University of Nevada, Reno

A Socio-Hydrologic Assessment of Mountain Water Supply Vulnerability to Changing Snowmelt

A Dissertation Submitted in Partial Fulfillment of
the Requirements for the Degree of Doctor of Philosophy
in Hydrogeology

By

Beatrice L. Gordon

Dr. Adrian A. Harpold/Dissertation Advisor

December 2022

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

BEATRICE L. GORDON

entitled

A Socio-Hydrologic Assessment of Mountain Water Supply Vulnerability to Changing Snowmelt

be accepted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

Adrian A. Harpold, Ph.D.
Advisor

Newsha K. Ajami, Ph.D.
Committee Member

Manuel Alejandro Andrade-Rodriguez, Ph.D.
Committee Member

Rosemary W.H. Carroll, Ph.D.
Committee Member

Elizabeth A. Koebele, Ph.D.
Graduate School Representative

Markus Kemmelmeier, Ph.D., Dean
Graduate School

December, 2022

ABSTRACT

Climate change is accelerating disconnects between snowmelt-driven water supply and downstream demand. Identifying what makes people and places vulnerable to these disconnects can improve understanding of present conditions and help anticipate future changes in water management. This dissertation seeks to understand the potential for increasing disconnects between downstream agriculturally productive regions and their primary water supply—higher elevation, mountainous (upland) environments. We do so by focusing on agriculturally productive regions in the western United States (US) that are heavily reliant on seasonal snowmelt-driven streamflow, and using interdisciplinary tools such as big data, conceptual modeling, social science, and computational hydrology to assess vulnerability from the source (mountains) to demand (agriculture) We find that a process-based framework isolating three dominant mechanisms linking snow to streamflow helps explain changes in snowmelt-driven streamflow in 537 upland catchments throughout the US. We then use a hydrogeological framework and optimized averaging in a subset of our initial 537 catchments, highlighting the critical and often overlooked role of groundwater contributions in high, arid, and deep mountain catchments. Equipped with a more robust understanding of surface water and groundwater supplies in the western US, we then quantify the benefits of adaptation to changing snow resources particularly in hay-dominated agriculturally productive systems with smaller declines in snow relative to reservoir storage. Finally, we derive a flexible approach for expanding vulnerability assessments beyond the mountains and show that robust consideration of multiple aspects of vulnerability requires better measures of the social value of water as well as demand.

ACKNOWLEDGEMENTS

The acknowledgements section of a dissertation should probably be the length of the dissertation itself. There are so many people without whom I could never have completed this research. I am first and foremost grateful to my partner, editor, cheerleader, critic, and rock: Austen Lorenz. I could never have done this without you—I am so grateful for you always. To my son Crawford, my nieces—Violet and Eloise, and nephew Everett: Thank you all for giving me a reason to care about the future and remember the magic of the world. To my endlessly supportive and loving parents, thank you for always encouraging me to pursue science and listening to me even when I'm insufferable. Legalize physics! To my siblings, thanks for always making me laugh even when things are hard. To my Aunt, Dr.[3] Susie Muir, thank you for always supporting me and validating that PhDs are hard. To Poppi, Kathi, and Emersen: Thank you for your love and support. To my friends, my appreciation for you knows no bounds—thank you for all you have done to support me.

I would also like to express my sincerest thanks to my advisor and my committee. Dr. Adrian Harpold: your ability to balance exacting standards with the bigger picture is a gift. Thank you for helping me grow and always treating my ideas with respect even when they are bad. Dr. Newsha Ajami: my greatest hope is that every young scientist finds a mentor, friend, and role model like you some day. Thank you for everything. Dr. Rosemary Carroll: Your perspective and warmth are such a delight and I can't wait to continue working with you. Dr. Elizabeth Koebele: I am so grateful for all the time you spent helping me understand a new field. To Dr. Alejandro Andrade-Rodriguez: Thank you so much for your approachability and thoughtfulness. I am so grateful to have connected.

To the countless other mentors I've been so lucky enough to encounter: Dr. Dusty Schroeder and Devon Ryan, Dr. Arthur Middleton and Anna Sale, Dr. Siva Sundareson, Dr. Alex Konings, Dr. Tara Moran, Leon Szeptycki, Dr. Ann Hild, Dr. James Dennedy-Frank, Dr. Sue Barsom and Paul Barsom—thank you for being so wonderful. Thank you also to all the folks who have meandered into and out of the Nevada Mountain Ecohydrology Lab while I've been in it—it has been such a gift getting to know you all.

Contents

TABLE OF TABLES

TABLE OF FIGURES

1 Chapter 1: Introduction

Snowmelt-driven water supplies are one of the fastest changing aspects of the global hydrologic cycle in response to climate change (Musselman *et al* 2017). Warmer winter and spring temperatures are decreasing the fraction of precipitation falling as snow (Knowles *et al* 2006, Klos *et al* 2014), delaying the initiation of consistent snow cover, increasing soil frost (Wobus *et al* 2017, Burakowski *et al* 2008), increasing water vapor exchanges between snowpack and the atmosphere (Harpold and Brooks 2018, Sexstone *et al* 2018), advancing the timing and slowing the rate of snowmelt (Musselman *et al* 2017), and decreasing the persistence of snow cover (Stewart 2009). In the United States (US) alone, snowmelt runoff from high elevation, mountainous (upland) catchments serves water to over 60 million people and supports billions of dollars in economic productivity (Barnett *et al* 2005, Sturm *et al* 2017) with additional impacts on ecosystem health (Allan and Castillo 2007), wildfires (Holden *et al* 2012, Westerling *et al* 2006), flood risk (Hamlet and Lettenmaier 2007), and reservoir management (Ehsani *et al* 2017).

Water supplied from mountain environments is perhaps most important in semi-arid regions of the northern hemisphere—particularly the western US—where the predictable cycle of seasonal snow accumulation and melt, together with vast networks of physical and legal infrastructure, enable development in otherwise water limited environments (Meybeck *et al* 2001, Qin *et al* 2020, Church 1933). These agricultural systems—including the infrastructure, institutions, and stakeholders (i.e., managers and users) encompassing them—support vast economic productivity (Barnett et al., 2008) and provide critical ecosystem services (Claes et al., 2021; Gordon et al., 2020, 2019) are thus effectively coupled to uplands where most precipitation falls as snow (Hansen *et al*., 2011; Li *et al*., 2017; Mankin *et al*., 2015; McCabe and Clark, 2005; Qin *et al*., 2020; Viviroli *et al*., 2007). This reliance leaves systems particularly vulnerable to both seasonal and multi-decadal changes in higher elevation snow processes (Swain *et al*., 2016). However, the capacity for these systems to adapt to these changes remains less well understood.

To address the adaptive capacity of human systems to shifts in upland snow dynamics, this dissertation asks: what elements make systems vulnerable to disconnects between water supply and demand? And secondly, what elements enhance system flexibility in the face of these changes? Contributing new information to existing and foundational examinations of system vulnerability to changing snow resources (e.g., Barnett *et al*., 2005; Mankin *et al*., 2015; Qin *et al*., 2020; Viviroli *et al*., 2007) requires a more holistic and interdisciplinary view of these systems themselves. Motivated by this pressing need, this research uses a mix of physical and social science methods and data to answer a pressing scientific question: How can we better characterize the vulnerability—and adaptive capacity— of socio-hydrologic systems to shifts in water supply and demand as a consequence of climate change?

We explore social-hydrologic vulnerability and adaptive capacity in four parts, which are outlined in Figure 1.1. Each Chapter is presented as a stand-alone, peer-reviewed publication. The first two Chapters characterize how climate change is stressing critical mountain water supply. In Chapter 3, we then examine how changes in mountain hydrology interact with society, specifically downstream agricultural production. Lastly in Chapter 4, we outline a more generalizable framework for characterizing vulnerability by integrating physical and social aspects of system vulnerability. Each Chapter builds on the previous Chapter to present a comprehensive understanding of changing mountain water resources—and critically, the mechanisms (Chapter 2), tools (Chapter 3), interactions (Chapter 4), and approaches (Chapter 5) that must be considered in assessing societal vulnerability to these changes.

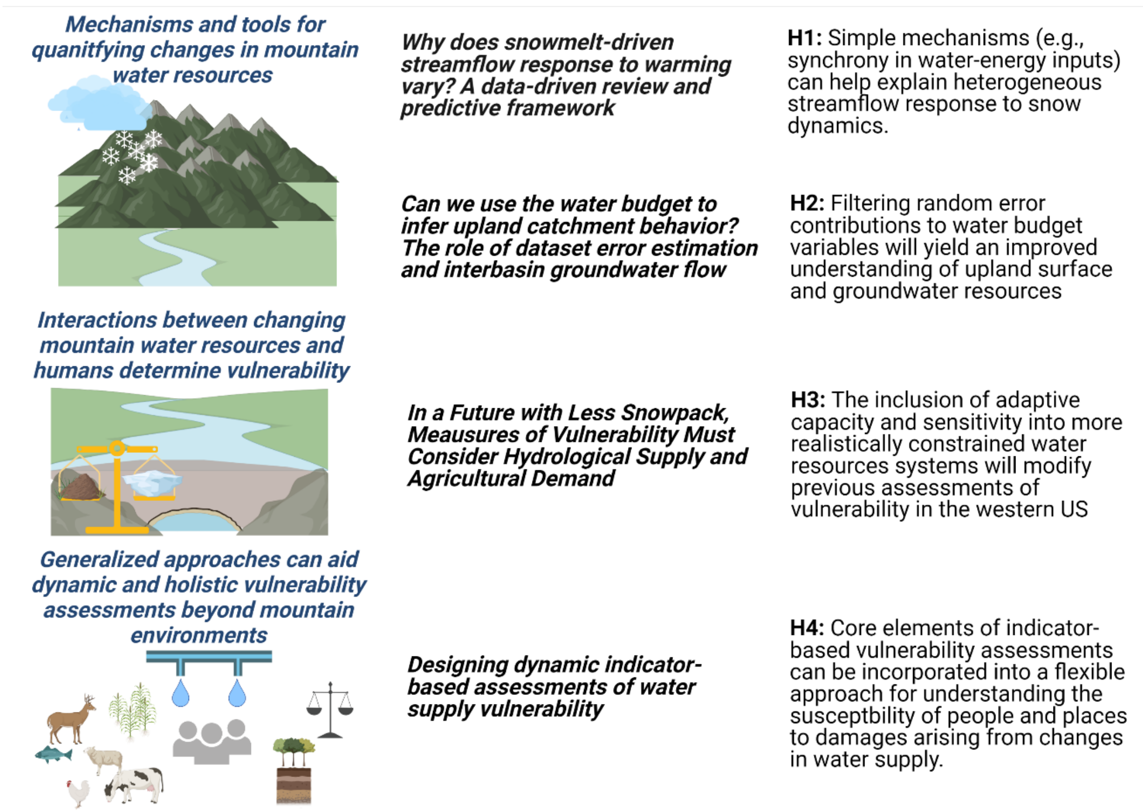

Figure 1-1:Schematic of Chapters of this dissertation and driving hypotheses.

1.1 Chapter 2: Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework

Research on social-hydrologic systems begins with an investigation of what is known, and more importantly what is not known, about changes in seasonal snow accumulation and ablation across mid-latitude mountainous regions in the Northern Hemisphere (Gordon et al., 2022a). A systematic literature review on seasonal snowmelt-driven streamflow and how it is altered by climate change serves to highlight unsettled questions about how annual streamflow volume is shaped by changing snow dynamics. From this literature review, a framework is developed based on three testable, inter-related mechanisms, (i) snow season mass and energy exchanges, (ii) the intensity of snow season liquid water inputs, and (iii) the synchrony of energy and water availability. Each mechanism is explored using data distributed across the United States. We show that streamflow prediction is more challenging in regions with multiple interacting mechanisms.

1.2 Chapter 3: Can we use the water budget to infer upland catchment behavior? The role of dataset error estimation and interbasin groundwater flow

Equipped with a broad understanding of the challenges and opportunities in predicting shifts in water resources as snow dynamics change, Chapter 3 investigates one commonly used tool—the water budget— for evaluating mountain water resources (Gordon et al., 2022b). The focus is on the often unappreciated role of groundwater in mountain hydrologic systems and measurement error (ε) in the characterization of upland water supply. We examination the shortcomings of closed water budgets (CWB), which ignore difficult-to-measure variables, including inter-basin groundwater fluxes (G) and ε to derive evapotranspiration (ET) from precipitation (P) and streamflow (Q) (e.g., the Budyko hypothesis). We contrast the shortcomings of CWB with open water budgets (OWB), which take advantage of remotely sensed ET products, physically-based frameworks for improving inferences about G, and tools to statistically characterize ε (Triple Collocation, TC). The value of these advances in upland settings is clarified by comparing standard land surface model, Ensemble Mean, and TC-Merged P and ET products in 114 upland catchments. When compared against a long-term OWB, we find that the CWB assumptions are unsupported in 75-100% of the catchments. Research highlights that groundwater resources is an important component of mountain hydrology, and tools like TC, a Fan (2019) framework, and *ET* products improve quantification of water resources in a changing climate.

1.3 Chapter 4: Water Management Can Reduce Agricultural Vulnerability to Decreasing Snowpack

We contextualize findings about physical hydrological changes in mountain water supplies from Chapters 2 and 3 into a socio-hydrologic analysis of vulnerability. Specifically, research examines how humans modify the hydrological cycle via adaptation to changes in the distribution and magnitude of vulnerability in the western US. Vulnerability is defined using at an operational scale using the Exposure, Sensitivity, and Adaptive Capacity framework (Cardona et al., 2012). The approach is tested in 13 basins experiencing declining snowpack across the western US. These basins rely on a mix of snow and reservoir storage for agricultural production. Research evaluates if these basins can adapt to projected declines in snow using two different strategies: 1) enhancing reservoir or groundwater storage capacity via tools like managed aquifer recharge or conjunctive use; and/or 2) reducing water use via demand management (i.e., fallowing). Results show that these strategies are most effective if implemented rapidly and in systems with a higher proportion of hay production relative to overall demand, and with smaller declines in snow relative to reservoir capacity. Adaptation yields the largest reductions in vulnerability in the near future (2020-2050) in higher elevation tributaries of the Missouri

Basin and the least benefit in the far future to certain tributaries of the California and Upper Colorado Basins.

1.4 Chapter 5: Designing dynamic indicator-based vulnerability assessments

Lastly, research in socio-hydrologic systems considers a framework to assess water supply vulnerability in a flexible and multidimensional manner across a range of hydrologic systems. Drawing from existing global assessments, we propose a conceptual model and then derive a general approach to water supply vulnerability assessment that can be used to evaluate multiple aspects of performance (e.g., social, environmental, etc.) in an ongoing manner. We then validate this approach using interdisciplinary analyses on a subset of indicators from 20 existing indices and find that multiple vulnerability frameworks can be integrated into indicator-based assessments. We show that certain key aspects of multidimensional system performance (termed domains) can capture a spectrum of existing indicators. However, we also underscore high potential risk for silo-ing when drawing upon these indicators—particularly with regard to measures of physical performance where redundancies and biases are substantial. Although several pathways for standardizing, aggregating, and evaluating multidimensional indicators exist, we highlight significant gaps in measures of cultural water use and values, urban water use, and groundwater; all of which lack widely available data for evaluation. Using both the raw data and the results of these analyses, we establish a database to operationalize our derived dynamic, multidimensional approach while maintaining the benefits of indicator-based assessments for water managers and policy-makers. We then provide a template for how this approach can be applied in practical settings.

1.5 References

Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature. https://doi.org/10.1038/nature04141

Barnett, T.P., Pierce, D.W., Hidalgo, H.G., Bonfils, C., Santer, B.D., Das, T., Bala, G., Wood, A.W., Nozawa, T., Mirin, A.A., Cayan, D.R., Dettinger, M.D., 2008. Human-induced changes in the hydrology of the Western United States. Science (80-.). 319, 1080–1083. https://doi.org/10.1126/science.1152538

Cardona, O.-D., van Aalst, M.K., Birkmann, Jörn, Fordham, Maureen, McGregor, Glenn, Perez, Rosa, Pulwarty, R.S., Lisa Schipper, E.F., Tan Sinh, B., Décamps, H., Keim, M., Davis, I., van Aalst, M., Birkmann, J, Fordham, M, McGregor, G, Perez, R, Pulwarty, R., Schipper, E., Sinh, B., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mach, K., Plattner, G., Allen, S., Tignor, M., Midgley, P., 2012. Coordinating Lead Authors: Lead Authors: Review Editors: Contributing Authors:: Determinants of risk: exposure and vulnerability. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation 2 Determinants of Risk: Exposure and Vulnerability.

Claes, N., Paige, G.B., Gordon, B.L., Parsekian, A.D., Miller, S.N., 2021. Hydrologic modeling of reach scale fluxes from flood irrigated fields. J. Hydrol. 598, 126254. https://doi.org/10.1016/J.JHYDROL.2021.126254

Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrame, G. F., Carroll, R. W., McNamara, J. P., & Harpold, A. A. (2022a). Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework. *Environmental Research Letters*.

Gordon, B. L., Crow, W. T., Konings, A. G., Dralle, D. N., & Harpold, A. A. (2022b). Can We Use the Water Budget to Infer Upland Catchment Behavior? The Role of Data Set Error Estimation and Interbasin Groundwater Flow. *Water Resources Research*, *58*(9), e2021WR030966.

Gordon, B.L., Kowal, V.A., Khadka, A., Chaplin-Kramer, R., Roath, R., Bryant, B.P., 2019. Existing Accessible Modeling Tools Offer Limited Support to Evaluation of Impact Investment in Rangeland Ecosystem Services. Front. Sustain. Food Syst. 3, 77. https://doi.org/10.3389/fsufs.2019.00077

Gordon, B.L., Paige, G.B., Miller, S.N., Claes, N., Parsekian, A.D., 2020. Field scale quantification indicates potential for variability in return flows from flood irrigation in the high altitude western US. Agric. Water Manag. 232, 106062. https://doi.org/10.1016/j.agwat.2020.106062

Hansen, Z.K., Libecap, G.D., Lowe, S.E., 2011. Chapter Title: Climate Variability and Water Infrastructure: Historical Experience in the Western United States. University of Chicago Press.

Li, D., Wrzesien, M.L., Durand, M., Adam, J., Lettenmaier, D.P., 2017. How much runoff originates as snow in the western United States, and how will that change in the future? Geophys. Res. Lett. 44, 6163–6172. https://doi.org/10.1002/2017GL073551

Mankin, J.S., Viviroli, D., Singh, D., Hoekstra, A.Y., Diffenbaugh, N.S., 2015. The potential for snow to supply human water demand in the present and future. Environ. Res. Lett. 10, 114016. https://doi.org/10.1088/1748-9326/10/11/114016

McCabe, G.J., Clark, M.P., 2005. Trends and variability in snowmelt runoff in the western United States. J. Hydrometeorol. 6, 476–482. https://doi.org/10.1175/JHM428.1

Qin, Y., Abatzoglou, J.T., Siebert, S., Huning, L.S., AghaKouchak, A., Mankin, J.S., Hong, C., Tong, D., Davis, S.J., Mueller, N.D., 2020. Agricultural risks from changing snowmelt. Nat. Clim. Chang. 10, 459–465. https://doi.org/10.1038/s41558-020-0746-8

Swain, D.L., Horton, D.E., Singh, D., Diffenbaugh, N.S., 2016. Trends in atmospheric patterns conducive to seasonal precipitation and temperature extremes in California. Sci. Adv. 2. https://doi.org/10.1126/SCIADV.1501344/SUPPL_FILE/1501344_SM.PDF

Viviroli, D., Dürr, H.H., Messerli, B., Meybeck, M., Weingartner, R., 2007. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water Resour. Res. 43, 7447. https://doi.org/10.1029/2006WR005653

2 Chapter 2: Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework

By: Beatrice L. Gordon, Paul D. Brooks, Sebastian A. Krogh, Gabrielle F.S. Boisrame, Rosemary W.H. Carroll, James P. McNamara, Adrian A. Harpold

Citation: Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrame, G. F., Carroll, R. W., McNamara, J. P., & Harpold, A. A. (2022). Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework. Environmental Research Letters.

Abstract

Climate change is altering the seasonal accumulation and ablation of snow across mid-latitude mountainous regions in the Northern Hemisphere with profound implications for the water resources available to downstream communities and environments. Despite decades of empirical and model-based research on snowmelt-driven streamflow, our ability to predict whether streamflow will increase or decrease in a changing climate remains limited by two factors. First, predictions are fundamentally limited by the high spatial and temporal variability in the processes that control net snow accumulation and ablation across mountainous environments. Second, we lack a consistent and testable framework to coordinate research to determine which dominant mechanisms influencing seasonal snow dynamics are most/least important for streamflow generation in different basins. Our data-driven review marks a step towards the development of such a framework. We first conduct a systematic literature review that synthesizes knowledge about seasonal snowmelt-driven streamflow and how it is altered by climate change, highlighting unsettled questions about how annual streamflow volume is shaped by changing snow dynamics. Drawing from literature, we then propose a framework comprised of three testable, inter-related mechanisms—snow season mass and energy exchanges, the intensity of snow season liquid

water inputs, and the synchrony of energy and water availability. Using data for 537 catchments in the United States, we demonstrate the utility of each mechanism and suggest that streamflow prediction will be more challenging in regions with multiple interacting mechanisms. This framework is intended to inform the research community and improve management predictions as it is tested and refined.

2.1 Introduction

Snowmelt-driven streamflow is a critical—and increasingly vulnerable—freshwater resource for downstream environments supporting agriculture, municipalities, and native ecosystems (Viviroli *et al* 2007, Immerzeel *et al* 2020, Viviroli *et al* 2020, Li *et al* 2017, Mankin *et al* 2015). In the United States (US) alone, snowmelt runoff from high elevation, mountainous (upland) catchments serves water to over 60 million people and supports billions of dollars in economic productivity (Barnett *et al* 2005, Sturm *et al* 2017) with additional impacts on ecosystem health (Allan and Castillo 2007), the extent and severity of wildfires (Holden *et al* 2012, Westerling *et al* 2006), flood risk (Hamlet and Lettenmaier 2007), and reservoir management (Ehsani *et al* 2017). Snowmelt-derived water resources are perhaps most important in semi-arid regions where the predictable cycle of seasonal snow accumulation and melt, together with vast networks of physical and legal infrastructure, enable development in otherwise water limited environments (Meybeck *et al* 2001, Qin *et al* 2020, Church 1933). Large interannual variability in snowmelt water supply is driven primarily by variability in winter precipitation (P), but also by 30% or higher variability in runoff efficiency across basins (Mote 2003, Brooks *et al* 2021, Harpold *et al* 2012). In spite of this variability, decades of observations have resulted in reasonably skilled water supply models dependent on readily observable metrics such as April 1 Snow

Water Equivalent (*SWE*) and temperature-driven melt models (Pagano *et al* 2004). The last few decades, however, have seen a large decrease in the skill of these models (Pagano *et al* 2004) leading to diverse suggestions about limitations in understanding and predictability of water supply (Milly *et al* 2008, Montanari and Koutsoyiannis 2014, Milly *et al* 2015). This review focuses on the mechanisms that give rise to spatial and temporal variability in runoff efficiency as snow cover changes in response to a warming climate.

Seasonal snowmelt-driven streamflow is one of the fastest changing aspects of the global hydrologic cycle in response to climate change (Musselman *et al* 2017). Warmer winter and spring temperatures are decreasing the fraction of precipitation falling as snow (*fs*, Knowles *et al* 2006, Klos *et al* 2014), delaying the initiation of consistent snow cover, increasing soil frost (Wobus *et al* 2017, Burakowski *et al* 2008), increasing water vapor exchanges between snowpack and the atmosphere (Harpold and Brooks 2018, Sexstone *et al* 2018), advancing the timing and slowing the rate of snowmelt (Musselman *et al* 2017), and decreasing the persistence of snow cover (Stewart 2009). In contrast to a general consensus on seasonal snow cover decline under a warming climate, predictions about streamflow response are much more diverse. For example, recent research suggests that changing snow dynamics may or may not lead to earlier streamflow timing (Fritze *et al* 2011, Stewart *et al* 2005, Moore *et al* 2007), and may or may not alter annual runoff efficiency (Berghuijs *et al* 2014, McCabe *et al* 2018). These uncertainties in how changes in seasonal snow dynamics will influence streamflow on the scale at which management decisions are made (e.g., 100-10000 km^2) remain and have been the focus of research seeking to improve empirical and model-based forecasting tools (Huang *et al* 2017; Siirila-Woodburn *et al* 2021).

The conflicting reports on snowmelt-driven streamflow response to a changing climate result in large part from regional differences in the mechanisms controlling seasonal snow accumulation and ablation. To better anticipate and adapt to changing upland snow resources requires a consistent and testable framework to characterize variability in the dominant processes that drive differences in energy-water coupling during the snow season. Such a framework would help focus research priorities as well as help the scientific community describe variability in how, where, and why streamflow will be impacted by changing snow dynamics.

Our data-driven review presents a consistent framework to answer a question fundamental to hydrologic sciences and water management: What are the mechanisms that give rise to variable streamflow response to changes in the amount, timing, accumulation, and ablation of winter snow cover? We first synthesize current understanding of the seasonal cycle of snow accumulation and ablation before summarizing prior work—and outstanding questions—on how changes in this cycle under climate change are altering the timing, intra-annual distribution, and annual volume of streamflow. We then use our literature review to derive a simple framework centered around three potentially competing mechanisms that capture how abiotic and biotic factors differentially impact the interactions between energy and water during the snow season. These mechanisms include: 1) snow season water vapor losses, 2) the intensity of liquid water inputs (*LWI*), which we define as the amount of liquid water that reaches the ground surface from either rain or melting snow, and 3) the synchrony of water and energy availability. We include a number of demonstrative experiments that highlight which regions across the continental United

States (CONUS) may be most sensitive to changes in a single snow metric (*fs*) before suggesting research opportunities.

2.2 Accumulation and ablation of seasonal snow cover

The annual cycle of snow accumulation and ablation is driven primarily by incoming solar radiation, secondarily by longwave radiation exchanges between the snowpack and atmosphere, and to a lesser degree by turbulent exchanges of sensible and latent heat and ground heat flux (Marks and Dozier 1992). Feedbacks between radiative and turbulent energy exchanges may either exacerbate or buffer the effects of climate change, resulting in spatially variable responses to the widespread warming observed across western North America (Harpold *et al* 2012, Harpold and Brooks 2018, Bormann *et al* 2018). To capture spatially and temporally variable responses in snow conditions, studies have proposed a number of snow metrics, which we review below.

2.2.1 Snow dynamics, terminology, and measurement

Point-scale metrics such as event snowfall, accumulated snow depth, and peak *SWE* have been broadly adopted to quantify the amount of snow on the ground (Bohr and Aguado 2001, Broxton *et al* 2016). Lapse rate models and assumptions regarding the relationship between temperature (*T*) and precipitation (*P*) phase have been widely used to estimate *fs* to quantify the contribution of snowfall versus rainfall to annual *P* (Karl *et al* 1993, Knowles *et al* 2006, Klos *et al* 2014). Recent progress in remote sensing (Lundquist *et al* 2008; Maggioni *et al* 2016; Skofronick-Jackson *et al* 2018) of snowfall in complex terrain is helping to distinguish between snow fall and redistribution or ablation of the snowpack. Contemporaneous advances in airborne and space-based remote sensing have also enhanced the measurement of snow cover and *SWE* at large spatial scales (JianCheng *et al* 2016, Tedesco *et al* 2014), facilitating improved gridded snow products (e.g., Broxton *et al* 2016). In spite of these advances, accurate estimation of the amount, phase, and intensity of *P* in complex mountain environments remains a challenge (Rasmussen *et al* 2012) with evidence that high-resolution atmospheric models may be particularly useful tools for enabling further improvement (Lundquist *et al* 2019). In this vein, combinations of process-based snow modeling (Painter *et al* 2016) and observations are beginning to uncover terrain-mediated complexities in snow cover (Dong *et al* 2005). Together, these advances—particularly with respect to satellite-based sensors like the Moderate Resolution Imaging Spectroradiometer (MODIS) —have improved our ability to characterize the temporal and spatial patterns of snow dynamics across regional scales. Measures facilitated by this progress include the extent (Karl *et al* 1993, Groisman and Legates 1994, Rupp *et al* 2013, Tennant *et al* 2017, Painter *et al* 2016), duration (Bulygina *et al* 2009), and persistence of snow cover (Hammond *et al* 2018).

Historical challenges in obtaining direct measurements and developing scalable metrics for snowfall, snow fraction, or *SWE* led to the focus on choosing a fixed date (e.g., April 1) to estimate net snow water input (Changnon *et al* 1991, Cayan 1996). However, Hamlet *et al* (2005) used trend analyses on March 1, April 1, and May 1 *SWE* to show how regions experience differential sensitivity to changes in *P* and *T* (e.g., higher sensitivity of *SWE* to *P*in cold regions and higher sensitivity of *SWE* to *T* in warm regions). Alternative, although less common, approaches have relied on the day of the water year (DoWY) of peak *SWE* to account for regional variance in snow accumulation and ablation (Bohr and Aguado

2001, Girotto *et al* 2014). We summarize several of these metrics and provide a standardized definition to assist in their evaluation in Table 1.

Table 2-1: Overview of key snow metrics considered in this paper. In the table, we present units of measurement in terms of t = time and l = length.

Term	Unit	Definition	Citation
SNOW METRICS			
Snow season	t	Water year: the length of time from the first occurrence of snow to the last occurrence of snow. Site average: site average snow season can be found between the 10^{th} and 90^{th} percentile of days with snow on the ground	(Hammond *et al* 2018)
Snow Fraction	$[l\ l^{-1}]$	Water year: the fraction of annual precipitation that falls as snow Snow season**:** the fraction of snow season precipitation that falls as snow, determined using the snow season metric above.	(Klos *et al* 2014)
Snow persistence	$[t\ t^{-1}]$	Water year: the fraction of time that snow is present on the ground	(Moore *et al* 2007, Hammond *et al* 2018)
Mean daily snowmelt rate	$[l\ t^{-1}]$	Water year: the average daily melt rate from peak *SWE* to the day of snow disappearance	(Trujillo and Molotch 2014)
Peak *SWE*	[l]	Water year: the maximum amount of *SWE* on the ground per snow season	(Bohr and Aguado 2001)
Day of peak *SWE*	[t]	Water year: the day of water year when peak *SWE* occurs	(Trujillo and Molotch 2014)

2.2.2 Snowpack energy balance

A warming climate interacts with seasonal snow cover through energy exchanges that occur between the snow surface and the atmosphere. We ground our discussion of the potential effects of these changes in the energy balance of a snowpack, written as:

$$\Delta \mathrm{q} = \sum \mathrm{F}\Delta \mathrm{t} \qquad (2\text{-}1)$$

Where Δq is the change in energy [J m^{-2}], t is time [s] and F is net energy flux [W m^{-2}] and:

$$\sum \mathrm{F} = \mathrm{R_n} + \mathrm{H} + \mathrm{LE} + \mathrm{C} + \mathrm{M} \qquad (2\text{-}2)$$

Net radiative fluxes (net radiation) [R_n; W m-2] dominates snowpack energy balance and is composed of incoming solar (positive, towards the snow surface) and bi-directional longwave radiation (see Eq. (2-6)) (Marks and Dozier 1992). H is net sensible heat flux [W m-2] and maybe be positive (downwards towards the snow surface) or negative (upwards away from the snow surface) depending on snow and air *T* (Marks and Dozier 1992). LE is net latent heat flux [W m^{-2}] and is typically negative (outgoing away from the snow surface) in continental snowpacks but may be positive in maritime environments. Both H and LE fluxes strongly relate to boundary layer turbulence and wind speed that drive air exchanges between snowpack and overlying atmosphere (Massman *et al* 1997; Lee *et al* 2004). C is typically small net conductive (soil) energy flux [W m^{-2}] and M is net advective energy flux typically associated with melt water loss [W m^{-2}] (Marks and Dozier 1992).

When $\sum F\Delta t = 0$, the snowpack is in thermal equilibrium, when $\sum F\Delta t < 0$, the snowpack is cooling or refreezing, and when $\sum F\Delta t > 0$, the snowpack is warming or melting. The change in energy state of the snowpack depends on the average snowpack temperature (T_s).

$$If\ \mathrm{T_s} < 0\ °\mathrm{C}: \Delta \mathrm{q} = \Delta \mathrm{q_{cc}}\ ,$$

$$If\ \mathrm{T_s} = 0\ °\mathrm{C}: \Delta \mathrm{q} = \Delta \mathrm{q_{melt}} \tag{2-3}$$

Where q_{cc} [J m^{-2}] is commonly known as the cold content and is the total energy required to raise the T_s to 0 °C:

$$\mathrm{q_{cc}} = -\mathrm{c_i}\rho_\mathrm{w}\mathrm{h}_{SWE}(\mathrm{T_s} - \mathrm{T_m}) \tag{2-4}$$

And q_{melt} [J m^{-2}] is the energy associated with phase change:

$$\mathrm{q_{melt}} = (\mathrm{h}_{SWE})\rho_\mathrm{w}\gamma_\mathrm{f} \tag{2-5}$$

Where c_i is the heat capacity of ice [2102 J kg^{-1} K^{-1}], T_s is the average *T* of the snowpack, T_m is the melting point of ice (0° C), ρ_w is the density of water [~1000 kg m^{-3}], and h_s is the snow depth [m], h_{SWE} is the snow water equivalent [m], and γ is the latent heat of fusion [J kg^{-1}].

Determining the response of snowpack to climate change is complicated by the fact that turbulent exchanges associated with *T* and Le are typically much small than radiative fluxes (Marks and Dozier, 1992), with Rn varying seasonally as a function of solar angle, day length, T, cloudiness, and spatially due to near surface topography, terrain, and vegetation structure. These complexities require a closer examination of snowpack radiative energy balance:

$$\mathrm{R_n} = (1 - \alpha)\ \mathrm{R_s} + \mathrm{R_{l-in}} - \mathrm{R_{l-out}} \tag{2-6}$$

Where R_s is incoming (positive) solar radiation [W m^{-2}], α is albedo [-], $R_{l\text{-}in}$ is incoming (positive) longwave radiation [W m^{-2}], and $R_{l\text{-}out}$ is outgoing (negative) longwave radiation [W m^{-2}]. During sunny days, R_s—which is driven by solar angle (a function of day of year, latitude), aspect, topographic reflectance, and both topographic and vegetative shading—is positive and typically the largest energy flux in Eq. (2-6). The albedo (α) of fresh snow is high although it can be modified by snow grain size (typically related to time since last snowfall) and impurities in the snowpack (Deems *et al* 2013, Skiles *et al* 2012). Net longwave radiation ($R_{l\text{-}in}$ - $R_{l\text{-}out}$) is a function of the T of the snowpack, T of the overlying atmosphere, and differences in emissivity of snow and air (Marks and Dozier 1992). For example, the atmosphere has a lower emissivity than snow resulting in a cooling of the snowpack below ambient air T, especially at night. In contrast, clouds have similar emissivity as snow which may prevent snowpacks from cooling, especially at night (Ambach 1974, Plüss and Ohmura 1997).

The amount, extent, persistence, and freshness of snow strongly influence R_n (e.g. Meira-Neto *et al* 2020) by altering the fraction of R_s that is reflected (Schneider and Dickinson 1974, Ingram *et al* 1989). Snow has much higher albedo than most terrestrial surfaces and thus, as R_n increases the climate system reduces snow cover in a positive feedback process known as the snow-albedo feedback (Thackeray and Fletcher 2016, Hall 2004, Déry and Brown 2007, Fletcher *et al* 2012, Qu and Hall 2014). If early season snowfall is sufficient to cover the lower albedo terrestrial surfaces, the snow-albedo feedback will reduce R_n and tend to preserve snow cover until solar angles increase in the spring (e.g., Koster *et al* 2010). A reduction in snow cover during spring when Rs is higher can enhance local warming (Thackeray and Fletcher 2016, Hall 2004, Déry and Brown 2007) with the future potential for a 1% reduction in surface albedo per degree of warming (Fletcher *et al* 2012, Hall *et al* 2008, Qu and Hall 2014). In contrast, spring snowfall events may increase albedo, reducing R_n, colling the local environment, and delaying melt. Superimposed on these larger scale radiative feedbacks are the effects of landscape heterogeneity, including aspect and vegetation (Harding and Pomeroy 1996, Broxton *et al* 2015, Tennant *et al* 2017) that remain difficult to measure in complex, mountainous terrain (Reba *et al* 2009) and under variable snow cover conditions (Schlögl *et al* 2018). These feedbacks may result in high local variability in the partitioning of available energy to sublimation fluxes, cooling the snowpack, versus greater energy fluxes causing melt, advancing snowmelt, and causing snow-albedo feedbacks (Sexstone *et al* 2018).

2.2.3 Water vapor fluxes between atmosphere and seasonal snowpacks

Complex interactions between the snow surface and the atmosphere drive variability in the amount of water vapor lost during the snow season. Strong T lapse rates associated with either orographic or convective uplift can result in snowfall during most months in high mountains; however, consistent seasonal snow cover does not begin to accumulate until solar angles are low and Rn favors net cooling of the land surface following a fresh snowfall (Bales *et al* 2006, Markovich *et al* 2019). During the snow season, air T is low and transpiration is typically assumed to be limited (Bowling *et al* 2018, Day *et al* 1989, Huxman *et al* 2003, Goulden and Bales 2014); however, other water vapor losses are possible. For example, exposed snowpacks above treeline, in meadows, forest gaps, fields, and in forest canopies are subject to considerable vapor loss over winter from sublimation (Sexstone *et al* 2018, Harpold *et al* 2012, Biederman *et al* 2015, Molotch *et al* 2009, Veatch *et al* 2009, Gustafson *et al* 2010, Rinehart *et al* 2008). In continental alpine systems, 20% to 30% of winter snowfall may sublimate before melt (Hood *et al* 1999, Sexstone *et al* 2018). Sublimation effects are most dominant during snowpack accumulation and increase with low atmospheric pressure, low humidity, increased solar radiation and high wind speed (Lundberg and Halldin 2001, Earman *et al* 2006, Stigter *et al* 2018). Sublimation of snow intercepted by forest canopies are estimated at roughly 30% of local snowfall depending on leaf area (Essery *et al* 2003, Storck *et al* 2002), but similar to sublimation from open exposed snowpack on the ground, their sensitivity to climate change is poorly characterized (Lundquist *et al* 2021). Potential water vapor losses from the snowpack to the atmosphere may increase due to greater energy availability (both R_n and *T*) (Meira-Neto *et al* 2020) and lower *LE* associated with evaporation of liquid water within the snowpack relative to sublimation of ice crystals in winter (Jambon-Puillet *et al* 2018).

2.2.4 Snowmelt and catchment liquid water input (*LWI*)

During the snow season, the intensity and timing of *LWI*—the amount of liquid from either rain or melting snow that reaches the ground surface in a given control volume (e.g., catchment) — is typically a function of snowmelt rates (Trujillo and Molotch 2014, Harpold and Brooks 2018, Musselman *et al* 2017; Harpold and Kohler 2017) and is thus relatively predictable (Harpold and Kohler 2017). However, the season-long interaction between *P*, snow accumulation and redistribution, and ablation processes causes highly heterogeneous snowmelt (Tennant *et al* 2017). Snowpacks become isothermal at 0 °C and begin to melt as solar angle increases, days lengthen, and surface albedo decreases (Cline 1997, Skiles *et al* 2012), which can be influenced by warming *T*, cloud cover, increased humidity (Clow 2010, Harpold and Brooks 2018), as well as snowpack depth and aspect

(Kormos *et al* 2014, Christensen *et al* 2021). For snowmelt water infiltration into the soil zone sufficient melt must occur to overcome matric forces within the snowpack (including preferential flowpaths), which is often referred to as the snowpack becoming "ripe" (Leroux and Pomeroy 2017, Marsh and Woo 1984). As such, large snow-covered areas in the catchment which receive sufficient energy to overcome cold content and become ripe experience a relatively predictable seasonal increase in *LWI* driven by snowmelt (Dunne and Black 1971). When driven by snowmelt, *LWI* tend to occur over a longer duration than when driven by rain and at an intensity well below the infiltration capacity of most mountain soils, particularly those with well-developed organic layers (Liu *et al* 2008). In the absence of rare extremely warm rainfall and condensation (rain on snow) events, for example, snowmelt rates in the western CONUS rarely exceed 10 cm per day and average closer to 2.5 cm per day (Harpold and Kohler 2017). An important exception is frozen soils, where lower infiltration rates can be exceeded by *LWI* (Shanley and Chalmers 1999, Bayard *et al* 2005). As a result, *LWI* driven primarily by snowmelt during the snow season are often associated with more consistent hydrological effects on shallow subsurface flow response, which is the dominant pathway for water redistribution and streamflow generation in upland catchments (Barthold and Woods 2015).

Because of moisture thresholds imposed by catchment-scale properties (e.g., infiltration capacity or soil water holding capacity), both the timing and intensity of *LWI* control how it is partitioned between subsurface and surface pathways (Harr 1981, Barnhart *et al* 2016, Berghuijs *et al* 2016). In general, snowmelt lags *P* inputs and there is some evidence suggesting that sequential melt results in more substantial soil moisture response—especially at deeper depths—than ephemeral snowmelt and rainfall (Kormos *et al* 2014, Petersky and Harpold 2018, Hammond *et al* 2019). However, it remains challenging to estimate how LWI will be partitioned between the atmosphere, streamflow, and subsurface storage at catchment scales (Meixner *et al* 2016, Frisbee *et al* 2012, Harpold *et al* 2012, Brooks *et al* 2015, Blöschl *et al* 2019). When the delivery of LWI exceeds catchment-specific thresholds such as soil water holding capacity, infiltration capacity, and/or rates of water vapor losses, it promotes subsurface drainage below the rootzone or activates subsurface and surface lateral flow (Seyfried *et al* 2009, Hammond *et al* 2019).

2.3 Streamflow response to changing snow dynamics

Interactions between snowmelt-driven LWI and the subsurface, atmosphere, and vegetation exert a complex control over streamflow generation in mountainous catchments. For example, subsurface and surface lateral flow arising from seasonal increases in snowmelt-driven LWI leads to seasonal increases in streamflow generation via a number of mechanisms including infiltration excess overland flow (Horton 1933), saturation excess overland flow (Dunne 1978), preferential subsurface flow, as well as fill and spill flow in certain cases (Meerveld and McDonnell 2006, McDonnell *et al* 2021). To help illustrate these connections, we adopt the below form of the snow season water budget for upland catchments following Godsey *et al* (2014), which we modify to include error:

$$\Delta S = LWI\ - ET - Q\ +/- \epsilon \tag{2-7}$$

Where ΔS is the change in storage within the catchment excluding storage in the snowpack itself [mm], LWI [mm] is the effective liquid rain and snowmelt water input into the catchment, ET [mm] is combined water vapor losses from the catchment, Q is streamflow [mm] exiting the catchment, and ϵ is any error, including unaccounted fluxes or stores within the catchment [mm]. Recent commentaries have noted that the term *ET* is linguistically imprecise and inconclusive with respect to interception fluxes (Miralles *et al* 2020). We adopt *ET* in this paper due to its widespread use, however, we clarify that it refers to the loss of water to the atmosphere via evaporation and sublimation including canopy interception effects and blowing snow sublimation (Mcmahon *et al* 2013). The theoretical maximum rate of vaporization from a saturated surface is often referred to as potential evaporation or evapotranspiration (*PET*), which is a function of both the available energy (primarily R_n, see also Eq. (2-5) and Eq. (2-6)) and the atmospheric water vapor pressure deficit (Meira-Neto *et al* 2020). Without limitations on available water, *ET* would be expected to equal *PET*. During the snow season, the Eq. (2-7) assumes that groundwater inflows and outflows from neighboring catchments are minimal although this assumption may be problematic in upland catchments over longer periods (Fan 2019). Interactions between variables in Eq. (2-7) lead to a distinct hydrograph in snowmelt-dominated systems typified by relative predictability in the timing and distribution of annual *Q* volume (Pagano *et al* 2004). Given the importance of stationarity for water supply prediction (Milly *et al* 2008), a diversity of metrics have been developed to better characterize the timing, distribution, and volume of *Q*. We detail several of these *Q* metrics below, including their derivation and significance for a discussion of the snowmelt-driven hydrograph and evidence for potential changes in *Q*.

2.3.1 Streamflow terminology and measurement

Studies assessing the connection between snow and streamflow timing have typically relied on the day of center of mass timing (DOQ_{50}) (Stewart *et al* 2004, McCabe and Clark 2005, Regonda *et al* 2005, Hidalgo *et al* 2009) with some studies including day of 25% (DOQ_{25})

and 75% of mass (DOQ_{75}) (Morán-Tejeda *et al* 2014). More recently, Krogh *et al* (2021) proposed the 20th percentile of snowmelt days to predict DOQ_{25} and DOQ_{50}. Other work has assessed changes in Q distribution throughout the water year via the seasonal or monthly fraction of Q (Aguado *et al* 1992, Dettinger and Cayan 1995, Stewart *et al* 2005), change in annual low Q (Godsey *et al* 2014), baseflow indices (Beck *et al* 2013), floods (Wenger *et al* 2010, Davenport *et al* 2020), maximum annual flows (Berghuijs *et al* 2016), and extreme runoff days (Li *et al* 2019). Mean annual Q (Hammond *et al* 2018, Berghuijs *et al* 2014, Stewart *et al* 2004, Barnhart *et al* 2016), runoff ratio or Q efficiency (e.g., the ratio of Q to P) (McCabe *et al* 2018, Li *et al* 2017), and Budyko Q anomaly—which quantify the difference between estimated and modeled Q based on a Budyko-type curve (Barnhart *et al* 2016, Berghuijs *et al* 2014, Ni *et al* 2015) — have also been used in mountainous environments. We summarize these metrics in Table 2-2.

Table 2-2: Overview of key streamflow dynamic metrics considered in this paper. In the table, we present units of measurement in terms of t = time and l = length.

Term	Unit	Definition	Citation
STREAMFLOW METRICS			
Mean annual runoff ratio	[$l\ l^{-1}$]	Water Year: The dimensionless ratio of streamflow to precipitation.	(Wenger *et al* 2010)
10-year flood	[l]	Water Year: Calculated by finding the largest flood for each year. The 90th percentile of annual maximum series defines the daily flow that occurs every 10 years on average.	(Wenger *et al* 2010)
25-year flood	[l]	Calculated by finding the largest flood for each year. The 96th percentile of annual maximum series defines the daily flow that occurs every 25 years on average.	(Wenger *et al* 2010)

Mean annual DOQ_{25}; DOQ_{50}; DOQ_{75}	[t]	Water Year: The day of the water year when cumulative discharge is at 25%, 50%, and 75% of its annual value.	(Wenger *et al* 2010)
Mean annual baseflow index	[l l^{-1}]	Water Year: The ratio of the lowest 7-day flow of summer (May 1 – September 30) to mean annual streamflow.	(Wenger *et al* 2010)
Mean annual Budyko streamflow anomaly	[l l^{-1}]	Water Year: The difference between estimated (1-*ET*/*P*) and modeled (1-f(*PET*/*P*)) streamflow using a Budyko-type equation.	(Berghuijs *et al* 2014, Ni *et al* 2015)
Mean annual streamflow volume	[l^3]	Water Year: Mean of yearly cumulative discharge	(Wenger *et al* 2010)
Seasonal or monthly fractional streamflow	[l^3 l^{-3}]	Seasonal: Fraction of annual streamflow occurring during a defined season (e.g., snow season or cool season, warm season or active growing season, winter, summer, etc.). Monthly: Fraction of annual streamflow occurring during a specific month of the year.	(Stewart *et al* 2005)

2.3.2 Changes in the snowmelt-dominated hydrograph

Climate change is altering the annual hydrograph in mountain environments with impacts to the timing, distribution and amount of streamflow (Stewart 2009, Lettenmaier and Gan 1990, Knowles and Cayan 2002, Gleick 1987, Hidalgo *et al* 2009, Rauscher *et al* 2008) with profound implications for downstream communities and environments (Westerling *et al* 2006, Viviroli *et al* 2007, Mankin *et al* 2015). Although there is strong evidence that the effects on different aspects of *Q* are linked (Aguado *et al* 1992, Fritze *et al* 2011, Nash and Gleick 1991, Dettinger and Cayan 1995), consensus about changes can also depend on the metric evaluated (Figure 2-1). There are relatively consistent findings about spring runoff or peak hydrograph timing (Figure 2-1A, 2-1D) and fraction of streamflow occurring during the snow season (Figure 2-1A, 2-1C) and the warm season (e.g., active growing season, Figure 2-1A). In contrast, there is less consensus with regard to changes in the annual volume of *Q* (Figure 2-1A, 2-1B) (McCabe *et al* 2018, Berghuijs *et al* 2014, Ni *et al* 2015, Stewart *et al* 2005, Nash and Gleick 1991, Das *et al* 2009, Jefferson 2011). As such, we treat these metrics separately in our review.

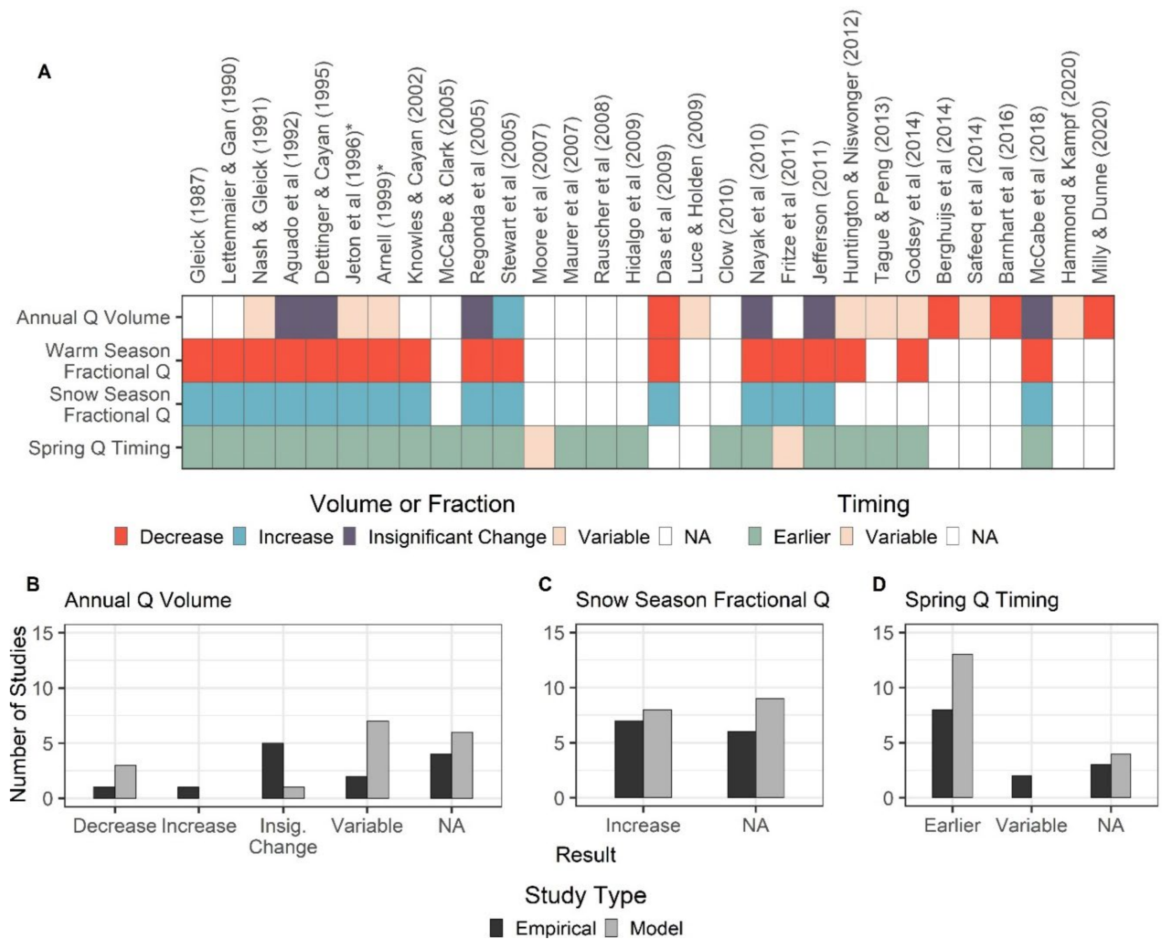

Figure 2-1: A) Synthesis of a subset of findings about Q response to climate change impacts on snow from literature reviewed in Section 2.3.3 and 2.3.4; B) Summary of findings from Figure 2-1A regarding changes in the mean annual volume of Q; C) Same as B), but for seasonal fraction of Q during the snow season, as an indicator of changes in the annual

distribution of Q; and, D) Summary of spring Q timing. We note that spring timing is measured in a multitude of ways (e.g., runoff timing, melt-out, peak Q, DOQ25) in the studies reviewed. We elected to use spring Q or runoff timing as an umbrella term to reflect the different ways in which changes in Q timing is measured. * Denotes a study that presented variable results, but where some results outside the scope of this review and thus categorized as earlier. In the case of Jeton *et al* (1996), we excluded their cooler scenario results, Arnell (1999) conducted a global analysis and we included only results for western North America. We also note that some results presented evidence for stronger trends in certain regions, as is the case for earlier spring Q or runoff timing at mid-elevation basins. In these cases, we followed the authors description of their results in categorizing them as earlier versus variable to the best of our ability.

2.3.2.1 Sensitivity of spring streamflow timing to changing snow dynamics

Strong evidence supports a trend towards earlier spring streamflow resulting from climate-induced changes in snow across the CONUS (Figure 2-1, McCabe and Clark 2005, Stewart 2009, Stewart *et al* 2005, Krogh *et al*, 2021). Initial studies highlighting western North America connected advances in *Q* timing to warmer winter and spring air T (Aguado *et al* 1992, Burn 1994), with follow on studies emphasizing the particular sensitivity of mid to lower elevation basins where air T was close to 0° C (Dettinger and Cayan 1995, McCabe and Clark 2005). Stewart *et al* (2005), for example, found that the timing of *Q* has shifted one to four weeks earlier in western North America in connection with widespread, monotonic increases in T, which contemporaneous studies (e.g., McCabe and Clark 2005, Regonda *et al* 2005) further connected to climate change. Knowles *et al* (2006) proposed that advances in spring *Q* timing were driven by decreases in both the amount and persistence of snow leading to earlier snowmelt. Research has also highlighted the potential for variability and/or statistically insignificant trends in the sensitivity of *Q* timing to climate change based on elevation in particular (Moore *et al* 2007, Fritze *et al* 2011, Stewart *et al* 2005, McCabe and Clark 2005).

Subsurface hydrological processes are typically invoked to explain catchment to regional scale streamflow timing sensitivity to changing snowpack. Tague and Grant (2009), Safeeq *et al* (2013), Maurer and Bowling (2014), and Harpold and Molotch (2015) all suggested that regional-scale subsurface hydrology provides a mechanistic explanation for the variable sensitivity of *Q* timing to climate change. Harpold and Molotch (2015), for example, emphasized that the timing of peak soil moisture can either exacerbate or moderate the sensitivity of *Q* to changes in snowmelt timing. Work in the Pacific Northwest highlights the role of bedrock properties and larger subsurface storage in moderating spring flows (Tague and Grant 2009, Safeeq *et al* 2013). Additional synthesis efforts have highlighted T and/or P, elevation, and atmospheric circulation variations to explain differences in *Q* timing sensitivity with mid to lower elevation basins exhibiting the greatest potential for earlier *Q* in western North America (Stewart 2009).

2.3.2.2 Sensitivity of annual streamflow volume to changing snow dynamics

There is clear evidence for the effects of changing snow on the distribution of annual streamflow (Dettinger and Cayan, 1995, Stewart *et al* 2005); however, the question of whether and how changes in snow will impact changes the annual volume of streamflow remains largely unsettled despite decades of research (Berghuijs *et al* 2014, McCabe *et al*

2018, Milly *et al* 2018, Ni *et al* 2015). Below, we outline previous contributions to these questions, focusing on changes in intra-annual *Q* distribution and changes in annual *Q* volume.

2.3.2.2.1 Snow effects on intra-annual streamflow distribution

Empirical work in the CONUS clearly connected advances in streamflow timing to an increase in the seasonal fraction of streamflow occurring during the snow season (Dettinger and Cayan, 1995, Stewart *et al* 2005). Dettinger and Cayan (1995) and Stewart *et al* (2005) found statistically significant increases in winter *Q* and decreases in warm season summer *Q*, which were echoed in smaller-scale analysis by Nayak *et al* (2010) in Idaho. Using hydrological modeling, Godsey *et al* (2014) later showed that simulated future changes in *fs* lead to a 10% decrease in the volume of warm season *Q* with evidence of considerable inter-site sensitivity to changes in *fs*. A hydrogeologic analysis by Safeeq *et al* (2014) suggested that future changes in *fs* might render areas with higher summer *Q* (greater subsurface storage) particularly vulnerable to climate change. On the whole, many of these lines of evidence about the intra-annual distribution of *Q* (Stewart *et al* 2005, Regonda *et al* 2005; Dettinger and Cayan, 1995) emphasized that they did not translate into statistically significant changes in the amount of interannual *Q*. Consistent with this conclusion, some later research reported changes in the seasonal fraction of *Q* without attendant changes in the magnitude of annual *Q* (Nayak *et al* 2010). The legacy of these findings is one line of evidence suggesting that changes the intra-annual distribution of *Q* without necessarily impacting the volume of annual *Q*.

2.3.2.2.2 Snow effects on annual streamflow volume

Multiple lines of evidence connecting changes in snow to changes in the annual volume of streamflow underscore conflicting results (Hammond and Kampf 2020, Berghuijs *et al* 2014, Barnhart *et al* 2016, Ni *et al* 2015). Some research has proposed that earlier streamflow timing and changes in the distribution of *Q* (e.g., increase in fractional *Q* during the snow season) increase the amount of runoff—and subsequently the annual volume of *Q*—assuming that the timing of energy inputs remains the same (Tague and Peng 2013). Jeton *et al* (1996), for example, used a process-based model to suggest that increased asynchrony between water and energy inputs may increase *Q* from high elevations basins and decrease *Q* in middle elevation basins. However, other research (Risbey and Entekhabi 1996, Dettinger *et al* 2004) found that advances in *Q* may also increase water-limitations on *ET*, thereby offsetting impacts on *Q*. Reflecting this uncertainty, other have recorded scattered trends in both the mean and median annual flow (Luce and Holden 2009, Stewart *et al* 2005).

More recent work by both Berghuijs *et al* (2014) and Ni *et al* (2015) marked a divergence from previous literature by hypothesizing that climate change driven declines in *fs* will lower streamflow efficiency. Specifically, Berghuijs *et al* (2014) supplemented an investigation of Budyko *Q* anomaly with more direct annual analysis to connect lower Budyko *Q* anomalies with lower *fs*. Parallel work on the role of increased *ET* in driving down *Q* (Milly and Dunne 2020, Goulden and Bales 2014)—particularly in higher elevation catchments with gentle slopes (Jepsen *et al* 2018) –offers some process explanation for this hypothesis.

However, physical explanations for the hypothesis that declines in snowfall will drive declines in streamflow remain elusive (Berghuijs *et al* 2014). Some research has posited that increases in spring *P* may also buffer *Q* against declines in *fs* (Pederson *et al* 2011) or that rainfall and mixed *P* inputs during the winter may countervail reductions in *Q* from declining *fs* (Hammond and Kampf, 2020) and model results suggest that higher snowmelt rates may have a larger effect on runoff ratios (Barnhart *et al* 2016). There is also recent evidence that efforts reliant on Budyko-based estimates of streamflow response to snow may be sensitive to poor assumptions about *PET* (Meira-Neto *et al* 2020). Complicating matters further, McCabe *et al* (2018) found that declines in *fs* have not altered runoff ratios in an empirical analysis of streamflow in the Pacific Northwest. These mixed findings on annual *Q* volume and runoff efficiency to changing snow conditions limit our ability to anticipate and respond to climate change.

2.4 Towards a framework linking snow processes and streamflow generation

Despite abundant research on changes in snowmelt-driven *Q* (Section 3), we lack a robust, consistent, and readily testable framework to explain varying *Q* response to climate change. Below, we present a conceptual framework that distills interactions between snow and the atmosphere, vegetation, and the subsurface into three inter-related mechanisms that can be tested using different snow (e.g., *fs* as in our demonstrations) and *Q* metrics (e.g., annual volume and runoff efficiency):

1. **Mechanism 1—Snow Season Water Vapor Fluxes**. Snow dynamics influence the available energy via Eq. (2-1 to 2-6), which influence the timing and amount of water vapor fluxes to the atmosphere during the snow season.

2. **Mechanism 2—Intensity of Liquid Water Inputs**. Because snow persists after it falls, snow dynamics can also modify the intensity of *LWI*, which in combination with site-specific thresholds (e.g., soil water holding capacity or infiltration capacity) can alter how water is partitioned to other water budget variables.
3. **Mechanism 3—Energy-Water Synchrony.** Snow enables the release of *LWI* after *P* has fallen. As such, snow dynamics facilitate greater temporal synchrony between *LWI* and *PET* during periods of higher radiation.

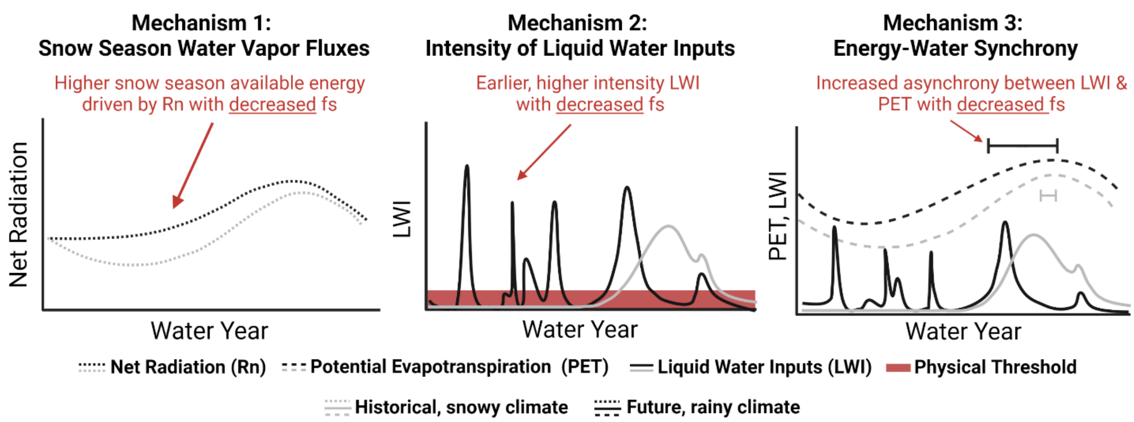

Figure 2-2: Conceptual figure outlining the three proposed mechanisms in this commentary. We emphasize that this is a diagram explaining each of these mechanisms individually and do not consider combined effects of different mechanisms. The threshold pictured for Mechanism 2 represents physical hydrological controls, which might include soil water holding capacity among other things. We acknowledge that future, rainy climate representations are speculative, particularly with respect to LWI intensity, which some research (Harpold and Kohler 2017, Godsey *et al* 2014) indicates may vary depending on environment.

We designed several data-driven demonstrations that rely on publicly available data from the Catchment Attributes for Large Sample studies (CAMELs) database (Addor *et al* 2017) to illustrate variable sensitivity to each mechanism across a range of study sites in the CONUS. Please see Text S2.1 for a full description of the data and Figure S2-1 for a map of study sites. Importantly, the intention of these experiments is not to quantitatively relate our mechanisms to *Q* or examine the effects of other snow dynamics, but rather to establish consistent mechanisms that can be further developed and tested in future empirical and process-based modeling work. As such, we focus explicitly on demonstrating the maximum potential sensitivity of each mechanism to a decline in *fs* to zero (e.g. an all rain future) because it is well-connected to both snow and *Q* metrics (see Section 2.4.1 below), reliably quantified without incorporating additional remotely-sensed data, and used to investigate *Q* response in several widely-cited studies (McCabe *et al* 2018, Berghuijs *et al* 2014). However, our framework leaves much room to be improved with additional snow and *Q* metrics to coordinate research on how and why *Q* response to changing snow dynamics varies. Additionally, our framework and demonstration are intended to establish a consistent set of mechanisms, thus we elect to treat each mechanism as distinct, but discuss the implications of interacting mechanisms in the conclusion.

2.4.1 Isolating snow metrics important for streamflow across the CONUS

Through correlation statistics, we assess relationships between snow metrics presented in Table 2-1 and *Q* metrics in Table 2-2 in 537 catchments to justify our focus on *fs*. Figure 2-3A illustrates that snow metrics are highly correlated to each other, with few metrics exhibiting Spearman correlation values below ~0.4. The mean annual *fs* captures a variety of snow dynamics similarly to snow persistence. Although each metric ultimately conveys different information about the characteristics of snow, both metrics exhibit the strongest and broadest correlation with other snow dynamics (Figure 2-3A) and *Q* characteristics (Figure 2-3B) of interest. As expected from past studies, relationships between the *fs* and *Q* volume metrics were weaker than timing metrics and *fs* negatively correlated to flood metrics (Davenport *et al* 2020). All of the metrics used in this evaluation with the exception of streamflow timing (e.g., $DOQ_{25,50,75}$), runoff ratio, baseflow index, and flood metrics rely heavily upon modeled data.

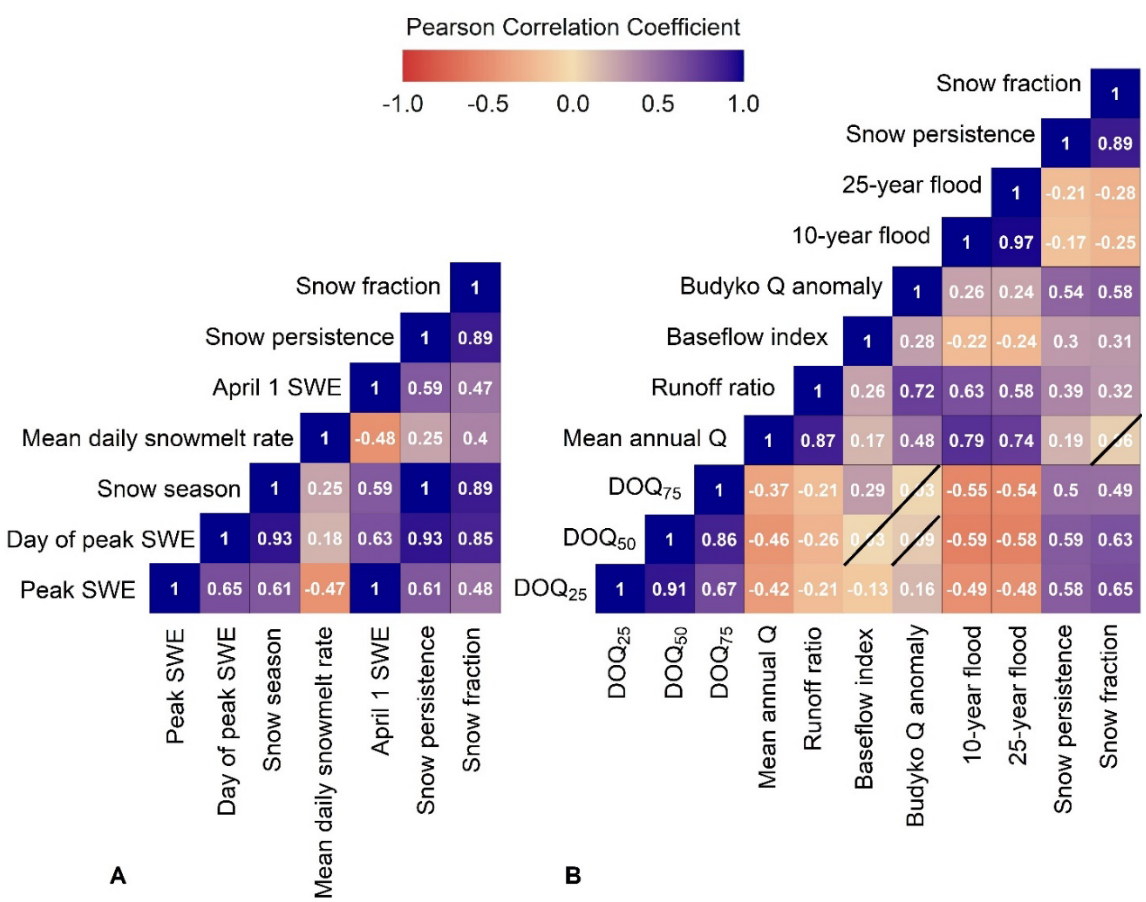

Figure 2-3: Correlograms of Pearson coefficients determined using streamflow data from the CAMELs database (please see Text S2.1) and NLDAS-2 (Xia *et al* 2012) forcing data

during the period 1980-2014 for: A) snow metrics defined in Table 2-1 and B) Q metrics defined in Table 2-2 as well as snow fraction and persistence. Black forward slash marks indicate statistically insignificant values (p> 0.05).

2.4.2 Mechanism 1: Changes in snow season water vapor fluxes

Snow dynamics influence abiotic interactions between the land surface and the atmosphere by exerting a first-order control over the amount of available energy that can drive water vapor fluxes (Mechanism 1). To investigate differences in the regional sensitivity to Mechanism 1, we performed several linear regressions between *fs* and $\mathrm{R_n}$ grouped by snow season *P* amount (Figure 2-4A to 2-4C). Grey bounds in Figure 2-4A to 2-4C reflect the 95% confidence interval for regressions. We refer to the reader to Figure S2-2 for scatterplots of the underlying data. Grouping by *P* accounts for the correlation between snow season *fs* and snow season length. Within snow season *P* groups, annual data were binned into nine groups of roughly equal number of catchments based on mean daily snow season solar radiation. We then used linear regression to identify the expected value of R_n when *fs*=0 (Figure 2-4D to 2-4F) and estimated the maximum potential change in R_n as the difference between modeled R_n using the mean historical snow season *fs* and the modeled R_n when snow season *fs* = 0. Assuming that the maximum potential change R_n was exclusively available to latent heat flux per Eq. (2-2), we then converted this value to a water flux (i.e., *ET* using latent heat of vaporization into mm d^{-1}) and normalized the resulting value by mean annual snow season *P*. We note that although the 95% confidence interval for our regressions is narrow in many cases (grey bounds in Figure 2-4A-C), our demonstration relies on a regression model (please see Figure S2-2 for more in-depth presentation of the underlying data).

