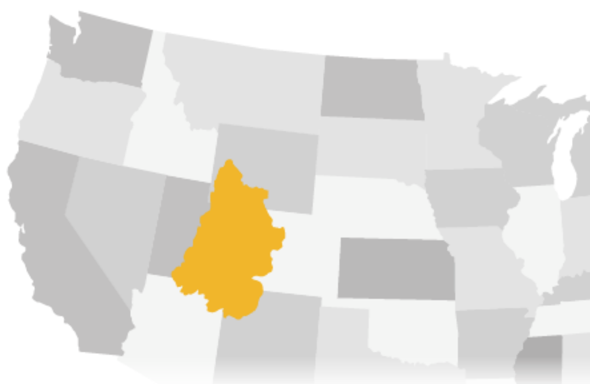
SEPTEMBER 2020

Project Hyperion - Narrative Case Study Report: *Colorado Headwaters*

Kripa Jagannathan (kajagannathan@lbl.gov) and **Andrew Jones** (adjones@lbl.gov)

Lawrence Berkeley Laboratory



Contributions

- Paul Ullrich, University of California, Davis (Project Team Leader)
- Bruce Riordan, Climate Readiness Institute (Engagement Facilitator)
- Alan Rhoades, Lawrence Berkeley Laboratory
- Binod Pokharel and Simon Wang, Utah State University
- Dana Veron and Sara Rauscher, University of Delaware
- Laurna Kaatz and Taylor Winchell, Denver Water
- Frank Kugel, Formerly with Upper Gunnison River Water Conservancy District
- James Eklund, Eklund Hanlon LLC
- Smitha Buddhavarapu, Lawrence Berkeley Laboratory

Contents

List of Tables

List of Figures

Introduction

This narrative case study report is a synthesis of key discussions and preliminary scientific results for the Colorado Headwaters region, undertaken as part of the Hyperion project (2016-19). Project Hyperion (now continuing as the HyperFACETS project) is a basic science project that aims to advance climate modelling by evaluating regional climate datasets for decision-relevant metrics. While there has been an explosive growth in the number of regional climate datasets available to users, there is limited understanding of the credibility and suitability of these datasets for use in different management decisions. Hyperion aims to address this need by developing comprehensive assessment capabilities to evaluate the credibility of regional climate datasets, understand the processes that contribute to model biases, and improve the ability of models to predict management relevant outcomes.

Since decision-relevance is a core motivation for the project, Hyperion is designed on the principles of co-production. The project brings together scientists from nine research institutions and managers from twelve water agencies in four watersheds: Sacramento/San Joaquin, Colorado Headwaters, South Florida, and Susquehanna. The project structure explicitly allows for both the groups to co-develop the science plan and research questions, in addition to co-producing the science itself. The scientists include atmospheric and earth system scientists as well as hydrologists. The water managers, depending on the agency, have functions including planning, operating and managing water quality, water supply, stormwater management, flood control, and water infrastructure design.

This narrative report provides an overview of the co-production process in Hyperion (Chapter 1), the regional hydro-climatic context and challenges (Chapter 2), broad climate information needs of water management agencies (Chapter 3), and short summaries of the key scientific activities undertaken for the region (Chapter 4). This information is based on the project's co-production engagements and preliminary scientific results. Some of the preliminary results may be updated or refined as they go through the peer-review process. While this report is based on the perspectives of water management agencies that were part of Hyperion, we hope that the insights and methodologies that were developed are broadly applicable to other agencies in the region as well.

1. Co-production in Hyperion

In Hyperion, as far as possible, the research questions, approaches and results were co-produced through regular structured and unstructured engagements between scientists and managers (Figure 1). Structured engagement methods included workshops, remote and in-person focus-group discussions, and quarterly project update calls. There were also continual less-structured, informal conversations over telephone calls and emails.

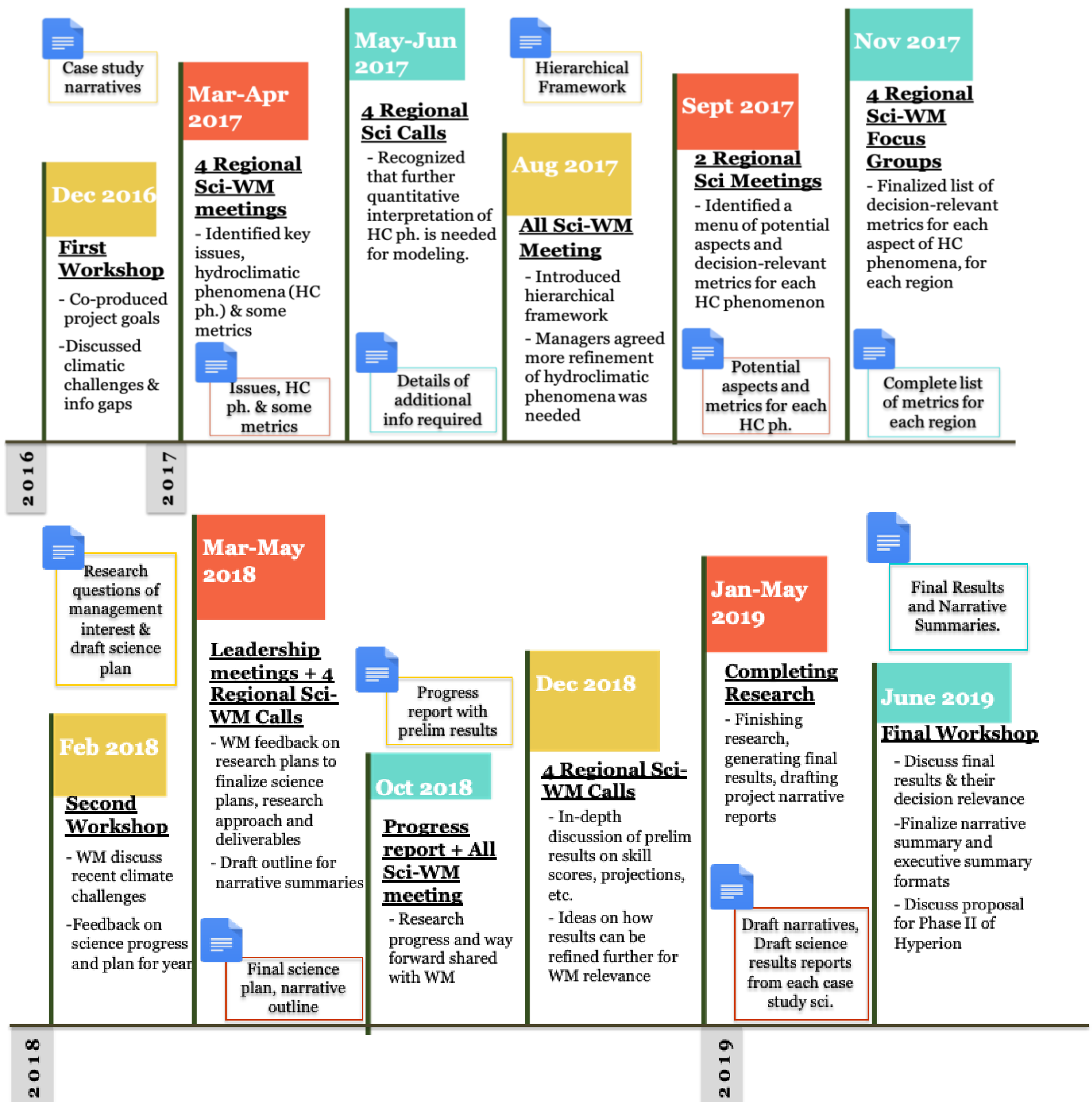


Figure 1: Co-production process and timeline
Summarizes key engagement activities along with important outcomes at each stage (depicted by the blue document icon). 'Sci' refers to Scientists, 'WM' refers to Water Manager and 'HC ph.' refers to Hydroclimatic Phenomena.

2. Regional hydro-climatic context & challenges

The Colorado River is the primary source of water for the Upper Colorado river basin (UCRB). Other important rivers for the state include South Platte, Arkansas, and Rio Grande, which interact with UCRB as far as water rights are concerned. In UCRB, streamflow is heavily dependent on snowpack, with a small amount of water coming from summer monsoon rains. The predominant water use of the basin is agriculture, ranching, and hydropower, in addition to other municipal, industrial and environmental uses. Two-thirds of water that originates in Colorado is committed by interstate law and international treaty to other states and Mexico.

Eighty percent of the state's precipitation is stored (as snowpack) in the western part of the state while 90% of the population resides in the eastern side. Colorado's Front Range municipalities and eastern plains agricultural irrigators have been importing water primarily via tunnels, from the UCRB to meet local water supply needs. For example, transbasin diversions account for more than 40% of the water that Denver Water delivers to its customers, while for the organizations belonging to the Front Range Water Council, water diverted from the UCRB is estimated to account for 72% of the water delivered to their customers.

One of the key climate-related challenges facing the region is a less-resilient snowpack, which has impacted the timing and magnitude of streamflow. For example, reduced late-summer flows due to snowpack melt, impacts farming and ranching operations. Extreme events such as droughts and floods are also of concern. Specifically, recent droughts have led to reduced and variable water supply over sustained periods of time. Temperature extremes and associated increases in evapotranspiration (ET) are also key challenges for the agricultural sector and particularly when combined with reduced snowpack.

