both dry and wet conditions (mildly, moderately, severe). Usability principles state that designs should aspire to have simplicity, consistency, learnability, and memorability. A consistent severity rating avoids the mixing of adjective forms present in drought indices such as the Palmer Drought Severity Index. These adjectives are common language, avoiding terms like abnormal or incipient. Not using the term "drought" avoids requiring that citizen scientists decide if conditions should be called a drought. Instead, the Condition Monitoring Scale Bar gauges the intensity of dryness or wetness in the area.

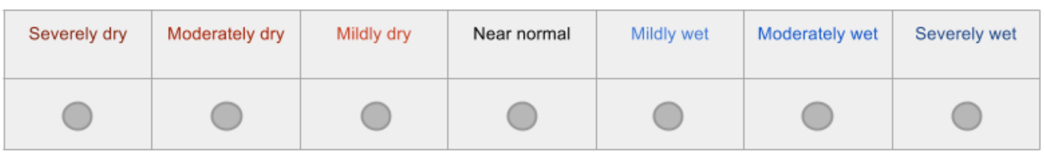


Figure 6: Likert scale-based Condition Monitoring Scale Bar prototype

An alternate scale bar was created based on a sliding scale design, also known as the visual analog scale. The second version retains the seven condition categories but uses a sliding scale for selection rather than radio buttons. Improvements in web development have made it simple to conduct visual analog scale questionnaires online. It is much easier to administer visual analog scale questionnaires because it eliminates the need to hand measure responses collected from a paper and pencil survey. Programing functionality available in a modern browser can quickly calculate the position of the sampled response. The primary argument for a sliding scale surveys is that they are less repetitive and more engaging to survey respondents than the traditional radio-button scale questions (i.e. Likert scales). This assumption is based on the argument that there is more interactivity with a sliding scale, which could lead to less survey fatigue and nonresponse. Proponents argue that sliding scales provide more precision because it measures responses at the interval level, rather than the ordinal level, leading to superior data.

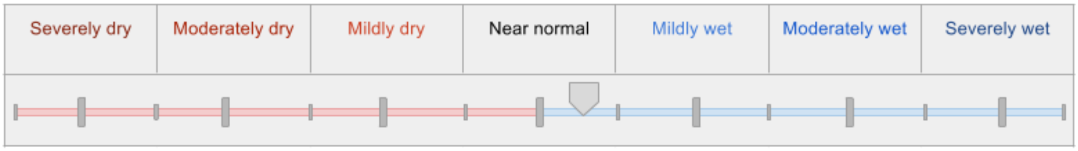


Figure 7: Sliding scale-based Condition Monitoring Scale Bar prototype

Likert Scale vs. Sliding Scale

Protocols for using the radio button or Likert scale question format have been well established due to years of research and use. Use of sliding scale bars on web-based surveys is relatively new and has underlying issues related to its utilization and arguments made by proponents. Alexander Dobronte (2012) reviewed the two primary arguments for sliding scales, and concluded a visual analog scale may create a more pleasing experience for survey respondents. He found that there was a lack of scientific evidence to support the argument that sliding scales produce a higher quality of data. Similarly, Roster et al. (2015) examined the two arguments in favor of sliding scales against published research and found no strong supporting empirical evidence for visual analog scales.

Only a small number of studies have directly assessed respondent satisfaction and engagement between the sliding scale and the traditional radio button scale (Roster et al., 2015). Stanley and Jenkins (2007) found that engagement scores for respondents were statistically higher in those who used a sliding scale survey than respondents who used the traditional radio-button survey, especially for adults between the ages of 25 - 34. However, Stanley and Jenkins (2007) also found that it took longer to complete the sliding scale survey and attributed the additional time to the need to read instructions on how to use and interpret the sliding scale.

Comparatively higher break-off rates and substantially higher response time have been attributed to sliding scales, suggesting that the sliding scale creates a higher cognitive load. Problems associated with sliding scales have been found to be more prevalent in respondents with less than average education (Funke, Reips, and Thomas, 2014). There is no clear evidence from current research supporting the argument that sliding scales improve respondent engagement and satisfaction. Roster et al. (2015) concluded that new forms of sliding scale surveys may be more engaging; however, visual analog scales may lose the novel effect as the survey completion time lengthens. Additionally, trained respondents (i.e. someone who completes a survey once a month or more) are more sensitive to the time it takes to complete a survey, making the respondent more likely to engage in satisfying behavior than a newly recruited respondent (Toepoel, Das, and Van Soest, 2008).

A larger body of research has examined the question of data quality between sliding scale surveys and traditional radio button surveys. Similar to the first argument, the findings are mixed among researchers and have identified issues specific to sliding scales. Unlike radio button scales, the respondent must place the slider at a starting point on the scale. Studies have shown that this inherent characteristic of sliding scales can lead to biased responses. A respondent must move the slider in order to register their response, which can lead to un-recorded responses due to respondents not interacting with the slider (Bayer and Thomas, 2004; Sellers 2013). Sliding scale proponents argue that sliding scale improves validity and reliability because it captures the respondent's opinion more accurately than the traditional Likert scale question (Taylor, 2012). This argument is predicated on the fact that a sliding scale uses a continuous range of values, i.e. 0 to 100, contributing to the claim that the increase in the resolution of the scale range or scale points leads to more precision. Roster et al. (2015) similarly found inconclusive evidence when testing for statistical difference in data quality between the two methods.

Roster et al. (2015) also illustrated two assumptions made by the proponents of sliding scales and improved data quality. The first assumes that the online sliding scale is an exact replica of the standard paper and pencil sliding scale designed in 1932. The second assumes that increasing scale points improves data reliability and validity in a linear fashion. Roster et al. (2015) pointed out that both assumptions ignore the fact that online sliding scale surveys can vary greatly in format, range, and presentation. Research comparing the two methods using Cronbach’s alpha has shown there is a comparatively higher correlation between reliability and the radio button scale; however, that the difference in reliability is small and statistically insignificant (Cook, Heath, Thompson, and Thompson, 2001).

There is no strong body of evidence to fully support either claim of increased engagement or data quality made by sliding scale advocates. Researchers agree that sliding scales can be more engaging in certain regards; however, there are many considerations when using a sliding scale (Puleston, 2011; Sikkel et al., 2014; Roster et al., 2015). There are issues relating to the starting point of the sliding scale, range and labeling protocols, varying slide size and appearance on different devices, higher break-off rates, and increased cognitive load (Puleston, 2011; Funke et al., 2014). If sliding scales are used in a survey, they should include clear instructions and be thoughtfully designed (Roster et al., 2015).

Decision Maker Interviews

Decision makers were interviewed using both forms (i.e., the Likert scale and sliding scale) of the revised Condition Monitoring Scale Bar. Decision makers were also shown the original three prototypes based on the RDI, PDSI, and USDM categories. The initial prototypes were shown to decision makers in order to explain the development process, reference objective indices during discussion, and compare the various versions of the Scale Bar. Feedback continued to be divided among interviewees. Some thought the Condition Monitoring Scale Bar should adhere to all categories used in the USDM scale, while others saw the value in providing a more approachable option for Condition Monitoring observers. Decision makers were not confident that citizen scientists could distinguish between eleven categories.

Decision makers were also divided between using a Likert scale and a sliding scale. Proponents of using the sliding scale argued that it would produce a better quality of data because the interval data would be easier to use in conjunction with other research and the precision would better capture the observers’ answers. Proponents like the idea of being able to place the sliding scale near or in between categories when conditions do not match up perfectly with categories. Decision makers in favor of the Likert scale option liked the simplicity of radio button for decision making and visual appearance.