Linear regressions support the moderating role of higher snow season *fs* on R_n in sunny, moderate to high *P* environments (orange and red lines in Figure 2-4B and 2-4C, Figure 2-4E to 2-4F, dark green circles). We summarize these linear regressions in Tables S2-1 to S2-3 Consistent with our linear regressions, we observe strong coherence between circle color (indicating higher potential changes in the ratio of *ET* to *P*) and circle size (indicating historically higher *fs*) in medium and high *P* environments (Figure 2-4E-F). Results suggest that areas with historically larger snowfall during the snow season have the highest potential sensitivity to increases in snow season water vapor losses. These dynamics are most important in the interior western and central CONUS with some evidence of sensitivity at higher latitudes in the northeastern CONUS.

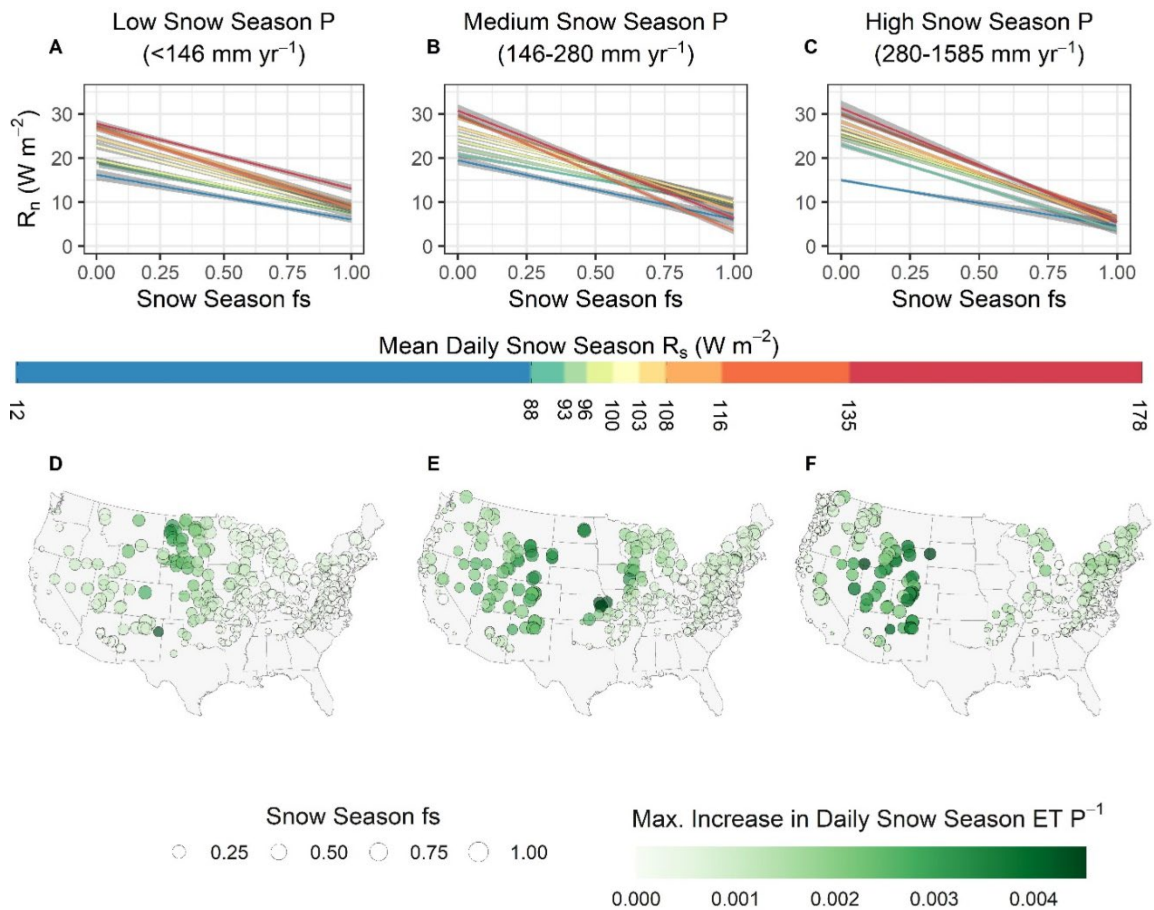

Figure 2-4: Potential for changes in snow season vapor fluxes (Mechanism 1) as illustrated by linear regression between snow season Rn and fs as illustrated using Daymet data from the CAMELs database (please see Text S2.1). A-C: Grouped site-year regression based on average daily snow season incoming shortwave (R_s) for low, medium, and high P environments (n ~5000 per low, medium, and high P). Grey bounds represent the 95% confidence interval for binned regressions. D-F: Map of the maximum potential daily increase in snow season ET is calculated between average historical fs and fs = 0, then normalized by snow season P.

2.4.3 Mechanism 2: Changes in *LWI* intensity

In certain environments, the unique characteristic seasonal of snow to persist after falling can facilitate the release of *LWI* later in time and at a lower or higher intensity than incoming *P*. Potential for changes in *LWI* together with physical hydrological thresholds (e.g., soil water holding capacity) exerts a first-order control on infiltration and runoff generation. We investigated the maximum potential change in peak *LWI* intensity (mm d^{-1}) under a complete transition from snowfall to rainfall (*fs* = 0) in Figure 2-5. We selected the annual peak intensity of *LWI* and *P* inputs for each catchment for a running 1-day, 3-day, and 14-day mean value. Windows were selected to capture different potential effects of *LWI* intensity on streamflow generation. Assuming stationarity in the intensity of *P*, we then estimated the maximum potential change in the intensity of *LWI* as the difference between the intensity of snowmelt-driven *LWI* and *P* inputs when *fs*=0 (no snow storage). Increases in rainfall intensity due to higher saturated vapor pressure with rising *T* (i.e., 7% increase per 1°C of warming) (Trenberth 2011) or melt during rain-on-snow events (Li *et al* 2019) could further intensify *LWI* beyond what we consider here.

Our results indicate that maximum potential *LWI* intensity at 1, 3, and 14-day durations will increase from historical snowmelt values as snowfall turns to rain, especially in catchments with higher historical *fs* (size of the grey circles in Figure 2-5A-C). This effect is particularly apparent for the 1 and 3-day *LWI* intensities (Figure 2-5A and 2-5B), with substantially lower shifts in the 14-day *LWI* intensities after shifts to rainfall (Figure 2-5C). Intuitively, in places with historically low *fs* there is already relatively little difference in *LWI* and *P* intensity, which is reflected in the clustering of small grey circles at x = 0 in Figure 2-5A-C. The map in Figure 2-5D-F highlights broad regional trends in sensitivity at the different temporal scales. The intensity of *LWI* in catchments at higher latitude and in the interior western CONUS appears most sensitive an increase as *fs* declines. However, the relationship between annual *fs* and maximum shift in *LWI* intensity does not fully explain the regional patterning in Figure 2-5D-F. For example, catchments along the central CONUS-Canadian border and some catchments along the western slope of the Appalachian Mountains also show large differences between *LWI* and *P* intensity. This suggests that other factors, such as intense spring rain, might explain or modulate the sensitivity to changes in *LWI* intensity.

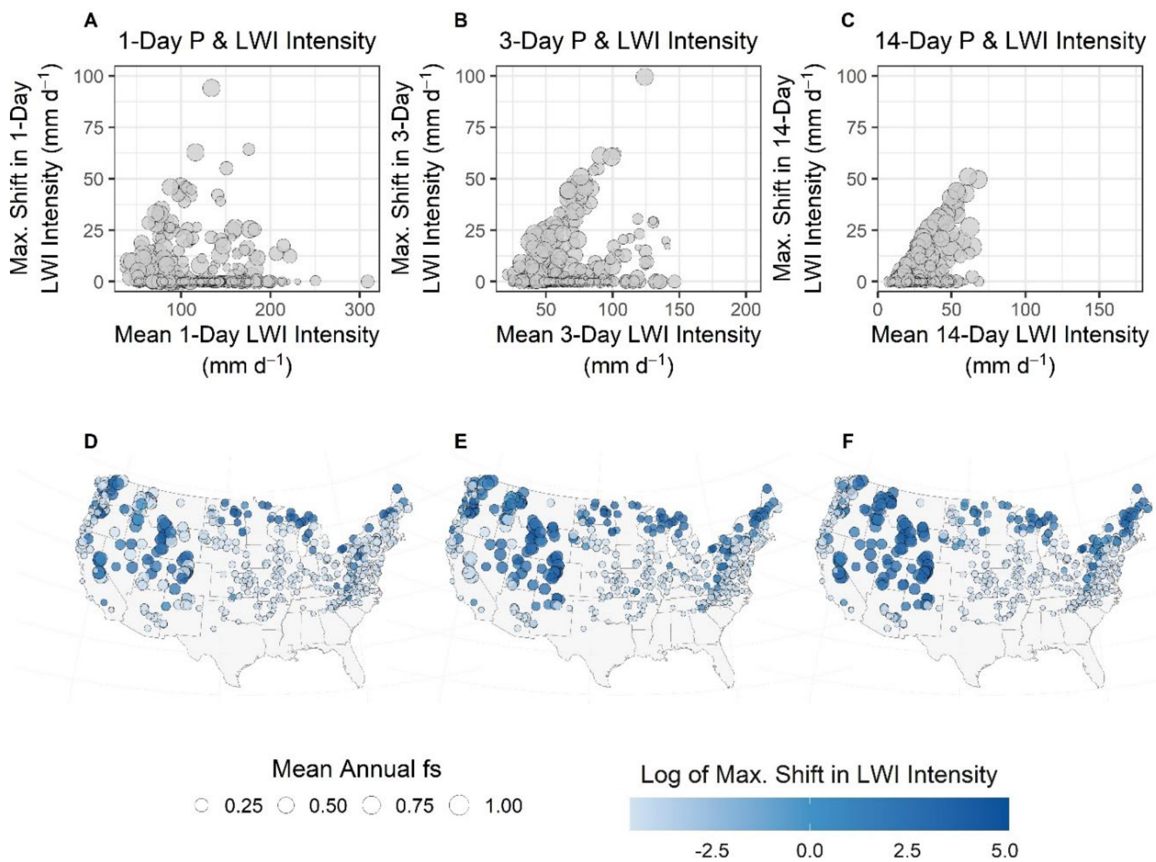

Figure 2-5: Potential for changes in LWI intensity (Mechanism 2) as illustrated using Daymet data from the CAMELs database (please see Text S2.1). Here we calculate the potential shift in LWI intensity (mm d-1) over 1-day, 3-day, and 14-day intervals as the difference between LWI and P. A-C: Scatterplots of the maximum LWI intensity for each site versus the shift in LWI intensity (mm d-1) for 1 day, 3 day, and 14 day intervals. D-F: Map of the maximum potential shift in LWI based on an fs = 0 for 1 day, 3 day, and 14 day intervals.

2.4.4 Mechanism 3: Water-energy synchrony

Snowpacks provide temporary storage of higher winter *P* that is released as *LWI* later in the season, which is a first-order control on *Q* via the partitioning of stored water to *ET* versus *Q* generation or groundwater. We simulate the maximum potential difference in the timing of *LWI* and *P* inputs under a complete transition from snow to rain (*fs* = 0). For each catchment, we calculated the mean DoWY in which the catchment achieved 25% and 50% of *LWI* and *P* inputs across all years. Using these data, we then approximated the maximum shift in the timing of *LWI* inputs as the difference between 25% or 50% *LWI* and *P* inputs, respectively (i.e., assuming that the timing of *LWI* would equal the timing of *P* inputs if *fs* = 0 and there is no snowmelt to modulate the timing of *LWI* inputs). Consistent with Tague and Peng (2013) and our own analysis of the *PET* timing variability, we assumed that changes in the timing of *LWI* was the largest driver of potential asynchrony between *LWI* and energy inputs.

We observed that catchments with a higher annual *fs* also experienced greater maximum shifts between 25% and 50% *LWI* and *P* inputs as indicated by increase in circle size along the y-axis in Figure 2-6A and 2-6B. This relationship between *fs* and the maximum potential shift in the timing of *LWI* is evident across catchments regardless of the DoWY they reach 25% or 50% of their annual *LWI* (e.g., x-axis in Figure 2-6A-B), although it does appear to scale the magnitude of the shift between *LWI* and *P* inputs. Intuitively, catchments with less annual snowfall experience relative harmony between *LWI* and *P* inputs. The translation of these results to geographic locations in Figure 2-6C-D highlights distinct regional trends in sensitivity, with largest potential shifts in the timing of 25% and 50% *LWI* in the interior western CONUS (dark purple circles). In both the 25% (early streamflow generation) and 50% *LWI* (peak streamflow generation) cases, the maximum risk for potential shifts in *LWI* timing were well connected to mean annual *fs*, as demonstrated by the coherence between circle size and shading in Figure2-6C-D.

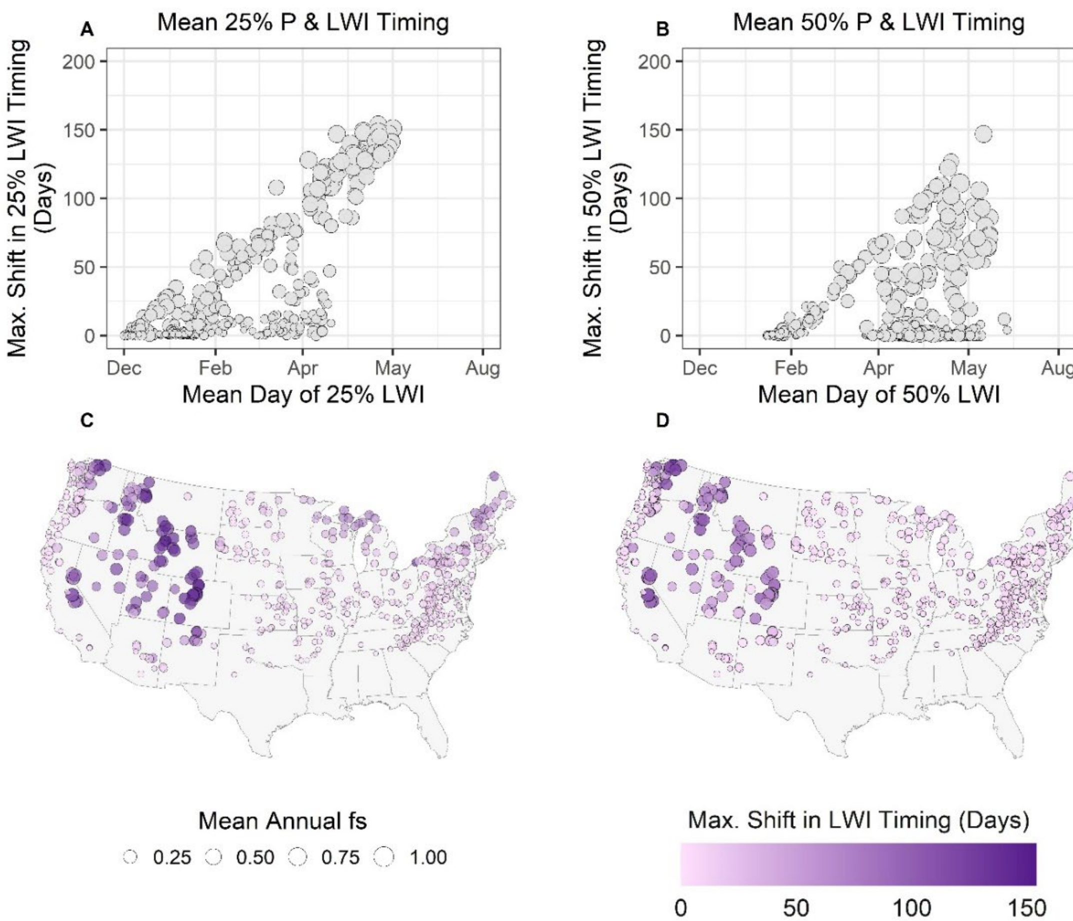

Figure 2-6: Potential for changes in the synchrony of water-energy inputs (Mechanism 1) as illustrated using Daymet data in the CAMELS database (please see Text S2.1). A-B: Mean annual historical DoWY when 25% and 50% of LWI occur versus maximum shift in days between 25% and 50% of LWI timing and P timing. C-D: Map of the potential maximum shift in the timing of 25% and 50% of LWI when fs = 0 (e.g., all LWI driven by rainfall).

2.5 Summary and conclusions

Our data-driven review focuses on how regional variability in climate differentially influences the partitioning of winter precipitation along a gradient of fractional snow cover within the CONUS. We identify three inter-dependent mechanisms based on how snowpack mass and energy balance interacts with local climate, infiltration, and catchment water storage. Our framework leads to testable hypotheses useful for evaluating regional variability in streamflow response under a warmer climate.

- Mechanism 1 addresses how a warmer and drier climate will impact snow season water vapor fluxes. Although often ignored, both snowpack sublimation and evaporation losses can be large components of the annual water budget, which are likely to increase in a warming climate. These losses are likely to be greatest in the Great Basin, Missouri, Upper and Lower Colorado, Souris-Red-Rainey, and Arkansas White-Red basins (Figure 2-7), with the potential to exacerbate summer drought stress and reduce annual *Q* consistent with Milly and Dunne (2020) who used a physically-based model in the Upper Colorado River Basin to estimate a 9.3% decrease in *Q* per degree Celsius of warming because of increased *ET* due primarily to snow albedo feedbacks.
- Mechanism 2 addresses how increased *LWI* intensity driven by snow season rainfall (e.g., *fs* = 0) will interact with physical hydrological controls on infiltration and routing. The maximum potential risk for more intense *LWI* are greatest in the Pacific Northwest, California, Great Basin, Upper Colorado, and Missouri Basins in the western CONUS as well as the Great Lakes and New England Basins in the eastern CONUS (Figure 2-7). These changes (Figure 2-5) would be expected to increase the amount of *Q* occurring during the snow season consistent with Davenport *et al* (2020), who showed that declines in *fs* led to proportionally larger

increases in streamflow and advance the timing of spring *Q*. As such, changes could potentially exacerbate summer drought stress per Harpold and Molotch (2015) with variable impacts on annual *Q* volume.

- Mechanism 3 addresses the role of subsurface storage in buffering earlier water inputs (energy-water synchrony) in a warmer, rainier climate. The maximum potential risk for decreased temporal synchrony between water and energy inputs is greatest in the Pacific Northwest, California, Great Basin, Upper Colorado, Rio Grande, and Missouri Basins in western CONUS (Figure 2-7). However, the extent to which temporal asynchrony may or may not impact seasonal and annual *Q* volume or drought stress remains difficult to parse consistent with Jeton *et al* (1996) who showed that higher and lower elevation basins experience bi-directional changes in annual *Q* volume in a warmer climate.

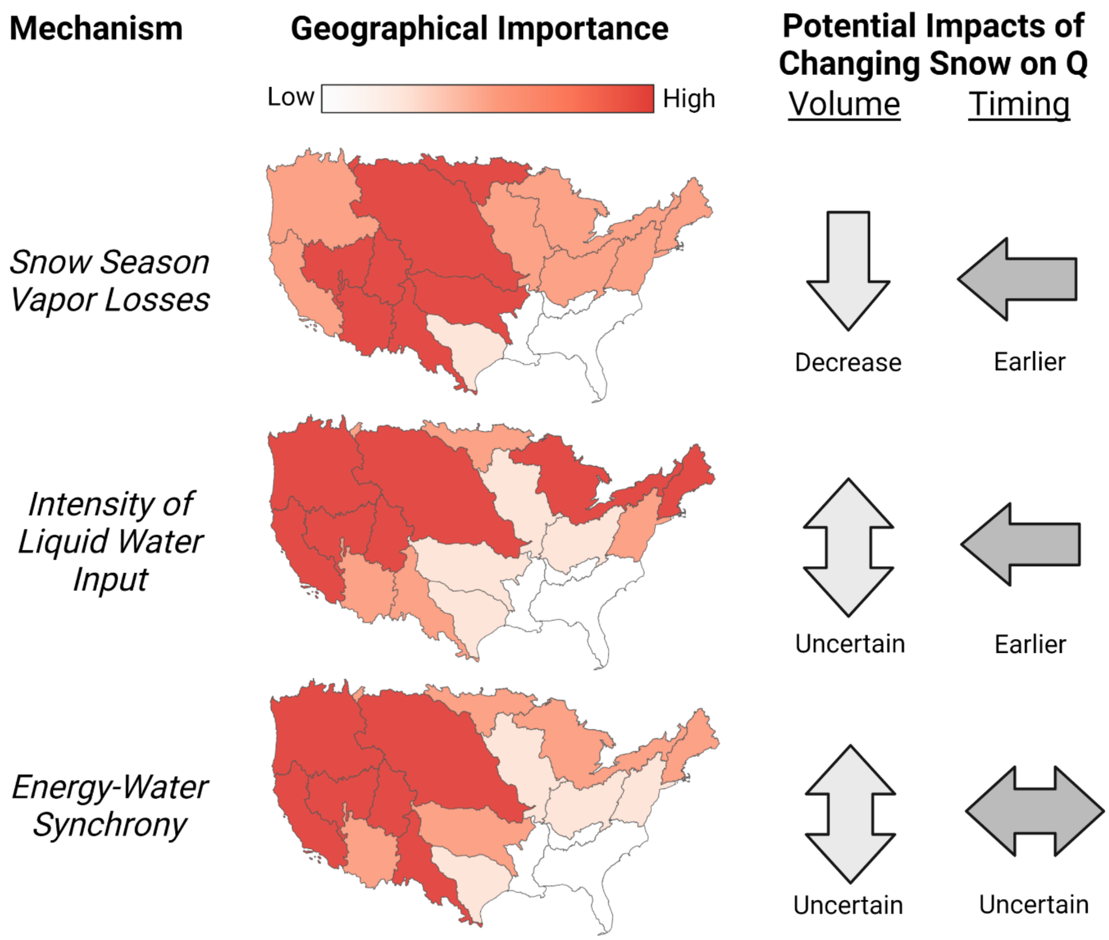

Figure 2-7: Summary graphic highlighting the potential utility of the framework proposed as part of this data-driven review. Here, we use both experimental results and literature review with experience of the authorship team to highlight where each Mechanism—and the set of processes it represents—may be most important. Hydrologic regions correspond to United States Geological Survey (USGS) HUC (Hydrologic Unit Code) 2 boundaries. We show the study sites and HUC 2 boundaries in Figure S2-1.

Our review provides a consistent framework for assessing the impacts of climate change on snowmelt-derived streamflow, highlighting differential risks across regions of the western U.S. There are, however, limitations in our initial demonstration worth considering. First, using one snow metric (*fs*) as a proxy for climate change may not capture all the potential nuances in each of our mechanisms, especially at smaller spatial scales. Related to this, there is also a clear opportunity for future research to evaluate how the mechanisms we identify influence streamflow generation using the metrics aggregated for this review. Finally, although mechanisms may interact, an in-depth investigation of these interactions was beyond the scope of our demonstration. However, our initial results can help to characterize end-members for future research by establishing a set of testable, inter-related mechanisms that reflect the dominant processes connecting snow to streamflow across CONUS. Snowmelt-driven streamflow in areas with either less persistent snow cover, small historical fs, and with low intensity *P* during the snow season (e.g., the HUCS in the northeastern CONUS in Figure 2-7) may be easier to predict than in areas with persistent snow cover, large historical *fs*, and higher intensity *P* during the snow season (e.g., HUCs in the interior western CONUS in Figure 2-7), where snowmelt-driven streamflow is likely to have unique feedbacks not completely captured in our framework.

Ultimately, hydrologic models need to capture potential interactions of these three mechanisms to accurately predict future changes in snowmelt-driven streamflow. Specifically, improvements in modeling capabilities should be focused in three areas: 1) representing complexities in snowpack-atmosphere energy fluxes and how this variability influences snow ablation; 2) representing variability in subsurface storage as soil and groundwater and how these stores are partitioned to either atmosphere fluxes or streamflow; and 3) continuing to develop high quality forcing datasets, including P, T, R_n, humidity, and wind speed, in complex terrain. Progress on these fronts requires advances in hydrologic models, process understanding, conceptual frameworks, and observations.

As a critical link in this chain, our mechanistic framework offers value for evaluating and communicating changes in critical mountain water supplies in an increasingly complex and uncertain hydrologic future.

2.6 References

Addor N, Newman A J, Mizukami N and Clark M P 2017 The CAMELS data set: Catchment attributes and meteorology for large-sample studies Hydrol. Earth Syst. Sci. 21 5293–313

Aguado E, Cayan D and Roos M 1992 Climate fluctuations and the timing of west coast streamflow J. Clim. 5

Allan J D and Castillo M M 2007 Stream Ecology (Springer Netherlands)

Ambach W 1974 The Influence of Cloudiness on the Net Radiation Balance of a Snow Surface with High Albedo J. Glaciol. 13 73–84

Arnell N W 1999 Climate change and global water resources Global Environmental Change vol 9 (Elsevier Ltd) pp S31–49

Bales R C, Molotch N P, Painter T H, Dettinger M D, Rice R and Dozier J 2006 Mountain hydrology of the western United States Water Resour. Res. 42 8432

Barnett T P, Adam J C and Lettenmaier D P 2005 Potential impacts of a warming climate on water availability in snow-dominated regions Nature 438 303–9

Barnhart T B, Molotch N P, Livneh B, Harpold A A, Knowles J F and Schneider D 2016 Snowmelt rate dictates streamflow Geophys. Res. Lett. 43 8006–16

Barthold F K and Woods R A 2015 Stormflow generation: A meta-analysis of field evidence from small, forested catchments Water Resour. Res. 51 3730–53

Bayard D, Stähli M, Parriaux A and Flühler H 2005 The influence of seasonally frozen soil on the snowmelt runoff at two Alpine sites in southern Switzerland J. Hydrol. 309 66–84

Beck H E, Van Dijk A I J M, Miralles D G, De Jeu R A M, Bruijnzeel L A, McVicar T R and Schellekens J 2013 Global patterns in base flow index and recession based on streamflow observations from 3394 catchments Water Resour. Res. 49 7843–63

Berghuijs W R, Woods R A and Hrachowitz M 2014 A precipitation shift from snow towards rain leads to a decrease in streamflow Nat. Clim. Chang. 4 583–6

Berghuijs W R, Woods R A, Hutton C J and Sivapalan M 2016 Dominant flood generating mechanisms across the United States Geophys. Res. Lett. 43 4382–90

Biederman J A, Somor A J, Harpold A A, Gutmann E D, Breshears D D, Troch P A, Gochis

D J, Scott R L, Meddens A J H and Brooks P D 2015 Recent tree die-off has little effect on streamflow in contrast to expected increases from historical studies Water Resour. Res. 51 9775–89

Blöschl G, Bierkens M F P, Chambel A, Cudennec C, Destouni G, Fiori A, Kirchner J W, McDonnell J J, Savenije H H G, Sivapalan M, Stumpp C, Toth E, Volpi E, Carr G, Lupton C, Salinas J, Széles B, Viglione A, Aksoy H, Allen S T, Amin A, Andréassian V, Arheimer B, Aryal S K, Baker V, Bardsley E, Barendrecht M H, Bartosova A, Batelaan O, Berghuijs W R, Beven K, Blume T, Bogaard T, Borges de Amorim P, Böttcher M E, Boulet G, Breinl K, Brilly M, Brocca L, Buytaert W, Castellarin A, Castelletti A, Chen X, Chen Y, Chen Y, Chifflard P, Claps P, Clark M P, Collins A L, Croke B, Dathe A, David P C, de Barros F P J, de Rooij G, Di Baldassarre G, Driscoll J M, Duethmann D, Dwivedi R, Eris E, Farmer W H, Feiccabrino J, Ferguson G, Ferrari E, Ferraris S, Fersch B, Finger D, Foglia L, Fowler K, Gartsman B, Gascoin S, Gaume E, Gelfan A, Geris J, Gharari S, Gleeson T, Glendell M, Gonzalez Bevacqua A, González-Dugo M P, Grimaldi S, Gupta A B, Guse B, Han D, Hannah D, Harpold A, Haun S, Heal K, Helfricht K, Herrnegger M, Hipsey M, Hlaváčiková H, Hohmann C, Holko L, Hopkinson C, Hrachowitz M, Illangasekare T H, Inam A, Innocente C, Istanbulluoglu E, *et al* 2019 Twenty-three unsolved problems in hydrology (UPH)–a community perspective Hydrol. Sci. J. 64 1141–58

Bohr G S and Aguado E 2001 Use of April 1 *SWE* measurements as estimates of peak seasonal snowpack and total cold-season precipitation Water Resour. Res. 37 51–60

Bormann K J, Brown R D, Derksen C and Painter T H 2018 Estimating snow-cover trends from space Nat. Clim. Chang. 2018 811 8 924–8

Bowling D R, Logan B A, Hufkens K, Aubrecht D M, Richardson A D, Burns S P, Anderegg W R L, Blanken P D and Eiriksson D P 2018 Limitations to winter and spring photosynthesis of a Rocky Mountain subalpine forest Agric. For. Meteorol. 252 241–55

Brooks P D, Chorover J, Fan Y, Godsey S E, Maxwell R M, McNamara J P and Tague C 2015 Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics Water Resour. Res. 51 6973–87

Brooks P D, Gelderloos A, Wolf M A, Jamison L R, Strong C, Solomon D K, Bowen G J, Burian S, Tai X, Arens S, Briefer L, Kirkham T and Stewart J 2021 Groundwater-Mediated Memory of Past Climate Controls Water Yield in Snowmelt-Dominated Catchments Water Resour. Res. 57

Broxton P D, Dawson N and Zeng X 2016 Linking snowfall and snow accumulation to generate spatial maps of *SWE* and snow depth Earth Sp. Sci. 3 246–56

Broxton P D, Harpold A A, Biederman J A, Troch P A, Molotch N P and Brooks P D 2015 Quantifying the effects of vegetation structure on snow accumulation and ablation in mixed-conifer forests Ecohydrology 8 1073–94

Bulygina O N, Razuvaev V N and Korshunova N N 2009 Changes in snow cover over Northern Eurasia in the last few decades Environ. Res. Lett. 4 045026

Burakowski E A, Wake C P, Bra*SWE*ll B, Brown D P, Burakowski C :, Wake C P, Bra*SWE*ll B and Brown D P 2008 Trends in wintertime climate in the northeastern United States: 1965–2005 J. Geophys. Res. Atmos. 113 20114

Burn D H 1994 Hydrologic effects of climatic change in west-central Canada vol 4

Cayan D R 1996 Interannual climate variability and snowpack in the western United States J. Clim. 9 928–48

Changnon D, McKee T B and Doesken N J 1991 HYDROCLIMATIC VARIABILITY IN THE ROCKY MOUNTAINS J. Am. Water Resour. Assoc. 27 733–43

Christensen L, Adams H R, Tai X, Barnard H R and Brooks P D 2021 Increasing plant water stress and decreasing summer streamflow in response to a warmer and wetter climate in seasonally snow-covered forests Ecohydrology 14

Church J E 1933 Snow Surveying: Its Principles and Possibilities vol 23

Cline D W 1997 Snow surface energy exchanges and snowmelt at a continental, midlatitude Alpine site Water Resour. Res. 33 689–701

Clow D W 2010 Changes in the timing of snowmelt and streamflow in Colorado: A response to recent warming J. Clim. 23 2293–306

Das T, Hidalgo H G, Dettinger M D, Cayan D R, Pierce D W, Bonfils C, Barnett T P, Bala G and Mirin A 2009 Structure and detectability of trends in hydrological measures over the western United States J. Hydrometeorol. 10 871–92

Davenport F V., Herrera-Estrada J E, Burke M and Diffenbaugh N S 2020 Flood Size Increases Nonlinearly Across the Western United States in Response to Lower Snow-Precipitation Ratios Water Resour. Res. 56 e2019WR025571

Day T A, Delucia E H and Smith W K 1989 Influence of cold soil and snowcover on photosynthesis and leaf conductance in two Rocky Mountain conifers Oecologia 80 546–52

Deems J S, Painter T H, Barsugli J J, Belnap J and Udall B 2013 Combined impacts of current and future dust deposition and regional warming on Colorado River Basin snow dynamics and hydrology Hydrol. Earth Syst. Sci. 17 4401–13

Déry S J and Brown R D 2007 Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback Geophys. Res. Lett. 34 L22504

Dettinger M D and Cayan D R 1995 Large-scale atmospheric forcing of recent trends toward early snowmelt runoff in California J. Clim 8 606–23

Dettinger M D, Cayan D R, Meyer M K and Jeton A 2004 Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900-2099 * Clim. Change 62 283–317

Dong J, Walker J P and Houser P R 2005 Factors affecting remotely sensed snow water equivalent uncertainty Remote Sens. Environ. 97 68–82

Dunne T 1978 Field studies of hillslope flow processes Hillslope Hydrology ed M J Kirby (John Wiley and Sons Inc.,New York, NY) pp 227–93

Dunne T and Black R D 1971 Runoff Processes during Snowmelt Water Resour. Res. 7 1160–72

Earman S, Campbell A R, Phillips F M and Newman B D 2006 Isotopic exchange between snow and atmospheric water vapor: Estimation of the snowmelt component of groundwater recharge in the southwestern United States J. Geophys. Res. 111 D09302

Ehsani N, Vörösmarty C J, Fekete B M and Stakhiv E Z 2017 Reservoir operations under climate change: Storage capacity options to mitigate risk J. Hydrol. 555 435–46

Essery R L H, Pomeroy J W, Parviainen J and Storck P 2003 Sublimation of Snow from Coniferous Forests in a Climate Model J. Clim. 16 1855–64

Fan Y 2019 Are catchments leaky? WIREs Water 6

Fletcher C G, Zhao H, Kushner P J and Fernandes R 2012 Using models and satellite observations to evaluate the strength of snow albedo feedback J. Geophys. Res. Atmos. 117 11117

Frisbee M D, Phillips F M, Weissmann G S, Brooks P D, Wilson J L, Campbell A R and Liu F 2012 Unraveling the mysteries of the large watershed black box: Implications for the streamflow response to climate and landscape perturbations Geophys. Res. Lett. 39 1–6

Fritze H, Stewart I T and Pebesma E 2011 Shifts in western North American snowmelt runoff regimes for the recent warm decades J. Hydrometeorol. 12 989–1006

Girotto M, Cortés G, Margulis S A and Durand M 2014 Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada Water Resour. Res. 50 6713–34

Gleick P H 1987 Regional hydrologic consequences of increases in atmospheric CO2 and other trace gases Clim. Change 10 137–160

Godsey S E, Kirchner J W and Tague C L 2014 Effects of changes in winter snowpacks on summer low flows: case studies in the Sierra Nevada, California, USA Hydrol. Process. 28 5048–64

Goulden M L and Bales R C 2014 Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion Proc. Natl. Acad. Sci. U. S. A. 111 14071–5

Groisman P Y and Legates D R 1994 The accuracy of United States precipitation data Bull.

- Am. Meteorol. Soc. 75 215–27

Gustafson J R, Brooks P D, Molotch N P and Veatch W C 2010 Estimating snow sublimation using natural chemical and isotopic tracers across a gradient of solar radiation Water Resour. Res. 46 12511

Hall A 2004 The role of surface albedo feedback in climate J. Clim. 17 1550–68

Hall A, Qu X and Neelin J D 2008 Improving predictions of summer climate change in the United States Geophys. Res. Lett. 35 1702

Hamlet A F and Lettenmaier D P 2007 Effects of 20th century warming and climate variability on flood risk in the western U.S. Water Resour. Res. 43 6427

Hamlet A F, Mote P W, Clark M P and Lettenmaier D P 2005 Effects of temperature and precipitation variability on snowpack trends in the Western United States J. Clim. 18 4545–61

Hammond J C, Harpold A A, Weiss S and Kampf S K 2019 Partitioning snowmelt and rainfall in the critical zone: effects of climate type and soil properties Hydrol. Earth Syst. Sci 23 3553–70

Hammond J C and Kampf S K 2020 Subannual Streamflow Responses to Rainfall and Snowmelt Inputs in Snow-Dominated Watersheds of the Western United States Water Resour. Res. 56

Hammond J C, Saavedra F A and Kampf S K 2018 How Does Snow Persistence Relate to Annual Streamflow in Mountain Watersheds of the Western U.S. With Wet Maritime and Dry Continental Climates? Water Resour. Res. 54 2605–23

Harding R J and Pomeroy J W 1996 The Energy Balance of the Winter Boreal Landscape J. Clim. 9 2778–87

Harpold A A and Brooks P D 2018 Humidity determines snowpack ablation under a warming climate Proc. Natl. Acad. Sci. U. S. A. 115 1215–20

Harpold A A and Kohler M 2017 Potential for Changing Extreme Snowmelt and Rainfall Events in the Mountains of the Western United States J. Geophys. Res. Atmos. 122 13,219-13,228

Harpold A A and Molotch N P 2015 Sensitivity of soil water availability to changing snowmelt timing in the western U.S. Geophys. Res. Lett. 42 8011–20

Harpold A, Brooks P, Rajagopal S, Heidbuchel I, Jardine A and Stielstra C 2012 Changes in snowpack accumulation and ablation in the intermountain west Water Resour. Res. 48

Harr R D 1981 Some characteristics and consequences of snowmelt during rainfall in western Oregon J. Hydrol. 53 277–304

Hidalgo H G, Das T, Dettinger M D, Cayan D R, Pierce D W, Barnett T P, Bala G, Mirin

A, Wood A W, Bonfils C, Santer B D and Nozawa T 2009 Detection and attribution of streamflow timing changes to climate change in the Western United States J. Clim. 22 3838–55

Holden Z A, Luce C H, Crimmins M A and Morgan P 2012 Wildfire extent and severity correlated with annual streamflow distribution and timing in the Pacific Northwest, USA (1984-2005) Ecohydrology 5 677–84

Hood E, Williams M and Cline D 1999 Sublimation from a seasonal snowpack at a continental, mid-latitude alpine site Hydrol. Process. 13 1781–97

Horton R E 1933 The role of infiltration in the hydrologic cycle Eos, Trans. Am. Geophys. Union 14 446–60

Huang C, Newman A J, Clark M P, Wood A W and Zheng X 2017 Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States Hydrol. Earth Syst. Sci. 21 635–50

Huxman T E, Turnipseed A A, Sparks J P, Harley P C and Monson R K 2003 Temperature as a control over ecosystem CO2 fluxes in a high-elevation, subalpine forest Oecologia 2003 1344 134 537–46

Immerzeel W W, Lutz A F, Andrade M, Bahl A, Biemans H, Bolch T, Hyde S, Brumby S, Davies B J, Elmore A C, Emmer A, Feng M, Fernández A, Haritashya U, Kargel J S, Koppes M, Kraaijenbrink P D A, Kulkarni A V., Mayewski P A, Nepal S, Pacheco P, Painter T H, Pellicciotti F, Rajaram H, Rupper S, Sinisalo A, Shrestha A B, Viviroli D, Wada Y, Xiao C, Yao T and Baillie J E M 2020 Importance and vulnerability of the world's water towers Nature 577 364–9

Ingram W J, Wilson C A and Mitchell J F B 1989 Modeling climate change: an assessment of sea ice and surface albedo feedbacks J. Geophys. Res. 94 8609–22

Jambon-Puillet E, Shahidzadeh N and Bonn D 2018 Singular sublimation of ice and snow crystals Nat. Commun. 2018 91 9 1–6

Jefferson A J 2011 Seasonal versus transient snow and the elevation dependence of climate sensitivity in maritime mountainous regions Geophys. Res. Lett. 38

Jepsen S M, Harmon T C, Ficklin D L, Molotch N P and Guan B 2018 Evapotranspiration sensitivity to air temperature across a snow-influenced watershed: Space-for-time substitution versus integrated watershed modeling J. Hydrol. 556 645–59

Jeton A E, Dettinger M D and Smith J L 1996 Potential effects of cliamte change on streamflow,easter and western slopes of the Sierra Nevada, California and Nevada (Vol., 95, No. 4260)

JianCheng S, Chuan X and LingMei J 2016 SCIENCE CHINA Review of snow water equivalent microwave remote sensing 59 731–45

Karl T R, Groisman P Y, Knight R W and Heim R R 1993 Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations J. Clim. 6 1327–44

Klos P Z, Link T E and Abatzoglou J T 2014 Extent of the rain-snow transition zone in the western U.S. under historic and projected climate Geophys. Res. Lett. 41 4560–8

Knowles N and Cayan D R 2002 Potential effects of global warming on the Sacramento/San Joaquin watershed and the San Francisco estuary Geophys. Res. Lett. 29 1891

Knowles N, Dettinger M D and Cayan D R 2006 Trends in snowfall versus rainfall in the western United States J. Clim. 19 4545–59

Kormos P R, Marks D, McNamara J P, Marshall H P, Winstral A and Flores A N 2014 Snow distribution, melt and surface water inputs to the soil in the mountain rain-snow transition zone J. Hydrol. 519 190–204

Koster R D, Mahanama S P P, Yamada T J, Balsamo G, Berg A A, Boisserie M, Dirmeyer P A, Doblas-Reyes F J, Drewitt G, Gordon C T, Guo Z, Jeong J H, Lawrence D M, Lee W S, Li Z, Luo L, Malyshev S, Merryfield W J, Seneviratne S I, Stanelle T, Van Den Hurk B J J M, Vitart F and Wood E F 2010 Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment Geophys. Res. Lett. 37

Krogh S A, Scaff L, Sterle G, Kirchner J W, Gordon B and Harpold A 2021 Diel streamflow cycles suggest more sensitive snowmelt-driven streamflow to climate change than land surface modeling

Kumar S V., Jasinski M, Mocko D M, Rodell M, Borak J, Li B, Beaudoing H K and Peters-Lidard C D 2019 NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment J. Hydrometeorol. 20 1571–93

Lee, X., Massman, W., & Law, B. (Eds.). (2004). Handbook of micrometeorology: a guide for surface flux measurement and analysis (Vol. 29). Springer Science & Business Media.

Leroux N R and Pomeroy J W 2017 Modelling capillary hysteresis effects on preferential flow through melting and cold layered snowpacks Adv. Water Resour. 107 250–64

Lettenmaier D P and Gan T Y 1990 Hydrologic sensitivities of the Sacramento-San Joaquin River Basin, California, to global warming Water Resour. Res. 26 69–86

Li D, Lettenmaier D P, Margulis S A and Andreadis K 2019 The Role of Rain-on-Snow in Flooding Over the Conterminous United States Water Resour. Res. 55 8492–513

Li D, Wrzesien M L, Durand M, Adam J and Lettenmaier D P 2017 How much runoff originates as snow in the western United States, and how will that change in the future? Geophys. Res. Lett. 44 6163–72

Liu F, Parmenter R, Brooks P D, Conklin M H and Bales R C 2008 Seasonal and interannual variation of streamflow pathways and biogeochemical implications in semi-arid, forested catchments in Valles Caldera, New Mexico Ecohydrology 1 239–52

Luce C H and Holden Z A 2009 Declining annual streamflow distributions in the Pacific Northwest United States, 1948–2006 Geophys. Res. Lett. 36 L16401

Lundberg A and Halldin S 2001 Snow interception evaporation. Review of measurement techniques, processes, and models vol 70 (Springer)

Lundquist J D, Neiman P J, Martner B, White A B, Gottas D J and Ralph F M 2008 Rain versus Snow in the Sierra Nevada, California: Comparing Doppler Profiling Radar and Surface Observations of Melting Level J. Hydrometeorol. 9 194–211

Lundquist J, Dickerson-Lange S, Gutmann E, Jonas T, Lumbrazo C and Reynolds D 2021 Snow Interception Modeling: Isolated Observations have led to Land Surface Models Lacking Appropriate Climate Sensitivities Authorea Prepr.

Lundquist J, Hughes M, Gutmann E and Kapnick S 2019 Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks Bull. Am. Meteorol. Soc. 100 2473–90

Maggioni V, Sapiano M R P and Adler R F 2016 Estimating uncertainties in high-resolution satellite precipitation products: Systematic or Random Error? J. Hydrometeorol. 17 1119–29

Mankin J S, Viviroli D, Singh D, Hoekstra A Y and Diffenbaugh N S 2015 The potential for snow to supply human water demand in the present and future Environ. Res. Lett. 10 114016

Markovich K H, Manning A H, Condon L E and McIntosh J C 2019 Mountain-Block Recharge: A Review of Current Understanding Water Resour. Res. 55 8278–304

Marks D and Dozier J 1992 Climate and energy exchange at the snow surface in the Alpine Region of the Sierra Nevada: 2. Snow cover energy balance Water Resour. Res. 28 3043–54

Marsh P and Woo M -K 1984 Wetting front advance and freezing of meltwater within a snow cover: 2. A simulation model Water Resour. Res. 20 1865–74

Massman, W. J., Sommerfeld, R. A., Mosier, A. R., Zeller, K. F., Hehn, T. J., & Rochelle, S. G. (1997). A model investigation of turbulence-driven pressure-pumping effects on the rate of diffusion of CO_2, N_2O, and CH_4 through layered snowpacks. Journal of Geophysical Research: Atmospheres, 102(D15), 18851-18863.

Maurer G E and Bowling D R 2014 Seasonal snowpack characteristics influence soil temperature and water content at multiple scales in interior western U.S. mountain ecosystems Water Resour. Res. 50 5216–34

McCabe G J and Clark M P 2005 Trends and variability in snowmelt runoff in the western United States J. Hydrometeorol. 6 476–82

McCabe G J, Wolock D M and Valentin M 2018 Warming is driving decreases in snow fractions while runoff efficiency remains mostly unchanged in snow-covered areas of the western United States J. Hydrometeorol. 19 803–14

McDonnell J J, Spence C, Karran D J, Meerveld H J (Ilja) van and Harman C J 2021 Fill-and-Spill: A Process Description of Runoff Generation at the Scale of the Beholder Water Resour. Res. 57 e2020WR027514

Mcmahon T A, Peel M C, Lowe L, Srikanthan R and Mcvicar T R 2013 Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis Hydrol. Earth Syst. Sci 17 1331–63

Meerveld H J T and McDonnell J J 2006 Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis Water Resour. Res. 42 2411

Meira-Neto A A, Niu G, Roy T, Tyler S and Troch P A 2020 Interactions between snow cover and evaporation lead to higher sensitivity of streamflow to temperature Commun. Earth Environ. 1 1–7

Meixner T, Manning A H, Stonestrom D A, Allen D M, Ajami H, Blasch K W, Brookfield A E, Castro C L, Clark J F, Gochis D J, Flint A L, Neff K L, Niraula R, Rodell M, Scanlon B R, Singha K and Walvoord M A 2016 Implications of projected climate change for groundwater recharge in the western United States J. Hydrol. 534 124–38

Meybeck M, Green P and Vörösmarty C 2001 A New Typology for Mountains and Other Relief Classes Mt. Res. Dev. 21 34–45

Milly P C ., Betancourt J, Falkenmark M, Hirsch R M, Kundzewicz Z W, Lettenmaier D P and Stouffer R J 2008 Climate change. Stationarity is dead: whither water management? Science (80-.). 319 573–4

Milly P C D, Betancourt J, Falkenmark M, Hirsch R M, Kundzewicz Z W, Lettenmaier D P, Stouffer R J, Dettinger M D and Krysanova V 2015 On Critiques of "Stationarity is Dead: Whither Water Management?" Water Resour. Res. 51 7785–9

Milly P C D and Dunne K A 2020 Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation Science (80-.). 367 1252–5

Milly P C D, Kam J and Dunne K A 2018 On the Sensitivity of Annual Streamflow to Air Temperature Water Resour. Res. 54 2624–41

Miralles D G, Brutsaert W, Dolman A J and Gash J H 2020 On the Use of the Term "Evapotranspiration" Water Resour. Res. 56 e2020WR028055

Molotch N P, Brooks P D, Burns S P, Litvak M, Monson R K, McConnell J R and Musselman K 2009 Ecohydrological controls on snowmelt partitioning in mixed-conifer sub-alpine forests Ecohydrology 2 129–42

Montanari A and Koutsoyiannis D 2014 Modeling and mitigating natural hazards: Stationarity is immortal! Water Resour. Res. 50 9748–56

Moore J N, Harper J T and Greenwood M C 2007 Significance of trends toward earlier snowmelt runoff, Columbia and Missouri Basin headwaters, western United States Geophys. Res. Lett. 34

Morán-Tejeda E, Lorenzo-Lacruz J, López-Moreno J I, Rahman K and Beniston M 2014 Streamflow timing of mountain rivers in Spain: Recent changes and future projections J. Hydrol. 517 1114–27

Mote P W 2003 Trends in snow water equivalent in the Pacific Northwest and their climatic causes Geophys. Res. Lett. 30

Mote P W, Parson E A, Hamlet A F, Keeton W S, Lettenmaier D, Mantua N, Miles E L, Peterson D W, Peterson D L, Slaughter R and Snover A K 2003 Preparing for climatic change: The water, salmon, and forests of the Pacific Northwest Clim. Change 61 45–88

Musselman K N, Clark M P, Liu C, Ikeda K and Rasmussen R 2017 Slower snowmelt in a warmer world Nat. Clim. Chang. 7 214–9

Nash L L and Gleick P H 1991 Sensitivity of streamflow in the Colorado Basin to climatic changes J. Hydrol. 125 221–41

Nayak A, Marks D, Chandler D G and Seyfried M 2010 Long-term snow, climate, and streamflow trends at the reynolds creek experimental watershed, Owyhee Mountains, Idaho, United States Water Resour. Res. 46

Newman A J, Clark M P, Sampson K, Wood A, Hay L E, Bock A, Viger R J, Blodgett D, Brekke L, Arnold J R, Hopson T and Duan Q 2015 Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance Hydrol. Earth Syst. Sci 19 209–23

Ni G, Yang D, Zhang D, Cong Z, Ni G, Yang D and Hu S 2015 Effects of snow ratio on annual runoff within the Budyko framework Hydrol. Earth Syst. Sci 19

Pagano T, Garen D and Sorooshian S 2004 Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922–2002 J. Hydrometeorol. 5 896–909

Painter T H, Berisford D F, Boardman J W, Bormann K J, Deems J S, Gehrke F, Hedrick A, Joyce M, Laidlaw R, Marks D, Mattmann C, McGurk B, Ramirez P, Richardson M, Skiles S M K, Seidel F C and Winstral A 2016 The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo Remote Sens. Environ. 184 139–52

Pederson G T, Gray S T, Ault T, Marsh W, Fagre D B, Bunn A G, Woodhouse C A and Graumlich L J 2011 Climatic controls on the snowmelt hydrology of the northern Rocky Mountains J. Clim. 24 1666–87

Petersky R and Harpold A 2018 Now you see it, now you don’t: A case study of ephemeral snowpacks and soil moisture response in the Great Basin, USA Hydrol. Earth Syst.

Sci. 22 4891–906

Plüss C and Ohmura A 1997 Longwave radiation on snow-covered mountainous surfaces J. Appl. Meteorol. 36 818–24

Qin Y, Abatzoglou J T, Siebert S, Huning L S, AghaKouchak A, Mankin J S, Hong C, Tong D, Davis S J and Mueller N D 2020 Agricultural risks from changing snowmelt Nat. Clim. Chang. 10 459–65

Qu X and Hall A 2014 On the persistent spread in snow-albedo feedback Clim. Dyn. 42 69–81

Rasmussen R, Baker B, Kochendorfer J, Meyers T, Landolt S, Fischer A P, Black J, Thériault J M, Kucera P, Gochis D, Smith C, Nitu R, Hall M, Ikeda K and Gutmann E 2012 How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed Bull. Am. Meteorol. Soc. 93 811–29

Rauscher S A, Pal J S, Diffenbaugh N S and Benedetti M M 2008 Future changes in snowmelt-driven runoff timing over the western US Geophys. Res. Lett. 35

Reba M L, Link T E, Marks D and Pomeroy J 2009 An assessment of corrections for eddy covariance measured turbulent fluxes over snow in mountain environments Water Resour. Res. 46 0–38

Regonda S K, Rajagopalan B, Clark M and Pitlick J 2005 Seasonal cycle shifts in hydroclimatology over the western United States J. Clim. 18 372–84

Rinehart A J, Vivoni E R and Brooks P D 2008 Effects of vegetation, albedo, and solar radiation sheltering on the distribution of snow in the Valles Caldera, New Mexico Ecohydrology 1 253–70

Risbey J S and Entekhabi D 1996 Observed Sacramento Basin streamflow response to precipitation and temperature changes and its relevance to climate impact studies J. Hydrol. 184 209–23

Rupp D E, Mote P W, Bindoff N L, Stott P A and Robinson D A 2013 Detection and attribution of observed changes in northern hemisphere spring snow cover J. Clim. 26 6904–14

Safeeq M, Grant G E, Lewis S L, Kramer M G and Staab B 2014 A hydrogeologic framework for characterizing summer streamflow sensitivity to climate warming in the Pacific Northwest, USA Hydrol. Earth Syst. Sci. 18 3693–710

Safeeq M, Grant G E, Lewis S L and Tague C L 2013 Coupling snowpack and groundwater dynamics to interpret historical streamflow trends in the western United States Hydrol. Process. 27 655–68

Schlögl S, Lehning M and Mott R 2018 How are turbulent sensible heat fluxes and snow melt rates affected by a changing snow cover fraction? Front. Earth Sci. 6 154

Schneider S H and Dickinson R E 1974 Climate modeling Rev. Geophys. 12 447–93

Sexstone G A, Clow D W, Fassnacht S R, Liston G E, Hiemstra C A, Knowles J F and Penn C A 2018 Snow Sublimation in Mountain Environments and Its Sensitivity to Forest Disturbance and Climate Warming Water Resour. Res. 54 1191–211

Seyfried M S, Grant L E, Marks D, Winstral A and McNamara J 2009 Simulated soil water storage effects on streamflow generation in a mountainous snowmelt environment, Idaho, USA Hydrol. Process. 23 858–73

Shanley J B and Chalmers A 1999 The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont { Hydrol. Process. 13 1843–57

Siirila-Woodburn E R, Rhoades A M, Hatchett B J, Huning L S, Szinai J, Tague C, Nico P S, Feldman D R, Jones A D, Collins W D and Kaatz L 2021 A low-to-no snow future and its impacts on water resources in the western United States Nat. Rev. Earth Environ. 2021 211 2 800–19

Skiles S M, Painter T H, Deems J S, Bryant A C and Landry C C 2012 Dust radiative forcing in snow of the Upper Colorado River Basin: 2. Interannual variability in radiative forcing and snowmelt rates Water Resour. Res. 48

Skofronick-Jackson, G., Kirschbaum, D., Petersen, W., Huffman, G., Kidd, C., Stocker, E., & Kakar, R. (2018). The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. Quarterly Journal of the Royal Meteorological Society, 144, 27-48.

Stewart I T 2009 Changes in snowpack and snowmelt runoff for key mountain regions Hydrol. Process. 23 78–94

Stewart I T, Cayan D R and Dettinger M D 2004 Changes in snowmelt runoff timing in western North America under a “business as usual” climate change scenario Clim. Change 62 217–32

Stewart I T, Cayan D R and Dettinger M D 2005 Changes toward earlier streamflow timing across western North America J. Clim. 18 1136–55

Stigter E E, Litt M, Steiner J F, Bonekamp P N J, Shea J M, Bierkens M F P and Immerzeel W W 2018 The Importance of Snow Sublimation on a Himalayan Glacier Front. Earth Sci. 6 1–16

Storck P, Lettenmaier D P and Bolton S M 2002 Measurement of snow interception and canopy effects on snow accumulation and melt in a mountainous maritime climate, Oregon, United States Water Resour. Res. 38 5-1-5–16

Sturm M, Goldstein M A and Parr C 2017 Water and life from snow: A trillion dollar science question Water Resour. Res. 53 3534–44

Tague C and Grant G E 2009 Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions Water Resour. Res. 45 7421

Tague C and Peng H 2013 The sensitivity of forest water use to the timing of precipitation and snowmelt recharge in the California Sierra: Implications for a warming climate J. Geophys. Res. Biogeosciences 118 875–87

Tedesco M, Derksen C, Deems J S and Foster J L 2014 Remote sensing of snow depth and snow water equivalent Remote Sens. Cryosph. 73–98

Tennant C J, Harpold A A, Lohse K A, Godsey S E, Crosby B T, Larsen L G, Brooks P D, Kirk R W Van and Glenn N F 2017 Regional sensitivities of seasonal snowpack to elevation, aspect, and vegetation cover in western North America Water Resour. Res. 53 6908–26

Thackeray C W and Fletcher C G 2016 Snow albedo feedback: Current knowledge, importance, outstanding issues and future directions Prog. Phys. Geogr. 40 392–408

Trenberth K E 2011 Changes in precipitation with climate change Clim. Res. 47 123–38

Trujillo E and Molotch N P 2014 Snowpack regimes of the Western United States Water Resour. Res. 50 5611–23

Veatch W, Brooks P D, Gustafson J R and Molotch N P 2009 'Quantifying the effects of forest canopy cover on net snow accumulation at a continental, mid-latitude site' Ecohydrology 2 115–28

Viviroli D, Dürr H H, Messerli B, Meybeck M and Weingartner R 2007 Mountains of the world, water towers for humanity: Typology, mapping, and global significance Water Resour. Res. 43 7447

Viviroli D, Kummu M, Meybeck M, Kallio M and Wada Y 2020 Increasing dependence of lowland populations on mountain water resources Nat. Sustain. 3 917–28

Wenger S J, Luce C H, Hamlet A F, Isaak D J and Neville H M 2010 Macroscale hydrologic modeling of ecologically relevant flow metrics Water Resour. Res. 46 9513

Westerling A L, Hidalgo H G, Cayan D R and *SWE*tnam T W 2006 Warming and earlier spring increase Western U.S. forest wildfire activity Science (80-.). 313 940–3

Wobus C, Small E E, Hosterman H, Mills D, Stein J, Rissing M, Jones R, Duckworth M, Hall R, Kolian M, Creason J and Martinich J 2017 Projected climate change impacts on skiing and snowmobiling: A case study of the United States Glob. Environ. Chang. 45 1–14

Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, Luo L, Alonge C, Wei H, Meng J, Livneh B, Lettenmaier D, Koren V, Duan Q, Mo K, Fan Y and Mocko D 2012 Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products J. Geophys. Res. Atmos. 117 3109

2.7 Supplemental Information

Text S2.1

In Section 2.4, we explore the difference in regional sensitivity to Mechanisms 1-3 outlined above across the CONUS using data from 1980-2014 in 537 catchments within the Catchment Attributes and Meteorology for Large Sample Studies (CAMELS) database (Addor *et al* 2017). The database of unimpaired, gauged catchments includes modeled snowmelt data from the SNOW-17 and Sacramento Soil Moisture Accounting (SAC-SMA) hydrologic modeling system forced with Daymet (Thornton *et al* 1997) data and Phase 2 of the North American Land Data Assimilation System (NLDAS-2). We obtained combined LWI and precipitation information from the CAMELS database. Because the CAMELS database does not include R_n, we supplement it with the NCA-LDAS (Kumar *et al* 2019) model to assess changes in the snow season surface energy balance (Mechanism 1). For Mechanism 1, we evaluate snow season *fs* (see Table 2-1 in the main Chapter). For Mechanisms 2 and 3, we evaluate annual fs.

For each mechanism, we establish maximum sensitivity to changes in *fs* by assuming the complete transition from snowfall to rainfall (*fs* = 0), which is highly improbable in the near-term, particularly for colder, higher elevation sites (O'Gorman 2014). However, this approach establishes a theoretical upper-limit for potential changes in each mechanism, which can be used in future research to explore controls on the variability in *Q* outlined in Section 2.3 of the main Chapter.

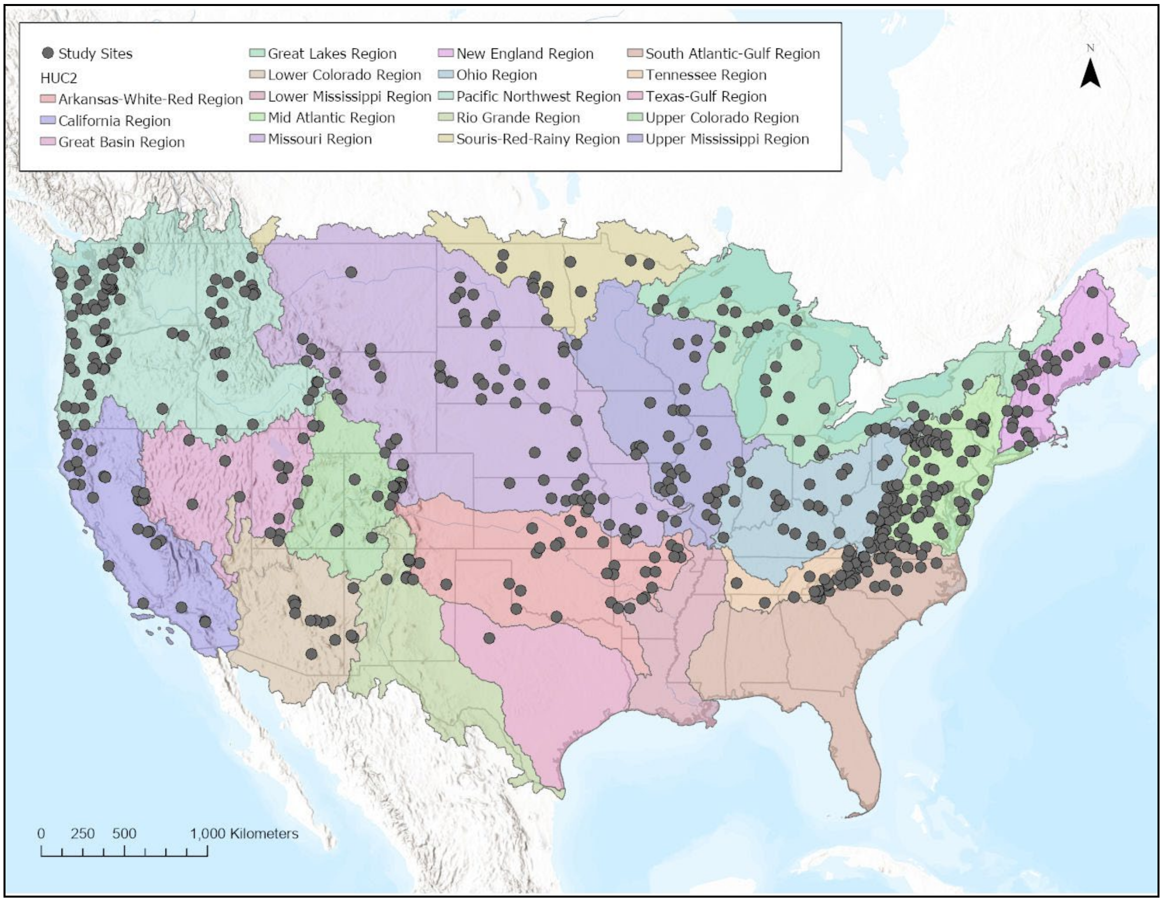

Figure S2-1: All study sites include in the demonstrations in the main text of the manuscript and an outline of the HUC2 Regions used to construct Figure 2-7.

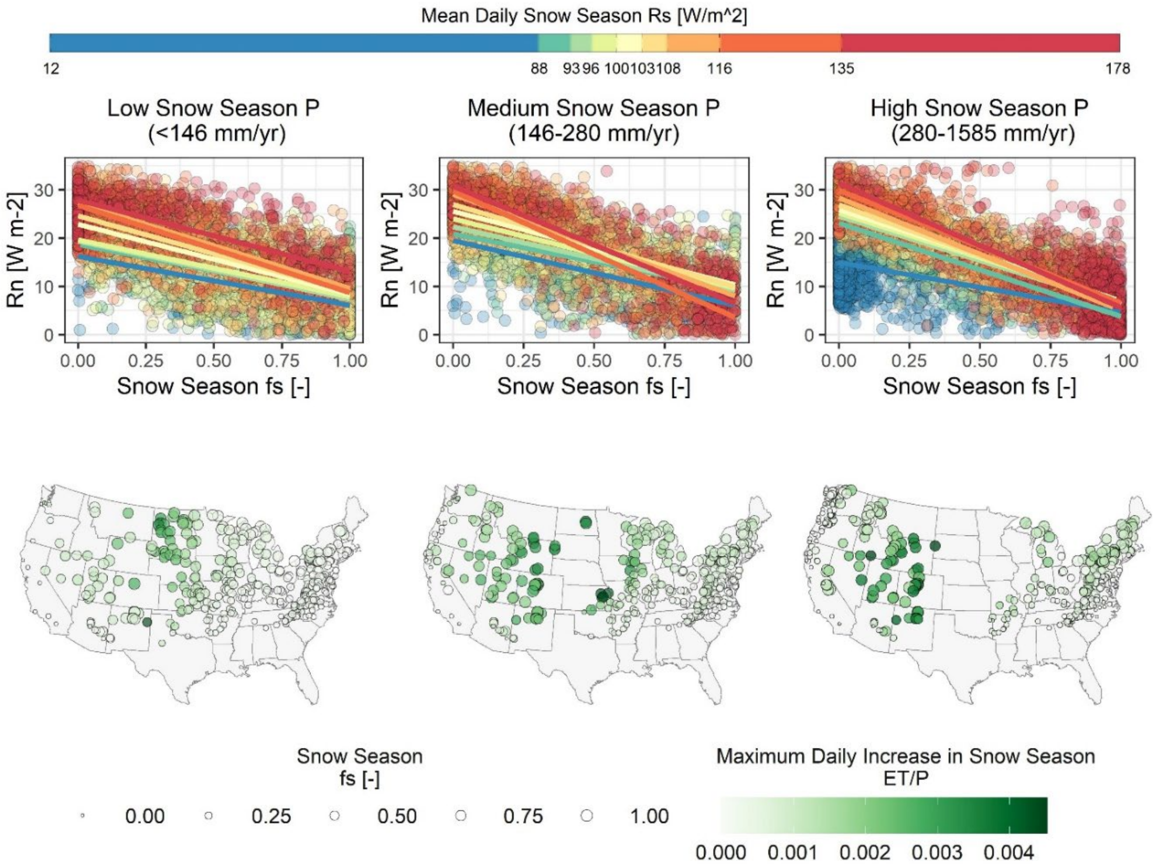

Figure S2-2: Linear regression between snow season *Rn* and *fs* as illustrated by Daymet data in the CAMELS database. A-C: Scatterplot of all points used to generate grouped site-year regression based on average daily snow season incoming shortwave (*Rs*) for low, medium, and high *P* environments (n = ~5000 per low, medium, and high *P*). Points are colored by average daily snow season *Rs*. D-F: Map of the maximum potential daily increase in snow season *ET* is based on an *fs* of 0 and normalized by snow season *P* for low, medium, and high *P* environments.

Table S2-1: Summary statistics for regressions between *fs* and *Rn* for the low snow season precipitation group in the main Chapter.

Low Snow Season P Regression Coefficients					
Mean Daily Snow Season Rs [W/m2]	Estimate	Statistic	p-value	Method	Alternative
12 to 88	-0.49551	22864735	2.56E-29	Spearman's rank correlation rho	two.sided
88 to 93	-0.58797	94325359	3.82E-67	Spearman's rank correlation rho	two.sided
93 to 96	-0.62385	1.2E+08	2.04E-83	Spearman's rank correlation rho	two.sided
96 to 100	-0.55945	2.96E+08	5.06E-87	Spearman's rank correlation rho	two.sided
100 to 103	-0.67774	41629287	1.59E-72	Spearman's rank correlation rho	two.sided
103 to 108	-0.69145	56438317	2.26E-84	Spearman's rank correlation rho	two.sided
108 to 116	-0.67532	89746230	2.22E-92	Spearman's rank correlation rho	two.sided
116 to 135	-0.67503	43959402	4.50E-73	Spearman's rank correlation rho	two.sided
135 to 178	-0.68266	20335309	1.65E-58	Spearman's rank correlation rho	two.sided

Table S2-2: Summary statistics for regressions between *fs* and *Rn* for the medium snow season precipitation group in the main manuscript.

Medium Snow Season P Regression Coefficients					
Mean Daily Snow Season Rs [W/m2]	**Estimate**	**Statistic**	**p-value**	**Method**	**Alternative**
12 to 88	-0.56489	10432945	3.20E-30	Spearman's rank correlation rho	two.sided
88 to 93	-0.55346	1.14E+08	2.74E-62	Spearman's rank correlation rho	two.sided
93 to 96	-0.66373	63249119	7.63E-79	Spearman's rank correlation rho	two.sided
96 to 100	-0.65834	1.67E+08	3.16E-106	Spearman's rank correlation rho	two.sided
100 to 103	-0.66875	47289744	4.45E-73	Spearman's rank correlation rho	two.sided
103 to 108	-0.64303	1E+08	1.14E-84	Spearman's rank correlation rho	two.sided
108 to 116	-0.64974	1.96E+08	2.40E-108	Spearman's rank correlation rho	two.sided
116 to 135	-0.79295	62945684	1.04E-129	Spearman's rank correlation rho	two.sided

135 to 178	-0.72474	21297029	1.29E-69	Spearman's rank correlation rho	two.sided

Table S2-3: Summary statistics for regressions between fs and Rn for the high snow season precipitation group in the main manuscript.

High Snow Season P Regression Coefficients					
Mean Daily Snow Season Rs [W/m2]	**Estimate**	**Statistic**	**p-value**	**Method**	**Alternative**
12 to 88	-0.22674	3.64E+08	1.35E-15	Spearman's rank correlation rho	two.sided
88 to 93	-0.79556	24974147	1.04E-96	Spearman's rank correlation rho	two.sided
93 to 96	-0.80783	10535294	1.33E-76	Spearman's rank correlation rho	two.sided
96 to 100	-0.78577	37203378	5.64E-106	Spearman's rank correlation rho	two.sided
100 to 103	-0.77775	24388552	2.39E-89	Spearman's rank correlation rho	two.sided
103 to 108	-0.82843	68829951	6.11E-155	Spearman's rank correlation rho	two.sided
108 to 116	-0.84513	59382434	8.33E-159	Spearman's rank correlation rho	two.sided
116 to 135	-0.81631	1.09E+08	3.68E-171	Spearman's rank correlation rho	two.sided
135 to 178	-0.41783	3.06E+08	2.70E-47	Spearman's rank correlation rho	two.sided

3 Chapter 3: Can we use the water budget to infer upland catchment behavior? The role of dataset error estimation and interbasin groundwater flow

By: Beatrice L. Gordon, Wade T. Crow, Alexandra G. Konings, David N. Dralle, Adrian A. Harpold

Citation: Gordon, B. L., Crow, W. T., Konings, A. G., Dralle, D. N., & Harpold, A. A. Can we use the water budget to infer upland catchment behavior? The role of dataset error estimation and interbasin groundwater flow. Water Resources Research, e2021WR030966.

