The making of a metric: Co-producing decision-relevant climate science

Kripa Jagannathan [a,b], Andrew D. Jones [a], Isha Ray [b]

[a] Lawrence Berkeley National Laboratory

Earth & Environmental Sciences Area – Berkeley Lab

1 Cyclotron Road

Berkeley, CA 94720

USA

[b] University of California, Berkeley

Energy and Resources Group,

Berkeley, CA 94720

USA

Corresponding author: Kripa Jagannathan (kripajagan@berkeley.edu)

Other authors: Andrew Jones (adjones@lbl.gov), Isha Ray (isharay@berkeley.edu)

Early Online Release: This preliminary version has been accepted for publication in *Bulletin of the American Meteorological Society*, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-19-0296.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Capsule Summary

Understanding how science is co-produced is a science unto itself. Using the case of Project Hyperion, we illustrate how co-production works (or does not work) in practice.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Abstract

Developing decision-relevant science for adaptation requires the identification of climatic parameters that are both actionable for practitioners as well as tractable for modelers. In many sectors, these decision-relevant climatic metrics and the approaches that enable their identification remain largely unknown. “Co-production” of science with scientists and decision-makers is one potential way to identify these metrics, but there is little research describing specific and successful co-production approaches. This paper examines the negotiations and outcomes from Project Hyperion, wherein scientists and water managers jointly developed decision-relevant climatic metrics for adaptive water management. We identify successful co-production strategies by analyzing the project’s numerous back-and-forth engagements and tracing the evolution of the science during these engagements. We found that effective mediation between scientists and managers needed dedicated “boundary spanners” with significant modeling expertise. Translating practitioners' information needs into tractable climatic metrics required direct and indirect methods of eliciting knowledge. We identified four indirect methods that were particularly salient for extracting tacitly-held knowledge and enabling shared learning: developing a hierarchical framework linking management issues with metrics; starting discussions from the planning challenges; collaboratively exploring the planning relevance of new scientific capabilities; and using analogies of other ‘good’ metrics. The decision-relevant metrics we developed provide insights into advancing adaptation-relevant climate science in the water sector. The co-production strategies we identified can be used to design and implement productive scientist-decision-maker interactions. Overall, the approaches and metrics we developed can help climate science to expand in new and more use-inspired directions.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Introduction

Adaptation practitioners across many sectors, including resource management, land-use planning, and public health, urgently need decision-relevant science to plan for and manage the impacts of climate change (ACCNRS 2015; Moss et al. 2013; Lemos and Morehouse 2005; Kirchhoff et al. 2013a; Kerr 2011). There have been several efforts towards developing actionable (or decision-relevant) science broadly, and more specifically towards providing scientific details of the climate impacts that planners need to account for (Mach et al. 2019; Bremer and Meisch 2017; Beier et al. 2017). Resource managers, however, still report that climate information that can help to develop adaptation decisions, is not readily available to them (Moss et al. 2019; Barsugli et al. 2013; USGAO 2015; Vogel et al. 2016). This is partly on account of unresolved mismatches between scientists' and decision-makers' perceptions of what constitutes 'actionable' climate information (Lemos et al. 2012; McNie 2007). One important example of this mismatch is that current climate modelling and model evaluation efforts typically focus on broad climatological metrics, such as averages or extremes in temperature and precipitation. However, in order to be actionable, resource managers need information on management-specific metrics, such as the start date of the rainy season or number of extreme heat days in the summer (Briley et al. 2015; Roncoli et al. 2009; Moss et al. 2019; Bornemann et al. 2019). This lack of focus on management-specific climate science can preclude its use in adaptation decisions, as even translation or communication of such broader information cannot move the science "off the shelf" to make it usable (Moss et al. 2019; Lemos et al. 2012; Hackenbruch et al. 2017).

The literature recognizes the importance of determining specific climatic metrics that could be most applicable for specific problems (Hackenbruch et al. 2017; Briley et al.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

2015; Bornemann et al. 2019). But this task is often assumed to be solely the decision-makers' responsibility (Briley et al. 2015), and is not considered a research problem *per se*. However, resource managers may not know, *a priori*, the types of climatic metrics that could be most useful, and scientists may not always know whether they can provide information on decision-relevant metrics with reasonable skill (Briley et al. 2015; Porter and Dessai 2017; Lemos et al. 2012). This means that directly asking decision-makers to explain the types of climate information they need is rarely sufficient. Therefore, few studies have systematically identified decision-relevant metrics for sectoral adaptations (Hackenbruch et al. 2017; Vano et al. 2019; Bornemann et al. 2019). 'Co-production', or iterative and continual engagement between scientists and decision-makers, is often suggested as a means to enable mutual learning and reconciliation between managers' needs and scientific priorities (Lemos 2015; Kirchhoff et al. 2013a; Weaver et al. 2014; Vogel et al. 2016; Kolstad et al. 2019). It can thus help to identify decision-relevant climatic metrics that are also tractable for modellers.

That being said, not all co-production efforts have led to positive outcomes (Lemos et al. 2018), or have been successful at understanding and responding to resource managers' needs (Lemos et al. 2018; Porter and Dessai 2017). The success of co-production is predicated on the level and quality of interactions between (and within) different groups (Porter and Dessai 2017; Wall et al. 2017; Kirchhoff et al. 2013b; Mach et al. 2019; Lemos et al. 2018; Meinke et al. 2006). While the literature provides rich guidance on the general principles and prerequisites for successful co-production (Hegger et al. 2012; Meadow et al. 2015; Lemos and Morehouse 2005; Beier et al. 2017), there is a dearth of empirically-grounded guidance on co-production processes that have worked in practice (Djenontin 2018; Lemos et al. 2018; Parker and Lusk 2019).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Hence, the process of co-production is often a black box; there is no clarity on the types of scientist-decision-maker engagement processes that can be expected to result in effective two-way communications and to enable the creation of usable climate science (Porter and Dessai 2017; Mach et al. 2019; Jagannathan et al. 2019a).

In this paper we present both the process of, and outcomes from, a case of co-production, Project Hyperion, that (eventually) led to the identification of decision-relevant climatic metrics for water management decisions. As a response to calls to detail the practice of 'how' co-production works (Porter and Dessai 2017; Lemos et al. 2018; Mach et al. 2019), we focus this paper on not just the knowledge outcomes from the effort (i.e. the decision-relevant metrics), but also on how the metrics evolved iteratively through multiple engagements over the course of a year. The rest of the paper details the boundary spanning and engagement strategies that enabled the project to overcome institutional and epistemological barriers, and allowed a shared understanding across professional communities to emerge.

