UC Berkeley
UC Berkeley Previously Published Works

Title
Assessing Mountains as Natural Reservoirs With a Multimetric Framework

Permalink

Journal

ISSN

Authors
Rhoades, Alan M
Jones, Andrew D
Ullrich, Paul A

Publication Date

DOI

Peer reviewed

Key Points:
- There is a need to evaluate snowpack data sets beyond conventional techniques such as climate average absolute bias and spatial correlations
- Multimetric evaluations elucidate processes that create disagreement in climate model snowpack estimates
- A consistent high bias in snowmelt rates across climate models limits their utility for water management applications

Supporting Information:
- Supporting Information S1

Correspondence to:
A. M. Rhoades,
arhoades@lbl.gov

Assessing Mountains as Natural Reservoirs With a Multimetric Framework

Alan M. Rhoades[1] , **Andrew D. Jones[1]** , and **Paul A. Ullrich[2]**

[1]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [2]Department of Land, Air, and Water Resources, University of California, Davis, CA, USA

Abstract Anthropogenic climate change will continue to diminish the unique role that mountains perform as natural reservoirs and alter long-held assumptions of water management. Climate models are important tools to help constrain uncertainty and understand processes that shape this decline. To ensure that climate model estimates provide stakeholder relevant information, the formulation of multimetric model evaluation frameworks informed by stakeholder interactions are critical. In this study, we present one such multimetric framework to evaluate snowpack data sets in the California Sierra Nevada: the snow water equivalent (SWE) triangle. SWE triangle metrics help to describe snowpack characteristics associated with total water volume buildup, peak water availability, and the rate of water release. This approach highlights compensating errors that would not be reflected in conventional large-scale spatiotemporal analysis. To test our multimetric evaluation framework, we evaluate several publicly available snow products including the Sierra Nevada Snow Reanalysis, Livneh (L15), and the North American Land Data Assimilation System version 2 data sets. We then evaluate regional climate model skill within the North American Coordinated Regional Climate Downscaling Experiment. All data sets analyzed show variation across the various SWE triangle metrics, even within observationally constrained snow products. This spread was especially shown in spring season melt rates. Melt rate biases were prevalent throughout most regional climate model simulations, regardless of snow accumulation dynamics, and will need to be addressed to improve their utility for water stakeholders.

Plain Language Summary Mountain snowpack is a key natural water reservoir. Due to anthropogenic climate change, this natural reservoir will likely decline over the next century. This decline will be nonlinear in both space and time. As such, climate model estimates will be key in constraining uncertainty surrounding this decline. These virtual laboratories allow us to understand what could be and not just what has been. The utility of climate models is apparent; however, uncertainties surrounding climate model estimates of mountain snowpack need to be understood. This article assesses uncertainties surrounding both observationally based and climate model estimates of mountain snowpack in the California Sierra Nevada. We use a new multimetric evaluation framework called the snow water equivalent (SWE) triangle. This framework elucidates agreement/disagreement in estimates of peak water volume and timing, accumulation, and melt rates, and the lengths of the accumulation and melt seasons. We found that spread across snowpack data sets were partly driven by differences in the accumulation to peak timing phase of the winter season but was dominated by melt season differences. To expand climate model utility for water management applications, more research is needed to understand why models tended to have earlier peak timing and abrupt snowmelt.

1. Introduction

Mountains are unique physical features that act as both natural dams that impede atmospheric moisture transport and natural reservoirs that extract, store, and release water over time. The western U.S. (WUS) water system is largely dependent upon the seasonal accumulation and melt of mountain snowpack (Bales et al., 2006) with three fourths of the WUS water supply originating from snowmelt (Palmer, 1988). For example, in the California Sierra Nevada mountain snowpack was found to contribute 72% additional storage compared with current man-made reservoirs (Dettinger & Anderson, 2015). Thus, water stakeholders are often beholden to the atmospheric, hydrologic, and land surface processes that shape snowpack dynamics to ensure they can meet annual water demands. The interplay between these components of the Earth system coupled with

anthropogenic climate change (Pierce et al., 2008) are leading to earlier peak snowpack timing (e.g., for the California Sierra Nevada 6–21 days earlier; Kapnick & Hall, 2010) evidence of enhanced warming with elevation (Pepin et al., 2015), and a slower snowmelt, which could alter hydrologic basin connectivity and decrease annual mean streamflow (Musselman et al., 2017). Snowpack diminution has been consistent across many of the Mediterranean mountain ranges with sensitivity studies showing varying reduction in mean snow water equivalent, SWE, (10–25% per degree Celsius increase) and snow duration (5 to >25 days per degree Celsius increase) based on the climate characteristics that drive snow-mass-energy balance (Fayad et al., 2017; López-Moreno et al., 2017). In the WUS, the rate of change in mountain snowpack is unprecedented in the paleoclimatic record (Belmecheri et al., 2016; Pederson et al., 2011). All said, changes in snowpack undermine the oft-used stationarity assumption (Milly et al., 2008) that has traditionally guided the design of water systems in the WUS and emphasizes the key role of snow drought (Harpold et al., 2017).

WUS scientists and water stakeholders have a need to understand seasonal snowpack dynamics and how related processes will be affected under anthropogenic climate change. Hossain et al. (2015) conducted a survey of 90 water experts with at least two decades of experience: 71% of respondents felt that conventional water management practices are not adequate to meet 21st century demand, and half of respondents desired more flexible water management strategies and/or better evaluation of the drivers of uncertainty in water resources, especially at the local-to-regional scale. Therefore, as the stationarity assumption in water management becomes less reliable with the onset of climate change, the utilization of models to identify and understand the various processes that shape snowpack supplies will be critical to mitigate future water supply risk. Models can act as virtual laboratories offering process understanding across global, regional, and local scales. This allows hydroclimate scientists and resource managers to pose questions of what could be and not just what has been. However, each model and scale presents its own set of potential biases that can have a cascading effect on modeled snowpack (i.e., an uncertainty chain). For example, at the global scale, the ability of a global climate model to realistically simulate the jet stream, semipermanent high-pressure regions, and atmosphere-ocean-land teleconnections will impact the timing and landfall location of storm systems. This in turn feeds into the regional scale, where land surface cover and topographic representation shape the intensity and phase of regional precipitation. Finally, at the local scale radiative and hydrologic processes shape the residence time of water that is then used within the ecology-energy-food-water nexus.

Evaluating the credibility of simulated climate and snow dynamics through a decision-relevant lens and attributing sources of bias across multiple processes, scales of analysis, and models poses a significant challenge to the scientific community. The WUS is a well-studied region with a vast literature in observed snowpack dynamics (Bales et al., 2006; Cristea et al., 2017; Henn et al., 2016; Kapnick & Hall, 2010; Margulis, Cortés, Girotto, & Durand, 2016; Mote, 2006; Mote et al., 2005; Painter et al., 2016; Serreze et al., 1999; Trujillo & Molotch, 2014) and modeling approaches (Ashfaq et al., 2013, 2016; Giorgi et al., 2009; Huang et al., 2016; Liu et al., 2017; McCrary et al., 2017; Minder et al., 2016; Naz et al., 2016; Pierce & Cayan, 2013; Rasmussen et al., 2014; Rhoades et al., 2016, 2017, 2018; Sun et al., 2016; Wu et al., 2017) dating back decades (Anderson, 1976; Palmer, 1988). Consequently, a staggering growth in publicly available observational and regional climate data sets have been produced, yet a lag in expert direction regarding their applicability for regional planning (Hall, 2014). Thus, a major need is not only climate data that can account for regional drivers that shape planning decisions, but a means of generating and evaluating this data informed through stakeholder interactions (Dilling & Lemos, 2011; Galford et al., 2016; Jones et al., 2016; Lowrey et al., 2009; McNeeley et al., 2012; Moss et al., 2013; Thompson et al., 2013).

One method to directly incorporate and produce iterative science with stakeholders is the use of decision-relevant metrics for model evaluation. Model evaluation metrics are derived quantities that can be computed from both model outputs and from observationally constrained data products in order to statistically characterize the fidelity of models. A decision-relevant metric is such a metric that characterizes some aspect of a physical phenomenon (e.g., seasonal streamflow) that has direct bearing on decisions (e.g., reservoir construction and operation). To date, several studies have used various metrics and methodologies to evaluate climate models at global scales (CCTAG, 2015; Gleckler et al., 2008; Hawkins & Sutton, 2009; Pierce et al., 2009), regional scales (CCTAG, 2015; Pierce et al., 2009; Rupp et al., 2013; Trujillo & Molotch, 2014), and local scales (CCTAG, 2015; Maurer et al., 2010). Many of these studies use large-scale climate phenomena to evaluate model performance, such as averages or extrema in temperature and precipitation, however did not evaluate mountain snowpack or the characteristic of its annual cycle as a model performance benchmark. There have been studies that have specifically assessed mountain snowpack as a climate model performance
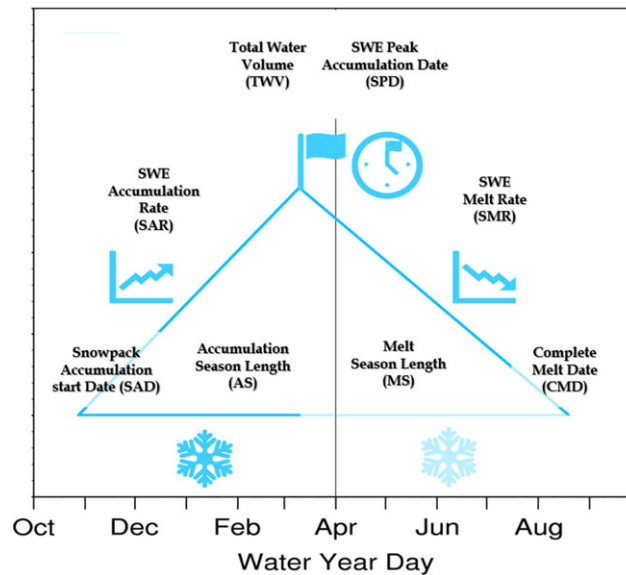
Figure 1. A graphical depiction of the multimetric framework known as the SWE triangle designed to characterize the annual cycle of snow accumulation and melt. Visual identifiers for each of the six SWE triangle metrics are provided next to their name and abbreviation and will be used in subsequent figures for clarity. SWE = snow water equivalent.

benchmark (Chen et al., 2014; Minder et al., 2016; Pavelsky et al., 2011; Rasmussen et al., 2011, 2014); however, it is not the standard in the climate community. Since mountain snowpack is an emergent phenomenon that depends on multiple processes across multiple scales, it provides a unique test bed for model performance that is sensitive to many upstream biases. These processes include topography and land surface cover, the timing, distribution, magnitude, and phase of precipitation and the time-dependent melting phases of snowpack shaped by the components of the radiative budget. Thus, if snowpack dynamics are properly modeled, it may infer confidence that other covariables and processes are well represented too. However, the emergent nature of snowpack also means that multiple metrics are required to characterize subprocesses and quantify the degree to which compensating errors create a false impression of model skill.