Abstract

Water budgets are essential for characterizing water supplies from snow-dominated upland catchments where data are sparse, groundwater systems are complex, and measurements are prone to error (ε). One solution is imposing water budget closure (CWB) by ignoring difficult-to-measure variables, including inter-basin groundwater fluxes (G) and ε. However, conventional CWB-based analyses, which derive evapotranspiration (ET) from precipitation (P) and streamflow (Q) (e.g., the Budyko hypothesis), are limited in their ability to take advantage of recent advances in ET products, physically-based frameworks for improving inferences about G, or tools to statistically characterize ε (Triple Collocation, TC); all of which offer promise for improved water supply predictions via open water budgets (OWB). We clarify the value of these advances in upland settings by comparing standard land surface model, Ensemble Mean, and TC-Merged P and ET products in 114 upland catchments. When compared against a long-term OWB, we find that the CWB assumptions are unsupported in 75-100% of our 114 catchments, depending on the product. We then show how applying these CWB assumptions in snowy, steep catchments where ε is large can inflate inferences about streamflow response to climate change by up to 9 times more than independent (OWB) estimates of ET using TC. Finally, we demonstrate how advances in OWB analysis reveal that high, arid settings with deep permeable substrate are

groundwater exporters while most other basins are groundwater importers. Our results highlight the advantages of OWB analyses that harness new products, tools, and frameworks for characterizing inter-basin groundwater fluxes in critical upland settings.

3.1 Introduction

Higher elevation (upland) catchments are critical for generating downstream water supplies (Barnett *et al*., 2008; Ehsani *et al*., 2017; Harpold *et al*., 2012; Li *et al*., 2017; Viviroli *et al*., 2007). As such, there is a pressing need to quantify how upland water supplies—including both surface and groundwater—will respond to changing climate (Gordon *et al*., 2022; Immerzeel *et al*., 2020; Mankin *et al*., 2015; Qin *et al*., 2020). Water budgets are foundational tools in this pursuit (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; Ni *et al*., 2015). However, accurately closing the water budget in upland settings remains elusive due to complex hydrologic pathways and difficult-to-measure variables such as groundwater fluxes (G) and evapotranspiration (ET). Moreover, measurement error—particularly biases (i.e., systematic error) in estimates of precipitation (P) associated with complex topography and snow plague water balance closure in upland catchments (Bales *et al*., 2006; Carroll *et al*., 2019; Henn *et al*., 2018). To circumvent these challenges, conventional approaches often assume that unknown or uncertain variables (including error) can be ignored by imposing a closed water budget (CWB, Fan, 2019; Safeeq *et al*., 2021, Kampf *et al*., 2020). To illustrate this point, we present a simple, but complete, catchment water budget following Fan (2019):

$$P - ET - Q = \frac{\Delta S}{\Delta t} + G + \varepsilon_P + \varepsilon_{ET} + \varepsilon_Q, \qquad \text{Eq. (3-1)}$$

where P is precipitation, ET is evapotranspiration, Q is streamflow, $\frac{\Delta S}{\Delta t}$ is change in terrestrial water storage, G is interbasin-groundwater flux into or out of the catchment, and ε_P, ε_{ET} , and ε_Q are combined systematic and random error in the measurement of *P, ET,* and *Q*, respectively. Here, we define systematic errors as correlated to the true value of the variable whereas random errors are uncorrelated to the true value of the variable. Because they are difficult to separate, we refer to the combination of ε_p, ε_{ET} , and ε_Q as a single term (ε) and assume that ε is a combination of systematic and random error. When a CWB is imposed, right-hand side terms in Eq. (3-1) including G, $\frac{\Delta S}{\Delta t}$, and ε are assumed to be zero, resulting in a simplified CWB form of the water budget in which $P - ET - Q = 0$. This simplification implies that *ET* can be calculated as $ET_{CWB} =$ P – Q.

Of the simple water budget-based tools that rely on CWB assumptions to derive ET_{CWB}, the Budyko hypothesis (Budyko, 1974) has emerged as a particularly useful and common framework for characterizing upland water resources (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; Greve *et al*., 2020). The Budyko hypothesis posits that water budget partitioning (e.g., the *ET* fraction or *ET/P* and runoff ratio or *Q/P*) can be determined solely based upon the ratio of available energy (often expressed as annual potential evapotranspiration, E_o, in equivalent water depth) to available water (expressed as annual *P,* also in water depth) (Sposito, 2017). Relying on this simplification, prior research has attributed observations of systematic patterns (i.e., statistically detectable trends) in catchment plotting relative to Budyko-type curves to underlying physical processes or catchment properties (e.g., Li *et al*., 2013; Padrón *et al*., 2017; Potter *et al*., 2005), such as streamflow response to shifts in precipitation phase measured as the annual fraction of P falling as snow (*fs*) and intensity under climate change (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; Ni *et al*., 2015).

To accurately characterize the water supplied from upland catchments, the above applications of the Budyko hypothesis require the imposition of (valid) CWB assumptions – specifically, that $G = 0$, $\varepsilon = 0 \rightarrow ET_{CWB} = P - Q$. However, there is growing evidence that these assumptions may not hold in upland catchments (Kampf *et al*., 2020; Safeeq *et al*., 2021). First, G—specifically interbasin groundwater exportation—is typically ignored in CWBs, but can be a significant component of the upland water budget (Frisbee *et al*., 2016; Safeeq *et al*., 2021) with potential to bias ET_{CWB} either high or low (Fan, 2019). Prior research has sought to characterize how non-zero groundwater fluxes and stores can impact catchment plotting in the Budyko space using models and densely instrumented catchments (Condon & Maxwell, 2017; Istanbulluoglu *et al*., 2012; Wang *et al*., 2009). Quantifying interbasin groundwater import or export, however, is notoriously difficult in upland settings due to a lack of data and the complexity of hydrologic pathways (Carroll *et al*., 2019; Fan, 2019; Maxwell & Condon, 2016). In the absence of available data, Fan (2019) proposed a new physically-based framework comprised of several explanatory criteria to condition expectations about the role of G in water budgets including: 1. catchment size, 2. catchment position, 3. aridity, 4. depth of permeable regolith, and 5. geological permeability (see Section 3.2). For example, the framework argues that headwater catchments with deep permeable regolith are likely to export groundwater ($G > 0$), violating CWB assumptions. Second, ε—particularly systematic error in P in snowy, steep upland catchments—can also impact upland water budget closure, violating CWB assumptions and impacting ET_{CWB} as a result. Upland gauges are often too sparse to represent spatial variability in P (Jing *et al*., 2017) and are plagued by under-catch bias in snowy and windy conditions (Rasmussen *et al*., 2012). Moreover, orographic enhancement and complex terrain can hamper both satellite retrievals and high-resolution models in these settings (Dettinger *et al*., 2004; He *et al*., 2019; Henn *et al*., 2018; Wrzesien *et al*., 2019). Given the above, there is a pressing need to better understand how evidence for non-zero G and ε interacts with CWB assumptions to influence inferences about upland water supplies derived from widely used tools like the Budyko hypothesis (Andréassian & Perrin, 2012; Valéry *et al*., 2010).

Open water budget (OWB) approaches, which require an independent estimate of ET (henceforth, ET_{OWB}), offer a pathway to more rigorously investigate the role poor assumptions about G and ε may play in inferences about upland water supplies (Kampf *et al*., 2020). In upland settings, advances in land data assimilation systems (LDASs) (e.g., Kumar *et al*., 2019) and remote sensing (e.g., Anderson *et al*., 2011; Mu *et al*., 2011) have given rise to a suite of new ET_{OWB} datasets. However, each of these ET_{OWB} datasets are subject to considerable uncertainties (Polhamus *et al*., 2013), which makes it challenging for users to determine if the benefits of adopting ET_{OWB} outweigh the previously described disadvantages of propagating assumptions about G and ε into ET_{CWB}. If users adopt an OWB-based approach at all, they are often left to rely on best judgement in selecting from a range of ET_{OWB} datasets. In some cases, users might simply select an ET_{OWB} dataset that uses other water budget variables like P and Q as inputs, hoping that this will reduce the total independent error sources (e.g., Barnhart *et al*., 2016). In other cases, users might elect to merge different estimates of ET_{OWB} together via an ensemble mean (e.g., Abolafia-Rosenzweig *et al*., 2021; Yilmaz *et al*., 2012). Doing so, however, implicitly assumes all estimates are equally error-prone. Information about the levels of errors in different ET_{OWB} estimates could facilitate a more accurate averaged best estimate. This can be achieved using triple collocation (TC)—a tool that objectively obtains random error estimates from three or more spatially and temporally collocated products to attain a single product with reduced random error (Gruber *et al*., 2017; Stoffelen, 1998; Yilmaz *et al*., 2012). Recent research has highlighted the value of TC for comparing different ET datasets (Khan *et al*., 2018). By improving estimates of P (Alemohammad *et al*., 2015), other research suggests that TC may also improve ET_{CWB} (Burnett *et al*., 2020) with additional benefits to water budget evaluations more generally. Despite the promise of TC, it is limited to providing statistical descriptions of random errors (see Section 3.3.1) and its limitations to describe or remove systematic errors impacting Eq. (3-1) in upland settings are unexplored.

OWBs—aided by recent advances in approaches to condition expectations about G (e.g., Fan (2019)), tools to characterize ε (e.g., TC) and products to estimate ET_{OWB}— offer promise in evaluating the benefits and limitations of using CWB assumptions to make inferences about critical upland water supplies (Kampf *et al*., 2020; Safeeq *et al*., 2021). Because the value of these advances remains largely speculative, this study aims to clarify what—if any—benefit they provide to potential users in 114 snow-dominated upland catchments that import and export G and have large potential for ε (Condon *et al*., 2020; Ying Fan, 2019; Henn *et al*., 2018; Rasmussen *et al*., 2012; Wrzesien *et al*., 2019). Using our study catchments, we first investigate the validity of CWB assumptions through an evaluation of long term water budget closure and the Fan (2019) framework. Second, we interrogate the consequences of improper CWB assumptions when inferring the response of upland surface water supplies to climate change—specifically upland streamflow response to changes in P. Here, we focus our analysis of these benefits on the Budyko hypothesis because it is a widely used CWB-based tool that has been used to set expectations about upland water resources in response to climate change (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; Greve *et al*., 2020). We follow Condon & Maxwell (2017) who more rigorously evaluate the effects of CWB assumptions in the Budyko space using ET_{CWB} and ET_{OWB}. Third, we examine whether these advances improve inferences about upland groundwater resources ignored in conventional applications of CWB-based tools like the Budyko hypothesis or otherwise masked by products with large ε. To do this we identify splits in our data using splits (e.g., thresholds) identified thorugh conditional inference trees—a type of unbiased recursive partitioning. Using these splits, we establish statistically-based rules to categorically classify our watersheds based on agreement with the Fan (2019) criterion. We distill the motivation of this study into three central questions:

1) Does long term OWB closure—assisted by the Fan (2019) framework, TC, and ET_{OWB} products— validate conventional CWB assumptions (i.e., $G = 0$, $\varepsilon = 0 \rightarrow ET_{CWB} = P - Q$) in upland catchments?
2) In upland catchments where CWB assumptions are invalid (e.g., $\varepsilon \neq 0$ due to P bias in steep, snowy catchments $\rightarrow ET_{CWB} \neq P - Q$), how do the Fan (2019) framework, TC, and ET_{OWB} products differ in Budyko-based inferences about

water supplies (i.e., streamflow efficiency in response to changes in precipitation phase, *fs*)?

3) When ε is characterized using TC, can the Fan (2019) framework and ET_{OWB} products improve insights about *G* export or import in upland settings?

3.2 Study Area and Data

3.2.1 Study Area

We focus on a large collection of upland catchments compiled using a subset of the Catchment Attributes and Meteorology for Large Sample studies (CAMELs) database (Newman *et al*., 2015). The CAMELS database is comprised of data for 671 catchments distributed throughout CONUS and includes forcing data for P, E_o, and Q. To select for upland catchments with high measurement uncertainty, we used the same subset of 268 catchments from the CAMELs database located throughout CONUS with fs > 0.15 used by Berghuijs *et al*. (2014). Catchments range from 6 to 2679 km^2 in drainage area with a mean size of 387 km^2 (see Figure 3-1). Additional details on candidate catchments are provided in Table S3-1.

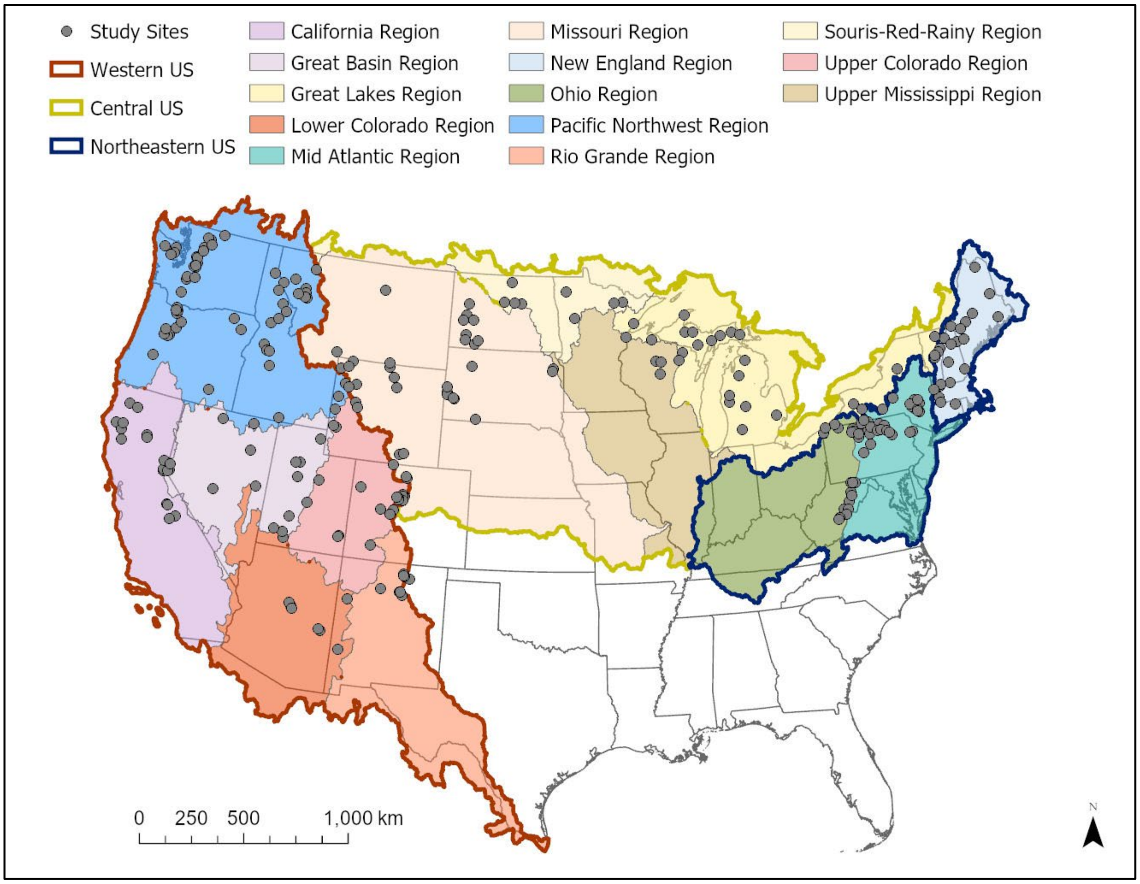

Figure 3-1: Map of candidate study sites broken out into the Western US (California, Great Basin, Pacific Northwest, Upper and Lower Colorado, and Rio Grande regions), Central US (Missouri, Upper Mississippi, Great Lakes, and Souris-Red-Rainey regions), and Northeastern US (Ohio, New England, and Mid-Atlantic regions). Here regions are abbreviated as follows: NE is New England, MA is Mid-Atlantic, OH is Ohio, GL is Great Lakes, UM is Upper Mississippi, MI is Missouri, SRR is Souris-Red-Rainey, UC is Upper Colorado, RG is Rio Grande, LC is Lower Colorado, GB is Great Basin, CA is California, and PNW is Pacific Northwest.

3.2.2 Data

In this study, we relied on multiple datasets (summarized in Table 3-1) to perform the analyses outlined in Section 3.3. For gridded time-series products, all data were obtained for the period from October 1, 2001 to September 30, 2016 for each catchment in Figure 1, which represents the maximum overlap between products allowable for the application of the TC approach outlined in Section 3.3.1. Please see Text S3.1 and S3.2 for more detail.

To assist users in evaluating the advantages and disadvantages associated with different estimates of ET_{OWB} and to test the value of TC-based merging in upland settings, we compared three different data products selected to represent the most common choices available to users in upland settings: 1) an NLDAS product (Section 3.2.2.1); 2) an Ensemble Mean product (Section 3.2.2.2); and 3) a TC-Merged product (described in Section 3.2.2.3 and 3.3.1 below).

Table 3-1: Summary of input data products used for this study including 5 P products and 4 ET products. Details about specific data products are summarized in Table 3-1 below. We note whether the data were included in the CAMELs database in the reference column. Data that were not available in the CAMELs database were independently estimated from the sources listed in Section 3.7. We provide more detail on each product in Text S3.2.

Product	Spatial Resolution	Spatial Extent	Temporal Resolution	Temporal Extent	Source	Included in Ensemble Mean?	Reference
WATER BUDGET AND BUDYKO VARIABLES							

PRECIPITATION							
ERA5	31 km	Global	Hourly	1979-Present	Reanalysis	No	(Hersbach *et al*., 2020)
PERSIANN-CDR	27.75 km	Quasi-Global	Daily	1983-Present	Satellite	No	(Sorooshian *et al*., 2014; Ashouri *et al*., 2015)
NLDAS-2	13.875 km	North America	Hourly	1979-Present	Gauge Based	Yes	(Xia *et al*., 2012) **Included in CAMELs database**
Daymet	1 km	North America	Daily	1980 - Present	Gauge Based	Yes	(Thornton *et al*., 2014) **Included in CAMELs database**
PRISM	4 km	CONUS	Daily	1981-Present	Gauge Based	Yes	(Daly *et al*., 1997, 2008)
EVAPOTRANSPIRATION							
ALEXI	10 km	CONUS	Daily	2001-Present	Thermal: TSEB	No	(Anderson *et al*., 2011)
SSEBop	1 km	CONUS	Daily	2000-Present	Thermal: Penma	Yes	(Senay *et al*., 2013)

					n-Monteith		
NCA-LDAS	12 km	CONUS	Daily	1979-2016	Land Surface Model: Penman-Monteith	Yes	(Kumar *et al*., 2019)
MODIS 16	0.5 km	Global	8-day	2001-Present	Satellite (Near Infrared): Penman-Monteith	Yes	(Mu *et al*., 2013)
POTENTIAL EVAPOTRANSPIRATION*							
NLDAS-2	13.875 km	North America	Hourly	1979-Present	Penman Based	N/A	(Xia *et al*., 2012) **Included in CAMELs database**
STREAMFLOW*							
USGS	catchment	North America	Daily	1980-2014	Gauge Based	N/A	(Addor *et al*., 2017; Newman *et al*., 2015) **Included in CAMELs database**

CATCHMENT AND SUBBASIN ATTRIBUTES							
SNOW FRACTION							
-	catchment	North America	Average Annual	1980-2014 (CAMELS); variable	-	N/A	(Addor *et al.*, 2017; Newman *et al.*, 2015) **Included in CAMELs database**
DEPTH TO BEDROCK							
Pelletier	catchment	Global	-	-	Model-Based	N/A	(Pelletier *et al.*, 2016) **Included in CAMELs database**
MAXIMUM SUB-BASIN ELEVATION							
SRTM	3 arc second	Global	-	-	Model-Based	N/A	(Yang *et al.*, 2011)
MEAN CATCHMENT ARIDITY							
-	catchment	CONUS	Average Annual	1980-2014 (CAMELs);variable	Model-Based	N/A	(Addor *et al.*, 2017; Newman *et al.*, 2015) **Included in CAMELs database**

GEOLOGIC PERMEABILITY							
GLHY MPS	~100 km^2 (average polygon size)	Global	-	-	Database synthesis	N/A	(Gleeson *et al.*, 2014) **Included in CAMELs database**

3.2.2.1 NLDAS product (Example of a single model)

The NLDAS product consists of NLDAS *P* forcing data, which is derived from a temporal disaggregation of gauge-only Climate Prediction Center (CPC) data with the PRISM topographical adjustment, CPC hourly CONUS gauge data, hourly Doppler radar precipitation data, half-hourly CPC data, 3-hourly North American Regional Reanalysis data (Xia *et al.*, 2012). The NLDAS product also includes NCA-LDAS *ET*, which is generated by running the National Climate Assessment-Land Data Assimilation System (LDAS) with NLDAS forcings, including *P* (Kumar *et al.*, 2019). The NLDAS product was selected to investigate the relative advantage of selecting a single dataset for *P* and *ET* (e.g., Barnhart *et al.*, 2016). NCA-LDAS was specifically selected based on the widespread use of LDASs for agricultural and water resources management applications and based on NCA-LDAS' explicit focus on the terrestrial water cycle. Unlike the other products, errors in NLDAS products are not mutually independent and are constrained to balance (over sufficiently long time periods). That is, unlike the two products below, the NLDAS product errors in *P* may be compensated by errors in *Q* and/or *ET*.

3.2.2.2 Ensemble Mean product (Example of merging of multiple datasets)

The Ensemble Mean product in this study was comprised of an ensemble mean of NLDAS, PRISM, and Daymet *P* datasets and an ensemble mean of MOD16, NCA-LDAS, and SSEBop *ET* datasets, which are noted in Table 3-1. The Ensemble Mean product was selected to investigate the common approach of merging multiple datasets. We elected to combine the *P* and *ET* products into the Ensemble Mean above based on their public availability, ease of access, and widespread use in studies of upland/mountain environments (Addor *et al*., 2017; Hahm *et al*., 2019; Newman *et al*., 2015; Velpuri *et al*., 2013).

3.2.2.3 TC-Merged product (Example of optimized merging of multiple datasets)

The TC-Merged product in this study was comprised of all *P* and *ET* datasets in Table 3-1 and was constructed following the methodology outlined in Section 3.3.1 below. The TC-Merged product was selected to investigate the merging of multiple datasets using an objective statistical representation of their random errors. Our constructed *P* triplets followed prior work (Massari *et al*., 2017) and were comprised of the following: 1) a reanalysis product; 2) a satellite-based product; and 3) a gauge-based interpolation or gauge-based interpolation/Land Surface Model (LSM) hybrid. *ET* triplets were constructed using: 1) a thermal product, 2) a LSM; and, 3) a near-infrared product.

3.3 Methods

To answer our three research questions, we first constructed new estimates of *ET* and *P* using TC-based merging (Section 3.3.1.), which required a robust characterization of cross-correlated estimation errors (Section 3.3.1.1). Once we characterized cross-correlated errors, we used TC outputs as the basis for optimized merging (Section 3.3.1.2) to obtain TC-Merged *P* & *ET* with minimized random error (Section 3.2.2.3). We then evaluated the performance of TC-Merged *P* & *ET* against NLDAS (Section 3.2.2.1) and Ensemble Mean (Section 3.2.2.2) *P* & *ET*.

We used TC-Merged, NLDAS, and Ensemble Mean *P* & *ET* to assess the validity of CWB assumptions, which are necessary for conventional application of the Budyko hypothesis. We did this by comparing 'inferred groundwater behavior' (e.g., $G + \varepsilon$) obtained using each of our three products in a long term OWB. We evaluated the long term OWB by rearranging Eq. (3-1) as:

$$P_{sum} - ET_{sum} - Q_{sum} = G_{sum} + \varepsilon_{sum} , \qquad \text{Eq. (3-2)}$$

where the subscript 'sum' indicates the 15-year sum of each variable. We neglected the contribution of $\frac{\Delta s}{\Delta t}$ because existing measurements of it were too coarse for our application (Tapley *et al*., 2004) and it was challenging to reliably separate *G* from $\frac{\Delta s}{\Delta t}$ (Enzminger *et al*., 2019). Consistent with reported common assumptions for USGS streamflow data from Hamilton & Moore (2012), we assumed Q_{sum} was accurate to within 5% at a 95% confidence interval. We normalized the resulting inferred groundwater behavior (i.e., $G + \varepsilon$) by *P* and sorted catchments into the three different categories described in Table 3-2: groundwater neutral, groundwater importer, or groundwater exporter.

Table 3-2: Summary of water budget closure categories adopted for this study based on Fan (2019) and their implications for G and ε. Psum is the sum of P over the 15-year period of record, ETsum is the sum of ET over the 15-year period of record, and so on.

Category	Mathematical Description per Eq. (2)	Implication for $\boldsymbol{G_{sum}} + \boldsymbol{\varepsilon_{sum}}$	Implication for CWB
Groundwater Neutral	$P_{sum} - ET_{sum} - Q_{sum} = 0$	$G_{sum} + \boldsymbol{\varepsilon}_{sum} = 0$	$ET_{CWB} = ET_{OWB}$
Groundwater Exporter	$P_{sum} - ET_{sum} - Q_{sum} > 0$	$G_{sum} + \boldsymbol{\varepsilon}_{sum} > 0$	$ET_{CWB} > ET_{OWB}$
Groundwater Importer	$P_{sum} - ET_{sum} - Q_{sum} < 0$	$G_{sum} + \boldsymbol{\varepsilon}_{sum} < 0$	$ET_{CWB} < ET_{OWB}$

Because each category in Table 3-2 can be influenced by true *G* and/or ε, we further tested the physical support for inferred groundwater behavior using five criteria proposed by Fan (2019) (described further in Section 3.3.2). We applied a type of unbiased recursive partitioning to establish splits in the five criteria related to differences in inferred groundwater behavior and then used the raw number of supporting criteria as the basis for classifying catchments with physically supported groundwater import or export (Section 3.3.2.1). Next, we used these results to examine the impacts of propagating poor CWB assumptions about *G* and ε into ET_{CWB} via the Budyko hypothesis (Section 3.3.3).

3.3.1 Characterization of random $\boldsymbol{\varepsilon}$: TC analysis

TC analysis requires the construction of a TC triplet with three distinct measurement systems (e.g., *X, Y,* and *Z*) of the same environmental variable (McColl *et al*., 2014; Stoffelen, 1998). Henceforth, we use TC[X-Y-Z] to refer to a generic TC triplet of a single environmental variable constructed from measurement systems *X*, *Y*, and *Z*. In order to obtain valid outputs for TC[X-Y-Z], TC requires that a number of assumptions are met (Gruber *et al*., 2017). The first of these assumptions (assumption 1) is that each measurement system or product (e.g., X, *Y*, *Z*) has a linear relationship to the "true" variable (C_T) which can be modeled for a generic variable *C* as follows:

$$C_i = \alpha_i \; C_T + \epsilon_i \qquad \text{Eq. (3-3)}$$

Here, I is the individual measurement system or product (e.g., X, Y, or Z), α_i is a measure of the relation between C_i and C_T, and ϵ_i are the respective random zero-mean errors associated with each measurement system or product. P errors are typically modeled as multiplicative in short-term and/or fine-scale applications (Alemohammad *et al*., 2015); however, recent work by Massari *et al*. (2017) and Dong *et al*. (2019) suggests that the assumption of a multiplicative error model for TC application at the daily timescale is not necessary. As such, we assumed that the underlying error model for both 8-day P and ET was linear and no logarithmic transformation was applied.

The second assumption (assumption 2) relates to signal and error stationarity, where stationarity is satisfied if the statistical properties of a geophysical process do not change over time. To satisfy assumption 2, we tested the performance of raw time-series data against time-series anomalies obtained by removing both the long-term and seasonal mean (e.g., DJF, MAM, JJA, and SON, Text S3.1). Because the TC results for all P and ET triplets were equivalent using raw and anomaly time-series data (not shown), we used raw time-series data in this analysis to avoid any effects introduced from seasonality determination on TC results.

The final two assumptions (assumptions 3 and 4) are that the signal and the error in the geophysical measurements are independent (error orthogonality) and that the errors in the selected geophysical measurements are independent (zero error cross-correlation) (Gruber *et al*., 2016). Additionally, obtaining valid ETC results requires that each triplet have at least 50 or more points and that all correlation outputs are positive for each of the input timeseries datasets (Chen *et al*., 2018). If assumptions 1 - 4 are satisfied, then the error variances (σ^2) can be determined using six unique elements from sample covariance matrix (e.g., A_{XY}) between three measurement systems using TC following Eq. (3-4 to 3-6):

$$\sigma_x^2 = A_{XX} - \frac{A_{XY}A_{XZ}}{A_{YZ}} \qquad \text{Eq. (3-4)}$$

$$\sigma_y^2 = A_{YY} - \frac{A_{XY}A_{YZ}}{A_{XZ}} \qquad \text{Eq. (3-5)}$$

$$\sigma_z^2 = A_{ZZ} - \frac{A_{XZ}A_{YZ}}{A_{XY}} \qquad \text{Eq. (3-6)}$$

The additional contribution of extended triple collocation (ETC) following McColl *et al*. (2014), is the estimation of Pearson's correlation coefficient (*R*) between C_i and C_T using Eq. (3-7), which estimates *R* between C_X and C_T as an example, and where σ_{XY} is the covariance between C_X and C_Y:

$$R_X = \sqrt{\frac{\sigma_{XY}\sigma_{YZ}}{\sigma_x^2\sigma_{XY}}} \qquad \text{Eq. (3-7)}$$

3.3.1.1 Estimation of uncertainties and 95% confidence interval calculations

Because of the 15-year period of record available across all data, considerable estimation errors arising from differences in our sample and the true variable were expected in our results. Furthermore, because two of the data products used were incorporated in all possible triplets, estimation errors were also assumed to be highly correlated across products (Chen *et al*., 2018). This correlation affects the R_X calculated by ETC, despite R_X representing a value that is only influenced by product X (C_X) and the true signal C_X. Comparison of R_X values across different triplets can thus be used to detect the influence of estimation errors on the TC results (Crow *et al*., 2017; Yilmaz & Crow, 2014). For example, with minimal bias from cross-correlated estimation errors, the R_X obtained from TC[X-Y-Z] would be expected to be the same (or very similar) to R_X obtained from TC[X-Y-W]. We would therefore anticipate small pairwise differences in calculated correlation values (ΔR). Thus, it is important to quantify when the difference in pairwise ΔR values is significant to appropriately interpret ETC results.

To detect cross-correlated estimation errors, we obtained uncertainty intervals for ΔR values sampled across all catchments using a 1000-member boot-strapping approach. Pairwise ΔR were assessed in this manner for all common products (e.g., the R_X obtained from TC[X-Y-Z] – R_X obtained from TC[X-Y-W]). Each boot-strapped sample in this approach was constructed using the exact same set of days. We then constructed 95% confidence intervals from the boot-strapped sampling distribution, which we defined as the range between the 2.5th and 97.5th percentile of boot-strapped ΔR values. We assumed that the results for any catchment that fell outside the 95% confidence interval in any pairwise ΔR comparison were unacceptably impacted by cross-correlated estimation errors and removed those catchments from further analysis.

3.3.1.2 TC-based merging

We then used a composite of the TC-based merging methodologies put forward by Yilmaz *et al*. (2012) and Gruber *et al*. (2017) to obtain an estimate of the true underlying values for both *P* and *ET* based on the ETC results. This method was also applied by Burnett *et al*. (2020). Using this methodology, TC-based error variance estimates (as opposed to ETC-based *R* estimates) can be used to obtain a single more accurate dataset for a given geophysical variable (denoted C_M , the TC-based merged estimate of C_T):

$$C_M = w_x\ C_X + w_y\ C_Y + w_z\ C_{Z,}, \qquad \text{Eq. (3-8)}$$

where w is a weight obtained from the TC-based error variances obtained from Eq. (3-4 to 3-6) obtained for each measurement system or product using Eq. (3-9 to 3-11):

$$w_x = \frac{\sigma_y^2\sigma_z^2}{\sigma_x^2\sigma_y^2 + \sigma_x^2\sigma_z^2 + \sigma_y^2\sigma_z^2} \qquad \text{Eq. (3-9)}$$

$$w_y = \frac{\sigma_x^2\sigma_z^2}{\sigma_x^2\sigma_y^2 + \sigma_x^2\sigma_z^2 + \sigma_y^2\sigma_z^2} \qquad \text{Eq. (3-10)}$$

$$w_z = \frac{\sigma_x^2\sigma_y^2}{\sigma_x^2\sigma_y^2 + \sigma_x^2\sigma_z^2 + \sigma_y^2\sigma_z^2} \qquad \text{Eq. (3-11)}$$

3.3.2 Characterization of G: Fan (2019) analysis

Because both *G* and/or ε can contribute to inferred groundwater behavior (Table 3-2) based on the evaluation of Eq. (3-2), we investigated the physical support for apparent groundwater neutrality, export, or import in each catchment using five criteria proposed by Fan (2019):

- **Criterion 1 (Catchment Scale).** Small catchment size increases the likelihood of groundwater importer or exporter behavior. If the catchment is small compared to its permeable regolith, the catchment has a greater likelihood of being a groundwater exporter. We assessed this factor in our study catchments using catchment size sourced from the CAMELS database and depth to bedrock estimates from Pelletier *et al*. (2016).
- **Criterion 2 (Catchment Position)**. Catchments situated at the high end of a regional elevation gradient are more likely to be groundwater exporters (Buss *et al*., 2013) while catchments situated at the low end of the gradient are more likely to be importers (Genereux *et al*., 2013). We assessed this factor in our study catchments using the ratio of mean catchment elevation relative to the maximum elevation of the sub-basin (HUC8) containing each study catchment (e.g., higher ratio corresponds to higher catchment position and vice versa). Mean catchment elevation was assessed using data from the CAMELs database and maximum sub-basin elevation was assessed using data from Yang *et al*. (2011).
- **Criterion 3 (Climate).** Headwater catchments under a dry climate with seasonal aridity or interannual droughts are more likely to be groundwater exporters. Fan (2019) argue that an arid climate leads deep local water tables below stream beds in the headwaters. Water from precipitation and losing streams in arid headwater catchments enters regional groundwater flow systems and resurfaces in lower basins to feed gaining systems (Käser & Hunkeler, 2016). They further hypothesize that the amount of groundwater export will increase if the climate is wetter in the headwaters and drier in the lower basin and decrease in the reverse case. We assessed this factor

in our study catchments using aridity data from the CAMELs database (Addor *et al*., 2017).

- **Criterion 4 (Substrate Properties).** Catchments underlain by thick regolith, fractured rock, or sediments in the headwaters facilitate are likely to be groundwater exporters (Buss *et al*., 2013). This is particularly true in tectonically active terrain under a humid climate. We assessed this factor using depth to bedrock from Pelletier *et al*. (2016).
- **Criterion 5 (Geologic Structure).** Catchments situated atop high-permeability, dipping sedimentary beds extending beyond the catchment, or shared sedimentary beds—particularly carbonate rocks—are more likely to be groundwater importers or exporters. We assessed the geologic permeability (expressed as mu(K) where K is hydraulic conductivity using data from the GLobal HYdrogeology MaPS (GLHYMPS) (Gleeson *et al*., 2014).

For each data product, we first assessed the statistical relationships between Fan (2019) criteria and inferred groundwater behavior based on Eq. (3-2) using Kruskal-Wallis and pairwise Wilcoxon rank sum tests. We binned catchments into roughly equal groups unless there were objective breakpoints (e.g., climate and substrate properties). We then used a Kruskal-Wallis test to determine if any statistically significant differences were observed between groups and OWB closure assuming an α of 0.05. In the case that statistically significant differences were observed using the Kruskal-Wallis test, we then used a pairwise Wilcoxon rank sum test to further characterize the statistical significance of differences between groups.

To relate the Fan (2019) criteria to our data, we established splits (e.g., thresholds) in each Fan (2019) variable used for categorical classification of our catchments. Splits were determined using conditional inference trees, which are a type of unbiased recursive partitioning (Hothorn, Hornik, & Zeileis, 2012; Hothorn, Hornik, Van De Wiel, *et al*., 2012). Following Hothorn *et al*. (2015), the algorithm we used first tested independence between input variables (e.g., catchment scale, position, and so on) and the response (i.e., OWB closure). The algorithm stopped if all variables are found to be independent of the response; otherwise, the variable with the strongest association –as measured by a p-value –was selected and a binary split was implemented. This process was then repeated recursively to obtain splits reported in Table 3-3. For more details on this method we refer to Hothorn *et al*. (2015) and for more details on its implementation in this study we refer to Figure S3-1 to S3-6. We report the qualitative Fan (2019) criterion, the split identified by unbiased recursive partitioning that was assumed to represent any qualitative threshold (e.g., catchment was positioned on the high end of a regional gradient) reported by Fan (2019), and the resulting rules for defining agreement between apparent groundwater behavior and the Fan (2019) criteria in Table 3-3.

Table 3-3: Summary of the classification rules for Fan (2019) factors developed via conditional inference trees as described by Hothorn *et al*. (2015) and reported in Figure S3-1 to S3-6. We define two rules for position as Fan (2019) clearly indicate that lower lying catchments are more likely to be importers and higher up catchments are more likely to be exporters. In the absence of agreement with any rules listed, catchments retained their apparent groundwater behavior (e.g., importer, exporter, or neutral), but were classified with lower confidence per Section 3.3.2.1.

Criterion	Split based on unbiased recursive partitioning	Agreement rule with Fan (2019)
Catchment Scale	No statistically significant relationship (Figure S3-1)	None defined
Position	Statistically significant split in the relationship between OWB closure and position (m/m) > 0.56 (Figure S3-2)	Apparent groundwater export if OWB + position > 0.56 Apparent groundwater import if OWB + position < 0.56
Climate	Statistically significant split in the relationship between OWB closure and aridity (mm/mm) > 0.42 (Figure S3-3)	Apparent groundwater export if OWB + aridity > 0.42
Substrate Properties	Statistically significant split in the relationship between OWB closure and soil depth (m) > 2.09 (Figure S3-4)	Apparent groundwater export if OWB + soil depth > 2.09
Geologic Structure	Statistically significant split in the relationship between OWB closure and permeability (mu(K)) > -12.6 (Figure S3-5)	Fan (2019) does not indicate clear expectation for *G,* so none defined

3.3.2.1 Classification of G or ε dominance analysis

Fan (2019) argued that multiple criteria provide the strongest evidence for groundwater leakage; in other words, satisfying one of the criteria outlined in Section 3.3.2.2. does not unambiguously support groundwater export or import. For example, they find that catchments are more likely to be leaky if positioned at the high end of a steep regional gradient, underlain by deep substrates, and in a drier climate. As such, we do not focus on whether a catchment satisfies a single criterion for our classification. We grouped catchments into four categories based on the agreement between inferred groundwater behavior based on the evaluation of Eq. (3-2) and agreement with the Fan (2019) criteria as defined in Table 3-3.

- **Self-Contained.** We defined catchments as self-contained if long term G_{sum} + ε_{sum} was 0% +/- Q uncertainty bounds. These catchments were not assessed for agreement with Fan (2019) criteria.
- **Dominant $\boldsymbol{\varepsilon}$ (No Fan (2019) criterion met).** We assumed that ε in the underlying water budget variables caused the lack of agreement between any explanatory physical criteria and inferred groundwater behavior. We had the lowest degree of confidence in the physical realism of our upland water budgets in these catchments.
- **$\boldsymbol{G}$ and $\boldsymbol{\varepsilon}$ (One Fan (2019) criterion met).** We assumed that both ε and G influenced the agreement between only one explanatory physical criterion and inferred groundwater behavior. We did not attempt to further disentangle the influence of ε and G due to data limitations and had a lower degree of confidence in the physical realism of our upland water budgets in these catchments.
- **Dominant $\boldsymbol{G}$ (Two or more Fan (2019) criteria met).** We assumed that agreement between multiple explanatory physical criteria and inferred groundwater behavior indicated true signal associated with G. We had the highest degree of confidence in the physical realism of our upland water budgets in these catchments.

3.3.3 Impacts of unsupported CWB assumptions in the Budyko space analysis

The impacts of broken CWB assumptions on water budget-based inferences are underexplored (Andréassian & Perrin, 2012; Koppa & Gebremichael, 2017). In order to investigate potential impacts, we first conducted a direct comparison of ET_{OWB} and ET_{CWB} using each of our three data products. We then assessed how broken CWB assumptions influenced deviation from a Budyko modeled ET (i.e., ET_{Budyko}) fraction by contrasting our observed ET_{CWB} fraction against our observed ET_{OWB} fraction. Finally, we evaluated how these differences impacted inferences about Budyko streamflow anomaly (Q_{anom}) to *fs* —a metric commonly used as a proxy for climate change in upland catchments (Gordon *et al*., 2022).

3.3.3.1 Budyko *ET* fraction analysis

We compared long term estimates of ET_{CWB} or ET_{OWB} across all study sites using our three different products. Using potential evapotranspiration data from NLDAS (Table 3-1, Text S3.2) and our three different realizations of *P*, we then evaluated how our different realizations of *ET* fraction interacted with *G* and ε in the Budyko space. To do this, we used the conventional form of the Budyko hypothesis presented in Eq. (3-12) to obtain a modeled estimate of ET_{Budyko} by multiplying the left-hand side by *P*:

$$\frac{\overline{ET_{Budyko}}}{\bar{P}} = \sqrt{\frac{\overline{E_o}}{\bar{P}} \tanh\left(\frac{\bar{P}}{\overline{E_o}}\right)\left(1 - \exp\left(-\frac{\overline{E_o}}{\bar{P}}\right)\right)}, \qquad \text{Eq. (3-12)}$$

where $\overline{ET}$, $\bar{P}$, and $\overline{E_o}$ are mean annual *ET, P,* and E_o over the 15-year period of record and $\overline{ET_{Budyko}}$ is the mean annual modeled Budyko *ET*.

3.3.3.2 Budyko Q_{anom} fraction analysis

In upland settings, analyses of Budyko streamflow anomalies (Q_{anom}) have indicated that greater-than-expected streamflow efficiency (or runoff ratio—*Q/P)* is correlated to higher *fs*, suggesting that streamflow will decline with an increase in winter rain under climate change (Berghuijs *et al*., 2014). However, to the best of our knowledge, these analyses have not considered whether findings are sensitive to assumptions about *G* and ε imbedded in ET_{CWB} despite evidence for large systematic ε due to *P* under-catch in snow-dominated catchments (Lundquist *et al*., 2021; Wrzesien *et al*., 2019). Here, we test how G and ε influence the relationship between Budyko streamflow anomaly and *fs* using the equation below:

$$\frac{\overline{Q_{anom}}}{\bar{P}} = \left(1 - \frac{\overline{ET}}{\bar{P}}\right) - \left(1 - \frac{\overline{ET_{Budyko}}}{\bar{P}}\right), \qquad \text{Eq. (3-13)}$$

where $\overline{ET}$ is mean annual ET calculated via ET_{CWB} or ET_{OWB} over the 15-year period of record and $\overline{ET_{Budyko}}$ is obtained from Eq. (3-12). We then examined the correlation between different realizations of streamflow anomaly and mean annual *fs*.

3.4 Results

3.4.1 Characterization of random ε: TC-merging results

A limitation of our TC-based analysis is the need to exclude catchments with cross-correlated estimation error (Section 3.3.1). Catchments where any pairwise ΔR values (Section 3.3.1.1) fell outside the 95% confidence intervals (indicated by black dashed lines in Figure S3-7 and S3-8, respectively) were excluded (see Table S3-2 and S3-3). This exclusion led to a reduction in the number of study catchments (from 268 to 114). In the valid 114 catchments, pairwise ΔR values were small and close to zero (see red lines in S3-7 and S3-8). The similarity in mean R values obtained for different products (e.g., ERA5 P in Figure 3-2A shaded in light purple) suggests that any bias from cross-correlated sampling errors in the constructed triplets had a small impact on ETC results. Based on this logic, ETC was deemed successful in 114 of the 268 candidate catchments. There were 143 valid catchments for ET and 170 catchments for P, with 114 catchments that had both. For successful catchments, individual performance—as measured by R—in P products (Figure

3-2A) was more varied than for *ET* (Figure 3-2B). The ERA5 most strongly correlated to the unknown "true" value of *P* in 85 of 170 catchments (~50%), NLDAS in 58 of 170 catchments (~34%), Daymet in 20 of 170 catchments (~12%), and PRISM in 7 of 170 catchments (~4%). In 124 of 143 catchments (~87%), NCA-LDAS exhibited the strongest correlation to the unknown "true" value of *ET*. In the remaining 19 catchments, ALEXI had the strongest correlation to the true *ET* in 8 catchments (~6%), MOD16 in 6 catchments (~4%), and SSEBop in 5 catchments (~3%).

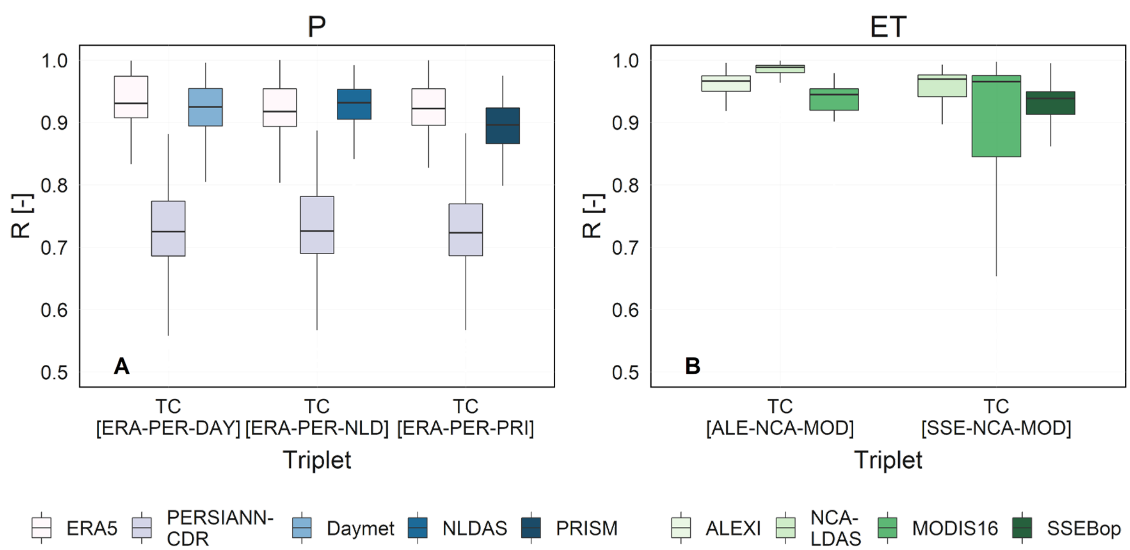

Figure 3-2: A) ETC R results for P products in all valid catchments (n = 170); and B) ETC R results for ET products in all valid catchments (n = 143). Boxplots describe variability across catchments. Here R is an estimation of the correlation between the product and the "true" underlying P or ET value as described in Section 3.3.1. Boxplots show the minimum, 25th percentile, median, 75th percentile, and maximum values. Catchments where any pairwise ΔR values (Section 3.3.1.1) fell outside the 95% confidence intervals (indicated

by black dashed lines in Figure S3-7 and S3-8, respectively) were excluded (see Tables S3-2 and S3-3).

3.4.2 Characterization of *G*: Fan (2019) results

We assessed Eq. (3-2) using TC-Merged *P* & *ET*, NLDAS *P* & *ET*, and Ensemble Mean *P* & *ET* to obtain an estimate of inferred groundwater behavior (i.e., $G + \varepsilon$). Median inferred groundwater behavior was -1.8% of *P* using TC-Merged *P* & *ET* , -9.7% of *P* using NLDAS *P* & *ET* , and -19.3% of *P* using Ensemble Mean *P* & *ET*; all of which are consistent with findings by Safeeq *et al*. (2021) in a smaller number of densely instrumented catchments. Table 3-4 reports the number of catchments classified as groundwater exporters (i.e., $P_{\mathrm{sum}} - Q_{\mathrm{sum}} - ET_{sum} > 0$), groundwater neutral (i.e., $P_{\mathrm{sum}} - Q_{\mathrm{sum}} - ET_{sum} \cong 0$), or groundwater importers (i.e., $P_{\mathrm{sum}} - Q_{\mathrm{sum}} - ET_{sum} < 0$) in accordance with the definitions outlined in Table 3-2. A greater number of catchments were characterized as groundwater exporters or groundwater neutral in the Northeastern and Central US (e.g., green and tan boxes in the top half of Figure 3-3) using TC-Merged *P* & *ET*. Conversely, all products indicated widespread regional groundwater importation in the Western US (e.g., red and blue colored boxes in the bottom half of Figure 3-3) and particularly the Pacific Northwest (labeled PNW in Figure 3-3) although this could be error related per our results in Figure 3-5.

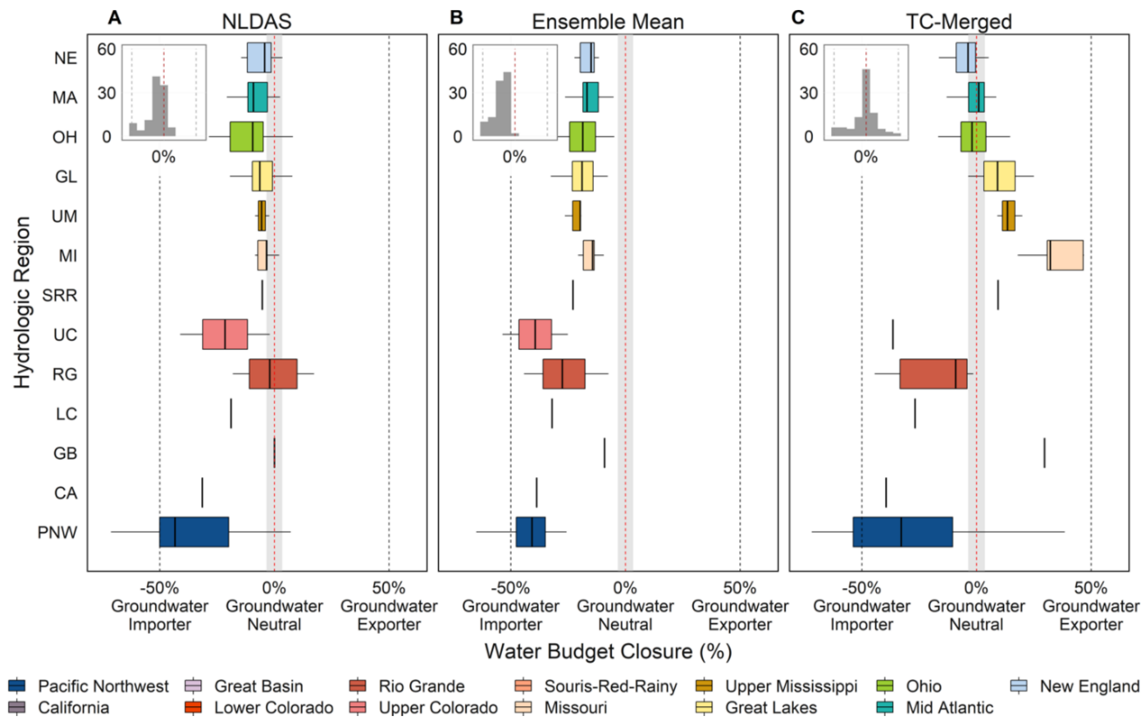

Figure 3-3: Regional variability in the evaluation of Eq. (3-2) where: A) groundwater importers are defined based on long-term water budget closure support for negative G_{sum} ($P_{\mathrm{sum}} - Q_{\mathrm{sum}} - ET_{sum} < 0$), groundwater neutrality is defined based on long-term water budget closure support for G_{sum} equal to 0 ($P_{\mathrm{sum}} - Q_{\mathrm{sum}} - ET_{sum} \cong 0$), and groundwater exporters are defined based long-term water budget support for positive G_{sum} ($P_{\mathrm{sum}} - Q_{\mathrm{sum}} - ET_{sum} > 0$) are estimated using the NLDAS P & ET; B) same as above but with Ensemble Mean P & ET; C) same as above but with TC-Merged P & ET. We plot the water budget closure as a percent of each respective P_{sum} to facilitate more intuitive and contextual interpretation of the results. Grey bounds indicate the potential uncertainty introduced by Q_{sum}. We refer the reader to Figure 3-1 for abbreviations. Inset histograms represent the distribution of OWB closure across catchments using each of the three different products.

Table 3-4: Summary of inferred groundwater behavior (e.g., $G_{sum} + \varepsilon_{sum}$) based on the evaluation of Eq. (3-2) with NLDAS, Ensemble Mean, and TC-Merged products.

Category	**Mathematical Description**	**NLDAS**		**Ensemble Mean**		**TC-Merged**	
		#	**% of total**	**#**	**% of total**	**#**	**% of total**
Groundwater Neutral	$G_{sum} + \varepsilon_{sum} = P_{sum} - Q_{sum} - ET_{sum} = 0$	25	21.9%	0	0%	29	25.4%
Groundwater Exporter	$G_{sum} + \varepsilon_{sum} > P_{sum} - Q_{sum} - ET_{sum} > 0$	7	6.1%	0	0%	33	28.9%
Groundwater Importer	$G_{sum} + \varepsilon_{sum} < P_{sum} - Q_{sum} - ET_{sum} < 0$	82	71.9%	114	100%	52	45.6%

We found that OWB closure using TC-Merged P & ET was the most consistent with physical reasoning based on Fan (2019). For example, when TC-Merged P & ET were used we observed that groundwater exporters were positioned higher up in the containing sub-basin (Figure 3-4F), were variably more arid (Figure 3-4I) and had deeper permeable regolith or fractured rock (Figure 3-4L), consistent with Fan (2019). TC-Merged P & ET indicated statistically significant differences between inferred groundwater behavior and catchment position (Table 3-5), with groundwater exportation (i.e., positive $G + \varepsilon$) observed for higher catchment positions (Figure 3-4F) as postulated by Fan (2019). Although NLDAS P & ET indicated a similar rightward shift in the boxplots as drainage position increased (Figure 3-4D), it supported widespread groundwater importation (i.e., negative $G + \varepsilon$) counter to our expectation. TC-Merged P & *Et al*so supported a statistically significant relationship between higher groundwater exportation (i.e., positive $G + \varepsilon$) and aridity (Figure 3-4I, Table 3-5). Conversely, NLDAS and Ensemble Mean P &

ET indicated a statistically significant relationship between groundwater importation (i.e., negative $G + \varepsilon$) and aridity (Figure 3-4G, 3-4F, Table 3-5). Consistent with the expectation that deep regolith or fractured rock increased groundwater export (Fan, 2019), TC-Merged *P* & *ET* supported statistically significant relationships between greater depth to bedrock and groundwater export (i.e., positive $G + \varepsilon$) (Figure 3-4L). Neither NLDAS nor Ensemble Mean *P* & *ET* (Figure 3-4J and 3-4L) indicated any statistically significant relationships between depth to bedrock and inferred groundwater behavior (Table 3-5).

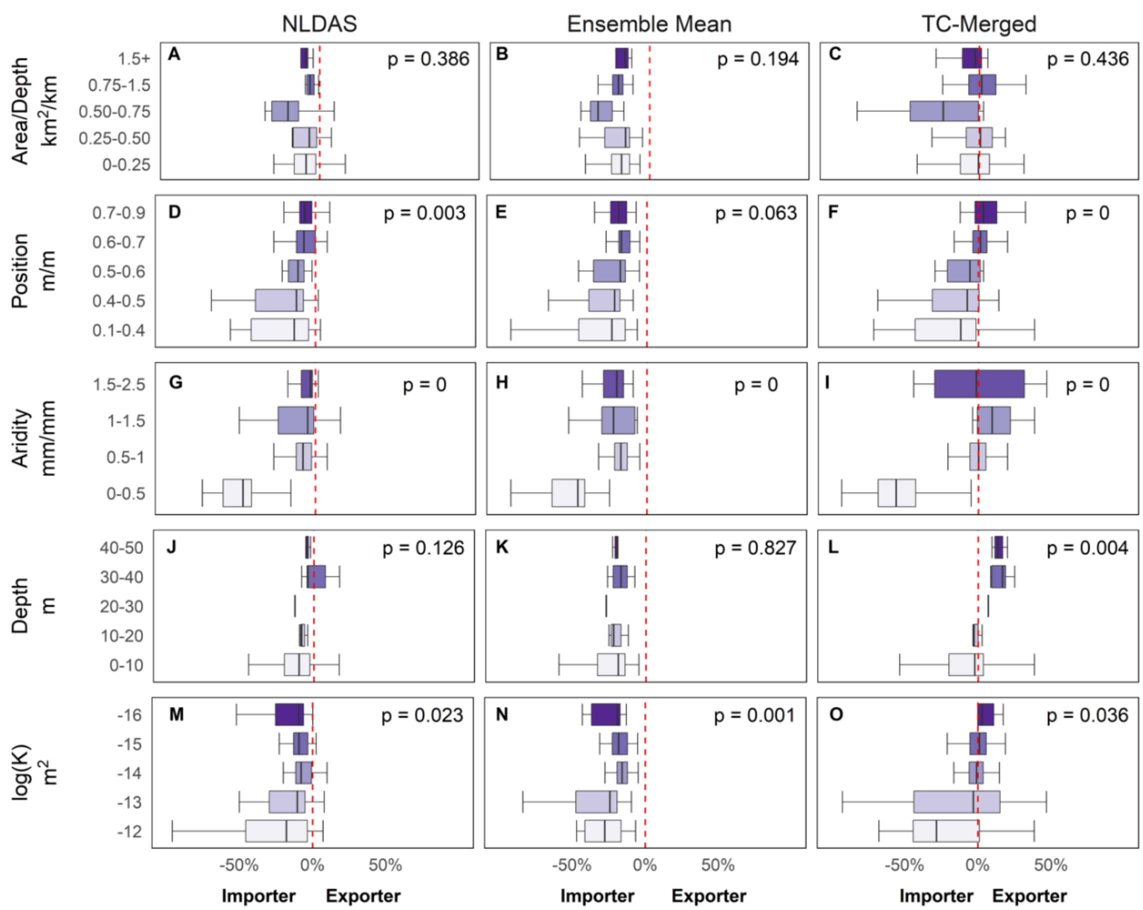

Figure 3-4: Boxplots showing the 25th percentile, 50th percentile, 75th percentile and the standard error bars for inferred groundwater behavior based on Eq. (2) versus each of the

five Fan (2019) criteria proposed in Section 3.3.2.1 to determine whether a catchment is an importer, neutral, or exporter. The criteria are plotted as follows: A-C) Catchment size relative to depth of permeable regolith (Criteria 1); D-F) Catchment position relative to a regional gradient measured as the ratio of mean catchment elevation to maximum sub-basin elevation; G-I) Catchment climate as measured using the aridity index; J-L) Catchment depth of permeable substrate or fractured rock as measured by depth to bedrock; and M-O) Catchment geological permeability as measured by the log of hydraulic conductivity (K). Inferred groundwater behavior is approximated for plots A, G, D, J, M using NLDAS *P* & *ET*, for plots B, E, H, K, N using Ensemble Mean *P* & *ET*, and plots C, F, I, L, and O using TC-Merged *P* & *ET*. We report the Kruskal-Wallis p-value assuming a significance level of α = 0.05 for each plot. If the Krushkal-Wallis p-value was statistically significant, we reported the results of a Pairwise Wilcoxon Rank Sum Test (Table 3-5).

Table 3-5: Table of the statistically significant results for Pairwise Wilcoxon Rank Sum Tests for each of the five Fan (2019) factors proposed in Section 3.3.2.1. We assumed a significance level of α = 0.05 and report significant p-values for pairwise comparisons within the groups above using a Pairwise Wilcoxon Rank Sum Test. We did not calculate pairwise statistics if the Kruskal-Wallis p-value was not significant.

Factor	**Pairwise Comparison**	**Group 1**		**Group 2**		**NLDAS**	**Ensemble Mean**	**TC-Merged**
		Name	**Sample Size**	**Name**	**Sample Size**	**p Value**	**p Value**	**p Value**

Position (m/m)	0.1-0.4 versus 0.6-0.7	0.1-0.4	20	0.6-0.7	21	0.008	0.048	0.005
	0.1-0.4 versus 0.7-0.9	0.1-0.4	20	0.7-0.9	38	0.008	*NS*	0.000
	0.4-0.5 versus 0.6-0.7	0.4-0.5	19	0.6-0.7	21	0.012	0.008	0.017
	0.4-0.5 versus 0.7-0.9	0.4-0.5	19	0.7-0.9	38	0.006	0.044	0.001
	0.5-0.6 versus 0.6-0.7	0.5-0.6	16	0.6-0.7	21	0.047	*NS*	0.025
	0.5-0.6 versus 0.7-0.9	0.5-0.6	16	0.7-0.9	38	0.020	*NS*	0.002
Aridity (mm/mm)	0-0.5 versus 0.5-1	0-0.5	17	0.5-1	79	0.000	0.000	0.000
	0-0.5 versus 1-1.5	0-0.5	17	1-1.5	7	0.002	0.002	0.002
	0-0.5 versus 1.5-2.5	0-0.5	17	1.5-2.5	11	0.000	0.000	0.000
Depth (m)	0-10 versus 30-40*	0-10	102	30-40	5	*NS*	*NS*	0.003
	0-10 versus 40-50*	0-10	102	40-50	3	*NS*	*NS*	0.023
	10-20 versus 30-40	10-20	3	30-40	5	*NS*	*NS*	0.036
log(K) (m^2)	-12 versus -14	-12	21	-14	29	0.007	0.01	0.019
	-12 versus -15	-12	21	-15	31	0.013	0.005	0.003
	-12 versus -16	-12	21	-16	8	*NS*	*NS*	0.032

	-13 versus -14	-13	25	-14	29	0.024	0.001	*NS*
	-13 versus -15	-13	25	-15	31	*NS*	0.002	*NS*

3.4.2.1 Classification of G or ε dominance results

We separated catchments into the four groups reported in Table 3-6 using the classification rules established via our unbiased recursive partitioning analysis (Table 3-3). Per Figure 3-5, we observed distinct regional patterning in our catchment classification. For example, of the 37 catchments located in the Western US (Table S3-2) that were not apparently self-contained, there was strong physical support for inferred groundwater behavior in only one catchment (~3% of 37) with weak support for 10 catchments (~27% of 37) and inconclusive support in 23 catchments (~62% of 37) using TC-Merged *P* & *ET* with NLDAS and Ensemble Mean *P* & *ET* yielding similar results. Some of these catchments in the Western US with the greatest implied ε were in steeper terrain and generally had a larger *fs* (consistent with gauge under-catch; see Figure S3-9), which would be expected to lead to the systematic under-prediction of *P* (Henn *et al*., 2018; Rasmussen *et al*., 2012; Wrzesien *et al*., 2019). Overall, TC-Merged *P* & *ET* yielded the highest number of catchments with strong physical support as assessed against the Fan (2019) framework, particularly in the Central and Eastern US (Figure 3-5, Table 3-6).

Table 3-6: Summary of inferred groundwater behavior (e.g., $G_{sum} + \varepsilon_{sum}$) based on the evaluation of Eq. (3-2) with NLDAS, Ensemble Mean, and TC-Merged products.

Classification	Agreement with Table 3 Rules	Assumptions	NLDAS		Ensemble Mean		TC-Merged	
			#	% of total	#	% of total	#	% of total

Self-Contained	N/A	$G = 0, \varepsilon = 0$	25	21.9%	0	0.0%	29	25.4%
Dominant *ε*	No rules satisfied	Very weak physical support for inferred groundwater behavior	41	36.0%	65	57.0%	19	16.7%
G* and *ε	One rule satisfied	Inconclusive physical support for inferred groundwater behavior	29	25.4%	30	26.3%	33	28.9%
Dominant *G*	Two or more rules satisfied	Strong physical support for inferred groundwater behavior	19	16.7%	19	16.7%	33	28.9%

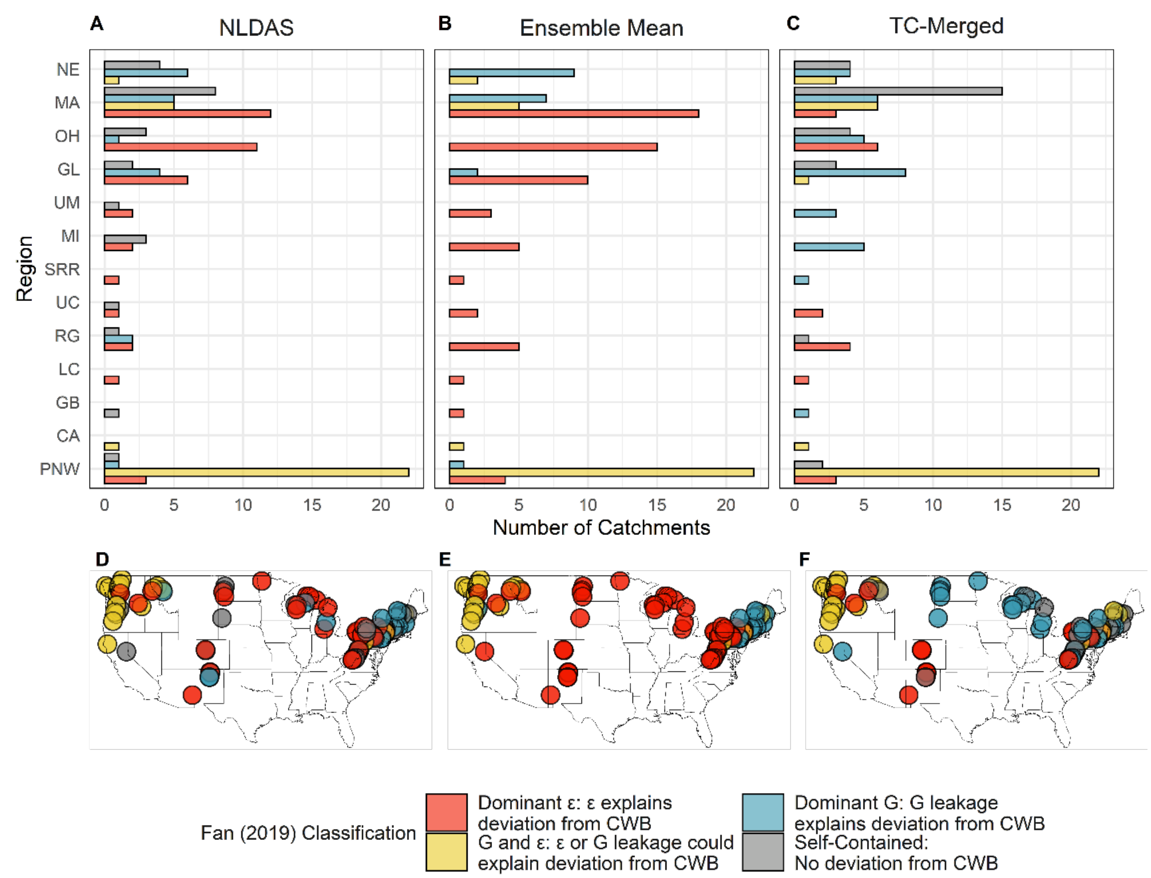

Figure 3-5: Variability in physical support for inferred groundwater behavior based on the Fan (2019) criterion presented in Section 3.3.2.1 using A) NLDAS P & ET; B) Ensemble Mean P & ET; and C) TC-Merged P & ET. Maps displaying catchment classification using D) NLDAS P & ET; E) Ensemble Mean P & ET; and C) TC-Merged P & ET. Abbreviations for the regions correspond to Figure 3-1.

3.4.3 Impacts of unsupported CWB assumptions

3.4.3.1 Budyko ET fraction results

Our results show that choices about *ET,* including whether it is evaluated independently (e.g., ET_{OWB}) or derived (e.g., ET_{CWB}) by ignoring G and ε, can lead to substantially different hydrologic inferences in upland catchments (Figure 3-6A to 3-6C). Across all products, we observed limited agreement between ET_{CWB} and ET_{OWB} with inset histograms reinforcing widespread potential for ET_{CWB} to underestimate ET_{OWB} (warmer colors in Figure 3-6A to 3-6C). This underestimation of ET_{OWB} was particularly pronounced in places (e.g., steep, snowy catchments in the western US, Figure S3-9) where the estimation of P is more variable (larger circles in Figure 3-6A to 3-6C). We suggest that this is consistent with large unconsidered ε (Figure 3-5) arising from systematic under-prediction of P (Henn *et al*., 2018; Rasmussen *et al*., 2012; Wrzesien *et al*., 2019). TC-Merged P & ET were less obviously biased towards underestimation than either NLDAS or Ensemble Mean P & ET, particularly in the central and eastern US. Here, the TC-Merged P & ET indicated potential for ET_{CWB} to overestimate ET_{OWB}, which is consistent with unconsidered G exportation and our classification results (Figure 3-5).