Project Hyperion and the process of co-production

Project Hyperion is a basic science project that aims to advance climate modelling by evaluating regional climate datasets for decision-relevant metrics. While there has been an explosive growth in the number of regional climate datasets available to users, there is limited understanding of the credibility and suitability of these datasets for use in different management decisions (Moss et al. 2019; Barsugli et al. 2013; Jones et al. 2016; Jagannathan et al. 2019b; Vandermolen et al. 2019). Hyperion aims to address this need by developing comprehensive assessment capabilities to evaluate the credibility of

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

regional climate datasets, understand the processes that contribute to model biases, and improve the ability of models to predict management relevant outcomes.

Since decision-relevance is a core motivation for the project, Hyperion is designed on the principles of co-production. The project brings together scientists from nine research institutions with managers from twelve water agencies in four watersheds: Sacramento/San Joaquin, Upper Colorado, South Florida, and Susquehanna. In addition, the project structure explicitly allows for both the groups to co-develop the science plan and research questions, in addition to co-producing the science itself. The scientists include atmospheric and earth system scientists as well as hydrologists. The water managers, depending on the agency, have functions including planning, operating and managing water quality, water supply, stormwater management, flood control, and water infrastructure design. These water managers have high levels of technical expertise in engineering, hydrology or other sciences, and were purposefully selected because of their interest in the project concept and their willingness to dedicate time to the engagement efforts. In addition, the project team for Hyperion includes three dedicated 'boundary spanners' (including two of the authors), i.e., people whose primary role is to facilitate and mediate the scientist-water manager boundary.

In this paper we focus on Phase 1 of the project, and describe how decision-relevant metrics in each of the study regions were co-produced by this group. From the water managers' perspective, such metrics quantitatively describe climatic phenomena that are directly related to practical management problems; changes in these quantities would necessitate shifts in water infrastructure planning and operations. From the scientists' perspective, these metrics can be used to test model fidelity for decision-relevant phenomena and hence push model development and scientific inquiry in more

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

use-inspired directions. To identify these metrics, a series of iterative engagement methods were used. Structured engagement methods included workshops, remote and in-person focus-group discussions, and quarterly project update calls. There were also continual less-structured, informal conversations between scientists, managers, and boundary spanners over phone calls or emails. Approval from Lawrence Berkeley Lab's Human Subjects Committee - Institutional Review Board was obtained for key engagements. The timeline of engagement activities, along with goals and milestones at each stage, is presented in Fig. 1.

The role of boundary spanners

The boundary spanners in Project Hyperion had varying degrees of social science, climate science and adaptation expertise; they also had prior experience in co-production and similar participatory research activities. It is generally acknowledged that boundary spanners are necessary for the translation of jargon and assumptions among different actors and across epistemic divides (Bednarek et al. 2016b; Kirchhoff et al. 2013b; Cash et al. 2003). At the same time, the literature recognizes that this role is challenging in practice (Bednarek et al. 2018; Safford et al. 2017) and that the functions and attributes of effective boundary spanning are not well understood (Goodrich et al. 2019; Bednarek et al. 2016a).

The challenges of boundary spanning are often discussed in instances where actors are resistant to crossing epistemic boundaries or "compromising" their expertise (Cash et al. 2003). In Hyperion, most of the water managers wanted to incorporate climate change information in their decisions, and most scientists were committed to developing decision-relevant science. This collective goodwill notwithstanding, several rounds of deliberations were needed to mediate differences in incentives and priorities,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

and to translate the water managers' needs into quantitative metrics and scientific research questions. The boundary spanners needed to actively ensure that feedback from both groups was not just heard and documented, but also incorporated into the overall science plan for the project.

The mediation of the scientist-manager boundary to arrive at actionable rainfall metrics illustrates these tensions and also their eventual resolution. Several of the managers wanted information on Intensity Duration Frequency (IDF) curves for rainfall events (Srivastava et al. 2019) that formed the basis of their flood-related decisions. The scientists, based on their expertise and modelling capabilities, prioritized metrics such as frequency and intensity of specific storm events (e.g. tropical cyclones) and associated rainfall. While these storm metrics were related to decision-relevant rainfall quantities, they were often one step 'upstream' (in both the hydrological and metaphorical senses) of what the water managers wanted for detailed planning. The upstream metrics represented drivers of phenomena of interest rather than the decision-relevant phenomena themselves. Recognizing this tension, the boundary spanners worked with the group to co-create a shared understanding of the term 'metric'. We introduced a hierarchical framework that distinguished decision-relevant from upstream metrics, illustrating the overlaps and linkages between the two, and showing how both types of metrics could fit within the project's larger goals. With the explicit linking of metric-types, managers could better appreciate the scientists' focus on upstream storm metrics for modelling causal processes that could eventually make IDF predictions more accurate. Scientists saw why it was necessary to include the metric of interest to managers, i.e. IDF curves, in the science plan, and how linking their storm metrics with IDF results added to the novelty and impact of their efforts.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

This and similar resolutions were highly dependent on the presence of a boundary spanner with domain expertise in climate modelling. While the literature recognises the importance of 'background and experience' in the subject matter (Safford et al. 2017; Meadow et al. 2015; Bednarek et al. 2016b), there is, we would argue, less appreciation of the technical expertise required to execute techno-scientific translations (Bednarek et al. 2018). For our project, having a boundary spanner who was also a modeller proved essential. Given the aims of Hyperion, many boundary functions towards the later stages of the project needed in-depth (and often painful) discussions on model parameters, types of simulations, decision-relevant thresholds, statistical measures of model performance, etc. which were beyond the technical capacities of the non-modeller boundary spanners (Fig. 2). In hindsight, we believe that a boundary spanner with expertise in water management could have been equally beneficial, and may have augmented our eventual list of metrics. Overall, we found that, depending on the nature of what is being co-produced, boundary spanners need considerably higher levels of domain expertise than is generally acknowledged in the literature.