Given there was no clear consensus among decision makers, the help of an human-computer interaction (HCI) expert, Dr. Jenay Beer, was solicited to provide more clarity in regards to the number of categories and the use of a sliding scale bar or radio buttons. Dr. Beer favored using seven categories with a Likert scale format. Five to seven choice options is the standard used in questionnaires. Dr. Beer pointed out that although the sliding scale and the radio button scale have the same number of categories, the sliding scale format places more cognitive load on the citizen scientist than the Likert scale format. Dr. Beer also provided advice on other aspects of the data entry form used to collect data. For example, at the bottom of the data entry form there is an option to report monetary damage from drought impacts, which is very rarely used by current Condition Monitoring observers. She advised that the monetary damage section be eliminated from the form, stating that it would be difficult for participants to answer that section of the form.

Dr. Beer made several additional suggestions regarding the Condition Monitoring Scale Bar. First, the concept of the Condition Monitoring Scale Bar could be extended to the impact categories section of the form. Each individual category could give the observer the option to select an impact severity from a scaled question format. Second, providing guidance to observers would be absolutely necessary for the Condition Monitoring Scale Bar and should be included on the data entry form. The project design should seek ways to continually reinforce the guidance to observers over time. Based on her experience, without reiterating guidance, an observer will look at the description a few times and never look at it again. Lastly, Dr. Beer suggested that we collect more feedback from observers through evaluation methods.

Observer Feedback

Based on the literature review and findings thus far, observers' input was solicited during a conference call regarding the seven-category radio button scale bar and the data entry form. The call was conducted similarly to a focus group exercise. The Condition Monitoring Scale Bar and guidance was emailed to observers prior to the call. Observers were provided time to view the Condition Monitoring scale bar and guidance during the call, then were asked about their perceptions, opinions, and attitudes on the material provide (Appendix: Figure C). Observer response was recorded during the call and also captured immediately after the call in a Google Forms survey.

The second survey was slightly more extensive than the survey used in Stage 2 and also questioned observers about the category guidance. CISA was concerned about overloading observers with too many questions. As such, survey participants were only asked four close-ended questions and given the option to provide open-end responses. No participant demographic details were recorded due to privacy restrictions and survey length concerns. Observers were presented with only one scale bar option (the modified USDM prototype, using seven categories), due to concern that showing observers too many options would create confusion. Close-end questions used a simple response option (i.e., yes, no, or not sure). Questions 1 and 2 from the first survey were repeated in the second survey to allow for comparison between the first and second survey.

Seventeen people participated in the second survey and responded more confidently to the idea of using a Condition Monitoring Scale Bar. When asked "Would this scale bar help you to describe or report conditions in your area?", fifteen respondents answered **yes** (88.2%) and two respondents answered **not sure** (11.8%). When asked "Would you feel confident selecting from these categories?*"*, fifteen respondents answered **yes** (88.2%). Two respondents answered **not sure** (11.8%).

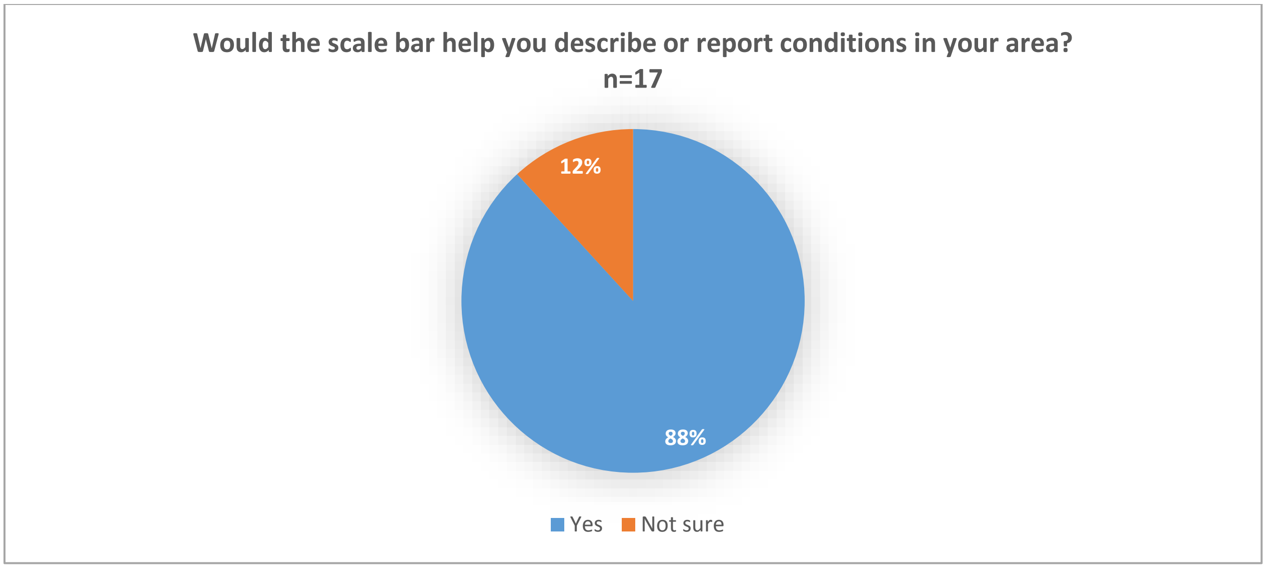


Figure 8: Observer Feedback Survey 2, Question 1

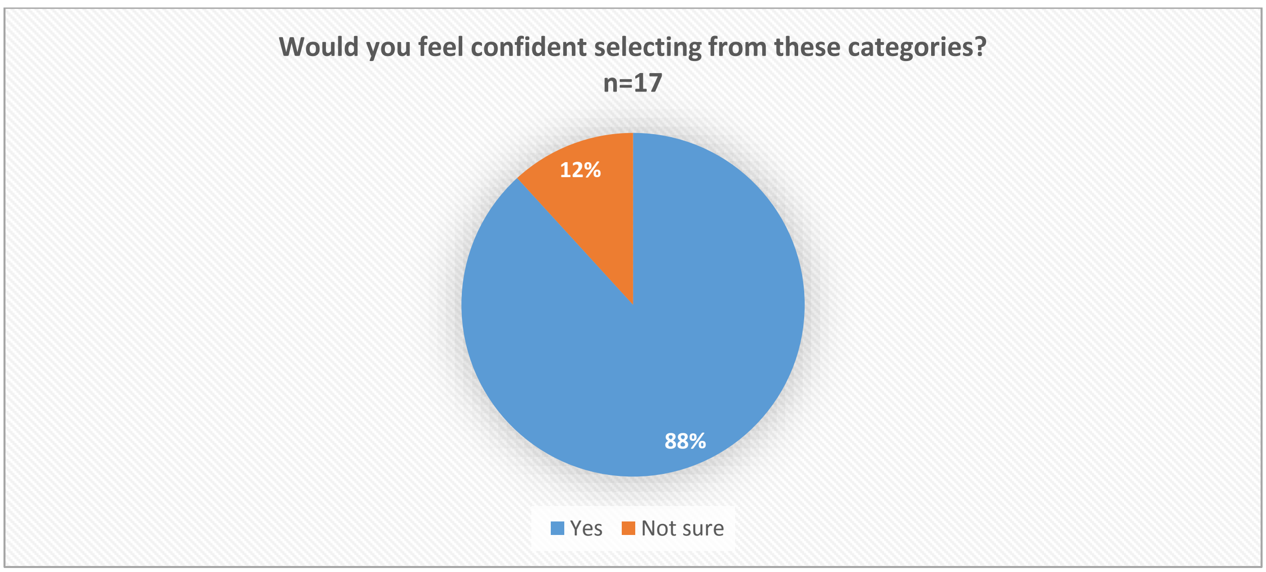


Figure 9: Observer Feedback Survey 2, Question 2

The second survey showed a higher percentage of people indicating they would feel more confident in using the Condition Monitoring Scale Bar and that the Likert scale would help them describe conditions in their area. Observers generally liked the idea, but felt the Condition Monitoring Scale Bar should only supplement the open-ended text box. Some observers felt the Condition Monitoring Scale Bar would help them report more often and identify conditions better than the text reports alone. Observers also suggested the comment space should be extended to the impact categories section of the data entry form.