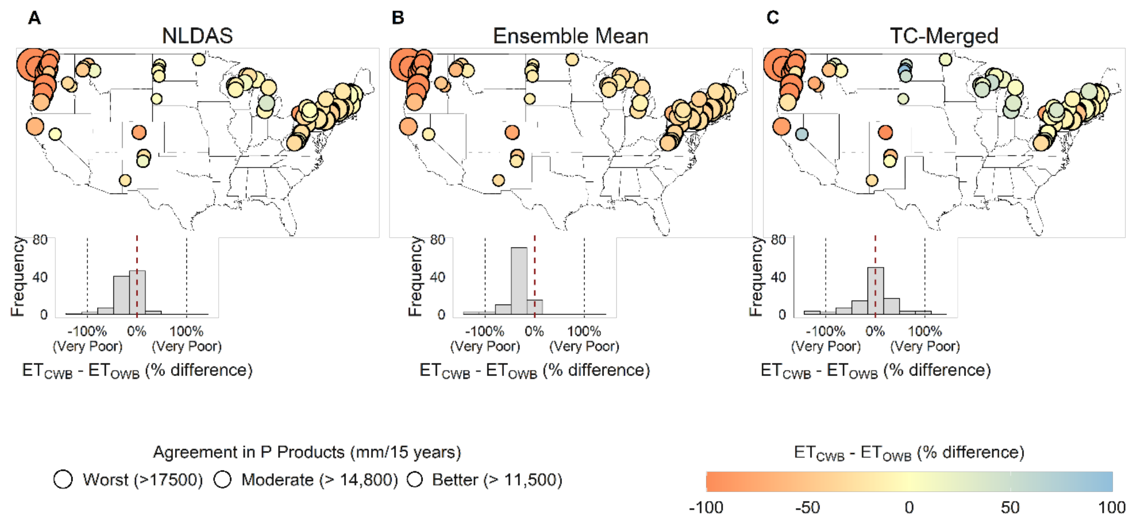

Figure 3-6: Differences between long-term ET_{CWB} and ET_{OWB} across all catchments (n = 114) using: A) NLDAS *P* & *ET* and USGS *Q*; B) Ensemble Mean *P* & *ET* and USGS *Q*; and C) TC-Merged *P* & *ET* and USGS *Q*. Coloring is based on observed on the percent difference between ET_{CWB} and ET_{OWB} with values less than -100% and more than 100% constrained to those bounds for plotting. Sizing is based on the maximum disagreement between long-term estimates of *P*. We present scatterplots of ET_{CWB} versus ET_{OWB} colored by aridity in Figure S3-10.

In the Budyko space, we found that the largest differences in plotted *ET* fraction were driven by the propagation of CWB assumptions about *G* and ε into ET_{CWB} with smaller differences based on the selection of an individual product (Figure 3-7A, C, E versus Figure 3-7B, D, F). When ET_{CWB} was used, we observed that *ET* fraction was substantially lower than expected based on Eq. (3-12) in catchments impacted by ε (triangles and inverted triangles in 3-7A, C, and E) when compared to smaller deviations in catchments with physically realistic *G* (squares in Figure 3-7) or that were self-contained (circles in Figure

3-7) consistent with Jones *et al*. (2012). Voepel *et al*. (2011) suggested that steeper catchments may be associated with lower ET fraction; however, we found that steepness may increase ε bias in observed ET fraction. That is, catchments impacted by ε also tended to be steeper, snowier, and in the Western US (see Figure S3-9), where other research suggests that there is large potential for ε —particularly systematic P under-prediction (Henn *et al*., 2018; Rasmussen *et al*., 2012; Wrzesien *et al*., 2019). When ET_{OWB} was used, G and ε had less obvious influence over ET fraction (Figure 3-7B, D, F), underscoring how these assumptions are propagated into the Budyko space via ET_{CWB} when not properly considered.

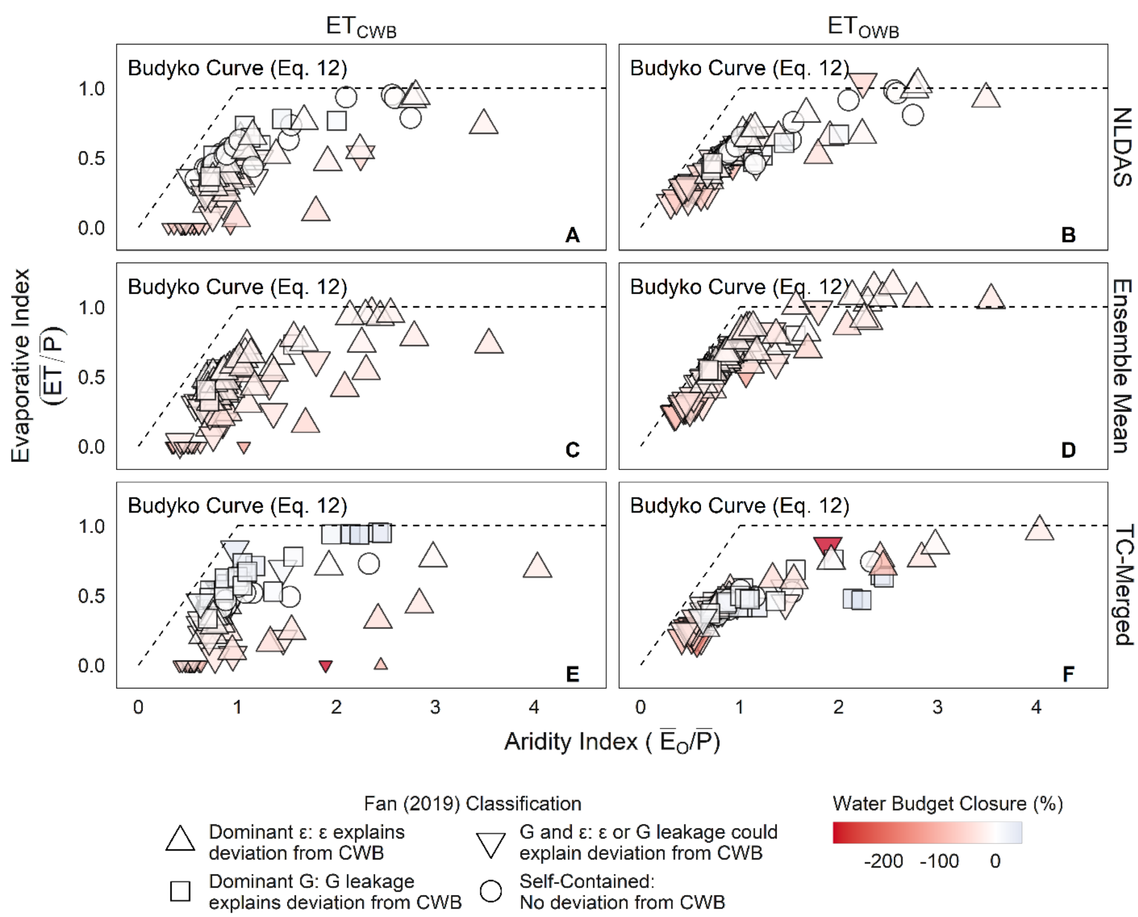

Figure 3-7: Variability in Budyko plots across all catchments (n = 114) using ET_{CWB} combined with: A) NLDAS P; C) Ensemble Mean P; and E) TC-Merged P. Variability in Budyko plots across all catchments using ET_{OWB} combined with: B) NLDAS P & ET; D) Ensemble Mean P & ET; and F) TC-Merged P & ET. Shapes correspond to Fan (2019) classification based on Figure 3-5. Coloring is based on observed OWB closure as calculated according to Section 3.3.2. Evaporative Index values less than 0 were forced to 0 for plotting purposes and are denoted as smaller symbols.

3.4.3.2 Budyko Q_{anom} results

Per Eq. (3-13), lower-than-expected *ET* fraction drives higher-than-expected streamflow anomaly in Budyko-based assessments. Although previous research has attributed higher-than-expected streamflow anomaly to larger *fs* in upland catchments (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; Ni *et al*., 2015), we found that the strength of this relationship is also systematically influenced by ε in particular when ET_{CWB} is used. For example, when TC-Merged *P* was used to obtain ET_{CWB}, we found that catchments impacted by ε (dashed lines in Figure 3-8E, Table 3-7) were 5.5 times more sensitive to *fs* than catchments with physically realistic *G* (dotted lines in Figure 3-8E, Table 3-7) and 3 times more sensitive than self-contained catchments (solid lines in Figure 3-8E, Table 3-7). This patterned response was similar using NLDAS *P* (Figure 3-8C), but more muted using Ensemble Mean *P* (Figure 3-8A). The contrast with ET_{OWB} across all products further highlighted the effect of poor CWB assumptions about ε versus *G* on the relationship between streamflow anomaly and *fs* (Figure 3-8B, D, F, Table 3-7). Notably, when TC-Merged P & ET were used in catchments impacted by *G* (i.e., with strong physical support for underlying water budget), streamflow sensitivity to *fs* was consistent between ET_{CWB} (slope = 0.13, Table 3-7) and ET_{OWB} (slope = 0.11, Table 3-7) and lower than previous results (e.g., slope = 0.37 from Berghuijs *et al*. (2014) and 0.32 from Barnhart *et al*. (2016)). Both TC-Merged *P* & *ET* and NLDAS *P* & *Et al*so yielded consistent slopes in self-contained catchments per Table 3-7.

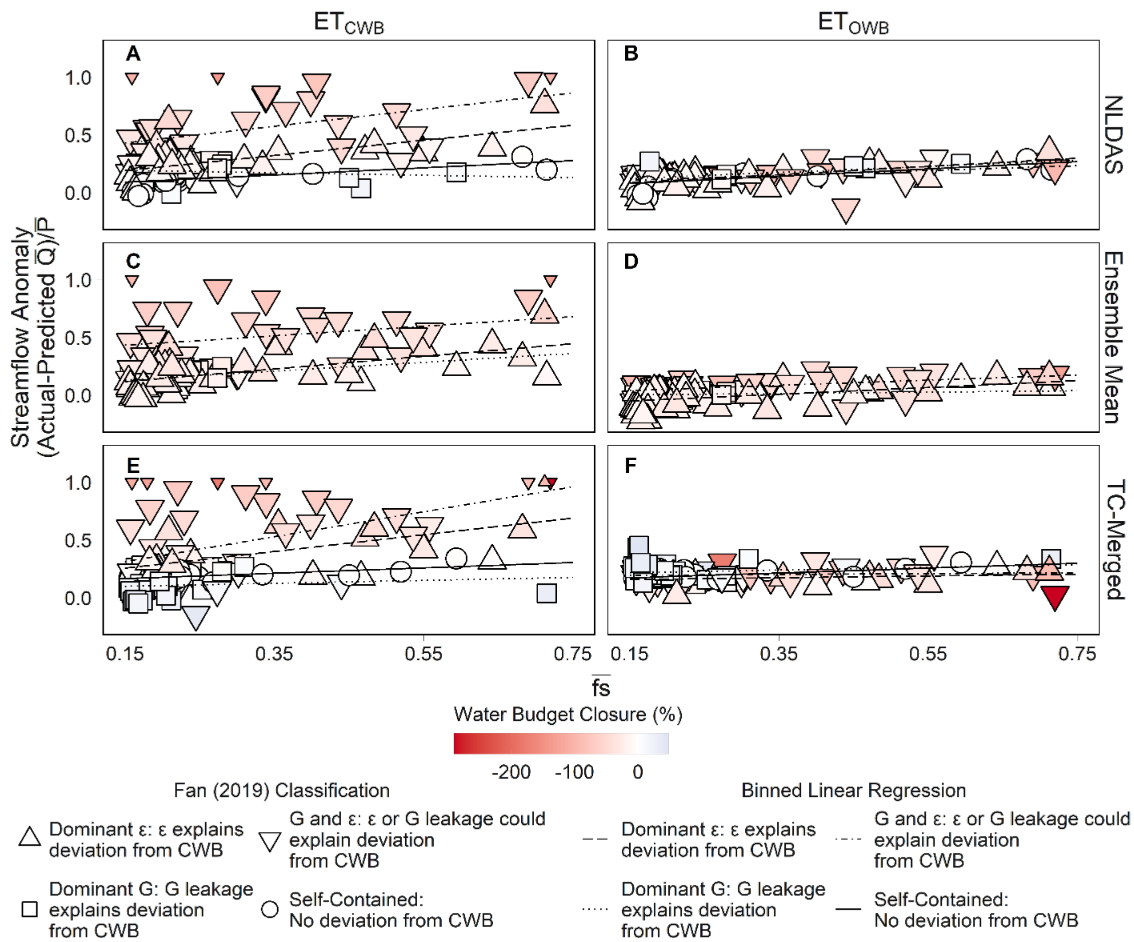

Figure 3-8: Variability in the relationship between Budyko streamflow anomaly and *fs* across all catchments (n = 114) using ET_{CWB} combined with: A) NLDAS *P*; C) Ensemble Mean *P*; and E) TC-Merged *P*. Variability in the relationship between Budyko streamflow anomaly and *fs* using ET_{OWB} combined with: B) NLDAS *P* & *ET*; D) Ensemble Mean *P*

& *ET*; and F) TC-Merged *P* & *ET*. Shapes correspond to Fan (2019) classification based on Figure 3-5. Coloring is based on the evaluation of Eq. (3-2) using ET_{OWB}. Line-type corresponds to binned linear regressions based on catchment grouping according to Fan (2019) classification based on Figure 3-5. Consistent with Figure 3-6, $\overline{Q_{anom}}$ values greater than 1 were forced to 1 for plotting purposes and are denoted as smaller symbols.

Table 3-7: Table of the binned linear regression equations and correlation coefficients corresponding with Figure 3-7A to 3-7F. Here, y = $\overline{Q_{anom}}$ and x = $\overline{fs}$.

Data Product	**Number of Fan (2019) Criteria Met**	ET_{CWB}		ET_{OWB}	
		Equation	**r**	**Equation**	**r**
NLDAS	Self-Contained	y = 0.31x + 0.04	0.66	y = 0.32x + 0.03	0.68
	Dominant ε	y = 0.66x + 0.09	0.62	y = 0.35x + 0.03	0.63
	Dominant ε or G	y = 0.65x + 0.38	0.32	y = 0.21x + 0.08	0.44
	Dominant G	y = -0.10x + 0.21	-0.15	y = 0.26x + 0.08	0.60
Ensemble Mean	Self-Contained	-	-	-	-
	Dominant ε	y = 0.55x + 0.04	0.560	y = 0.29x – 0.09	0.53
	Dominant ε or G	y = 0.81x + 0.27	0.44	y = 0.22x + 0.01	0.46
	Dominant G	y = 0.38x + 0.07	0.27	y = 0.06x – 0.01	0.11
TC-Merged	Self-Contained	y = 0.23x + 0.13	0.58	y = 0.21x + 0.14	0.62
	Dominant ε	y = 0.72x + 0.15	0.58	y = 0.08x + 0.15	0.28

	Dominant ε or G	y = 1.68x + 0.14	0.41	y = 0.04x + 0.2	0.09
	Dominant G	y = 0.13x + 0.08	0.13	y = 0.11x + 0.20	0.14

3.5 Discussion

3.5.1 Does long term OWB closure— assisted by the Fan (2019) framework, TC, and ET_{OWB} products — validate conventional CWB assumptions in upland catchments?

Neglecting the contributions of G and ε to upland water budgets by imposing closure conveniently reduces data requirements to P and Q . However, our results show that these assumptions are unsupported in a range of upland settings (see Figure 3-3, 3-5) with significant implications for CWB-based tools, including conventional forms of the Budyko hypothesis. We show that when a CWB is inappropriately applied, the true G and ε magnitudes are deposited into a calculated ET (ET_{CWB}) (Figure 3-6), leading to systematic biases in catchment plotting in Budyko space (Figure 3-7, 3-8) as discussed in Section 3.5.2. We contrast these results with an evaluation of Eq. (3-2) using an OWB-based approach that leverages a physically-based framework for characterizing G proposed by Fan (2019), TC-based merging, and several independent estimates of ET_{OWB}. The combined result of these advances indicated widespread non-zero $G + \varepsilon$ in a range of upland settings: for example, inferred groundwater behavior (i.e., $G + \varepsilon$) as determined by the evaluation of Eq. (3-2) using TC-Merged P & ET suggested that 85 of 114 upland catchments were not self-contained. These findings provide broad and quantitative evidence in support of recent calls to revisit common assumptions about G and ε used to close the water budget and derive ET_{CWB} in upland settings (Fan, 2019; Kampf *et al*., 2020; Safeeq *et al*., 2021).

At the same time, our results point to the very real challenge of disentangling signal associated with G from ε (Figure 3-3, 3-5, Tables 3-4 to 3-6) in upland settings that lack comprehensive groundwater datasets (Fan, 2019) and are more susceptible to the systematic under-prediction of P (Lundquist *et al*., 2019; Rasmussen *et al*., 2012; Wrzesien *et al*., 2019). This is particularly true for steeper, snowier catchments in the western US due to orographic enhancement and complex terrain (Dettinger *et al*., 2004; He *et al*., 2019; Henn *et al*., 2018; Wrzesien *et al*., 2019), sparse precipitation measurements (Jing *et al*., 2017), and P under-catch (Rasmussen *et al*., 2012). Consistent with this body of research, we found that the evaluation of Eq. (3-2) in these catchments was disproportionately impacted by ε (Figure 3-3, 3-5, S9), limiting confidence in the resulting water budgets regardless of the products, approaches, and tools used. Although this challenge is a significant one, as bias correction methodologies improve estimations of P (Beck *et al*., 2020) the benefits of using TC-Merged estimates of ET_{OWB} to evaluate Eq. (3-2) may enable more robust water budget-based supply predictions in these particular settings.

3.5.2 In upland catchments where CWB assumptions are invalid, how do the Fan (2019) framework, TC, and ET_{OWB} products differ in Budyko-based inferences about water supplies?

Recent research has highlighted the need to better connect systematic shifts in catchment location in the Budyko space to physical explanation (Berghuijs *et al*., 2020). Previous work (Hahm *et al*., 2019; Istanbulluoglu *et al*., 2012; Jones *et al*., 2012) has suggested that groundwater losses can lead to systematic deviations in catchment plotting below the Budyko curve due to lower ET fractions and illustrated impacts to streamflow (Han *et al*., 2021), with a diminished focus on ε. Meanwhile, other work has examined the influence of ε without consideration of G (Andréassian & Perrin, 2012; Koppa & Gebremichael, 2017; Valéry *et al*., 2010). Our results underscore that there is a pressing need for future work to simultaneously consider both (Figure 3-3 to 3-8, Tables3-4 to 3-7)—particularly when ε is systematic in nature (Section 3.4.2.1)—as it can potentially bias estimates of ET_{CWB} (Figure 3-5 to 3-6), which are widely used in upland settings (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; Condon & Maxwell, 2017; Greve *et al*., 2020). Consistent with prior work on the influence of groundwater in the Budyko space (Hahm *et al*., 2019; Istanbulluoglu *et al*., 2012; Jones *et al*., 2012), we observed that catchments tracked slightly below the Budyko curve. We observed slightly lower-than-expected ET fractions based on modeled ET_{Budyko} per Eq. (3-12) when constrained to catchments with strong physical support for inferred groundwater behavior (see squares in Figure 3-7). However, large ε in the underlying water budget led to more uniform and dramatic deviations from ET on modeled ET_{Budyko} (see triangles and inverted triangles in Figure 3-7).

Understanding whether and how changes in precipitation phase will influence streamflow remains a pressing challenge as reflected in contrasting literature results (Barnhart *et al*., 2016; Berghuijs *et al*., 2014; McCabe *et al*., 2018; Milly & Dunne, 2020; Gordon *et al*., 2022). By imposing CWB assumptions in upland catchments, some research (Berghuijs *et al*., 2014; Ni *et al*., 2015) has used the Budyko hypothesis to posit that higher *fs* influences lower-than-expected ET fractions (higher-than-expected Q_{anom}) based on modeled

ET_{Budyko}. However, snow-dominated upland catchments where *fs* is higher are also likely to experience non-zero G (Carroll *et al*., 2019) and are particularly susceptible to systematic under-prediction of P (Dettinger *et al*., 2004; He *et al*., 2019; Henn *et al*., 2018; Wrzesien *et al*., 2019). We also observed non-zero G and ε in the majority of our catchments (Table 3-4), the balance of which was propagated into the relationship between Q_{anom} and *fs* via ET_{CWB} (Figure 3-8, Table 3-7). Across all products, we found that the relationship between Q_{anom} and *fs* was highly sensitive to broken assumptions about ε when ET_{CWB} was used. Using TC-Merged ET_{CWB}, the sensitivity of Q_{anom} to *fs* was 5.5 times greater in catchments impacted by ε than in catchments with physically supported G and 3 times greater than in self-contained catchments (Figure 3-8, Table 3-7). Within catchments impacted by ε, we further observed that the use of ET_{CWB} increased the sensitivity of Q_{anom} to *fs* by 9 times when compared to ET_{OWB}. Catchments impacted by ε tended to be steeper, snowier catchments in the western US (Figure 3-5, S9), highlighting that systematic under-prediction of P can bias Budyko-based inferences about upland water supplies in important and unconsidered ways.

Using TC-Merged P & ET in catchments with strong physical support for OWB closure (Section 3.4.2) also led to weaker relationships between Q_{anom} and *fs* than in previous results (slope = 0.13 for ET_{CWB}, slope = 0.11 for ET_{OWB} per Table 3-7 versus slope = 0.37 from Berghuijs *et al*. (2014) and 0.32 from Barnhart *et al*. (2016)). These findings are consistent with recent research showing that climate change-driven declines in *fs* have not led to anticipated declines in streamflow efficiency in the Western US (McCabe *et al*., 2018) and/or that rain may countervail reductions in streamflow from declining *fs*

(Hammond & Kampf, 2020). Given this, we suggest that caution is warranted when applying the conventional (CWB-based) form of the Budyko hypothesis to make predictions about water supplies in upland settings where disentangling the effects of physical (e.g., *G*) and non-physical (e.g., ε) factors remains a challenge. Expansion of the Budyko equation and subsequently Eq. (3-13) to include *G* and ε variables following prior work by Istanbulluoglu *et al*. (2012) may be one way to enable more robust inferences.

3.5.3 When ε is characterized using TC, can the Fan (2019) framework and ET_{OWB} products improve insights about *G* export or import in upland settings?

CWBs and associated tools like the Budyko hypothesis ignore *G* to derive ET_{CWB} and make inferences about upland surface water supplies over longer timescales. However, evidence increasingly suggests that upland catchments can import/export substantial groundwater resources even over longer periods of time (Condon *et al*., 2020; Fan & Schaller, 2009; Fan, 2019). We found that TC-Merged *P* & *ET* facilitated insights about *G* that were not possible with other products tested (Figure 3-5) and importantly, that are ignored altogether in many applications of CWBs and CWB-based tools including the Budyko hypothesis (Section 3.5.1 and 3.5.2). Consistent with expectations based on relevant literature (Fan & Schaller, 2009; Welch & Allen, 2012), TC-Merged *P* & *ET* substantially increased the number of catchments classified as groundwater exporters based on inferred groundwater behavior (i.e., positive $G + \varepsilon$ based on Eq. (3-2) over other products (33 with TC-Merged *P* & *ET* versus 7 with NLDAS *P* & *ET* versus 0 with Ensemble Mean per Table 3-4). When inferred groundwater behavior was further evaluated using the criteria outlined in Section 3.3.2., TC-Merged *P* & *ET* yielded 33 catchments with strong physical support for inferred

G (Table 3-6), which was almost two times the catchments NLDAS P & ET or Ensemble Mean P & ET yielded (19 catchments each, respectively). Overall, statistically characterizing random ε via TC substantially increased the number of catchments with physically supported G based on five testable physical criteria drawn from Fan (2019) (Figure 3-5, Table 3-6). Furthermore, results using TC-Merged P & ET confirmed the profile of groundwater exporting catchments drawn by Fan (2019): that they are positioned higher up in their containing sub-basins, are drier, and have deeper permeable regolith or fractured rock (Figure 3-4 and 3-5, Table 3-5 and 3-6). When combined with Section 3.5.1 and 3.5.2, this suggests that there is robust potential for novel combinations of TC-based merging and the Fan (2019) framework to harness advances in independent estimates of ET_{OWB} to improve predictions about critical surface and groundwater resources originating in upland settings.

3.5.4 Limitations and paths forward

Accurately closing the water budget in data-limited upland settings has eluded hydrologists for decades (Kampf *et al*., 2020) and is likely to remain elusive for many more (Safeeq *et al*., 2021). However, incremental progress that weaves together novel tools like TC and physically based frameworks like the one presented by Fan (2019) into OWBs can facilitate critical insight about upland water supplies. Nevertheless, several limitations are worthy of further discussion.

First, near term measurement limitations in interbasin groundwater fluxes and stores (Condon & Maxwell, 2017; Fan, 2019) undeniably exist and were a limitation in this study. Due to these limitations, we were restricted in our validation of inferred groundwater behavior and neglected the contribution of $\frac{\Delta s}{\Delta t}$ due to the too coarseness of existing data (Tapley *et al*., 2004) and challenges associated with distinguishing *G* from $\frac{\Delta s}{\Delta t}$ (Enzminger *et al*., 2019). However, a number of promising paths are being pursued to estimate groundwater fluxes and stores using gravity based measurements such as the Gravity Recovery and Climate Experiment (GRACE) mission (Tapley *et al*., 2004) and follow-on mission (Flechtner *et al*., 2014). Here too, there is potential value in microwave-based soil moisture retrievals (Crow *et al*., 2017b), which could serve as a down-scaling tool. Other approaches, such as hydro-geophysical characterization (Gordon *et al*., 2020; Schmidt & Rempe, 2020; Smith *et al*., 2017), or some combination of the above in combination with the water budget (Hahm *et al*., 2019) may also be helpful in improving our understanding of upland groundwater resources. At larger scales, however, physically based frameworks like the one put forward by Fan (2019) can help to improve and refine our understanding of the characteristics that promote interbasin groundwater import or export beyond those illustrated here. There are, however, challenges associated with linking qualitative criteria to quantitative data in a robust and repeatable manner. In this study, we used a Kruskal-Wallis and pairwise Wilcoxon rank sum to first establish statistical relationships between water budget closure and continuous data used to approximate the physical criteria proposed by Fan (2019). We then use conditional inference trees to establish rules relating observed water budget closure to physical criteria. Although our results indicate strong statistical support for the use of Fan (2019), future work could improve linkages between this generalizable framework and continuous data. When incorporated into the template put forward by Istanbulluoglu *et al*. (2012), advances in the quality and availability of groundwater data could assist simple tools like the Budyko hypothesis in more robustly accounting for conventionally ignored water budget variables.

Second, the use of TC in our study reduced the number of study catchments from 268 to 114, which underscores an important limitation of this method, at least based on the levels of correlated errors in currently available products. Because our methodology required valid TC outputs (e.g., outputs that do not violate assumptions listed in Section 3.1) for both *P* and *ET*, it is not surprising that many catchments had to be excluded in our analysis. Interestingly, valid *ET* estimates proved to be more challenging for TC (n = 143 valid estimates) than valid *P* outputs (n =170). These challenges could be attributed in part to the pervasiveness of systematic errors in the underlying *P* data, including persistent under-catch, which tend to be more resilient to filtering via averaging (Yilmaz & Crow, 2014). Despite these challenges, TC enabled physically supported insight about *G* and ε that were otherwise obscured using other products (Figure 3-8). Improvements in bias-correction methodologies and underlying estimations (Beck *et al*., 2020) could work in concert with advances in data and modeling to expand the number of products with independent error sources available in upland settings. Of particular interest is the role dynamic atmospheric models like the Weather Research and Forecasting (WRF) model may have in improving estimates of *P* by better accounting for topographical features and mountain-precipitation interactions (He *et al*., 2019; Lundquist *et al*., 2019). Advances in remote sensing may also improve detection of solid and/or mixed phase *P* in complex topography (Lundquist *et al*., 2008; Maggioni *et al*., 2016), which is a goal of the on-going Global Precipitation Measurement (GPM) mission (Skofronick-Jackson *et al*., 2018). Continued improvement in measurement technology can help further eliminate systematic errors, enhancing the efficiency of TC for statistically characterizing and filtering random error in upland water budgets.

3.6 Conclusions

Water budgets and associated tools like the Budyko hypothesis are likely to remain central to investigations of upland catchment behavior. Recent advances in approaches to condition expectations about G (e.g., Fan (2019)), tools to characterize ε (e.g., TC), and products to estimate ET_{OWB} offer an opportunity for users to pivot away from the restrictive—and increasingly fragile—assumptions required to impose water budget closure. Motivated by this opportunity, our study sought to better understand the value of these advances for understanding the validity and consequences—in terms of inferences about surface and groundwater —of CWB assumptions (i.e., $G = 0$, $\varepsilon = 0 \rightarrow ET_{CWB} = P - Q$) in a range of upland settings.

Here we find that propagating largely unsupported CWB assumptions into ET_{CWB} had a profound effect on inferences about upland surface and groundwater resources when assessed against an OWB assisted by a physical framework proposed by Fan (2019), TC, and independent estimates of ET_{OWB}. In particular, we observed that unconsidered ε can unrealistically alter expectations about streamflow response to climate change and mask groundwater contributions. Long term OWB closure supported non-zero G and ε in 85 of 114 catchments (~75%) using TC-Merged P & ET, 89 of 114 catchments (~78%), using NLDAS P & ET, and all 114 catchments using Ensemble Mean P & ET over a 15-year period. When these were screened based using five testable criteria proposed by Fan (2019), TC-Merged P & ET led to physically realistic G in 33 of 114 catchments (~29%)—

the most of any product tested. Using TC-Merged estimates of ET_{CWB}, we observed inflated (5.5 times higher) streamflow response to *fs* in catchments with large ε compared to those with smaller suspected ε. Furthermore, within catchments with large ε, we found that TC-Merged ET_{CWB} increased the sensitivity of streamflow to *fs* by 9 times more than TC-Merged ET_{OWB}. Because ET_{CWB} is commonly used to make these types of inferences, our results highlight the need for users to consider the effects of ε—in addition to physical factors such as *G*—in discussions about catchment behavior in the Budyko space. More fundamentally though, the widespread conflict between CWB assumptions and observed OWB closure in upland catchments points to the need for users to critically evaluate the Budyko assumptions when error and groundwater flow are expected to be high.

In the vital pursuit of improved predictions about upland water supplies, our results demonstrate the value of TC and a physically-based framework proposed by Fan (2019) for harnessing advances in the estimation of ET_{OWB} and expanding the use of OWBs. While CWB approaches—and to a lesser extent conventional products (e.g., a standard model and ensemble mean)—overlooked strong evidence for *G*, the combination of TC-based merging and the physically-based framework proposed by Fan (2019) revealed groundwater exportation in high positioned, arid catchments with deep substrates. Albeit challenging, larger and more comprehensive groundwater datasets in upland settings could help to further refine insights about *G* in Budyko-type analyses. As modeling, remote sensing, data assimilation, and bias correction techniques in upland settings advance in tandem, combining TC and evidence-based frameworks is a promising path forward to improve predictions about upland water supplies.

3.7 Acknowledgments, Samples, and Data

Support for this research is provided the USDA NIFA (Project # NEVW-2017-08812), NSF EAR #2012310, and the Lincoln Institute's Babbitt Dissertation Fellowship Program. We would also like to thank the three anonymous reviewers, Associate Editor, and Editor for their constructive and detailed feedback on this manuscript. Many thanks to Dr. Gabrielle Boisrame for her constructive and thoughtful feedback on this manuscript and guidance on statistics. We are also grateful for Dr. Christopher Hain at NASA for generously providing and assisting with the ALEXI data, Dr. Sebastian Krogh and Aidan Manning for constructive feedback on an earlier draft of this manuscript, and Gary Sterle for assistance with figures. The USDA ARS is an equal opportunity employer.

We provide our TC-Merged data in SI. With the exception of ALEXI data (Anderson *et al*., 2011), the original data products used in this study are freely available to the public: streamflow, Daymet precipitation, NLDAS precipitation, catchment shapefiles, and attributes can be retrieved from https://ral.ucar.edu/solutions/products/camels (last access: 20 July 2021) (Addor *et al*., 2017; Newman *et al*., 2015). ERA5 precipitation data were accessed at https://cds.climate.copernicus.eu/cdsapp#!/home (last access: 1 February 2020) (Hersbach *et al*., 2020) via Google Earth Engine, Persiann-CDR precipitation data were accessed at https://doi.org/10.7289/V51V5BWQ (last access: 1 February 2020) (Sorooshian *et al*., 2014; Ashouri *et al*., 2015), NLDAS precipitation data were accessed via https://ral.ucar.edu/solutions/products/camels and https://doi.org/10.1029/2010EO340001 (last access: 22 July 2020) (Xia *et al*., 2012) via Google Earth Engine, Daymet precipitation data were accessed via https://ral.ucar.edu/solutions/products/camels and https://doi.org/10.3334/ORNLDAAC/1840 (last access: 22 July 2020) (Thornton *et al.*, 2014) via Google Earth Engine, PRISM precipitation data were accessed at https://doi.org/10.1371/journal.pone.0141140 (last access: 9 September 2019). SSEBop evapotranspiration data were accessed at https://earlywarning.usgs.gov/fews (last access: 22 July 2020) (Senay *et al.*, 2013), NCA-LDAS data were accessed at https://disc.gsfc.nasa.gov/ (last access: 10 July 2020) (Kumar *et al.*, 2019), and MODIS16 evapotranspiration data were accessed at https://www.ntsg.umt.edu/project/modis/mod16.php (last access: 2 February 2020) (Mu *et al.*, 2013) via Google Earth Engine. Potential Evapotranspiration data were accessed via https://ral.ucar.edu/solutions/products/camels and https://doi.org/10.1029/2010EO340001 (last access: 1 August 2020) (Xia *et al.*, 2012). Streamflow data were accessed at https://ral.ucar.edu/solutions/products/camels and https://waterdata.usgs.gov/nwis/rt (last access: June 20, 2020) (USGS, 2016). Snow fraction data were accessed at https://ral.ucar.edu/solutions/products/camels (last access: July 20, 2021).

3.8 References

Abolafia-Rosenzweig, R., Pan, M., Zeng, J. L., & Livneh, B. (2021). Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques. *Remote Sensing of Environment*, *252*, 112191. https://doi.org/10.1016/J.RSE.2020.112191

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., & Stoffelen, A. (2015). Characterization of precipitation product errors across the United States using multiplicative triple collocation. *Hydrology and Earth System Sciences*, *19*(8), 3489–3503. https://doi.org/10.5194/hess-19-3489-2015

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., *et al*. (2011). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences*, *15*(1), 223–239. https://doi.org/10.5194/hess-15-223-2011

Andréassian, V., & Perrin, C. (2012). On the ambiguous interpretation of the Turc-Budyko nondimensional graph. *Water Resources Research*, *48*(10). https://doi.org/10.1029/2012WR012532

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., *et al*. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society*, *96*(1), 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, *42*(8), 8432. https://doi.org/10.1029/2005WR004387

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., *et al*. (2008). Human-induced changes in the hydrology of the Western United States. *Science*, *319*(5866), 1080–1083. https://doi.org/10.1126/science.1152538

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., & Schneider, D. (2016). Snowmelt rate dictates streamflow. *Geophysical Research Letters*, *43*(15), 8006–8016. https://doi.org/10.1002/2016GL069690

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., *et al*. (2020). Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments. *Journal of Climate*, *33*(4), 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1

Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, *4*(7), 583–586. https://doi.org/10.1038/nclimate2246

Berghuijs, Wouter R., Gnann, S. J., & Woods, R. A. (2020). Unanswered questions on the Budyko framework. *Hydrological Processes*, hyp.13958. https://doi.org/10.1002/hyp.13958

Burnett, M. W., Quetin, G. R., & Konings, A. G. (2020). Data-driven estimates of evapotranspiration and its controls in the Congo Basin. *Hydrology and Earth System Sciences*, *24*(8), 4189–4211. https://doi.org/10.5194/hess-24-4189-2020

Buss, H. L., Brantley, S. L., Scatena, F. N., Bazilievskaya, E. A., Blum, A., Schulz, M., *et al*. (2013). Probing the deep critical zone beneath the luquillo experimental forest, Puerto Rico. *Earth Surface Processes and Landforms*. https://doi.org/10.1002/esp.3409

Carroll, R. W. H., Deems, J. S., Niswonger, R., Schumer, R., & Williams, K. H. (2019). The Importance of Interflow to Groundwater Recharge in a Snowmelt-Dominated Headwater Basin. *Geophysical Research Letters*, *46*(11), 5899–5908.

https://doi.org/10.1029/2019GL082447

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., & Aida, K. (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sensing of Environment*, *214*, 1–13. https://doi.org/10.1016/j.rse.2018.05.008

Condon, L. E., & Maxwell, R. M. (2017). Systematic shifts in Budyko relationships caused by groundwater storage changes. *Hydrol. Earth Syst. Sci*, *21*, 1117–1135. https://doi.org/10.5194/hess-21-1117-2017

Condon, L. E., Markovich, K. H., Kelleher, C. A., McDonnell, J. J., Ferguson, G., & McIntosh, J. C. (2020, March 1). Where Is the Bottom of a Watershed? *Water Resources Research*. Blackwell Publishing Ltd. https://doi.org/10.1029/2019WR026010

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., *et al*. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(22). https://doi.org/10.1029/2002jd003118

Crow, W. T., Han, E., Ryu, D., Hain, C. R., & Anderson, M. C. (2017). Estimating annual water storage variations in medium-scale (2000-10000km2) basins using microwave-based soil moisture retrievals. *Hydrology and Earth System Sciences*, *21*(3), 1849–1862. https://doi.org/10.5194/HESS-21-1849-2017

Daly, C., Taylor, G., & Gibson, W. (1997). *The Prism Approach to Mapping Precipitation and Temperature*.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., *et al*. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Dettinger, M., Redmond, K., & Cayan, D. (2004). Winter orographic precipitation ratios in the Sierra Nevada - Large-scale atmospheric circulations and hydrologic consequences. *Journal of Hydrometeorology*, *5*(6), 1102–1116. https://doi.org/10.1175/JHM-390.1

Dettinger, M. D., Cayan, D. R., Meyer, M. K., & Jeton, A. (2004). Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900-2099 *. *Climatic Change*, *62*(1–3), 283–317. https://doi.org/10.1023/B:CLIM.0000013683.13346.4f

Dong, J., Crow, W. T., Duan, Z., Wei, L., & Lu, Y. (2019). A double instrumental variable method for geophysical product error estimation. *Remote Sensing of Environment*, *225*, 217–228. https://doi.org/10.1016/j.rse.2019.03.003

Ehsani, N., Vörösmarty, C. J., Fekete, B. M., & Stakhiv, E. Z. (2017). Reservoir operations under climate change: Storage capacity options to mitigate risk. *Journal of Hydrology*, *555*, 435–446. https://doi.org/10.1016/j.jhydrol.2017.09.008

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., *et al*. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(22). https://doi.org/10.1029/2002jd003296

Enzminger, T. L., Small, E. E., & Borsa, A. A. (2019). Subsurface Water Dominates Sierra Nevada Seasonal Hydrologic Storage. *Geophysical Research Letters*, *46*(21), 11993–12001. https://doi.org/10.1029/2019GL084589

Fan, Y., & Schaller, M. F. (2009). River basins as groundwater exporters and importers: Implications for water cycle and climate modeling. *Journal of Geophysical Research Atmospheres*, *114*(4). https://doi.org/10.1029/2008JD010636

Fan, Ying. (2019). Are catchments leaky? *WIREs Water*, *6*(6). https://doi.org/10.1002/wat2.1386

Flechtner, F., Morton, P., Watkins, M., & Webb, F. (2014). Status of the GRACE Follow-On mission. *International Association of Geodesy Symposia*, *141*, 117–121. https://doi.org/10.1007/978-3-319-10837-7_15/COVER

Flinchum, B. A., Holbrook, W. S., Grana, D., Parsekian, A. D., Carr, B. J., Hayes, J. L., & Jiao, J. (2018). Estimating the water holding capacity of the critical zone using near-surface geophysics. *Hydrological Processes*, *32*(22), 3308–3326. https://doi.org/10.1002/hyp.13260

Frisbee, M. D., Tysor, E. H., Stewart-Maddox, N. S., Tsinnajinnie, L. M., Wilson, J. L., Granger, D. E., & Newman, B. D. (2016). Is there a geomorphic expression of interbasin groundwater flow in watersheds? Interactions between interbasin groundwater flow, springs, streams, and geomorphology. *Geophysical Research Letters*, *43*(3), 1158–1165. https://doi.org/10.1002/2015GL067082

Genereux, D. P., Nagy, L. A., Osburn, C. L., & Oberbauer, S. F. (2013). A connection to deep groundwater alters ecosystem carbon fluxes and budgets: Example from a Costa Rican rainforest. *Geophysical Research Letters*. https://doi.org/10.1002/grl.50423

Gleeson, T., Moosdorf, N., Hartmann, J., & Van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*. https://doi.org/10.1002/2014GL059856

Gordon, B. L., Paige, G. B., Miller, S. N., Claes, N., & Parsekian, A. D. (2020). Field scale quantification indicates potential for variability in return flows from flood irrigation in the high altitude western US. *Agricultural Water Management*, *232*, 106062. https://doi.org/10.1016/j.agwat.2020.106062

Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrame, G. F. S., Carroll, R. W. H., McNamara, J. P., & Harpold, A. A. (2022). Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework. *Environmental Research Letters*, *17*(5), 053004. https://doi.org/10.1088/1748-

9326/AC64B4

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Greve, P., Burek, P., & Wada, Y. (2020). Using the Budyko Framework for Calibrating a Global Hydrological Model. *Water Resources Research*, *56*(6), e2019WR026280. https://doi.org/10.1029/2019WR026280

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 200–211. https://doi.org/10.1016/j.jag.2015.09.002

Gruber, Alexander, Dorigo, W. A., Crow, W., & Wagner, W. (2017). Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), 6780–6792. https://doi.org/10.1109/TGRS.2017.2734070

Hahm, W. J., Dralle, D. N., Rempe, D. M., Bryk, A. B., Thompson, S. E., Dawson, T. E., & Dietrich, W. E. (2019). Low Subsurface Water Storage Capacity Relative to Annual Rainfall Decouples Mediterranean Plant Productivity and Water Use From Rainfall Variability. *Geophysical Research Letters*, *46*(12), 6544–6553. https://doi.org/10.1029/2019GL083294

Hahm, W Jesse, Rempe, D. M., Dralle, D. N., Dawson, T. E., Lovill, S. M., Bryk, A. B., *et al*. (2019). Lithologically Controlled Subsurface Critical Zone Thickness and Water Storage Capacity Determine Regional Plant Community Composition. *Water Resources Research*, *55*(4), 3028–3055. https://doi.org/10.1029/2018WR023760

Hamilton, A. S., & Moore, R. D. (2012). Quantifying Uncertainty in Streamflow Records. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *37*(1), 3–21. https://doi.org/10.4296/cwrj3701865

Hammond, J. C., & Kampf, S. K. (2020). Subannual Streamflow Responses to Rainfall and Snowmelt Inputs in Snow-Dominated Watersheds of the Western United States. *Water Resources Research*, *56*(4). https://doi.org/10.1029/2019WR026132

Han, P. F., Istanbulluoglu, E., Wan, L., & Wang, X. S. (2021). A New Hydrologic Sensitivity Framework for Unsteady-State Responses to Climate Change and Its Application to Catchments With Croplands in Illinois. *Water Resources Research*, *57*(8), e2020WR027762. https://doi.org/10.1029/2020WR027762

Harpold, A., Brooks, P., Rajagopal, S., Heidbuchel, I., Jardine, A., & Stielstra, C. (2012). Changes in snowpack accumulation and ablation in the intermountain west. *Water Resources Research*, *48*(11). https://doi.org/10.1029/2012WR011949

He, C., Chen, F., Barlage, M., Liu, C., Newman, A., Tang, W., *et al*. (2019). Can Convection-Permitting Modeling Provide Decent Precipitation for Offline High-

Resolution Snowpack Simulations Over Mountains? *Journal of Geophysical Research: Atmospheres*, *124*(23), 12631–12654. https://doi.org/10.1029/2019JD030823

Henn, B., Newman, A. J., Livneh, B., Daly, C., & Lundquist, J. D. (2018). An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, *556*(March), 1205–1219. https://doi.org/10.1016/j.jhydrol.2017.03.008

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., *et al*. (2020). The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3803

Higgins, R. W., Schemm, J. K. E., Shi, W., & Leetmaa, A. (2000). Extreme precipitation events in the Western United States related to tropical forcing. *Journal of Climate*, *13*(4), 793–820. https://doi.org/10.1175/1520-0442(2000)013<0793:EPEITW>2.0.CO;2

Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2012). A Lego System for Conditional Inference. *Http://Dx.Doi.Org/10.1198/000313006X118430*, *60*(3), 257–263. https://doi.org/10.1198/000313006X118430

Hothorn, T., Hornik, K., & Zeileis, A. (2012). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Http://Dx.Doi.Org/10.1198/106186006X133933*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Hothorn, T., Zeileis, A., Cheng, E. :, & Ong, S. (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*, *16*, 3905–3909. Retrieved from http://cran.r-project.org/package=

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., *et al*. (2020). Importance and vulnerability of the world's water towers. *Nature*, *577*(7790), 364–369. https://doi.org/10.1038/s41586-019-1822-y

Istanbulluoglu, E., Wang, T., Wright, O. M., & Lenters, J. D. (2012). Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis. *Water Resources Research*, *48*(3). https://doi.org/10.1029/2010WR010100

Jing, X., Geerts, B., Wang, Y., & Liu, C. (2017). Evaluating seasonal orographic precipitation in the interior western United States using gauge data, gridded precipitation estimates, and a regional climate simulation. *Journal of Hydrometeorology*, *18*(9), 2541–2558. https://doi.org/10.1175/JHM-D-17-0056.1

Jones, J. A., Creed, I. F., Hatcher, K. L., Warren, R. J., Adams, M. B., Benson, M. H., *et al*. (2012). Ecosystem Processes and Human Influences Regulate Streamflow Response to Climate Change at Long-Term Ecological Research Sites. *BioScience*, *62*(4), 390–404. https://doi.org/10.1525/BIO.2012.62.4.10

Kampf, S. K., Burges, S. J., Hammond, J. C., Bhaskar, A., Covino, T. P., Eurich, A., *et al*. (2020, June 1). The Case for an Open Water Balance: Re-envisioning Network Design and Data Analysis for a Complex, Uncertain World. *Water Resources Research*. Blackwell Publishing Ltd. https://doi.org/10.1029/2019WR026699

Käser, D., & Hunkeler, D. (2016). Contribution of alluvial groundwater to the outflow of mountainous catchments. *Water Resources Research*. https://doi.org/10.1002/2014WR016730

Khan, M. S., Liaqat, U. W., Baik, J., & Choi, M. (2018). Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. *Agricultural and Forest Meteorology*, *252*, 256–268. https://doi.org/10.1016/j.agrformet.2018.01.022

Koppa, A., & Gebremichael, M. (2017). A Framework for Validation of Remotely Sensed Precipitation and Evapotranspiration Based on the Budyko Hypothesis. *Water Resources Research*, *53*(10), 8487–8499. https://doi.org/10.1002/2017WR020593

Kumar, S. V., Jasinski, M., Mocko, D. M., Rodell, M., Borak, J., Li, B., *et al*. (2019). NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment. *Journal of Hydrometeorology*, *20*(8), 1571–1593. https://doi.org/10.1175/JHM-D-17-0125.1

Li, Dan, Pan, M., Cong, Z., Zhang, L., & Wood, E. (2013). Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, *49*(2), 969–976. https://doi.org/10.1002/wrcr.20107

Li, Dongyue, Wrzesien, M. L., Durand, M., Adam, J., & Lettenmaier, D. P. (2017). How much runoff originates as snow in the western United States, and how will that change in the future? *Geophysical Research Letters*, *44*(12), 6163–6172. https://doi.org/10.1002/2017GL073551

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, *100*(12), 2473–2490. https://doi.org/10.1175/BAMS-D-19-0001.1

Lundquist, J., Dickerson-Lange, S., Gutmann, E., Jonas, T., Lumbrazo, C., & Reynolds, D. (2021). Snow Interception Modeling: Isolated Observations have led to Land Surface Models Lacking Appropriate Climate Sensitivities. *Authorea Preprints*. https://doi.org/10.22541/AU.161443094.42414018/V1

Lundquist, J. D., Neiman, P. J., Martner, B., White, A. B., Gottas, D. J., & Ralph, F. M. (2008). Rain versus Snow in the Sierra Nevada, California: Comparing Doppler Profiling Radar and Surface Observations of Melting Level. *Journal of Hydrometeorology*, *9*(2), 194–211. https://doi.org/10.1175/2007JHM853.1

Maggioni, V., Sapiano, M. R. P., & Adler, R. F. (2016). Estimating uncertainties in high-resolution satellite precipitation products: Systematic or Random Error? *Journal of Hydrometeorology*, *17*(4), 1119–1129. https://doi.org/10.1175/JHM-D-15-0094.1

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., & Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*, *10*(11), 114016. https://doi.org/10.1088/1748-9326/10/11/114016

Massari, C., Crow, W., & Brocca, L. (2017). An assessment of the performance of global rainfall estimates without ground-based observations. *Hydrol. Earth Syst. Sci*, *21*, 4347–4361. https://doi.org/10.5194/hess-21-4347-2017

Maxwell, R. M., & Condon, L. E. (2016). Connections between groundwater flow and transpiration partitioning. *Science*, *353*(6297), 377–380. https://doi.org/10.1126/science.aaf7891

McCabe, G. J., Wolock, D. M., & Valentin, M. (2018). Warming is driving decreases in snow fractions while runoff efficiency remains mostly unchanged in snow-covered areas of the western United States. *Journal of Hydrometeorology*, *19*(5), 803–814. https://doi.org/10.1175/JHM-D-17-0227.1

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters*, *41*(17), 6229–6236. https://doi.org/10.1002/2014GL061322

Milly, P. C. D., & Dunne, K. A. (2020). Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. *Science*, *367*(6483), 1252–1255. https://doi.org/10.1126/science.aax0194

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019

Mu, Q., Zhao, M., & Running, S. W. (2013). *MODIS Global Terrestrial Evapotranspiration (ET) Product*.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., *et al*. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci*, *19*, 209–223. https://doi.org/10.5194/hess-19-209-2015

Ni, G., Yang, D., Zhang, D., Cong, Z., Ni, G., Yang, D., & Hu, S. (2015). Effects of snow ratio on annual runoff within the Budyko framework. *Hydrol. Earth Syst. Sci*, *19*. https://doi.org/10.5194/hess-19-1977-2015

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., *et al*. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*, *116*(12). https://doi.org/10.1029/2010JD015139

NLDAS-2 Forcing Dataset Information | LDAS. (n.d.). Retrieved August 21, 2020, from https://ldas.gsfc.nasa.gov/nldas/v2/forcing

Padrón, R. S., Gudmundsson, L., Greve, P., & Seneviratne, S. I. (2017). Large-Scale Controls of the Surface Water Balance Over Land: Insights From a Systematic Review and Meta-Analysis. *Water Resources Research*, *53*(11), 9659–9678. https://doi.org/10.1002/2017WR021215

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., *et al*. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems*, *8*(1), 41–65. https://doi.org/10.1002/2015MS000526

Peng, L., Li, D., & Sheffield, J. (2018). Drivers of Variability in Atmospheric Evaporative Demand: Multiscale Spectral Analysis Based on Observations and Physically Based Modeling. *Water Resources Research*, *54*(5), 3510–3529. https://doi.org/10.1029/2017WR022104

Polhamus, A., Fisher, J. B., & Tu, K. P. (2013). What controls the error structure in evapotranspiration models? *Agricultural and Forest Meteorology*, *169*, 12–24. https://doi.org/10.1016/j.agrformet.2012.10.002

Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A., & Jakeman, A. J. (2005). Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. *Water Resources Research*, *41*(6), 1–11. https://doi.org/10.1029/2004WR003697

Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., *et al*. (2020). Agricultural risks from changing snowmelt. *Nature Climate Change*, *10*(5), 459–465. https://doi.org/10.1038/s41558-020-0746-8

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., *et al*. (2012). How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bulletin of the American Meteorological Society*, *93*(6), 811–829. https://doi.org/10.1175/BAMS-D-11-00052.1

Rui, H., & Mocko, D. (2018). *README for NCA-LDAS Version 2.0 Data Product 2*.

Safeeq, M., Bart, R. R., Pelak, N. F., Singh, C. K., Dralle, D. N., Hartsough, P., & Wagenbrenner, J. W. (2021). How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds. *Hydrological Processes*. https://doi.org/10.1002/hyp.14199

Schmidt, L., & Rempe, D. (2020). Quantifying Dynamic Water Storage in Unsaturated Bedrock with Borehole Nuclear Magnetic Resonance. *Geophysical Research Letters*, *47*(22). https://doi.org/10.1029/2020GL089600

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., & Verdin, J. P. (2013). Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. *Journal of the American Water Resources Association*, *49*(3), 577–591. https://doi.org/10.1111/jawr.12057

Smith, R. G., Knight, R., Chen, J., Reeves, J. A., Zebker, H. A., Farr, T., & Liu, Z. (2017). Estimating the permanent loss of groundwater storage in the southern San Joaquin Valley, California. *Water Resources Research*, *53*(3), 2133–2148. https://doi.org/10.1002/2016WR019861

Soroosh Sorooshian, Hsu, K., Braithwaite, D., Ashouri, H., & Program, N. C. (2014). PERSIANN-CDR, 4–6. https://doi.org/10.7289/V51V5BWQ

Sposito, G. (2017). Understanding the Budyko Equation. *Water*, *9*(4), 236. https://doi.org/10.3390/w9040236

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, *103*(3334), 7755–7766. https://doi.org/10.1029/97jc03180

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505. https://doi.org/10.1126/science.1099192

Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, *190*(3–4), 214–251. https://doi.org/10.1016/S0022-1694(96)03128-9

Tugrul Yilmaz, M., & Crow, W. T. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology*, *15*(3), 1293–1302. https://doi.org/10.1175/JHM-D-13-0158.1

Valéry, A., Andréassian, V., & Perrin, C. (2010). Regionalization of precipitation and air temperature over high-altitude catchments – learning from outliers. *Hydrological Sciences Journal*, *55*(6), 928–940. https://doi.org/10.1080/02626667.2010.504676

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., & Verdin, J. P. (2013). A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sensing of Environment*, *139*, 35–49. https://doi.org/10.1016/j.rse.2013.07.013

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*, *43*(7), 7447. https://doi.org/10.1029/2006WR005653

Wang, T., Istanbulluoglu, E., Lenters, J., Scott, D., Wang, T., Istanbulluoglu, E., *et al*. (2009). On the role of groundwater and soil texture in the regional water balance: An investigation of the Nebraska Sand Hills, USA. *Water Resources Research*, *45*(10). https://doi.org/10.1029/2009WR007733

Welch, L. A., & Allen, D. M. (2012). Consistency of groundwater flow patterns in mountainous topography: Implications for valley bottom water replenishment and for defining groundwater flow boundaries. *Water Resources Research*, *48*(5). https://doi.org/10.1029/2011WR010901

Wrzesien, M. L., Durand, M. T., & Pavelsky, T. M. (2019). A Reassessment of North American River Basin Cool-Season Precipitation: Developments From a New Mountain Climatology Data Set. *Water Resources Research*, *55*(4), 3502–3519. https://doi.org/10.1029/2018WR024106

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., *et al*. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research Atmospheres*, *117*(3), 3109. https://doi.org/10.1029/2011JD016048

Xu, C. Y., & Singh, V. P. (2002). Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. *Water Resources Management*, *16*(3), 197–219. https://doi.org/10.1023/A:1020282515975

Yang, L., Meng, X., & Zhang, X. (2011). SRTM DEM and its application advances. *International Journal of Remote Sensing*, *32*(14), 3875–3896. https://doi.org/10.1080/01431161003786016

Yilmaz, M. T., Crow, W. T., Anderson, M. C., & Hain, C. (2012). An objective methodology for merging satellite- and model-based soil moisture products. *Water Resources Research*, *48*(11). https://doi.org/10.1029/2011WR011682

3.9 Supplemental Information

Text S3.1: Spatial and Temporal Upscaling

All raw data for *P* and *ET* products were upscaled to an 8 day temporal resolution to match with the MODIS16 product resolution. For products with a daily temporal resolution, missing data for more than 3 days within the 8 day period resulted in the 8 day aggregate being considered ‘not available’ in the resulting timeseries. On average, *P* and *ET* triplets were comprised of 581 of a maximum 689 collocated data points with a standard deviation of 85 points.

Text S3.2: Data

<u>ERA5 *P*</u>

ERA5 is the fifth atmospheric reanalysis product generated by the European Centre for Medium-Range Weather Forecasts (ECMWF), which combines model data with global observations to create a complete and consistent record of selected variables, including precipitation (Hersbach *et al*., 2020). It is based on the Integrated Forecasting System Cy41r2, which leverages developments in model physics, data assimilation, and core dynamics. Details about the spatial and temporal resolution of ERA5 and all other datasets are provided in Table 2 below. For this analysis, we used the daily aggregate ERA5 product available through Google Earth Engine (Gorelick *et al*., 2017) and extracted the mean daily total precipitation for each candidate catchment from October 1, 2001 to September 30, 2016.

PERSIANN-CDR *P*

PERSIANN-CDR is a continuous, long-term precipitation data product that is generated using gridded satellite (GridSat-B1) infrared data (Ashouri *et al*., 2015). PERSIANN-CDR is adjusted using the Global Precipitation Climatology Project monthly product to ensure consistency in the data sets. Like the ERA5 product, the mean daily total precipitation from PERSIANN-CDR was extracted for each candidate catchment from October 1, 2001 to September 30, 2016 using Google Earth Engine.

NLDAS-2 P

NLDAS-2 precipitation data set is an hourly product based on temporal disaggregation of Climate Prediction Center CONUS gauge data (Cosgrove *et al*., 2003; Higgins *et al*., 2000; "NLDAS-2 Forcing Dataset Information | LDAS," 2020), CPC hourly gauge data, hourly Doppler Stage II radar precipitation data, half-hourly CMORPH data, and 3-hourly North

American Regional Reanalysis precipitation data. A more complete account of the NLDAS-2 precipitation forcing data are provided by NASA (Xia *et al*., 2012). NLDAS-2 precipitation data were obtained from the CAMELS database for October 1, 2001 to December 31, 2014. Precipitation data from 2015-2016 were obtained from Google Earth Engine following the same methods as for ERA5 and PERSIANN-CDR.

Daymet P

Daymet product provides gridded weather parameters for North America and includes continuous daily precipitation (Thornton *et al*., 1997). The Daym*et al*gorithm uses ground observations from meteorological stations throughout the United States sourced from the Cooperative Summary of the Day network run by the National Climate Data Center (NCDC) and the SNOwpack and TELemetry (SNOTEL) dataset managed by the Natural Resources Conservation Service (NRCS) (Thornton *et al*., 1997). Daymet additionally requires a digital elevation model and land mask. Daymet precipitation data were obtained using the same methodology as above.

PRISM P

PRISM generates gridded estimates of climatic parameters using a combination of point data, elevation models, and spatial datasets (Daly *et al*., 1997). PRISM data are generated using a process called climatologically aided interpolation (CAI) and in some cases, Doppler radar data. Within the PRISM model Point data for precipitation are taken from 19 different networks, which include SNOTEL and NOAA's Cooperative Observer Network (COOP) among others. A station-weighted climate-elevation regression is calculated for each 4 km grid cell across CONUS, with weighting based on station characteristics like elevation, position, and orographic effectiveness of terrain. For more details, we refer to Daly *et al*. (2008). Daily PRISM precipitation data were obtained from October 1, 2001 to September 30, 2016 using Google Earth Engine as above.

ALEXI ET

ALEXI maps ET using multi-sensor thermal infrared (TIR) remote sensing of LST (Anderson *et al*., 2011). ALEXI couples a two-source (soil and canopy) land-surface model with an atmospheric boundary layer model to map daily fluxes in canopy transpiration and soil evaporation (combined in ET) across CONUS at 5 to 10 km resolution. Daily ALEXI ET data were obtained from the National Aeronautics and Space Administration (NASA) and the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) and extracted for each candidate catchment from October 1, 2001 to September 30, 2016.

SSEBop ET

SSEBop combines remotely sensed thermal imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) with reference ET. Data are parameterized using pre-defined, seasonally dynamic boundary conditions for each pixel. SSEBop estimates transpiration and soil evaporation (combined in ET) every 8 days (Senay *et al*., 2013) at a 1 km resolution. Daily SSEBop ET data were downloaded from the United States Geological Survey and extracted for each candidate catchment from October 1, 2001 to September 30, 2016 using the same methods as described for ALEXI.

NCA-LDAS ET

NCA-LDAS couples the Noah model (Ek *et al*., 2003; Niu *et al*., 2011) with the Weather Research Forecasting (WRF) regional atmospheric model, the NOAA coupled Climate Forecast System, and the Global Forecast System (GFS) (Rui & Mocko, 2018). Daily data, including evapotranspiration, are simulated using the Noah-3.3 LSM and mapped to a grid with 12 km spacing. NCA-LDAS data measures canopy transpiration and soil evapotranspiration (combined in ET). ET data were downloaded from the Goddard Earth Sciences Data and Information Services Center (GES DISC) and extracted for each candidate catchment from October 1, 2001 to September 30, 2016 using the method described above.

MOD16 ET

MOD16 is based on the Penman-Monteith equation and uses daily meteorological reanalysis data from NASA Global Modeling and Assimilation Office (GMAO) combined with MODIS products for vegetation characteristics, land cover, and albedo (Mu *et al*., 2013). MOD16 produces 8-day soil evaporation and canopy transpiration (combined in *ET*) at a 1km resolution. 8-day MOD16 data for each catchment from October 1, 2001 to September 30, 2016 were obtained using Google Earth Engine.

NLDAS-2 E_o

Although there are different ways to estimate E_o (Xu & Singh, 2002) that can affect the absolute aridity index values across catchments although uncertainty in E_o is not the focus of this study. Because not all above ET datasets contained a readily available estimate of E_o, we elected to use a common E_o from NLDAS-2 because it is a Penman-based calculation which is understood to estimate apparent atmospheric demand (Peng *et al*., 2018). Daily E_o values for all candidate watersheds were obtained from the CAMELS database as described above and were aggregated to obtain an annual value.

USGS *Q*

Daily values for Q from October 1, 2001 to December 31, 2014 were obtained from the CAMELS database for each catchment. More details are provided in Addor *et al*. (2017). Daily values for Q from January 1, 2015 to September 30, 2016 were obtained from the USGS streamflow database (USGS, 2020). While *Q* is also subject to uncertainty, it is widely presumed to be the most certain water budget component (Bales *et al*., 2006) with an assumed uncertainty ($\pm$) 5% at the 95% confidence interval (CI) for gauged streamflow in North America (Hamilton & Moore, 2012).

Snow Fraction

Mean snow fraction (*fs*) was obtained from the CAMELS database for each catchment. More details are provided in Addor *et al*. (2017), *fs* values use Daymet data and leverage pervious work (Newman *et al*., 2015). We use the CAMELS *fs* dataset in the main manuscript but tested the *fs* using the same methodology as in the CAMELS database with Daymet precipitation and temperature data from 2001-2016, which had no significant impact on our findings.

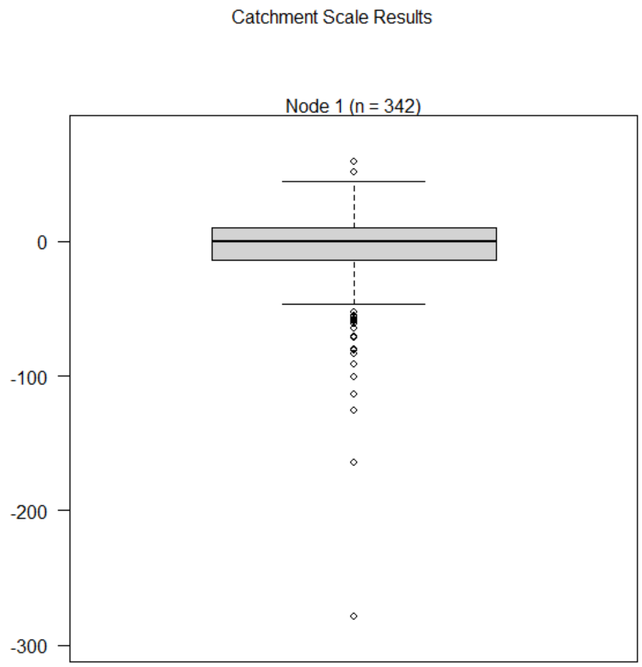

Figure S3-1: Results for unbiased recursive partitioning between OWB closure and Fan (2019) Criterion 1 (Catchment Scale) in the main Chapter. Here we use observed OWB closure using all three products (n = 342). To account for biases, we adjust observed OWB closure by the median observed OWB closure across all sites (-12.8%).

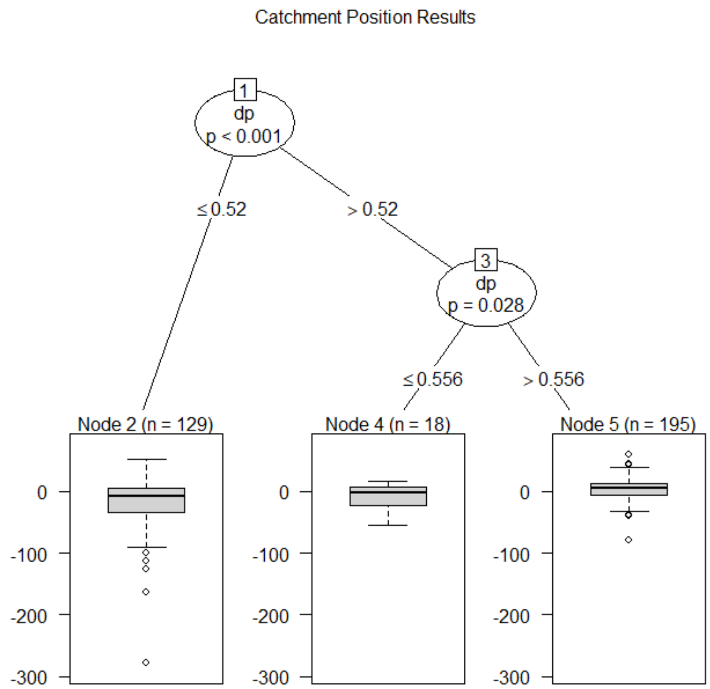

Figure S3-2: Results for unbiased recursive partitioning between OWB closure and Fan (2019) Criterion 2 (Catchment Position) in the main manuscript. Here we use observed OWB closure using all three products (n = 342). To account for biases, we adjust observed OWB closure by the median observed OWB closure across all sites (-12.8%).

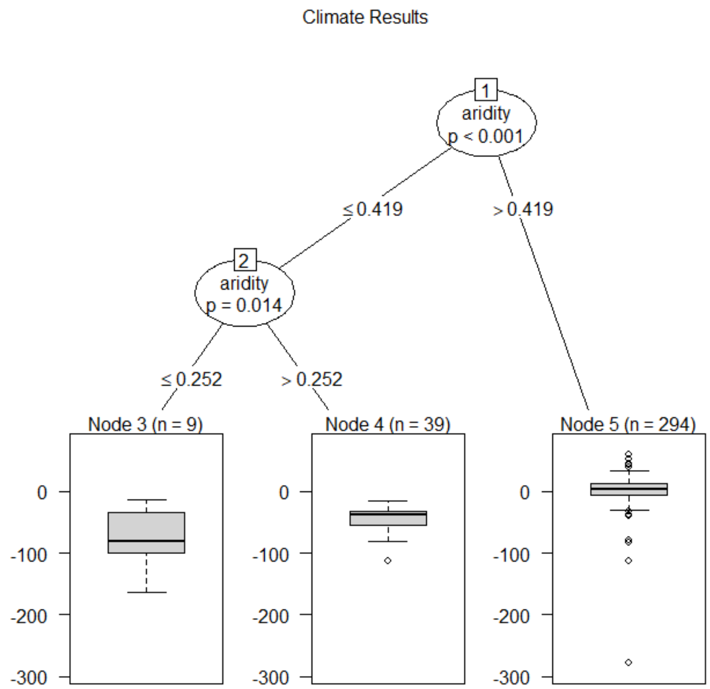

Figure S3-3: Results for unbiased recursive partitioning between OWB closure and Fan (2019) Criterion 3 (Climate) in the main manuscript. Here we use observed OWB closure using all three products (n = 342). To account for biases, we adjust observed OWB closure by the median observed OWB closure across all sites (-12.8%).

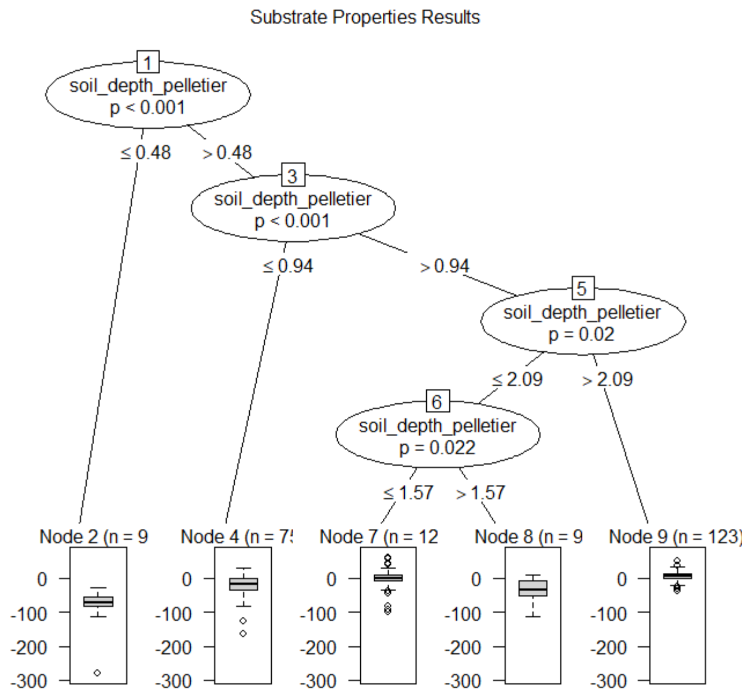

Figure S3-4: Results for unbiased recursive partitioning between OWB closure and Fan (2019) Criterion 4 (Substrate Properties) in the main manuscript. Here we use observed OWB closure using all three products (n = 342). To account for biases, we adjust observed OWB closure by the median observed OWB closure across all sites (-12.8%).