Direct and indirect approaches to 'making' metrics

A common approach to user needs assessments in conventionally-designed as well as co-production projects is to directly ask decision-makers for the types of information they want (Hudlicka 1996; Briley et al. 2015). This approach is based on the prevalent assumption that decision-makers not only know the climatic metrics they want, but are also able to articulate their knowledge in response to direct questions (Hudlicka 1996). Neither of these assumptions is true for every engagement. We found that determining the quantitative details of decision-relevant information required both direct and indirect approaches. We did explicitly ask managers to identify any metrics for which

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

they required projections, and this direct approach was partially successful. But it put the onus of metric identification on the water managers, who did not always know what to ask for or what the scientists had to offer by way of quantification. For example, the direct approach revealed water supply and floods as key climate-related management issues in California, with snowpack, snowmelt, streamflow, dry spells and rainfall as hydroclimatic phenomena of interest. But managers were not used to translating these phenomena into tractable parameters or thresholds (Briley et al. 2015; Hackenbruch et al. 2017).

We therefore supplemented the direct approach with an indirect approach that assumed that relevant knowledge cannot be revealed by direct questions, but needs to be extracted through more open-ended scenario analysis and contextual inquiry. Although such discussions are a time-intensive way to access internal knowledge structures (Hudlicka 1996), combining direct and indirect conversational methods have been shown to be an effective way of eliciting user needs (Zhang 2007). This indirect approach is used in software development for user requirements engineering (Hudlicka 1996; Zhang 2007), but is not commonly used in the co-production or actionable environmental science literatures. Partly guided by research on tacitly-held knowledge, and partly through trial and error, we developed four indirect strategies that enabled scientists and water managers to collaboratively identify decision-relevant metrics.

1. *Developing hierarchical frameworks:* There was often confusion among scientists and managers on how specific a 'metric' needs to be to have an unambiguous interpretation from a modelling perspective. For example, in the initial engagements, the whole group understood 'peak streamflow' or 'flooding' to be potential metrics. However, when modelling methods were being developed, the scientists had

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

questions as to what 'peak' might mean or how 'flooding' was defined by the managers. Further direct questions that probed the managers for "more specific" metrics were unsuccessful in eliciting the details that scientists were looking for. At the same time, scientists were not able to clearly articulate what constituted an unambiguous metric. To resolve this stalemate, the boundary spanners asked the scientists to provide examples of what might constitute a specific metric for their modelling exercises. The group then decided to contextualize metrics by developing a hierarchical framework: a management issue came first, then the hydroclimatic phenomena related to the issue, then the aspects of each phenomenon that were of most relevance to the water managers, and finally a tractable metric for each aspect (Fig. 3) (see also(Maraun et al. 2015)). For Hyperion, the hierarchy represented a logical framework that helped us to understand that peak streamflow could have varied interpretations for modelling; it could be daily maximum flow, or the high end of streamflow distribution, or values above certain thresholds. Each interpretation represented a very different 'metric' with unique results. Through the framework we collectively understood that peak streamflow was best characterized as an 'aspect' of a hydroclimatic phenomenon, and one step ahead of being an unambiguous metric, which required further quantitative details describing the characteristics of the peak that were important to managers.

2. *Starting from the planning challenge/goal rather than the science question:* A focus on current and future planning challenges or goals as they related to different hydroclimatic phenomena was a productive path towards metric identification. For example, when asked about planning goals with respect to streamflow quantity, some managers suggested that the aim was to have a full reservoir on July 1st. Through this

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

exchange we identified cumulative run-off on July 1 as a decision-relevant metric. Another discussion centred on recent climate- or weather-related planning challenges (such as Hurricane Irma, or the Oroville dam failure) in the managers' regions. One of the managers discussed an ice-jam related flooding event and described how warm temperatures and heavy rain conditions in early spring caused the snow to melt rapidly, leading to flooding. This prompted a collective discussion about whether frequency of rain-on-snow events and the associated run-off could be an actionable metric to help anticipate and manage such events. These results support recommendations from other studies that also suggest starting the co-production process from the management goal rather than from a scientific "puzzle" (Beier et al. 2017; Kolstad et al. 2019).

3. *Collaboratively exploring the planning relevance of new models, tools, or datasets:* It is often assumed that practitioners are mainly interested in pragmatic solutions and may be less open to exploring novel models and tools (Vogel et al. 2016). However, in Hyperion, collaboratively and critically examining whether and how new models, datasets or tools could be relevant to managers' contexts, proved to be a productive strategy for identifying metrics. For example, one of the scientists sought the water managers' opinion on a new type of satellite data on terrestrial water storage (TWS) that had the potential to aid in flood/drought prediction. Managers responded that their agencies mainly used 10-year ground water (GW) baseflow as a key metric for drought predictions, but that it was not easy to collect data for computing GW baseflow. They were interested in alternatives to this metric, whereupon the scientist explained that new findings suggested that TWS can be a good predictor of GW flow (in some regions). The group collectively agreed that both TWS and 10-year GW

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

baseflow would be good metrics, and that TWS would be explored as a potential proxy or upstream metric to GW baseflow.

4. *Using analogies for 'good' metrics:* Finally, some of the new metrics identified in our project came from discussions of other 'good' metrics. For example, one well-received set of metrics was visualized through the 'Snow Water Equivalent (SWE) triangle', which uses a fitted triangle to characterize the annual cycle of snow accumulation and melt (Rhoades et al. 2018). The SWE triangle represents a composite of six metrics of management relevance: peak water volume and timing, snow accumulation and melt rates, and the lengths of the accumulation and melt seasons. Each metric is tractable as well as decision-relevant, and the triangle itself presents a visually digestible linear approximation of all six metrics comprising the snow cycle (Rhoades et al. 2018). The water managers thought this was a "nifty" multi-metric representation as it allowed for both a comprehensive and an individual examination of the management-relevant components of seasonal snow dynamics. Their response led to discussions on whether a similar set of metrics describing the annual cycle of rainfall would also be useful. A new composite approach, tentatively termed 'rainfall geometry' (to signify whatever geometric figure fits the annual cycle of rainfall in a given location), and which includes the start date of the wet season, peak rainfall, and length of the wet season, was co-developed as a promising multi-metric representation of key management-relevant components of rainfall.

Overall, we found that the making of decision-relevant metrics needed an iteratively-derived mix of direct and indirect engagement approaches to capture the information needs of the water managers, and to translate them into tractable quantitative metrics for

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

the scientists. Fig. 4 shows the evolution of two decision-relevant metrics using different direct and indirect strategies.

Decision-relevant metrics and their characteristics

Table 1 presents examples of the metrics identified in the project (Supplement Table 1 has the full list for all four regions). In some cases, these metrics already existed in other contexts (such as in engineering or hydrology manuals), but had not been recognized as metrics relevant for climate modelling prior to our co-production process. We also observed that not every identified metric mapped onto a specific management decision. Some metrics, such as deviations from historical mean snowpack, were more useful for understanding the future state of watersheds than for making decisions. The interest in snowpack shows that there are overlaps between upstream and decision-relevant metrics; several water managers were, in fact, interested in understanding upstream processes in addition to working with actionable metrics (Vano et al. 2019).