Climate and hydrological data is used in planning for water dependent systems (e.g., from the Colorado River Forecast Center). The state's 2003 strategic water plan detailing water conservation efforts was modified to address climate change issues. The state's drought and flood response planning efforts also include climate and hydrological data. Some long-term infrastructure plans (such as reservoir operations) use hydrological forecasts in their planning.

In terms of information gaps, uncertainties in modeling the region is a key problem, which is related to large interannual climate variability. Spatial and temporal resolution is also an issue, as precipitation varies sharply over short distances. Responses to events such as El Niño vary regionally due to the influence of topography. Specific impacts of only temperature (independent of precipitation) could be better understood. Improved understanding of projected increases in extreme events is required. Snowpack and soil moisture data for the region are very limited. Better understanding of existing data and models is also required, e.g. which data/models/approaches/products are most appropriate for what specific uses. A complicating factor is that, typical metrics used to characterize drought in the climate literature (and in the Drought Monitor) (e.g. PDSI) are not commonly applied in practice.

3. Climate information needs for water management

3.1. Overview

The key decision-relevant hydroclimate phenomenon for Colorado is Streamflow. Since most of the state relies on water in the rivers, planning models and systems are based on the amount of water that is or will be in the river. Since most water in the rivers comes from snowpack, therefore snowpack dynamics are another key hydroclimate phenomenon of importance. Rainfall metrics, while interesting, are less important than snowpack. Droughts are also important events, and while not as important as streamflow, certain drought metrics can be informative for specific water management decisions.

In terms of temporal scale for planning, stakeholders suggested that they would plan 50 years out into the future +/- 15 years. From a spatial scale, reservoir in-flow estimates are of importance (one example is the Taylor Park reservoir). Sub-basin level spatial information is of interest for certain drought and water metrics, specifically in Upper Colorado, where two regions within close proximity can also differ substantially in their characteristics. In addition, the structure of water rights in the region makes sub-basin level spatial resolution important. In terms of the types of simulations that would be useful, managers acknowledged that currently most planning is done based on simple 50-100 year simulations of the future, but simulations that recreate a worst-case scenario event could be of potential interest to them. However, the limitations of climate models in predicting such events should be made explicit and transparent.

3.2. List of decision-relevant metrics and their importance

In order for science to be actionable, resource managers need information on decision-relevant climatic metrics. Therefore, one of the first goals of Hyperion was to co-produce the decision-relevant metrics for different management decisions in each of the case study regions. From the water managers' perspective, such metrics quantitatively describe climatic phenomena that are directly related to practical management problems; changes in these quantities would necessitate shifts in water infrastructure planning and operations. From the scientists' perspective, these metrics can be used to test model fidelity for decision-relevant phenomena and hence push model development and scientific inquiry in more use-inspired directions. Table 1 represents the decision relevant metrics, along with their potential importance, that were developed through iterative engagements between Dec 2016 to Nov 2017. This table is referred from the published journal article titled "The making of a metric: Co-producing decision-relevant climate science" by Jagannathan, Jones and Ray.[1]

Table 1: Examples of decision-relevant metrics for each region
The table highlights management issues, hydroclimatic phenomena, aspect of phenomena and then each decision-relevant metric. The last column also describes some of the potential decisions or uses for these metrics that were identified by the case study water managers.

Issue	Hydroclimatic Phenomenon	Aspect of Phenomenon	Decision-relevant Metric	Decision/Use
Water Supply	Streamflow	Seasonal Streamflow amount (in snowmelt season)	Cumulative run-off from April-July (Reservoir in-flow estimates)	Water supply planning in terms of forecasted water volumes.
Water Supply	Streamflow	Seasonal Streamflow amount (in snowmelt season)	Cumulative run-off on July 1 and August 1	Annual water supply planning for the year done based on July 1 or August 1 reservoir level estimates (depending on the reservoir).
Floods	Streamflow	Seasonal Streamflow amount (in snowmelt season)	% of average annual inflow for Apr-July	Reservoir management. This metric is an input into some reservoir operations models.
Water Supply	Streamflow	Inter-annual variability in summer streamflow	10^{th}, 50^{th} and 90^{th} percentile volumes of Apr-May-June-July sum, over several years	Understanding range of possibilities while planning for water supply on timescales of 1 or more years.
Water Supply	Streamflow	Low-end Streamflow	7-day 10 year low flows	Water quality management (issuing discharge permits), and water supply planning during dry years (determining permit limits for water withdrawals).
Water Supply	Streamflow	Low-end Streamflow	Bottom 10^{th} or 25^{th} percentile volumes of streamflow (especially in dry years)	Water supply planning during dry years (or droughts).
Water Supply	Streamflow	Streamflow Timing	Center of mass	Water supply planning for time-sensitive uses (crops, timed water diversions, water in ecosystems), and reservoir operations management.

Issue	Hydroclimatic Phenomenon	Aspect of Phenomenon	Decision-relevant Metric	Decision/Use
Water Supply	Streamflow	Streamflow Timing	Quartiles: Day on which 25%, 50%, and 75% flow (beginning at start of water year) has passed	Water supply planning for time-sensitive uses (crops, timed water diversions, water ecosystems), and reservoir operations management.
Floods	Streamflow	Peakflow	Intensity Duration and Frequency curves for different flow events	Storm water management & design, and flood protection.
Floods	Streamflow	Variability of Streamflow	Probabilities of exceedance for certain run-off thresholds (ASPE design thresholds, which vary by region and infrastructure type)	Storm water management & design, and flood protection.
Water Supply	Snowpack	Seasonal snowpack volume	Monthly snow water equivalent: Total Water availability metric on a monthly scale. (especially Dec-Jul)	Understanding streamflow characteristics and the state of the water system - and its potential to meet demands. E.g. Winter months SWE (Dec-Jan-Feb) is used in seasonal forecasting of reservoir levels on July 1.
Water Supply	Snowpack	Seasonal snowpack volume	% of annual snowpack accumulated in different months	Seasonal reservoir in-flow forecasting and understanding expectations with regard to early versus late season snow accumulation.
Water Supply	Snowpack	Annual cycle of snow accumulation and melt	SWE triangle[1] - Peak snow (amount and timing), and its relationship with average snow-accumulation and -melt rates, and timing and length of accumulation and melt seasons	On-stream reservoir management, and to understand future streamflow characteristics. Shape of the triangle helps better understand the changing dynamics of the snow season, and what to expect in terms of runoff timing and amounts.

Issue	Hydroclimatic Phenomenon	Aspect of Phenomenon	Decision-relevant Metric	Decision/Use
Water Supply	Snowpack	Inter-annual Variability in Snowpack	Deviations from historical mean in monthly SWE	Water supply planning, especially for supply restrictions.
Water Supply	Snowpack	Inter-annual Variability in Snowpack	Upper and lower end of distribution: 10^{th} and 90^{th} percentile of annual snowpack for wet and dry years (or top and bottom 25 pc)	Water supply planning, especially for supply restrictions.
Water Supply	Droughts/Dry Spells	Low precipitation	Standardized Precipitation Index (SPI): Number of dry or wet years	Drought planning and understanding dry spells.
Water Supply	Droughts/Dry Spell	Net water availability	SPEI (Standardised Precipitation-Evapotranspiration Index)	Drought planning and understanding dry spells.

4. Key scientific activities and results from Hyperion

From the above long list of decision-relevant metrics, project Hyperion's managers and scientists collectively developed case study science plans that identified a shorter list of scientific activities and metrics that will be a focus of the project (Table A1 in the Appendix).

Out of this long list of decision-relevant metrics, project Hyperion's stakeholders and scientists collectively decided to focus on the following metrics and scientific activities, as outlined in the case study science plans. These key scientific activities are as follows: Snow Water Equivalent (SWE), North American Monsoons, and Water Supply. The rest of this section presents a narrative description of these three short-listed scientific activities. The key motivation, methods, results and limitations from each of the three scientific activities, are summarised below.

4.1. SWE triangle

Summary

- This work examines the skill of various climate model datasets, and analyses future projections for SWE triangle metrics.
- Snow Water Equivalent (SWE) triangle uses a fitted triangle to characterize the annual cycle of snow accumulation and melt through six metrics of management relevance: peak water volume and timing, snow accumulation and melt rates, and the lengths of the accumulation and melt seasons (Rhoades et al. 2018).[2]
- The study finds that models are better at representing the accumulation portion of the snow season than the melt season: snow melt rate being the most common failure mode across models.
- Average peak SWE volume is expected to dramatically decline by the end of the century, with a general reduction in accumulation rate and accumulation season length. Interestingly, and surprisingly there is a lengthening of the melt season by the end of the century, due to the earlier start of the melt season.