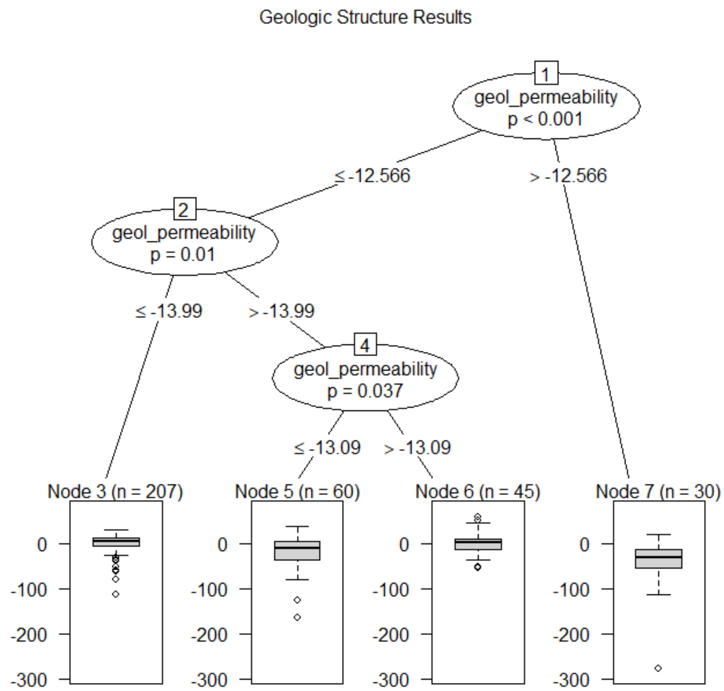

Figure S3-5: Results for unbiased recursive partitioning between OWB closure and Fan (2019) Criterion 4 (Geological Structure) in the main manuscript. Here we use observed OWB closure using all three products (n = 342). To account for biases, we adjust observed OWB closure by the median observed OWB closure across all sites (-12.8%).

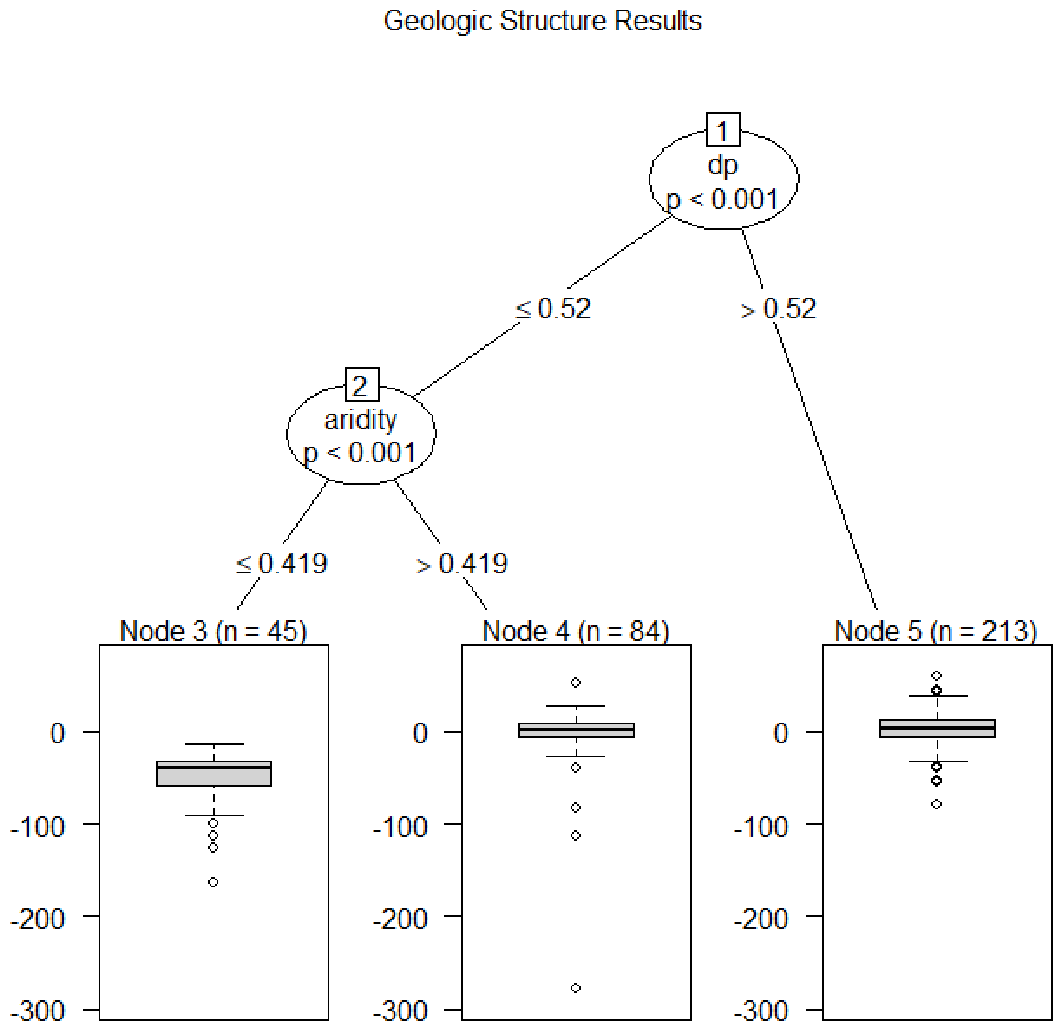

Figure S3-6: Results for unbiased recursive partitioning between OWB closure and all Fan (2019) in the main manuscript. Here we use observed OWB closure using all three products (n = 342). To account for biases, we adjust observed OWB closure by the median observed OWB closure across all sites (-12.8%).

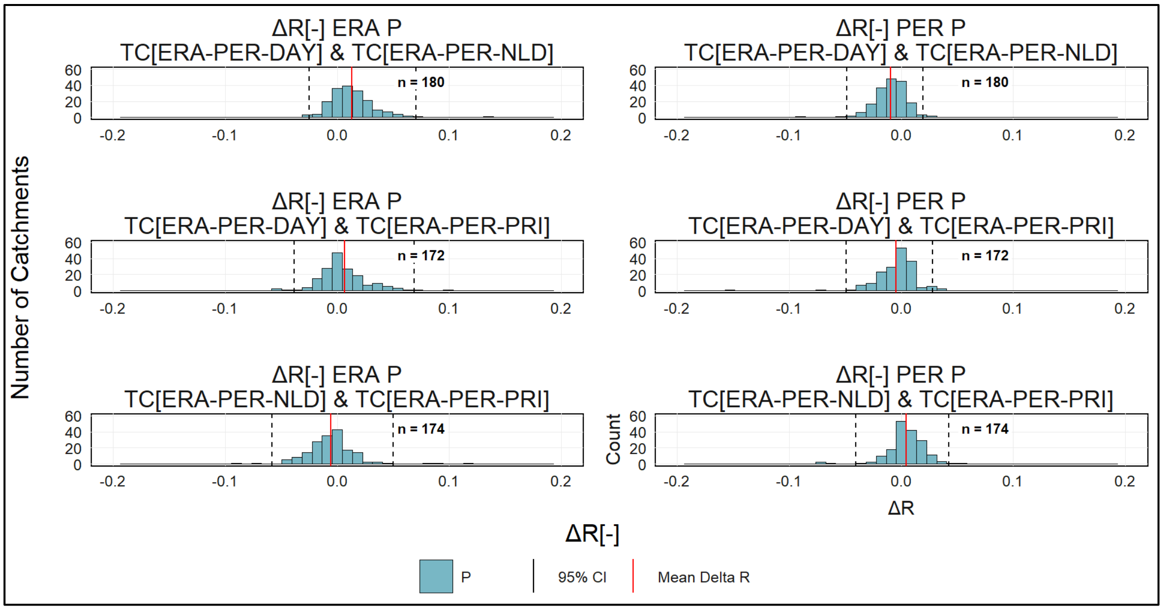

Figure S3-7: ΔR histograms for common elements in constructed ET Triplets and 95% CI.

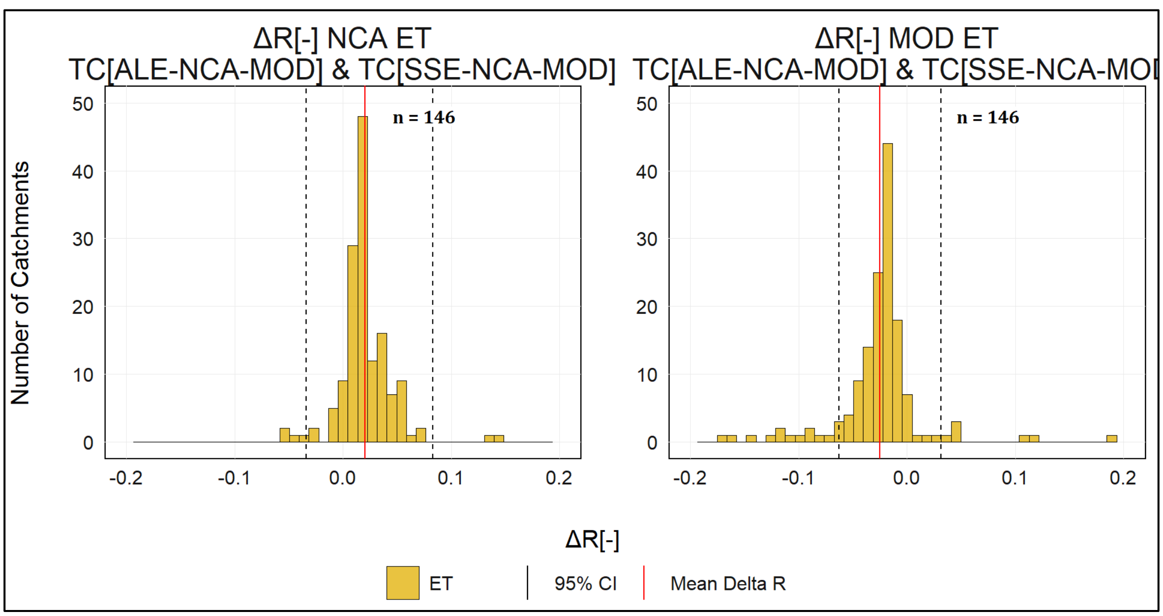

Figure S3-8: ΔR histograms for common elements in constructed ET Triplets and 95% CI.

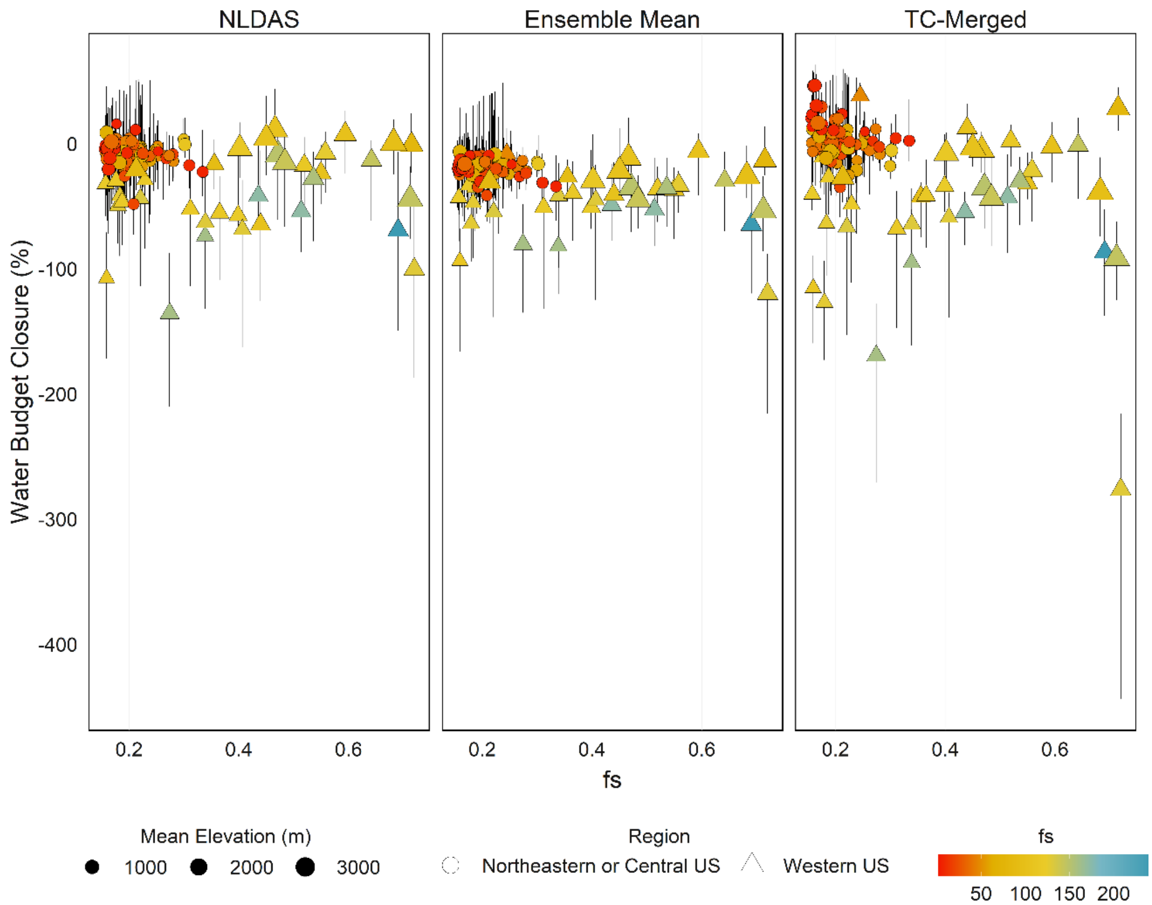

Figure S3-9: Fraction of precipitation as snowfall versus min, max, and mean (ε + ΔS)/P shaded by mean slope (m/km) and sized by elevation. Circles indicate catchments within the northeastern and central US regions (see Figure 1 in manuscript for definition) and diamonds indicate catchments in the western US region (see Figure 1 in manuscript for definition).

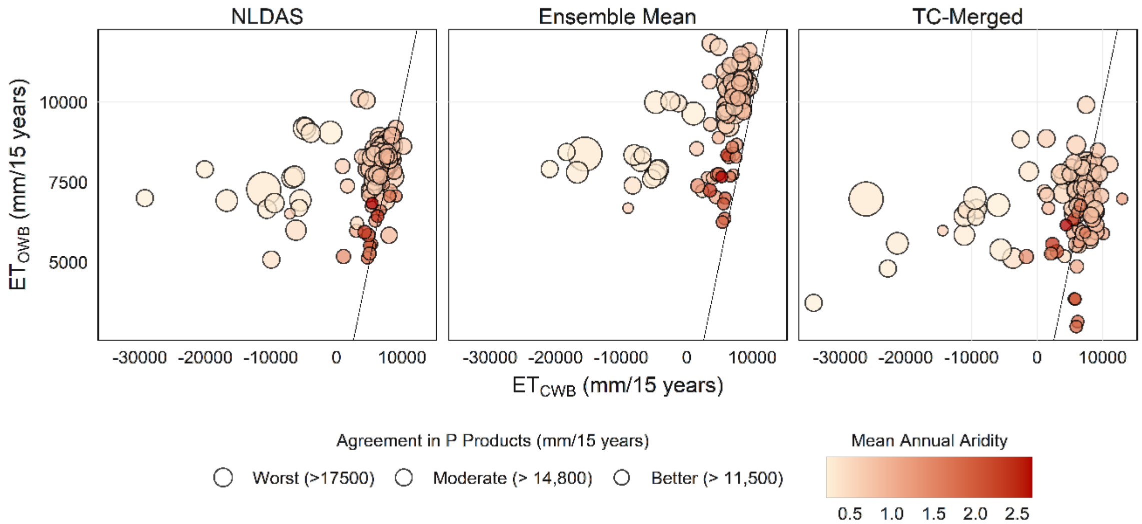

Figure S3-10: Long-term ET_{CWB} versus ET_{OWB} across all catchments (n = 114) using: A) NLDAS P & ET and USGS Q; B) Ensemble Mean P & ET and USGS Q; and C) TC-Merged P & ET and USGS Q. Coloring is based on observed on aridity. Sizing is based on the maximum disagreement between long-term estimates of P.

Table S3-1: Summary of candidate catchments per hydrologic region; ecoregion. * Indicates Temperate. State abbreviations are used where appropriate.

Catchment Count by Hydrologic Region													
AWR	CA	GB	GL	LC	MA	MO	NE	OH	PNW	RG	SRR	UC	UM
2	12	18	28	8	32	34	23	15	61	7	8	17	3
Catchment Count by Ecoregion													
Eastern Forest*	Great Plain	Marine West Coast Forest	North American Deserts	Northern Forests	North-western Forests	Sierra*							
41	18	3	14	63	122	7							

Table S3-2: Summary of ETC success in candidate catchments broken out by Hydrologic Region.

Hydrologic Region	# of Valid Catchments	# of Candidate Catchments	% Valid Catchments	US Region
Great Basin Region	1	18	5.6	Western
Upper Colorado Region	2	17	11.8	Western
California Region	1	12	8.3	Western
Lower Colorado Region	1	8	12.5	Western
Pacific Northwest Region	27	61	44.2	Western
Rio Grande Region	5	7	71.4	Western
Souris-Red-Rainy Region	1	8	12.5	Central
Missouri Region	5	34	14.7	Central
Great Lakes Region	12	28	42.9	Central
Upper Mississippi Region	3	3	100.0	Central
Arkansas White Red	0	2	0.0	Central
New England Region	11	23	47.8	Northeastern
Mid-Atlantic Region	30	32	93.8	Northeastern
Ohio Region	15	15	100.0	Northeastern

Table S3-3: Summary of ETC success in candidate catchments broken out by Ecoregion.

Ecoregion	# of Valid Catchments	# of Candidate Catchments	% Valid Catchments
North American Deserts	0	14	0
Temperate Sierras	1	7	14.3
Great Plains	5	18	27.8
Northwestern Forested Mountains	35	122	28.7
Northern Forests	40	63	63.5
Marine West Coast Forest	2	3	66.7
Eastern Temperate Forests	31	41	75.6

Supplemental References

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., *et al*. (2011). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences*, *15*(1), 223–239. https://doi.org/10.5194/hess-15-223-2011

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., *et al*. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society*, *96*(1), 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, *42*(8), 8432. https://doi.org/10.1029/2005WR004387

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., *et al*. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(22). https://doi.org/10.1029/2002jd003118

Daly, C., Taylor, G., & Gibson, W. (1997). *The Prism Approach to Mapping Precipitation and Temperature*.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., *et al*. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., *et al*. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(22). https://doi.org/10.1029/2002jd003296

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hamilton, A. S., & Moore, R. D. (2012). Quantifying Uncertainty in Streamflow Records. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *37*(1), 3–21. https://doi.org/10.4296/cwrj3701865

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., *et al*. (2020). The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3803

Higgins, R. W., Schemm, J. K. E., Shi, W., & Leetmaa, A. (2000). Extreme precipitation events in the Western United States related to tropical forcing. *Journal of Climate*, *13*(4), 793–820. https://doi.org/10.1175/1520-0442(2000)013<0793:EPEITW>2.0.CO;2

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters*, *41*(17), 6229–6236. https://doi.org/10.1002/2014GL061322

Mu, Q., Zhao, M., & Running, S. W. (2013). *MODIS Global Terrestrial Evapotranspiration (ET) Product*.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., *et al*. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci*, *19*, 209–223. https://doi.org/10.5194/hess-19-209-2015

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., *et al*. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*, *116*(12). https://doi.org/10.1029/2010JD015139

NLDAS-2 Forcing Dataset Information | LDAS. (2020). Retrieved August 21, 2020, from https://ldas.gsfc.nasa.gov/nldas/v2/forcing

Peng, L., Li, D., & Sheffield, J. (2018). Drivers of Variability in Atmospheric Evaporative Demand: Multiscale Spectral Analysis Based on Observations and Physically Based Modeling. *Water Resources Research*, *54*(5), 3510–3529. https://doi.org/10.1029/2017WR022104

Rui, H., & Mocko, D. (2018). *README for NCA-LDAS Version 2.0 Data Product 2*.

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., & Verdin,

J. P. (2013). Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. *Journal of the American Water Resources Association*, *49*(3), 577–591. https://doi.org/10.1111/jawr.12057

Soroosh Sorooshian, Hsu, K., Braithwaite, D., Ashouri, H., & Program, N. C. (2014). PERSIANN-CDR, 4–6. https://doi.org/10.7289/V51V5BWQ

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, *103*(3334), 7755–7766. https://doi.org/10.1029/97jc03180

Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, *190*(3–4), 214–251. https://doi.org/10.1016/S0022-1694(96)03128-9

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., *et al*. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research Atmospheres*, *117*(3), 3109. https://doi.org/10.1029/2011JD016048

Xu, C. Y., & Singh, V. P. (2002). Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. *Water Resources Management*, *16*(3), 197–219. https://doi.org/10.1023/A:1020282515975

4 Chapter 4: Water Management Can Reduce Agricultural Vulnerability to Decreasing Snowpack

By: Beatrice L. Gordon, Gabrielle F.S. Boisrame, Newsha K. Ajami, Rosemary W.H. Carroll, Bryan Leonard, Christine Albano, Naoki Mizukami, Alejandro Andrade-Rodriguez, Elizabeth Koebele, Adrian A. Harpold

Abstract:

By focusing on physical changes in mountain water supplies as snow declines, prior research has shown high potential for catastrophic damage to downstream water users and ecosystems (i.e., vulnerability) in certain river basins in the western US. However, humans can also modify the hydrological cycle via adaptation (socio-hydrology), altering the distribution and magnitude of vulnerability in this region. Here, we present a new paradigm for indexing the susceptibility of agricultural systems to damages arising from declines in snowpack at an operational (e.g., district or water user group) scale and test this approach in 13 basins with declining snowpack in the western US. Each of these basins relies on both snow storage and reservoir storage to meet agricultural production and contains no impairment above their reservoirs. Research evaluates ability of each basin to adapt to projected declines in snow storage using two different strategies : 1) enhancing reservoir or groundwater storage capacity via tools like managed aquifer recharge or conjunctive use; and/or 2) reducing water use via demand management (i.e., fallowing). Results show that these strategies are most effective if implemented rapidly; and if applied to systems with a higher proportion of hay production relative to overall demand, and with smaller declines in snow relative to reservoir capacity. Adaptation reduces vulnerability values by a median of 3.6 times in the near future (2020-2050), 1.9 times in the mid future (2050-2080), and 1.8 times in the far future (2080-2100) with the largest benefits for higher

elevation tributaries of the Missouri Basin and the least benefit to certain tributaries of the California and Upper Colorado Basins. As climate change continues to alter snow storage throughout the western US, findings present a roadmap for identifying priority areas for adaptation in critical basins.

4.1 Introduction

Agricultural production is the western US depends on snow, which is one of the fastest changing aspects of the hydrological cycle in response to climate change (Musselman *et al* 2017). Warmer winter and spring temperatures are decreasing the fraction of precipitation falling as snow in headwater catchments, reducing the size (Knowles *et al* 2006, Klos *et al* 2014) and persistence of seasonal snowpacks (Stewart 2009), and altering the timing and rate of snowmelt (Barnett *et al* 2008, Rauscher *et al* 2008). Because the amount of water temporarily held in snow has long exceeded built storage capacity in this region (Nijssen *et al* 2001, Barnett *et al* 2005), declines in snow leave water management to face a transformational change in how surface water is stored in mountain environments. In order to reduce damage to people and the environment arising from this change (henceforth, termed vulnerability), headwater management must confront two interacting stresses. First, more streamflow will occur earlier during the winter when flood risk is also higher, forcing tradeoffs between water capture and release (Davenport *et al* 2020, Herrera-Estrada *et al* 2019). Second, reductions in summer streamflow connected to declines in snow storage will enhance reliance on stored surface water during the growing season (Harpold *et al* 2012, Ehsani *et al* 2017, Lundquist *et al* 2008, Barnett *et al* 2008).

Prior research has emphasized how these interacting stresses will leave large portions of the western US vulnerable to continued declines in snow storage (Immerzeel *et al* 2020, Barnett *et al* 2008, Mantkin *et al* 2015, Qin *et al* 2020). Early research by Barnett *et al* (2005) showed that, given insufficient reservoir storage capacity, earlier winter streamflow will be passed to the oceans. They suggest this will enhance water demand competition in places like the Columbia River Basin. Mankin *et al* (2015) incorporated historical demand data into an analysis of projected changes in snowmelt-driven streamflow at the basin scale, finding that the San Joaquin, Sacramento, Rio Grande, Colorado, Klamath, and Upper Great Basin in the western US could experience substantial (albeit uncertain) increases in unmet demand as snow storage declines. More recently work (Qin *et al* (2020) has analyzed relationships between seasonal snowmelt-driven water supply and projected agricultural water demand to demonstrate the vulnerability of irrigated agriculture (specifically, wheat, maize, and rice) in the San Joaquin, Colorado, and Columbia River Basins to declines in snow. These findings were echoed in a global vulnerability assessment conducted by Immerzeel *et al* (2020a), who underscored the pressing need for water management to adapt to these stresses via increased buffering capacity (i.e., storage management) and enhanced water-use efficiency (i.e., demand management).

To date, however, there is limited information about where adapting to declines in snow storage by enhancing water storage (e.g., reservoir expansion or groundwater banking) or reducing demand (e.g., fallowing or crop switching) throughout the western US (Dilling *et al* 2015, He *et al* 2021, Rising and Devineni 2020). Decisions about adaptive strategies require information about how physical hydrology is modified by infrastructure, institutions, and stakeholders (Kellner and Brunner 2021). In order to account for these interactions, vulnerability must be quantified in a robust manner and at a local and operationally meaningful scale (Dilling *et al* 2015, Dilling and Berggren 2015, Sullivan 2011). While previous assessments highlighting basin-scale vulnerability can help inform areas of priority, better information about the distribution of vulnerability across the western US at smaller scales is necessary to plan and implement effective adaptation strategies to declines in snow storage.

Here, we present a novel approach for addressing this research-to-application gap by assessing the adaptive capacity—or the flexibility of agricultural systems to adapt to declines in snow using storage and/or demand management—at an operational scale in the western US. In contrast with previous basin-scale assessments, we identify 13 snow-dominated, headwater reservoir systems across western US that serve downstream agriculturally productive regions. Using historical and projected supply, demand, and storage data, we explore their vulnerability to changing snow using the exposure, sensitivity, and adaptive capacity (ESAC) framework for indexing vulnerability (Cardona *et al* 2012). We quantified exposure as the likelihood of a headwater reservoir system to experience stress related to the impacts of climate change on snow resources, sensitivity as the system's ability to meet with agricultural demand in response to declining snowpack (Cardona *et al* 2012, Luers *et al* 2003), and adaptive capacity as system's flexibility to adapt storage and/or demand under declining snowpack. We then combined these elements to evaluate vulnerability with and without adaptive capacity, which allowed us to quantify where, how, and when storage and demand management can buffer headwater reservoir systems in the western US against declines in snow storage.

4.2 Study Area and Data

4.2.1 Study Area

We identified 13 systems comprised of 28 individual demand regions (i.e., irrigation districts, water conservation districts, water users' associations, cooperative units, or reclamation districts) connected to 23 points of surface water supply distributed across the western US (Table 4-1). Each resulting system listed in Table 4-2 was selected based on three criteria:

1. Basins were selected based on surface water originating in mountainous catchments with at least one headwater type reservoir identified using the National Inventory of Dams (USACE, 2021) and the Global Reservoir and Dam database (GRanD, Lehner *et al* 2011). Headwater type reservoirs were defined as any reservoir located within or directly adjacent to mountains in the western CONUS without significant upstream impairment or managed inflows.
2. Basin selection was limited to those with agriculture designated as their primary or secondary use.
3. A continuous record of streamflow of 26 years for major points of surface water inflow into the system with 25% percent tolerance for NAs was required. Water supply points were defined as reservoirs or streams with explicit rights to direct withdrawal granted to users in a given demand region. Points of surface supply in the system were identified using a combination of publicly available materials and personal communication with water managers.

The approximate area of each demand region was manually digitized by georeferencing publicly available maps. The approximate contributing area for each identified point of surface water supply for each system was delineated using the Terrain analysis using digital elevation (TauDEM) toolbox (Tarboton, 2005).

Table 4-3: Summary of headwater reservoirs, including demand regions and specific points of water supply, adopted for this study. The Kern County Water Agency is abbreviated as KCWA. * Indicates a source of water that was omitted from consideration due to lack of data or complexity.

Demand Region	State	Points of Surface Water Supply	System
Stanfield Irrigation District	OR	Umatilla River, McKay Reservoir	Umatilla
Westland Irrigation District	OR	Umatilla River, McKay Reservoir	
Greenfields Irrigation District	MT	Pishkun Dike*, Willow Creek Reservoir, Sun River via Gibson Reservoir	Sun
Lakeview Irrigation District*	WY	South Fork of the Shoshone River	Shoshone
Deaver Irrigation District	WY	Shoshone River via Buffalo Bill Reservoir	
Willwood Irrigation District	WY	Shoshone River via Buffalo Bill Reservoir	
Shoshone Irrigation District	WY	Shoshone River via Buffalo Bill Reservoir	
Heart Mountain Irrigation District	WY	Shoshone River via Buffalo Bill Reservoir	
Midvale Irrigation District	WY	Bull Lake, Pilot Butte Reservoir	Wind
Bridger Valley Water Conservancy District	WY	Blacks Fork River via Meeks Cabin Reservoir, Smiths Fork River via Stateline Reservoir	Bridger
North Fork Water Conservancy District	CO	North Fork of the Gunnison River via Anthracite Creek and Paonia Reservoir	Paonia

Rio Costilla Cooperative Livestock Association	NM	Rio Costilla via Costilla Dam	Costilla
Price Valley Water Users Association	UT	Price River via Scofield Reservoir, White River, and Willow Creek	Price
Little Wood River Irrigation District	ID	Little Wood River via Little Wood River Reservoir	Little Wood
Walker River Irrigation District	NV	West Walker River via Topaz Lake, East Walker River via Bridgeport Reservoir	Walker
Olcese Water District (KCWA)	CA	Kern River via Isabella Reservoir	Kern
North Kern Water Supply District (KCWA)	CA	Kern River via Isabella Reservoir	
Kern Delta Water District (KCWA)	CA	Kern River via Isabella Reservoir, State Water Project*	
Henry Miller Water District (KCWA)	CA	Kern River via Isabella Reservoir, State Water Project*	
Rosedale Rio Bravo Water Supply District (KCWA)	CA	Kern River via Isabella Reservoir, State Water Project*	
Buena Vista Water Supply District (KCWA)	CA	Kern River via Isabella Reservoir, State Water Project*	
Kaweah Delta Water Conservation District	CA	Kaweah River via Kaweah Lake, Federal Central Valley Project*, Dry Creek* and Yokohl Creek*	Kaweah
Kittitas Reclamation District	WA	Keechelus and Kachness Lakes	Kittitas

4.2.2 Data

4.2.2.1 Water supply (*Q*)

Water supply is defined as streamflow, including reservoir inflows, (*Q*) *to* which the demand region is granted access. We used the following data to quantify historical and projected *Q* into each system identified in Table 4-4.

4.2.2.2 Reservoir inflows and streamflow (*Q*)

Historical streamflow is defined using a combination of data from United States Geological Survey (USGS), the United States Bureau of Reclaimation (USBR), and the United States

Army Corps of Engineers (USACE). We obtained simulated historical and projected streamflow from the localized constructed analogue (LOCA, Pierce *et al* 2014) CMIP5 hydrologic projections (Vano *et al*, 2020) routed in MizuRoute (Mizukami *et al* 2016) using the Kinematic Wave tracking option.

Streamflow was modified using the statistical change factor method to account for biases in simulated water supply (Minville *et al* 2008, Chen *et al* 2011, Mankin and Diffenbaugh 2015). Twenty-six 14-day windows (or, in the case of leap years, a single 15-day window) were defined based on a consistent day of the water year. The cumulative Q and median ($\bar{Q}$) value for each of these windows uses the observed and simulated data over the historical period (1979-2005); and simulated data for the near future (2020-2050), mid future (2050-2080), and far future (2080-2100). Simulated data includes the 64 RCP-GCM pairs in the CMIP5 ensemble. A statistical change factor was applied to account for biases in the CMIP5 ensemble:

$$\overline{Q_{\text{bias,future}}} = \overline{Q_{\text{observed,historical}}}\ x\ \left(\frac{\overline{Q_{\text{simulated,future}}}}{\overline{Q_{\text{simulated,historical}}}}\right) \qquad \text{Eq. (4-1)}$$

Where $\overline{Q_{\text{bias,future}}}$ is the bias corrected median simulated streamflow over the near, mid, or far future period, $\overline{Q_{\text{observed,historical}}}$ is the median observed streamflow over the 1979-2005 historical period, $\overline{Q_{\text{simulated,historical}}}$ is the median observed streamflow over the 1979-2005 historical period, and $\overline{Q_{\text{simulated,future}}}$ is the median simulated streamflow over the near, mid, or far future period. Bias correction at sub-annual timesteps can change the magnitude of annual streamflow (Q_{WY}) per Zhu *et al* (2005) and Hamlet and Lettenmaier

(1999). To preserve simulated Q_{WY}, we thus applied a second bias correction to each $\overline{Q_{\text{bias,future}}}$ value obtained from Eq. (4-1) as described in Eq. (4-2):

$$\overline{Q_{\text{corrected,future}}} = \overline{Q_{\text{bias,future}}}\ x \left(\frac{\frac{\overline{Q_{\text{WY,simulated,future}}}}{\overline{Q_{\text{WY,simulated,historical}}}}}{\frac{\overline{Q_{\text{WY,bias,future}}}}{\overline{Q_{\text{WY,observed,}historical}}}} \right) \qquad \text{Eq. (4-2)}$$

Where $\overline{Q_{\text{corrected,future}}}$ is the simulated $\bar{Q}$ bias-corrected using the statistical change factor and adjusted to preserve the annual simulated $\overline{Q_{WY}}$, and $\overline{Q_{\text{WY,simulated,future}}}$ was obtained as the sum of all $\overline{Q_{\text{simulated,future}}}$ values each *i* 14-day window in the water year:

$$\overline{Q_{\text{WY,simulated,future}}} = \sum_{0}^{i=26} \overline{Q_{\text{simulated,future,}i}} \qquad \text{Eq. (4-3)}$$

We then used the resulting estimates of $\overline{Q_{\text{corrected,future}}}$ on the left-hand side of Eq. (4-3) to obtain corrected median annual streamflow ($\overline{Q_{\text{WY,future}}}$) for each of our three future time periods.

4.2.2.3 Water demand (*D*)

4.2.2.3.1 Net Irrigation Water Demand ($\boldsymbol{NIWD}$)

Water demand (*D*) is defined as the sum of net irrigation water demand (NIWR) and reservoir evaporation demand ($E_{reservoir}$) for each basin identified in Table 4-1. Data from the United States Department of Agriculture (USDA) and the USBR (Huntington *et al*., 2014; Allen *et al* (2020)) is used to simulate water demand over each future period. To account for changes in atmospheric water demand, a west-wide irrigation net irrigation water requirement (NIWR) dataset was selected from the USBR (Huntington *et al* 2014) The USBR NIWR dataset was selected because it is the only product that provides a dynamic estimate of NIWR (m) and that accounts for differences among the major crop types for each demand region. The dataset contains 10 simulations of projected future agricultural water requirement for major crop types in each demand region using the climate from 2020-2050, from 2050-2080, and from 2080-2100 as well as a historical irrigation water demand from 1950-1999.

The USBR NIWR data was combined with area estimates for each crop type in each demand regions based on Cropland Data Layer (CDL) data (Huntington *et al* 2014, Allen *et al* 2020) to produce a volumetric demand rate. Following Lark *et al* (2017), CDL data were bias corrected using published error super matrices (USDA NASS RDD Spatial Analysis Research Section, 2016), which were then multiplied by the estimated crop class area to obtain a bias corrected estimate of the total area for each crop class for each year in each demand region (A_{crop}). Using the same 26 14-day windows in the water year, a cumulative 14-day NIWR was calculated for each crop type in each demand region. From these estimates, a median 14-day NIWR ($\overline{NIWR_{crop,historical}}$) was calculated based on the historical period (1950-1999) and the near, mid, and far future periods ($\overline{NIWR_{crop,future}}$). Because NIWR data over the historical period are already bias-corrected per Huntington *et al* (2014) and Allen *et al* (2020), no further bias-correction was applied.

The CDL data were used to estimate a 'Business-as-Usual' or BAS cropping mix to simulate a no-change future scenario and a 'Demand Management' or DM cropping mix to simulate reduced water demand for each demand region. For each of the 12 years in the CDL record, an estimate of NIWD ($NIWD_{CDL}$) was calculated as:

$$\mathrm{NIWD_{CDL}} = \overline{\mathrm{NIWR_{crop,historical}}}\ x\ \mathrm{A_{crop}} \qquad \text{Eq. (4-4)}$$

Where $\overline{\mathrm{NIWR_{crop,historical}}}$ refers to the median historical NIWR for each crop in the USBR NIWR (m) and $\mathrm{A_{crop}}$ refers to the area for each crop for each year based on the CDL data (m^2). Because of the harmony between the two datasets, there is good agreement between crops represented in the USBR NIWR dataset and the CDL dataset. However, in the case of an imperfect cross-reference, crop coefficient data were used to approximate the closest fit with available data (e.g., miscellaneous vegetables were classified as field corn in the case that miscellaneous vegetables was not a crop class provided in the USBR data). All 12 values of $\mathrm{NIWD_{CDL}}$ were used to estimate a BAS crop mix for each demand region ($\mathrm{A_{BAS}}$) or the crop mix corresponding to the median system-wide NIWD and a DM crop mix for each demand region ($\mathrm{A_{DM}}$) or the crop mix corresponding to the 25th percentile system-wide NIWD.

The median historical NIWD is calculated assuming a BAS cropping scenario for each demand region ($\overline{\mathrm{NIWD_{historical}}}$):

$$\overline{\mathrm{NIWD_{historical}}} = \sum_{crop=1}^{crop=n} \overline{\mathrm{NIWR_{historical}}}\ x\ \mathrm{A_{crop,BAS}} \qquad \text{Eq. (4-5)}$$

Where *n* is the number of individual crops (*crop*) represented in the BAS crop mix (A_{BAS}) and $\mathrm{A_{crop,BAS}}$ is the area of each individual crop contained in the BAS cropping scenario. The median future NIWD assuming a BAS cropping scenario for each demand region ($\overline{\mathrm{NIWD_{BAS,future}}}$) is calculated as:

$$\overline{\mathrm{NIWD}_{\mathrm{BAS,future}}} = \sum_{crop=1}^{crop=n} \overline{\mathrm{NIWR}_{\mathrm{crop,future}}}\ x\ \mathrm{A}_{\mathrm{crop,BAS}} \qquad \text{Eq. (4-6)}$$

Where $\overline{NIWD_{BAS,future}}$ is the median *NIWD* under the BAS cropping scenario for each of the three future periods. Lastly, the median future NIWD was calculated assuming a DM cropping scenario:

$$\overline{\mathrm{NIWD}_{\mathrm{DM,future}}} = \sum_{crop=1}^{crop=n} \overline{\mathrm{NIWR}_{\mathrm{crop,future}}}\ x\ \mathrm{A}_{\mathrm{crop,DM}} \qquad \text{Eq. (4-7)}$$

Where $\overline{NIWD_{DM,future}}$ is the median *NIWD* under the DM cropping scenario for each of the three future periods, *n* is the number of individual crops (*crop*) represented in the DM crop mix (A_{DM}), and $\mathrm{A}_{\mathrm{crop,DM}}$ is area of each individual crop contained in the DM cropping scenario.

4.2.2.3.2 Reservoir Evaporation Demand ($\boldsymbol{E_{reservoir}}$)

A combination of datasets is used to estimate volumetric reservoir evaporation ($E_{reservoir}$). In the absence of future evaporation data for each headwater reservoir listed in Table 4-1, we assumed that potential evapotranspiration from the USBR NIWR dataset was a reasonable approximation of evaporative demand (E in m). The maximum surface area ($A_{reservoir}$ in m^2) for each headwater reservoir (Lehner *et al,* 2011) was combined with a cumulative 14-day E for each demand region to obtain a median 14-day E ($\overline{E_{historical}}$) based on the historical period (1950-1999) and the near, mid, and far future periods ($\overline{E_{future}}$). Because E data over the historical period are already bias-corrected (Huntington *et al,* 2014); Allen *et al,* 2020), no further bias-correction was applied.

Because the rules of future reservoir operations are unknown, our projections of $A_{reservoir}$ use a gross annual consumption approach proposed by Hogeboom *et al* (2018) where:

$$\overline{E_{reservoir,future}} = 10\ x\ \overline{E_{future}}\ x\ A_{reservoir}\ x\ k \qquad \text{Eq. (4-8)}$$

Where k [-] is a correction factor to account for differences in the filling condition and $A_{reservoir}$, which Hogeboom *et al* (2018) propose should be set to 0.5625 for most reservoirs based on Kohli & Frenken (2015). The median historical $\overline{E_{reservoir}}$ uses Eq. (4-8) but replaces $\overline{E_{historical}}$ on the right-hand side.

4.2.2.3.3 Annual Demand (D_{WY})

NIWD and $E_{reservoir}$ are used to construct median annual estimated of water demand ($\overline{D_{\mathrm{WY}}}$) for each system over the historical and future periods. $\overline{D_{\mathrm{WY,historical}}}$ is calculated as:

$$\overline{D_{\mathrm{WY,historical}}} = \sum_{0}^{i=26} \overline{\mathrm{NIWD}_{\mathrm{historical},i}} + \overline{E_{reservoir,historical,i}} \qquad \text{Eq. (4-9)}$$

Where *i* refers to each 14-day window in the water year. For the near, mid, and far future periods, $\overline{D_{\mathrm{WY,BAS,future}}}$ was estimated using the same equation but with $\overline{\mathrm{NIWD}_{\mathrm{BAS,future},i}}$ and $\overline{D_{\mathrm{WY,DM,future}}}$ was estimated using the same equation but with $\overline{\mathrm{NIWD}_{\mathrm{DM,future},i}}$. The fraction of $\overline{E_{reservoir,future}}$ to $\overline{D_{\mathrm{WY,historical}}}$ is reported in Table S4-1.

4.2.3 Water Storage (*S*)

Surface water storage (*S*) is defined as the combination of natural snow storage (S_{snow}), approximated as snow water equivalent (*SWE*) and built reservoir storage (S_{built}).

A storage transition metric ($\Delta S_{transition}$) is the difference between the future and historical fraction of natural (S_{snow}) to built (S_{built}) storage using the data sources described below.

4.2.3.1 Natural Storage (S_{snow})

To quantify observed historical S_{snow}, we used SWE from the National Climate Assessment-Land Data Assimilation System (NCA-LDAS, Kumar *et al* 2019) available from 1979-2005. From NCA-LDAS *SWE,* a daily accumulated depth of *SWE* is calculated as $SWE_j = SWE_j - SWE_{j-1}$, where j is a day of the water year. Each value of SWE_j is multiplied by the contributing area for each source of water supply outlined in Table 4-1. The same process is used to obtain simulated historical and projected S_{snow}, except with SWE_j obtained from LOCA (Pierce *et al* 2014) CMIP5 hydrologic projections (Vano *et al*, 2020). Using the same 26 14-day windows as for *Q* and *D,* cumulative observed S_{snow} is calculated over the historical period (1979-2005) and simulated S_{snow} over the historical and future periods. The median observed value of S_{snow} over the historical period ($\overline{S_{snow-observed,historical}}$) and for each of the 64 simulated RCP-GCM pairs in the CMIP5 ensemble over the historical period ($\overline{S_{snow-simulated,historical}}$) and for each of our future periods $\overline{S_{snow-simulated,future}}$). Simulated future $\overline{S_{snow}}$, is bias-corrected using the same methodology as for *Q. It is* aggregated following Eq. (4-3) to obtain median annual S_{snow} over the three future periods ($\overline{S_{\text{snow}-WY,\text{corrected,future}}}$). Median observed annual S_{snow} ($\overline{S_{\text{snow-WY,observed,historical}}}$) was evaluated similarly.

4.2.3.2 Built Storage ($\boldsymbol{S_{Built}}$)

Following Masia *et al* (2018), a single annual volume of S_{Built} for each basin is calculated as a function of the maximum capacity for each headwater reservoir in Table 4-1 reported by Lehner *et al* (2011).

4.3 Methods

Using the data outlined in Section 4-2, the vulnerability of our headwater systems is assessed to a storage transition metric ($\Delta S_{surface}$), which evaluates changes in S_{snow} relative to S_{built}. Vulnerability assumes the ESAC framework adopted by the IPCC (Cardona *et al* 2012), which relies on quantification of exposure (Section 4.3.1), sensitivity (Section 4.3.2), and adaptive capacity (4.3.3).

4.3.1 Exposure Analysis

Following Cardona *et al* (2012), exposure is defined as the likelihood of a headwater reservoir system to experience stress related to the impacts of climate change on snow resources. Existing literature suggests that although the impacts of change snow resources on water supply are varied, changes in the center of water supply mass (DoQ_{50}) and changes in the annual volume of water supply (Q_{WY}) are critical sentinels for water management (Stewart 2009, Stewart *et al* 2004, Regonda *et al* 2005, Gordon *et al* 2022). Exposure is defined via Eq. (4-10 to 4-11) below. A timing exposure metric is calculated as the center of water supply mass timing:

$$Timing\ I = \left|\overline{\mathrm{DoQ}_{50,corrected,future}} - \overline{\mathrm{DoQ}_{50,\mathrm{observed,historical}}}\right| \qquad \text{Eq. (4-10)}$$

Where $\overline{\mathrm{DoQ}_{50-corrected,future}}$ is the median simulated day of water supply center of mass timing for the near, mid, or far future periods and $\overline{\mathrm{DoQ}_{50-\mathrm{observed,historical}}}$ is the median observed day of water supply center of mass timing for the historical period (1979-2005).

A magnitude metric based on $\mathrm{Q_{WY}}$ per is given as:

$$Magnitude\ I = \frac{\overline{\mathrm{Q_{WY,corrected,future}}}}{\overline{\mathrm{Q}_{\mathrm{WY,observed,}historical}}} \qquad \text{Eq. (4-11)}$$

The timing and magnitude indicators are rescaled following Gonzales and Ajami (2017a) as:

$$Rescaled\ I = 1 + (10 - 1) * \frac{(I - A)}{(B - A)} \qquad \text{Eq. (4-12)}$$

Where *I* refers to a generic indicator, A and B are the upper and lower bound of the original scale, respectively. The rescaled *I* retains its original ranking, but on a 1 to 10 scale. In the case of the *Magnitude I,* for example, higher values are associated with decreased exposure and vice versa, which is retained in the *Rescaled Magnitude I*. *Rescaled Timing I* and *Magnitude I* then combined into estimate of Exposure for each system over each future period using the geometric mean following Gonzales and Ajami (2017a) where:

$$Exposure = [\,(11 - Rescaled\ Magnitude\ I)* (Rescaled\ Timing\ I)]^{1/2} \qquad \text{Eq. (4-13)}$$

4.3.2 Sensitivity Analysis

Sensitivity is defined as the system's response (i.e., its ability to meet water demand from agricultural regions listed in Table 4-1) under exposure from declining snowpack (Cardona *et al* 2012, Luers *et al* 2003). Following Luers *et al* (2003), Sensitivity is defined as:

$$Sensitivity = \frac{|dW/dX|}{W/W_o} \qquad \text{Eq. (4-14)}$$

Where $dW$is the change in well-being (W) with respect to the change in the stressor dX and W_o is the threshold value of W below which the system is assumed to incur damage. W is defined as $\frac{\overline{\mathrm{Q_{WY,corrected,future}}}}{\overline{D_{\mathrm{WY,BAS,future}}}}$ and the dW is evaluated against $\frac{\overline{\mathrm{Q_{WY,observed,historical}}}}{\overline{D_{\mathrm{WY,historical}}}}$. dX is evaluated as ($Exposure$ - 0). We then evaluate the threshold below which the system incurs damage (W_o) as $\frac{\overline{\mathrm{Q_{WY,observed,historical}}}}{\overline{D_{\mathrm{WY,historical}}}}$, meaning that the system is susceptible to damage if it falls below the historical value of W_o, which is selected to avoid penalizing basins that rely on groundwater and import (Table 4-1). This, in effect, quantifies system sensitivity to changes in headwater supply. Sensitivity was then rescaled to the same 1-10 range as Exposure using Eq. (4-6).

4.3.3 Adaptive Capacity Analysis

Adaptive capacity is defined as system flexibility to meet water demand from agricultural regions under stress (Cardona *et al* 2012, Luers *et al* 2003). Two different adaptive capacities are defined for the western US: demand management and storage management (He *et al* 2021, Consulting 2020, Heikkila 2003, Olmstead 2014, Elliott *et al* 2014). Demand management is the fraction of water savings obtained by reducing and/or altering crop type and area following work by Gonzales and Ajami (2017b) in urban water systems where:

$$Demand\ Management\ I = \frac{\overline{D_{\mathrm{WY,DM,future}}} - \overline{D_{\mathrm{WY,BAS,future}}}}{\overline{D_{\mathrm{WY,BAS,future}}}} \qquad \text{Eq. (4-15)}$$

The adaptive potential of storage management follows the storage recharge indicator proposed by Masia *et al* (2018) where:

$$Storage\ Management\ I = \frac{\overline{(Q_{\mathrm{WY,future}} - D_{\mathrm{WY,BAS,future}})} - \overline{(Q_{\mathrm{WY,future}} - D_{\mathrm{WY,BAS,future}})}}{S_{Built}} \qquad \text{Eq. (4-16)}$$

Both *Demand Management* and *Storage Management I* were rescaled following Eq. (4-12) and combined using the geometric mean where:

$$Adaptive\ Capacity = [(11 - Rescaled\ Storage\ Management\ I) * (Rescaled\ Demand\ Management\ I)]^{1/2} \qquad \text{Eq. (4-17)}$$

4.3.4 Vulnerability Analysis

Using results from Section 4.3.1 to 4.3.3, vulnerability follows Cardona *et al* (2012) as:

$$Vulnerability = (Exposusure + Sensitivity) - Adaptive\ Capacity \qquad \text{Eq. (4-18)}$$

Where *Exposure, Sensitivity*, and *Adaptive Capacity* are the re-scaled values obtained from Eq. (4-10 to 4-17). Vulnerability with and without adaptive capacity is calculated to identify systems where adaptive strategies reduce vulnerability. Based on Eq. (4-18), the highest value of vulnerability for an individual system is 20 based on maximum values of 10 for exposure and sensitivity and removal of adaptive capacity consistent with previous work (Qin *et al* 2020, Mankin *et al* 2015, Immerzeel *et al* 2020b). The lowest value of vulnerability for an individual system is -8 based on minimum values of 1 for exposure and sensitivity and a maximum value of 10 for adaptive capacity. Vulnerability scores were used to establish three groups based on terciles representing high vulnerability (upper tercile of scores), moderate vulnerability (middle tercile of scores), and low vulnerability (lowest tercile of scores).

4.4 Results

The vulnerability of 13 headwater reservoir systems to climate change were examined by focusing on declines in snow storage (S_{snow}) relative to built storage (S_{built}) in order to evaluate changes in surface water storage ($\Delta S_{surface}$). $\Delta S_{surface}$ is defined as a change from historical S_{snow}/S_{built} over the three future periods (early, mid, far). Historical headwater storage conditions are provided in Figure 4-1. Historical S_{snow} ranges from a low of ~28 million m^3 in Costilla to a high of ~583 million m^3 in Wind with a median value of ~164 million m^3 across all systems (Figure 4-1A). Historical S_{Built} ranges from a low of ~20 million m^3 in Costilla to a high of ~746 million m^3 in Shoshone with a median value of ~110 million m^3 all systems (Figure 4-1B). Given that median S_{snow} exceeds median S_{Built} across all systems, it is unsurprising that $S_{surface}$ is 1.54 across all systems. This result is consistent with previous findings at larger scales suggesting that S_{snow} generally exceeds S_{Built} in this region (Nijssen *et al* 2001, Barnett *et al* 2005). The least S_{snow} dependent system is Kittitas ($S_{surface} = 0.19$) and the S_{snow} dependent system is Paonia ($S_{surface} = 3.92$).

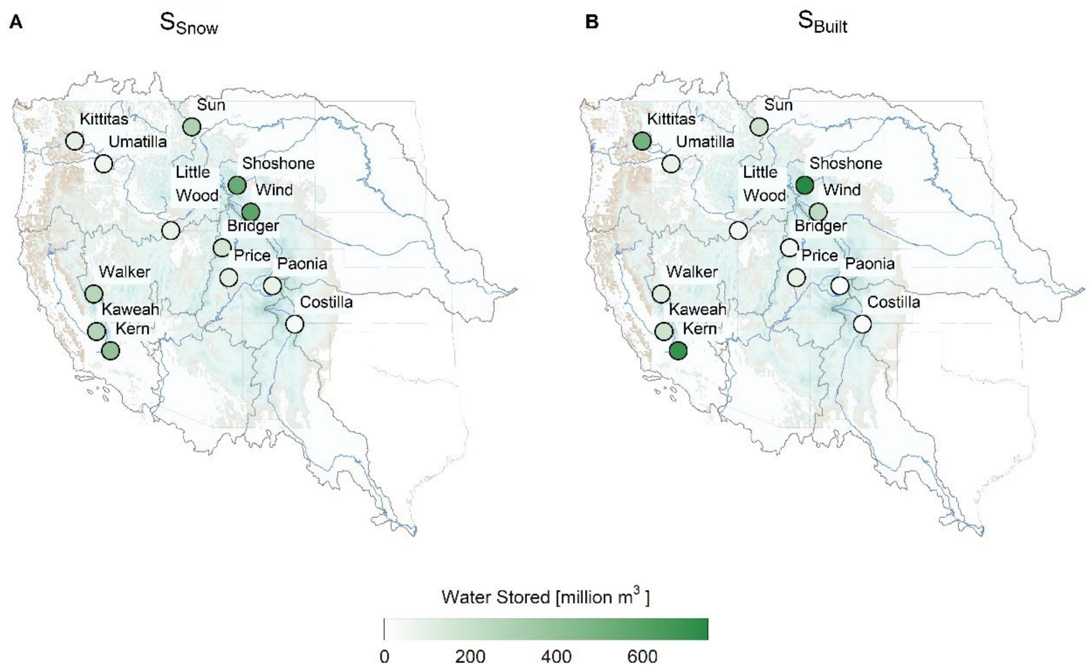

Figure 4-1: A) Historical median S_{snow} (1979-2005) and B) maximum S_{built} for all reservoirs within each system.

4.4.1 Exposure

Exposure (Figure 4-2A to 4-2C), or the likelihood of experiencing stress from declining S_{snow}, is a function of changes in the timing and magnitude of Q (Figure 4-2D to 4-2F). Results indicate that our systems are exposed to declines in S_{snow} following two different typologies: 1) higher near future exposure that does not increase into the mid and far future periods and 2) lower near future exposure that accelerates over the mid and far future. Exposure appears largely a function of geography dictating differences in snowpack and climate. For example, lower elevation coastal systems (e.g., Kittitas, Umatilla, Walker, Kern, and Kaweah) experience high but stable exposure, as indicated by consistent darker red symbol coloring in Figure 4-2A to 4-2C. Conversely, the majority of our higher elevation interior systems follow the second exposure typology with low near future exposure that accelerates over the century—and in some cases (i.e., Costilla), eventually overtaking the exposure of more coastal systems. Spearman correlation (ρ) indicates that exposure increases as ΔS_{snow} grows larger (Figure 4-2, Figure S4-4, ρ = -0.38, slope = -4.31), but is not strongly associated with $\Delta S_{Surface}$ (represented by symbol sizing, Figure S4-5, ρ = - 0.03, slope = 0.17). Exposure is also higher in systems with a lower fraction of water demand from hay production (Figure S4-6, ρ = - 0.23, slope = -2.58).

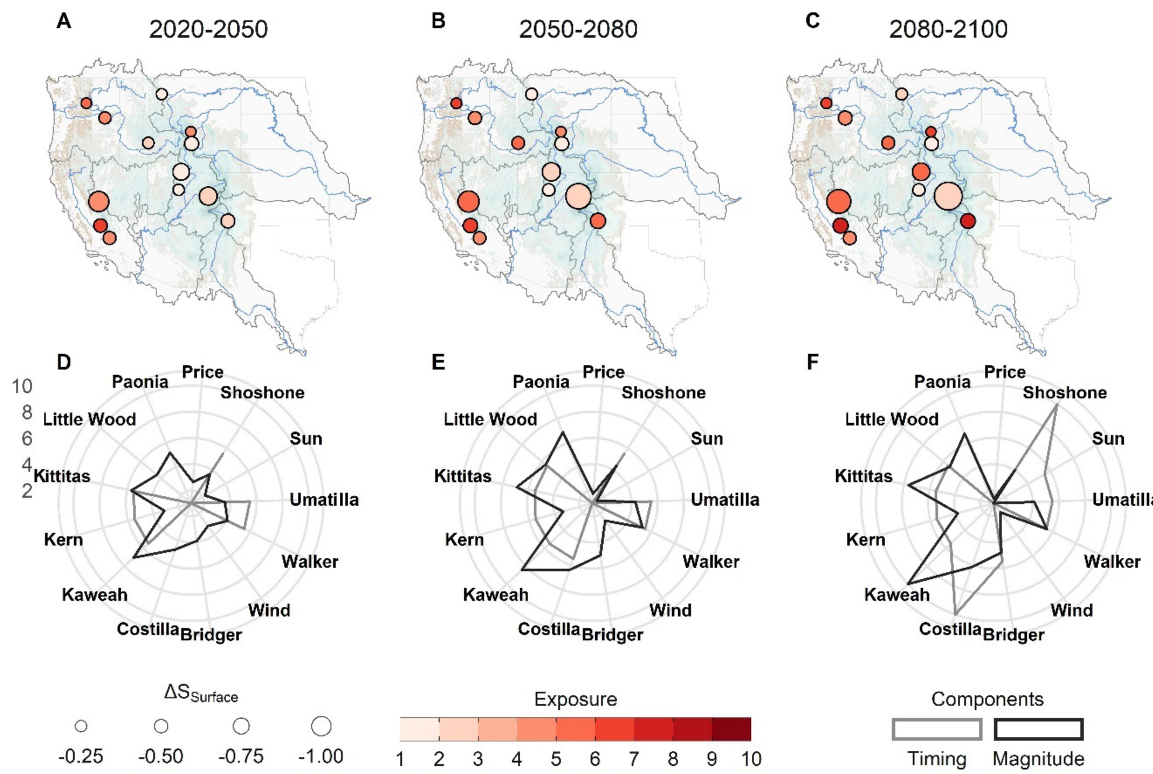

Figure 4- 2: Exposure results for 13 systems based on the geometric mean of changes in streamflow timing and magnitude for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Values of re-scaled timing and magnitude indicators are presented for: D) the near future (2020-2050); E) the mid future (2050-2080);

F) the far future (2080-2100). Higher values suggest larger changes in magntidue or timing. Large symbols indicate larger $\Delta S_{Surface}$ values.

4.4.2 Sensitivity

The sensitivity of each system's supply-to-demand ratio (termed well-being, W) under projected changes in Q arising from declines in S_{snow} (exposure) are provided in Figure 4-2. It is assumed that systems are more sensitive to damage in the future as the projected supply-to-demand ratio falls further below the median historical ratio of supply-to-demand. Results highlight how projected changes in demand can moderate or amplify the system's response to changes in water supply (exposure). We present the underlying dW values used to assess sensitivity in Figure S4-7. For example, Costilla, Kern, and Kaweah are all highly exposed to changes in water supply per Figure 4-2; however, supply-demand interactions leave these systems less sensitive to changes in water supply. Conversely, Paonia is less exposed, but highly sensitive to changes in water supply, particularly in the mid and far future (Figure 4-3B and 4-3C). Sensitivity is unevenly distributed throughout major river basins, with more sensitive systems located in the tributaries of the Upper Colorado, Columbia, and Missouri Basins. Temporally, system sensitivity accelerates less from climate change compared to exposure (Figure 4-3). Sensitivity is very weakly associated with ΔS_{Snow}, but results indicate that the well-being of systems with smaller ΔS_{Snow} is more sensitive to changes in water supply than in systems with larger declines in snow storage (Figure S4-8, ρ = 0.04, slope = 1.17). The relationship between sensitivity and $\Delta S_{Surface}$ is also weak, but more intuitive suggesting that sensitivity increases with larger values of $\Delta S_{Surface}$ (represented by symbol sizing, Figure S4-9, , ρ = -0.12, slope = -1.04).

Systems with a larger fraction of water demand from hay production are substantially more sensitive to changes in water supply (Figure S4-10, ρ = -0.56, slope = 5.01).

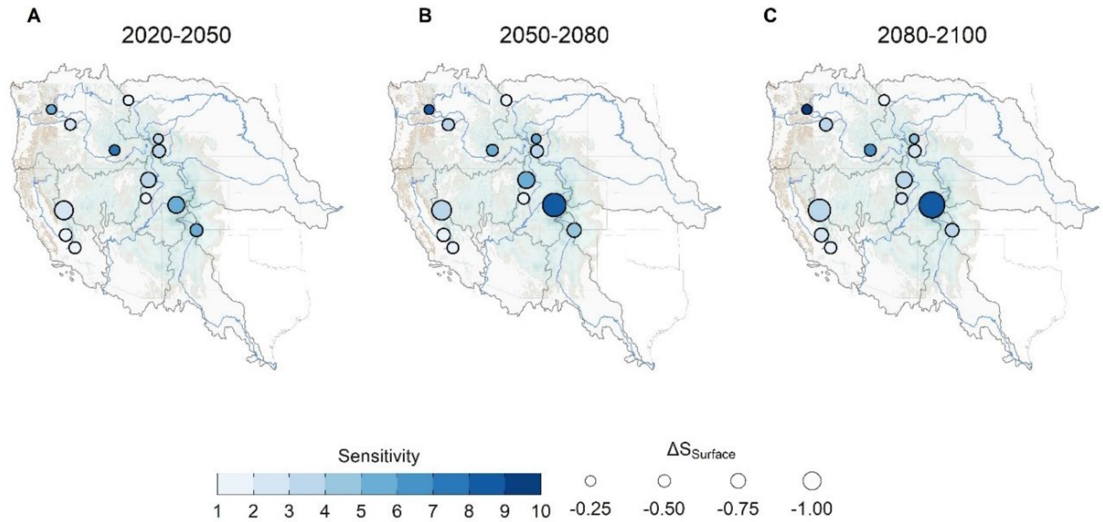

Figure 4-3: Sensitivity results for 13 systems based on changes in the system well-being (W) or the ratio of supply to demanwith respect to exposure assuming that damage incurs if the historical system well-being cannot be met for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). High values suggest greater sensitivity to changes in S_{snow}.

4.4.3 Adaptive Capacity

We evaluated adaptive capacity as the flexibility of the system to meet agricultural demand under stress (Figure 4-4) using two relevant adaptive strategies (Consulting 2020, He *et al* 2021): demand management and storage management. In general, our results highlight robust potential for adaptation via combined storage and demand management in interior and higher elevation systems (i.e., Bridger, Costilla, Price, Wind, and Shoshone) particularly in the near and mid future (Figure 4-4A to 4-4C). We present the underlying storage and demand management indicator values used to assess exposure in Figure S4-11 and S4-12, respectively. However, there are several systems, such as Paonia and Kaweah and to a lesser extent, Walker, which have fewer opportunities for adaptation through demand management and/or storage management as the century progresses. Geographically, the three of the least adaptive systems (Kaweah, Kern, and Walker located in the California and Great Basin) are located in lower elevation areas along the west coast (Figure 4-4) with the outlier in this group being Paonia (a higher elevation interior tributary in the Colorado Basin). Temporally, our results suggest a general decline in overall adaptive capacity for most systems over the century —driven largely by declines in storage management capacity (Figure 4-4D to 4-4F). Adaptive capacity increases with smaller values of both ΔS_{snow} (Figure S4-13, slope = 2.68, ρ = 0.32) and $\Delta S_{Surface}$ (Figure S4-14, slope = 2.59, ρ = 0.46). Hay production (i.e., alfalfa, pasture, or grass hay) also appears to determine overall adaptive capacity, with larger demand from hay associated with larger adaptive capacity (Figure S4-15, slope = 2.76, ρ = 0.21).

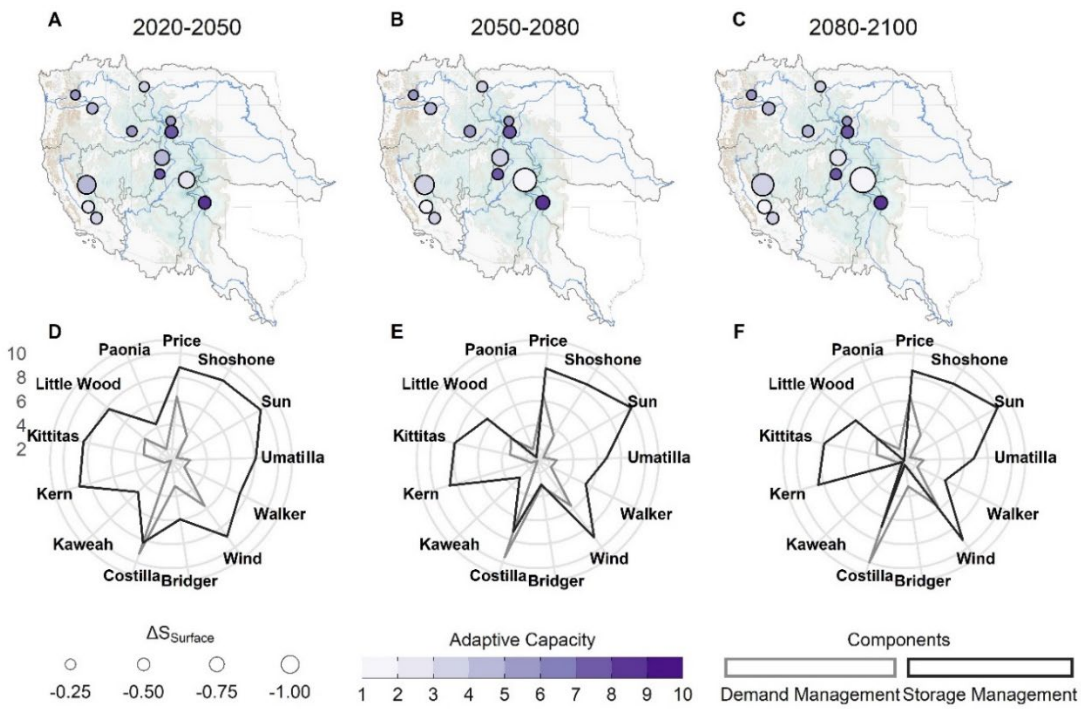

Figure 4-4: Adaptive capacity results for 13 systems based on the geometric mean of supply management and demand management capacity for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100).

4.4.4 Vulnerability

We evaluate system vulnerability to exposure from changing snowpacks in two different ways: 1) first ignoring adaptive capacity and using only exposure (Figure 4-2) and sensitivity (Figure 4-3) results to evaluate Eq. (4-18) and 2) including adaptive capacity results (Figure 4-4) in Eq. (4-18). This allows us to measure the potential reduction in vulnerability from adaptation (Figure 4-5). Vulnerability scores without considering potential adaptive capacity are presented in Figure 4-5A to C with cooler colors representing lower vulnerability scores and warmer colors representing higher vulnerability scores. Results show that without adaptation vulnerability is higher in systems with larger ΔS_{Snow} (Figure S4-16, slope = -3.13, ρ = -0.2) and larger fraction of water demand from hay production (Figure S4-17, slope = 2.42, ρ = 0.16). Vulnerability is not associated with $\Delta S_{Surface}$ (Figure S4-18, slope = -0.87, ρ = -0.07). Median vulnerability with no adaptation is 7.1 in the near future, 8.7 in the mid future, and 9.5 in the far future.

When adaptive capacity is considered, a number of systems see substantive reductions in their vulnerability (median = 4.4 on a scale from -8 to 20 across all systems) per coloring in Figure 4-5D to F. Results show that when adaptation is considered, higher vulnerability is associated with larger ΔS_{Snow} (Figure S4-19, slope = -5.81, ρ = -0.25) and larger $\Delta S_{Surface}$ (Figure S4-20, slope = -3.46, ρ = -0.19), but not with hay production (not shown, slope = -0.31, ρ = -0.01).

Reductions in vulnerability from adaptation are not unform across all systems (Figure 4-4) and potential benefits decline over time per Figure 4-5 (median = 4.6 in the near future, median = 4.2 in the mid future, median = 4.2 in the far future). In the near future, adaptation reduces the number of highly vulnerable systems (e.g., systems in the upper tercile for all vulnerability scores across all systems and periods) from 4 to 0 per Table S4-1 with the largest reductions in vulnerability (> 4.4) for Costilla, Price, Wind, Shoshone, Little Wood, and Kittitas. Midcentury adaptation reduces the number of highly vulnerable systems by 4.5 times (Table S4-1), with the benefits for the same systems. Albeit more modest, far future adaptive capacities also reduce the number of highly vulnerable systems by 3 times, with the largest reductions in the same six systems. Geographically, the systems with the largest reductions in vulnerability across time are higher elevation, interior systems with the exception of Kittitas. Smaller reductions in vulnerability through adaptation (Figure 4-

5G to 4-5I) occurred in lower elevation and coastal systems (i.e., Umatilla, Kern, Walker, and Kaweah) with the exception of Sun, Bridger, and Paonia. Accounting for adaptation via adaptive capacity were correlated to smaller values of ΔS_{Snow}, smaller values of $\Delta S_{Surface}$, and in systems where hay (e.g., alfalfa, grass, or pasture) accounts for a larger fraction of D per Section 4.3.4.