Finally, we found that the relevance of metrics depends on, and evolves with, the availability of climate information. In regions with limited availability of climate data even simple climatic metrics such as monthly or annual run-off were considered relevant enough. In regions with more information such simple metrics were not as useful; managers identified more detailed metrics, such as the runoff associated with highest snow melt rate, or maximum daily or 3-day flow volumes, as actionable. An analysis of how and why the characteristics of decision-relevant metrics differed among the water management agencies is planned for the next phase of the project.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Discussion and Conclusions

In this paper, we open up the black box of co-production and document in detail the strategies that enabled (and did not enable) the creation of decision-relevant science. We illustrate how co-production works in practice by analyzing the numerous back-and-forth collaborative engagements of Project Hyperion, and describing how the science changed and evolved during the process. By describing how climate scientists and water managers (eventually) crossed the boundaries of both mandate and epistemology to co-produce decision-relevant metrics, we add to the sparse literature on 'how and when' co-production works. To our knowledge, this is the first study to document in detail the actionable climatic metrics for adaptive water management, and the co-production processes needed to arrive at such metrics. Our outcomes (i.e. the co-produced decision-relevant metrics), can be used as inputs for developing actionable climate science for adaptation in the water sector. Our learnings on engagement approaches provide co-production scholars with insights on how to design and implement productive scientist-decision-maker interactions.

We found that identifying problem-specific climatic metrics is even more iterative, and needs more social and technical negotiations, than is generally implied in the literature promoting co-production. These metrics often represent new scientific directions for the scientists as well as new ways of management for the water managers. The commonly used direct approach to identifying decision-makers' information needs was insufficient for getting at the quantitative details of climatic metrics, even when the decision-makers had high levels of scientific knowledge. We found that the task of translating user needs into quantitative metrics needs the expertise of both resource managers and climate scientists, as well as an enabling process for both groups'

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

knowledge(s) to evolve. Hence, a judicious mix of direct and indirect approaches was needed to "make" these metrics. The indirect methods, in particular, revealed the groups' tacitly-held knowledge and allowed a comprehensive set of shared learnings to emerge. Key indirect strategies included developing a hierarchical framework linking management issues with actionable metrics and upstream phenomena; starting discussions from the planning challenges and then moving to the model-specific metrics; collaboratively exploring the planning relevance of new models, datasets and scientific findings that managers did not yet know about; and using analogies of good metrics from other hydroclimatic phenomena. Eventually, the twin functions of the metrics -- of being decision-relevant and extending model capability -- spoke to both the decision-makers' and the scientists' priorities, and allowed both groups to co-exist within the project. Additionally, the institutionalization of the boundary spanning role, and the domain expertise of at least one boundary spanner (an under-appreciated phenomenon in the co-production literature), proved to be crucial for effective trans-boundary translation.

Although the co-production was time-consuming, the richness of our understanding came from analyzing the many iterative back-and-forth engagements, where even the processes that did not fully work were essential to get to the processes that did eventually work. Co-production is often presented as an outcome in itself, rather than as a means to an end (Lemos et al. 2018). This perspective may have its merits, but we argue that the ability to achieve desired outcomes is quite sensitive to how the co-production process is structured and implemented. More critical assessments of specific co-production processes would help to move the practice forward more efficiently, and to meet the growing need for actionable climate science across many sectors of society.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Acknowledgements

The authors are deeply grateful to Project Hyperion's water managers and scientists who patiently participated in the many back-and-forth engagements that form the basis of this paper. We are also thankful to Bruce Riordan who co-led the engagements and Paul Ullrich for his agile leadership of the project. The authors would also like to thank the Water+ Group at UC Berkeley, James Arnott, Margaret Torn and Alastair Iles for their detailed feedback on the manuscript. This work was supported by the Office of Science, Office of Biological and Environmental Research, Climate and Environmental Science Division, of the U.S. Department of Energy under contract DE-AC02-05CH11231 as part of the Hyperion Project, An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US (award DE-SC0016605).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

References

ACCNRS, 2015: *Report to the Secretry of the Interior - Advisory Committee on Climate Change and Natural Resource Science*. 86 pp.

Barsugli, J. J., and Coauthors, 2013: The practitioner's dilemma: How to assess the credibility of downscaled climate projections. *Eos, Trans. Am. Geophys. Union*, **94**, 424–425, https://doi.org/10.1002/2013EO460005.

Bednarek, A., C. Wyborn, R. Meyer, A. Parris, P. Leith, B. Mcgreavy, and M. Ryan, 2016a: *Practice at the Boundaries policy and society for environmental outcomes*. 1–27 pp.

Bednarek, A. T., B. Shouse, C. G. Hudson, and R. Goldburg, 2016b: Science-policy intermediaries from a practitioner's perspective: The Lenfest Ocean Program experience. *Sci. Public Policy*, **43**, 291–300, https://doi.org/10.1093/scipol/scv008.

Bednarek, A. T., and Coauthors, 2018: Boundary spanning at the science–policy interface: the practitioners' perspectives. *Sustain. Sci.*, **13**, 1175–1183, https://doi.org/10.1007/s11625-018-0550-9.

Beier, P., L. J. Hansen, L. Helbrecht, and D. Behar, 2017: A How-to Guide for Coproduction of Actionable Science. *Conserv. Lett.*, **10**, 288–296, https://doi.org/10.1111/conl.12300.

Bornemann, F. J., and Coauthors, 2019: Future changes and uncertainty in decision-relevant measures of East African climate.

Bremer, S., and S. Meisch, 2017: Co-production in climate change research: reviewing different perspectives. *Wiley Interdiscip. Rev. Clim. Chang.*, **8**, 1–22, https://doi.org/10.1002/wcc.482.

Briley, L., D. Brown, and S. E. Kalafatis, 2015: Overcoming barriers during the co-production of climate information for decision-making. *Clim. Risk Manag.*, **9**, 41–49,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

https://doi.org/10.1016/j.crm.2015.04.004.

Cash, D. W., W. C. Clark, F. Alcock, N. M. Dickson, N. Eckley, D. H. Guston, J. Jäger, and R. B. Mitchell, 2003: Knowledge systems for sustainable development. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 8086–8091, https://doi.org/10.1073/pnas.1231332100.