4.1.1. Background and Methods

Mountain snowpack is integral to water supply and security in the Western US and has wide ranging impacts on agricultural productivity, ecological function, hydroelectric power, and tourism. Yet, the simple question of how much snow is in the mountains still can't be easily answered. Although many techniques have been developed to produce snowpack estimates, there is very limited inter-comparison of these techniques, particularly for decision-relevant metrics that focus on water resource management. This has resulted in a large degree of uncertainty in both historical and future snowpack estimates, which has constrained both model development and decisions on water management based on climate change. A critical examination of the different observational and modeled snowpack datasets can provide a better understanding of the relative credibility of these datasets. Further, an analysis of future projections in relation to the relative skill of the different modeled datasets can provide a better understanding of the different uncertainties and reasons for these uncertainties in estimates of future snow predictions. In this way, the evaluation of snowpack can also act as a great litmus test for climate model performance as it requires appropriate representation of both temperature and precipitation, particularly with elevation. The key research questions addressed in this work are as follows:

1. ***What are snowpack related metrics that are both decision-relevant for water resources management and tractable from a climate model development perspective?***
2. ***How well are different snowpack processes and SWE metrics represented across different snowpack datasets (both observed and modeled), and what is their relative credibility?***

3. ***What do the different climate models say about snowpack decline in the future and does model choice, resolution, and boundary forcing, matter? Do these dynamically downscaled results differ from those derived from statistical downscaling methods?***

As decision-relevant SWE metrics, six simple snowpack metrics (known as the SWE triangle) (Refer to Figure A1 in the Appendix) were developed that linearize the major components of the snow season and can be easily applied across any gridded snow product. The goal of the SWE triangle metric framework was to find a middle ground between usefulness in evaluating model performance and distilling the management relevant points in the life cycle of snowpack. After identification of SWE metrics, an extensive intercomparison of publicly available snowpack products (that had data on daily snow water equivalent across at least 15-20 years) was conducted using a z-score analysis. The products that were evaluated included observationally constrained, model derived (i.e., bounded by atmospheric reanalysis and global climate model data), and statistically derived (i.e., observations update global climate model results) datasets. A detailed description of the products examined are provided in Appendix 1 - Table A2. Particular focus was also provided on evaluating SWE triangle metrics for snow products in regions that are more relevant to water resource management (e.g., upstream regions of major reservoirs and/or regions central to water conveyance networks; Refer to Figure A2 in the Appendix). If a Z score is 0, the data set mean is exactly the same as the observed Sierra Nevada Snow Reanalysis (SNSR) data, and if the Z score is positive (negative) the data set mean is higher (lower) than SNSR. If a Z score falls outside of the range of 2 to −2 the given data set's mean is substantially different than the observed.

4.1.2. Key results

Figure 2 (and Figures A3 and A4) provides detailed skill scores for the different observational and modeled snow datasets, for decision-relevant SWE triangle metrics including Snowpack accumulation rate (SAR), Total water volume at peak accumulation (TWV), SWE peak accumulation date (SPD), Snowpack melt rate (SMR), length of the accumulation season (AS), and length of the melt season (MS). In CO, we found that the observationally constrained snow products were in agreement with each other. This was interesting considering that in CA, there was a 2x difference in peak SWE volume among some of the observational products. Overall, in CO there was a general agreement across the observationally constrained snow products, and regions, for peak SWE timing and accumulation and melt season length. Model products that were bounded by atmospheric reanalysis data were generally low biased, for all SWE triangle metrics, highlighting the major role that atmospheric boundary conditions play in the simulation of snowpack. SWE triangle metric performance and interannual variability in SWE is much more varied when assessing model products bounded by global climate model data. In general, across model products and regions, the snow melt rate was too fast.

Across all regions assessed, a high-emissions scenario results in a large decline in average peak SWE volume by the end of the century (Figure 3). There is a general reduction in both accumulation rate and accumulation season length, likely due to changes in precipitation phase from snowfall to rainfall as a result of changes in surface temperature. Melt season length shows a general lengthening, which likely related to the interaction between the average elevation and surface temperature of the region assessed (number of days at or below freezing) and an earlier peak SWE timing (shorter days lead to less available energy to melt the snow). It is to be noted that for the future projections, an ensemble of the evaluated models were used and no skill-based model weighting was conducted.

Lake Powell		SAR	TWV	SMR	SPD	AS	MS
L15 (reference)		0.00	0.00	0.00	0.00	0.00	0.00
NLDAS_2_SAC	*Observationally constrained*	-0.53	-0.94	-1.46	-1.06	-1.11	0.51
NLDAS_2_VIC		-0.05	-0.36	0.43	-0.83	-0.92	-0.73
ECMWF_ERAINT_CRCM5 at 12km	*RCM forced by atmospheric reanalysis*	0.10	-0.13	0.71	-0.66	-0.85	-0.53
ECMWF_ERAINT_CRCM5 at 25km		-0.12	-0.46	1.29	-0.91	-1.18	-0.97
ECMWF_ERAINT_CRCM5 at 50km		0.04	-0.41	1.56	-1.11	-1.28	-1.24
ECMWF_ERAINT_CanRCM4 at 25km		0.14	-0.01	1.63	-0.53	-0.72	-1.16
ECMWF_ERAINT_HIRHAM5 at 50km		-0.25	-1.10	-0.46	-2.08	-2.23	-0.56
ECMWF_ERAINT_WRF at 12km		-0.14	-0.69	-0.37	-1.31	-1.16	-0.45
ECMWF_ERAINT_WRF at 25km		0.29	-0.30	0.55	-1.50	-1.31	-0.73
ECMWF_ERAINT_WRF at 50km		0.16	-0.44	1.30	-1.57	-1.56	-1.36
CanESM2_CanRCM4 at 50 km	*RCM forced by GCM*	0.51	0.24	3.05	-0.75	-1.00	-1.27
CanESM2_CRCM5 at 50 km		0.99	1.16	6.71	-0.08	-0.41	-1.72
MPI_ESM_CRCM5 at 50 km		2.57	3.59	9.25	0.59	-0.23	-1.47
MPI_ESM_RegCM4 at 50 km		7.64	10.34	20.45	0.67	-0.20	-1.23
GFDL_WRF at 50 km		2.57	2.61	5.47	-0.63	-1.06	-0.86
GFDL_WRF at 25 km		2.56	2.59	4.33	-0.61	-0.98	-0.39
HadGEM2_WRF at 50 km		0.35	-0.13	2.27	-1.27	-1.51	-1.47
HadGEM2_WRF at 25 km		0.49	-0.01	1.34	-1.32	-1.42	-0.86
VR28_EPAC_141 at 25 km	*Variable-Resolution Global Climate Model*	1.00	0.65	6.79	-1.28	-1.87	-1.83
VR28_EPAC_94 at 25 km		1.08	0.70	5.92	-1.07	-1.62	-1.96
VR28_EPAC_47 at 25 km		0.84	0.23	4.15	-1.43	-1.95	-1.59
VR28_NATL_EXT at 25 km		0.36	-0.34	12.26	-1.66	-2.59	-2.34
VR28_NATL_REF at 25 km		0.17	-0.37	5.17	-1.45	-2.30	-2.36
VR28_NATL_WAT at 25 km		0.38	-0.44	9.85	-2.23	-2.95	-2.29

Figure 2: Z scores for SWE triangle metrics for the Lake Powell region
Z scores for the six SWE triangle metrics evaluated for the Lake Powell upstream analysis region. The metrics include snowpack accumulation rate (SAR), total water volume at peak accumulation (TWV), snowpack peak accumulation date (SPD), snowpack melt rate (SMR), the length of the accumulation season (AS), and the length of the melt season (MS). The Z score is computed by using the mean and standard deviation from SNSR. Red (blue) indicates positive (negative) Z score bias, and saturation indicates the magnitude of bias. Text color is used to distinguish resolution in (b) and global climate model forcing data set in (c). SWE = snow water equivalent.

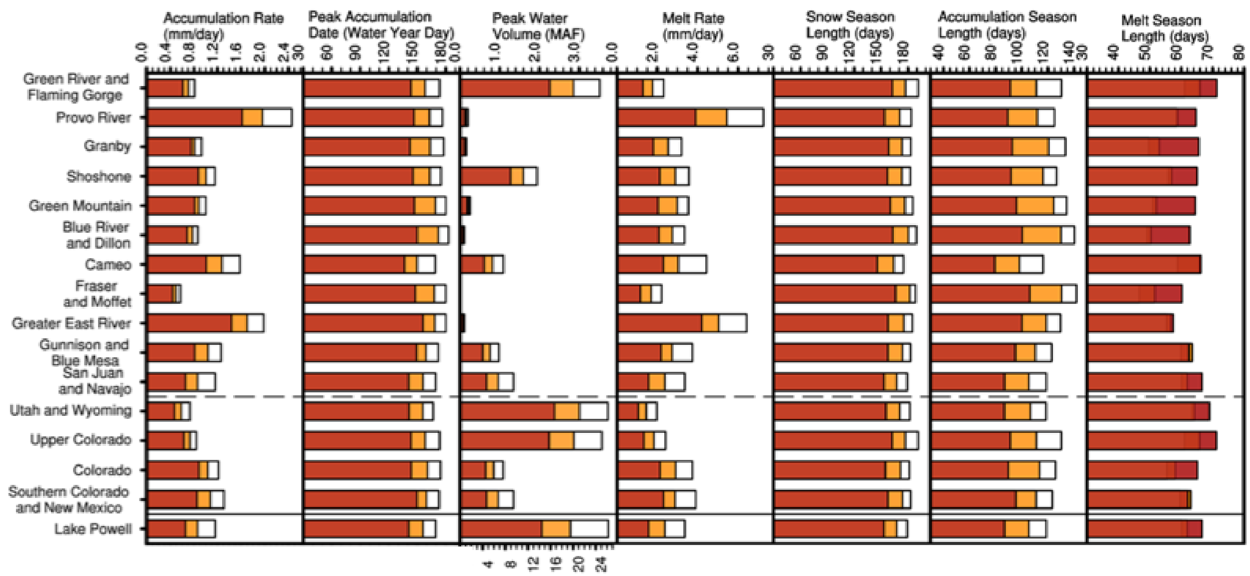


Figure 3: Snow water equivalent triangle metrics for key watershed regions in Colorado
Color is used to distinguish 1985–2005 (white), 2039–2059 (orange), and 2079–2099 (red). Future conditions in these simulations assume a high-emission scenario (or RCP8.5). For Peak Water Volume the top x axis is used for individual regions and the bottom x axis is used for aggregate regions. MAF = million acre-feet.