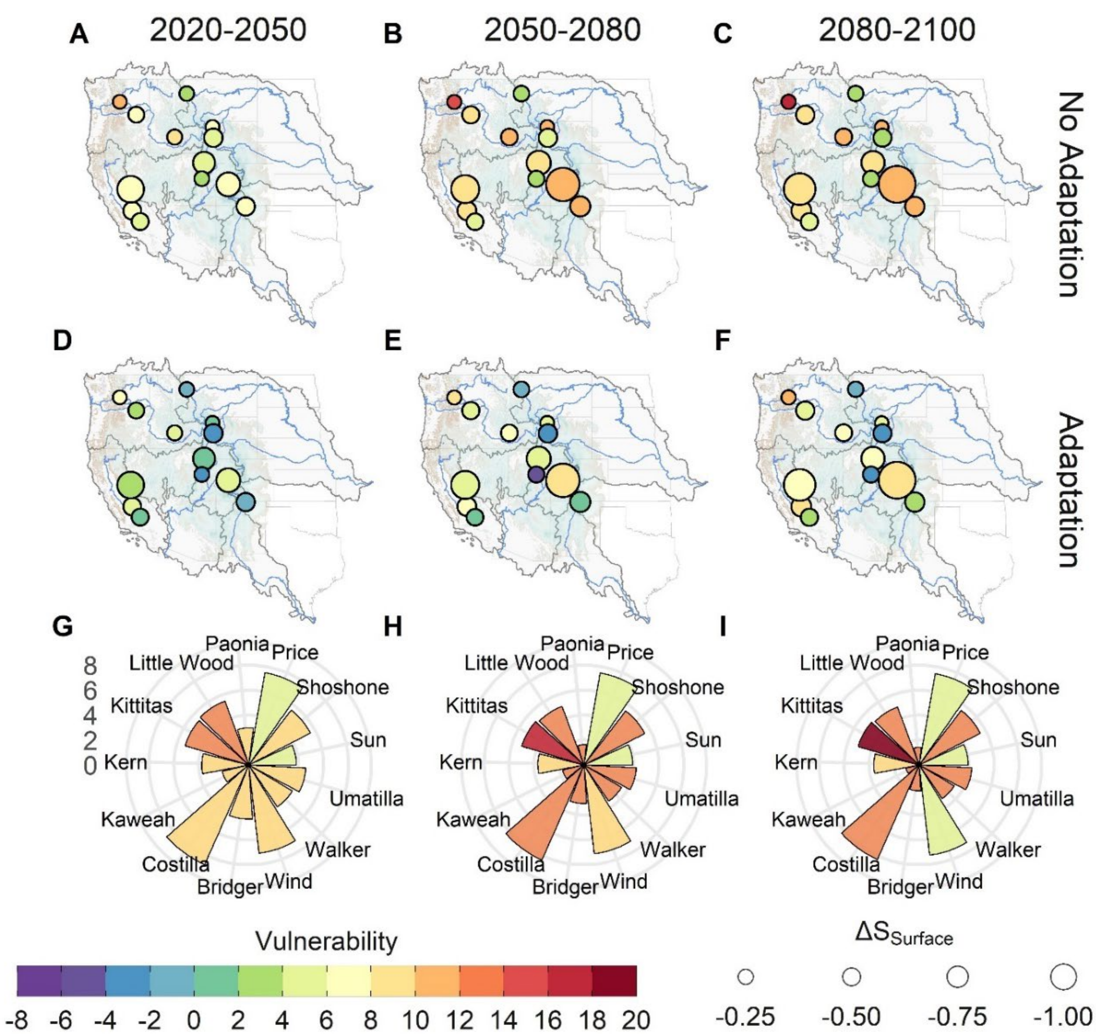

Figure 4-5: Vulnerability results for 13 systems withouth adaptive capacity for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100) and with

adaptive capacity for the same periods D-F. We then present the difference (adaptive capacity) in G through I with symbol coloring from A to C results.

4.5 Discussion

Agricultural production in the western US will be substantially impacted by ongoing and accelerating changes in mountain snow storage (Barnett *et al* 2005, Immerzeel *et al* 2020a, Qin *et al* 2020, Mankin *et al* 2015). However, comparatively few studies have examined how complex interactions between humans and the environment can modify the damages associated with less snow at a scale that is useful to water managers (Kellner and Brunner 2021, Immerzeel *et al* 2020a). Using 13 systems located in the headwaters of major river Basins in the western US , we show that vulnerability can be substantially altered by demand and adaptive capacity over the century. This is particularly apparent in certain tributaries to the Rio Grande (Costilla), Upper Colorado (Bridger and Price), Columbia (Umatilla, Kittitas, and Little Wood), and Missouri Basins (Shoshone and Wind). However, our results also indicate that vulnerability—and adaptive capacity—are unevenly distributed throughout large basins. Possible adaptive pathways for greater reduction in vulnerability include larger hay production relative to overall water demand and smaller changes in surface water storage (i.e., snow relative to built storage). Results indicate that larger transitions from snow dependence to built storage dependence are just as important—if not more important—than overall declines in snow storage.

Consistent with other work at larger scales, lower elevation coastal systems appear to be particularly exposed to declines in snow storage—as measured by changes in streamflow timing (DoQ_{50}) and amount (Q_{WY})—in in the near future (Fritze *et al* 2011, Stewart 2009, Stewart *et al* 2004, McCabe and Clark 2005, Regonda *et al* 2005). However, interior and higher elevation systems see their exposure accelerate in the mid and far future. These findings are also consistent with relevant basin-scale analyses suggesting large potential changes in water supply in the California, Columbia, Upper Colorado, and Rio Grande Basins (Qin *et al* 2020, Mankin *et al* 2015). Possible explanations for systems system buffering against exposure may be compensatory increases in the amount of winter and spring rain and/or mixed precipitation (Hammond and Kampf 2020).

Critically, results indicate that how a system responds to changes in the amount and timing of water supply is more associated with demand characteristics than with snow. That is, sensitivity is more correlated to the projected fraction of hay production relative to overall water demand than to either changes in surface water storage conditions (e.g., increased built storage dependence) or overall declines in snow storage This directly underscores the essential—and often under considered role— of demand in vulnerability analyses (Qin *et al* 2020) and, thus the need for higher quality estimates of projected demand in the western US particularly when it comes to future land use (Mu *et al* 2018, Prestele *et al* 2016). For example, the only crop-level demand data available at the spatial and temporal scale necessary for our analysis was the USBR NIWR data generated using the 3rd phase of the Coupled Model Intercomparison Project (CMIP3) (Meehl *et al* 2007) as opposed to the newer 5th phase of the Coupled Model Intercomparison Project (CMIP5) (Taylor *et al* 2012). Incorporating new advances in climate models into crop-level data available at the western US scale could further improve understanding of supply-demand interactions going forward.

Due to their geography, headwater reservoir systems have two primary pathways for adaptation to changing water supply and demand: 1) enhancing or building new reservoirs and/or groundwater banking using tools like managed aquifer recharge or conjunctive use; and/or 2) reducing water use via reducing cropping acreage or changing crop types (Immerzeel *et al* 2020, He *et al* 2021, Kellner 2021). As the century progresses opportunities for storage management decrease for a number of systems including tributaries of the California Basin (Kaweah), the Upper Colorado Basin (Bridger and Paonia), and Great Basin (Walker) in particular. Far future storage management is promising for a smaller number of systems include tributaries of the Missouri Basin in particular (Shoshone, Sun, and Wind). Our storage metric is not specific to the kind of management strategy pursued, but rather focuses on whether—on average—water recharges existing reservoirs, assuming that if supply is insufficient the system is a poor candidate for storage management. Follow on work could incorporate legal and institutional analysis (as well as hydrogeology if groundwater storage is pursued) in order to determine the most feasible types for storage management. Some types of management (i.e., conjunctive use and managed aquifer recharge) may be more feasible where there is existing infrastructure for flood irrigation and likely more feasible than expansion of built reservoirs in many locations (He *et al* 2021, Kellner 2021). Due to the complexity of water allocation in California, we also exclude California State Water Project and Federal Central Valley Water Project from our analysis; both of which subsidize demand in those systems in addition to substantial groundwater withdrawals. As a result, storage management opportunities in Kern and Kaweah may be more limited if outside sources of water decline which is likely given recent groundwater regulations (California Department of Water Resources 2014).

Demand management opportunities are smaller albeit more consistent over the century with the largest potential in tributaries of the Rio Grande (Costilla), Upper Colorado (Price), Missouri (Wind), and Columbia (Little Wood and Kittitas). Because our demand management land use scenario reflects both changes in crop type and crop acreage, we assume that our indicator captures potential for both crop switching and reduction in acreage. Overall, we show that systems with larger hay production are more adaptive. The caveat to these findings is that our demand management scenario is based on historical changes in crop type and acreage and projected changes in land use are notoriously uncertain (Mu *et al* 2018, Medellín-Azuara *et al* 2007). With that said, our cropping scenarios are assumed to integrate the region's response ongoing megadrought (Ault *et al* 2018).

Without considering adaptive capacity, vulnerability is consistently highest in tributaries of the Columbia (Kittitas and Little Wood), Upper Colorado (Paonia) and Rio Grande (Costilla) consistent with other findings (Qin *et al* 2020, Mankin *et al* 2015, Immerzeel *et al* 2020a). We also find that vulnerability is consistently lowest in tributaries of the Missouri (Sun and Wind) and Upper Colorado (Price). In this, findings reveal the heterogeneity of vulnerability within Basins, nuancing the findings of larger scale conclusions about agricultural vulnerability in particular (e.g., Qin *et al* 2020). Rather than geography, our results emphasize that systems with larger transitions in surface water storage, larger declines in snow storage, and with a higher fraction of water demand driven by hay production tend to be more vulnerable when adaptive capacity is ignored. Accounting for adaptative capacity reduces vulnerability index values by a median of 4.6 points in the near future with a median reduction of 4.2 in the mid and far future. Interestingly, we find that although systems with large hay production tend to be more vulnerable, they are also more adaptive likely because they can respond to changing water supplies in ways that systems with more perennial crops like nuts and fruits cannot.

Continued declines in snow coupled with ongoing intensification of the hydrologic cycle are straining freshwater resources in the western US and beyond. Our results show that continued supply-focused analyses of vulnerability can mask opportunities for adaptation that are currently at hand, reducing the resiliency of tributaries to critical western river basins. Although we find that these opportunities are unevenly distributed within major river basins of the western US, potential reductions in vulnerability through adaptation are largest in hay-dominated systems undergoing a smaller transition in snow storage relative to overall built storage capacity (e.g., several tributaries of the Missouri Basin in particular). While more work is needed to integrate institutional and legal considerations, results indicate that the benefits of adaptation are largest in the immediate future across all systems. On a broader level, we show that vulnerability analyses in these environments must adopt a more demand-focused systems perspective in order to provide managers and policy-makers with actionable information at an operationally meaningful (e.g., irrigation district) scale. Such a pivot could aid managers and policy-makers in in deploying established tools such as MAR (He *et al*., 2021; Sallwey *et al*., 2019; Scanlon *et al*., 2016), demand management via temporary fallowing and/or crop switching (Consulting, 2020; Rising and Devineni, 2020; Schaible and Aillery, 2013), and multi-objective reservoir management via forecast informed reservoir operations (FIRO, Delaney *et al.*, 2020). By rapidly identifying and targeting systems with the largest potential benefits from adaptation, research can help promote more resilient headwater systems in critical basins throughout the western US.

4.6 Supplemental Information

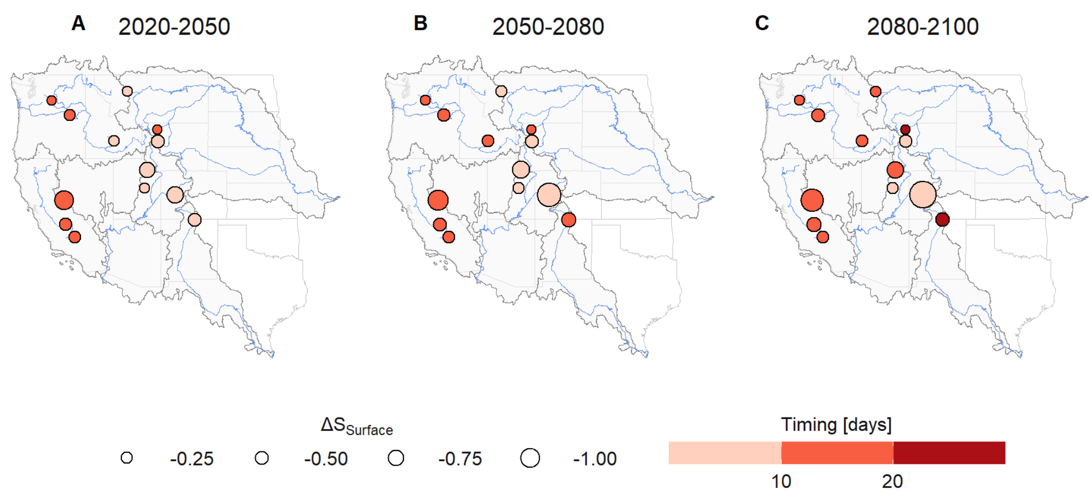

Figure S4- 1: Timing results for 13 systems for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Symbol size is based on $\Delta S_{Surface}$.

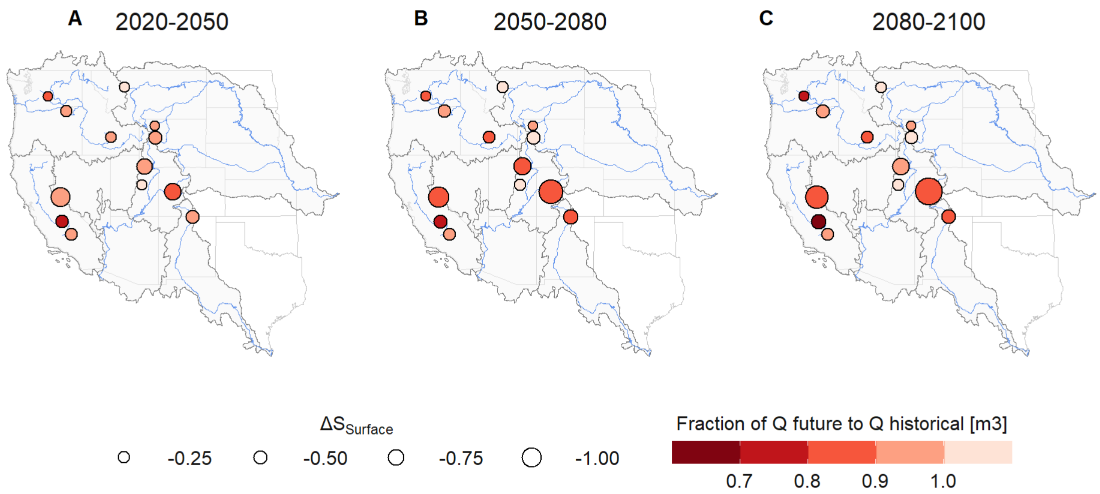

Figure S4- 2: Magnitude results for 13 systems for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Symbol size is based on $\Delta S_{Surface}$.

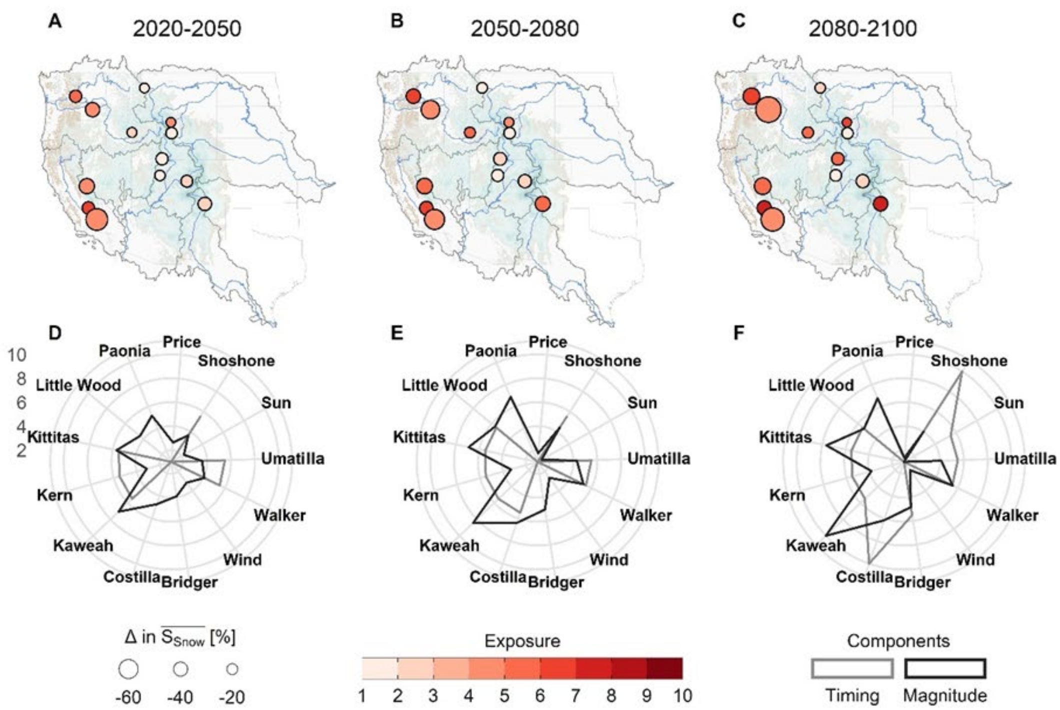

Figure S4-3: Exposure results for 13 systems based on the geometric mean of changes in streamflow timing and magnitude for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Values of re-scaled timing and magnitude indicators are presented for: D) the near future (2020-2050); E) the mid future (2050-2080); F) the far future (2080-2100). Symbol size is based on ΔS_{Snow}.

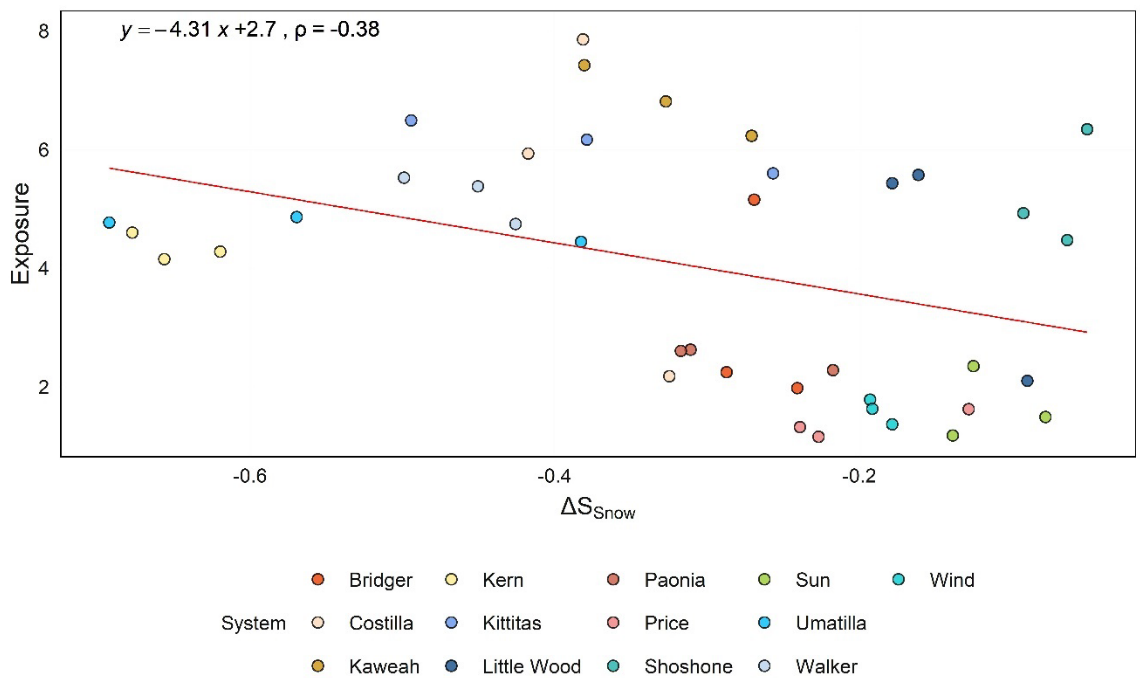

Figure S4-4: ΔS_{Snow}.versus exposure for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

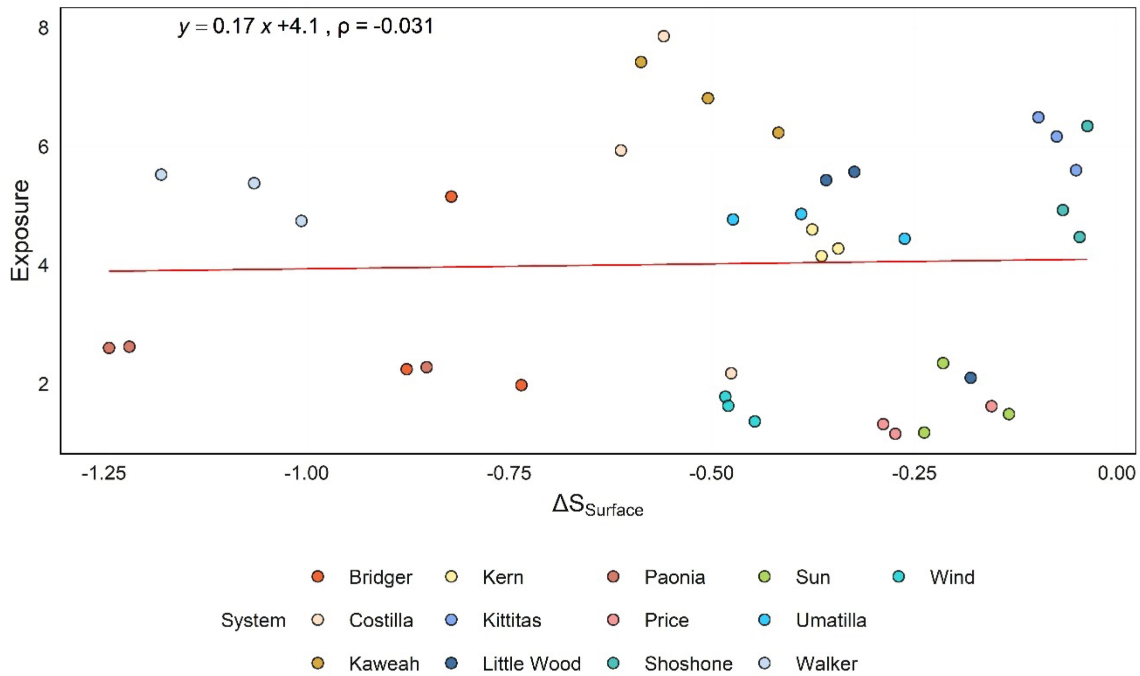

Figure S4-5: $\Delta S_{Surface}$.versus exposure for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

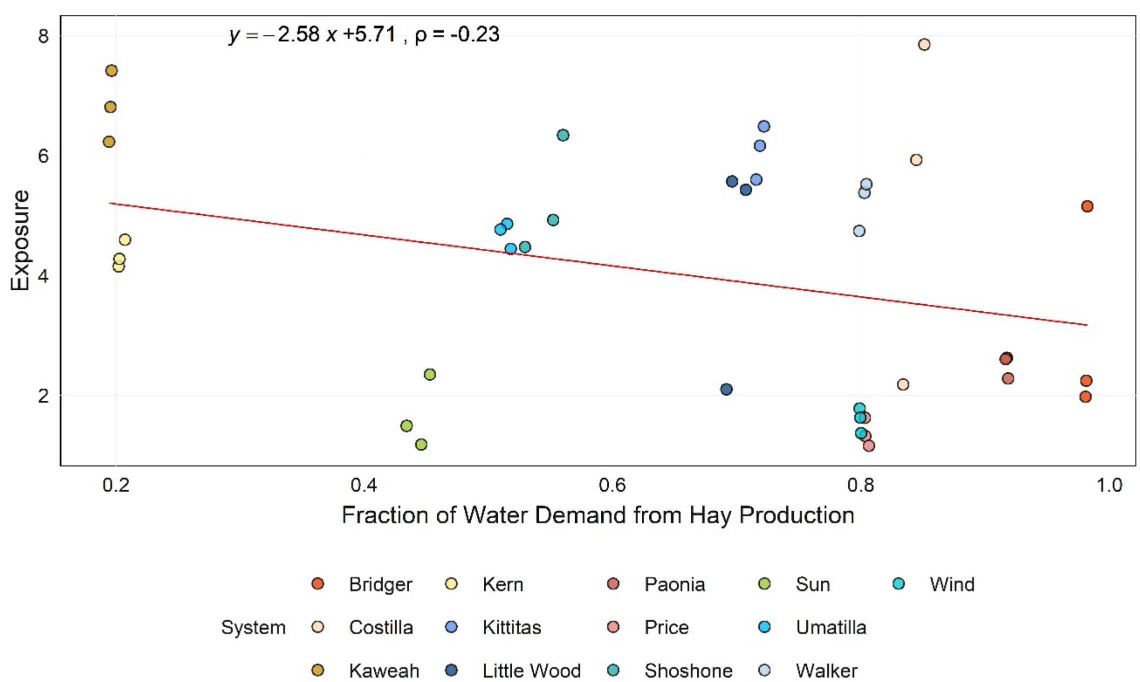

Figure S4-6: Fraction of water demand from hay production versus exposure for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

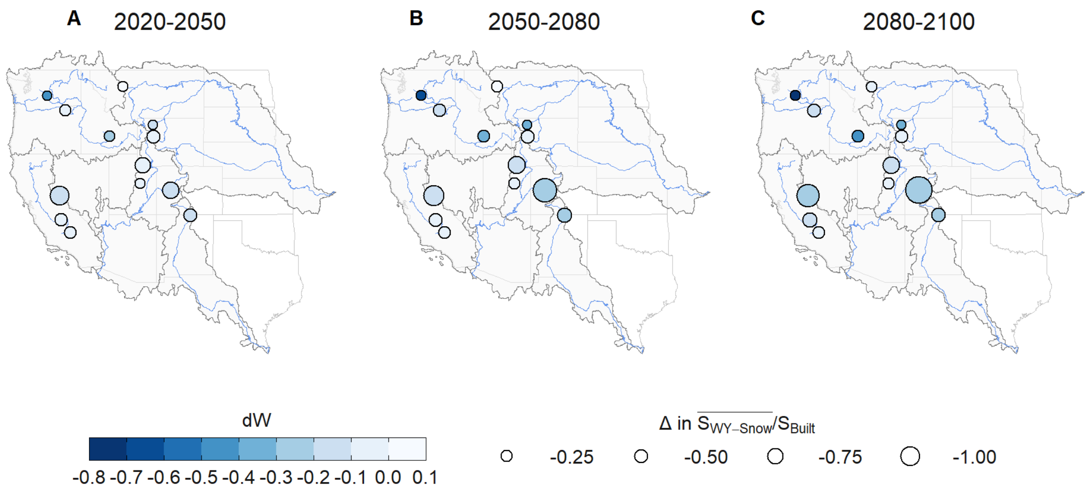

Figure S4-7: Change in well-being (dW) results for 13 systems for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Symbol size is based on $\Delta S_{Surface}$.

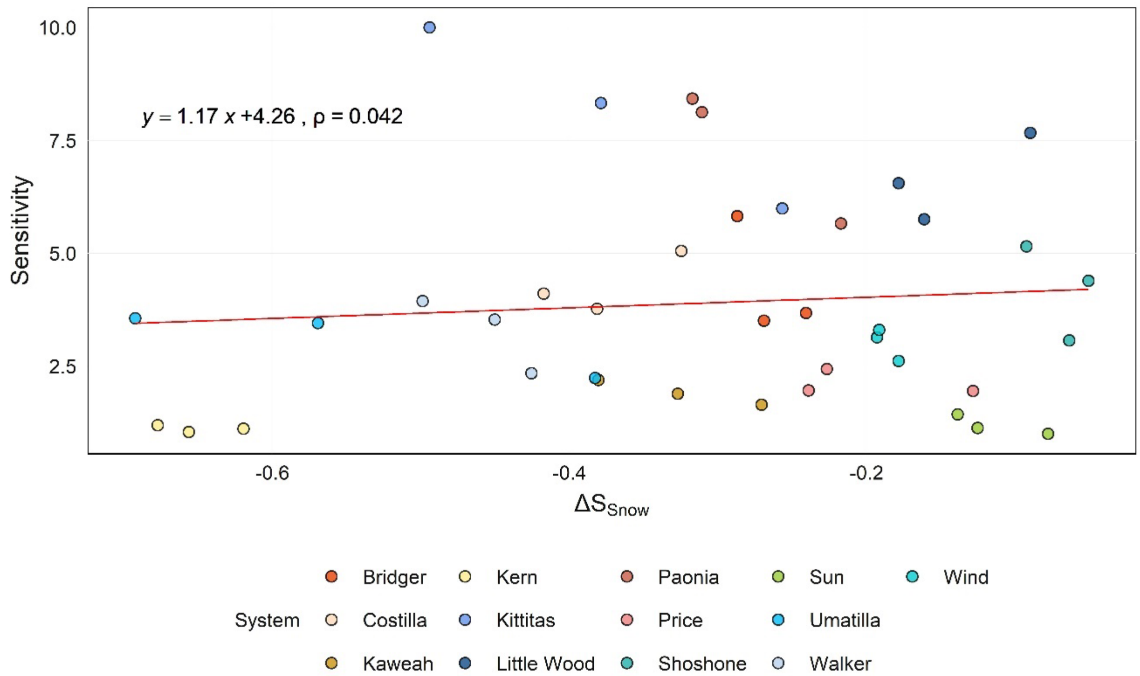

Figure S4-8: ΔS_{Snow}.versus sensitivity for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

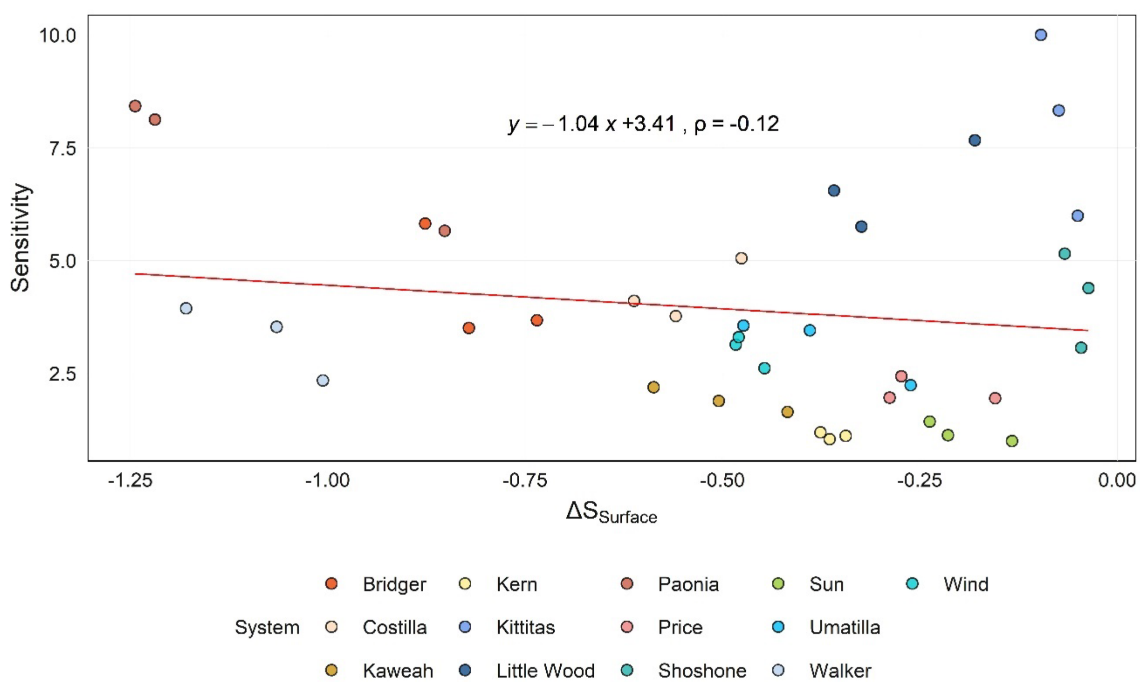

Figure S4-9: $\Delta S_{Surface}$.versus sensitivity for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

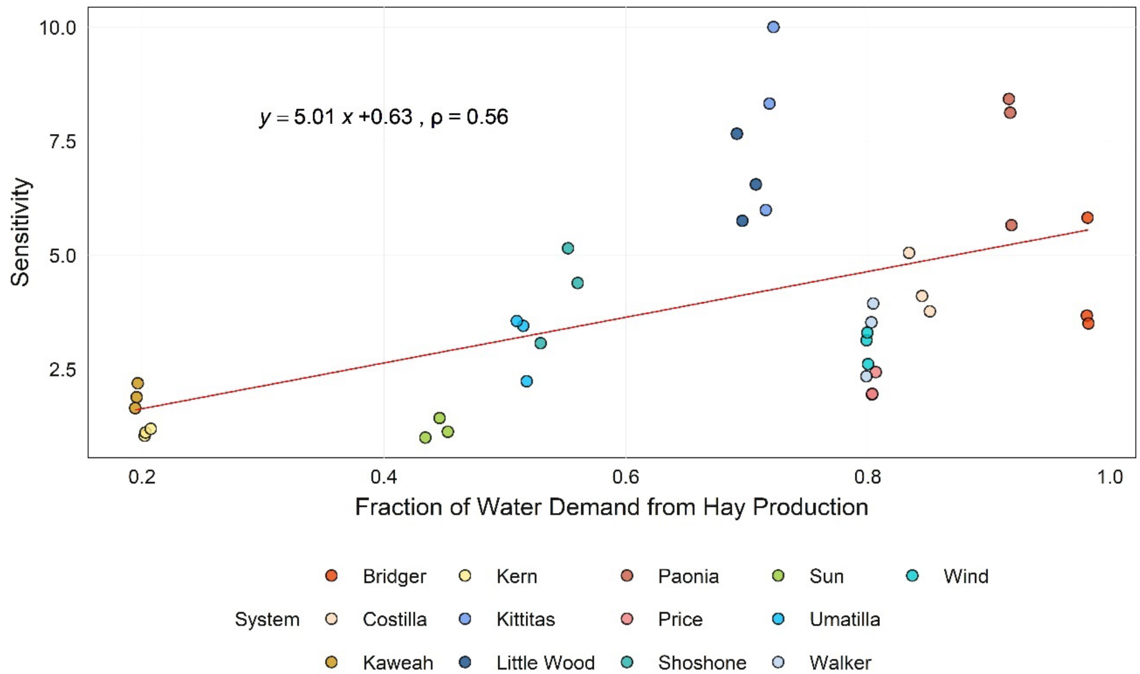

Figure S4- 10: Fraction of water demand from hay production versus sensitivity for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

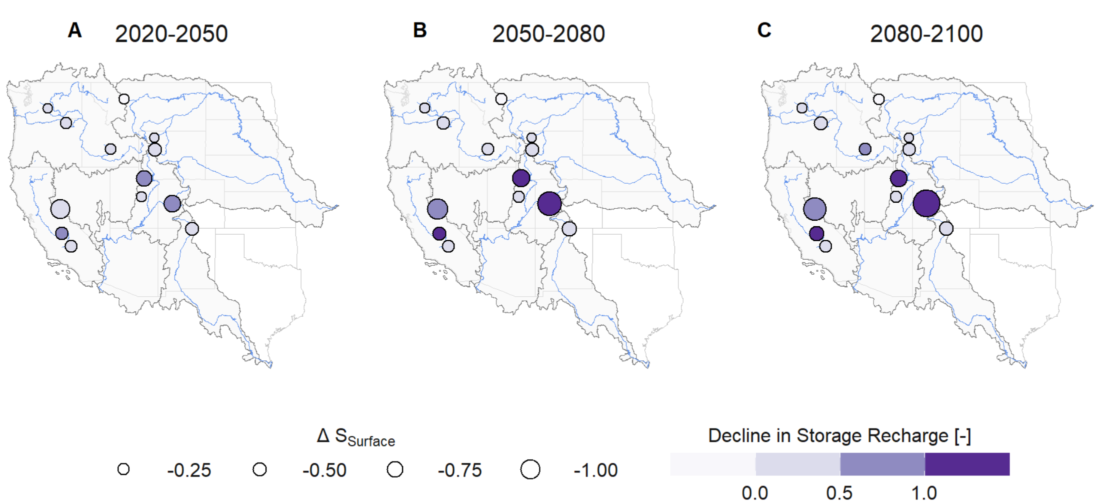

Figure S4-11: Storage management results for 13 systems for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Symbol size is based on $\Delta S_{Surface}$.

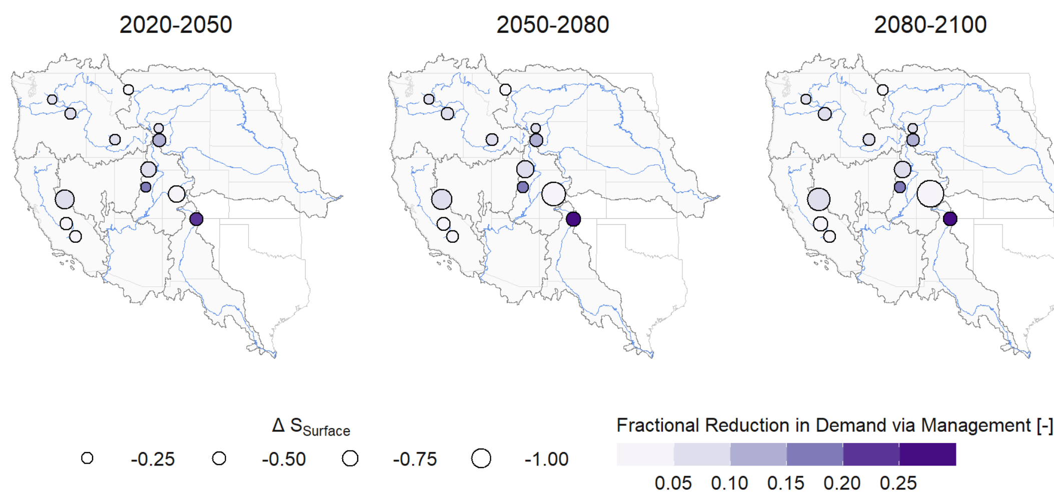

Figure S4-12: Demand management results for 13 systems for: A) the near future (2020-2050); B) the mid future (2050-2080); C) the far future (2080-2100). Symbol size is based on $\Delta S_{Surface}$.

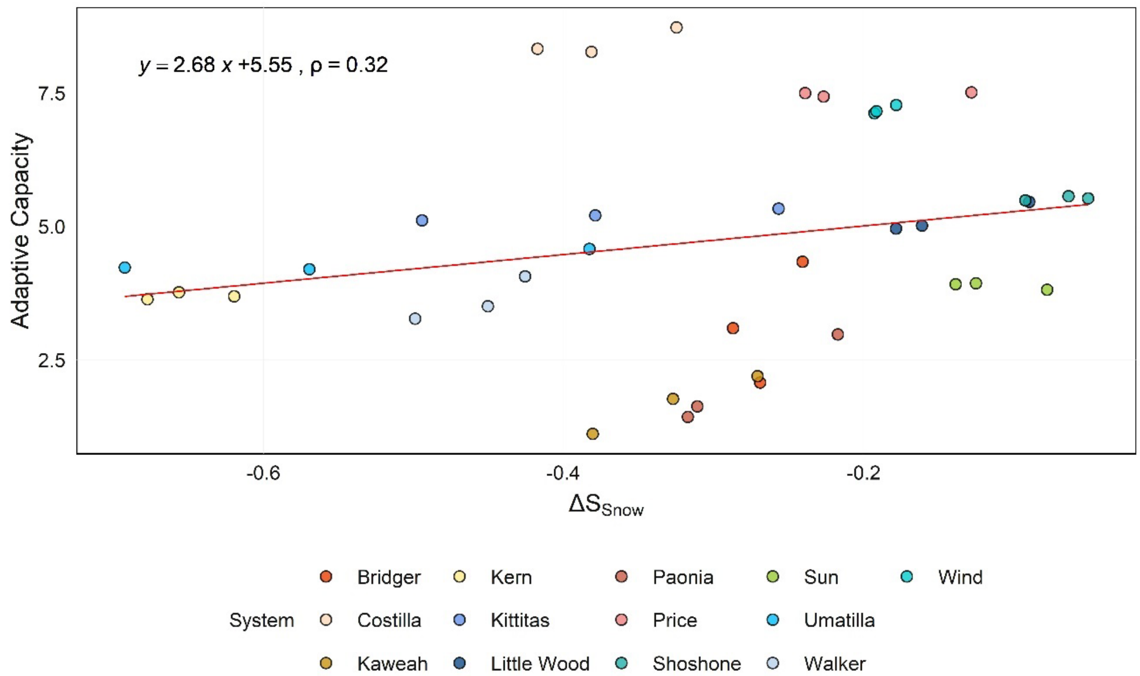

Figure S4-13: ΔS_{Snow}.versus adaptive capacity for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

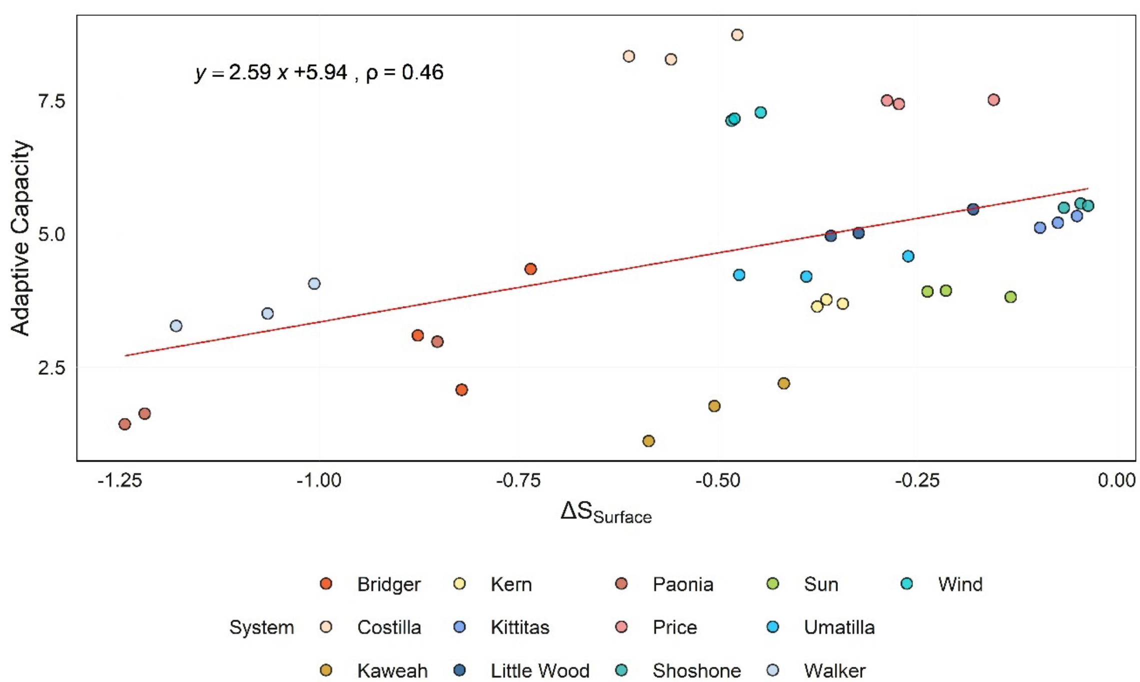

Figure S4-14: $\Delta S_{Surface}$.versus adaptive capacity for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

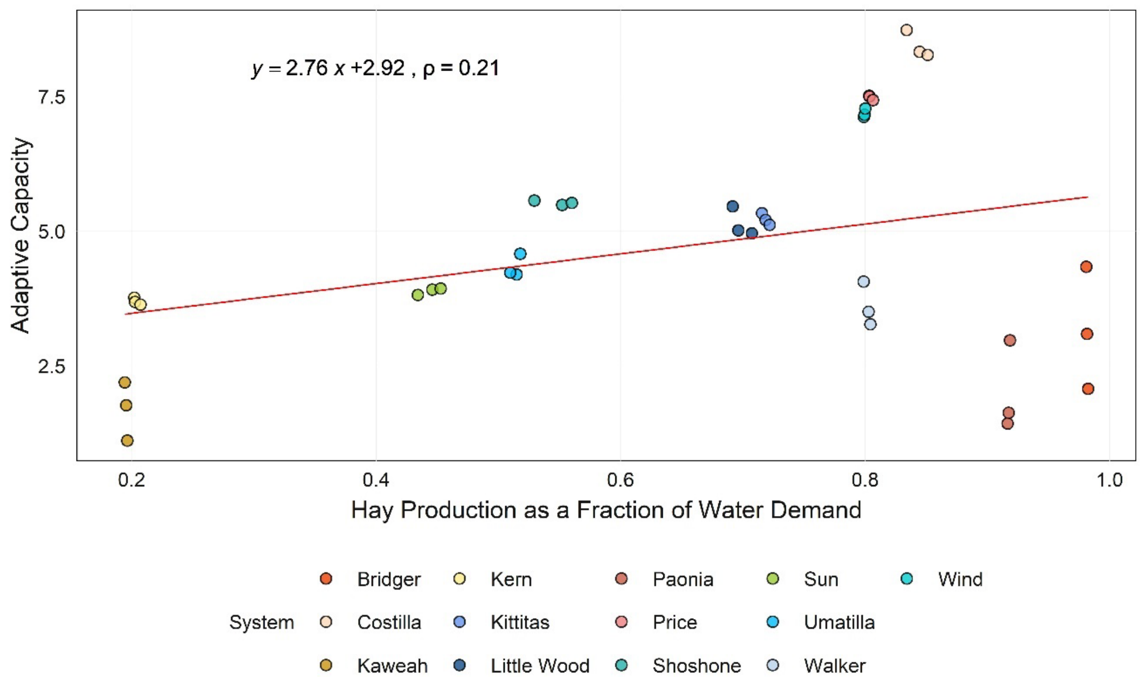

Figure S4-15: Hay production versus adaptive capacity for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

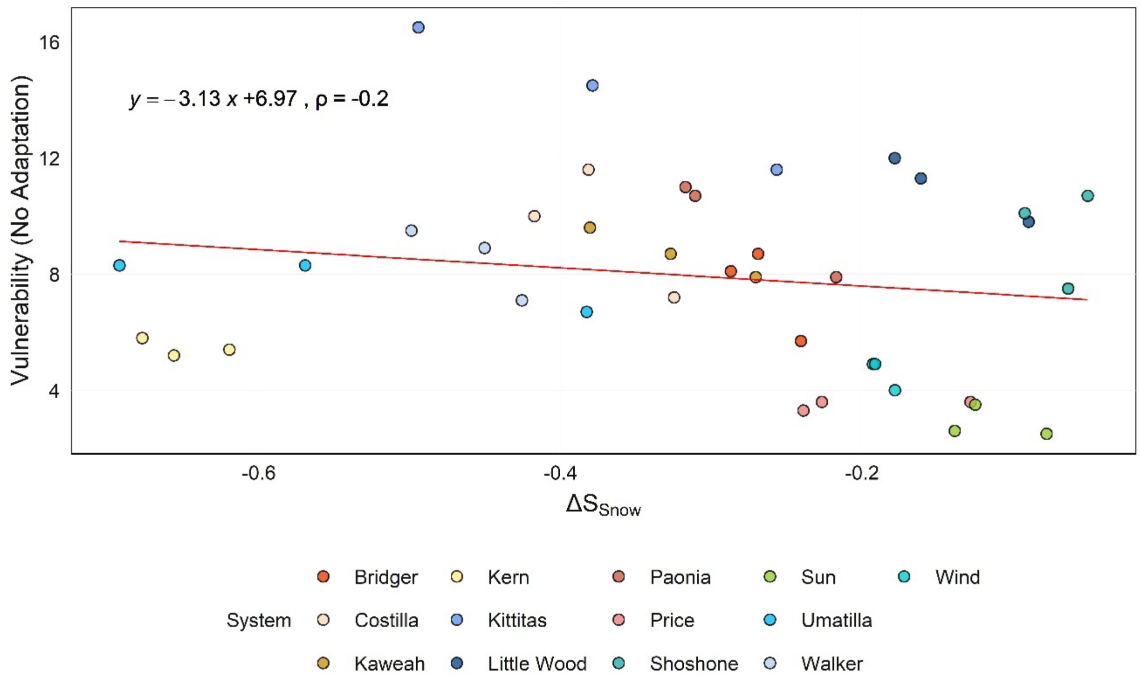

Figure S4-16: ΔS_{Snow}.versus vulnerability without adaptation for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

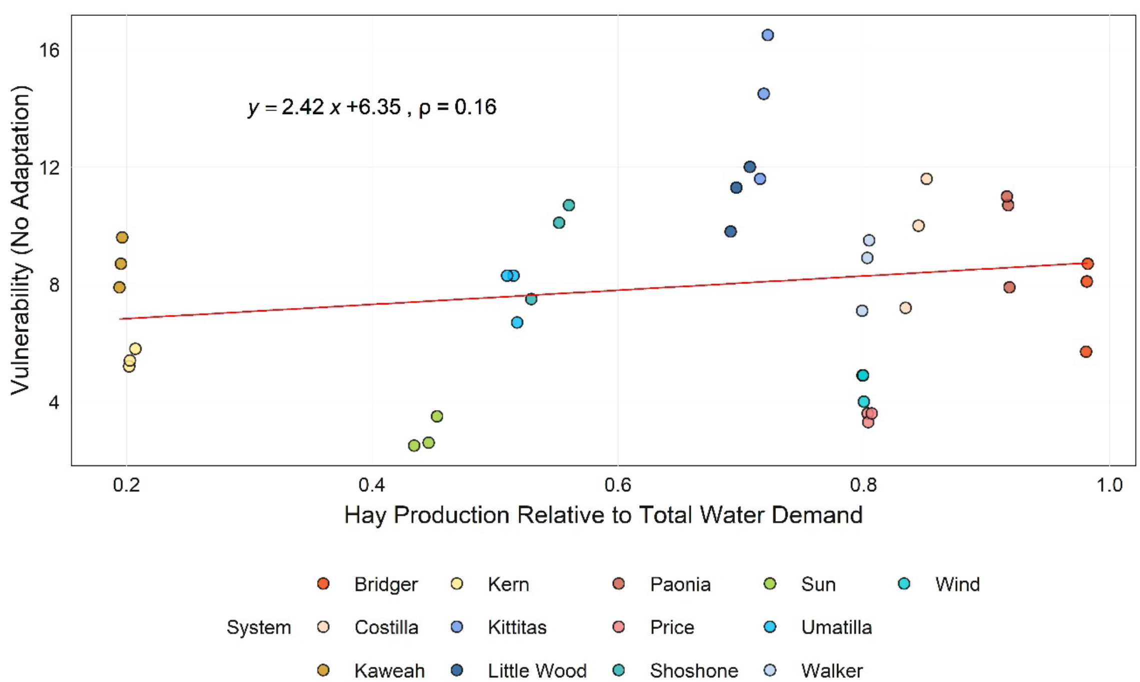

Figure S4- 17: Fraction of water demand from hay production versus vulnerability without adaptation for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

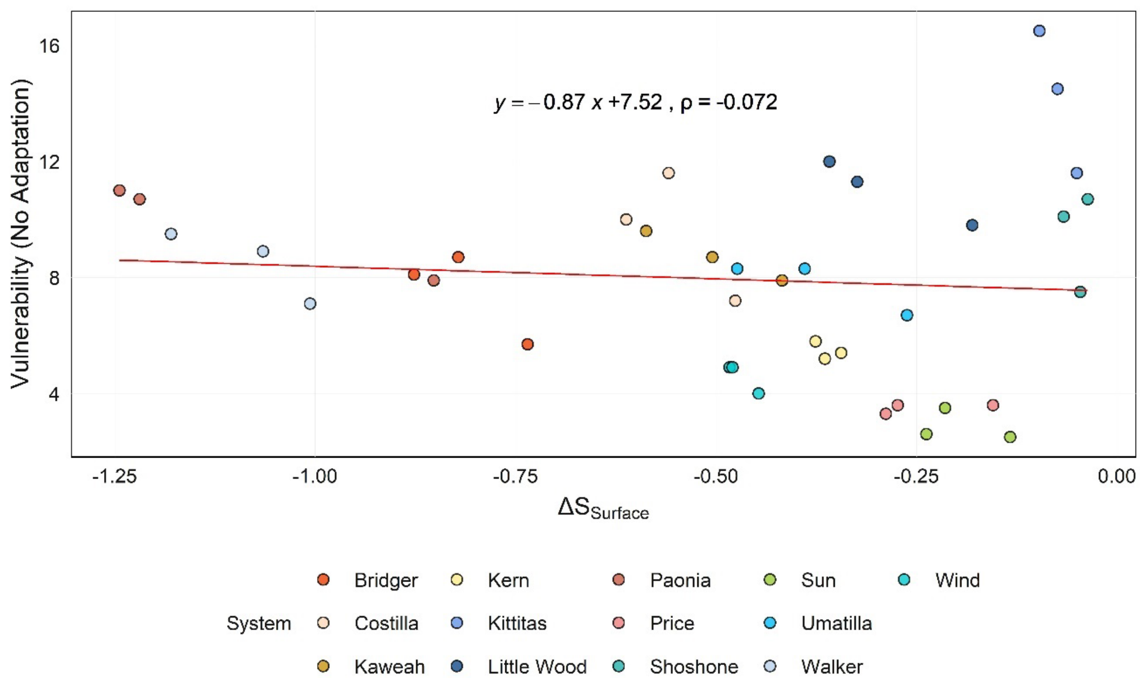

Figure S4-18: $\Delta S_{surface}$.versus vulnerability without adaptation for all systems in all future periods. ρ denotes Spearman rank correlation coefficient.

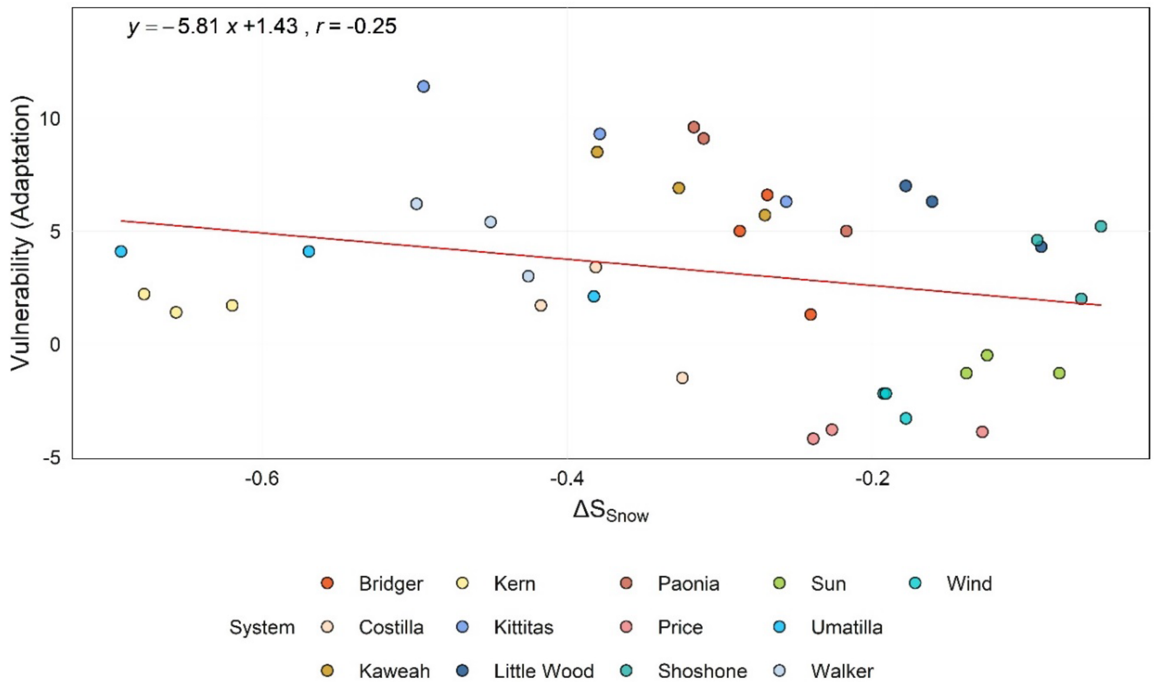

Figure S4-19: ΔS_{Snow}.versus vulnerability with adaptation for all systems in all future periods.

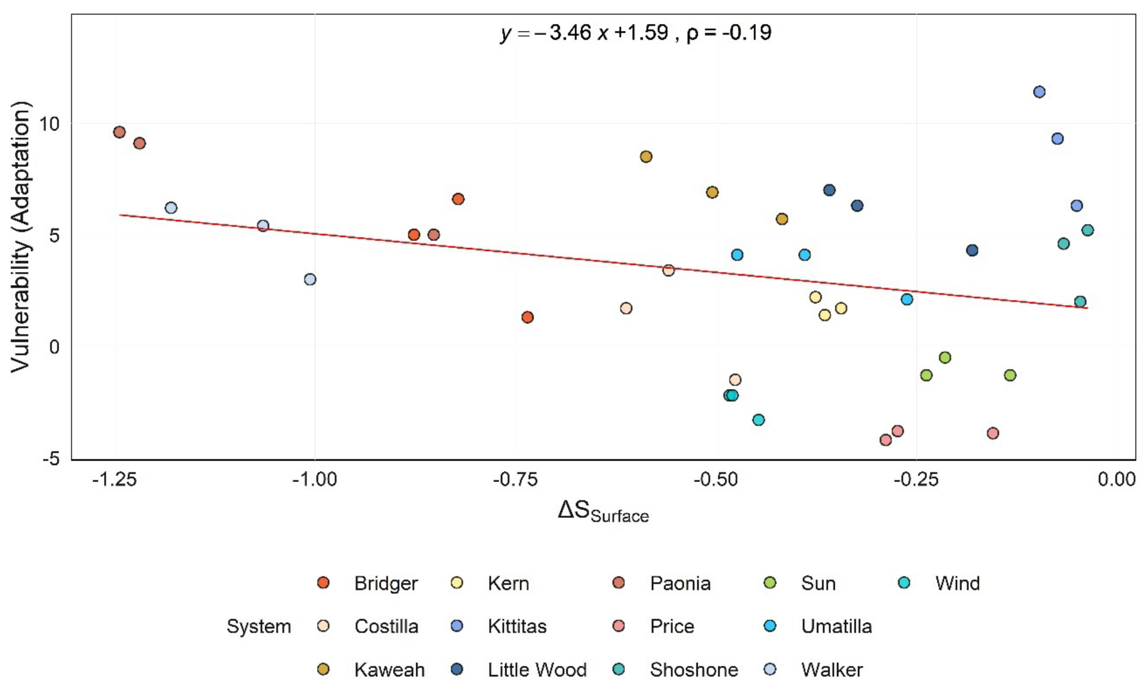

Figure S18: $\Delta S_{Surface}$.versus vulnerability with adaptation for all systems in all future periods.

Table S4- 1: Vulnerability results for 13 systems without and with adaptive capacity based on terciles for all values where Low Vulnerability is defined as the bottom tercile of vulnerability scores (<33rd percentile), Moderate Vulnerability is the middle tercile of vulnerability scores (33rd percentile to 66th percentile), and High Vulnerability is the upper tercile of vulnerability scores (> 66th percentile).

Period	**High Vulnerability (# of systems)**	**Moderate Vulnerability (# of systems)**	**Low Vulnerability (# of systems)**
No Adaptation			
Near Future	4	7	2
Mid Future	9	2	2
Far Future	9	1	3
Adaptation			
Near Future	0	4	9
Mid Future	2	5	6
Far Future	3	5	5

4.7 References

Allen R G, Robison C W, Huntington J, Wright J L and Kilic A 2020 Applying the FAO-56 Dual *Kc* Method for Irrigation Water Requirements over Large Areas of the Western U.S. *Trans. ASABE* **63** 2059–81 Online:

https://elibrary.asabe.org/azdez.asp?JID=3&AID=51945&CID=t2020&v=63&i=6&T=1

Ault T R, George S S, Smerdon J E, Coats S, Mankin J S, Carrillo C M, Cook B I and Stevensong S 2018 A Robust Null Hypothesis for the Potential Causes of Megadrought in Western North America *J. Clim.* **31** 3–24 Online: https://journals.ametsoc.org/view/journals/clim/31/1/jcli-d-17-0154.1.xml

Barnett T P, Adam J C and Lettenmaier D P 2005 Potential impacts of a warming climate on water availability in snow-dominated regions *Nature* **438** 303–9 Online: https://www.nature.com/articles/nature04141

Barnett T P, Pierce D W, Hidalgo H G, Bonfils C, Santer B D, Das T, Bala G, Wood A W, Nozawa T, Mirin A A, Cayan D R and Dettinger M D 2008 Human-induced changes in the hydrology of the Western United States *Science (80-.).* **319** 1080–3 Online: http://science.sciencemag.org/

California Department of Water Resources 2014 SGMA Groundwater Management Online: https://water.ca.gov/Programs/Groundwater-Management/SGMA-Groundwater-Management

Cardona O-D, van Aalst M K, Birkmann J, Fordham M, McGregor G, Perez R, Pulwarty R S, Lisa Schipper E F, Tan Sinh B, Décamps H, Keim M, Davis I, van Aalst M, Birkmann J, Fordham M, McGregor G, Perez R, Pulwarty R, Schipper E, Sinh B, Barros V, Stocker T, Qin D, Dokken D, Ebi K, Mach K, Plattner G, Allen S, Tignor M and Midgley P 2012 Coordinating Lead Authors: Lead Authors: Review Editors: Contributing Authors:: Determinants of risk: exposure and vulnerability. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation 2 Determinants of Risk: Exposure and Vulnerability

Chen J, Brissette F P and Leconte R 2011 Uncertainty of downscaling method in quantifying the impact of climate change on hydrology *J. Hydrol.* **401** 190–202

Consulting B R & 2020 *Upper Basin Demand Management Economic Study in Western Colorado Revised Final Report* Online: www.bbcresearch.com

Davenport F V., Herrera-Estrada J E, Burke M and Diffenbaugh N S 2020 Flood Size Increases Nonlinearly Across the Western United States in Response to Lower Snow-Precipitation Ratios *Water Resour. Res.* **56** e2019WR025571 Online: https://onlinelibrary.wiley.com/doi/full/10.1029/2019WR025571

Delaney, C.J., Hartman, R.K., Mendoza, J., Dettinger, M., Delle Monache, L., Jasperse, J., Ralph, F.M., Talbot, C., Brown, J., Reynolds, D., Evett, S., 2020. Forecast Informed Reservoir Operations Using Ensemble Streamflow Predictions for a Multipurpose Reservoir in Northern California. Water Resour. Res. 56, e2019WR026604. https://doi.org/10.1029/2019WR026604

Dilling L and Berggren J 2015 What do stakeholders need to manage for climate change and variability? A document-based analysis from three mountain states in the Western USA *Springer* Online:

https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/content/pdf/10.1007/s10113-014-0668-y.pdf&casa_token=RBZc3jvrpZAAAAAA:DT4CI4o0sbG0xuzW3Z5D7uyAmwwwvSOPi3D_dkxyV35TvOVe2tjbHm7JZcmy2k-mj7LI7RBJDL9fHg

Dilling L, Lackstrom K, Haywood B, Dow K, Lemos M C, Berggren J and Kalafatis S 2015 What Stakeholder Needs Tell Us about Enabling Adaptive Capacity: The Intersection of Context and Information Provision across Regions in the United States *Weather. Clim. Soc.* **7** 5–17 Online: https://journals.ametsoc.org/view/journals/wcas/7/1/wcas-d-14-00001_1.xml

Ehsani N, Vörösmarty C J, Fekete B M and Stakhiv E Z 2017 Reservoir operations under climate change: Storage capacity options to mitigate risk *J. Hydrol.* **555** 435–46

Elliott J, Deryng D, Müller C, Frieler K, Konzmann M, Gerten D, Glotter M, Flörke M, Wada Y, Best N, Eisner S, Fekete B M, Folberth C, Foster I, Gosling S N, Haddeland I, Khabarov N, Ludwig F, Masaki Y, Olin S, Rosenzweig C, Ruane A C, Satoh Y, Schmid E, Stacke T, Tang Q and Wisser D 2014 Constraints and potentials of future irrigation water availability on agricultural production under climate change *Proc. Natl. Acad. Sci. U. S. A.* **111** 3239–44 Online: www.pnas.org/cgi/doi/10.1073/pnas.1222474110

Fritze H, Stewart I T and Pebesma E 2011 Shifts in western North American snowmelt runoff regimes for the recent warm decades *J. Hydrometeorol.* **12** 989–1006 Online: http://jisao.washington.edu/pdo/PDO.latest

Gonzales P and Ajami N 2017 Social and Structural Patterns of Drought-Related Water Conservation and Rebound *Water Resour. Res.* **53** 10619–34 Online: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017WR021852

Gordon B L, Brooks P D, Krogh S A, Boisrame G F S, Carroll R W H, McNamara J P and Harpold A A 2022 Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework *Environ. Res. Lett.* **17** 053004 Online: https://iopscience.iop.org/article/10.1088/1748-9326/ac64b4

Hamlet A F and Lettenmaier D P 1999 Effects of climate change on hydrology and water resources in the Columbia River Basin *Journal of the American Water Resources Association* vol 35 (American Water Resources Assoc) pp 1597–623 Online: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1752-1688.1999.tb04240.x

Hammond J C and Kampf S K 2020 Subannual Streamflow Responses to Rainfall and Snowmelt Inputs in Snow-Dominated Watersheds of the Western United States *Water Resour. Res.* **56** Online: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019WR026132

Harpold A, Brooks P, Rajagopal S, Heidbuchel I, Jardine A and Stielstra C 2012 Changes in snowpack accumulation and ablation in the intermountain west *Water Resour. Res.* **48** Online: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012WR011949

He X, Bryant B P, Moran T, Mach K J, Wei Z and Freyberg D L 2021 *Climate-informed hydrologic modeling and policy typology to guide managed aquifer recharge* vol 7 Online: http://advances.sciencemag.org/

Heikkila T 2003 Coordination in water resource management: The impact of water rights institutions *Water Policy* **5** 331–48

Herrera-Estrada J E, Martinez J A, Dominguez F, Findell K L, Wood E F and Sheffield J 2019 Reduced Moisture Transport Linked to Drought Propagation Across North America *Geophys. Res. Lett.* **46** 5243–53 Online: https://onlinelibrary.wiley.com/doi/full/10.1029/2019GL082475

Hogeboom R J, Knook L and Hoekstra A Y 2018 The blue water footprint of the world's artificial reservoirs for hydroelectricity, irrigation, residential and industrial water supply, flood protection, fishing and recreation *Adv. Water Resour.* **113** 285–94

Immerzeel W W, Lutz A F, Andrade M, Bahl A, Biemans H, Bolch T, Hyde S, Brumby S, Davies B J, Elmore A C, Emmer A, Feng M, Fernández A, Haritashya U, Kargel J S, Koppes M, Kraaijenbrink P D A, Kulkarni A V., Mayewski P A, Nepal S, Pacheco P, Painter T H, Pellicciotti F, Rajaram H, Rupper S, Sinisalo A, Shrestha A B, Viviroli D, Wada Y, Xiao C, Yao T and Baillie J E M 2020a Importance and vulnerability of the world's water towers *Nature* **577** 364–9 Online: http://dx.doi.org/10.1038/s41586-019-1822-y

Immerzeel W W, Lutz A F, Andrade M, Bahl A, Biemans H, Bolch T, Hyde S, Brumby S, Davies B J, Elmore A C, Emmer A, Feng M, Fernández A, Haritashya U, Kargel J S, Koppes M, Kraaijenbrink P D A, Kulkarni A V, Mayewski P A, Nepal S, Pacheco P, Painter T H, Pellicciotti F, Rajaram H, Rupper S, Sinisalo A, Shrestha A B, Viviroli D, Wada Y, Xiao C, Yao T and Baillie J E M 2020b Importance and vulnerability of the world's water towers *364* | *Nat.* | **577** Online: https://doi.org/10.1038/s41586-019-1822-y

Jeton A E, Dettinger M D and Smith J L 1996 *Potential effects of cliamte change on streamflow,easter and western slopes of the Sierra Nevada, California and Nevada (Vol., 95, No. 4260)*

Kellner E 2021 The controversial debate on the role of water reservoirs in reducing water scarcity *Wiley Interdiscip. Rev. Water* **8** e1514 Online: https://doi.org/10.1002/wat2.1514

Kellner E and Brunner M I 2021 Reservoir Governance in World's Water Towers Needs to Anticipate Multi-purpose Use *Earth's Futur.* **9** 1–19 Online: https://doi.