Additionally, observers were surveyed on their perception of the Condition Monitoring Scale Bar guidance. In general, observers responded strongly in favor of the guidance, feeling that it would be crucial when using the Condition Monitoring Scale Bar. When asked "Would this guidance help you make a selection from dry categories in the scale bar?", all seventeen survey participants answered **yes** (100%). When asked "Would this guidance help you make a selection from wet categories in the scale bar?", sixteen respondents answered **yes** (94.1%) and one answered **not sure** (5.9%).

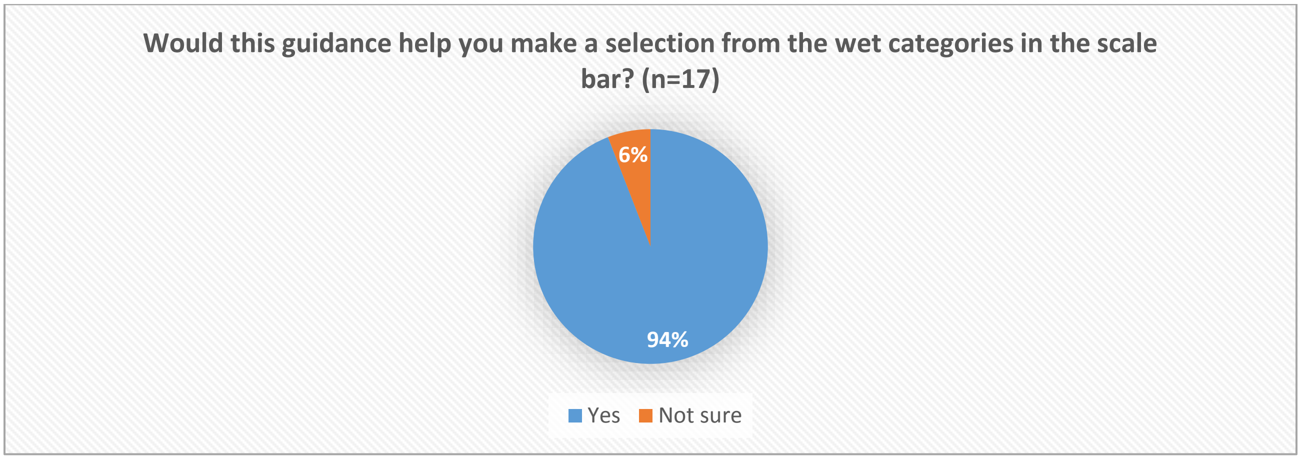


Figure 10: Observer Feedback Survey 2, Question 4

Using subject matter expert interviews (SMEs), additional feedback was solicited in November 2015 from observers who regularly submit condition monitoring reports. SME interviews are often used to obtain insights into the functionality of an existing or updated system, with the goal of improving that system. Seven observers participated in the SME interviews. No participant population details were recorded due to privacy restrictions and concerns about survey length.

Observers completed a mock entry of the Condition Monitoring report form that incorporated the seven-category Condition Monitoring Scale Bar based on the USDM categories (Appendix: Figure D). The process was intended to replicate the process normally used by observers to submit reports. Observers completed the form as they normally would, but were required to make a selection using the scale bar. While completing the form observers talked aloud, detailing their thought process, especially in regards to making a selection on the scale bar. After completing the form, observers were immediately asked a series of open-ended questions (Appendix: Figure E). Depending on the level of response of an observer, more time was taken to discuss particular topics. The audio report and the completed Condition Monitoring report were recorded and saved.

Observer responses indicated that they felt comfortable using the Condition Monitoring Scale Bar and thought the guidance helped. When asked if having more categories would allow them to report more information, all observers said having seven categories was sufficient. They did not feel that adding additional categories would improve their ability to report conditions. The only observer that expressed a desire for more categories was a professional meteorologist, but he also stated that the seven-category Scale Bar was a good balance between simplicity and too many choices on the scale. One observer suggested that a comment space should be extended to the impact categories section of the data entry form, while another wanted the wet categories to have more distinct divisions in the guidance. Many observers reported they did not use the monetary damage portion of the report form and questioned why it was there. Some suggested that it be removed from the data entry form.

Guidance

During the conceptual development stage for the Condition Monitoring Scale Bar it became clear that there would be a need for guidance for the citizen scientists. Clear definitions as to what constitutes each category of dry and wet conditions are necessary in order to obtain consistent responses from condition monitoring observers. This sentiment was independently echoed by some drought decision makers in feedback interviews during the conceptual development stage. Citizen science research also states that projects need to develop protocols for participation. Thorough guidance documents should explain to amateur scientists what they are expected and not expected to do and how they are expected to do it (Conrad and Hilchey, 2011).

Guidance was created for each dry, wet, and neutral category for the Condition Monitoring Scale Bar used in the second stage of evaluation, producing seven descriptions of likely conditions. Guidance is a composite of the “Possible Impacts” categories used in the U.S. Drought Monitor Classification Scheme and the categories derived from the qualitative coding of Condition Monitoring reports during Phase I of the Condition Monitoring pilot project. Guidance for Mildly Dry is written to match D0 (Abnormally Dry), the least severe USDM category. Moderately Dry is tied to the possible impacts list for D1 (Moderate Drought). Severely Dry is associated primarily with D2 (Severe Drought) but extends to include the impacts listed under D3 (Extreme Drought) and D4 (Exceptional Drought). (Appendix: Figure F)

Decision makers and observers provided feedback on the guidance and both agreed that it would be beneficial when using the Condition Monitoring Scale Bar. Specific issues with the guidance arose during feedback discussions. For example, how do you account for variation in regional difference in drought or dry conditions, when what is defined as dry would be very different in Arizona versus South Carolina? Within South Carolina, the understanding of what is dry can vary greatly from the Low Country to the Upstate. How can a temporal aspect be included in the guidance? Seasonal changes, for example, can have a significant impact on local conditions.

Web Map

Two options were initially considered for the online mapping component for the Condition Monitoring pilot project. Option 1 looked at using a mapping API, Leaflet, a leading and popular open source Javascript library for building web mapping applications. Leaflet is used by GitHub, Foursquare, Pinterest, Facebook, Flickr, and Data.gov among many other users (Leaflet). Leaflet’s wide use and popularity make it comparable to major web mapping providers like Google Maps API, Bing Maps API, and OpenLayers. Vladimir Agafonkin is the original developer of Leaf let and made the initial release in 2011. Development is still active, with the first stable release in 2015. Leaflet works well across all major desktop and mobile platforms, and comes with a well-documented API and numerous plugins (Leaflet).

Option 2 looked at using products developed by ESRI, the foremost company in geospatial software. ESRI has web mapping APIs for various languages, which allows developers to create a custom mapping application delivered from ArcGIS Server or ArcGIS Online. ArcGIS online is a cloud-based mapping platform, which allows users to easily load content. In addition to feeding web services to external applications, content can be used within ArcGIS Online to create various cartographic products which can then be shared with user groups, networks, or the general public (ESRI). ArcGIS Online can be used in combination with ESRI's Web AppBuilder. Web AppBuilder provides a foundation for building web applications. Developers and non-developers alike can use Web AppBuilder to configure templates and widgets to create semi-customizable web applications (ESRI).