Djenontin, I. N. S., 2018: The art of co-production of knowledge in environmental sciences and management : lessons from international practice. *Environ. Manage.*, 885–903, https://doi.org/10.1007/s00267-018-1028-3.

Goodrich, K. A., K. D. Sjostrom, A. Bednarek, C. Vaughan, L. Nichols, and M. C. Lemos, 2019: Who are boundary spanners? Attributes, opportunities and constraints for engaging knowledge across boundaries. *Curr. Opin. Environ. Sustain.*, Under Review.

Hackenbruch, J., T. Kunz-Plapp, S. Müller, and J. Schipper, 2017: Tailoring Climate Parameters to Information Needs for Local Adaptation to Climate Change. *Climate*, **5**, 25, https://doi.org/10.3390/cli5020025.

Hegger, D., M. Lamers, A. Van Zeijl-Rozema, and C. Dieperink, 2012: Conceptualising joint knowledge production in regional climate change adaptation projects: Success conditions and levers for action. *Environ. Sci. Policy*, **18**, 52–65, https://doi.org/10.1016/j.envsci.2012.01.002.

Hudlicka, E., 1996: Requirements elicitation with indirect knowledge elicitation techniques: comparison of three methods. *Proceedings of the Second International Conference on Requirements Engineering*, 4–11.

Jagannathan, K., J. C. Arnott, C. Wyborn, N. Klenk, K. J. Mach, R. H. Moss, and K. D. Sjostrom, 2019a: Great expectations? Reconciling the aspiration, outcome, and possibility of coproduction. *Curr. Opin. Environ. Sustain.*, Under review.

——, A. D. Jones, and A. C. Kerr, 2019b: Selecting Climate Projections for Decision - relevant Metrics : A Case Study of Chill Hours in California. *Clim. Serv.*, Under

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Review.

Jones, A., K. Calvin, and J.-F. Lamarque, 2016: Climate Modeling with Decision Makers in Mind. *Eos (Washington. DC).*, **97**, 2–5, https://doi.org/10.1029/2016EO051111.

Kerr, R. A., 2011: Time to Adapt to a Warming World, But Where's the Science? *Science (80-.).*, **334**, 1052–1053, https://doi.org/10.1126/science.334.6059.1052.

Kirchhoff, C. J., M. Carmen Lemos, and S. Dessai, 2013a: Actionable Knowledge for Environmental Decision Making: Broadening the Usability of Climate Science. *Annu. Rev. Environ. Resour.*, **38**, 393–414, https://doi.org/10.1146/annurev-environ-022112-112828.

——, M. C. Lemos, and N. L. Engle, 2013b: What influences climate information use in water management? The role of boundary organizations and governance regimes in Brazil and the U.S. *Environ. Sci. Policy*, **26**, 6–18, https://doi.org/10.1016/j.envsci.2012.07.001.

Kolstad, E. W., and Coauthors, 2019: Trials, Errors, and Improvements in Coproduction of Climate Services. *Bull. Am. Meteorol. Soc.*, **100**, 1419–1428, https://doi.org/10.1175/BAMS-D-18-0201.1.

Lemos, M. C., 2015: Usable climate knowledge for adaptive and co-managed water governance. *Curr. Opin. Environ. Sustain.*, **12**, 48–52, https://doi.org/10.1016/j.cosust.2014.09.005.

——, and B. J. Morehouse, 2005: The co-production of science and policy in integrated climate assessments. *Glob. Environ. Chang.*, **15**, 57–68, https://doi.org/10.1016/J.GLOENVCHA.2004.09.004.

——, C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nat. Clim. Chang.*, **2**, 789–794, https://doi.org/10.1038/nclimate1614.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

——, and Coauthors, 2018: To co-produce or not to co-produce. *Nat. Sustain.*, **1**, 722–724, https://doi.org/10.1038/s41893-018-0191-0.

Mach, K. J., and Coauthors, 2019: Actionable knowledge and the art of engagement. *Curr. Opin. Environ. Sustain.*, Under Review.

Maraun, D., and Coauthors, 2015: VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Futur.*, **3**, 1–14.

McNie, E. C., 2007: Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environ. Sci. Policy*, **10**, 17–38, https://doi.org/10.1016/j.envsci.2006.10.004.

Meadow, A. M., D. B. Ferguson, Z. Guido, A. Horangic, G. Owen, and T. Wall, 2015: Moving toward the Deliberate Coproduction of Climate Science Knowledge. *Weather. Clim. Soc.*, **7**, 179–191, https://doi.org/10.1175/WCAS-D-14-00050.1.

Meinke, H., R. Nelson, P. Kokic, R. Stone, R. Selvaraju, and W. Baethgen, 2006: Actionable climate knowledge: From analysis to synthesis. *Clim. Res.*, **33**, 101–110, https://doi.org/10.3354/cr033101.

Moss, R. H., and Coauthors, 2013: Hell and High Water: Practice-Relevant Adaptation Science. *Science (80-.).*, **342**, 696–698, https://doi.org/10.1126/science.1239569.

Moss, R. H., and Coauthors, 2019: Evaluating Knowledge to Support Climate Action: A Framework for Sustained Assessment. Report of an Independent Advisory Committee on Applied Climate Assessment. *Weather. Clim. Soc.*, **11**, 465–487, https://doi.org/10.1175/WCAS-D-18-0134.1.

Parker, W. S., and G. Lusk, 2019: Incorporating user values into climate services. *Bull. Am. Meteorol. Soc.*, 1–23, https://doi.org/10.1175/bams-d-17-0325.1.

Porter, J. J., and S. Dessai, 2017: Mini-me: Why do climate scientists' misunderstand users and their needs? *Environ. Sci. Policy*, **77**, 9–14,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

https://doi.org/10.1016/j.envsci.2017.07.004.

Rhoades, A. M., A. D. Jones, and P. A. Ullrich, 2018: Assessing Mountains as Natural Reservoirs With a Multimetric Framework. *Earth's Futur.*, **6**, 1221–1241, https://doi.org/10.1002/2017EF000789.

Roncoli, C., and Coauthors, 2009: From accessing to assessing forecasts: An end-to-end study of participatory climate forecast dissemination in Burkina Faso (West Africa). *Clim. Change*, **92**, 433–460, https://doi.org/10.1007/s10584-008-9445-6.

Safford, H. D., S. C. Sawyer, S. D. Kocher, J. K. Hiers, and M. Cross, 2017: Linking knowledge to action: the role of boundary spanners in translating ecology. *Front. Ecol. Environ.*, **15**, 560–568, https://doi.org/10.1002/fee.1731.