A comprehensive multimetric evaluation approach for mountain snowpack was done by Trujillo and Molotch (2014) who aimed to understand and contextualize the lifecycle of snowpack from season-to-season and across various regions of the WUS using in situ measurements from SNOTEL stations. Based on this work, it was identified that the snowpack lifecycle in a given water year resembles a triangle that can be described by a period of linear accumulation (left side), a point of maximum accumulation (vertex), a period of snowmelt (right side), and the distance between the two vertices which describe the snow season length (bottom side; Figure 1). Thus, the complex global-regional-local interactions that drive mountain snowpack can be cast into easily digestible linear approximations, the SWE triangle.

The SWE triangle allows one to discern changes in peak water volume and timing, accumulation and melt rates, and the lengths of the accumulation and melt seasons. This type of multimetric evaluation technique is useful to WUS regional water planners who monitor snowpack reserves throughout the winter and spring season to balance competing interests such as maintaining high reservoir levels during peak demand, lower reservoir levels for flood risk attenuation, reservoir temperature, and release times for species habitats and constant hydroelectric base loads to meet energy demand (Bartos & Chester, 2015; Pagán et al., 2016; Tarroja et al., 2016). Moreover, within a modeling framework, the SWE triangle can provide insight into which phenomena are not well represented by models. For instance, errors in spring melt rate may point to radiation and/or temperature biases whereas changes in peak SWE point to errors in precipitation magnitude and/or phase.

The premise of this study is to evaluate SWE biases across a wide range of data sets using a multimetric framework that characterizes mountains as natural reservoirs and helps to elucidate water management relevant topics such as seasonal water storage, storage residence time, and interannual variability. We evaluate the SWE triangle multimetric framework in the California Sierra Nevada. We leverage and expand upon the initial

concept of Trujillo and Molotch (2014) to analyze SWE biases across a myriad of publicly available gridded SWE data products. Since spatially continuous SWE observations are themselves uncertain and often derived various in situ and/or remote sensing measurements, we include several observationally constrained snow products in our comparison (Livneh et al., 2013, 2015; Margulis, Cortés, Girotto, & Durand, 2016; Xia, Mitchell, Ek, Cosgrove, et al., 2012; Xia, Mitchell, Ek, Sheffield, et al., 2012). It is understood that these four data sets use distinctly different techniques to estimate SWE in both space and time and are broadly lumped together by the term observationally constrained snow product. We do this to simplify discussion and identify that all four of these data sets used some form of meteorological observations that either directly or indirectly influenced the eventual SWE estimate. For example, the data sets derived from Xia, Mitchell, Ek, Cosgrove, et al. (2012) and Xia, Mitchell, Ek, Sheffield, et al. (2012) employ a common atmospheric reanalysis data set to bound several land surface models that then produce SWE estimates. In the SWE data set produced by Livneh et al. (2015), atmospheric conditions are derived from in situ observations and interpolated in space to then drive a stand-alone land surface model that estimates SWE. Last, Margulis, Cortés, Girotto, and Durand 2016) blend a combination of satellite observations and Bayesian inference techniques to produce estimates of SWE. More details about how these SWE data sets were generated are presented in the methods section. By leveraging the North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX; Mearns, 2017), we can compare regional climate models (RCMs) forced with identical atmospheric reanalysis data sets and/or global climate model data sets to gain insight into errors arising from the representation of global-to-regional atmosphere processes and how this influences SWE at the land surface. We start by comparing models based on seasonal mean SWE accumulation, then break the differences down using the SWE triangle approach to gain more process-oriented and decision-relevant insights into the nature of these errors. By doing this, we effectively sample along an uncertainty chain that allows us to

1. identify the range of uncertainty across observationally constrained and RCM SWE products;
2. assess the ability of regional downscaling methods to properly characterize seasonal SWE water storage, residence time, and interannual variability relative to observationally constrained snow products; and
3. assess the components of the SWE triangle to understand what aspects of the model snowpack dynamics lead to agreement/disagreement.

It is our hope that using this decision-relevant multimetric framework will lead to new insights in model evaluation, thus improving the usability of scientific data products for decision-making and offering process-level insight into model representations that affect SWE dynamics.

2. Experimental Design

2.1. Multimetric Framework

A set of metrics were created to characterize key features of the snow mass-energy balance annual cycle, similar to those discussed in Anderson (1976) and Trujillo and Molotch (2014). Figure 1 visually depicts the multimetric framework used in this study including the snowpack accumulation start date (SAD), the snowpack accumulation rate (mm/day, SAR), the snowpack accumulation peak date (SPD), total water volume at this date (million acre feet, TWV), the snowpack melt rate (mm/day, SMR), the complete melt date (CMD), and the length of the accumulation and melt season (days, AS and MS). These metrics were chosen not only for their physical significance but also for their potential relevance to water management stakeholders. Thus, they describe characteristics such as the total water volume buildup (SAD, SAR, and SPD), maximum water volume available for storage and/or use in summer months (TWV), the rate of water release (SMR, CMD, and MS), and the time available for water accumulation as input into downstream reservoirs (AS and MS). Taken as a whole, the metrics describe a triangle whose base is equivalent to the total snow season length (AS plus MS) and whose peak is defined by TWV and SPD. Hereafter, we refer to the group of metrics as the SWE triangle. A formal description of the thresholds used for each metric are presented in Table 1. Of note, we evaluated several strategies for determining the beginning and end of the snow season (SAD and CMD) including absolute SWE depth thresholds ranging from 1.0 to 40 mm and relative thresholds ranging from 5% to 30% of max SWE for a given year. We also evaluated the use of SWE depth percentiles across a given water year (i.e., lower decile to lower tercile). For each threshold considered, we calculated a metric that describes the degree of divergence between the SWE triangle linear approximation and actual SWE depth data from the benchmark Sierra Nevada Snow Reanalysis (SNSR) data set, that is, the absolute sum differences between the actual snow data and the SWE triangle approximation. We found that this metric of divergence was minimized by using either a relative threshold of 10% of max SWE or an absolute threshold of 20 mm. We chose to apply the 10%

Table 1
A Description of the Metrics Used to Formulate the Components of the Snow Water Equivalent (SWE) Triangle

Metric	Units	Assessment thresholds
SWE demarcation dates		
Snowpack accumulation start date (SAD)	date	day when SWE $\geq$10% of maximum SWE
Snowpack peak accumulation date (SPD)	Julian water year day	day of maximum SWE
Complete melt date (CMD)	date	day when SWE $\leq$10% of maximum SWE
SWE triangle metrics		
Snowpack accumulation rate (SAR)	mm/day	$\dfrac{SWE\ depth\ at\ SPD}{day\ of\ SPD - day\ of\ SAD}$
Total water volume at peak accumulation (TWV)	million acre feet	areal average SWE at SPD $\times$ region area
Snowpack melt rate (SMR)	mm/day	$\dfrac{SWE\ depth\ after\ SPD - SWE\ depth\ at\ SPD}{day\ of\ CMD - day\ of\ SPD}$
The length of the accumulation season (AS)	days	sum of days from SAD to SPD
The length of the melt season (MS)	days	sum of days from SPD to CMD

relative threshold across all data sets. Therefore, the SAD is defined as the first date at which 10% of max SWE occurs prior to max SWE and CMD is the first date at which 10% of max SWE occurs after max SWE.

2.2. SWE Data Sets

We consider two primary types of gridded SWE data sets in this study: observationally constrained snow products and RCM simulations. Observationally constrained snow products are generated using a variety of methods to derive SWE but are bounded by hydrometeorological observations, whereas RCM simulations estimate SWE based on internal representations of regional atmospheric dynamics driven by larger-scale meteorological boundary conditions. The RCMs can be forced by either large-scale atmospheric reanalysis conditions or conditions drawn from global climate models (GCMs). RCMs driven by historical reanalysis allow us to evaluate how well the RCMs recreate the observed climate, whereas RCMs driven by historical GCM simulations allow us to identify compounding errors associated with the coupling of the GCM and RCM. These compounding errors become important when assessing future projections of climate change. A summary of the data sets considered and their primary sources of uncertainty can be found in Figure 2. By contrasting the skill across this large ensemble of data sets, we can gain insight into their sources of bias. For instance, by contrasting the performance of the same RCM forced with atmospheric reanalysis versus a GCM, we can learn something about the bias introduced by that GCM.