4.1.3. Discussion and Conclusions

Identifying exactly how much snowpack is in the mountains is a messy business given its multi-scale dependencies. In general, models appear to better represent the accumulation portion of the snow season than the melt season. In particular, snow melt rate appears to be the most common failure mode across models, with melt rates generally too fast. Snow melt rate issues may be model resolution dependent, as this study sampled an uneven number of simulations from 50 km (8 simulations), 25 km (5 simulations), and 12 km (2 simulations). Further research is needed to explore this thoroughly, particularly through filling in gaps in the simulation matrix (e.g., use more atmospheric reanalysis boundary forcings and more cross-resolution simulations) produced by important coordinated regional downscaling efforts such as the North American Coordinated Regional Downscaling Experiment (NA-CORDEX). In addition, a framework to isolate the relative contributions of snowpack simulation error associated with inaccuracies in precipitation, surface temperature, topography, etc. was developed for California.[3] This framework could be used in Colorado as well and enables model developers to better isolate cause and effect in model bias and inform stakeholders on why a given model got the right/wrong answer for the right/wrong reasons. Although not done here, the combination of the aforementioned methods (i.e., Rhoades et al., 2018 and Xu et al., 2019) can be used to understand the confidence in model sub-selection or model weighting for the region. Until a more detailed model sub-selection analysis is available, this multi-metric, multi-dataset SWE triangle evaluation framework samples across many different datasets and uses an ensemble of models rather than choosing a specific skill-based subset. More research is needed on different ways to use the skill evaluation results to inform future projections.

Overall, snowpack is highly sensitive to surface temperature (and its impact on precipitation phasing) and most climate reports are virtually certain that surface temperatures have and will continue to increase heading into the next century. Hence, there is reasonable confidence in these projected changes in peak SWE volume by mid- and end-of-century, if the high emission scenario comes to fruition.

4.2. North American Monsoons (NAMS)

Summary

- This study examines the uncertainty within different observational datasets for the timing of NAMS. It also evaluates how well CMIP5 and CORDEX models predict NAMS timing.
- Different observational datasets differ by about a week or so, on average for onset dates for NAMS.
- Individual models may have large biases in their simulation of onset dates. In the overall ensemble of models (both CMIP5 and CORDX), onset occurs earlier in the southern part and later in the northern part.
- Higher resolution did not lead to smaller biases in onset calculations.

4.2.1. Background and Methods

The North American Monsoon or NAMS contributes to over 30% of annual precipitation in the U.S. Southwest, and more than 40-50% in central and northwestern Mexico. The monsoon is important for the agricultural and the fire seasons (as it typically marks the end of fire season). However, there is limited understanding of the timing of the NAMS. These limitations extend both to the observations of NAMS as well as to the climate models that predict NAMS. This is because there are not many credible monsoon timing algorithms or metrics that can be applied to both observational data sets and models. This means that there is not enough understanding of the observational uncertainty of the NAMS timing. For example, the National Weather Service considers onset to occur after three consecutive days of daily average dewpoint temperature above a threshold. However, if models display a substantial bias, such metrics based on absolute values will not accurately depict onset. Further, there is limited knowledge about how well models represent the NAMS timing. It is hypothesized that, because of the importance of the Gulf of California and the Sierra Madre Occidental to NAMS formation, models with higher horizontal resolution (finer grid spacing) should better simulate the NAMS and its timing. However, this hypothesis has not been tested. Therefore, this study asks the following questions:

1. ***What is the observational uncertainty of North American monsoon (NAMS) onset and how does it compare to the CMIP5 and CORDEX model spread? (This is done by developing monsoon timing metrics that can be applied to both observational data sets and models)***
2. ***How well do models represent NAMS timing?***

For the observational analysis, daily precipitation data from three observational datasets and two reanalyses for the years 1981-2016 were collected. These were from Climate Prediction Center (CPC), Global Precipitation Climatology Center (GPCC), Tropical Rainfall Measuring Mission (TRMM) (data starts in 1998), NCEP North American Regional Reanalysis (NARR), and the ECMWF ERA-Interim Reanalysis (ERAI). For the model evaluations, daily precipitation data from 12 models of the Coupled Model Intercomparison Project version 5 (CMIP5) data set (historical (1985-2004) and RCP 8.5 (2041-2060 and 2080-2099) was examined. In addition, select CORDEX models forced by ERA-Interim (1989-2008) with ~50 km and ~25 km was examined. Two NAMS timing calculation methods were used:

- The Liebmann et al. (2008)[4] method (modified) computes onset by summing the daily precipitation anomaly from the long-term annual daily mean precipitation, beginning during the dry season (the month with the lowest precipitation during the annual cycle at that grid cell). The date on which this sum is a minimum is the date of onset, while the date of the maximum sum marks the rainy season withdrawal (when the monsoon rains end). This method is both objective and defined locally, that is, based on the climate of the area of interest. Onset must occur by October 27. However, the date of onset can vary slightly depending on the start date of the summation.
- The Grantz et al. (2007) method[5] (modified) computes the day on which the daily precipitation reaches 10% of the monsoon precipitation (count begins on May 1st).

4.2.2. Key results

The Liebmann et al. (2008) and Grantz et al. (2007) onset methods produce similar average onset days in the southern and central NAMS region, but in the northern monsoon region the Liebmann method has onset occurring later, whereas the Grantz method has earlier onsets at higher latitudes. Figure 4 (left) shows that observed uncertainty (from Liebmann method) in monsoon dates is typically about a week or so between data sets compared to a mean of the three purely observation-based datasets (TRMM, CPC, and GPCC). NARR is similar to the other observational data sets while ERA-Interim has an early onset in the southern and central parts of the monsoon and later onset in the northern reaches of the monsoon region. The ERA-Interim behavior is similar to CMIP5. Onset dates vary more than withdrawal dates (the date on which the monsoon rains end, shown on the right); onset occurs gradually from south to north, whereas monsoon withdrawal happens more suddenly across all latitudes of the NAMS region. Interannual variability is higher in onset than withdrawal (not shown).

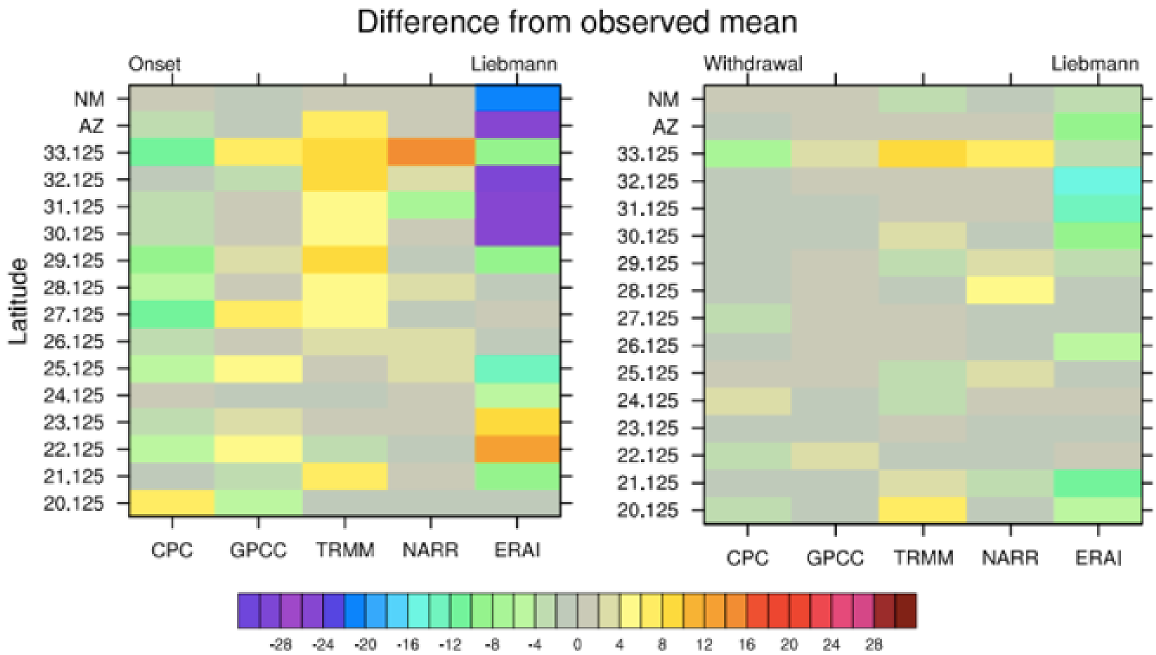


Figure 4: Biases in onset and withdrawal dates of NAMs in observed datasets
Biases in onset (left) and withdrawal (right) dates calculated using the Liebmann method, compared to a mean of three observed data sets (CPC, GPCC, and TRMM) for three different observed data sets and two reanalyses. The dates are shown for a transect progressing from south to north, as well as for the two US states of Arizona and New Mexico. Latitude or state is shown on the left, and different data sets are shown on the bottom. Warm colors indicate that the data set shows a later date than observed, and cool colors indicate an earlier date than observed. The biases are small for the three observed data sets and the NARR (about a week for onset), but the ERA-Interim reanalysis shows a substantially later onset and withdrawal than the observations. Differences in onset are larger because onset occurs gradually throughout the monsoon region, progressing from south to north, whereas monsoon withdrawal tends to occur all at once across the region.