Srivastava, A., R. Grotjahn, P. A. Ullrich, and M. Risser, 2019: A unified approach to evaluating precipitation frequency estimates with uncertainty quantification : Application to Florida and California watersheds. *J. Hydrol.*, **In Press**.

USGAO, 2015: *Climate information: A national system could help federal, state, local, and private sector decision makers use climate information*. 49 pp.

Vandermolen, K., T. U. Wall, and B. Daudert, 2019: A call for the evaluation of web-based climate data and analysis tools. *Bull. Am. Meteorol. Soc.*, **100**, 257–268, https://doi.org/10.1175/BAMS-D-18-0006.1.

Vano, J. A., K. Miller, M. D. Dettinger, R. Cifelli, D. Curtis, A. Dufour, J. R. Olsen, and A. M. Wilson, 2019: Hydroclimatic Extremes as Challenges for the Water Management Community: Lessons from Oroville Dam and Hurricane Harvey. *Bull. Am. Meteorol. Soc.*, **100**, S9--S14.

Vogel, J., E. Mcnie, and D. Behar, 2016: Co-producing actionable science for water utilities. *Clim. Serv.*, **2–3**, 30–40, https://doi.org/10.1016/j.cliser.2016.06.003.

Wall, T. U., E. McNie, and G. M. Garfin, 2017: Use-inspired science: making science usable

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

by and useful to decision makers. *Front. Ecol. Environ.*, **15**, 551–559, https://doi.org/10.1002/fee.1735.

Weaver, C. P., and Coauthors, 2014: From global change science to action with social sciences. *Nat. Clim. Chang.*, **4**, 656–659, https://doi.org/10.1038/nclimate2319.

Zhang, Z., 2007: Effective requirements development-A comparison of requirements elicitation techniques. *Softw. Qual. Manag. XV Softw. Qual. Knowl. Soc. E. Berki, J. Nummenmaa, I. Sunley, M. Ross G. Staples Br. Comput. Soc.*, 225–240.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Figure Captions List

FIG. 1. Co-production process and timeline.

FIG. 2. Dialogue between Water manager (W) and Boundary spanner (B).

FIG. 3. Hierarchical framework with examples.

FIG. 4. Examples showing the evolution of decision-relevant metrics.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Figures and Tables

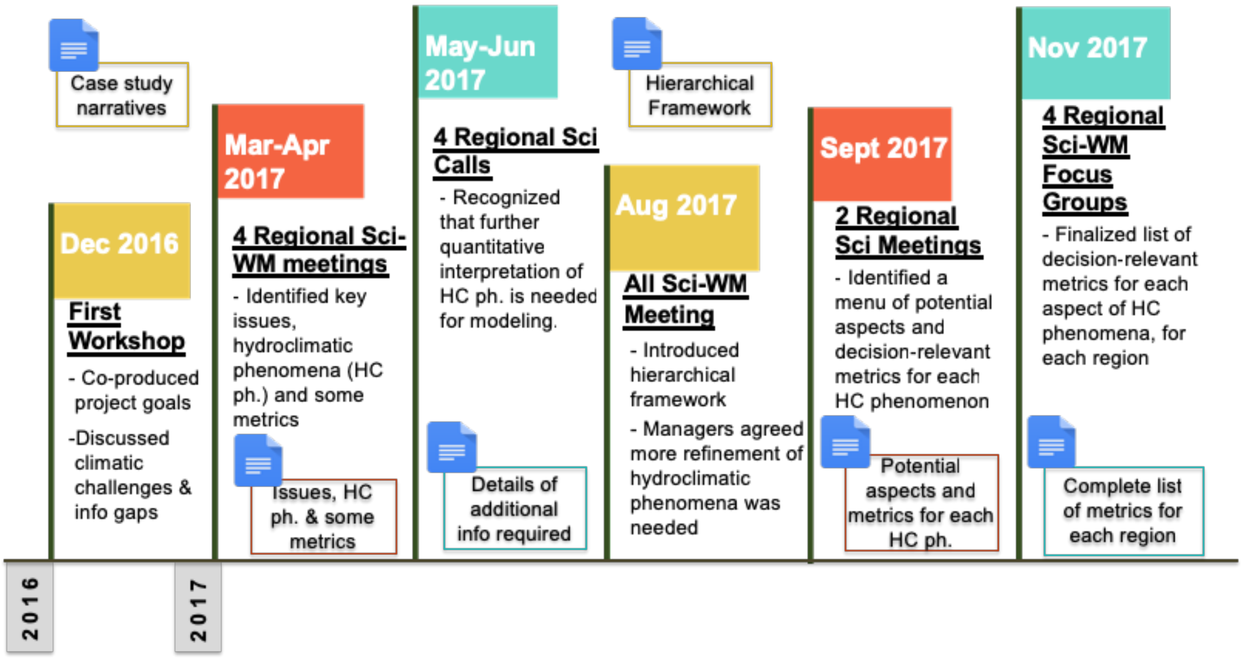

FIG. 1. Co-production process and timeline summarising key engagement activities over the course of a year, along with the most important outcomes at each stage (depicted by the blue document icon). 'Sci' refers to Scientists, 'WM' refers to Water Manager and 'HC ph.' refers to Hydroclimatic Phenomena. For details of each of these activities please see the Supplement. There was constant boundary spanning work during and between each of these activities.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