We have chosen the Landsat-ERA SNSR 90-m data set by Margulis, Cortés, Girotto, and Durand (2016) as the benchmark data set for evaluating the skill of model simulations as well as additional observationally based snow products. SNSR characterizes SWE across 20 watersheds of the California Sierra Nevada from 1985 to 2015. The SWE estimates are derived from a Bayesian data assimilation technique that uses 30-m elevation estimates from the Advanced Spaceborne Thermal Emission and Reflection Radiometer and the National Land Cover Database, hourly 14-km meteorological inputs from the North American Land and Data Assimilation Database phase 2 (NLDAS-2), and snow cover area and vegetation cover fractions derived from the NASA Landsat 5, 7, and 8 satellite data. For quality assessment, SNSR was compared with 108 snow pillows and 202 snow course in situ measurements and found to have correlations of 0.97 throughout most of the sites (Margulis, Cortés Girotto, & Durand, 2016; Margulis, Cortés, Girotto, Huning et al., 2016).

We leverage two of the four NLDAS-2 reanalysis data sets in our SWE evaluation. The two NLDAS-2 simulations that were used include the Sacramento Soil Moisture Accounting (SAC) model (NLDAS_2_SAC) and the Princeton version of the Variable Infiltration Capacity (VIC) Land Model (NLDAS_2_VIC) as they provided SWE at daily temporal resolution. Each of the NLDAS-2 data sets were generated using atmosphere prescribed observational/reanalysis data sets (i.e., the National Centers for Environmental Prediction North American Regional Reanalysis) that bounded land surface models run over the continental United States (CONUS) at 14-km resolution from 1985 to 2014 (Xia, Mitchell, Ek, Cosgrove, et al., 2012; Xia, Mitchell, Ek, Sheffield, et al., 2012).
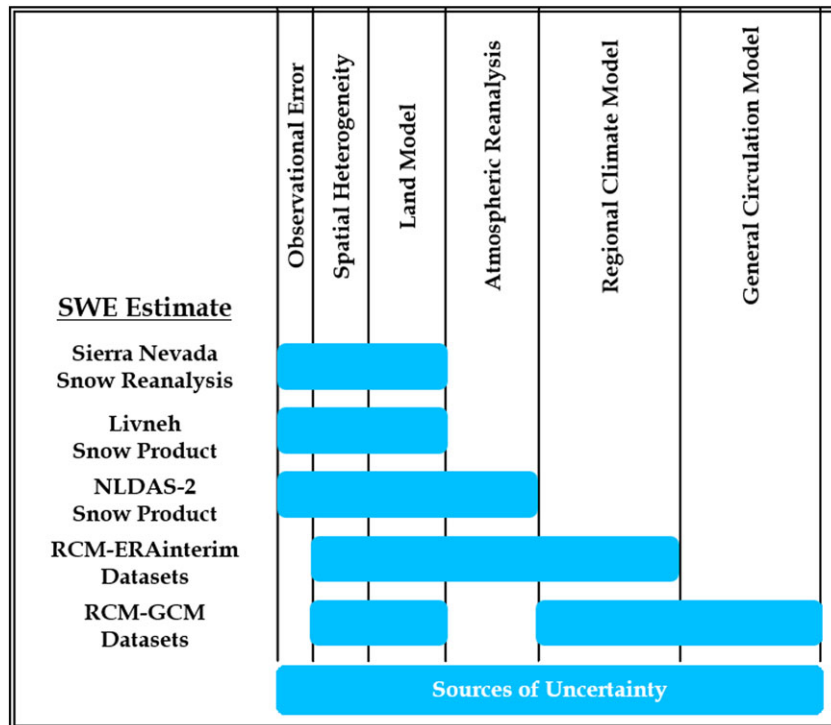
Figure 2. A conceptual diagram of the hierarchy of uncertainty in predictability of SWE across several publicly available products. SWE = snow water equivalent; NLDAS-2 = North American Land and Data Assimilation Database phase 2; RCM = regional climate model; GCM = global climate model.

The most recently updated Livneh gridded reanalysis data set (L15) is derived from 21,137 observational stations that span Canada, CONUS, and Mexico and have been spatially interpolated to 6 km resolution over 1950–2013 (Livneh et al., 2013, 2015). The quality-controlled spatial interpolation is updated using atmospheric variables from National Centers for Environmental Prediction-National Center for Atmospheric Research reanalysis and augmented with orographic scaling procedures (i.e., lapse rate and orographic precipitation), which is then used to bound simulations using the VIC land surface model.

The NA-CORDEX ensemble is composed of six GCMs and one reanalysis data set that drive six RCMs over the CONUS at 50 and 25 km with prescribed sea surface temperatures and sea ice fractions as boundary conditions (Mearns, 2017). In addition, one simulation from the new regional downscaling ensemble Framework for Assessing Climate's Energy-Water-Land Nexus using Targeted Simulations (FACETS) was used. FACETS builds upon the successes of NA-CORDEX with a new set of GCM-RCM coupled simulations over the CONUS at even higher resolutions (e.g., 12 km) and with several new variable resolution global climate models. Table 2 provides a list of the simulations we downloaded and assessed. All of the completed simulations from NA-CORDEX/FACETS are shown. Emboldened resolutions were used in this analysis. The set of prerequisites for using an NA-CORDEX/FACETS member in this evaluation was that it had to have outputted SWE at daily temporal resolution and spanned at least 15 years between a common overlapping time frame of 1985–2005. Therefore, 16 of the NA-CORDEX members and one of the FACETS members were used. Three European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim (ECMWF_ERAINT) forced simulations spanned less than 1985–2005. HIRHAM5 at 50-km and CanRCM4 at 25-km simulations were run from 1989 to 2005 and WRF at 12 km were run at 1990–2005, whereas all other data sets spanned 1985–2005. For simplicity, we do not acknowledge the difference between CRCM5 and MPI-ESM simulations in the NA-CORDEX. The CRCM5 simulations used in this study were performed by Université du Québec à Montréal. Also, only RCM simulations forced by the Max-Planck-Institut für Meteorologie Earth System Model running on low resolution grid (MPI-M-MPI-ESM-LR) global model were used, shortened to MPI-ESM hereafter.

2.3. Regions of Analysis

We conduct our analysis over 10 water source regions that supply surface water to 10 major reservoirs in California (Figure 3). These 10 regions represent 12% of California's area (54,000 km^2), yet provide 16 million

Table 2
The North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX) and Framework for Assessing Climate's Energy-Water-Land Nexus using Targeted Simulations (FACETS) Simulations

Regional climate model	Global data set					
	ERAINT	HadGEM2	CanESM2	MPI-ESM	EC-EARTH	GFDL
CRCM5	**50, 25, 12 km**		**50 km**	**50**, 25 **km**		25 km
RCA4	50 km		50 km		50 km	
RegCM4	50, 25 km	50, 25 km		**50**, 25 **km**		50, 25 km
WRF	**50, 25, 12 km**	**50, 25 km**		50, 25 km		**50, 25 km**
CanRCM4	50, **25 km**		**50**, 25 **km**			
HIRHAM5	**50 km**				**50 km**	

Note. Emboldened climate model horizontal resolutions indicate the simulations included in this study as they provided daily snow water equivalent outputs.

acre feet (MAF) of water storage or 40% of the total surface water storage capacity (41 MAF; Mount & Hanak, 2015). Arranged from north to south these reservoirs are Shasta, Oroville, Folsom, New Melones, Don Pedro, Exchequer, Pine Flat, Terminus, Success, and Isabella. Their metadata are presented in supporting information Table S1. The upstream regions that feed surface water to these reservoirs were generated by Teklu Tesfa and Ruby Leung at Pacific Northwest National Laboratories using EPA ecoregion classifications and a hydrologic connectivity routine that ensures all surface water that would historically flow into a given reservoir location is spatially accounted for (Tesfa et al., 2011).

To standardize each of the observationally constrained snow products and RCM data sets for the intercomparison, they have been regridded to a common 12-km resolution using bilinear interpolation via the Earth System Modeling Framework software and masked with the 10 upstream analysis regions. The intercomparison was chosen to span 1985–2005 to encapsulate all overlapping years across data sets and ensure a proper comparison of sampled water years. Twelve kilometers was chosen as the regrid resolution as it is representative
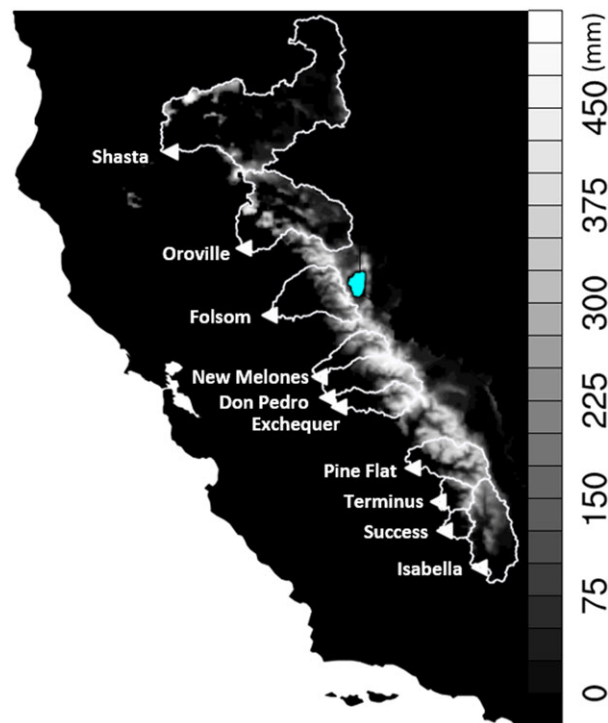

Figure 3. A December-January-February average snapshot of the Sierra Nevada Snow Reanalysis data set, which is used in this study as the standard data set for intercomparison. The 10 upstream analysis regions (major dams of California) are highlighted as white lines (white triangles).

of the finest resolution of the RCMs analyzed in this study. A plot of the December, January, and February (DJF) climate average SWE across each of the data sets used in this study is shown in supporting information Figure S1.