Both Leaflet and ESRI's Web Mapping API give developers the ability to create web applications from the ground up. The main advantage of this approach is a developer has much greater control over the design and construction of their mapping application. However, creating a web application from the group up is much more time intensive during and after development. Relatively advanced computer science skills are needed to build and maintain a web application with advanced functionality, requiring knowledge of programming languages and browser-based software development. In contrast, ESRI's combination of ArcGIS Online and Web AppBuilder provide the option of developing a web application with little programming experience required. No programming knowledge is required for the continual maintenance of the application.

Several factors and questions were considered in determining the best approach to developing the Condition Monitoring Web Map. CISA's goal was to develop and have a web map available within a five-month period, which is a short time period to implement a research and design process. Questions to investigate with observers and decision makers included: How should the citizen science data be visualized? What additional data would be useful for decision makers? And, how do citizen scientists perceive the web map? The project also required that the CISA team work with CoCoRaHS to determine how the Web Map could access the citizen science-condition monitoring data that is collected and stored by CoCoRaHS at Colorado State University. Technical questions included: How will the data be updated in the web map as more condition monitoring data is collected on a weekly basis? How will this external data be retrieved and processed to make it usable in the web map? In addition, the CISA team has limited experience in web development, making it difficult to overcome any learning curves related to the initial web map development. Any new mapping application would need to be easily maintained by CISA team members and rotating graduate research assistants.

In light of these constraints, utilizing the combination of ESRI's ArcGIS Online and Web AppBuilder was the logical choice. It allowed for the creation of a semi-custom web map application with relatively little demand for site development experience. A basic prototype of the map was created using existing condition monitoring data. The data was modified to include hypothetical examples of Condition Monitoring Scale Bar data. This enabled CISA, during interviews with decision makers and citizen scientists, to more easily explain the concepts related to the Condition Monitoring Scale Bar, as well as the intended use of the data in the mapping application. The prototype used during feedback only included two layers: the Condition Monitoring pilot project area and a subset of condition monitoring reports that was modified to include hypothetical data collected in the Condition Monitoring Scale Bar (Appendix: Figure G).

During interviews in Stage 2 and Stage 3, decision makers indicated that they liked the ability to quickly see the spatial pattern of dry and wet conditions and appreciated seeing reports in a spatial context. The Web Map would allow decision makers to quickly ascertain the conditions, a sentiment which was repeated throughout interviews. Most decision makers indicated having some form of temporal analysis would greatly enhance the utility of data and suggested including additional data (e.g., precipitation data) as a layer in the Web Map. Aggregation of the data was suggested at various levels: county, river basins, or drought management regions. Overall, decision makers responded very positively to the Web Map and felt it would be a good way to utilize Condition Monitoring Scale Bar data. The final version of the online map will be made publicly available to both decision makers and citizen scientists.

Stage 4: Final Recommendations

There are limitations with the current work that are important in considering future project directions and decisions. Decision makers provided a large amount of feedback through open-ended discussions of the Condition Monitoring Scale Bar, the accompanying guidance, and the Web Map. Input was captured through the recording of interviews, providing valuable qualitative data. However, decision makers were never systematically surveyed on their opinions and the qualitative data is difficult to process. Structured questionnaires gauging decision maker preference on the different types of scale bars would provide quantitative data that could support the qualitative findings in the audio transcriptions.

No participant population details were recorded during observer feedback due to privacy restrictions and concerns regarding survey length. CISA has worked closely with the Condition Monitoring volunteers and is familiar with the demographics of these citizen scientists. Most observers are older working professionals or retirees living in North Carolina or South Carolina. The omission of population data is a significant limitation to the observer feedback interviews. It obscures potential insight that might be derived from more comprehensive surveying. The sample size for all surveys (12, 17, and 7 respectively) is very small and polls mostly the same group of people. This is not an issue for the SME interviews due to the nature of the evaluation method but could be important in the other questionnaires.

There are 40 observers who regularly participate in the Condition Monitoring pilot project, making it difficult to get a high participation rate in an optional survey. Furthermore, all observers are already very familiar with the project and live within a region that experiences very similar dry and wet conditions. It could be useful to conduct more comprehensive usability studies and survey larger populations who do not have existing knowledge of condition monitoring and who inhabit areas with different conditions than the Carolinas.

Despite these limitations, efforts over the past five months are significant enough to produce conclusions and recommendations for the Condition Monitoring Scale Bar. The literature review and feedback from decision makers and Condition Monitoring observers will provide direction for the Condition Monitoring pilot project in the future. A standardized drought metric in the form of the Condition Monitoring Scale Bar will improve reporting and should be implemented using a seven-category Likert scale format. This scale bar design will also limit cognitive load for the observers.

Striking a balance between decision maker utility and observer usability is important; the proposed scale bar achieves this balance. The Condition Monitoring Scale Bar categories and category guidance should be based on the USDM. In assessing the proposed Condition Monitoring Scale Bar, it could be argued that the final, recommended Scale Bar has elements of the three original quantitative indices: the RDI, the PDSI, and the USDM (Appendix: Figure H). The proposed scale bar is similar to the RDI in the small number of categories in both scales (7 and 6 respectively) and the category descriptions are similar (RDI: mild, moderately, extremely versus scale bar: mildly, moderately, severely). The proposed scale bar is similar to the PDSI because both scales have the near normal category. The three dry and wet categories in the proposed scale bar are similar to the PDSI if the Incipient Dry Spell category is eliminated. The proposed Condition Monitoring Scale Bar is constructed to capture the initial three categories of the USDM categories. In theory, this is accomplished primarily by creating guidance that will classify conditions in a similar fashion to that of the possible impacts in the USDM classification scheme. Creating guidance that is comparable to the USDM classification scheme will help produce citizen science reports that are integrated with the USDM categories.

The initial three USDM categories are used to construct the Condition Monitoring scale bar and the abnormal category is changed to “mildly.” In this form the Condition Monitoring Scale Bar does not capture all the possible USDM categories, but does use the D0 (abnormally dry) to D2 (severely dry) ratings. The original purpose of the pilot project was to develop a tool to support a drought early warning system. The Condition Monitoring Scale Bar adheres to the DEWS goals by focusing its design and construction on the USDM categories and impacts that are related to drought onset and recovery. Adding extreme and exceptional to the scale does not necessarily provide an early warning. If a region is in extreme or exceptional drought, it is likely that decision makers and stakeholders have already been aware of dry conditions for quite some time (Appendix: Figure H).

Ideas for future research include a variety of issues related the concept of a condition monitoring scale bar, the existing data entry page on the CoCoRaHS website, condition monitoring guidance, and the Condition Monitoring Project protocol. Decision makers may find it desirable to capture very extreme dry or wet conditions. A second Likert scale bar could be developed that captures more extreme conditions and is deployed on the data entry page only when triggered by certain conditions. For example, if a D2 (severe) declaration or above (D3, D4) is made by the USDM, a second scale bar utilizing D2 - D4 could be made available for citizen scientists, who are using a region-specific portal of the CoCoRaHS website. There are many questions related to use of an extreme Condition Monitoring scale bar; it may prove beneficial to the collection of citizen science data.

The monetary reporting for category impacts should be eliminated from the report form. The lack of monetary values reported in the data entry, as well as observer feedback, shows that this portion of the data entry page is not used. Observers do not feel comfortable using this report option and likely will not use the monetary reporting feature in the future. Instead the Condition Monitoring Scale Bar should extend to the impacts section of the form. For each individual category, the observer could select an impact severity from a scaled question format. This would capture category impacts and replace the unutilized monetary reporting with a reporting option that would likely be used by citizen scientists. The resulting data could be incorporated into condition summaries and mapping layers, visualizing data based on the severity rating selected from a Likert scale question.