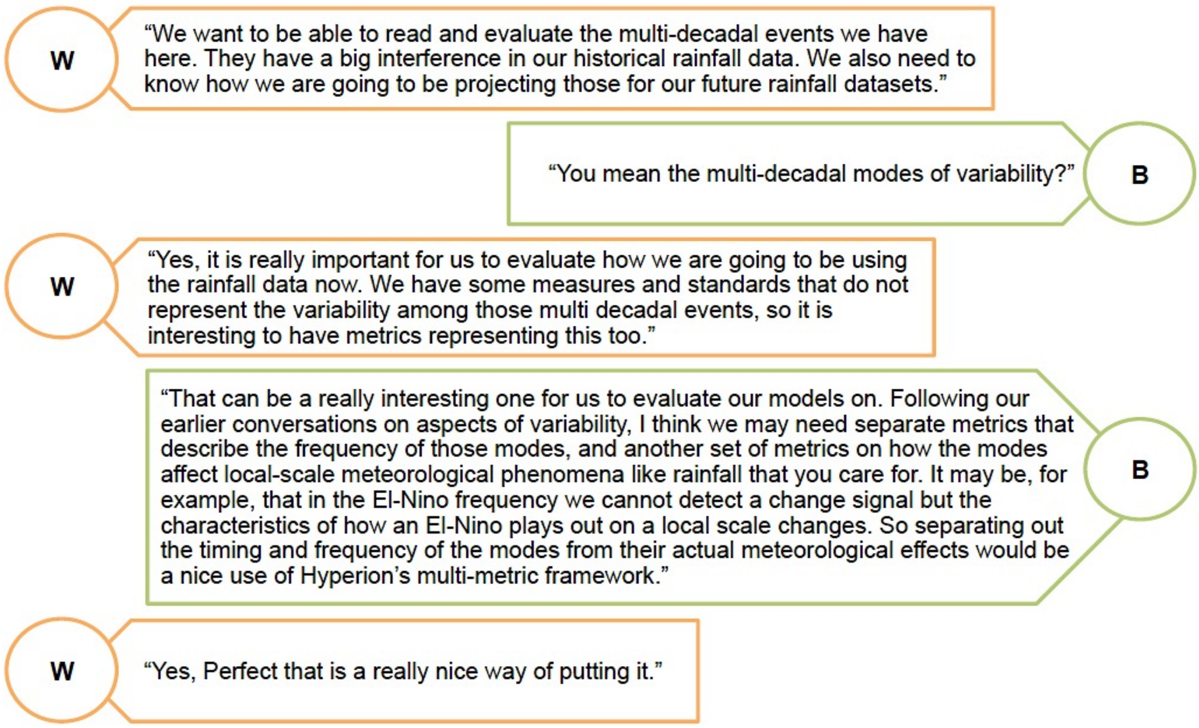

FIG. 2. Dialogue between Water manager (W) and Boundary spanner (B) showing the benefits of having a modeller as a “translator” of the water manager’s description of information needs into quantitative metrics that can be pursued by modellers.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

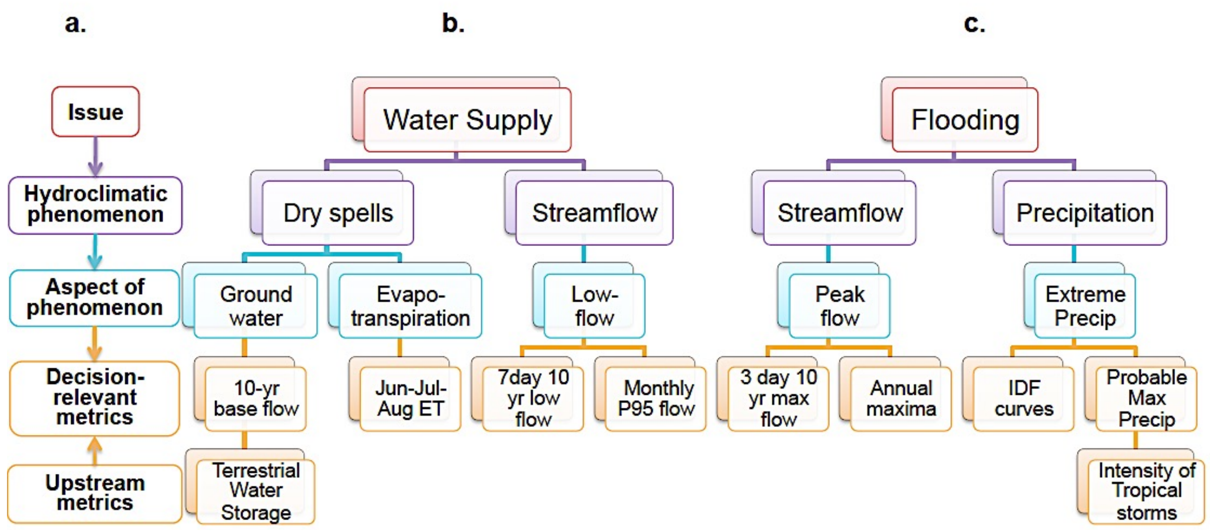

FIG. 3. Hierarchical framework with examples. (a) illustrates the hierarchical framework starting from a management issue and ending in the metrics. (b) and (c) are examples of metrics and how they fit within the framework. The hierarchy starts with the 'Issue' or topic of management relevance in the region (e.g. Flooding), and moves to the 'Hydroclimatic Phenomenon' related to the issue (e.g. Precipitation is a hydroclimatic phenomenon related to the issue of Flooding), and then to the 'Aspect of the Phenomenon' that is of specific interest for the management decision (e.g. Extreme precipitation is the aspect of precipitation that is of specific management interest). Finally, the hierarchy yields the actual 'Decision-relevant metric', which refers to a quantity that has potential use for the water managers and has an unambiguous formula or algorithm that can be applied to both observation-based data and model outputs (e.g. Probable Maximum Precipitation (PMP) is a metric related to extreme precipitation). We also identified upstream metrics that describe phenomena hypothesized to be important drivers of the decision-relevant phenomena (e.g. Intensity of tropical storms of certain durations or return periods are an upstream driver of PMP).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