Klos P Z, Link T E and Abatzoglou J T 2014 Extent of the rain-snow transition zone in the western U.S. under historic and projected climate *Geophys. Res. Lett.* **41** 4560–8 Online: http://doi.wiley.com/10.1002/2014GL060500

Knowles N, Dettinger M D and Cayan D R 2006 Trends in snowfall versus rainfall in the western United States *J. Clim.* **19** 4545–59 Online:

http://journals.ametsoc.org/jcli/article-pdf/19/18/4545/3798835/jcli3850_1.pdf

Kumar S V., Jasinski M, Mocko D M, Rodell M, Borak J, Li B, Beaudoing H K and Peters-Lidard C D 2019 NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment *J. Hydrometeorol.* **20** 1571–93

Lehner B, Liermann C R, Revenga C, Vörösmarty C, Fekete B, Crouzet P, Döll P, Endejan M, Frenken K, Magome J, Nilsson C, Robertson J C, Rödel R, Sindorf N and Wisser D 2011 Global Reservoir and Dam (GRanD) database Online: http://sedac.ciesin.columbia.edu/pfs/grand.html

Luers A L, Lobell D B, Sklar L S, Addams C L and Matson P A 2003 A method for quantifying vulnerability, applied to the agricultural system of the Yaqui Valley, Mexico *Glob. Environ. Chang.* **13** 255–67

Lundquist J D, Neiman P J, Martner B, White A B, Gottas D J and Ralph F M 2008 Rain versus Snow in the Sierra Nevada, California: Comparing Doppler Profiling Radar and Surface Observations of Melting Level *J. Hydrometeorol.* **9** 194–211 Online: https://journals.ametsoc.org/view/journals/hydr/9/2/2007jhm853_1.xml

Mankin J S and Diffenbaugh N S 2015 Influence of temperature and precipitation variability on near-term snow trends *Clim. Dyn.* **45** 1099–116 Online: www.earthsystemgrid.org;

Mankin J S, Viviroli D, Singh D, Hoekstra A Y and Diffenbaugh N S 2015 The potential for snow to supply human water demand in the present and future *Environ. Res. Lett.* **10** 114016 Online: https://iopscience.iop.org/article/10.1088/1748-9326/10/11/114016

Masia S, Sušnik J, Marras S, Mereu S, Spano D and Trabucco A 2018 Assessment of irrigated agriculture vulnerability under climate change in Southern Italy *Water (Switzerland)* **10**

McCabe G J and Clark M P 2005 Trends and variability in snowmelt runoff in the western United States *J. Hydrometeorol.* **6** 476–82 Online: http://www.cru.uea.ac.uk/cru/

Medellín-Azuara J, Harou J J, Olivares M A, Madani K, Lund J R, Howitt R E, Tanaka S K, Jenkins M W and Zhu T 2007 Adaptability and adaptations of California's water supply system to dry climate warming *Clim. Change* **87** 75–90 Online: http://cee.engr.ucdavis.edu/faculty/lund/CALVIN/

Meehl G A, Covey C, Delworth T, Latif M, McAvaney B, Mitchell J F B, Stouffer R J and Taylor K E 2007 THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research *Bull. Am. Meteorol. Soc.* **88** 1383–94 Online: https://journals.ametsoc.org/view/journals/bams/88/9/bams-88-9-1383.xml

Minville M, Brissette F and Leconte R 2008 Uncertainty of the impact of climate change on the hydrology of a nordic watershed *J. Hydrol.* **358** 70–83

Mizukami N, Clark M P, Sampson K, Nijssen B, Mao Y, McMillan H, Viger R J, Markstrom S L, Hay L E, Woods R, Arnold J R and Brekke L D 2016 MizuRoute version 1: A river network routing tool for a continental domain water resources applications *Geosci. Model Dev.* **9** 2223–8

Mu J E, McCarl B A, Sleeter B, Abatzoglou J T and Zhang H 2018 Adaptation with climate uncertainty: An examination of agricultural land use in the United States *Land use policy* **77** 392–401

Musselman K N, Clark M P, Liu C, Ikeda K and Rasmussen R 2017 Slower snowmelt in a warmer world *Nat. Clim. Chang.* **7** 214–9 Online: www.nature.com/natureclimatechange

Nijssen B, O'donnell G M, Hamlet A F and Lettenmaier D P 2001 Hydrologic sensitivity of global rivers to climate change *Clim. Change* **50** 143–75 Online: https://link.springer.com/article/10.1023/A:1010616428763

Olmstead S M 2014 Climate change adaptation and water resource management: A review of the literature *Energy Econ.* **46** 500–9

Pierce D W, Cayan D R and Thrasher B L 2014 Statistical Downscaling Using Localized Constructed Analogs (LOCA) *J. Hydrometeorol.* **15** 2558–85 Online: https://journals.ametsoc.org/view/journals/hydr/15/6/jhm-d-14-0082_1.xml

Prestele R, Alexander P, Rounsevell M D A, Arneth A, Calvin K, Doelman J, Eitelberg D A, Engström K, Fujimori S, Hasegawa T, Havlik P, Humpenöder F, Jain A K, Krisztin T, Kyle P, Meiyappan P, Popp A, Sands R D, Schaldach R, Schüngel J, Stehfest E, Tabeau A, Van Meijl H, Van Vliet J and Verburg P H 2016 Hotspots of uncertainty in land-use and land-cover change projections: a global-scale model comparison *Glob. Chang. Biol.* **22** 3967–83 Online: https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.13337

Qin Y, Abatzoglou J T, Siebert S, Huning L S, AghaKouchak A, Mankin J S, Hong C, Tong D, Davis S J and Mueller N D 2020 Agricultural risks from changing snowmelt *Nat. Clim. Chang.* **10** 459–65 Online: https://doi.org/10.1038/s41558-020-0746-8

Rauscher S A, Pal J S, Diffenbaugh N S and Benedetti M M 2008 Future changes in snowmelt-driven runoff timing over the western US *Geophys. Res. Lett.* **35** Online: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2008GL034424

Regonda S K, Rajagopalan B, Clark M and Pitlick J 2005 Seasonal cycle shifts in hydroclimatology over the western United States *J. Clim.* **18** 372–84 Online: https://journals.ametsoc.org/view/journals/clim/18/2/jcli-3272.1.xml

Rising J and Devineni N 2020 Crop switching reduces agricultural losses from climate change in the United States by half under RCP 8.5 *Nat. Commun. 2020 111* **11** 1–7 Online: https://www.nature.com/articles/s41467-020-18725-w

Sallwey, J., Bonilla Valverde, J.P., Vásquez López, F., Junghanns, R., Stefan, C., 2019. Suitability maps for managed aquifer recharge: A review of multi-criteria decision analysis studies.

Environ. Rev. https://doi.org/10.1139/er-2018-0069

Scanlon, B.R., Reedy, R.C., Faunt, C.C., Pool, D., Uhlman, K., 2016. Enhancing drought resilience with conjunctive use and managed aquifer recharge in California and Arizona. Environ. Res. Lett. 11, 035013. https://doi.org/10.1088/1748-9326/11/3/035013

Schaible, G., Aillery, M., 2013. Water Conservation in Irrigated Agriculture: Trends and Challenges in the Face of Emerging Demands. SSRN Electron. J. https://doi.org/10.2139/ssrn.2186555

Stewart I T 2009 Changes in snowpack and snowmelt runoff for key mountain regions *Hydrol. Process.* **23** 78–94

Stewart I T, Cayan D R and Dettinger M D 2004 Changes in snowmelt runoff timing in western North America under a "business as usual" climate change scenario *Clim. Change* **62** 217–32 Online: https://link.springer.com/article/10.1023/B:CLIM.0000013702.22656.e8

Stewart I T, Cayan D R and Dettinger M D 2005 Changes toward earlier streamflow timing across western North America *J. Clim.* **18** 1136–55 Online: http://tao.

Sullivan C A 2011 Quantifying water vulnerability: A multi-dimensional approach *Stoch. Environ. Res. Risk Assess.* **25** 627–40 Online: https://link.springer.com/article/10.1007/s00477-010-0426-8

Tarboton, D. G. (2005). Terrain analysis using digital elevation models (TauDEM). Utah State University, Logan, 3012, 2018.

Taylor K E, Stouffer R J and Meehl G A 2012 An Overview of CMIP5 and the Experiment Design *Bull. Am. Meteorol. Soc.* **93** 485–98 Online: https://journals.ametsoc.org/view/journals/bams/93/4/bams-d-11-00094.1.xml

U.S. Geological Survey, 2016, National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), accessed [June 10, 2012], at URL [http://waterdata.usgs.gov/nwis/].

Vano, J., Hamman, J., Gutmann, E., Wood, A., Mizukami, N., Clark, M., ... & Arnold, J. (2020). Comparing Downscaled LOCA and BCSD CMIP5 Climate and Hydrology Projections-Release of Downscaled LOCA CMIP5 Hydrology. N/A, 96.

Zhu T, Jenkins M W and Lund J R 2005 ESTIMATED IMPACTS OF CLIMATE WARMING ON CALIFORNIA WATER AVAILABILITY UNDER TWELVE FUTURE CLIMATE SCENARIOS1 *JAWRA J. Am. Water Resour. Assoc.* **41** 1027–38 Online: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1752-1688.2005.tb03783.

5 Chapter 5: Designing dynamic indicator-based assessments of water supply vulnerability

Abstract:

Climatic and societal stressors are straining freshwater resources for both people and the environment. One lens for assessing the impacts of these stressors on water resources systems is *vulnerability*—commonly defined as the susceptibility of people and places to damage. To quantify vulnerability, water managers and policy-makers have long turned to vulnerability indices, which rely on proxy measures (i.e., indicators) of system performance. However, these indices are often poorly equipped to measure vulnerability in a multidimensional (i.e., physical, social, political, etc.) and dynamic way. Here, we develop a generalizable approach for evaluating water supply vulnerability that can be implemented by managers and policy-makers in a range of systems. We construct an open-source database of existing indices and indicators from global literature and evaluate these data to identify core elements for indicator-based assessment. We then perform a number of analyses on our data to ensure that our approach facilitates multidimensional assessment of vulnerability. Through a feedback loop with our database, we then show how this approach can be updated to account for changes in vulnerability due to social, political, and environmental stresses. Results outline a pathway for users to construct holistic, flexible, and practical vulnerability assessments in a range of settings. However, we find that users will need to overcome strong bias towards physical measures of system and gaps in measures of cultural water use and values, urban water use, and groundwater if relying on existing data. . In this, our work underscores the need for the research community to pivot away from the continued development of one-off indices and towards the

construction and evaluation of more diverse and locally-relevant indicators that can be integrated into flexible approaches like the one we present here. Critically, this points to the pressing need for more comprehensive data to evaluate the social value of water.

5.1 Introduction

Climatic and social stressors are straining the availability of freshwater supply for people and the environment (Vörösmarty *et al*., 2010, 2000). Climate change is, for example, altering historical precipitation dynamics in critical mountain environments (Gordon *et al*., 2022; Immerzeel *et al*., 2020; Mankin *et al*., 2015; Qin *et al*., 2020), increasing extreme hydrologic behavior (Davenport *et al*., 2020; Herrera-Estrada *et al*., 2019), challenging reservoir operations (Ehsani *et al*., 2017), and straining groundwater resources (Kumar, 2012). These changes in physical water supply interact with social, political, and economic conditions to either amplify or moderate resulting damages to people and the environment (Enqvist *et al*., 2019; O'Brien and Leichenko, 2000; Savelli *et al*., 2021; Scott *et al*., 2021; Zuniga-Teran *et al*., 2021). As such, there is longstanding interest in assessing the relative susceptibility of water resource systems—or the hydrologic, ecologic, infrastructural, and human processes involving water in a given place (Brown *et al*., 2015; Marlow *et al*., 2013)—to damage arising from socio-hydrologic stress (Gleick, 1993; Hashimoto *et al*., 1982; Hurd *et al*., 1999).

The susceptibility of such systems to damages arising from these interacting stressors is often described in terms of its *vulnerability* (Adger, 2006; Brooks, 2003). In practice, vulnerability is assessed to identify systems of greatest concern and subsequently prioritize adaptation activities intended to reduce potential damages (Luers, 2005). Although clear consensus on a single definition of vulnerability has long proven elusive (Cutter and Finch, 2008), it is often characterized as a framework that incorporates at least one of the following: sensitivity, exposure, stress or disturbance, state of the system relative to a threshold, or ability of the system to adapt to changing conditions (Eakin and Luers, 2006; Luers, 2005). The practicality of these numerous conceptual vulnerability frameworks (e.g., Adger, 2006; Adger and Kelly, 1999; Birkmann *et al*., 2013; Blaikie *et al*., 2005; Brooks, 2003; Cardona, 2011; Kasperson *et al*., 2012; Kelly and Adger, 2000; Luers *et al*., 2003; Turner *et al*., 2003), however, remains challenging due to a lack of clear and precise guidance on how they can actually be implemented in a bottom-up manner (i.e., by managers and policy-makers to assess vulnerability in a specific system) (Hughes *et al*., 2012; Luers, 2005; Sullivan, 2011).

A frequently prescribed approach for practical evaluations of vulnerability is the indicator-based assessment, which quantifies vulnerability using a set number of indirect measures assumed to capture the 'spirit of vulnerability'—termed proxy indicators (henceforth, indicators) (Sullivan, 2011). This assumption, coupled with the long-standing use of indicators in policy and decision-making (Plummer *et al*., 2012; Sullivan and Meigh, 2005), has led to the development of myriad indices for assessing water supply vulnerability in a diversity of contexts and scales (e.g., Hamouda *et al* 2009, Okpara *et al* 2016, Alessa *et al* 2008, Sullivan 2011). However, many of these indices were developed in a top-down manner and designed to facilitate relative comparisons at the municipal to national-level using numerical targets (e.g., Water Poverty Index (Lawrence *et al*., 2002), Water Vulnerability Index (Sullivan, 2011)). Increasingly, however, indices have pivoted away from this 'one-size-fits-all' approach using normative measures and towards assessments that capture the context, place, and time-specific nature of vulnerability (e.g., Composite Climate-Water Conflict Vulnerability Index (Okpara *et al*., 2017). Recent research (e.g., Anandhi and Kannan, 2018) has laid a foundation for how flexible conceptual approaches to index design can aid in this transition away from 'one-size-fits-all' approaches.

Despite these advances, the acceleration and complexity of socio-hydrologic stresses challenge the utility of existing paradigms for bottom-up indicator-based water supply vulnerability assessments in two specific ways. First, developed indices—and as a result, the indicators they rely upon to quantify vulnerability—are often poorly equipped to consider multidimensional (i.e., social, economic, physical, cultural, environmental, and institutional) aspects of vulnerability even when implemented in a bottom-up manner. For example, many existing indices place a strong emphasis on physical measures of system performance where data for evaluation are more readily available (de Ruiter and van Loon, 2022; Notaro *et al*., 2014), and a reduced focus on more difficult-to-measure social, political, and economic measures. Furthermore, there is often limited guidance around how to integrate missing context-appropriate and multidimensional indicators of system performance when deficiencies are identified by potential users (Okpara *et al*., 2016). Second, despite recognition of vulnerability as a dynamic concept that must be evaluated with feedback-loops in a comprehensive manner (Birkmann *et al*., 2013; Cardona, 2011), developed indices often adopt a static view of water supply vulnerability in relation to climate and societal change (Adger, 2006; de Ruiter and van Loon, 2022). Existing indices can, for example, account for change with respect to each indicator (e.g., annual precipitation can vary as more data are incorporated). However, they fail to include an emphasis on the underlying dynamics of vulnerability (e.g., immigration, instability, displacement), long-lasting disasters (e.g., drought, pandemics), and/or compounding vulnerability (e.g., multiple disasters) (de Ruiter and van Loon, 2022). Critically, failure to consider the multidimensional and dynamic nature of vulnerability when assessing vulnerability can lead to an uncomprehensive snapshot of vulnerability (Birkmann *et al*., 2013), hindering the development of effective, locally-relevant, and just adaptation strategies to changing water resources (Dilling and Berggren, 2015; van den Berg and Keenan, 2019; Zuniga-Teran *et al*., 2022, 2021).

As such, there is a need to integrate these aspects of multidimensionality and dynamic vulnerability into new assessments while retaining the practical benefits of indicators for water managers and policy-makers. This requires that conceptual advances be clearly translated into practical approaches—and, critically, resources for implementing these approaches—that can be implemented in a bottom-up manner across different systems (Anandhi and Kannan, 2018; Dilling and Berggren, 2015; Gallopín, 2006; Sullivan, 2011). In recognition of this gap, our study is motivated to develop a practical, bottom-up approach for quantifying water supply vulnerability in a multidimensional and dynamic manner. We distill this motivation into a central motivating question:

1. How can indicator-based assessments be adapted to measure water supply vulnerability in a dynamic and multidimensional manner?

5.2 Methods

In light of ongoing socio-hydrologic changes, we answer our research question in two parts. We first evaluate gaps in existing data for evaluating water supply vulnerability and then use this information to develop a generalizable approach for measuring vulnerability using indicators in a multidimensional and dynamic manner. We outline our process for achieving these objectives in Figure 1 below.

We first develop an open-source database of all existing data for evaluating water supply vulnerability from global water supply literature (Section 5.2.1). Using these data, we construct a conceptual model of common elements for indicator-based assessments of water supply vulnerability. We do this to ensure that our approach captures core elements for water supply vulnerability assessment (Section 5.2.2). Next, we link these core elements together in an approach that can be implemented by managers and policy-makers to assess vulnerability using indicators. We then perform a number of analyses using our data to provide users with guidance around ensuring the multidimensionality of assessments developed using our approach (Section 5.2.3). Finally, we show how this approach, when paired with our open-source database, can be used to update assessments of vulnerability in response to social, political, and environmental conditions.

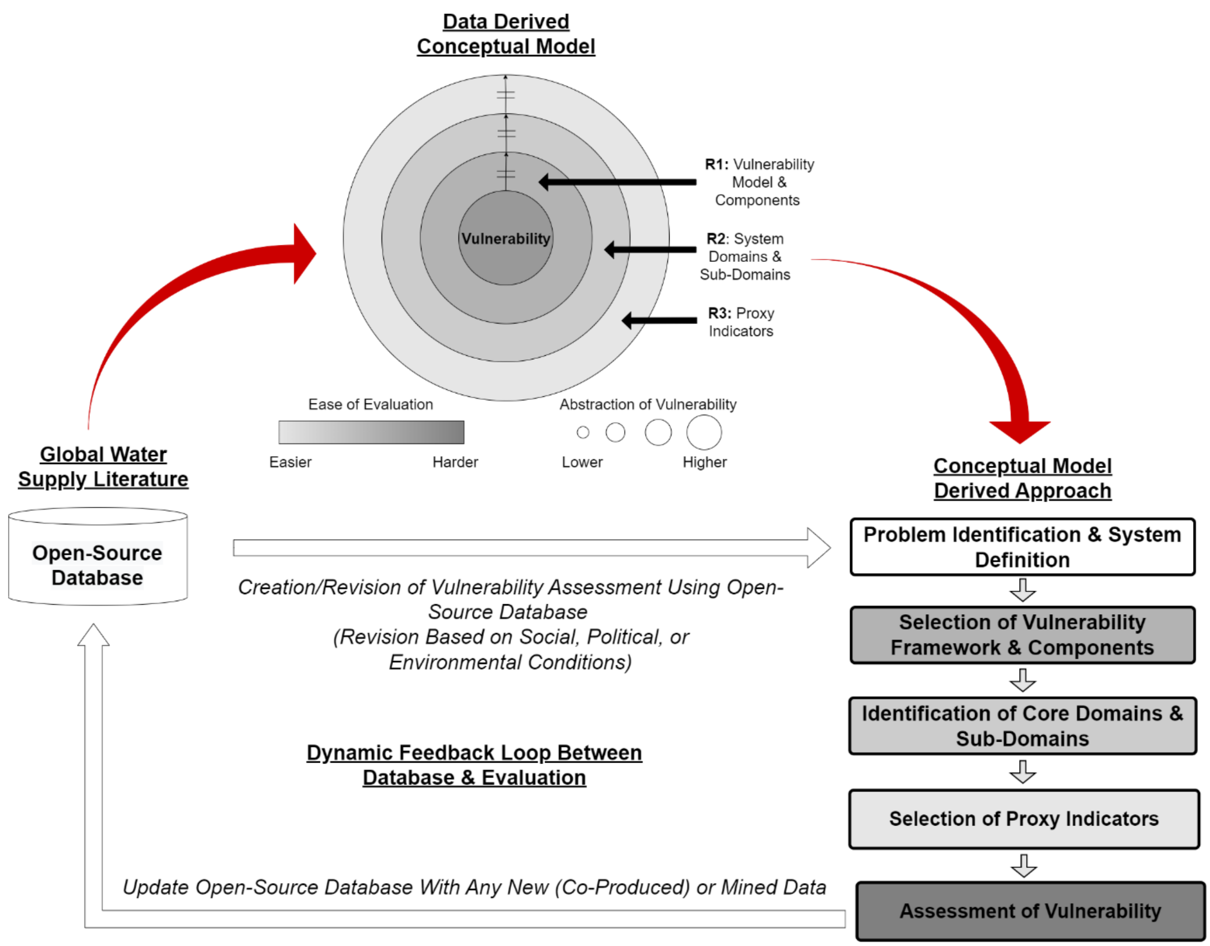

Figure 5-1: Outline of the process for developing our conceptual model of core elements for index-based vulnerability assessment and translation into a dynamic approach. Red arrows indicate the contributions of this paper that do not need to be revisited on an ongoing basis. White arrows demarcate the feedback loop between our living, open-source database and our dynamic approach.

5.2.1 Open-Source Database

The first step in our approach was a systematic review of global literature using a Google Scholar search for the terms “water supply vulnerability index” and “water supply vulnerability indicator." From this global review, we drew on a smaller number of indices that met the following criteria:

1. explicitly considered water;
2. was not a new application of an existing index (although modified indices were included); and,
3. included a list of indicators either in the main text or in the supplemental information.

We then constructed an open-source database (Gordon *et al,* 2021) comprised of 20 global indices that relied on a combined 504 indicators to quantify water supply vulnerability (henceforth, data). We refer to individual data components of the database (e.g., indicators) where appropriate throughout this manuscript. We assumed that our data are a reasonable representation of available water supply indices and indicators based on their global nature. By making our database open-source, our goal is for it to be a living database that can be updated on an ongoing basis.

5.2.2 Conceptual Model

To ensure that our approach is generalizable and transparent, we used our data to identify common elements and their linkages for global indicator-based water supply vulnerability assessments. Our model is comprised of a target—the system's true vulnerability—and three concentric rings (M = 3, labeled R1-R3 in Figure 5-1). We used a circular network model, adapted from work on wireless cluster networks by Tandon (2012), to capture the interdependence of model aspects and illustrate tradeoffs associated with the level of abstraction and difficulty of measurement. These core elements are as follows:

R1. The first ring outside of the target is comprised of discrete components of vulnerability derived from a theoretical framework that characterizes (or defines) vulnerability. A diversity of vulnerability frameworks exist, spanning various disciplinary views (e.g., Adger and Kelly 1999, Luers *et al* 2003, Cardona 2011, Turner *et al* 2003) from which users can select. R1 requires the lowest level of abstraction of the target (vulnerability) but is also the most challenging to directly measure (Luers, 2005). However, a vulnerability framework provides a theoretical grounding for the next step in the conceptual model: selecting system domains and sub-domains.

R2. Based on a selected vulnerability framework, users can organize complex systems into components of vulnerability. This requires conceptualization and prioritization of the most relevant aspects of the system for assessment, which we define as domains consistent with Füssel (2007). Domains are a related collection of interlinked factors that determine how the system responds to and performs under various stressors (the third concentric circle in Figure 1). The determination of important domains and sub-domains for a target system drives the next step in the model: the selection of indicators. R2 requires a moderate level of abstraction of the target (vulnerability) and is challenging to directly measure without the incorporation of indicators (Sullivan, 2011).

R3. The outermost circle represents indicators, which are used to measure the performance or functionality of a domain or subdomain (R2) and thus to assess the characteristics of vulnerability (R1) and ultimately the target (vulnerability). While a significant number of indicators for assessing water supply vulnerability exist (Plummer *et al*., 2012), there are limited tools to support users in selecting and assessing these indicators across diverse systems. Moreover, such indicators must often be standardized and aggregated in order to evaluate R2. As such, indicators are both the most easily measured and require the highest abstraction of vulnerability.

5.2.3 Dynamic Approach

We then linked the core elements for water supply vulnerability assessments together in an approach for evaluating water supply vulnerability (Figure 1). We performed a number of analyses using our data to ensure that assessments based on our approach are: 1) practical and 2) consider multiple dimensions of vulnerability. We describe these analyses for each step in our approach below.

5.2.3.1 Vulnerability framework analysis

Frameworks are essential for evaluating the structure of vulnerability in order to characterize potential damage to people and the environment (Luers, 2005). We first analyzed common vulnerability frameworks and their frequency of use based on our data per Figure 1. We then used extant theory to outline potential advantages and disadvantages for users interested in applying our approach.

5.2.3.2 Core domains and sub-domains analysis

To ensure that assessments based on our approach consider multiple aspects of vulnerability, we identified and defined general domains and sub-domains for indicator-based water supply vulnerability assessments based on categories and concepts from literature (see Supplemental Codebook). These domains and sub-domains are defined as follows (see Table 5-1 for more detail):

1. **Factors Influencing Water Supply (FIWS):** This includes any aspects of the system related to the availability or accessibility of water resources.
2. **Factors Influencing Water Demand (FIWD):** This domain includes any aspects of the system associated with productivity or demand for water resources**.**
3. **Factors Influencing the Social Value of Water (FISVW):** This domain includes any aspects of the system associated with the anthropogenic valuation and allocation of water resources.

Table 5- 1: Summary and definitions for domains and sub-domains.

Sub-Domain	**Description**
	FIWS

Water Source	• Relates to the origin of water supplies
	• Includes measures of the volume of water obtained from:
	  o Conventional (e.g., surface and groundwater)
	  o Non-conventional water sources (e.g., recycled water or treated wastewater)

Water Quality	• Relates to characteristics of water that are required to adequately meet various sectoral end uses
	• Includes measures of:
	  o Water quality associated with standards or requirements for a given use (e.g., drinking water)
	  o Factors contributing to water quality (e.g., discharge)
	  o Water treatment and sanitation

Water Infrastructure and Distribution	• Relates to the transportation of water from a source or treatment location to storage or the end use or users
	• Includes measures of:
	  o Physical aspects of the distribution system (e.g., canals, reservoirs)
	  o Distance and time to access water supply

Physical Environment	• Relates to the condition of the broader physical environment that can directly and/or indirectly impact water supply
	• Includes measures of:
	  o The health and functionality of the environment
	  o Perceptions of change
	  o Measures without direct attribution to water quality or quantity

<table>
<tr><th colspan="2">FIWD</th></tr>
<tr><td>Industrial Land and Water Use</td><td>Relates to any measures associated with fabricating, processing, washing, diluting, or transportation to assist smelting, refineries, and industries producing products
Includes measures of:
Land use and land cover</td></tr>
<tr><td>Urban and Municipal Land and Water Use</td><td>Relates to any human consumption or domestic uses such as bathing, cooking, cleaning, or watering lawns or gardens within a specific region (e.g., property owners in a specific place) or within the urban environment (e.g., region surrounding a city or developed area)
Includes measures of:
Land use and land cover</td></tr>
<tr><td>Agricultural Land and Water Use</td><td>Relates to any produce or crop production (including the use of fertilizers and pesticides) or livestock rearing
Includes measures of:
Land use and land cover
Food security</td></tr>
<tr><td>Cultural and Environmental Land and Water Use</td><td>Relates to the functionality of physical systems to support spiritual and/or religious practices as well as the health and well-being of flora and fauna
Includes measures of:
Land use and land cover
Subsistence fishing
Spiritual and/or cultural value</td></tr>
<tr><td>General Land and Water Use</td><td>Relates to any non-sectoral or non-specific measures of water or land use
Includes measures of:
Land and water use without attribution to other demand sub-domains (e.g., general groundwater withdrawal)</td></tr>
<tr><th colspan="2">FISVW</th></tr>
<tr><td>Institutions and Management</td><td>Relates to the laws, policies, and customs that govern how water is allocated, distributed, and managed at a variety of scales
Includes measures of:
Operational (e.g., local use or control) water governance</td></tr>
</table>

	○ Organizational (e.g., coordination or reduction of conflict between competing uses via administration of rules) water governance ○ Constitutional (e.g., laws, policies, and legislation) water governance
Socio-Culture	• Relates to the social and cultural elements of how a group of individuals interact around water • Includes measures of: ○ Societal structure (e.g., demographics, education, age, land ownership) ○ Culture (e.g., beliefs practices values and norms) ○ Adaptivity (e.g., innovation, access to information, and capacity for adaptation and change)
Economics	• Relates to the production, distribution, and/or consumption of goods as well as services within a society • Includes measures of: ○ Material assets ○ Monetary assistance and capacity ○ Economic diversity and dependence (sometimes measured via employment)

We then validate our proposed core domains and sub-domains using our data as described in Steps 1 and 2 below.

Step 1: Iterative consensus building—Multiple researchers hand coded all indicators included in our database into domains and sub-domains based on proposed definitions in Table 1 to establish intercoder reliability and repeatability. Any disagreements among coders were recorded and discussed until consensus was achieved. The definitions of each domain and sub-domain were iteratively refined during this process to enhance reproducibility (see Supplemental Codebook).

Step 2: Assessing robustness—We next implemented a text analysis of the hand-coded indicators obtained from Step 1 to verify the robustness of proposed definitions for domains and sub-domains. First, we identified a list of stop words to be excluded from indicators and standardized punctuation, capitalization, established synonyms, and abbreviations (Feinerer, 2022). We also removed duplicate words from a single indicator (e.g., if 'precipitation' was mentioned twice for a single indicator, we recorded 'precipitation' only once). We then counted the frequency of each word in each domain and sub-domain and compared these results to definitions proposed in Table 1 to validate consistency between hand coded indicators and proposed definitions.

5.2.3.3 Indicator analysis

Previous work has underscored the importance of selecting indicators of water supply vulnerability that are: 1) appropriate and relevant; 2) transparent (i.e., not complex), 3) feasible to using available data, 4) match system considerations (i.e., spatial and temporal scale), and 5) are consistent with the level of detail required for the desired assessment (Anandhi and Kannan, 2018; de Ruiter and van Loon, 2022; Hurd *et al*., 1999). In order to screen indicators in this way and select indicators that capture vulnerability in a multidimensional manner, users need publicly available and rigorous information about the breadth and focus of existing indicators. Such information is also assumed to enhance efficiency in the development of new indicators by enabling a bigger picture view of where new indicators may be needed and how indicators from different domains can be quantitatively evaluated.

To do this, we first calculated similarity for indicators assigned to the same sub-domain (Section 5.3.2) based on the Jaro-Winkler Fuzzy match algorithm. This algorithm has been shown to perform well on relatively short strings (Leonardo and Hansun, 2017) consistent with the majority of our indicators. We obtained the Jaro similarity (sim_j as proposed by Jaro, 1989) for all pairwise combinations of unique strings (e.g., s_1 and s_2) as:

$$sim_j = \begin{cases} 0 & if\ m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & otherwise \end{cases} \quad \text{Eq. (5-1)}$$

Where m was the number of matching characters between strings s_1 and s_2, defined as:

$$m = \left\lfloor \frac{max(|s_1|,|s_1|)}{2} \right\rfloor - 1 \quad \text{Eq. (5-2)}$$

And t is the number of transpositions, which was obtained by dividing the number of matching characters in the wrong sequence order by two. We then calculated the Jaro-Winkler similarity or sim_w using sim_j from above with a modification that assigned favorable ranking to strings with matching prefixes up to a set length (L) per Eq. (5-3):

$$sim_w = sim_j + Lp(1 - sim_j) \quad \text{Eq. (5-3)}$$

Where p was a scaling factor used to the score for common prefixes. The standard value of p is 0.1, which was also adopted in this study. The Jaro-Winkler distance (d_w) was then determined using Eq. (5-4):

$$d_w = 1 - sim_w \quad \text{Eq. (5-4)}$$

As suggested by Eq. (5-4), lower values of d_w correspond to greater similarity between strings s_1 and s_2 whereas higher values of d_w correspond to less similarity between strings s_1 and s_2.

We then examined clusters within our data using a matrix based on the Jaro-Winkler Fuzzy match algorithm. We used a complete linkage hierarchical cluster (Amine *et al*., 2010), which forms clusters based on the maximum distance between two clusters X and Y:

$$D(X,Y) = \max_{x \in X, y \in Y} d(x,y) \qquad \text{Eq. (5-5)}$$

Where $d(x,y)$ is the distance between elements $x \in X$ and $y \in Y$.

Complete linkage clustering produces well-separated and compact clusters and has been employed using text, string, and record data (Mamun *et al*., 2016; Rajalingam and Ranjini, 2011; Ram *et al*., 2005) in larger datasets (Saraçli *et al*., 2013). However, there is limited guidance on how to determine the optimal number of clusters for our analysis (Orford, 1976). In the absence of a clear set of best practices, we implemented a tiered methodology to determine the optimal number of clusters for each sub-domain to avoid subjectively determining cutoff points. First, we determined whether there was any consensus (defined as >50% agreement) between multiple estimation methods described in more detail by Lüdecke *et al* (2020). In the absence of consensus (<50% agreement), we then determined the optimal number of clusters by visually comparing a Silhouette score plot (Rousseeuw, 1987) and an elbow method plot. Once the optimal number of clusters was determined, we followed Tseng and Tsay (2013) to obtain cluster labels analysis using maximally repeated words or word sequences within indicators. Two researchers reviewed and agreed upon all resulting labels.

5.2.3.4 Indicator evaluation

We next investigated how indices aggregate indicators from different domains and sub-domains. We noted whether there was guidance around weighting, which is an integral and variable part of vulnerability assessments (Hinkel, 2011).

5.2.3.5 Catalyst for revision analysis

Conceptual work on vulnerability has highlighted the need to move towards a complex system theory approach to account for the dynamic nature of the vulnerability (Birkmann *et al*., 2013; Cardona, 2011; Pelling, 2010). However, there is some tension between considering numerous complex and nonlinear processes and the practical benefits of indicator-based assessment for managers and policy-makers. We thus elected to focus specifically on whether and how indices identify a socio-hydrologic catalyst for revising, updating, or expanding the indicators, domains, or framework selected for assessing water supply vulnerability (see Figure 5-1 for feedback between our database and approach).

5.3 Results

5.3.1 Vulnerability framework results

Indices included in database (n = 20) relied upon a diversity of vulnerability frameworks with various associated components for assessment (Table 5-2). Both the Exposure-Sensitivity-Adaptive Capacity (ESAC) and the integrated and social ecology frameworks were the most widely adopted (n = 6, 30% each) with slightly fewer indices (n = 4, 20%)

adopting a Pressure-State-Response based framework. Two indices made use of combined frameworks and two indices did not clearly indicate a framework.

Table 5-2: Vulnerability frameworks and components. Columns include the framework name, definition, advantages, and disadvantages as well as the frequency of use in our data.

Model	Definition	Advantages	Disadvantages	↓ Frequency of Use
Exposure-Sensitivity-Adaptive Capacity	Vulnerability is a function of exposure, sensitivity, and adaptive capacity.	Model integrates adaptive capacity, is widely used, and has clear guidance around implementation.	Model requires subjective categorization into exposure, sensitivity, and adaptive capacity.	30%
Integrated and social ecology	Vulnerability is an interactive coupling between a system's social and environmental subsystems.	Model emphasizes coupling of human and environmental systems (often place-based).	Model can neglect interaction between socioeconomic and biophysical stressors (e.g., double exposure).	30%
Pressure-State-Response	Vulnerability arises from pressures that lead to astate with responses alleviating pressures and improving the state.	Model is flexible and has clearer guidance around implementation.	Model can neglect coupling between social and environmental systems and underemphasize feedback.	20%
Risk-Hazard	Vulnerability arises from exposure to hazards of a particular type and magnitude.	Model is widely used by engineers and economists to refer to a physical vulnerability.	Model can exclude society's ability to modify hazards.	15%
Holistic	Vulnerability is comprehensive, accounting for causal factors as well as different thematic dimensions.	Model allows for a dynamic conceptualization of vulnerability.	Model can lack clear guidance around implementation.	5%

Number of Indices

0 1 2 3 4 5 6

Overall, the results validated the importance of established frameworks for assessing water supply vulnerability and also suggest that a diversity of frameworks can be robustly incorporated into indicator-based assessments. However, the analysis revealed a number of advantages and disadvantages that should be considered before adopting any of the frameworks outlined in Table 5-2. For example, despite its frequent use, ESAC may be challenging for users to implement given the subjective categorization of domains, sub-domains, and ultimately indicators into three categories (Fortini and Schubert, 2017).

5.3.2 Domain and sub-domain results

Hand coding analysis on our data confirmed the general applicability of our proposed multidimensional domains and sub-domains outlined in Table 5-1. At the domain level, initial hand coding results yielded agreement on 399 of 504 indicators (~79%) with follow-up discussion leading to consensus on all 504 indicators. At the sub-domain level, initial hand coding results were similar with agreement on 403 of 504 indicators (~80%). Here too, follow-up discussion led to consensus on all 504 indicators and clarification were recorded in the definitions outlined above (see Supplemental Codebook).

Text analysis further supported the robustness of proposed domains and sub-domains for a broad range of systems. For example, text analysis on the hand coded results for the water infrastructure and distribution sub-domain revealed that existing indicators were heavily focused on storage and reservoirs (Figure 5-2A) consistent with Table 5-1. Although included in our definitions, few indicators grouped into the FIWS domain emphasized the water infrastructure and distribution sub-domain (20 of 217 indicators, 9.2%), which we explore in more detail in Section 5.4.3. Results were similarly robust for the FIWD domain. Figure 5-2B, for example, highlights the importance of arable land (cover), food, and irrigation consistent with the proposed definitions (Table 5-1). Here the comparison between our proposed definitions and results in Figure 5-2 highlighted a diminished emphasis on Table 5-1 terms associated with urban and municipal water and land use sub-domain (10 of 105 indicators, 9.5%). We found that the institutions and management sub-domain emphasized indicators such as conflict, government, and public and civil measures consistent with the proposed definitions in Table 5-1.

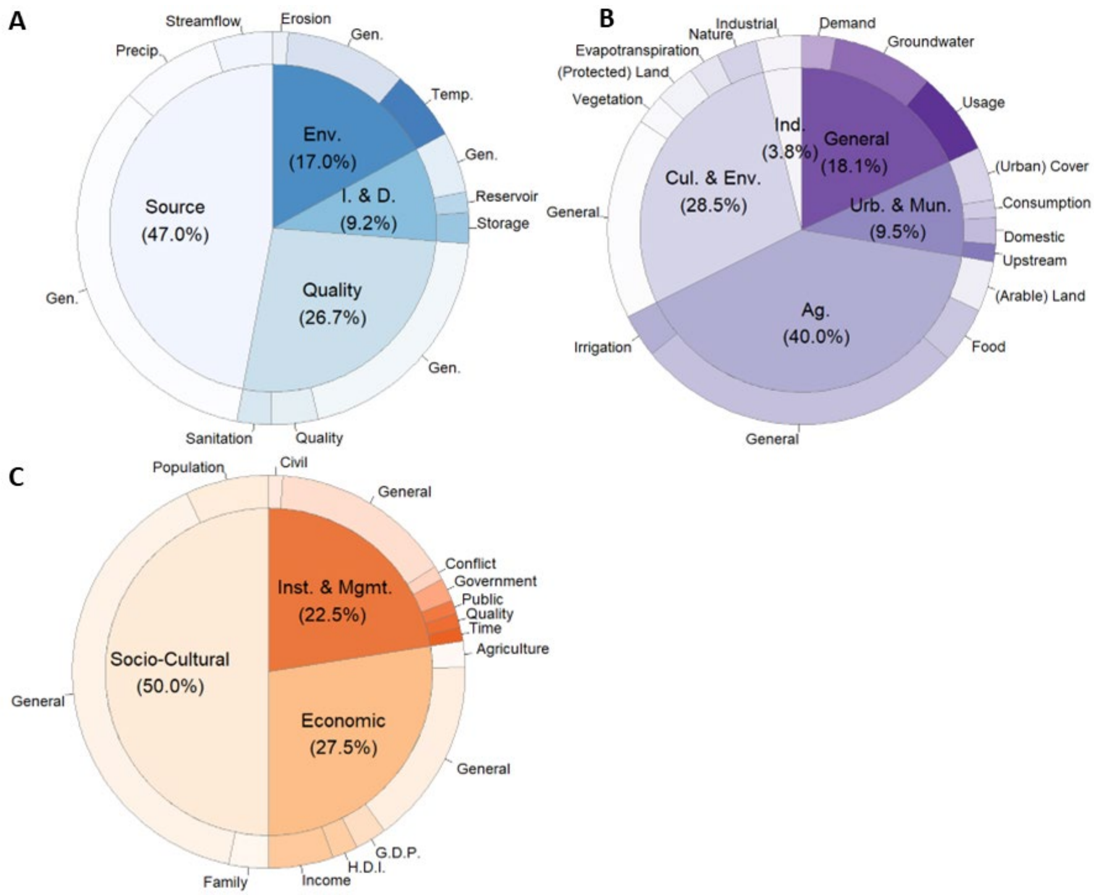

Figure 5-2: Distribution of existing indicators (n = 504) across proposed sub-domains based on hand coding including a robustness check based on text analysis of word frequency for each of the three domains: A) FIWS (n = 217); B) FIWD (n = 105); and C) FISVW (n = 182).

Overall, hand coding and text analysis validated our proposed domains and sub-domains outlined in Table 5-1 for the broad range of global systems. Importantly, we also observed that the indicators included database—which we assume are a reasonable representation of available water supply indicators more generally—prioritized certain domains over others. Evidence of this bias suggests that existing indices may be prone to silo-ing particularly with regard to water supply. We also observed that there were a number of general indicators for each of our domains, which did not fit into any particular category.

5.3.3 Indicators results

Our analysis of indicators included in our database revealed reinforced the biases towards physical indicators observed in Section 5.4.2 (see also Figure 5-2) can help users quickly identify existing indicators and prioritize areas for development of additional indicators. For the FIWS domain and associated sub-domains (Figure 5-3A to 5-D), cluster results emphasized a strong focus on surface water for water source sub-domain indicators (Figure 5-3A) with clusters supporting the evaluation of precipitation, unregulated flows, surface water stress, surface water source, and streamflow. Indicators in our database—represented by individual lines in each dendrogram—tended to be similar based on the height at which distinct clusters (represented by different colors in Figure 5-3A to 5-D) emerged. See, for example, indicators focused on the evaluation of surface water sources (see the height at which individual indicators were observed to diverge in the 'Water Source cluster in Figure 5-3A).

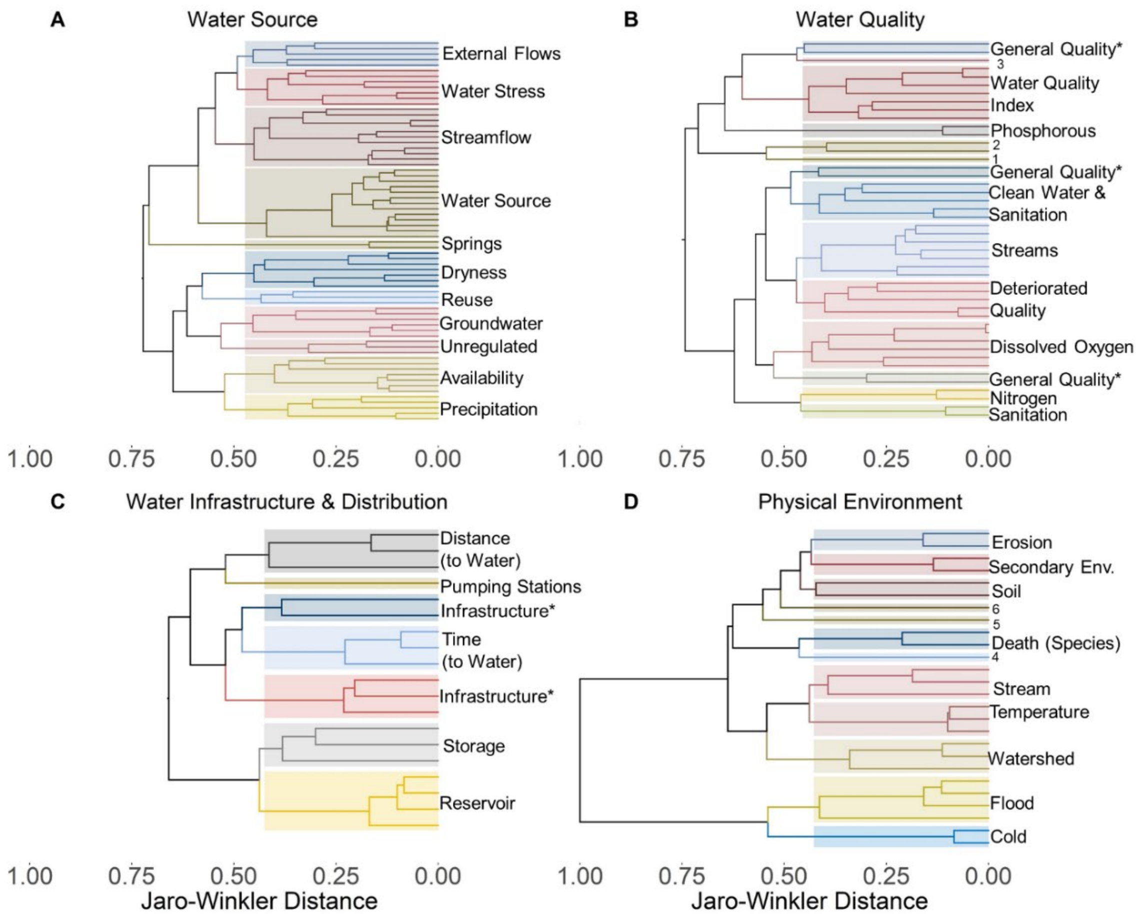

Figure 5-3: Results of the complete linkage agglomerative hierarchical clustering and labeling analysis presented in dendrograms for all four sub-domains in the factors influencing the water supply domain: A) the water source sub-domain; B) the water quality sub-domain; C) the water infrastructure and distribution sub-domain; and D) the physical environment sub-domain. * indicates an inconclusive or repeated cluster label where best judgement was used to generate a unique cluster name, [1] Algal Bloom, [2] Lake Clarity, [3] Zebra Mussel, [4] Glacial Lake Outburst, [5] Humidity Index, and [6] Transmissivity. Plots with indicator label names are included in SI (Figure S5-1 to S5-4).

For the FIWD domain and associated sub-domains (Figure 5-4A to 5-4C), we found a smaller number of distinct clusters with differences tending to emerge well below a Jaro-Winkler distance of 1 (i.e., complete dissimilarity). Clusters in the agricultural land and water use domain highlighted a broad existing focus on food consumption, irrigation and cropping, livestock, and food scarcity supported by multiple moderately dissimilar indicators (see corresponding labels in Figure 5-4A). In particular, we found that a large number of unique indicators were associated with irrigation and crops, as evidenced by the divergence of individual lines within this cluster at Jaro-Winkler distances between 0.4 and 0.5 in the majority of cases (see 'Irrigation/Crops' label in Figure 5-4A).

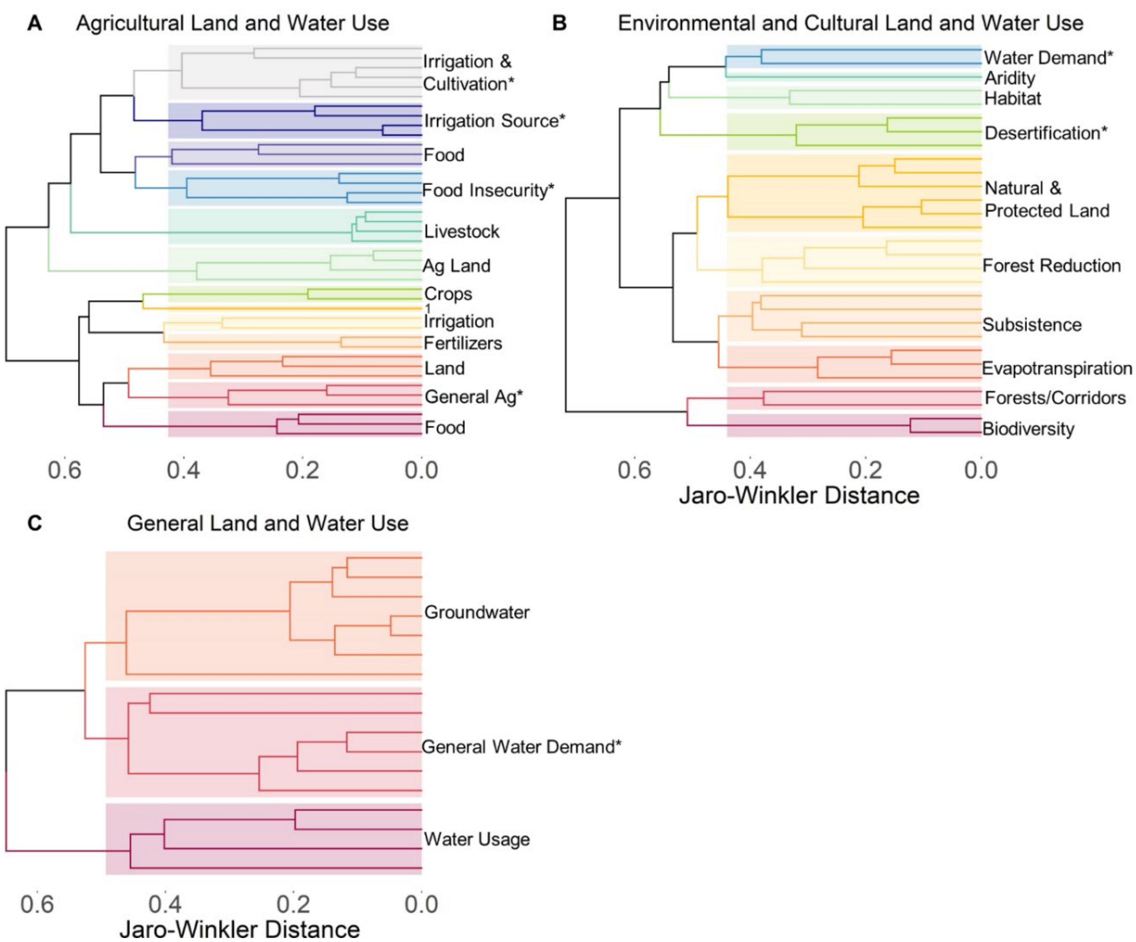

Figure 5-4: Results of the complete linkage agglomerative hierarchical clustering and labeling analysis presented in dendrograms for three of the five sub-domains in the factors influencing water demand domain: A) the agricultural land and water use sub-domain; B) the environmental and cultural land and water use sub-domain; C) the general land and water use sub-domain. Cluster optimization results for the urban and municipal land and water use sub-domain and the industrial land and water use sub-domain were inconclusive. As a result, both sub-domains were excluded from further analysis. * indicates an inconclusive or repeated cluster label where best judgement was used to generate a unique cluster name, [1] Growing Season. Plots with indicator label names are included in SI (Figure S5-5 to S5-7).

Indicators for evaluating FISVW domain (Figure 5-5A to 5-5C) were more complex than either the FIWS (Figure 5-3) or FIWD (Figure 5-4) domains as evidenced by a high number of optimal clusters. Specifically, we found that the socio-culture (Figure 5-5B) and economic sub-domains (Figure 5-5C) tended to be both dissimilar and complex (see height at which distinct clusters emerge and labels in Figure 5-5B and 5-5C). There were, however, exceptions with results for the socio-culture sub-domain (Figure 5-5B) where there was a strong existing emphasis on relatively similar population-level indicators (see ‘Population’ label in Figure 5-5B).

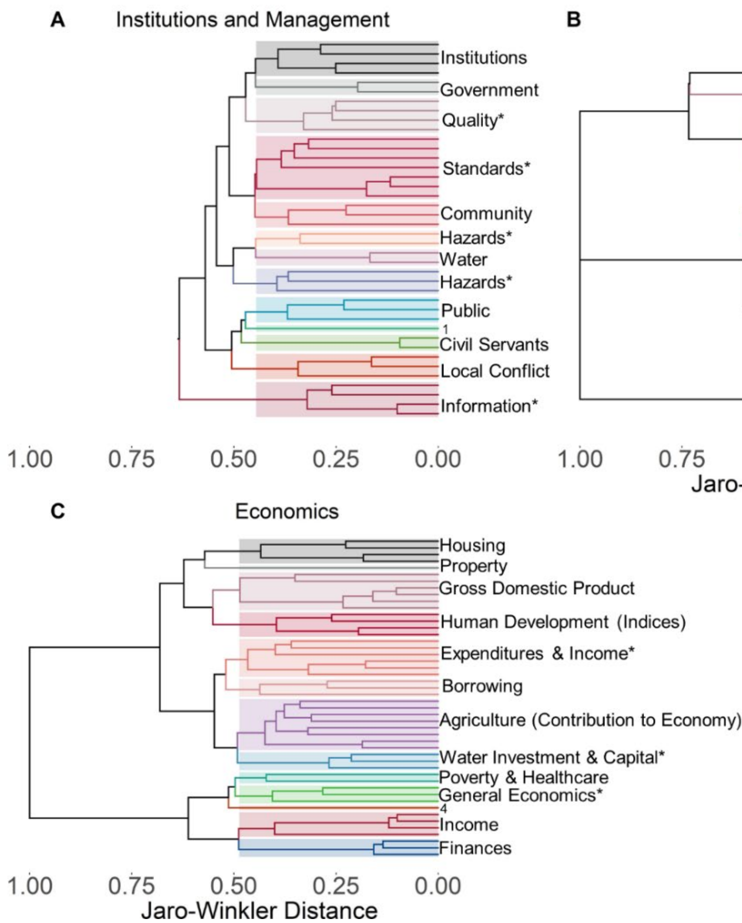

Figure 5-5: Results of the complete linkage agglomerative hierarchical clustering and labeling analysis presented in dendrograms for all three clusters in the factors influencing the social value of water domain: A) the institutional and management sub-domain; B) the socio-culture sub-domain; C) the economics sub-domain. * indicates an inconclusive or repeated cluster label where best judgement was used to generate a unique cluster name, [1] Association Membership, [2] Life Expectancy, [3] Orphans, [4] Vehicles. Plots with indicator label names are included in SI (Figure S5-8 to S5-10).

Overall, we found that existing indicators—as represented by our data—are well equipped to capture surface water supply, agricultural water use, and population-level socioeconomic aspects of water supply vulnerability. However, many important gaps exist and are likely related to data availability. Few indicators in the FIWS domain focus on groundwater, which likely corresponds with the lack of high resolution global groundwater datasets available for smaller-scale assessments (Tapley *et al*., 2004). Likewise, we found a particularly acute lack of indicators associated with the measurement of reused or recycled water resources (Figure 5-3A). Like groundwater, reused and recycled water is likely to become increasingly critical in many arid and semi-arid regions (He *et al*., 2021; Toze, 2006) and is plagued by data limitations (Wiener *et al*., 2016). Within this domain, our analysis also highlighted the narrow focus of water infrastructure and distribution indicators on built storage (Figure 5-4C), which may be easier to evaluate but is also likely to be increasingly stressed and controversial under continued climate change (Ehsani *et al*., 2017; Kellner, 2021; Kellner and Brunner, 2021; Steyaert *et al*., 2022). Prioritization of other aspects of water infrastructure and distribution (e.g., groundwater recharge) within these systems could enhance understanding of not only vulnerability but also where opportunities for adaptation may exist (He *et al*., 2021). However, the feasibility of these efforts is closely tied to the availability and/or development of sufficient datasets. We also observed a particularly weak and narrow emphasis on cultural water needs (Figure 5-5B), which are increasingly recognized as important (Immerzeel *et al*., 2020) and threatened by climate change (Vuille *et al*., 2018). Reconciling this gap is likely to require improvements in data (Smith and Ali, 2006) as well as conscientious engagement with communities in order to understand the spectrum of cultural uses for different groups of people (e.g., Chief *et al*., 2016). Indicators for evaluating the FISVW domain prioritized population-level measures, which we hypothesize is an artifact of the type of data available for evaluating this domain (e.g., national or regional scale census data). Here too, we observed a distinct lack of indicators associated with the influence of values and cultural norms on vulnerability.

5.3.3.1 Indicator assessment results

We summarize our findings with regard to indicator standardization and aggregation in Figure 5-6 and Table 5-3, noting the advantages and disadvantages associated with the different methods based on extant theory. Most indices included in our database normalized indicators based on minimum and maximum values (Min-Max) and then aggregated these indicators based on Composite Index Approach (CIA) with equal weighting (dark red stream in Figure 5-6, details in Table 5-3). However, we found that minimum-maximum based standardizing presents problems when the vulnerability is not uniformly high or low based on the raw indicator values (e.g., 10 could indicate either low or high vulnerability based on different indicators). Rating scale-based approaches may circumvent this challenge but require expert knowledge of the system for realistic values. We present the most common combinations of methods for standardizing, evaluating, and weighting indicators based on our database per Figure 5-6.

Table 5-3: Summary of standardization and aggregation methodologies for evaluating diverse indicators of water supply vulnerability.

Description	Evaluation	Advantages	Disadvantages
Standardization			
Rating Scale	Raw values are grouped and then re-assigned a value from 0-1, where 0	• Simple way to assign different	• Requires expert input and/or a strong

	is highly vulnerable and 1 is highly resilient.	indicators values on a 0-1 scale. • Standardizes low values as vulnerable and high values as resilient.	understanding of system thresholds.
Min-Max Normalization	$I = \frac{(\mathrm{I}_o - \mathrm{I}_{min})}{(\mathrm{I}_{max} - \mathrm{I}_{min})} * C$ Where I is the re-scaled indicator, I_o is the initial indicator, I_{max} is the upper lower bound of the original scale, and I_{min} is the upper bound of the original scale. In some cases, users may multiply the normalized indicator by a scalar (C) in order to obtain values within a desired range (e.g., 0-100). Indicators can also be normalized as: $I = \frac{(\mathrm{I}_{max} - \mathrm{I}_o)}{(\mathrm{I}_{max} - \mathrm{I}_{min})}$	• Can include the option to re-scale indicators, which may be useful for further standardization. • Substantial guidance exists, making it easier for users to implement.	• Requires subjective input to transform (via the inverse of I_o) in some cases.
Basket Approach	$D = n^{-1} \sum_{i}^{n} \mathrm{I}_{o\,i}$ Where D is the domain value, I_o is the raw indicator, n is the number of indicators, and i is the i-th value of out of n.	• Circumvents challenges associated with standardizing individual indicators.	• May not consider that raw indicators can reflect vulnerability at both high and low values.
Threshold Normalization	Indicators are manipulated to ensure that all high values are associated with vulnerability and then standardized relative to a threshold.	• Vulnerability is consistently associated with high values.	• Requires subjective definition of a threshold and lacks rigorous mathematical guidance.

Aggregation			
Multi-Criteria Analysis (MCA)	$S = w_s \sum_{i}^{n} r_i I_i$ Where S is the domain, n is the number of indicators, i is the i-th value of out of n, r is the risk of that indicator increasing vulnerability based on either statistical analysis or expert opinion, and I is the raw indicator. In the case that there are multiple domains, S is multiplied by w_s, which is obtained by dividing 1 by the number of domains if all domains are equally weighted.	• Theoretical basis is well established in multiple disciplines, including natural resource management.	• Requires either statistical or expert-based knowledge of risk.
Composite Programming Approach (CPA)	$SD = \left[\sum_{i=1}^{l} w_i I_i^P\right]^P$ Where SD is a given sub-domain, l is the number of indicators grouped into the given SD, i is the i-th value of out of l indicators, w is the weight assigned to each normalized indicator (I), and P is a balancing factor among indicators selected to reflect the importance of maximal deviation. SD can then be used to evaluate a domain (D) as: $D = \left[\sum_{j}^{m} w_j SD_j^{P_j}\right]^{\frac{1}{P_j}}$	• Similar to MCA • Does not require a risk assessment based on statistical evaluation or expert opinion.	• Requires subjective determination of weighting factor and balancing factor.

	Where m is the number of sub-domains grouped into the given D and j is j-th value out of m sub-domains. A composite value (V) can then be obtained as: $$V = 1 - \left[\sum_{k}^{n} w_k D_k^n\right]^{\frac{1}{n}}$$ Where n is the number of domains included in the vulnerability assessment and k is the k-th value out of n domains.		
Composite Index Approach (CIA)	$$V = \sum_{i}^{n} \frac{w_i G_i}{w_i}$$ Where *w* is the weight assigned to each group (e.g., sub-domain, domain) and V is the composite value of vulnerability.	• Similar to MCA • Explicitly allows for unequal weighting of domains.	• May not consider the ways in which raw indicators can reflect vulnerability at both high and low values

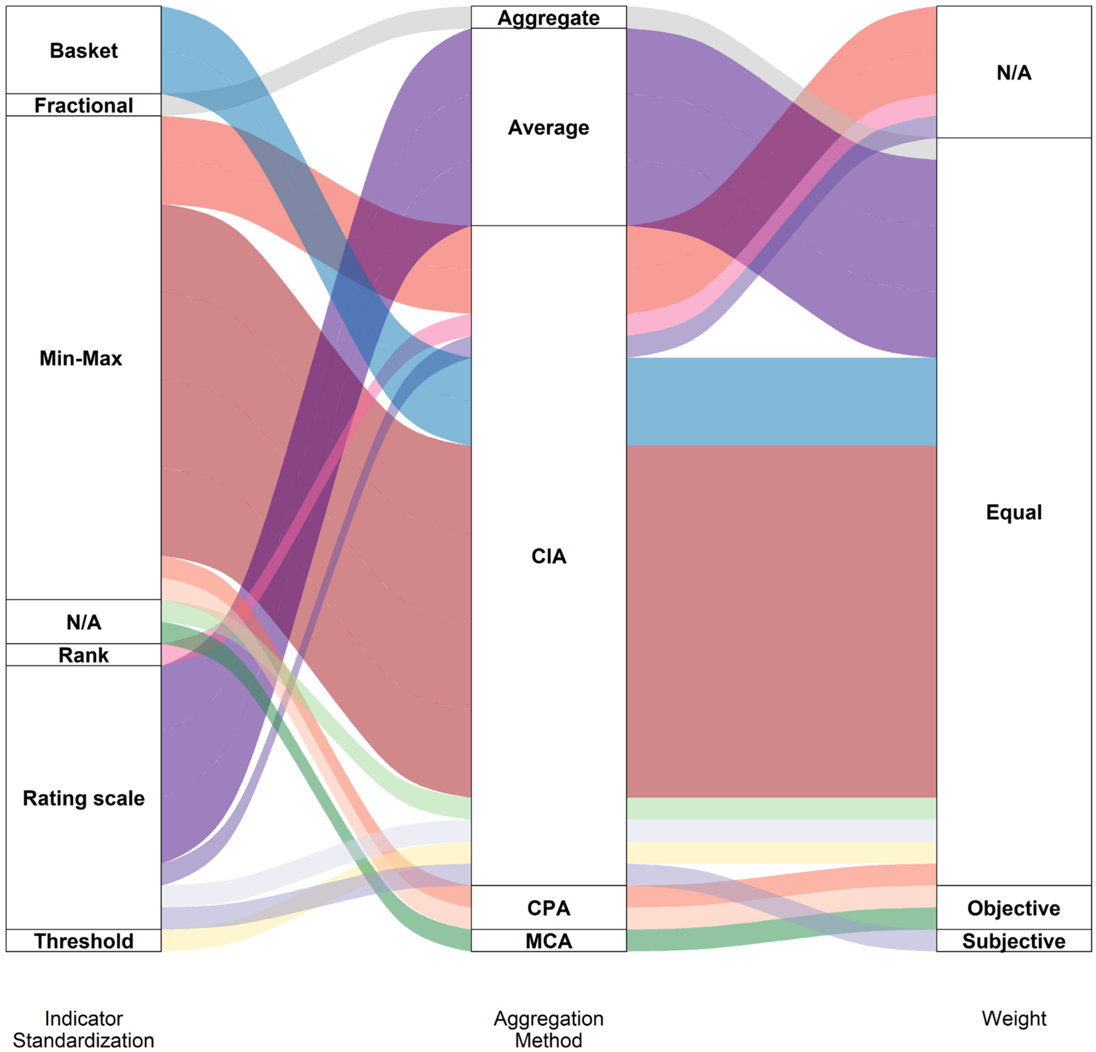

Figure 5- 6: Alluvial diagram of results for indicator standardization, sub-domain and domain aggregation, and weighting based on analysis of existing indices collected per Section 5.3.1.

We observed multiple pathways for aggregating across social, economic, physical, cultural, environmental, and institutional indicators of system performance (Figure 5-6, Table 5-3). This suggests that if biases toward physical indicators can be corrected, there is a clear pathway for assessing vulnerability in a multidimensional manner.

5.3.4 Catalyst for revision results

Critically, we observed that no indices included in our database explicitly noted when or how they should be revised to capture the underlying dynamics of vulnerability, long-lasting vulnerability, and/or compounding vulnerability (see Table S5-1). These findings underscore a pressing need to for future indices and approaches to consider the conditions under which users should revise the indicators—or in more extreme cases, domains and sub-domains or vulnerability frameworks—adopted for vulnerability assessment.

5.3.5 Approach and Database

When integrated into the approach derived from our conceptual model in Figure 5-1, results provide practical guidance for water managers and policy-makers interested in implementing a bottom-up assessment of water supply vulnerability as shown in Figure 5-7. The analyses and results described in the sections above help ensure that assessments of vulnerability are multidimensional and when paired with our database, can be revisited in response to social, political, and environmental stresses per feedbacks in Figure 5-1 as follows:

1. Per previous work by Anandhi and Kannan (2018), the target system is defined based on its: 1) spatial and social bounds; 2) the level of detail required to address vulnerability (e.g., rapid, intermediate, or comprehensive); and, 3) the data and/or resources available (e.g., measurements, models, national statistics, stakeholder interviews, etc);
2. A vulnerability framework (R1, Figure 5-1) is selected drawing upon the results presented in Section 5.3.1, specifically Table5-2;

3. Core domains and sub-domains (corresponding with R2 in Figures 5-1) are identified based on Section 5.3.2, Figure 5-2, and Table 5-1;
4. Using the open-source database accompanying this manuscript, indicators available to evaluate the performance of identified domains and sub-domains (R3 in Figure 5-1) are then screened based on their relevance, transparency, feasibility, system considerations, and the level of detail required for the desired assessment (Anandhi and Kannan, 2018; Hurd *et al*., 1999). The results presented in Section 5.4.3, specifically Figures 5-3 to 5-5 can be used to identify gaps where users the co-production or collaborative development of indicators is necessary to ensure local relevance;
5. Vulnerability is then assessed by evaluating indicators using the results presented in Section 5.3.3.1, specifically Table 5-3 and Figure 5-6;
6. External (e.g., exogenous stressors imposed by the physical environment) or internal (e.g., endogenous stressors imposed by society) catalysts that would trigger a revision of indicators (grey arrow in Figure 5-7) or in more extreme cases the selection of a vulnerability framework, domains and sub-domains, and indicators (black arrow in Figure 5-7).

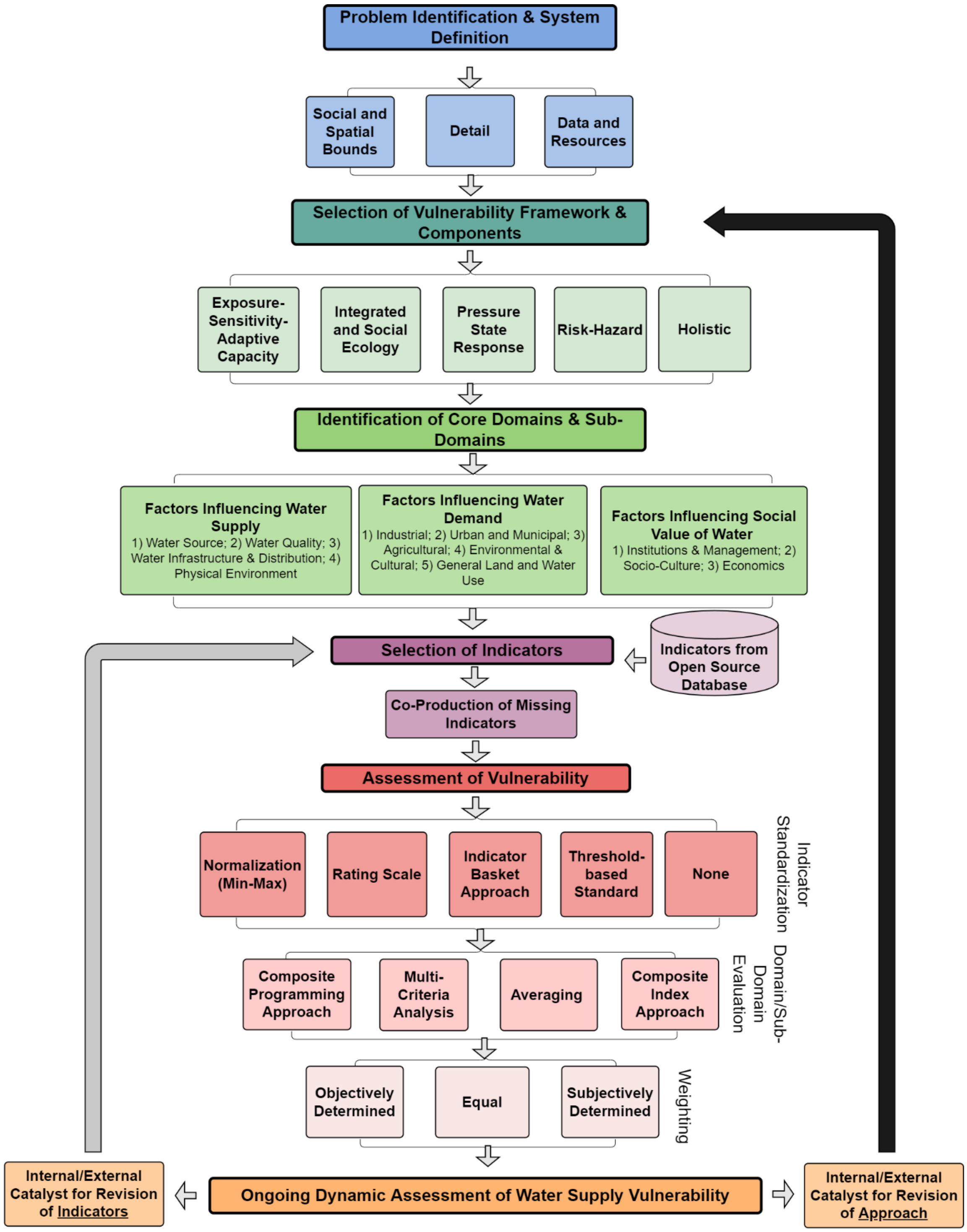

Figure 5- 7: A scalable approach to multidimensional and dynamic indicator-based vulnerability assessments for bottom-up implementation in water resource systems.

5.4 Conclusions

Complex, interacting, and accelerating socio-hydrologic stresses are straining freshwater supplies around the world, leaving water resources systems increasingly vulnerable to damage. To design just, efficient, and locally-relevant adaptation strategies, water managers and policy-makers need dynamic and multidimensional assessments of system vulnerability that can be implemented in a bottom-up manner and in a diversity of places (Dilling *et al*., 2015; Dilling and Berggren, 2015; Sullivan, 2011). Drawing from a diverse body of existing water supply indices and indicators, we distill these approaches from around the globe into a conceptual model comprised of key elements for assessing vulnerability. We then link these elements via an approach that retains the practical benefits of indicator-based assessment while ensuring that vulnerability can measured in a multidimensional and dynamic manner (Figure 5-1).

When combined with our supporting analyses and open-source database, this approach can be implemented for a variety of reasons in a diversity of systems. For example, per Figure 5-7, we used this approach to assess the water supply vulnerability of agricultural systems to changing snow (Gordon *et al, In Prep*). As part of this, we first defined the system bounds based on irrigation demand and selected the ESAC framework based on Table 5-2. We then identified the FIWS and FIWD domains, specifically water source and agricultural demand sub-domains respectively, based on available data. We then used the database to review available indicators for these sub-domains and drew from additional literature to add additional indicators appropriate for our analysis. These indicators were then added to our database. We then assessed these indicators using the min-max normalization which were aggregated using the geometric mean with equal weighting to evaluate vulnerability as a function of exposure, sensitivity, and adaptive capacity.