Guidance has been shown to be very crucial to the implementation of the Condition Monitoring Scale Bar. Moving forward, standard and recurring guidance needs to be developed that is general enough to be applicable to all regions. However, it is clear that regional guidance would also be helpful. There is enough climate variation from region to region to support regional guidance. If the pilot project is scaled to a national level, national and regional guidance could be presented in unison to citizen scientists. Temporal and seasonal elements and dry/wet season guidance should be considered as well.

The Condition Monitoring Project is a pilot project. If Condition Monitoring is to be scaled to a national level, project design and protocols must be considered at the regional and national level. To illustrate regional differences in Condition Monitoring, a structured regional version of the website could be extended from the main CoCoRaHS site. Or citizen scientists could be directed to a user profile and data entry page specific to their user account. These methods are currently used by both ebird and USA-NPN and could provide direction for Condition Monitoring if it is scaled to a national level.

Lastly, more work needs to be done to improve the quality of Condition Monitoring reporting. For example, protocols need to instruct citizen scientists when to submit Condition Monitoring reports, perhaps on Saturdays and Sundays, and what time range the report should capture, perhaps one week. Until this happens, there will be inconsistencies across reports, i.e., observers will report at different frequencies (e.g., weekly or monthly) and/or on different days of the week or month. Spatial reporting guidelines could also be introduced. Currently, condition monitoring observers primarily report on conditions near their home or neighborhood, but some observers report on conditions at broader spatial scales. Some form of guidance on spatial considerations could be incorporated into the observer guidance.

Conclusion

The current form of Condition Monitoring has succeeded in creating a baseline of local conditions. Citizen science reports have helped to improve the understanding of how the level of precipitation impacts local ecosystems and communities. Moving forward, standardized reporting for Condition Monitoring will make it easier and quicker to assess changing conditions in a short time frame. Data will be more consistent and easier to summarize and display through web map applications. This will enhance the utility of the data for decision makers and stakeholders, thus furthering the development of an early warning system for drought onset and recovery.

Sources

Alley, W. M. 1984. The Palmer drought severity index: limitations and assumptions. *Journal of Climate and Applied Meteorology 23*(7): 1100-1109.

Bayer, L. R. and R. K. Thomas. 2004. A comparison of sliding scales with other scale types in online surveys. *RC33 Sixth International Conference on Social Science Methodology*, August 16-20, 2004. Amsterdam, Netherlands.

Boehm, B. and W. Hansen. 2001. The spiral model as a tool for evolutionary acquisition. *CrossTalk 14*(5): 4-11.

Community Collaborative Rain, Hail & Snow Network. About Us. http://www.cocorahs.org/Content.aspx?page=aboutus

Conrad, C. C. and K. G. Hilchey. 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment 176*(1-4): 273-291.

Cook, C., F. Heath, R. L. Thompson, and B. Thompson. 2001. Score reliability in web- or Internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement 61*(4): 697-706.

Couper, M. P., R. Tourangeau, F. G. Conrad, and E. Singer. 2006. Evaluating the effectiveness of visual analog scales a web experiment. *Social Science Computer Review 24*(2): 227-245.

eBird. *About eBird*. http://ebird.org/content/ebird/about/

Elliott, G. and J. Strachan. 2004. *Global Business Information Technology*, 87.

Dobronte, A. 2012. Likert scales vs. slider scales in commercial market research. August 21, 2012. https://www.checkmarket.com/2012/08/likert_v_sliderscales

ESRI. ArcGIS Online. http://www.esri.com/software/arcgis/arcgisonline

ESRI. Web AppBuilder for ArcGIS. http://www.esri.com/software/web-appbuilder

Eveleigh, A., C. Jennett, A. Blandford, P. Broha, and A. L. Cox. 2014. Designing for dabblers and deterring drop-outs in citizen science. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 2985-2994.

Funke, F., U.-D. Reips, and R. K. Thomas. 2011. Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review 29*(2): 221-231.

Hayes, M. J. n.d. Comparison of Drought Indices. National Drought Mitigation Center. http://drought.unl.edu/Planning/Monitoring/ComparisonofDroughtIndices.aspx

Kero, P. and D. Lee. 2015. Slider Scales and Web-Based Surveys: A Cautionary Note. *Journal of Research Practice 1*(1): 1.

Leaflet. *Overview*. http://leafletjs.com/

Matell, M. S., and J. Jacoby. 1971. Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity. *Educational and Psychological Measurement 31*: 657-674.

National Drought Mitigation Center. *U.S. Drought Monitor Background*. http://droughtmonitor.unl.edu/AboutUSDM/Background.aspx

National Drought Mitigation Center. *U.S. Drought Monitor Classification Scheme*. http://droughtmonitor.unl.edu/aboutus/classificationscheme.aspx

National Integrated Drought Information System. *What is NIDIS?* https://www.drought.gov/drought/what-nidis

Palmer, W. C. 1965. *Meteorological drought*. Research Paper No. 45. Washington, DC: US Department of Commerce Weather Bureau.

Preston, C. C., and A. M. Colman. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica 104*(1): 1-15.

Roster, C. A., L. Lucianetti, and G. Albaum. 2015. Exploring slider vs. categorical response formats in web-based surveys. *Journal of Research Practice. 11*(1). Article D1. http://jrp.icaap.org/index.php/jrp/article/view/509/413

Sellers, R. 2013. How sliders bias survey data. *Alert! 53*(3): 56-57.

Sikkel, D., R. Steenbergen, and S. Gras. 2014. Clicking vs. dragging: Different uses of the mouse and their implications for online surveys. *Public Opinion Quarterly 78*: 177-190

Sullivan, B. L., J. L. Aycrigg, J. H. Barry, R. E. Bonney, N. Bruns, C. B. Cooper, et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation 169*: 31-40.

Stanley, N., and S. Jenkins. 2007. Watch what I do: Using graphical input controls in web surveys. In M. Trotman, T. Burrell, L. Gerrard, K. Anderton, G. Basi, M. Couper, et al. (Eds.), *The challenges of a Changing World. Proceedings of the Fifth International Conference of the Association for Survey Computing*, pp. 81-92. Berkeley, Glos, UK: Association for Survey Computing.

Taylor, I. 2012, Use slider scales for a more accurate rating. June 8, 2012. https://blog.questionpro.com/2012/06/08/use-slider-scales-for-a-more-accurate-rating/

Toepoel, V., M. Das, and A. V. Soest. 2008. Effects of design in web surveys: Comparing trained and fresh respondents. *Public Opinion Quarterly 72*(5): 985-1007.

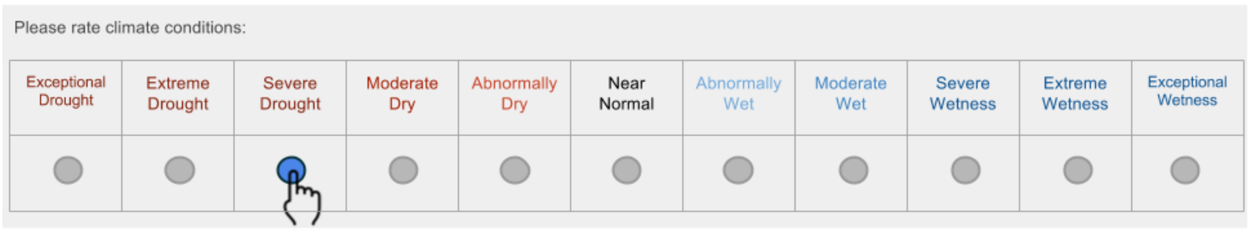
USA National Phenology Network. About Us. https://www.usanpn.org/about

Wiggins, A. and K. Crowston. 2011. From conservation to crowdsourcing: A typology of citizen science. HICSS '11 Proceedings of the 2011 44th Hawaii International Conference on System Sciences, pp. 1-10. Washington, DC: IEEE Computer Society.