Figure 5 shows that the historical CMIP5 simulations simulate an early onset in the southern monsoon region and a late monsoon onset in the northern region (similar to ERA-Interim). The CMIP5 model spread is much wider than the observational spread. Figure 6 shows that the ERA-Interim-driven CORDEX runs also (for the most part) indicate an early onset in the south and a late onset in the north. The regional models seem to retain the biases from the ERA-Interim. Despite their finer horizontal resolution, the CORDEX models have larger biases in onset date compared to the CMIP5 models. This was unexpected due to the use of “perfect” reanalysis boundary conditions and the higher horizontal grid spacing, which allows the models to have a better representation of the Gulf of California and the Sierra Madre Occidental. Regional models inherit biases from their boundary conditions. Typically, reanalysis offers “perfect” boundary conditions that are as close to observations as one can get in a three-dimensional consistent, gridded framework. Using reanalysis as boundary conditions therefore gives us an idea of the overall skill of the regional climate model.
In terms of future projections, there are very small changes in calculated onset date in the future CMIP5 simulations, but these differences are smaller than the model spread. The large errors

make it difficult for future scenarios to have any credibility when it comes to changes in NAMS onset timing.

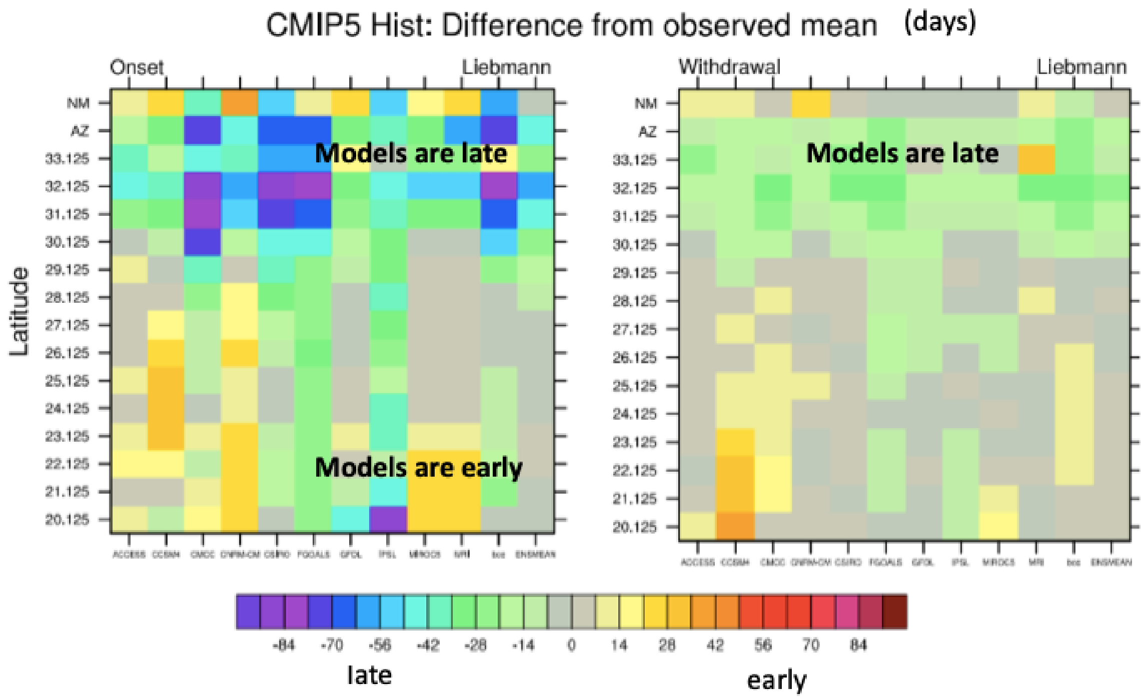


Figure 5: Biases in onset and withdrawal dates of NAMs in CMIP5 models
Biases in onset (left) and withdrawal (right) dates calculated using the Liebmann method, compared to a mean of three observed data sets (CPC, GPCC, and TRMM) for 12 different CMIP5 models. The dates are shown for a transect progressing from south to north, as well as for the two US states of Arizona and New Mexico. Latitude or state is shown on the left, and different data sets are shown on the bottom. Warm colors indicate that data set shows a later date than observed, and cool colors indicate an earlier date than observed. The models tend to simulate an early monsoon in the southern monsoon region, and a late monsoon in the northern region. Thus, it takes too long for the monsoon to penetrate into the northern part of the region. Withdrawal is also late in the northern part of the monsoon region, although the bias is smaller than the onset date, resulting in a shorter-than-observed monsoon.

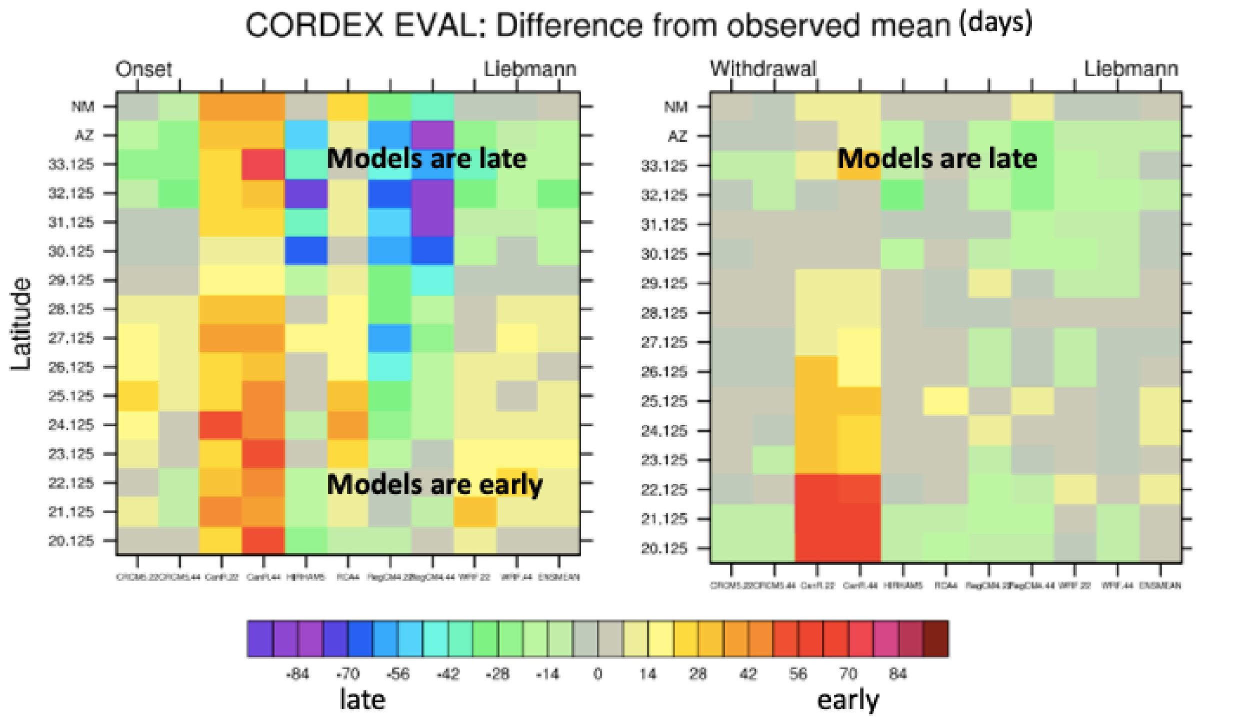


Figure 6: Biases in onset and withdrawal dates of NAMs in CORDEX models
Biases in onset (left) and withdrawal (right) dates calculated using the Liebmann method, compared to a mean of three observed data sets (CPC, GPCC, and TRMM) for 11 different CORDEX regional models. The dates are shown for a transect progressing from south to north, as well as for the two US states of Arizona and New Mexico. Latitude or state is shown on the left, and different data sets are shown on the bottom. Warm colors indicate that the data set shows a later date than observed, and cool colors indicate an earlier date than observed. The regional models behave somewhat similarly to the global models, although there is more variability amongst the models. The biases are still fairly large, however, so increased resolution does not seem to improve the simulation of monsoon timing.

4.2.3. Discussion and Conclusions

Overall, different observation datasets differ by about a week or so on average, for onset dates. Individual models may have large biases in their simulation of onset dates. In fact, observational uncertainty is much smaller than the CMIP5 and CORDEX model spreads. Models simulate an annual cycle with a summer peak in precipitation that is lower than observed. In the overall ensemble of models, onset occurs earlier in the southern part and later in the northern part. Particularly in the northern part of the monsoon region, the onset occurs later by more than a month on average, in some models. This occurs in both the global and regional models, as well as the ERA-Interim reanalysis.

The tendency for the models and ERA-Interim to have early onset in the south and late onset in the north needs to be explored to understand the reason for this happening. In the case of regional models, the bias could be due to the boundary forcing. The regional model result is disappointing as they are often expected to do a better job with monsoon timing than the global models. It is seen that finer horizontal grid spacing (even with “perfect” boundary conditions from reanalysis) does not lead to smaller biases in onset calculations, thereby countering the “higher resolution is always better” narrative. Many of the regional or global NAMS simulations lack a Gulf of California, which therefore limits how well they can represent gulf surges which are often critical to monsoon onset and wet phases, which represents a key gap. But higher resolution does not seem to help. More work is needed to understand whether this is due to the boundary conditions, or some other reason. Understanding what forcing (e.g. MJO, Gulf surges, soil moisture, snow cover) causes changes in monsoon timing should help elucidate model issues.