3. Results and Discussion
3.1. Mean Wintertime Snow Biases

We begin our analysis by evaluating the skill of observationally constrained snow products and RCM data sets in capturing the spatial properties of the DJF average SWE across the 10 regions of interest. We use the skill metrics embedded in the Taylor Diagram and encourage readers to reference Taylor (2001) for mathematical details. These include the following:

1. A measure of linear spatial similarity between the standard data set and any other snow product, the Pearson pattern spatial correlation (PPC). In our case, this identifies the DJF seasonal spatial error in SWE across data sets within our analysis regions. A value of 1 indicates no error.
2. The average difference between the standard data set and another snow product, the root-mean-square error (RMSE). This performance measure contextualizes how much error in DJF SWE accumulation there was between data sets from season to season. A value of 0 indicates a perfect match.
3. The ratio of variance (RV) describes the DJF seasonal variability match between the standard data set and another snow product. RV measures the ability of a given SWE product to emulate the SWE DJF spatial variability of the standard data set. A value of 1 identifies a perfect match.
4. In addition to the three Taylor Diagram uncertainty metrics, the average sum bias (ASB) in MAF is also computed for each upstream analysis region to assess the spread in the total amount of SWE. A negative value indicates that snow accumulation was too low, a value of 0 indicates a perfect match, and a positive value indicates that snow accumulation was too high when compared with the standard data set. All computed values are converted from SWE depths (mm) to SWE volumes (MAF).

The interplay between several of the Taylor diagram metrics are summarized in Figure 4 for DJF SWE across all 10 of the upstream analysis regions. Each Taylor Diagram highlights a particular group of data sets that share common sources of uncertainty (see Figure 2), including the observationally constrained snow products (panel a), the regional model simulations forced by ERA-Interim (RCM-ERA Interim; panel b), and the regional model simulations forced by global model data sets (RCM-GCM; panel c). Table 3 quantifies each data sets performance across the uncertainty metrics. The standard for each uncertainty metric and Taylor Diagram is the SNSR data set.

The observationally constrained snow products have the most skill in representing the DJF climate average performance measures across all 10 of the upstream analysis regions (Figure 4a). This agreement presumably stems from the various observationally constrained snow products' use of similar observed meteorology that bound the data set generation process to be more in line with SNSR which uses Landsat satellite data. DJF climate average ASB, RMSE, PPC, and RV ranged from −2.66 to 1.00 MAF, 70.1 to 119 mm, 0.77 to 0.90, and 1.07 to 2.21, respectively. Across all uncertainty measures, the L15 data set is most similar to the SNSR data set across the 10 upstream analysis regions with an ASB of 1.00 MAF, a 70.1-mm RMSE, a 0.90 PPC, and an RV of 1.07. The NLDAS_2 reanalysis products have consistently low biased estimates of SWE (ASB and RMSE) and higher winter season spatial variability (RV) with nine (two) of the 1985–2005 winter season RV ≥ 2.50 in NLDAS_2_SAC (NLDAS_2_VIC). To place the spread across reanalysis data sets in a water management context, the ASB range of 1.00 to 2.66 MAF can be interpreted as being 6 to 17% of the 16 MAF of storage capacity of the 10 reservoirs assessed in this study.

We assess the RCM-ERA Interim simulations in Figure 4b. Several RCM-ERA Interim simulations fall within the range of values exhibited by the observationally constrained snow products for DJF climate average ASB, RMSE, and RV. Particularly, ECMWF_ERAINT_CRCM5 and ECMWF_ERAINT_WRF at 12 (25) km whose ASB were low biased at −2.02 (−2.16) and −1.61 (−2.26) MAF, yet less so than both NLDAS_2 reanalysis data sets. In addition, PPC values were constrained between 0.59 and 0.82, which were almost all lower than even the lowest PPC value found across the observationally constrained snow products. This is an indication that although the RCM-ERA Interim simulations relative to SNSR have a similar bias in the total amount of SWE compared with the other observationally constrained snow products, the spatial distribution of SWE (i.e., PPC) is less skillful up to an absolute difference of 0.41. A common trend across all RCM-ERA Interim simulations were their higher RV. This is an indication that the RCMs produce greater seasonal spatial variability in SWE across
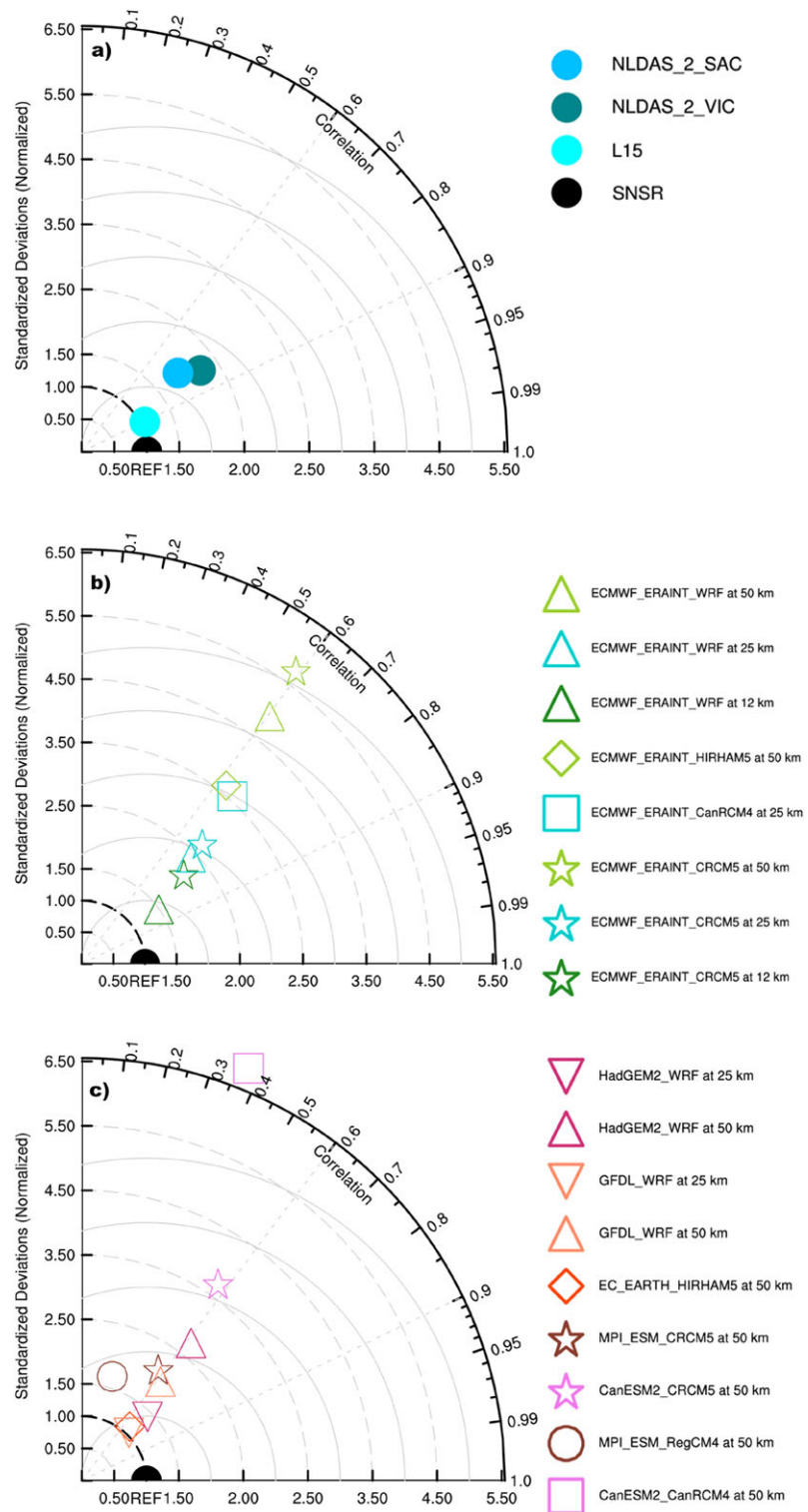
Figure 4. Taylor diagrams of SWE across (a) observationally constrained snow products, (b) regional models forced by ERA-Interim, and (c) regional models forced by global models for all 10 upstream analysis regions. Each symbol represents the 1985–2005 December, January, and February climate average SWE for a given RCM in comparison to SNSR. Colors are used to distinguish resolution in (b) and GCM forcing data set in (c). The *x* and *y* axes designate the ratio of variance; the azimuthal angle along the circular axis connecting the two axes indicates the Pearson pattern correlation, and the root-mean-square error is indicated by the distance away from REF on the *x* axis. A perfect match with SNSR would fall on the *x* axis at REF. SWE = snow water equivalent; RCM = regional climate model; SNSR = Sierra Nevada Snow Reanalysis; NLDAS = North American Land and Data Assimilation Database.