Appendix

Figure A: Form for decision maker feedback in Stage 2

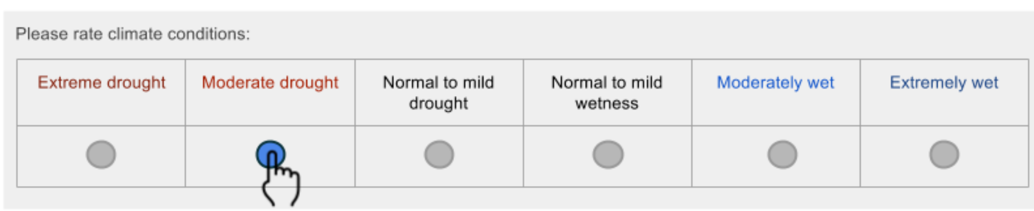
US Drought Monitor-based Condition Monitoring Scale Bar prototypes



Scale 1: All USDM dry categories and created matching wet categories

Scale 2: Condensed version the USDM categories (**the version used in the questionnaire**)

Reclamation Drought Index-based Condition Monitoring Scale Bar prototypes



Scale 1: Based on all of the original RDI categories

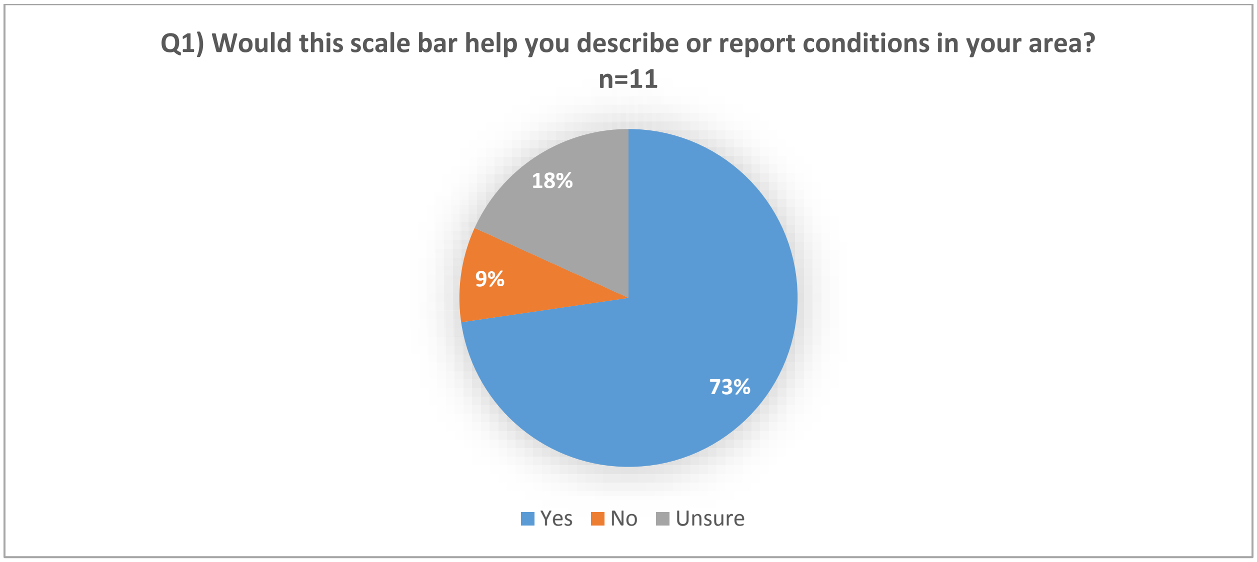
Scale 2: Expands upon the RDI categories by adding a "Near normal" category in the center and shortens both of the mild categories

Palmer Drought Severity Index

Scale 1: The Likert scale is based on all Palmer Index categories

Scale 2: Condensed version of the Palmer Index categories

Figure B: Additional results from Observer Feedback Survey 1



*Note: The respondents that selected “Not sure” or “No” in the first question are the same respondents that selected “Not sure” in the second question, collectively comprising 27.3% of the responses.

Q3) Please provide feedback if you have any suggestions, comments, or questions about the idea to incorporate this scale bar into the CoCoRaHS report form.

Comments from observers who responded “yes” in the first and second questions:

- “Sorry for the double answer in #2. It would depend on the definitions of each condition and how hard it is to measure something tangible to get the "right" answer.”
- “It is a good idea. Our area, for example, seemed to not get rained on as much as the city just 3-4 miles away during the past summer. This must impact the small farmers down the road from us (hwy 21 in Northwest of Orangeburg bordering with Calhoun County).”
- “You have too many choices. Difficult to distinguish between say severe and extreme wet. I would suggest normal, moderate wet, extreme wet and follow suit on the dry side. But this rating system is a good idea.”
- “Nature is fuzzy and does not always follow discrete models. These scales offer a choice of discrete values that I often find are hard to quantify with specificity and confidence. If you allow more than one radio button to be selected, I would be more comfortable submitting a report. For example, if some of the indicators I'm looking at point to one value (for example, "Normal" while at the same time other indicators may point to another value that may be adjacent to the original value or even two or more places away, it is a simple algorithm to average the values I've identified to get one overall score. I realize a combined rating of, say, Normal and

Abnormally Dry may seem confusing to some, however, in my experience, I think it would better fit what I am observing in the real world. "

- "I have found over the time I have been reporting on climate conditions, that sometimes, at THE moment you are writing your report, the condition is/will-be/or has just recently changed. I think the participants would appreciate a way to note this fact; as your star of this month noted in one of his reports... "After ample rain in June and July. The past 11 days have been.... "
  To just have "the" bar, would be frustrating if we could not also note some comments such as "Just yesterday, after weeks of... (whatever)"
  Just a thought for you to consider"

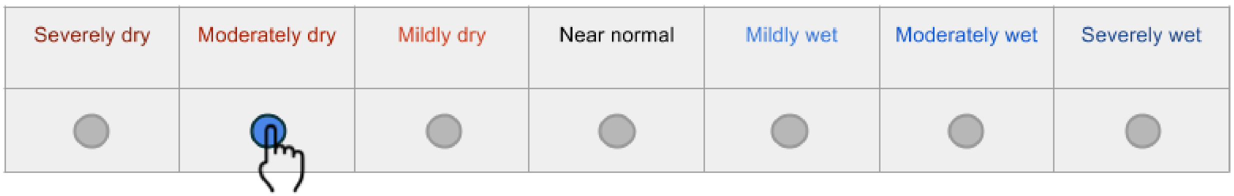
Comments from observers who responded "no" or "not sure" to the first and second question:

- "Since I run an irrigation system at my home, it is not possible to evaluate what actual drought conditions are."
- "I could not answer your questions without more information. I would like to see the guidance/training that goes with the scale in order to answer the questions. "
- "To me the Moderate Wet and Moderate Drought should logically be placed near Normal and Abnormally Wet and Abnormally Dry outside of the Moderate items. Just a thought. "

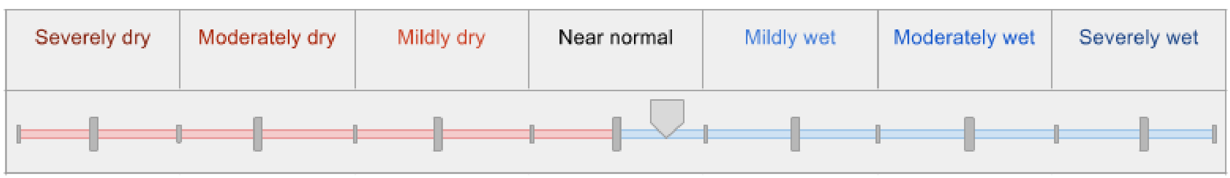
Figure C: Form used for second round of observer feedback