4.3. Decadal water supply

Summary

- The study examines whether a credible technique for decadal forecast of the Colorado River water supply can be developed. The study also tests an earth system model's capability in predicting the total soil water and whether it tracks the Colorado River's water supply.
- A statistical method in predicting the Colorado River's water supply out to 10 years was developed.
- Overall the results show that predicting the Colorado River's water supply up to 10 years is possible, supported by skillful prediction of the Great Salt Lake level that shares a high level of coherence with the Colorado River's water supply.

4.3.1. Background and Methods

Multi-year droughts impact water management in and around the Colorado River. Therefore, being able to predict the water supply for the next 5-10 years is critical. However, there are not many such decadal projections. It is hypothesized that based on the known climate cycle and the low-frequency variability of water storage in the Intermountain Region, decadal prediction for water supply is feasible. This hypothesis is based on prior studies from this team, which has developed forecasting techniques to predict the lake levels of the Great Salt Lake. These forecasting techniques have demonstrably been able to predict the Great Salt Lake's low-frequency fluctuation since 2009. The studies also showed that the Great Salt Lake (GSL) can be a proxy for depicting the Colorado water supply. Therefore, this study asks the question:

1. ***Taking guidance from the Great Salt Lake forecasting techniques, can a credible technique for "decadal forecast" for the Colorado River water supply be developed? The study also tests an earth system model's capability in predicting the total soil water and whether it tracks the Colorado water supply.***

In order to conduct this work, water storage and water supply, consisting of all liquid water above bedrock (groundwater, soil moisture, streamflow, reservoirs etc.) were examined from many datasets. A statistical model that was developed based on regressive methods and the GSL water level data, could predict the Colorado water supply. The skill of GCM CESMv1, in addition to many reanalysis and hydrological datasets of decadal prediction for Colorado water supply were also examined. An annual timescale and Intermountain region domain (Great Basin and Colorado Basins) were considered, to examine the change in water storage/supply of the

Colorado River. In addition to water supply prediction, this work is being expanded through a collaboration with Prof. Yoshi Chikamoto (USU) and Ruby Leung (PNNL) who are working to provide a streamflow forecast utilizing the CESM model output.

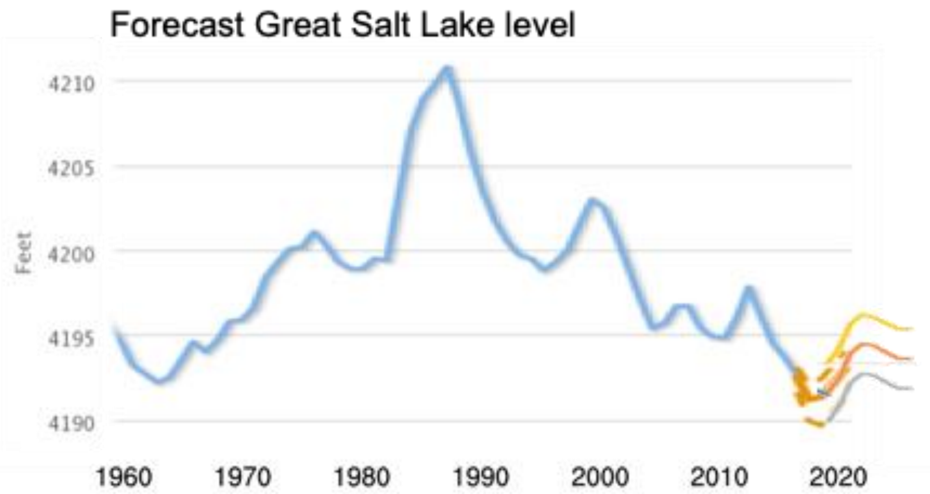
4.3.2. Key results

A statistical method in predicting the Colorado water supply out to 10 years was developed (Refer lower-panel graph in Figure 7, Plucinski et al., 2019).[6] Figure 7 further confirms that the two time series i.e. the Great Salt Lake on top and the smoothed water supply in the bottom, track each other nicely. Their forecasts are agreeable, suggesting an upturn around 2020. Although it is observed that the winter snowpack is much above normal and the river flow should respond to it, the models predicted this upturn 5 years earlier. Predictions from the model show that Colorado water supply did not go up as the projections from a 2012 Bureau of Reclamation's (BOR) report shows, but rather, went down and will go up by 2020. This multi-year upturn and downturn was not revealed by the CMIP5-derived projection but is depicted by the decadal forecast.

As a next important step, the study evaluated a climate model “dynamical” forecast. Figure 8 shows that VR-CESM appears to predict the upturn in total soil water in the Colorado Basin, consistent with the statistical forecast. Overall, the statistical model developed is able to simulate the surface and subsurface features, though the model does not perform well in representing atmospheric teleconnection leading to shifted jet stream and/or storm tracks. Therefore, model prediction might not be able to represent the year-to-year variation.

4.3.3. Discussion and Conclusions

This study shows that predicting the Colorado water supply up to 10 years is possible, supported by skillful prediction of the Great Salt Lake level that shares a high level of coherence with the Colorado water supply. Water supply is an integrated metric and modeling the decadal variability (like the Bureau of Reclamation does) allows for multi-year prediction of water supply which can potentially be applied to other regions. Further engagement in data assimilation techniques can help to enable such decadal predictions. Future research needs to integrate hydrological models with climate models while also working on increasing climate models’ resolution.



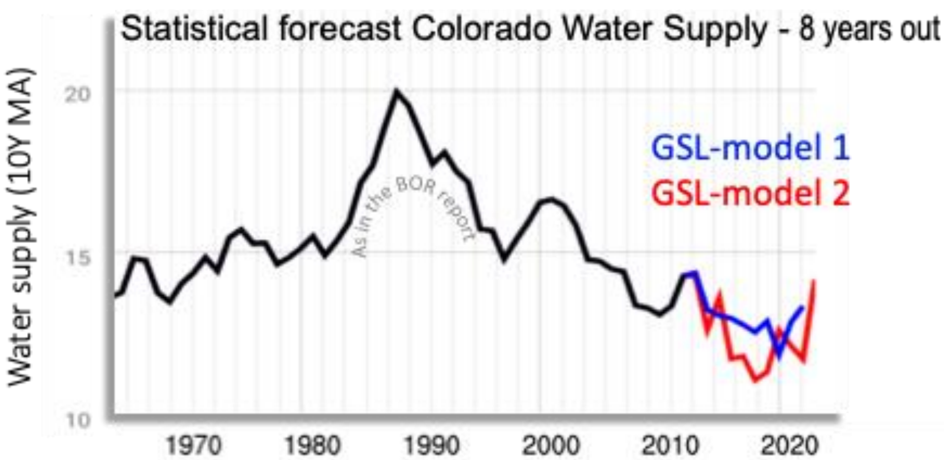


Figure 7: Dynamical model forecast of GSL versus statistical model forecast of Colorado Water Supply
Time-series of Dynamical model forecast of Great Salt Lake (GSL) level (upper panel) and observed values (black line) with statistical models forecast (blue and red) of Colorado Water Supply (lower panel). The black line in lower panel is from the Bureau of Reclamation (BOR) report that released a comprehensive Colorado River basin water supply and demand study in 2013.

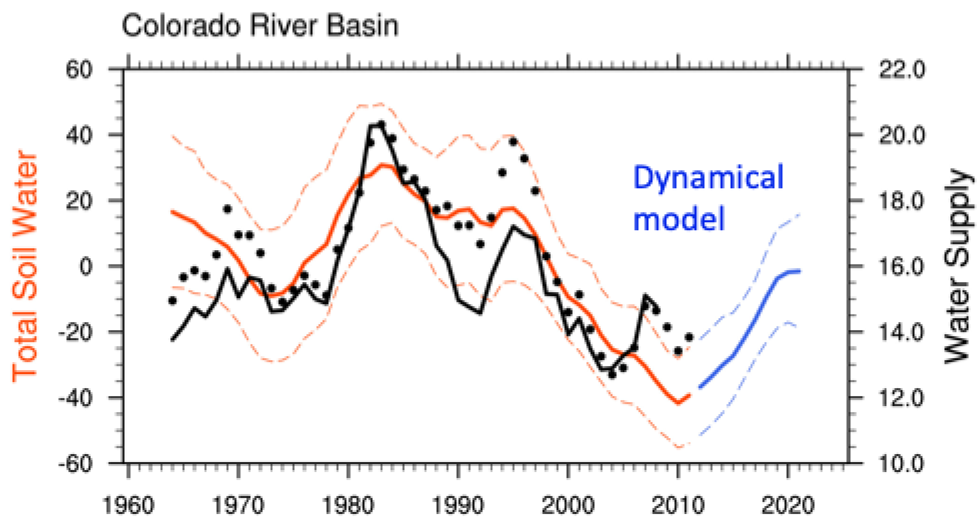


Figure 8: Total Soil Water in Colorado Basin
Time-series of total soil water (solid orange) with 95% confidence bounds (dashed orange) in the Colorado Basin with values in left y-axis. Similarly, right y-axis shows the time-series of Colorado River water supply from observation (black dots) and from statistical model prediction (black line). Blue lines (solid average and dashed 95% confidence bounds) show the VR-CESM prediction of total soil water in the Colorado Basin. Since total soil water of the Colorado River Basin tracks the Colorado water supply, soil water from dynamical model can be used to predict the Colorado water supply.