Table 3
SWE Uncertainty Metrics for The DJF Climate Average Across the 10 Upstream Analysis Regions

Data set name	Average sum bias (MAF)	Root-mean-square error (mm)	Pearson pattern correlation	Variance ratio	Number of DJF seasons RV ≥ 2.50
Observationally constrained snow products					
SNSR	0.00	0.00	1.00	1.00	0
L15	1.00	70.1	0.90	1.07	0
NLDAS_2_SAC	−2.66	119	0.82	2.21	9
NLDAS_2_VIC	−2.63	119	0.77	1.91	2
Observationally constrained snow product absolute average	2.09 ± 1.10	103 ± 32.6	0.83 ± 0.08	1.73 ± 0.68	4 ± 5
RCM-ERA Interim					
ECMWF_ERAINT_CRCM5 at 12 km	−2.02	119	0.75	2.14	9
ECMWF_ERAINT_CRCM5 at 25 km	−2.16	129	0.71	2.69	13
ECMWF_ERAINT_CRCM5 at 50 km	−3.22	156	0.59	5.74	20
ECMWF_ERAINT_CanRCM4 at 25 km	−2.97	145	0.67	3.56	13
ECMWF_ERAINT_HIRHAM5 at 50 km	−3.38	151	0.63	3.63	13
ECMWF_ERAINT_WRF at 12 km	−1.61	97.2	0.82	1.49	4
ECMWF_ERAINT_WRF at 25 km	−2.26	127	0.72	2.41	9
ECMWF_ERAINT_WRF at 50 km	−3.31	155	0.60	4.91	20
RCM-ERA Interim absolute average	2.61 ± 0.48	135 ± 14.7	0.69 ± 0.05	3.32 ± 1.02	13 ± 4
RCM-GCM					
CanESM2_CanRCM4 at 50 km	−3.20	163	0.37	6.89	18
MPI_ESM_RegCM4 at 50 km	4.49	185	0.28	1.68	6
CanESM2_CRCM5 at 50 km	−1.56	138	0.57	3.71	15
MPI_ESM_CRCM5 at 50 km	0.45	127	0.57	2.09	9
EC_EARTH_HIRHAM5 at 50 km	0.81	121	0.66	1.12	3
GFDL_WRF at 50 km	0.47	121	0.62	1.96	9
GFDL_WRF at 25 km	2.78	132	0.70	1.04	2
HadGEM2_WRF at 50 km	−1.65	131	0.62	2.72	12
HadGEM2_WRF at 25 km	0.13	106	0.72	1.42	3
RCM-GCM absolute average	1.73 ± 0.99	136 ± 15.9	0.57 ± 0.10	2.51 ± 1.23	9 ± 4

Note. SNSR is used as the standard data set. SWE = snow water equivalent; DJF = December, January, and February; MAF = million acre feet; SNSR = Sierra Nevada Snow Reanalysis; NLDAS = North American Land and Data Assimilation Database; RCM = regional climate model; GCM = global climate model.

the 10 analysis regions. RV absolute differences from SNSR ranged between 0.50 and 4.74. Model horizontal resolution plays a large role in determining RV as it directly relates to model topography which dictates both the magnitude of orographic forcing and the elevation and variability of the freezing isotherm, both of which are key processes in snowpack dynamics. For every 2 times refinement of model horizontal resolution in the RCM-ERA Interim simulations RV decreased by 1.47 and the number of DJF seasons that RV values were ≥2.50 decreased by 6. Similarly, ASB decreased by 0.74 MAF, RMSE decreased by 23.1 mm and PPC increased by 0.09. Although a larger sample across the three model resolutions is desirable (i.e., three simulations were at 50 km, three simulations were at 25 km, and two simulations were at 12 km), this potentially highlights some of the direct benefits of model horizontal resolution. Overall, the ASB range was low biased across all of the RCM-ERA Interim simulations (−1.61 to −3.38 MAF) and translated to 10 to 21% of the storage capacity of the 10 reservoirs assessed in this study. As a group, the five RCM-ERA Interim simulations increased ASB, RMSE, and RV by 0.52 MAF, 32.0 mm, 1.59, and decreased PPC by 0.15 against the observationally constrained snow product data set average.

We assess the RCM-GCM simulations in Figure 4c. As shown previously, model horizontal resolution played a large role in shaping all uncertainty metrics. Unfortunately, the relatively coarse resolution that these simula-

Table 4
SWE Triangle Metrics Average ± Margin of Error (or 95% Confidence Intervals) Across the 10 Upstream Analysis Regions

Data set name	SAR (mm/day)	SPD (water year day)	TWV (MAF)	SMR (mm/day)	AS (days)	MS (days)	Ratio of SAR/SMR	Ratio of AS/MS
Observationally constrained snow products								
SNSR	2.65 ± 0.67	166 ± 11	9.78 ± 1.60	2.29 ± 0.28	96 ± 16	97 ± 11	1.16	0.99
L15	3.15 ± 0.77	154 ± 14	11.2 ± 1.91	3.00 ± 0.49	94 ± 16	90 ± 13	1.05	1.04
NLDAS_2_SAC	1.83 ± 0.43	137 ± 14	4.56 ± 0.83	1.34 ± 0.29	66 ± 13	86 ± 13	1.36	0.77
NLDAS_2_VIC	1.15 ± 0.25	158 ± 15	4.55 ± 0.81	1.11 ± 0.22	98 ± 16	96 ± 12	1.03	1.02
Observationally constrained snow product absolute average	2.19 ± 0.89	154 ± 12	7.51 ± 3.46	1.93 ± 0.87	88 ± 15	92 ± 5	1.15 ± 0.15	0.95 ± 0.12
RCM-ERA Interim								
ECMWF_ERAINT_CRCM5 at 12 km	3.27 ± 0.84	133 ± 14	6.79 ± 1.10	2.80 ± 1.18	60 ± 15	76 ± 16	1.17	0.79
ECMWF_ERAINT_CRCM5 at 25 km	2.93 ± 0.81	126 ± 13	6.24 ± 1.15	4.29 ± 3.09	62 ± 16	69 ± 16	0.68	0.90
ECMWF_ERAINT_CRCM5 at 50 km	2.88 ± 1.27	126 ± 12	4.29 ± 0.69	5.59 ± 4.25	53 ± 13	44 ± 12	0.52	1.21
ECMWF_ERAINT_CanRCM4 at 25 km	2.27 ± 0.71	133 ± 13	4.35 ± 0.72	1.75 ± 0.38	58 ± 14	67 ± 16	1.29	0.87
ECMWF_ERAINT_HIRHAM5 at 50 km	5.65 ± 3.05	121 ± 18	5.16 ± 0.97	6.63 ± 6.71	44 ± 14	48 ± 17	0.85	0.91
ECMWF_ERAINT_WRF at 12 km	3.60 ± 1.21	118 ± 21	6.44 ± 1.33	3.02 ± 1.58	55 ± 20	66 ± 18	1.19	0.84
ECMWF_ERAINT_WRF at 25 km	3.42 ± 1.01	119 ± 13	5.97 ± 0.85	2.85 ± 1.33	52 ± 13	66 ± 13	1.20	0.79
ECMWF_ERAINT_WRF at 50 km	2.33 ± 0.75	119 ± 14	3.94 ± 0.54	3.49 ± 1.99	55 ± 13	46 ± 12	0.67	1.18
RCM-ERA Interim absolute average	3.32 ± 0.76	125 ± 4	5.45 ± 0.83	3.83 ± 1.13	55 ± 4	61 ± 9	0.95 ± 0.21	0.94 ± 0.12
RCM-GCM								
CanESM2_CanRCM4 at 50 km	2.01 ± 0.52	118 ± 14	3.47 ± 0.66	1.66 ± 0.54	51 ± 13	64 ± 14	1.21	0.78
MPI_ESM_RegCM4 at 50 km	5.58 ± 0.71	154 ± 11	19.4 ± 2.92	6.55 ± 1.14	83 ± 13	73 ± 11	0.85	1.14
CanESM2_CRCM5 at 50 km	2.77 ± 0.69	136 ± 14	7.39 ± 1.73	3.06 ± 0.83	71 ± 13	62 ± 10	0.90	1.15
MPI_ESM_CRCM5 at 50 km	3.28 ± 0.96	157 ± 13	11.7 ± 1.88	4.17 ± 0.77	95 ± 13	69 ± 9	0.79	1.39
EC_EARTH_HIRHAM5 at 50 km	4.37 ± 1.01	134 ± 14	12.0 ± 2.27	3.39 ± 0.79	72 ± 14	89 ± 12	1.29	0.81
GFDL_WRF at 50 km	4.43 ± 1.27	145 ± 11	10.9 ± 1.67	4.96 ± 3.76	74 ± 15	84 ± 13	0.89	0.88
GFDL_WRF at 25 km	4.81 ± 0.94	151 ± 10	16.4 ± 2.46	7.71 ± 6.83	85 ± 12	88 ± 13	0.62	0.96
HadGEM2_WRF at 50 km	4.02 ± 1.18	115 ± 10	6.64 ± 1.11	2.11 ± 0.49	49 ± 11	80 ± 12	1.90	0.61
HadGEM2_WRF at 25 km	4.48 ± 1.31	119 ± 11	9.33 ± 1.48	2.64 ± 0.53	57 ± 11	87 ± 11	1.70	0.66
RCM-GCM absolute average	3.97 ± 0.73	136 ± 11	10.8 ± 3.26	4.03 ± 1.36	71 ± 11	77 ± 7	1.13 ± 0.29	0.93 ± 0.17

Note. The six SWE triangle 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). SWE = snow water equivalent; NLDAS = North American Land and Data Assimilation Database; RCM = regional climate model; GCM = global climate model.