Proposed Condition Scale:

Based on your local knowledge and experience, please select a condition that best describes your Condition Monitoring area from the options below:

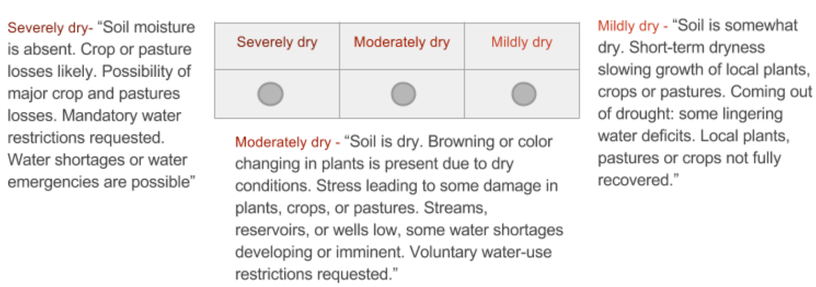


Option 1



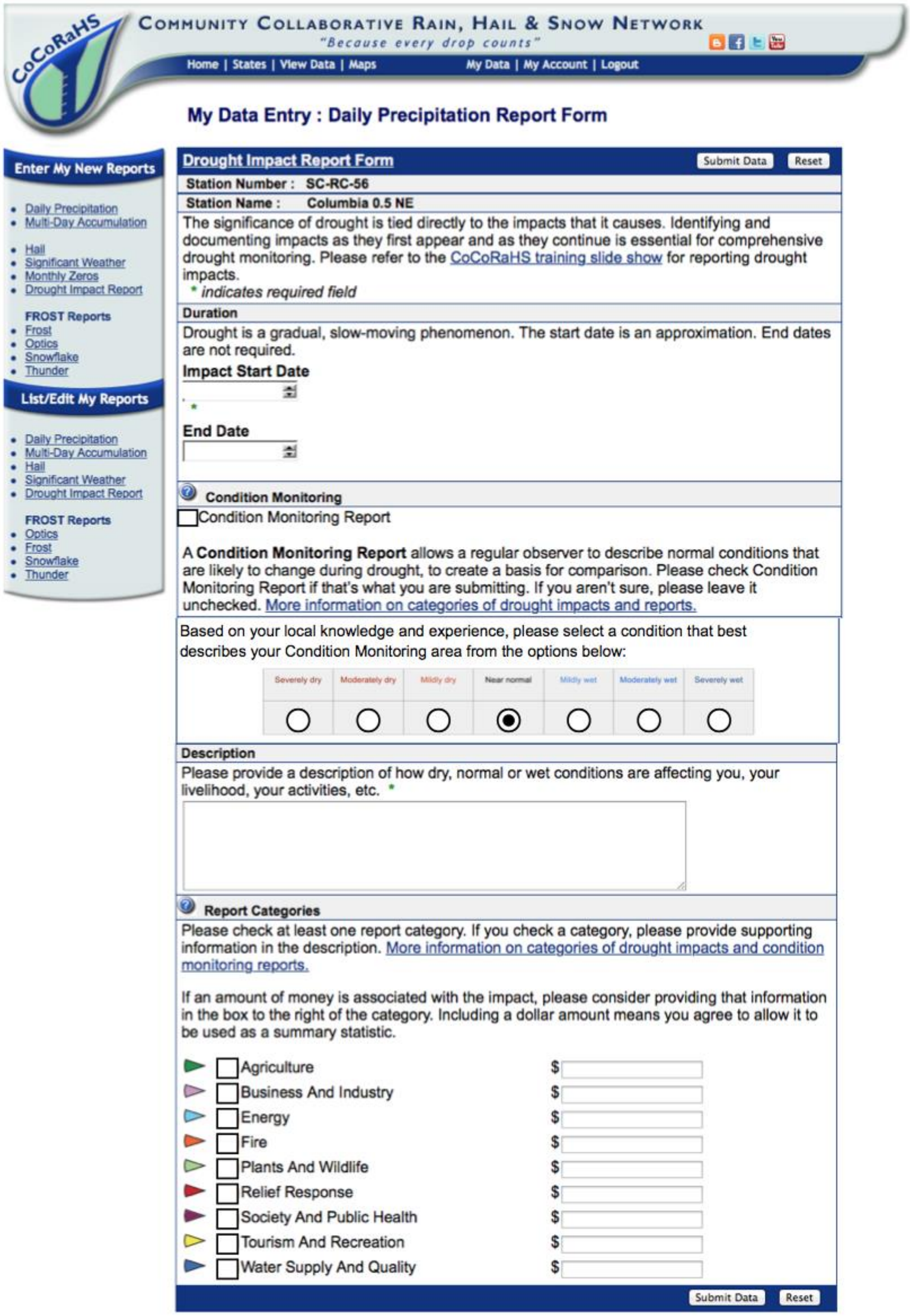
Option 2

Dry guidance - general descriptive information on dry conditions will be included with each category in order to provide help in making a selection.



Wet guidance - general descriptive information on wet conditions will be included with each category in order to provide help in making a selection.

Figure D: Screenshot of the mock condition monitoring report form



CoCoRaHS

COMMUNITY COLLABORATIVE RAIN, HAIL & SNOW NETWORK

"Because every drop counts"

Home | States | View Data | Maps   My Data | My Account | Logout

My Data Entry : Daily Precipitation Report Form

Enter My New Reports

- Daily Precipitation
- Multi-Day Accumulation
- Hail
- Significant Weather
- Monthly Zeros
- Drought Impact Report

FROST Reports

- Frost
- Optics
- Snowflake
- Thunder

List/Edit My Reports

- Daily Precipitation
- Multi-Day Accumulation
- Hail
- Significant Weather
- Drought Impact Report

FROST Reports

- Optics
- Frost
- Snowflake
- Thunder

Drought Impact Report Form   Submit Data   Reset

Station Number : SC-RC-56

Station Name : Columbia 0.5 NE

The significance of drought is tied directly to the impacts that it causes. Identifying and documenting impacts as they first appear and as they continue is essential for comprehensive drought monitoring. Please refer to the CoCoRaHS training slide show for reporting drought impacts.

* *indicates required field*

Duration

Drought is a gradual, slow-moving phenomenon. The start date is an approximation. End dates are not required.

Impact Start Date

*

End Date

Condition Monitoring

☐ Condition Monitoring Report

A **Condition Monitoring Report** allows a regular observer to describe normal conditions that are likely to change during drought, to create a basis for comparison. Please check Condition Monitoring Report if that's what you are submitting. If you aren't sure, please leave it unchecked. More information on categories of drought impacts and reports.

Based on your local knowledge and experience, please select a condition that best describes your Condition Monitoring area from the options below:

Severely dry	Moderately dry	Mildly dry	Near normal	Mildly wet	Moderately wet	Severely wet
○	○	○	◉	○	○	○

Description

Please provide a description of how dry, normal or wet conditions are affecting you, your livelihood, your activities, etc. *

Report Categories

Please check at least one report category. If you check a category, please provide supporting information in the description. More information on categories of drought impacts and condition monitoring reports.

If an amount of money is associated with the impact, please consider providing that information in the box to the right of the category. Including a dollar amount means you agree to allow it to be used as a summary statistic.

- ☐ Agriculture $
- ☐ Business And Industry $
- ☐ Energy $
- ☐ Fire $
- ☐ Plants And Wildlife $
- ☐ Relief Response $
- ☐ Society And Public Health $
- ☐ Tourism And Recreation $
- ☐ Water Supply And Quality $

Submit Data   Reset

Figure E: Script and questions used during observer SME interviews

Hi ________ , this is David Eckhardt from CISA. How are you doing this morning?

Thanks for setting aside time to participate in the usability for the condition monitoring scale bar.