Acknowledgements and way forward

We are deeply grateful to all of Project Hyperion's water managers and scientists who patiently participated in the many back-and-forth engagements that form the basis of this report. We are also thankful to Bruce Riordan who co-led the engagements, Paul Ullrich for his agile leadership of the project, and Smitha Buddhavarapu for her careful review and edits of this report.
Hyperion's successor project "HyperFACETS" is currently underway (2019-present) and will expand on the project's research activities and further work on creating broadly applicable tools for co-producing actionable climate science.

This work was supported by the Office of Science, Office of Biological and Environmental Research, Climate and Environmental Science Division, of the U.S. Department of Energy under contract DE-AC02-05CH11231 as part of the Hyperion Project, An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US (award DE-SC0016605).

Appendix 1

Table A1: List of metrics and summary of scientific activities pursued by Hyperion project

S No.	Science Activity	Lead Scientists	Description of Research & Papers/Conference Abstracts based on the work
1.	Observational uncertainties and model skill for predicting North American Monsoon timings	Sara Rauscher	This study examines the uncertainty within different observational datasets for the timing of NAMS. It also evaluates how well CMIP5 and CORDEX models predict the NAMS timing. Different observational datasets differ by about a week or so, on average for onset dates for NAMS. Individual models may have large biases in their simulation of onset dates. In the overall ensemble of models (both CMIP5 and CORDX), onset occurs earlier in the southern part and later in the northern part. Higher resolution did not lead to smaller biases in onset calculations. **Papers:** ● Yadav, P., S. A. Rauscher, and D. E. Veron (2019) The Influence of the Madden-Julian Oscillation on the North American Monsoon onset and its variability. https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/598454 ● Yadav, P., S. A. Rauscher, and D. E. Veron (2019) The Influence of the Madden-Julian Oscillation on the North American Monsoon onset and its variability. Manuscript in preparation. ● Rauscher, S. A., P. Yadav, D. E. Veron, E. Aiken, and E. Hoeflich (2019) How well do models simulate the timing of the North American Monsoon? Manuscript in preparation.
2.	Snow Water Equivalent (SWE) triangle: Model skill and future projections	Alan Rhoades	This work examines the skill of various climate model datasets, and analyses future projections for SWE triangle metrics. The study finds that models are better at representing the accumulation portion of the snow season than the melt season: snow melt rate being the most common failure mode across models. Average peak SWE volume is expected to dramatically decline by the end of the century, with a general reduction in accumulation rate, accumulation season length. Interestingly, and surprisingly there is a lengthening of the melt season by the end of the century due to the earlier start to the melt season.
3.	Decadal variability and predictability in water supply	Simon Wang	The study examines whether a credible technique for decadal forecast of the Colorado River water supply be developed. The study also tests an earth

S No.	Science Activity	Lead Scientists	Description of Research & Papers/Conference Abstracts based on the work
			system model's capability in predicting the total soil water and whether it tracks the Colorado water supply. A statistical method in predicting the Colorado water supply out to 10 years was developed. Overall the results show that predicting the Colorado water supply up to 10 years is possible, supported by skillful prediction of the Great Salt Lake level that shares a high level of coherence with the Colorado water supply. **Papers:** ● Wang, S.-Y., R. R. Gillies, O.-Y. Chung, and C. Shen, 2018: Cross-Basin Decadal Climate Regime connecting the Colorado River and the Great Salt Lake. Journal of Hydrometeorology, DOI:10.1175/JHM-D-17-0081.1 ● Plucinski, B., Y. Sun, **S.-Y. Wang**, R. R. Gillies, **J. Eklund,** and C.-C. Wang, 2019: Feasibility of Multi-Year Forecast for the Colorado River Water Supply: Time Series Modeling. Water, DOI:10.3390/w11122433 (this paper is co-authored with one of our water managers - James Eklund) ● Chikamoto et al., 2019: Assessing the multi-year predictability of Colorado River water supply using a drift-free decadal climate prediction system. In preparation.
4.	Variable resolution domain sensitivity experiments	Alan Rhoades	Variable-resolution global climate models are a new means by which to provide dynamically downscaled climate data and are currently being vetted for resolution, refinement domain size, and parameterization-based sensitivities. This study assesses the role of refinement domain size over the North Pacific Ocean, particularly longitudinal extent, in shaping variable-resolution in the Community Earth System Model (VR-CESM) simulations of winter (DJF) hydroclimate of the western U.S. through modifications in dynamical and/or thermodynamical drivers. The study finds that there is minimal impact of refinement domain size on model fidelity in representing western U.S. hydroclimate and show that topographic resolution and land-surface model choice have a greater influence. **Papers:** ● AGU2018 abstract: https://ui.adsabs.harvard.edu/abs/2018AGUFM.A23N3111R/abstract ● Rhoades, Alan M., et al. "Influences of North Pacific Ocean Domain Extent on the Western

S No.	Science Activity	Lead Scientists	Description of Research & Papers/Conference Abstracts based on the work
			US Winter Hydroclimatology in Variable-Resolution CESM." *Journal of Geophysical Research: Atmospheres* 125.14 (2020): e2019JD031977.

Table A2: List of datasets evaluated for SWE triangle metrics

Downscaling method	Dataset name	Product summary	Resolution (evaluated at 12km)	Time period(s) assessed
Observationally constrained snow products	Livneh, 2015 (L15) **(reference dataset for all regions)**	In-situ observations, NCEP reanalysis meteorological data, MT-CLIM based spatial interpolation, and bounded simulations of VIC land-surface model. Estimates updated with PRISM normals.	6km – 1 product	1985-2005
	North American Land Data Assimilation System, phase 2 (NLDAS-2)	Land-surface models bounded by NCEP reanalysis meteorological data	14km – 2 products	1985-2005
Dynamical Downscaling	North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX)	Five regional climate models bounded by ERA-interim atmospheric reanalysis data.	50km – 3 products 25km – 3 products 12km – 2 products	1985-2005
	NA-CORDEX	Six regional climate models bounded by five global climate model datasets.	50km – 7 products 25km – 2 products	1985-2005 2039-2059 2079-2099
Variable-Resolution Global Climate Model	Variable-Resolution in the Community Earth System Model (VR-CESM)	Global climate model simulation (atmosphere-land coupling) with monthly prescribed sea-ice and sea-surface temperatures.	50km – 1 product 25km – 3 products 12km – 1 product 6km – 1 product	1985-2005 2000-2015

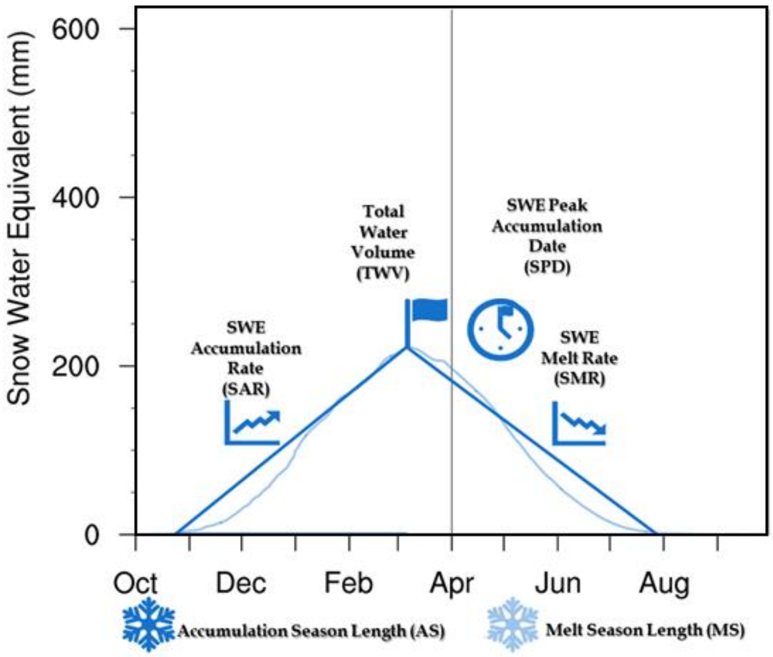


Figure A1: Snow water equivalent (SWE) triangle metrics.
The six snow water equivalent (SWE) triangle metrics visually represented and overlaid on top of the observationally constrained Livneh, 2015 dataset historical average snowpack life cycle for the California Sierra Nevada.

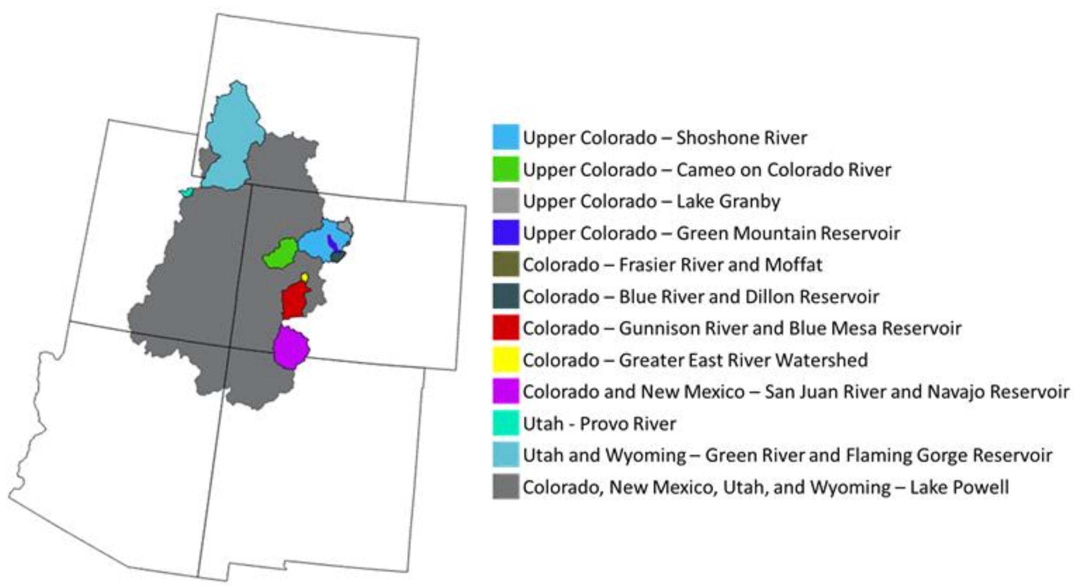


Figure A2: The Colorado regions that are relevant to water management operations and that were used to evaluate SWE datasets.