tions were run at was found to be a persistent issue: seven of the RCM-GCM simulations were only run at 50 km and two were run at 25 km. However, this is the most up-to-date coordinated multimodel RCM experiment that is publicly available. Interestingly, this did not lead to a systematic low bias in DJF climate average ASB across all simulations. In fact, the RCMs forced by MPI_ESM, EC_EARTH, GFDL, and HadGEM2 had high-biased ASB values of +0.13 to +4.49 MAF when compared with SNSR across the 10 upstream regions. The other two simulations forced by CanESM2 exhibited a low bias between −1.56 and −3.20 MAF, which is still less than or within the range exhibited by the RCM-ERA Interim simulations. The difference in ASB between SNSR and the RCM-GCM simulations was equivalent ≤28% of the storage capacity of the 10 reservoirs assessed. Across the other uncertainty metrics a large discrepancy in simulated DJF climate SWE was found based on which GCM was used to drive the RCM simulations. RMSE values ranged from 106 to 185 mm, PPC spanned 0.28 to 0.72, RV was between 1.04 and 6.89, and the number of DJF seasons whose RV ≥2.50 was between 2 and 18. Across all uncertainty metrics HadGEM2_WRF at 25 km had the closest approximation to SNSR with an ASB of 0.13 MAF, RMSE of 106 mm, PPC of 0.72, and an RV of 1.42 with only three DJF seasons exhibiting RV values ≥2.50. The similarity across uncertainty metrics is quite surprising given the majority of simulations were run at coarse resolution (50 km) and lack of periodic bounding with atmospheric reanalysis forcing. Overall, when
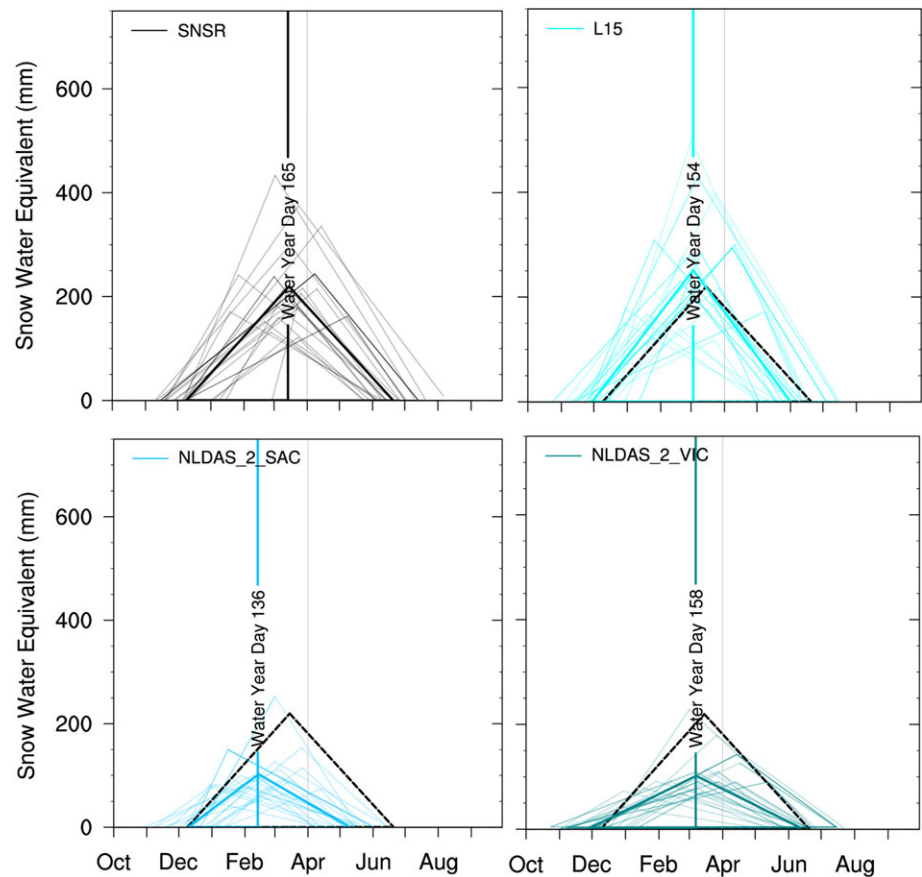
Figure 5. SWE triangles for each of the observationally constrained snow products across the 10 upstream analysis regions. Each SWE triangle represents a linearized daily average approximation of snowpack dynamics in an individual water year within 1985–2005. The 1 April is shown with a vertical gray line to highlight the date of historical peak SWE timing in the western United States. The SWE triangle component average is emboldened and the snowpack peak accumulation date for a given snowpack product is shown with a vertical line. The standard reference data set used for this study (SNSR) is overlaid on all figures with black dashed lines. SWE = snow water equivalent; SNSR = Sierra Nevada Snow Reanalysis.

compared against the observationally constrained snow product average, the nine RCM-GCM simulations increased DJF climate ASB, RMSE, and RV by 0.37 MAF, 33.3 mm, and 0.78, and decreased PPC by 0.27.

3.2. Multimetric Analysis of Subannual Snow Biases

To decipher the snowpack dynamics driving disagreement across data sets in Figure 2, we employ the SWE triangle metrics outlined in Table 4. As mentioned previously, the SWE triangle linearizes aspects of daily snowpack dynamics over a given water year. This allows for an easily decipherable and quantifiable means to intercompare various snowpack products and assess skill.

3.2.1. Observationally Constrained Snow Products

In Figure 5, SWE triangles for the four observationally constrained snow products are depicted for the 1985–2005 water years within the 10 upstream analysis regions. The solid black line at 1 April marks the oft-assumed historical peak accumulation date for SWE in the WUS according to in situ measurements that have been maintained for decades. This date is more a rule-of-thumb than a definitive measurement as shown by Bohr and Aguado (2001) for the Rocky Mountains. The statistics for the average and margin of error about the mean (or 95% confidence intervals) across 1985–2005 for each of the SWE triangle components are given across all SWE data sets in Table 4. As with mean DJF SWE, there is a clear delineation in SWE estimates presented by SNSR and L15 compared with the NLDAS_2 SAC and VIC data sets. These differences are driven by the 2 times difference in SAR and TWV. The maximum (minimum) values in SAR and TWV were shown in L15 (NLDAS_2_VIC) at 3.15 ± 0.77 (1.15 ± 0.25) mm/day and 11.2 ± 1.91 (4.55 ± 0.81) MAF. Although there were significant differences in both SAR and TWV, SMR showed even more disagreement across reanalysis
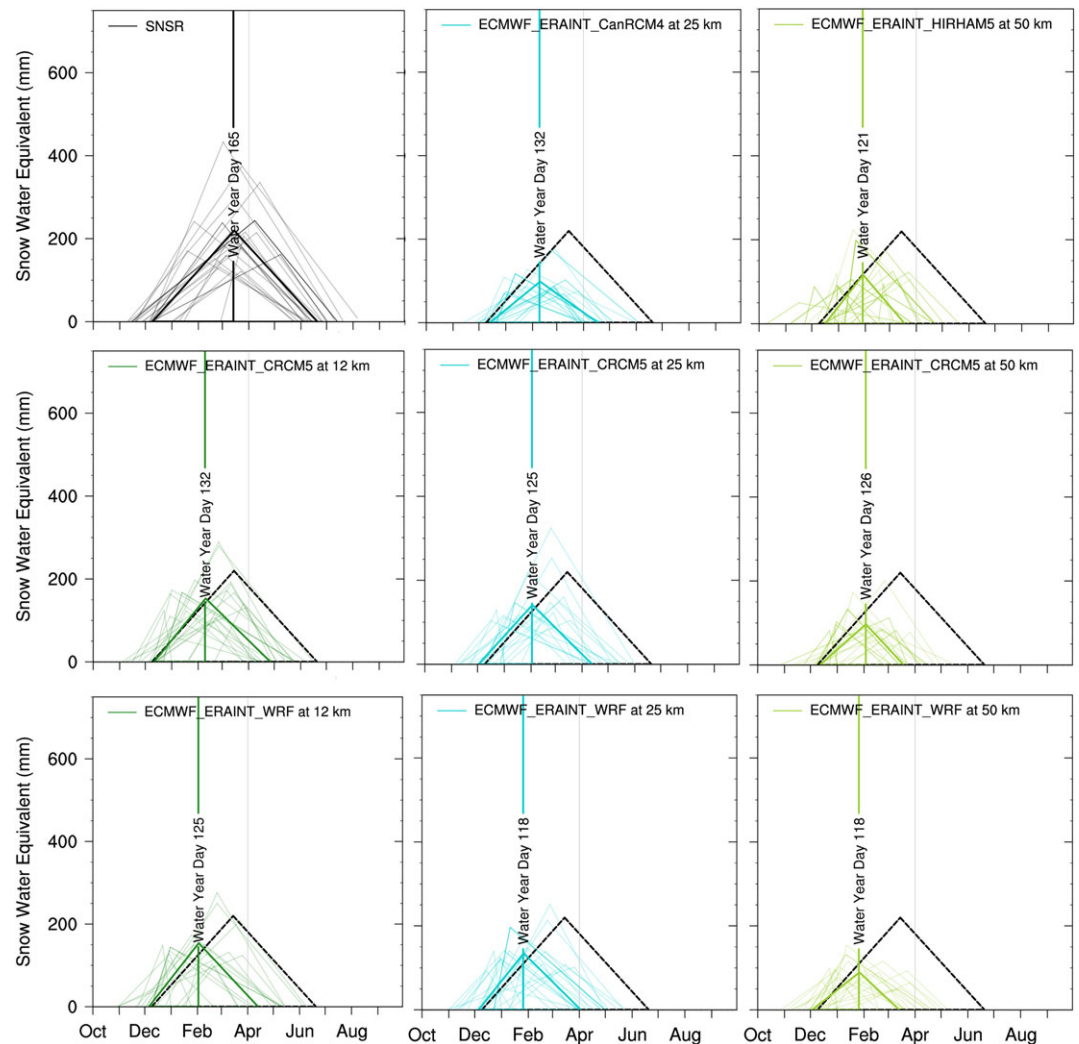
Figure 6. SWE triangles for the eight RCM-ERA Interim simulations across the 10 upstream analysis regions. Colors are used to distinguish resolution. Light green represents model resolutions of 50 km, cyan represents 25 km, and dark green represents 12 km. Each SWE triangle represents a linearized daily average approximation of snowpack dynamics in an individual water year within 1985–2005. The 1 April is shown with a vertical gray line to highlight the date of historical peak SWE timing in the western United States. The SWE triangle component average is emboldened and the snowpack peak accumulation date for a given snowpack product is shown with a vertical line. The standard reference data set used for this study (SNSR) is overlaid on all figures with black dashed lines. SWE = snow water equivalent; SNSR = Sierra Nevada Snow Reanalysis.