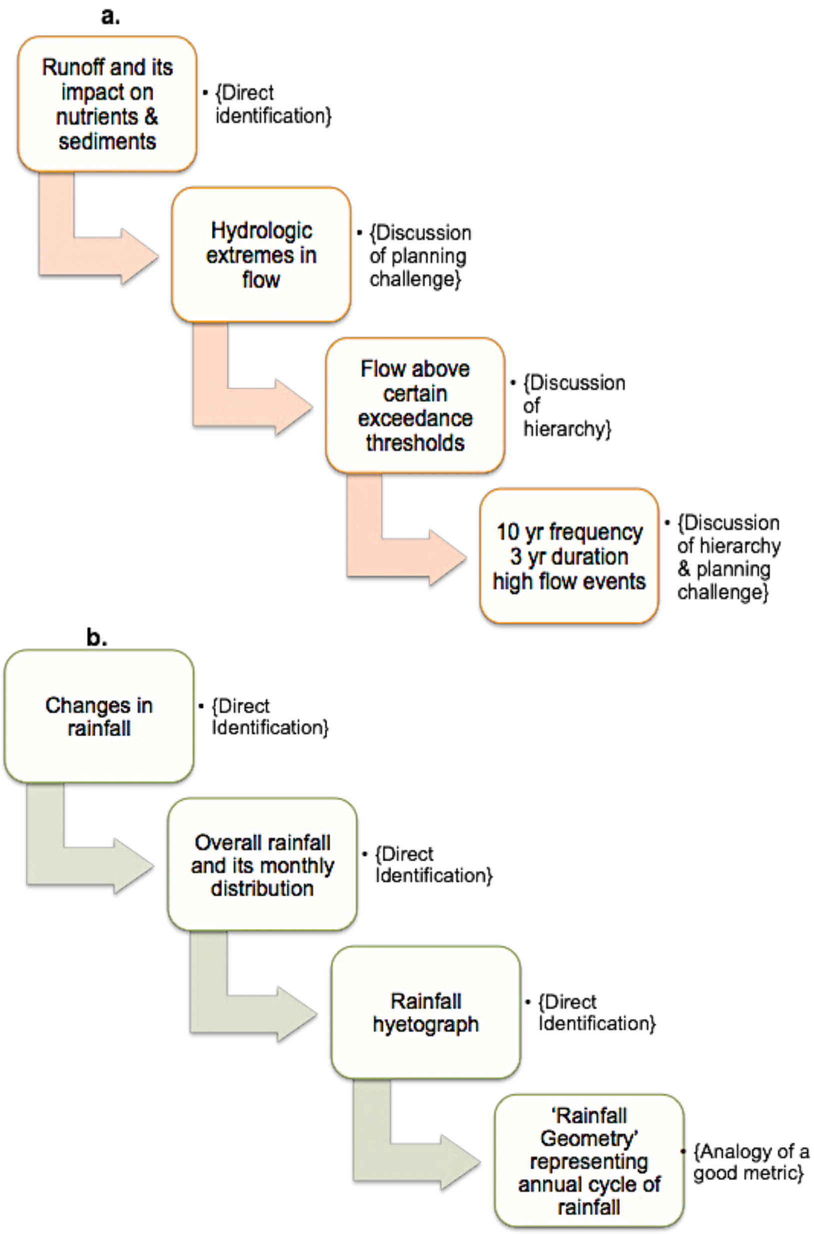

FIG. 4. Examples showing the evolution of decision-relevant metrics. (a) shows the evolution of the metric that represents the 3-year critical duration of October-March high flows at a 10-year recurrence interval. The initial direct identification approach gave a broad understanding of the importance of runoff for nutrients and sediments, and then a discussion of runoff-based planning led to identifying hydrologic extremes as one of the important components of runoff. Using the hierarchy (Fig. 3), we came to understand that 'extremes' were an 'aspect of phenomenon', and we probed further to find that extremes actually meant flows above certain thresholds. We derived the final unambiguous metric at the next iteration where we

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

interrogated the types of exceedance thresholds that impact water quality management in the region. (b) shows the making of a rainfall metric. First, the direct approach highlighted that changes in rainfall patterns were an important challenge for the region. In the next 2 iterations, which also used direct engagements, we identified the specific aspects of rainfall that were of importance. Finally, with the analogy of the 'good metrics' of the SWE triangle, we identified "Rainfall Geometry" as a promising concept for additional decision-relevant metrics.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

Table. 1. Examples of decision-relevant metrics for each region, highlighting management issues, hydroclimatic phenomena, aspect of phenomena and then each decision-relevant metric. 'CA' refers to the Sacramento/San Joaquin watershed, 'CO' is Upper Colorado, 'FL' is South Florida, and 'SQ' is Susquehanna. The last column also describes some of the potential decisions or uses for these metrics that were identified by the case study water managers. Supplement Table 1 has the full list for all four regions.

Region	**Issue**	**Hydroclimatic Phenomenon**	**Aspect of Phenomenon**	**Decision-relevant Metric**	**Decision/Use**
CA	Water Supply	Snowpack	Annual cycle of snow accumulation and melt	Snow Water Equivalent (SWE) triangle (Rhoades et al. 2018) - Peak snow (amount and timing), and its relationship with average snow-accumulation and -melt rates, and timing and length of accumulation and melt seasons	On-stream reservoir management, and understanding future streamflow characteristics. Shape of the triangle shows the changing dynamics of the snow season, and what to expect in terms of runoff timing and amounts.
CA	Flooding	Streamflow	Peakflow {Pulse events}	Frequency of Rain-on-snow events and magnitude of associated run-off	Reservoir operations and flood management.
CA	Water Supply	Snowpack	Inter-annual Variability in Snowpack	Deviations from historical mean in SWE, Snowpack and Snowmelt (amount and timing)	Multi-year water supply planning and drought preparedness.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

CO	Water Supply	Streamflow	Seasonal Streamflow amount (in snowmelt season)	Cumulative run-off on July 1 and August 1	Annual water supply planning for the year done based on July 1 or August 1 reservoir level estimates (depending on the reservoir).
CO	Floods	Streamflow	Seasonal Streamflow amount (in snowmelt season)	% of average annual inflow for Apr-July	Reservoir management - this metric is an input into some reservoir operations models.
CO	Water Supply	Streamflow	Low-end Streamflow	7-day 10 year low flows	Water quality management (issuing discharge permits), and water supply planning during dry years (determining permit limits for water withdrawals).
FL	Flooding	Rainfall	Extreme Rainfall	Intensity Duration Frequency or IDF curves, specifically, 1-day, 3-day and up to 7-day rainfall events, for 10, 25, 50 and 100 year frequency intervals.	To calculate applicable discharge rates for different storm water management infrastructure. Design criteria used for drainage and flood protection are in terms of IDFs. In other words, designing of standard engineering practices for infrastructure.
FL	Flooding	Rainfall	Extreme Rainfall	Probable maximum precipitation. For 1-day, 3-day and maybe up to 7-day events	Large storage infrastructure design (like high dams).
FL	Water Supply	Rainfall	Variability in Rainfall	Rainfall anomalies at Monthly time scales	Water supply planning, and drought monitoring

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.

SQ	Water Supply	Streamflow	Peakflow	10- year frequency 3-year duration high flows for Oct-March	Water quality management in terms of monitoring Chesapeake Bay water quality standards.
SQ	Flooding	Streamflow	Average/ cumulative flows	Mean annual flow and harmonic mean flow	Water supply planning, for monitoring passby flows and conservation releases associated with water withdrawal permits. Water quality management for calculating design flows for effluent limitations based on water quality criteria.
SQ	Water Supply	Streamflow	Low-end Streamflow	7-day10-year low flow	Water quality management in terms of wastewater assimilation standards for discharge permits. Water supply planning in terms of passby flows or conservation releases for water withdrawal permits.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0296.1.