The call should not take much of you time. I would like to give you a quick overview of the user study. Then I will have you complete the Data Entry Form. Lastly, I will get you to answer a few questions after you have completed the form. Does that sound good?

Okay, please open the PDFs called the *Data Entry Form* and *Guidance*. These PDF forms are attached to the email I sent on Friday. The data entry form is designed to simulate the standard condition monitoring report submission process. At this point in the report submission process you have already logged into the CoCoRaHS website and navigated to the *My Data Entry* page for condition monitoring reports. The Data Entry Form is an editable PDF. This means you be able to click and write in portions of the PDF, just like you would when submitting your report on the CoCoRaHS website.

To prepare for the usability test we asked you to observe conditions in your area as you normally would. We also asked you to consider how observed conditions fall within the condition monitoring categories. Please think back to those observations when filling out the form. The attached document, Guidance, provides information on the condition monitoring scale bar and its categories. You can use the guidance to help you fill out data entry form.

Do you have the form open? Okay, great. You are ready to complete and submit the condition monitoring report using the data entry form. As you are filling out the form please think out loud. Talk about your thought process when making a selection on the condition scale bar and when writing report. Talk about anything that comes to mind when completing the form. Just, tell me what section you are on when filling out the form. Do you have any questions? Okay, you can start whenever you are ready.

These are open ended questions. You can elaborate on your answer as much you want, or provided a simple yes or no. It is up to you.

1. Did you think the condition monitoring scale bar was easy to use?

2. Was the scale bar well integrated into the condition monitoring report form?

3. Did the condition monitoring scale bar help you describe or report conditions in your area?

4. Do you feel that having more than 3 choices of either wet or dry conditions would allow you to provide more information?

5. Did the scale bar limit your ability to describe or report conditions in your area?

6. Did the guidance help you make a selection from dry or wet categories on the scale bar?

7. Would you need the support of more technical guidance to be able to use the condition monitoring scale bar?

8. Was there any inconsistency in the condition monitoring scale bar categories?

9. Do you think that most people would learn to use the condition monitoring scale bar very quickly or easily?

10. Did you feel confident selecting from the condition monitoring scale bar?

11. Is there any changes you would make the CMSB or the CM report form?

Figure F: US Drought Monitor categories used in Scale Bar Guidance development

Condition Monitoring Dry Values	Category	Description	Possible Impacts	Palmer Drought Severity Index (PDSI)	CPC Soil Moisture Model (Percentiles)	USGS Weekly Streamflow (Percentiles)	Standardized Precipitation Index (SPI)	Objective Drought Indicator Blends (Percentiles)
Mildly dry (D0)	D0	Abnormally Dry	Going into drought: • short-term dryness slowing planting, growth of crops or pastures Coming out of drought: • some lingering water deficits • pastures or crops not fully recovered	-1.0 to -1.9	21 to 30	21 to 30	-0.5 to -0.7	21 to 30
Moderately dry (D1)	D1	Moderate Drought	• Some damage to crops, pastures • Streams, reservoirs, or wells low, some water shortages developing or imminent • Voluntary water-use restrictions requested	-2.0 to -2.9	11 to 20	11 to 20	-0.8 to -1.2	11 to 20
Severely dry (D2 - D4)	D2	Severe Drought	• Crop or pasture losses likely • Water shortages common • Water restrictions imposed	-3.0 to -3.9	6 to 10	6 to 10	-1.3 to -1.5	6 to 10
	D3	Extreme Drought	• Major crop/pasture losses • Widespread water shortages or restrictions	-4.0 to -4.9	3 to 5	3 to 5	-1.6 to -1.9	3 to 5
	D4	Exceptional Drought	• Exceptional and widespread crop/pasture losses • Shortages of water in reservoirs, streams, and wells creating water emergencies	-5.0 or less	0 to 2	0 to 2	-2.0 or less	0 to 2

Figure G: Prototype Web Map

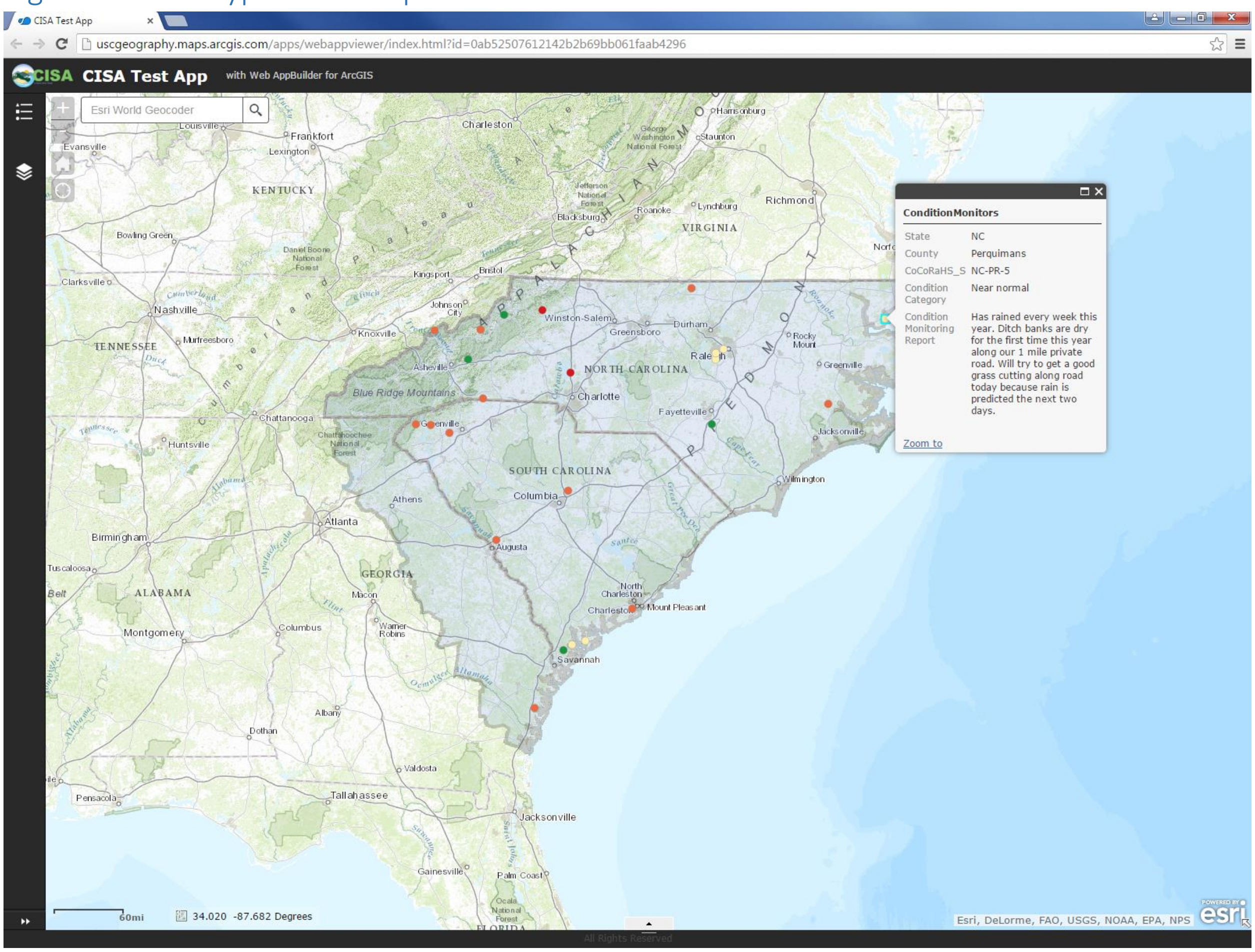


Figure H: Construction of Condition Monitoring Scale Bar categories