data sets. SMR differed between NLDAS_2_VIC at -1.11 ± 0.22 mm/day and L15 at -3.00 ± 0.49 mm/day. This drove a 42-day difference in the estimation of the total length of the snow season (i.e., AS + MS). For example, NLDAS_2_SAC had the lowest number of accumulation and melt season days at 66 ± 13 and 86 ± 13, whereas NLDAS_2_VIC had the highest number of accumulation and melt season days at 98 ± 16 and 96 ± 12. Since the magnitudes for many of the SWE triangle metrics varied by 2 times across the observationally constrained snow products, we also compute the ratio of SAR/SMR and AS/LS to get an idea of the congruency between the snow accumulation and melt portions of the water year. All observationally constrained snow products are in agreement that the ratio of AS/MS is 0.8 to 1.0, meaning the accumulation season occurs at or less than half of the winter season. However, there were clear differences in SAR/SMR across the observationally constrained snow products. L15 and NLDAS_2_VIC snow season melt rates were generally the same (3–5% slower) compared with accumulation rates, whereas SNSR (NLDAS_2_SAC) had a 16–36% slower melt rate compared with accumulation rate. This indicates that although SAC and VIC were driven by the same North American Regional Reanalysis atmospheric forcing data set and their DJF climate average tendencies

were very similar, large disagreements in the sub-annual processes that shape accumulation and melt rate dynamics were apparent.

To elucidate interseasonal snowpack dynamics, supporting information Figure S2 shows the time series for the various components of the SWE triangle for all 10 upstream analysis regions. Since the analysis period only spans 21 seasons, inferring trends related to large-scale teleconnections and/or climate change would be misleading. However, the time series analysis in this study helps to highlight erroneous snowpack dynamics that persist into subsequent winter seasons. As shown in SNSR, no clear trend across SWE triangle metrics is observed. This generally holds for L15, NLDAS_2_SAC, and NLDAS_2_VIC as well. An exception is shown in L15, with an average +5 MAF in TWV and a decrease in SMR by −1.00 mm/day from 1985 to 2005.

3.2.2. Atmospheric Reanalysis-Driven RCM Simulations

Figure 6 highlights the SWE triangles for all five of the RCM-ERA Interim simulations for the 10 upstream analysis regions. SNSR is also shown for juxtaposition. In general, the RCM-ERA Interim simulations had shorter accumulation season lengths (44 ± 14 to 62 ± 16 days), melt season lengths (44 ± 12 to 76 ± 16 days), and a peak SWE timing occurring too soon compared with SNSR (Table 4). In addition, accumulation rates were more varied (−15% to +113%) and, importantly, peak water volumes were substantially lower (−30% to −60%) than SNSR. Therefore, RCM-ERA Interim data sets can be characterized as having a later start to the accumulation season, a more varied accumulation rate, a lower peak water volume, a generally higher melt rate (+67%), on average, and a total winter season length that was −76 days shorter than SNSR. Possible hypotheses for the source of the simulated SWE biases are model temperature biases, the covariability of precipitation and temperature biases, and model structural/parameter uncertainty in the snow dynamics such as differences in the land surface models and how they parameterize latent heat loss, snow layers and melt, and albedo. A more formal attribution and causation study as to why these abrupt snow melt events occurred will need to be conducted in future research. The relative roles of the atmosphere-ocean-land coupled system in shaping these abrupt melt rates also remains to be addressed.

Interseasonal snowpack dynamics for each of the RCM-ERA Interim simulations is shown in supporting information Figure S3 with SNSR shown for comparison. As discussed before, no trend was observed in SNSR over the 1985–2005 period for any of the SWE triangle metrics. Most of the RCM-ERA Interim simulations had much more year-to-year variability in SAR, yet the average trend in SPD tended toward an earlier timing each year and TWV was consistently smaller than was shown in SNSR. For example, three SNSR water years had SWE that was ≥15 MAF yet only one RCM-ERA Interim simulation (ECMWF_ERAINT_CRCM5 at 25 km) topped 14 MAF in a single water year. Again, SMR showed the most disagreement in interseasonal trend across the eight RCM-ERA Interim simulations. Only one of the RCM-ERA Interim simulations (ECMWF_ERAINT_CRCM5 at 25 km) showed similar year-to-year trend in SMR to SNSR. The other simulations, especially at 50 km, had much more variance between water years than SNSR with SMR radically faster 9 times across RCM-ERA Interim simulations (i.e., ≥10 mm/day, or 5 times any SMR value found in SNSR). The refinement of model horizontal resolution to ≤25 km generally helped to constrain erroneous year-to-year variance in SMR within most of the RCM-ERA Interim simulations.

3.2.3. GCM-Driven RCM Simulations

The last set of SWE triangles are presented in Figure 7 for the RCM-GCM simulations across the 10 upstream analysis regions. SNSR is also shown for comparison. The RCM-GCM simulations SAR and TWV estimates widely varied (Table 4). The lowest SAR was 2.01 ± 0.52 mm/day, or 1/4 slower than SNSR, while the highest SAR was 5.58 ± 0.71 mm/day, or 2 times faster than SNSR. Further, total water volumes varied between 3 times lower (3.47 ± 0.66 MAF) to 2 times higher (19.37 ± 2.92 MAF) than was found in SNSR. Across all nine of the RCM-GCM simulations, CanESM2_CRCM5 at 50 km was most closely aligned with SNSR. CanESM2_CRCM5 at 50 km had an SAR within 4% (2.77 ± 0.69 mm/day) and TWV (7.39±1.73 MAF) within 24% of SNSR. However, this correspondence did not translate to SMR, AS, and/or MS. Akin to RCM-ERA Interim simulations, melt rates across RCM-GCM data sets were generally higher than SNSR, regardless of SAR or TWV. The lowest SMR was 1.66 ± 0.54 mm/day and the highest was 7.71 ± 6.83 mm/day. Additionally, all RCM-GCM simulations had a shorter total snow season length (on average, 44 less days) when compared with SNSR. The RCM-GCM simulation snow season congruency (AS/MS) varied more so than previous simulations and/or data sets. MPI_ESM_CRCM5 at 50 km indicated its snow season had a longer accumulation season (1.39), whereas HadGEM2_WRF at 50 km had a longer melt season (0.66).

Supporting information Figure S4 highlights the interseasonal snowpack dynamics across the 10 upstream analysis regions for the RCM-GCM simulations. All of the RCM-GCM simulations agree with SNSR on a
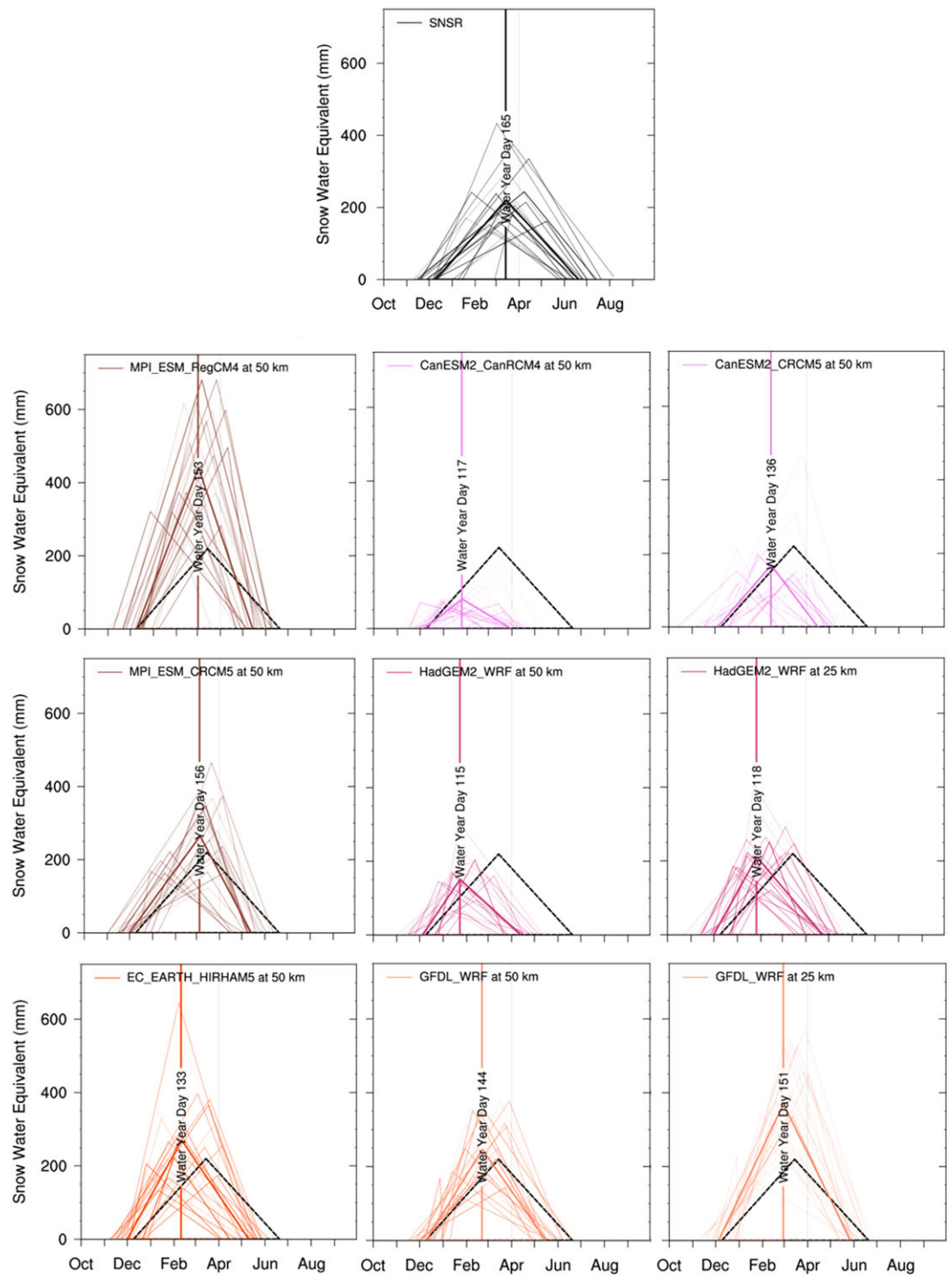
Figure 7. SWE triangles for the nine RCM-GCM simulations across the 10 upstream analysis regions. Colors are used to distinguish GCM forcing data set. Orchid represents CanESM2, coral is MPI_ESM, orange red is EC_EARTH, salmon is GFDL, and violet red is HadGEM2. Each SWE triangle represents a linearized daily average approximation of snowpack dynamics in an individual water year within 1985–2005. The 1 April is shown with a vertical gray line to highlight the date of historical peak SWE timing in the western United States. The SWE triangle component average is emboldened and the snowpack peak accumulation date for a given snowpack product is shown with a vertical line. The standard reference data set used for this study (SNSR) is overlaid on all figures with black dashed lines. SWE = snow water equivalent; SNSR = Sierra Nevada Snow Reanalysis.