Shoshone		SAR	TWV	SMR	SPD	AS	MS
L15 (reference)		0.00	0.00	0.00	0.00	0.00	0.00
NLDAS_2_SAC	*Observationally constrained*	-0.63	-0.87	-1.02	-1.23	-1.39	1.09
NLDAS_2_VIC		0.30	-0.19	0.13	-1.45	-1.52	-0.58
ECMWF_ERAINT_CRCM5 at 12km	*RCM forced by atmospheric reanalysis*	-0.61	-0.80	-0.16	-1.12	-1.53	-0.79
ECMWF_ERAINT_CRCM5 at 25km		-1.08	-1.25	-0.41	-1.53	-1.73	-1.12
ECMWF_ERAINT_CRCM5 at 50km		-1.40	-1.57	-0.75	-1.91	-2.17	-1.38
ECMWF_ERAINT_CanRCM4 at 25km		-0.49	-0.76	-0.21	-1.33	-1.42	-0.74
ECMWF_ERAINT_HIRHAM5 at 50km		-1.44	-2.02	-1.02	-4.32	-4.65	-0.94
ECMWF_ERAINT_WRF at 12km		0.21	-0.59	-0.65	-2.64	-2.55	0.37
ECMWF_ERAINT_WRF at 25km		0.07	-0.64	-0.38	-2.47	-2.31	-0.44
ECMWF_ERAINT_WRF at 50km		-0.26	-0.86	-0.08	-2.43	-2.18	-1.39
CanESM2_CanRCM4 at 50 km	*RCM forced by GCM*	-0.82	-0.88	0.57	-0.77	-0.95	-2.14
CanESM2_CRCM5 at 50 km		-0.90	-0.91	0.82	-0.58	-1.10	-2.30
MPI_ESM_CRCM5 at 50 km		0.22	0.18	1.49	0.00	-0.77	-1.75
MPI_ESM_RegCM4 at 50 km		2.06	1.85	2.66	0.13	-0.58	-1.01
GFDL_WRF at 50 km		1.40	0.96	2.28	-0.64	-1.11	-1.45
GFDL_WRF at 25 km		1.92	1.35	1.82	-0.71	-1.28	-0.66
HadGEM2_WRF at 50 km		-0.25	-0.77	0.32	-2.07	-2.33	-1.71
HadGEM2_WRF at 25 km		0.15	-0.48	-0.02	-2.08	-2.22	-0.84
VR28_EPAC_141 at 25 km	*Variable-Resolution Global Climate Model*	1.07	0.13	3.69	-2.36	-2.97	-2.78
VR28_EPAC_94 at 25 km		1.11	0.20	4.47	-2.20	-3.13	-2.82
VR28_EPAC_47 at 25 km		0.39	-0.42	2.93	-2.42	-3.17	-2.56
VR28_NATL_EXT at 25 km		-1.59	-1.97	3.07	-3.38	-4.11	-3.72
VR28_NATL_REF at 25 km		-1.69	-1.97	1.83	-3.23	-4.53	-3.73
VR28_NATL_WAT at 25 km		-1.84	-2.14	1.83	-3.84	-4.77	-3.59

Figure A3: Z scores for SWE triangle metrics for the Shoshone watershed
Z scores for the six SWE triangle metrics evaluated for the Shoshone region. The metrics include snowpack accumulation rate (SAR), total water volume at peak accumulation (TWV), snowpack peak accumulation date (SPD), snowpack melt rate (SMR), the length of the accumulation season (AS), and the length of the melt season (MS). The Z score is computed by using the mean and standard deviation from SNSR. Red (blue) indicates positive (negative) Z score bias, and saturation indicates the magnitude of bias. Text color is used to distinguish resolution in (b) and global climate model forcing data set in (c). SWE = snow water equivalent.

Gunnison River and Blue Mesa		SAR	TWV	SMR	SPD	AS	MS
L15 (reference)	Observationally constrained	0.00	0.00	0.00	0.00	0.00	0.00
NLDAS_2_SAC		-1.51	-1.61	-1.61	-0.56	-0.55	0.60
NLDAS_2_VIC		-0.59	-0.75	-0.33	-0.48	-0.57	-0.42
ECMWF_ERAINT_CRCM5 at 12km	RCM forced by atmospheric reanalysis	-1.21	-1.43	-0.94	-0.88	-0.94	-0.14
ECMWF_ERAINT_CRCM5 at 25km		-1.71	-1.89	-1.20	-0.97	-1.08	-0.32
ECMWF_ERAINT_CRCM5 at 50km		-1.19	-1.45	-0.52	-0.95	-1.17	-0.71
ECMWF_ERAINT_CanRCM4 at 25km		-0.96	-1.07	-0.03	-0.45	-0.51	-0.96
ECMWF_ERAINT_HIRHAM5 at 50km		-1.02	-1.40	-0.92	-1.01	-1.13	-0.11
ECMWF_ERAINT_WRF at 12km		-1.55	-1.78	-0.60	-1.02	-0.95	-0.79
ECMWF_ERAINT_WRF at 25km		-1.00	-1.49	-1.02	-1.34	-1.15	-0.35
ECMWF_ERAINT_WRF at 50km		-0.72	-1.30	-0.55	-1.42	-1.40	-0.75
CanESM2_CanRCM4 at 50 km	RCM forced by GCM	-0.46	-0.39	1.89	-0.01	-0.15	-1.45
CanESM2_CRCM5 at 50 km		-0.27	-0.41	1.13	-0.37	-0.69	-0.74
MPI_ESM_CRCM5 at 50 km		1.56	1.90	3.83	0.55	-0.09	-1.12
MPI_ESM_RegCM4 at 50 km		4.14	4.44	5.73	0.47	-0.09	-0.82
GFDL_WRF at 50 km		2.58	1.97	2.32	-0.47	-0.88	-0.42
GFDL_WRF at 25 km		1.94	1.46	1.57	-0.39	-0.71	-0.27
HadGEM2_WRF at 50 km		-0.09	-0.83	0.07	-1.36	-1.48	-0.65
HadGEM2_WRF at 25 km		-0.44	-1.06	-0.51	-1.26	-1.34	-0.32
VR28_EPAC_141 at 25 km	Variable-Resolution Global Climate Model	1.97	0.99	4.04	-1.10	-1.65	-1.29
VR28_EPAC_94 at 25 km		1.93	1.09	4.80	-0.92	-1.56	-1.43
VR28_EPAC_47 at 25 km		1.66	0.72	6.98	-1.07	-1.71	-1.36
VR28_NATL_EXT at 25 km		-0.31	-1.14	9.45	-1.67	-2.40	-2.14
VR28_NATL_REF at 25 km		-0.56	-1.20	4.17	-1.55	-2.46	-2.09
VR28_NATL_WAT at 25 km		-0.50	-1.34	8.62	-1.88	-2.69	-2.13

Figure A4: Z scores for SWE triangle metrics for the Gunnison River and Blue Mesa watershed

Z scores for the six SWE triangle metrics evaluated for the Gunnison River and Blue Mesa watersheds. The metrics include snowpack accumulation rate (SAR), total water volume at peak accumulation (TWV), snowpack peak accumulation date (SPD), snowpack melt rate (SMR), the length of the accumulation season (AS), and the length of the melt season (MS). The Z score is computed by using the mean and standard deviation from SNSR. Red (blue) indicates positive (negative) Z score bias, and saturation indicates the magnitude of bias. Text color is used to distinguish resolution in (b) and global climate model forcing data set in (c). SWE = snow water equivalent.

References

[1] Jagannathan, K., A. D. Jones, and I. Ray, The making of a metric: Co-producing decision-relevant climate science. *Bull. Amer. Meteor. Soc.*, doi: https://doi.org/10.1175/BAMS-D-19-0296.1.

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

[3] Xu, Yun, Andrew Jones, and Alan Rhoades. "A quantitative method to decompose SWE differences between regional climate models and reanalysis datasets." *Scientific reports* 9.1 (2019): 1-11.

[4] Liebmann, B., I. Bladé, N.A. Bond, D. Gochis, D. Allured, and G.T. Bates, 2008: Characteristics of North American Summertime Rainfall with Emphasis on the Monsoon. J. Climate, 21, 1277–1294, https://doi.org/10.1175/2007JCLI1762.1

[5] Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona, 2007: Seasonal Shifts in the North American Monsoon. J. Climate, 20, 1923–1935, https://doi.org/10.1175/JCLI4091.1

[6] Plucinski, Brian, et al. "Feasibility of Multi-Year Forecast for the Colorado River Water Supply: Time Series Modeling." *Water* 11.12 (2019): 2433.