In this study, we also highlight several fundamental gaps in existing data for evaluating water supply vulnerability that can be confronted in future research. In this pursuit, the success of our interdisciplinary approach may be particularly helpful. For example, our text clustering analysis had to be complemented by hand coding, reinforcing the need for robust, interdisciplinary approaches to characterize the risks and opportunities for global water resource systems. Rather than focusing on the development of new static, top-down indices for relative comparisons across systems, our findings illustrate the need for ongoing research and management efforts to propose, test, and refine more diverse, locally-relevant indicators—particularly as they relate to cultural aspects of water use—in collaboration with stakeholders to ensure that outcomes are just and efficient. This need intersects with the broader challenge of comprehensive data for evaluating the social value of water. When incorporated into practical approaches, advances on both of these fronts can further assist in more comprehensive evaluations of vulnerability in order to improve the local relevance, justness, and efficacy of critical adaptation activities.

5.5 Supplemental Information

Table S5-1: Complete indices included in our test data per the main manuscript along with geographic region (if applicable) and catalyst.

Index Name	Citation	Location specific?	If, yes specify?	Includes temporal extent?	Includes catalyst?

Water Vulnerability Index	Sullivan (2011)	No		Yes	No
	Jun *et al.* (2011)	No		No	No
Arctic Water Resource Vulnerability Index	Alessa *et al.* (2008)	Yes	Communities in the circumpolar Arctic	Yes	No
The Livelihood Vulnerability Index	Hahn *et al.* (2009)	Yes	Mabote and Moma Districts of Mozambique	No	No
CWCVI	Okpara *et al.* (2016)	Yes	South-eastern shores of Lake Chad in the Republic of Chad	No	No
DART	Dennis & Dennis (2011)	Yes	South Africa	No	No
CVI	Sullivan and Meigh (2005)	No		No	No
N/A	Khajuria and Ravindranath (2012)	No		No	No
GCVI	Jubeh and Mimi (2012)	Yes	Israel, Jordan, Lebanon, Palestine and Syria	No	No
Climate Vulnerability Index for Water (CVIW)	Pandey *et al.* (2015)	Yes	Nepali Himalaya	No	No
WDNR	State of Wisconsin (2014)	Yes	Wisconsin	No	No
WR-VISTA	Anandhi and Kannan (2018)	No		Yes	No

No title	Chhetri *et al*. (2020)	Yes	Hilly Region of Nepal	No	No
The watershed sustainability index	Chaves and Alipaz (2007)	Yes	South America, Oceania, Africa	No	No
The Water, Economy, Investment and Learning Assessment Indicator (WEILAI)	Cohen and Sullivan (2010)	Yes	Rural China	No	No
Water Poverty Index	Lawrence *et al*. (2002)	No		No	No
N/A	Hamouda *et al*. (2009)	Yes	East Nile Basin countries	Yes	No
N/A	Hurd *et al* (1999)	Yes	United States	No	No
N/A	Chang *et al* (2013)	Yes	Columbia River Basin	No	No
N/A	Kim *et al* (2013)	Yes	South Korea	No	No

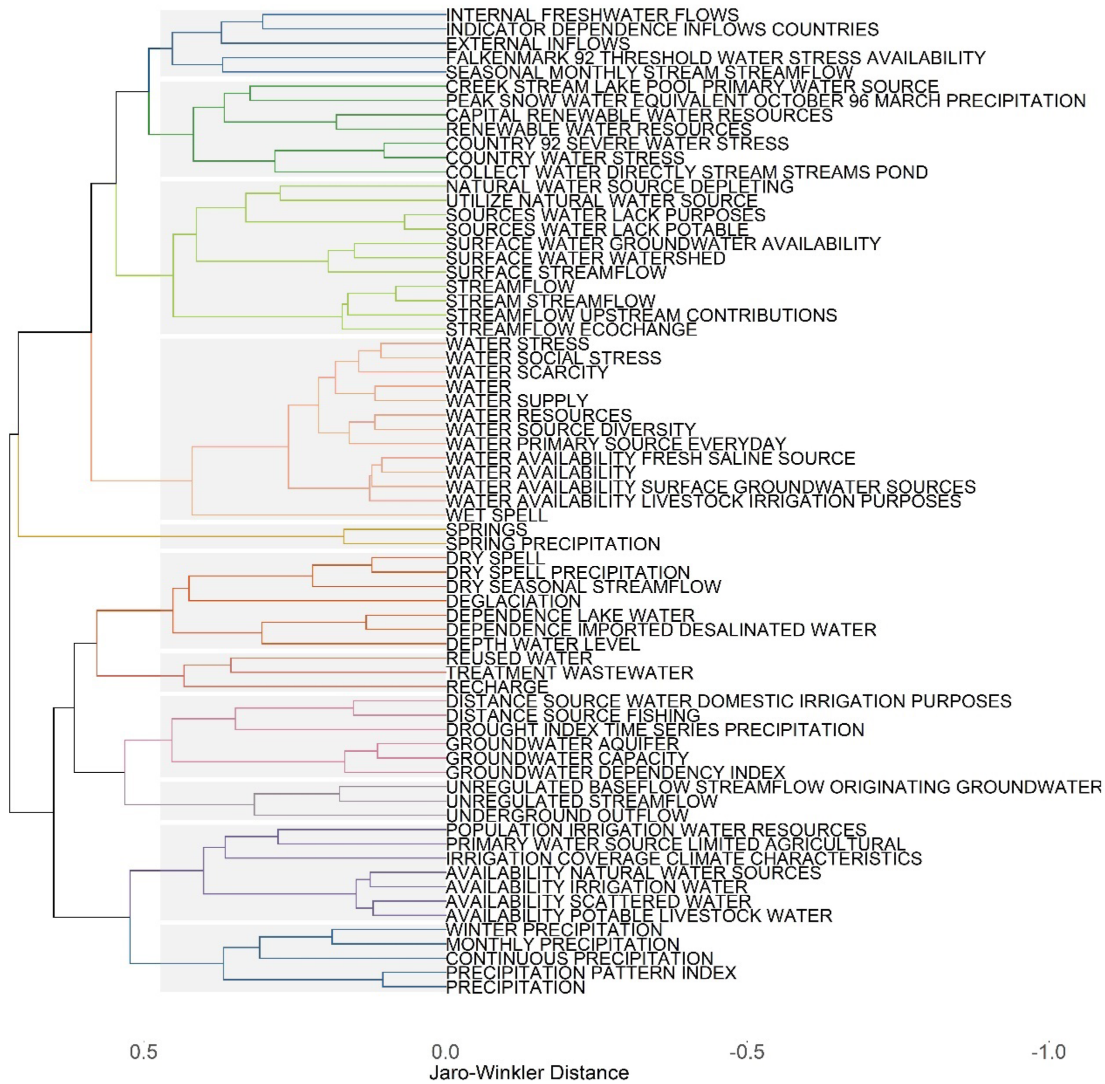

Figure S5- 1: Full cluster results for the water source sub-domain.

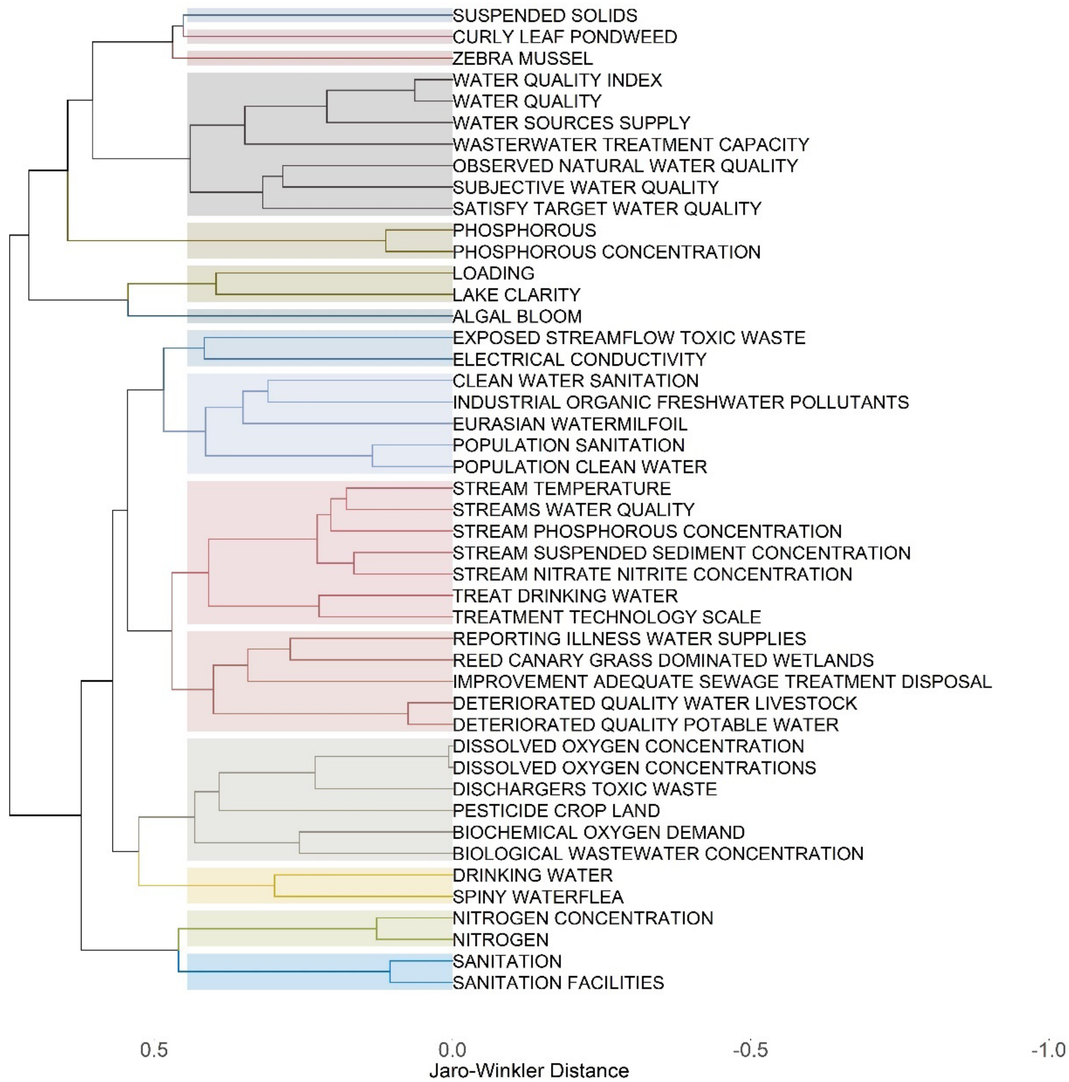

Figure S5-2: Full cluster results for the water quality sub-domain.

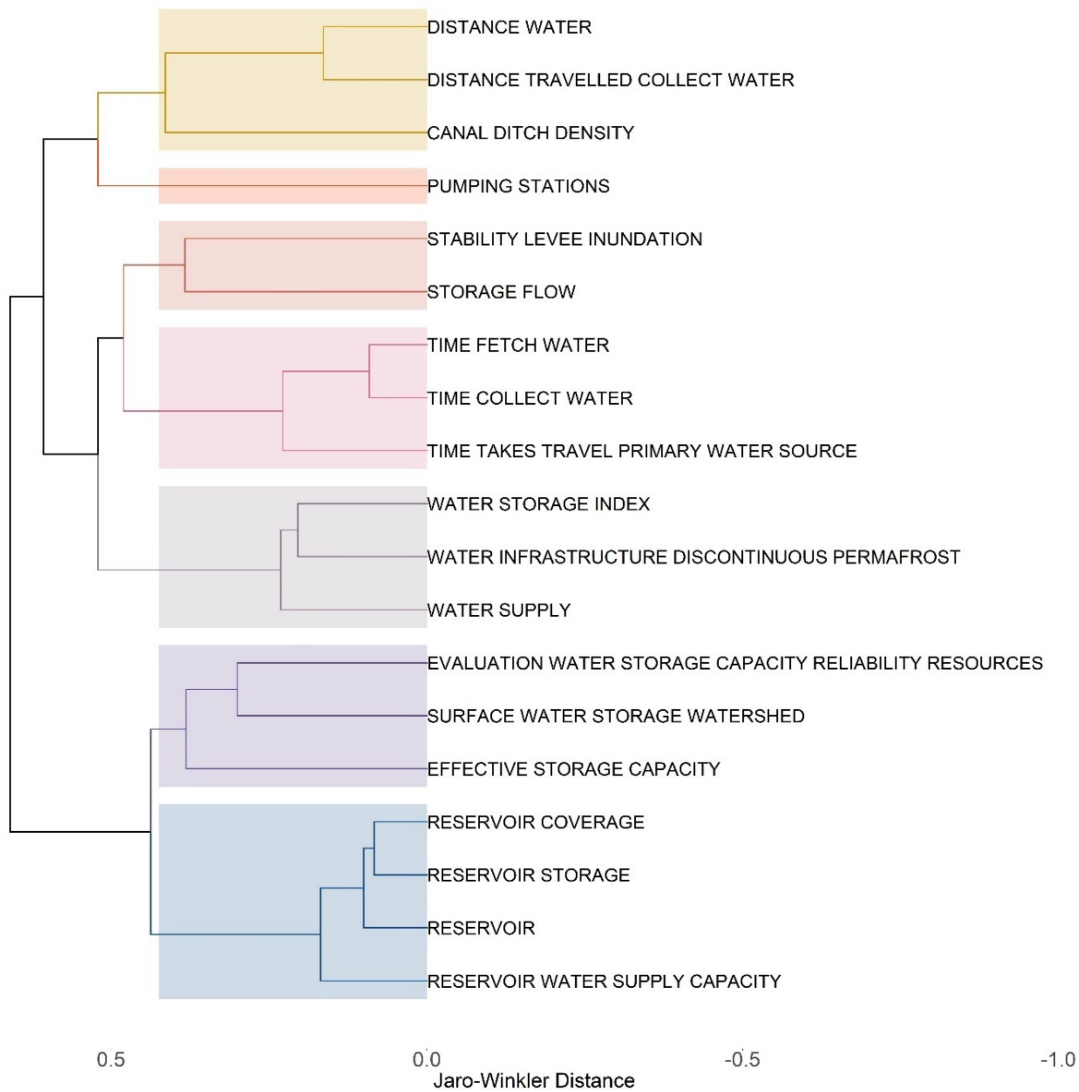

Figure S5-3: Full cluster results for the water infrastructure and distribution sub-domain.

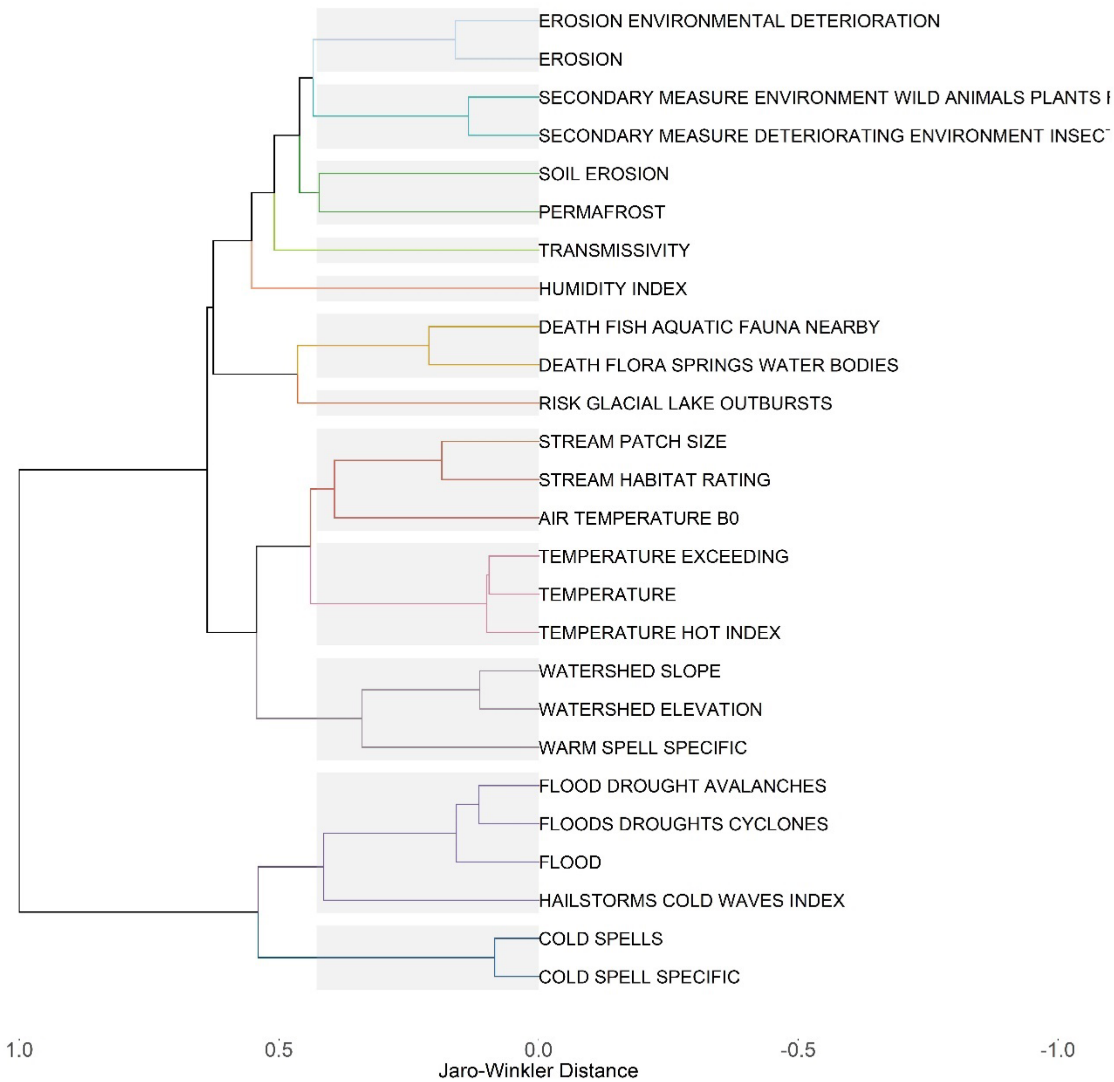

Figure S5-4: Full cluster results for the physical environment sub-domain.

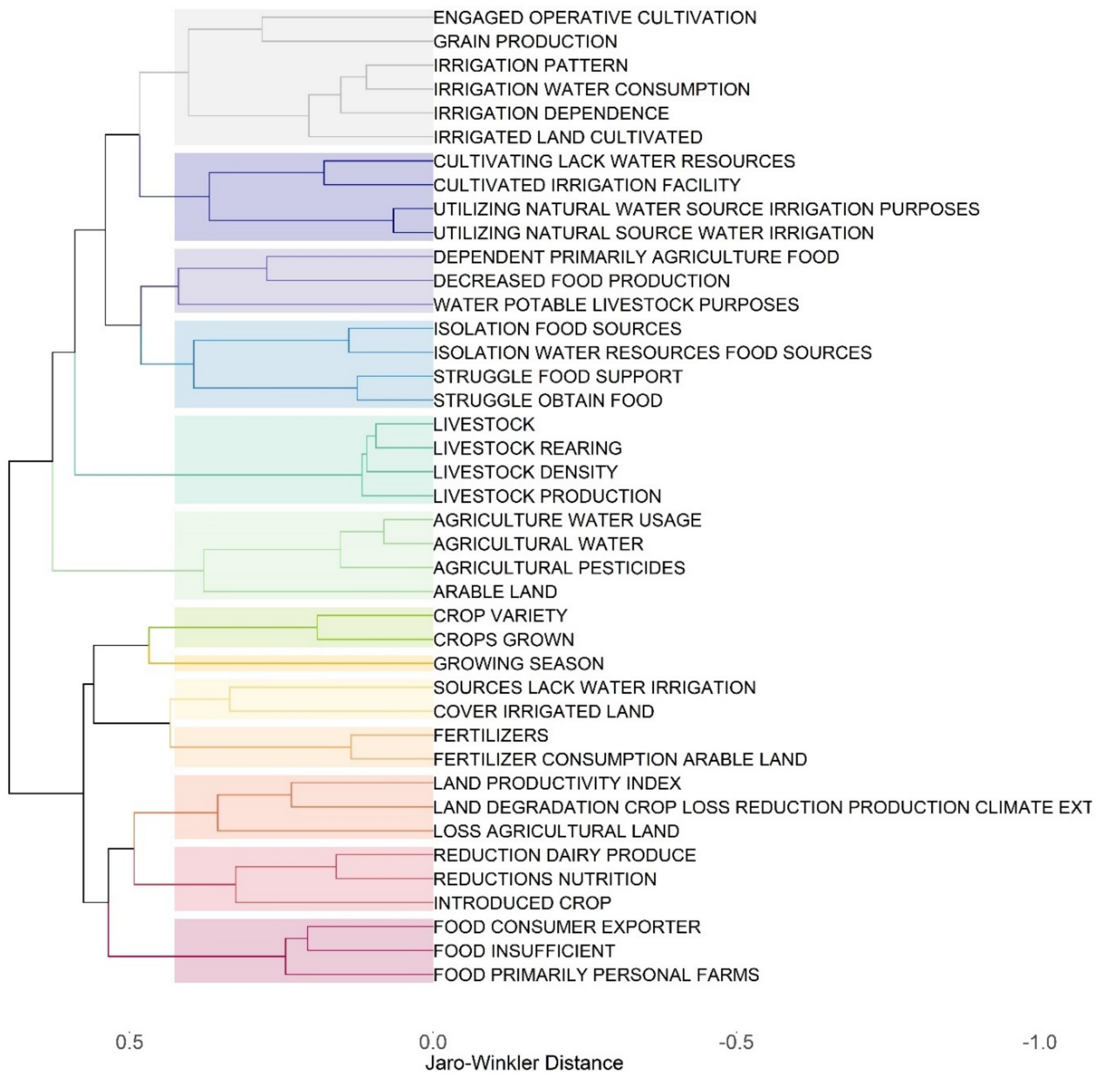

Figure S5- 5: Full cluster results for the agricultural land and water use sub-domain.

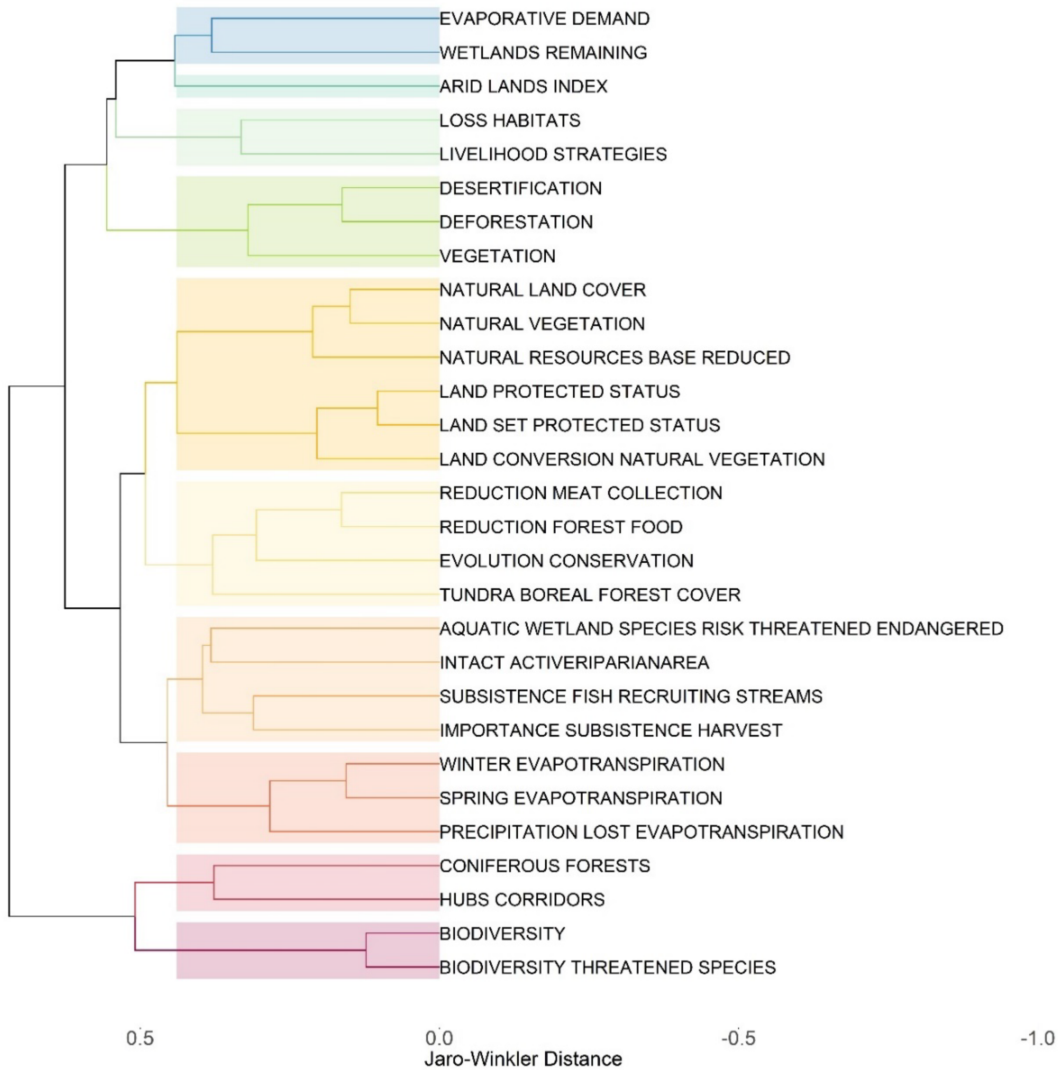

Figure S5-6: Full cluster results for the environmental and cultural land and water use sub-domain.

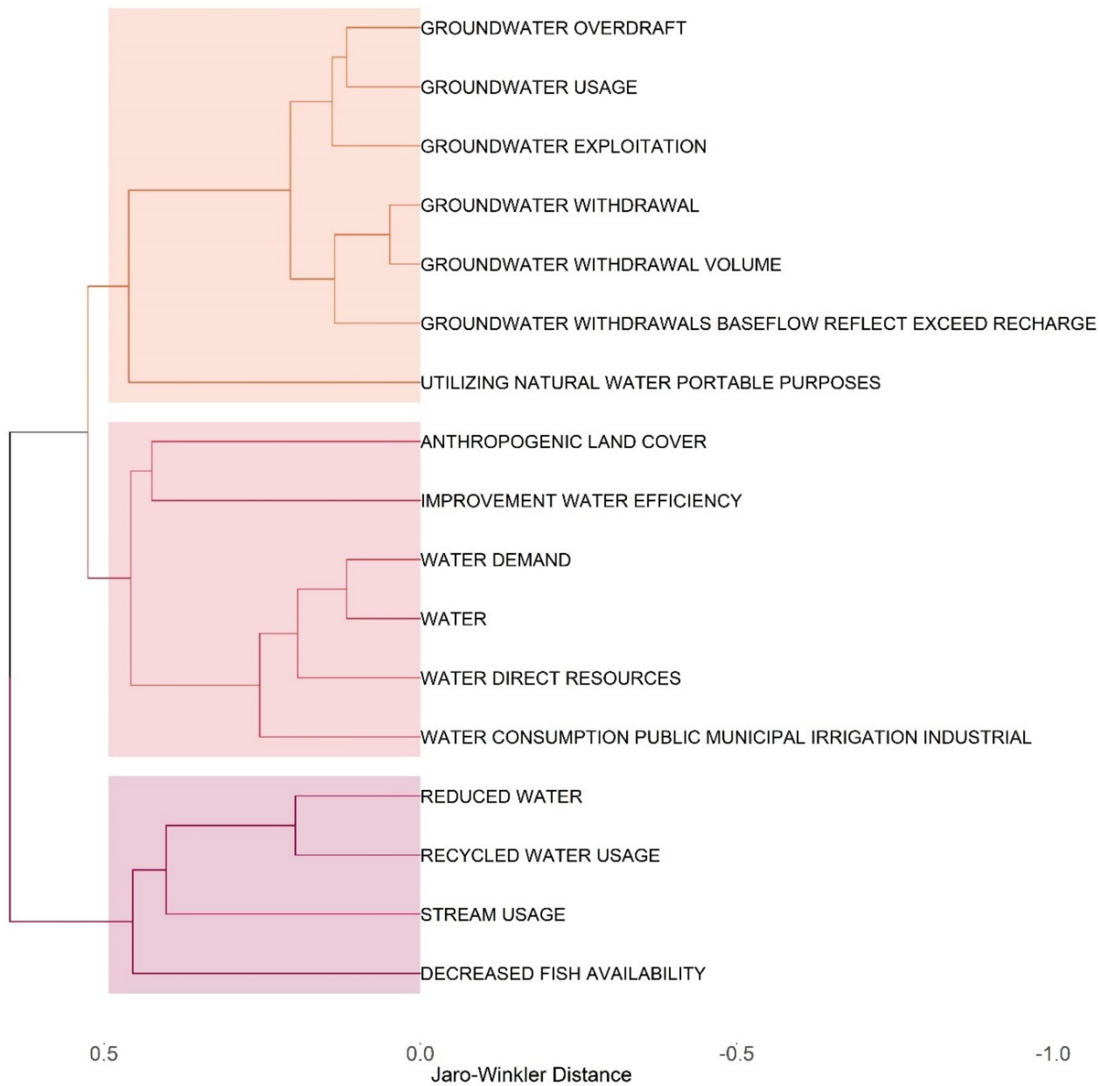

Figure S5- 7: Full cluster results for the general land and water use sub-domain.

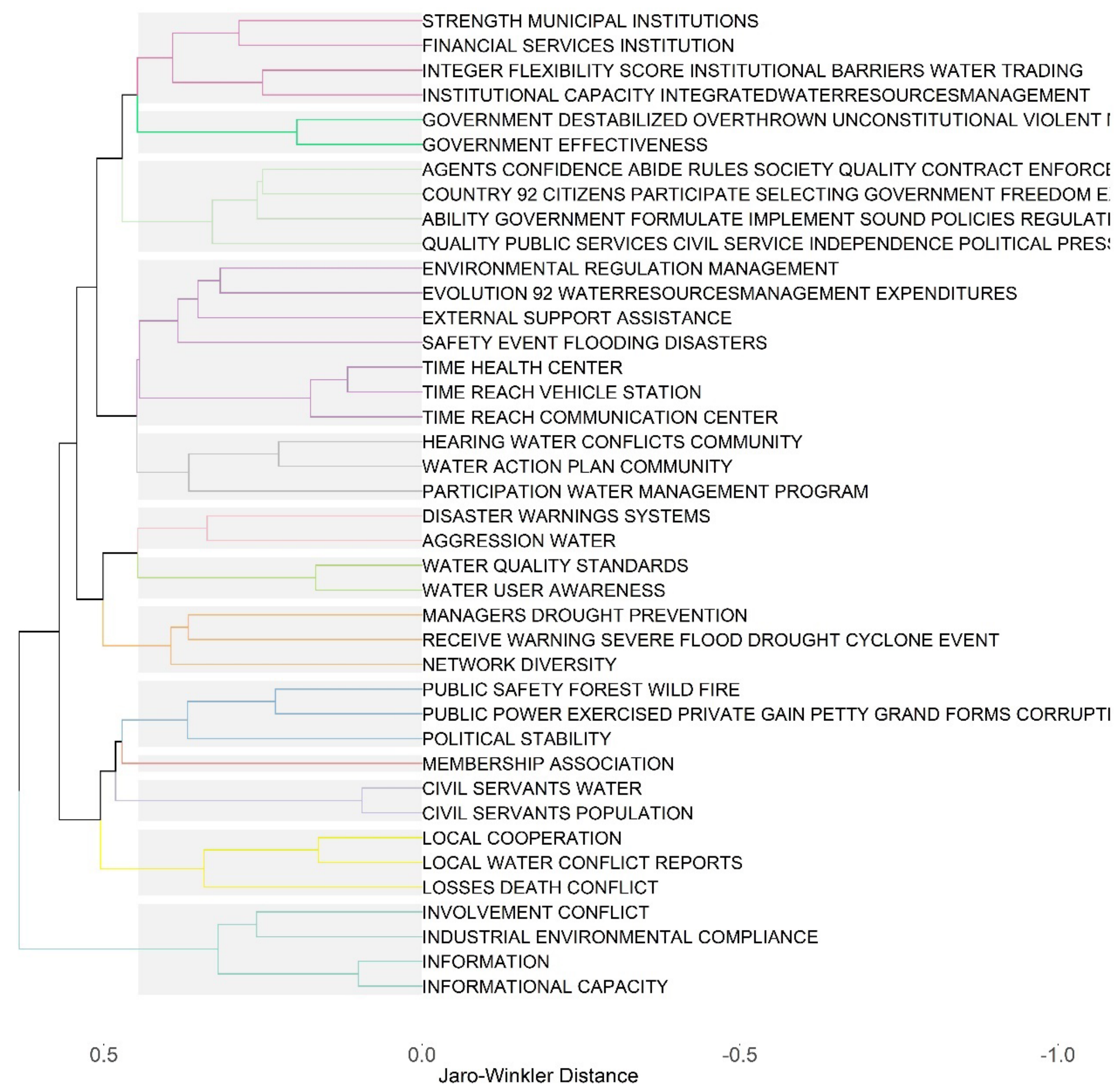

Figure S5-8: Full cluster results for the institutions and management sub-domain.

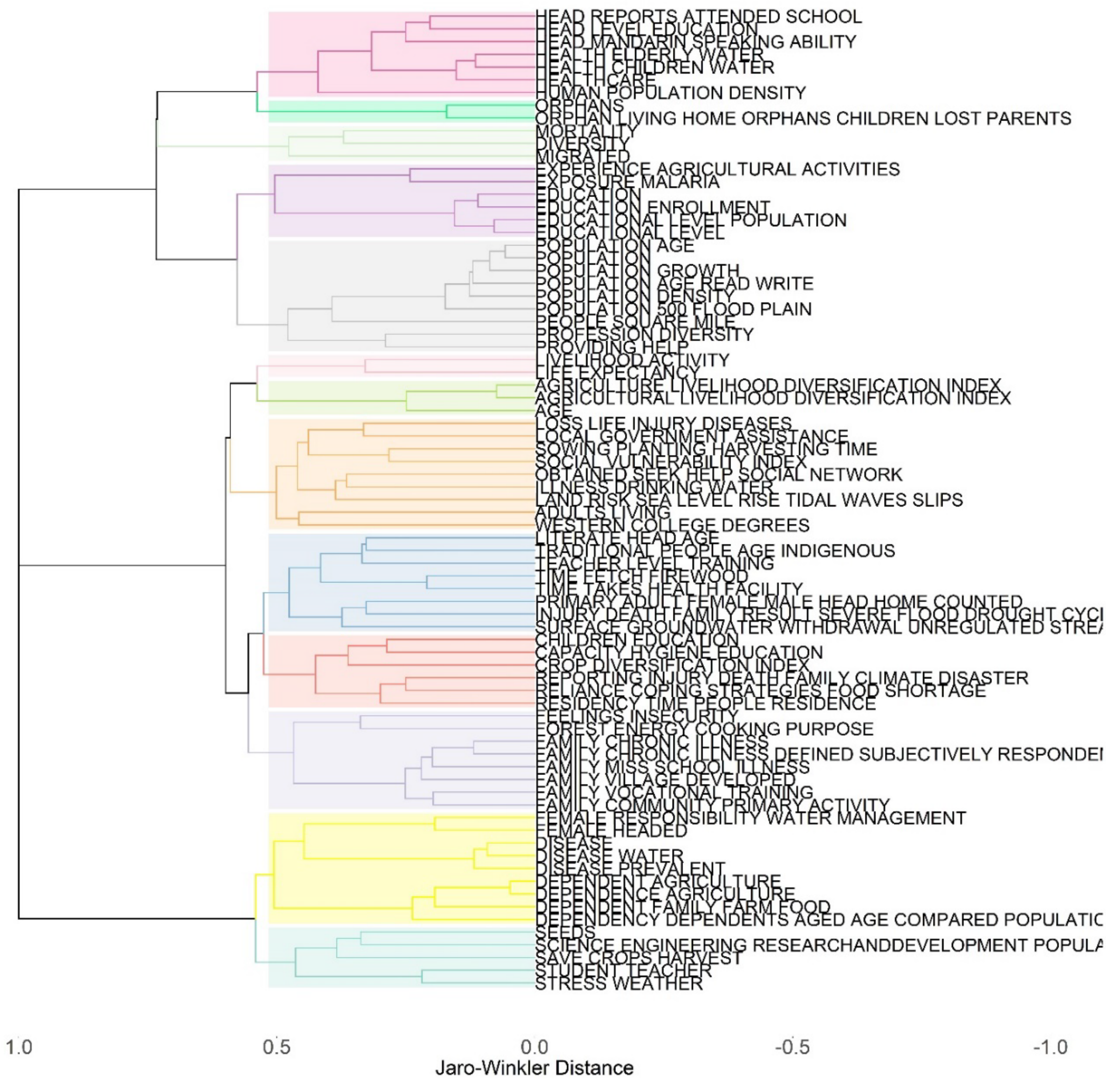

Figure S5-9: Full cluster results for socio-culture sub-domain.

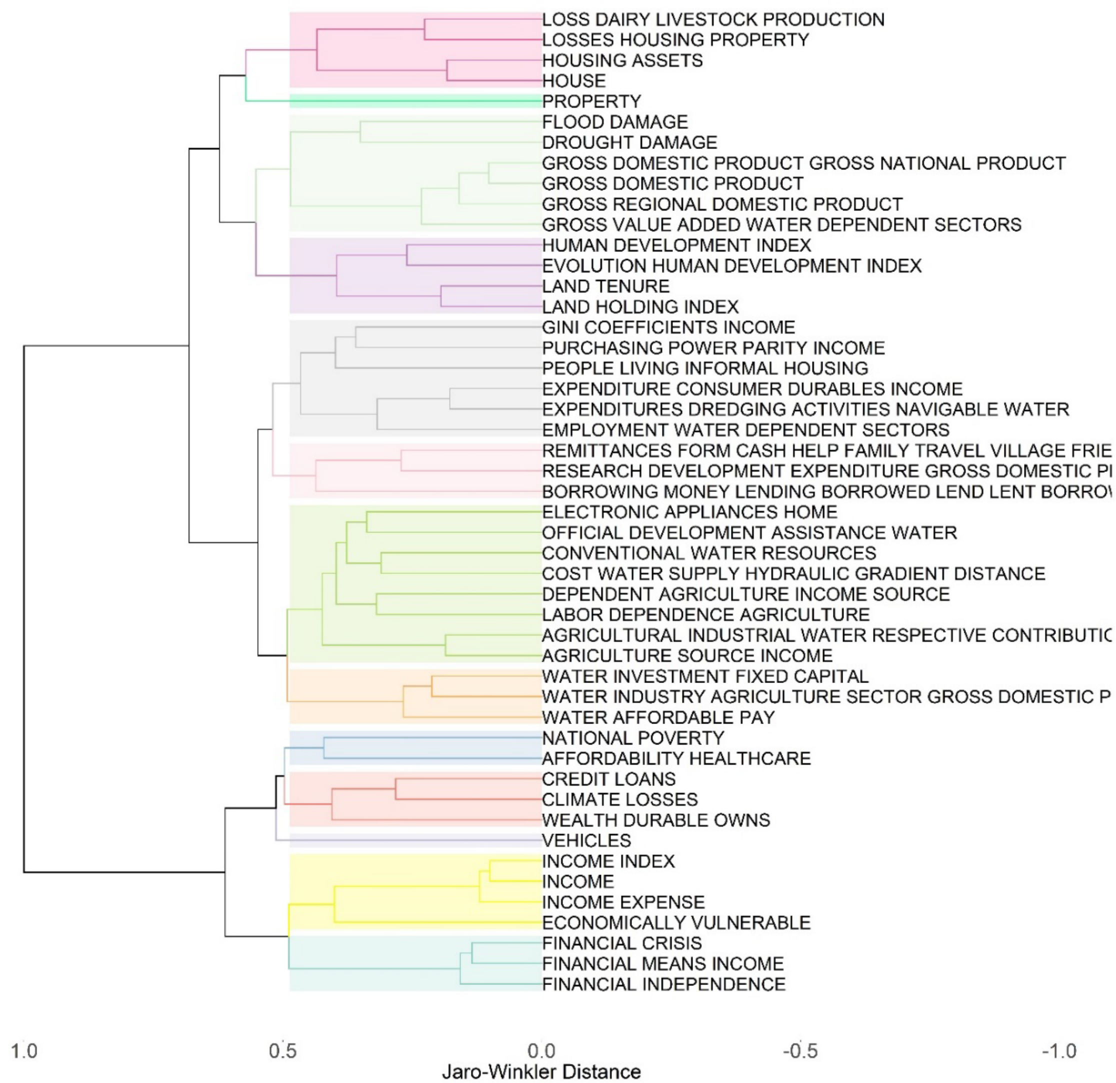

Figure S5- 10: Full cluster results for economics sub-domain.

5.6 References

Adger, W.N., 2006. Vulnerability. Glob. Environ. Chang. 16, 268–281. https://doi.org/10.1016/J.GLOENVCHA.2006.02.006

Adger, W.N., Kelly, P.M., 1999. Social vulnerability to climate change and the architecture of entitlements. Mitig. Adapt. Strateg. Glob. Chang. 4, 253–266. https://doi.org/10.1023/a:1009601904210

Alessa, L., Kliskey, A., Lammers, R., Arp, C., White, D., Hinzman, L., Busey, R., 2008. The arctic water resource vulnerability index: An integrated assessment tool for community resilience and vulnerability with respect to freshwater. Environ. Manage. 42, 523–541. https://doi.org/10.1007/s00267-008-9152-0

Amine, A., Elberrichi, Z., Simonet, M., 2010. Evaluation of text clustering methods using WordNet. Int. Arab J. Inf. Technol. 7, 349–357.

Anandhi, A., Kannan, N., 2018. Vulnerability assessment of water resources – Translating a theoretical concept to an operational framework using systems thinking approach in a changing climate: Case study in Ogallala Aquifer. J. Hydrol. 557, 460–474. https://doi.org/10.1016/j.jhydrol.2017.11.032

Birkmann, J., Cardona, • O D, Carreñ, • M L, Barbat, • A H, Pelling, • M, Schneiderbauer, • S, Kienberger, • S, Keiler, • M, Alexander, • D, Zeil, • P, Welle, • T, Welle, Á.T., Cardona, O.D., Carreño, M.L., Barbat, A.H., Pelling, M., Schneiderbauer, S., Kienberger, S., Zeil, Á.P., Zeil, P., Keiler, M., Alexander, D., 2013. Framing vulnerability, risk and societal responses: the MOVE framework 67, 193–211. https://doi.org/10.1007/s11069-013-0558-5

Blaikie, P., Cannon, T., Davis, I., Wisner, B., 2005. At Risk : Natural Hazards, People's Vulnerability and Disasters. Risk. https://doi.org/10.4324/9780203974575

Brooks, N., 2003. Vulnerability, risk and adaptation: A conceptual framework. Tyndall Centre for Climate Change Research and Centre for Social and Economic Research on the Global Environment (CSERGE). Tyndall Cent. Clim. Chang. Res. Working Pa, 20.

Brown, C.M., Lund, J.R., Cai, X., Reed, P.M., Zagona, E.A., Ostfeld, A., Hall, J., Characklis, G.W., Yu, W., Brekke, L., 2015. The future of water resources systems analysis: Toward a scientific framework for sustainable water management. Water Resour. Res. 51, 6110–6124. https://doi.org/10.1002/2015WR017114

Cardona, O.-D., van Aalst, M.K., Birkmann, Jörn, Fordham, Maureen, McGregor, Glenn, Perez, Rosa, Pulwarty, R.S., Lisa Schipper, E.F., Tan Sinh, B., Décamps, H., Keim, M., Davis, I., van Aalst, M., Birkmann, J, Fordham, M, McGregor, G, Perez, R, Pulwarty, R., Schipper, E., Sinh, B., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mach, K., Plattner, G., Allen, S., Tignor, M., Midgley, P., 2012. Coordinating Lead Authors: Lead Authors: Review Editors: Contributing Authors:: Determinants of risk: exposure and vulnerability. In: Managing the Risks of Extreme Events and

Disasters to Advance Climate Change Adaptation 2 Determinants of Risk: Exposure and Vulnerability.

Cardona, O.D., 2011. Disaster Risk and Vulnerability: Concepts and Measurement of Human and Environmental Insecurity 107–121. https://doi.org/10.1007/978-3-642-17776-7_3

Chief, K., Meadow, A., Whyte, K., 2016. Engaging Southwestern Tribes in Sustainable Water Resources Topics and Management. Water 2016, Vol. 8, Page 350 8, 350. https://doi.org/10.3390/W8080350

Cutter, S.L., Finch, C., 2008. Temporal and spatial changes in social vulnerability to natural hazards. Proc. Natl. Acad. Sci. 105, 2301–2306. https://doi.org/10.1073/PNAS.0710375105

Davenport, F. V., Herrera-Estrada, J.E., Burke, M., Diffenbaugh, N.S., 2020. Flood Size Increases Nonlinearly Across the Western United States in Response to Lower Snow-Precipitation Ratios. Water Resour. Res. 56, e2019WR025571. https://doi.org/10.1029/2019WR025571

de Ruiter, M.C., van Loon, A.F., 2022. The challenges of dynamic vulnerability and how to assess it. iScience 25, 104720. https://doi.org/10.1016/J.ISCI.2022.104720

Dilling, L., Berggren, J., 2015. What do stakeholders need to manage for climate change and variability? A document-based analysis from three mountain states in the Western USA. Springer. https://doi.org/10.1007/s10113-014-0668-y

Dilling, L., Lackstrom, K., Haywood, B., Dow, K., Lemos, M.C., Berggren, J., Kalafatis, S., 2015. What Stakeholder Needs Tell Us about Enabling Adaptive Capacity: The Intersection of Context and Information Provision across Regions in the United States. Weather. Clim. Soc. 7, 5–17. https://doi.org/10.1175/WCAS-D-14-00001.1

Eakin, H., Luers, A.L., 2006. ASSESSING THE VULNERABILITY OF SOCIAL-ENVIRONMENTAL SYSTEMS. https://doi.org/10.1146/annurev.energy.30.050504.144352

Ehsani, N., Vörösmarty, C.J., Fekete, B.M., Stakhiv, E.Z., 2017. Reservoir operations under climate change: Storage capacity options to mitigate risk. J. Hydrol. 555, 435–446. https://doi.org/10.1016/j.jhydrol.2017.09.008

Enqvist, J.P., Ziervogel, G., Johan Enqvist, C.P., Climate and, A., 2019. Water governance and justice in Cape Town: An overview. Wiley Interdiscip. Rev. Water 6, e1354. https://doi.org/10.1002/WAT2.1354

Feinerer, I., 2022. Introduction to the tm Package Text Mining in R.

Fortini, L., Schubert, O., 2017. Beyond exposure, sensitivity and adaptive capacity: a response based ecological framework to assess species climate change vulnerability. Clim. Chang. Responses 2017 41 4, 1–7. https://doi.org/10.1186/S40665-017-0030-Y

Füssel, H.M., 2007. Vulnerability: A generally applicable conceptual framework for climate change research. Glob. Environ. Chang. 17, 155–167. https://doi.org/10.1016/J.GLOENVCHA.2006.05.002

Gallopín, G.C., 2006. Linkages between vulnerability, resilience, and adaptive capacity. Glob. Environ. Chang. 16, 293–303. https://doi.org/10.1016/J.GLOENVCHA.2006.02.004

Gleick, P.H., 1993. Water and Conflict: Fresh Water Resources and International Security. Int. Secur. 18, 79. https://doi.org/10.2307/2539033

Gordon, B.L., Rego, J.J., Koebele, E.K. (2019). Vulnerability Indicators (Version 2.0.4) [Github]. https://github.com/BeaGordon/VulnerabilityIndicators

Gordon, B.L., Brooks, P.D., Krogh, S.A., Boisrame, G.F.S., Carroll, R.W.H., McNamara, J.P., Harpold, A.A., 2022. Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework. Environ. Res. Lett. 17, 053004. https://doi.org/10.1088/1748-9326/AC64B4

Hamouda, M.A., Nour El-Din, M.M., Moursy, F.I., 2009. Vulnerability assessment of water resources systems in the Eastern Nile Basin. Water Resour. Manag. 23, 2697–2725. https://doi.org/10.1007/s11269-009-9404-7

Hashimoto, T., Stedinger, J.R., Loucks, D.P., 1982. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. Water Resour. Res. 18, 14–20. https://doi.org/10.1029/WR018i001p00014

He, X., Bryant, B.P., Moran, T., Mach, K.J., Wei, Z., Freyberg, D.L., 2021. Climate-informed hydrologic modeling and policy typology to guide managed aquifer recharge, Sci. Adv.

Herrera-Estrada, J.E., Martinez, J.A., Dominguez, F., Findell, K.L., Wood, E.F., Sheffield, J., 2019. Reduced Moisture Transport Linked to Drought Propagation Across North America. Geophys. Res. Lett. 46, 5243–5253. https://doi.org/10.1029/2019GL082475

Hinkel, J., 2011. "Indicators of vulnerability and adaptive capacity": Towards a clarification of the science–policy interface. Glob. Environ. Chang. 21, 198–208. https://doi.org/10.1016/J.GLOENVCHA.2010.08.002

Hughes, S., Yau, A., Max, L., Petrovic, N., Davenport, F., Marshall, M., McClanahan, T.R., Allison, E.H., Cinner, J.E., 2012. A framework to assess national level vulnerability from the perspective of food security: The case of coral reef fisheries. Environ. Sci. Policy 23, 95–108. https://doi.org/10.1016/J.ENVSCI.2012.07.012

Hurd, B., Leary, N., Jones, R., Smith, J., 1999. Relative regional vulnerability of water resources to climate change. J. Am. Water Resour. Assoc. 35, 1399–1409. https://doi.org/10.1111/J.1752-1688.1999.TB04224.X/FORMAT/PDF

Immerzeel, W.W., Lutz, A.F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B.J., Elmore, A.C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J.S., Koppes, M., Kraaijenbrink, P.D.A., Kulkarni, A. V, Mayewski, P.A., Nepal, S., Pacheco, P., Painter, T.H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A.B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., Baillie, J.E.M., 2020. Importance and vulnerability of the world's water towers. 364 | Nat. | 577. https://doi.org/10.1038/s41586-019-1822-y

Jaro, M.A., 1989. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. J. Am. Stat. Assoc. 84, 414–420. https://doi.org/10.1080/01621459.1989.10478785

Kasperson, J.X., Kasperson, R.E., Turner, B., Hsieh, W., Schiller, A., 2012. Vulnerability to global environmental change. Soc. Contours Risk Vol. II Risk Anal. Corp. Glob. Risk 245–285. https://doi.org/10.4324/9781849772549

Kellner, E., 2021. The controversial debate on the role of water reservoirs in reducing water scarcity. Wiley Interdiscip. Rev. Water. https://doi.org/10.1002/wat2.1514

Kellner, E., Brunner, M.I., 2021. Reservoir Governance in World's Water Towers Needs to Anticipate Multi-purpose Use. Earth's Futur. 9, 1–19. https://doi.org/10.1029/2020EF001643

Kelly, P.M., Adger, W.N., 2000. Theory and practice in assessing vulnerability to climate change and facilitating adaptation. Clim. Change 47, 325–352. https://doi.org/10.1023/A:1005627828199

Kumar, C.P., 2012. Climate Change and Its Impact on Groundwater Resources.

Lawrence, P., Meigh, J., Sullivan, C., 2002. The Water Poverty Index: an International Comparison.

Leonardo, B., Hansun, S., 2017. Text Documents Plagiarism Detection using Rabin-Karp and Jaro-Winkler Distance Algorithms. Artic. Indones. J. Electr. Eng. Comput. Sci. 5, 462–471. https://doi.org/10.11591/ijeecs.v5.i2.pp462-471

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Makowski, D., 2020. Extracting, Computing and Exploring the Parameters of Statistical Models using R. J. Open Source Softw. 5, 2445. https://doi.org/10.21105/JOSS.02445

Luers, A.L., 2005. The surface of vulnerability: An analytical framework for examining environmental change. Glob. Environ. Chang. 15, 214–223. https://doi.org/10.1016/j.gloenvcha.2005.04.003

Luers, A.L., Lobell, D.B., Sklar, L.S., Addams, C.L., Matson, P.A., 2003. A method for quantifying vulnerability, applied to the agricultural system of the Yaqui Valley, Mexico. Glob. Environ. Chang. 13, 255–267. https://doi.org/10.1016/S0959-3780(03)00054-2

Mamun, A. Al, Aseltine, R., Rajasekaran, S., 2016. Efficient Record Linkage Algorithms Using Complete Linkage Clustering. PLoS One 11.

https://doi.org/10.1371/JOURNAL.PONE.0154446

Mankin, J.S., Viviroli, D., Singh, D., Hoekstra, A.Y., Diffenbaugh, N.S., 2015. The potential for snow to supply human water demand in the present and future. Environ. Res. Lett. 10, 114016. https://doi.org/10.1088/1748-9326/10/11/114016

Marlow, D.R., Moglia, M., Cook, S., Beale, D.J., 2013. Towards sustainable urban water management: A critical reassessment. Water Res. 47, 7150–7161. https://doi.org/10.1016/J.WATRES.2013.07.046

Notaro, V., De Marchis, M., Fontanazza, C.M., La Loggia, G., Puleo, V., Freni, G., 2014. The Effect of Damage Functions on Urban Flood Damage Appraisal. Procedia Eng. 70, 1251–1260. https://doi.org/10.1016/J.PROENG.2014.02.138

O'Brien, K.L., Leichenko, R.M., 2000. Double exposure: assessing the impacts of climate change within the context of economic globalization. Glob. Environ. Chang. 10, 221–232. https://doi.org/10.1016/S0959-3780(00)00021-2

Okpara, U.T., Lindsay, •, Stringer, C., Dougill, A.J., 2016. Using a novel climate-water conflict vulnerability index to capture double exposures in Lake Chad. Reg. Environ. Chang. 17. https://doi.org/10.1007/s10113-016-1003-6

Okpara, U.T., Stringer, L.C., Dougill, A.J., 2017. Using a novel climate–water conflict vulnerability index to capture double exposures in Lake Chad. Reg. Environ. Chang. 17, 351–366. https://doi.org/10.1007/s10113-016-1003-6

Orford, J.D., 1976. Implementation of criteria for partitioning a dendrogram. J. Int. Assoc. Math. Geol. 8, 75–84. https://doi.org/10.1007/BF01039686

Pelling, M., 2010. Adaptation to Climate Change: From Resilience to Transformation. Adapt. to Clim. Chang. From Resil. to Transform. 1–203. https://doi.org/10.4324/9780203889046/ADAPTATION-CLIMATE-CHANGE-MARK-PELLING

Plummer, R., de Loë, R., Armitage, D., 2012. A Systematic Review of Water Vulnerability Assessment Tools. Water Resour. Manag. 26, 4327–4346. https://doi.org/10.1007/s11269-012-0147-5

Qin, Y., Abatzoglou, J.T., Siebert, S., Huning, L.S., AghaKouchak, A., Mankin, J.S., Hong, C., Tong, D., Davis, S.J., Mueller, N.D., 2020. Agricultural risks from changing snowmelt. Nat. Clim. Chang. 10, 459–465. https://doi.org/10.1038/s41558-020-0746-8

Rajalingam, N., Ranjini, K., 2011. Hierarchical Clustering Algorithm-A Comparative Study. Int. J. Comput. Appl. 19, 975–8887.

Ram, S., Hwang, Y., Zhao, H., 2005. A clustering based approach for facilitating semantic web service discovery. 15th Work. Inf. Technol. Syst. WITS 2005 3–8. https://doi.org/10.2139/ssrn.893985

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Saraçli, S., Doğan, N., Doğan, I., 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. J. Inequalities Appl. 2013, 1–8. https://doi.org/10.1186/1029-242X-2013-203/TABLES/2

Savelli, E., Rusca, M., Cloke, H., Di Baldassarre, G., 2021. Don't blame the rain: Social power and the 2015–2017 drought in Cape Town. J. Hydrol. 594, 125953. https://doi.org/10.1016/J.JHYDROL.2020.125953

Scott, C.A., Zilio, M.I., Harmon, T., Zuniga-Teran, A., Díaz-Caravantes, R., Hoyos, N., Perillo, G.M.E., Meza, F., Varady, R.G., Neto, A.R., Velez, M.I., Martín, F., Escobar, J., Piccolo, M.C., Mussetta, P., Montenegro, S., Rusak, J.A., Pineda, N., 2021. Do ecosystem insecurity and social vulnerability lead to failure of water security? Environ. Dev. 38, 100606. https://doi.org/10.1016/J.ENVDEV.2020.100606

Smith, A., Ali, M., 2006. Understanding the impact of cultural and religious water use. Water Environ. J. 20, 203–209. https://doi.org/10.1111/J.1747-6593.2006.00037.X

Steyaert, J.C., Condon, L.E., W.D. Turner, S., Voisin, N., 2022. ResOpsUS, a dataset of historical reservoir operations in the contiguous United States. Sci. data 9. https://doi.org/10.1038/S41597-022-01134-7

Sullivan, C., Meigh, J., 2005. Targeting attention on local vulnerabilities using an integrated index approach: the example of the climate vulnerability index. Water Sci. Technol. 51, 69–78. https://doi.org/10.2166/WST.2005.0111

Sullivan, C.A., 2011. Quantifying water vulnerability: A multi-dimensional approach. Stoch. Environ. Res. Risk Assess. 25, 627–640. https://doi.org/10.1007/s00477-010-0426-8

Tandon, R., 2012. Determination of optimal number of clusters in wireless sensor networks 4, 235–249.

Tapley, B.D., Bettadpur, S., Ries, J.C., Thompson, P.F., Watkins, M.M., 2004. GRACE measurements of mass variability in the Earth system. Science (80-.). 305, 503–505. https://doi.org/10.1126/science.1099192

Toze, S., 2006. Reuse of effluent water—benefits and risks. Agric. Water Manag. 80, 147–159. https://doi.org/10.1016/j.agwat.2005.07.010

Tseng, Y.H., Tsay, M.Y., 2013. Journal clustering of library and information science for subfield delineation using the bibliometric analysis toolkit: CATAR. Scientometrics 95, 503–528. https://doi.org/10.1007/s11192-013-0964-1

Turner, B.L., Kasperson, R.E., Matson, P.A., Mccarthy, J.J., Corell, R.W., Christensen, L., Eckley, N., Kasperson, J.X., Luers, A., Martello, M.L., Polsky, C., Pulsipher, A., Schiller, A., Turner, B.L., Kasperson, R.E., Matsone, P.A., Mccarthy, J.J., Corell, R.W., Christensene, L., Eckley, N., Kasperson, J.X., Luers, A., Martello, M.L., Polsky, C., Pulsipher, A., Schiller, A., 2003. A framework for vulnerability analysis in sustainability science, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences. https://doi.org/10.1073/pnas.1231335100

van den Berg, H.J., Keenan, J.M., 2019. Dynamic vulnerability in the pursuit of just adaptation processes: A Boston case study. Environ. Sci. Policy 94, 90–100. https://doi.org/10.1016/J.ENVSCI.2018.12.015

Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: Vulnerability from climate change and population growth. Science (80-.). 289, 284–288. https://doi.org/10.1126/science.289.5477.284

Vörösmarty, C.J., Mcintyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467. https://doi.org/10.1038/nature09440

Vuille, M., Carey, M., Huggel, C., Buytaert, W., Rabatel, A., Jacobsen, D., Soruco, A., Villacis, M., Yarleque, C., Timm, O.E., Condom, T., Salzmann, N., Sicart, J.-E., 2018. Rapid decline of snow and ice in the tropical Andes-Impacts, uncertainties and challenges ahead A R T I C L E I N F O. Earth-Science Rev. 176, 195–213. https://doi.org/10.1016/j.earscirev.2017.09.019

Wiener, M.J., Jafvert, C.T., Nies, L.F., 2016. The assessment of water use and reuse through reported data: A US case study. Sci. Total Environ. 539, 70–77. https://doi.org/10.1016/J.SCITOTENV.2015.08.114

Zuniga-Teran, A.A., Fisher, L.A., Meixner, T., Le Tourneau, F.M., Postillion, F., 2022. Stakeholder participation, indicators, assessment, and decision-making: applying adaptive management at the watershed scale. Environ. Monit. Assess. 194, 1–19. https://doi.org/10.1007/S10661-021-09741-4/FIGURES/4

Zuniga-Teran, A.A., Mussetta, P.C., Lutz Ley, A.N., Díaz-Caravantes, R.E., Gerlak, A.K., 2021. Analyzing water policy impacts on vulnerability: Cases across the rural-urban continuum in the arid Americas. Environ. Dev. 38, 100552. https://doi.org/10.1016/J.ENVDEV.2020.100552

6 Chapter 6: Conclusions

In mountain environments, climate change has already substantially and rapidly altered snow resources (Musselman *et al*., 2017). Evidence suggests that changes in the persistence and amount of snow are likely to be ongoing and more complex under continued climate change (Barnett *et al*., 2008; Rauscher *et al*., 2008). As a result, society must adapt to uncertainty in the amount and timing of mountain water supplies (Adam *et al*., 2009; Stewart, 2009), which will impact people, agriculture (Qin *et al*., 2020), economic productivity (Barnett *et al*., 2008; Sturm *et al*., 2017), ecosystem health (Allan and Castillo, 2007), wildfires (Holden *et al*., 2012; Westerling *et al*., 2006), flood risk (Davenport *et al*., 2020; Hamlet and Lettenmaier, 2007), spiritual and cultural practices (Immerzeel *et al*., 2020; Vuille *et al*., 2018), and reservoir management (Ajami *et al*., 2008; Ehsani *et al*., 2017) to name only a few. To assist scientists, water managers, and policy-makers in the grand challenge of adapting socio-hydrologic systems to these multifaceted changes, this dissertation is motivated to answer a single question: How can we better characterize the vulnerability—and adaptive capacity— of socio-hydrologic systems to shifts in mountain water supplies driven by climate change?

We do answer this question in four parts, coupling technical investigations of hydrology with robust and interdisciplinary methods to illustrate:

1. the mechanisms contributing to changing mountain water supplies;
2. the tools available for quantifying water supply contributions from mountain environments;

3. the ways in which humans interact with water supplies from mountain environments to amplify or moderate vulnerability; and,
4. the approaches available for measuring water supply vulnerability in a dynamic and multidimensional manner.

As a whole, each of these Chapters contributes novel information for society as we seek to adapt to changes in mountain water resources. We summarize the major conclusions of this dissertation below:

Key Conclusions

- Predicting changes in streamflow arising from changes in snow is uniquely critical and uniquely challenging in the western US. Due to high potential for interaction between the mechanisms controlling how changes in snow are translated into changes in streamflow, mountainous catchments in the western US are likely to be impacted by changes in the timing and intensity of water inputs as well as increases in the amount of water lost to the environment during the snow season.
- Existing tools—and data—for characterizing water supplies are imperfect and particularly so when it comes to quantifying groundwater contributions from mountain environments. Simplifications with regard to the impacts of measurement error and the negligibility of groundwater are often used to quantify and make predictions about streamflow. However, in doing so these tools neglect potential groundwater contributions from high, arid upland catchments with deep permeable substrates.

- Foregrounding the effects of dynamic physical changes in mountain water supplies overlooks how human activity can moderate the consequences of these changes, particularly in the near future if storage and demand management solutions are pursued.
- Water supply vulnerability must be considered—and assessed—in a dynamic and multidimensional manner. Static assessments of physical vulnerability neglect the social value of water and often diminish the active role demand plays in determining vulnerability. Maintaining the practical benefits of indicator-based assessments is essential for continued vulnerability assessment in the management and policy space. However, we need flexible approaches—and, critically, more comprehensive indicators and data for evaluating these indicators—in order to move closer to the true vulnerability of socio-hydrologic systems in the face of changing water supplies.

Key Themes and Recommendations

1. **The benefits of simplification.** This dissertation focuses broadly on investigating the benefits and the costs of simplification in different ways: first through metrics, then through metrics and tools, and finally through metrics, tools, and systems. In Chapter 2, we highlight how simple mechanisms that reduce complex interactions between the subsurface, snow, and the atmosphere can be used to explain variability in streamflow response to changing snow. In Chapter 3, we illustrate the benefits of a simple framework proposed by Fan (2019) for conditioning expectations about groundwater contributions from mountainous catchments. In Chapters 4 and 5, we

rely upon demonstrations using indicators of system performance to capture elements of vulnerability in a policy and management relevant manner. We recommend the following:

- Continued testing of simple mechanisms to explain complex hydrological processes and incorporation of these mechanisms into modeling and observational efforts.
- Continued development of simple metrics, particularly with regard to the social value of water. Here, specifically metrics to enhance consideration of the cultural value of water would enhance holistic vulnerability assessment.
- Development of more robust data for evaluating the social value of water across different systems. In this dissertation, we hypothesize that the availability of widespread gridded data products for evaluating physical aspects of vulnerability have led to somewhat lopsided evaluations of vulnerability that rely heavily on physical measures (e.g., precipitation) and can neglect the social value of water. Efforts to reconcile this gap could thus promote a more complete understanding of system vulnerabilities to climate induced changes in water supply.

2. **The costs of simplification, specifically with regard to the water budget.**

   At the same time, our results highlight the need to weigh decisions about simplification carefully. In Chapters 2 and 3 in particular, this dissertation illustrates the pitfalls of simplifying and conventionally accepted assumptions about water budget closure. In Chapter 2, we show that simplification of the water budget over time can mask important contributions from groundwater and lump measurement error into physical inferences about surface water. In Chapter 3, we illustrate how the spatial simplification of the water budget ignores the critical role institutions and laws in moving water around arid landscapes like the western US. In this way, spatial simplification can alter conclusions about the heterogeneity of vulnerability by specifically ignoring adaptive capacities. To remedy these challenges, we recommend the following:

- Reconsideration of closed water budgets in mountain environments. This dissertation reinforces that closed water budgets are particularly fragile in higher elevation, arid catchments in the western US with deep permeable substrates.
- Continued development of improved forcing data and bias correction methodologies particularly with regard to precipitation.
- Reconsideration of assumptions about the spatial coherence between supply and demand. We recommend the adoption of systems perspectives for water budget analyses in the western US, which we discuss in more detail below.

3. **The importance of systems perspectives for unraveling the impacts of changing mountain water resources on agriculture.**

Numerous hydrologic analyses have investigated and articulated the physical effects of climate change on agriculture through impacts to the timing and amount of mountain water resources. As outlined above, such analyses rely on assumptions about the spatial coherence between supply and demand at catchment, sub-basin, or basin scales. However, such an approach overlooks how socio-hydrologic systems have co-evolved not only with their physical hydrology, but also with policy, infrastructure, socio-economic conditions, and water demand; all of which can lead to variable different outcomes in response to changing water supplies. In addition to broad reconsideration of assumptions about supply and demand coherence at the catchment, sub-basin, or basin scale, we recommend that the next phase of water supply investigations adopt a more robust systems thinking approach. Such a shift would specifically include the following:

- More robust consideration of how society influences hydrology through land cover change and infrastructure—specifically reservoirs, water transfers, and evolving tools like managed aquifer recharge (MAR).
- Incorporation of the ways in which institutions and laws shape physical hydrology by moving water around from areas of comparatively higher availability to lower availability in the arid western US. Specifically, this recommendation would require analyses to center demand rather than supply in defining socio-hydrologic systems as is done in Chapter 4 of this dissertation. Such a shift will undoubtedly require more work and would be specifically aided by:
    - Improved spatial information about demand regions in the western US;
    - Standardized data about critical points of water supply for these demand regions. In this dissertation, we used a mix of grey literature and direct contact to construct 13 socio-hydrologic systems in the western US and found that information varied across states. For example, Wyoming maintains a fairly complete database of all

demand regions in the state and major points of water supply. Adoption of this template across state lines may be one way to promote more robust systems thinking in this space.

- Better data for water transfers and environmental flow requirements. We observed that data about water transfers into and out of the systems identified in Chapter 4 were highly variable. Encouragement of public reporting of these activities accompanied by a database may be one way in which this information can be better incorporated in analyses of system vulnerability and resilience in the western US.

4. The importance of inter and transdisciplinary research.

The adoption of systems thinking, which incorporates more robust consideration of how hydrology influences society and vice versa in a dynamic manner, must be accompanied by inter and transdisciplinary research in order to be effective. Here, we define interdisciplinary research as efforts that "analyze, synthesize, and harmonize links between disciplines into a coordinated and coherent whole" (Choi, 2006). Following Toomey *et al*. (2015) we define transdisciplinary work as engaging directly with knowledge production and use outside of academia. This dissertation is more interdisciplinary than transdisciplinary in nature and thus our recommendations are largely targeted towards enhancing work in that space. We recommend the following:

- Recognition of the time and effort required for interdisciplinary mentorship. The writing of this dissertation and incorporation of robust and defensible interdisciplinary research is the product of immense time and effort on the part of committee members and co-authors. The success of this dissertation's interdisciplinary Chapters (i.e., Chapters 4 and 5) was entirely dependent on the dedication of experts from different fields in developing, designing, and critiquing these Chapters.

- Early engagement with researchers from different fields. Chapter 4 of this dissertation would not have adopted a more realistic demand-centered view of socio-hydrologic systems were it not for early and active participation from researchers from different fields. As such, the establishment of an interdisciplinary team *prior to* conducting research in order to incorporate different perspectives in study design appears essential to successful outcomes.
- Recognition of the value of grey literature. This dissertation relies heavily on grey literature and personal communication in Chapter 4 in order to identify critical points of water supply. In order to move beyond simplifying assumptions and towards a systems perspective for water resources management, we need to recognize the value of and mine new sources of data. In this pursuit, an all-of-the-above approach to potential data sources that includes grey literature is essential and is one way to avoid stakeholder fatigue in the pursuit of more transdisciplinary research.