10 upstream analysis regions	SAR	TWV	SMR	SPD	AS	MS
a) SNSR (reference)	0.00	0.00	0.00	0.00	0.00	0.00
L15	0.32	0.38	1.10	-0.48	-0.06	-0.28
NLDAS_2_SAC	-0.98	-1.43	-1.84	-0.30	0.07	-0.02
NLDAS_2_VIC	-0.54	-1.43	-1.48	-1.18	-0.83	-0.46
b) ECMWF_ERAINT_CRCM5 at 12 km	0.40	-0.82	0.80	-1.35	-0.99	-0.86
ECMWF_ERAINT_CRCM5 at 25 km	0.18	-0.97	3.12	-1.64	-0.94	-1.14
ECMWF_ERAINT_CRCM5 at 50 km	0.15	-1.50	5.14	-1.62	-1.20	-2.18
ECMWF_ERAINT_CanRCM4 at 25 km	-0.25	-1.48	-0.84	-1.35	-1.06	-1.24
ECMWF_ERAINT_HIRHAM5 at 50 km	1.96	-1.26	6.76	-1.80	-1.45	-2.00
ECMWF_ERAINT_WRF at 12 km	0.77	-0.80	1.43	-1.64	-1.04	-1.11
ECMWF_ERAINT_WRF at 25 km	0.50	-1.04	0.87	-1.92	-1.21	-1.27
ECMWF_ERAINT_WRF at 50 km	-0.21	-1.60	1.86	-1.92	-1.16	-2.09
c) CanESM2_CanRCM4 at 50 km	-0.42	-1.73	-0.99	-1.96	-1.27	-1.34
CanESM2_CRCM5 at 50 km	0.07	-0.65	1.20	-1.20	-0.70	-1.45
MPI_ESM_CRCM5 at 50 km	0.41	0.52	2.93	-0.37	-0.01	-1.16
MPI_ESM_RegCM4 at 50 km	1.91	2.62	6.64	-0.50	-0.35	-0.99
EC_EARTH_HIRHAM5 at 50 km	1.12	0.60	1.72	-1.30	-0.66	-0.33
GFDL_WRF at 50 km	1.16	0.30	4.15	-0.86	-0.60	-0.53
GFDL_WRF at 25 km	1.41	1.81	8.45	-0.60	-0.29	-0.35
HadGEM2_WRF at 50 km	0.89	-0.86	-0.28	-2.06	-1.30	-0.68
HadGEM2_WRF at 25 km	1.19	-0.12	0.55	-1.92	-1.09	-0.42

Figure 8. Z scores for the six SWE triangle metrics used in this study across the 10 upstream analysis regions. The six SWE triangle 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. Similar to previous figures, text color is used to distinguish resolution in (b) and global climate model forcing data set in (c). SWE = snow water equivalent; SNSR = Sierra Nevada Snow Reanalysis.

limited-to-no trend in SAR, although the accumulation rate and year-to-year variability were higher and less constrained than SNSR. In addition, the TWV values also widely varied with some RCM-GCM simulations producing too little TWV, such as CanESM2_CanRCM4 at 50 km with only a single water year above 7 MAF, and some producing consistently high TWV, such as MPI_ESM_RegCM4 at 50 km with 13 water years at ≥15 MAF. Akin to RCM-ERA Interim simulations, SMR year-to-year variability was larger than the observationally constrained snow products. However, unlike the RCM-ERA Interim simulations, SMR was only radically faster (≥10 mm/day) than that found in SNSR 3 times across the RCM-GCM simulations.

3.3. Comparison Across SWE Metrics

To summarize agreement/disagreement in snowpack dynamics across observationally constrained snow products and/or model simulations, we have tabulated and color coded Z scores in Figure 8 with the six SWE triangle metrics. Z scores allow for metrics with widely varying magnitudes to be intercompared in the same statistical space (Wilks, 2011). To standardize the Z score across SWE products, SNSR's temporal mean and standard deviation is used against each individual data set. Therefore, SNSR is shown on each plot at the 0 Z score line in emboldened black. If a Z score is 0, the data set mean is exactly the same as SNSR, 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 SNSR. A summary of the Z scores for various SWE triangle metrics are given in supporting information Table S2.

As highlighted previously, SMR exhibits the most spread across observationally constrained snow products when evaluating all 10 of the upstream analysis regions (Figure 8). Within observationally constrained snow products, SMR Z scores varied between 1.10 (L15) and −1.84 (NLDAS_2_VIC; supporting information Table S2). This likely indicates that spring-dependent melting factors such as days above freezing and/or cloud/surface shortwave and longwave radiation feedbacks widely differed across data sets. Unfortunately, not all reanalysis

data sets provided temperature and cloud/surface radiative feedback variables along with their SWE esti-mates, so this could not be explored further. SMR also dominated the differences in RCM-ERA Interim and RCM-GCM simulations. All but three simulations (ECMWF_ERAINT_CanRCM4 at 25 km, CanESM2_CanRCM4 at 50 km, and HadGEM2_WRF at 50 km) had SMR Z scores that were positive. This was even shown for the extreme winters simulated by MPI_ESM_RegCM4 at 50 km, which had the highest positive bias in SAR (1.91) and TWV (2.62) across all simulations. Both MPI_ESM_RegCM4 at 50 km and GFDL_WRF at 25 km had extremely high SMR bias at 6.64 and 8.45, respectively. Overall, RCM-ERA Interim and RCM-GCM simulations had an absolute average bias in SMR of 1.47 ± 0.43 and 2.99 ± 1.92. This presents a challenge to model devel-opers as spring snowmelt rate is particularly important for water management decision-makers who balance competing interests between flood management and water supply and can ill-afford biased forecasts.

Supporting information Figure S5 shows three of the 10 upstream analysis regions, Shasta, Don Pedro, and Isabella, which are representative of the northern, central, and southern portions of the Sierra Nevada. Unfor-tunately, the three northernmost upstream analysis regions (i.e., Shasta, Oroville, and Folsom), which provide 22% of water storage capacity in the state, were the regions with the most spread in SWE triangle metrics, especially SMR. This is likely due to the region having an average elevation lower than the Southern Sierra Nevada which induces higher interannual variability in rain-snow partitioning (Bales et al., 2006). Shasta is shown in supporting information Figure S5 as it is the largest region and had the most mismatch in SMR. Within observationally constrained snow products, SMR ranged from −2.21 to 2.69, whereas the RCM-ERA Interim (4.18 to 14.3) and RCM-GCM (4.84 to 23.9) simulations were all high biased. RCM-ERA Interim and RCM-GCM absolute average bias in SMR was 8.38 ± 2.97 and 11.2 ± 4.45, respectively. Conversely, Isabella, the southernmost and third largest upstream analysis region, had SMR values that were some of the least biased across the 10 upstream analysis regions. The observationally constrained snow products' SMR were at 0.85 ± 0.24 and, similarly, RCM-ERA Interim and RCM-GCM were low biased at 0.74 ± 0.43 and 1.11 ± 0.92.

The Z scores in Figure 8 and supporting information Figure S5 highlight that model simulations show more agreement in the Southern Sierra Nevada than the Northern Sierra Nevada. However, across all 10 upstream regions, systemic biases in how the interactions between simulated snowpack and spring-dependent melting factors such as days above freezing and cloud/surface shortwave and longwave radiation feedbacks may have unrealistically shortened their total winter season lengths. This is further shown in supporting information Table S2 where Z-scores across all simulations are most in agreement during winter accumulation (AS and SAR) and in most disagreement during the melt phase (SMR and MS).

4. Discussion

A multimetric approach to characterizing mountains as natural reservoirs offers several insights regarding the processes that contribute to biases in both observationally constrained snow products and the repre-sentation of snow dynamics within climate models. The observationally constrained snow products that we evaluated have greater skill than RCMs when examining DJF SWE averages (Figure 4). However, the more detailed process-oriented metrics encapsulated in the SWE triangle (Figure 8) reveal more subtle biases and inconsistencies across these data sets, including the potential for compensating errors that would not be reflected in the DJF SWE alone. The observationally constrained snow products show variation in all of the SWE triangle metrics, reflecting a diversity of data sources, assumptions, and methodologies used in creat-ing these data sets. Using the six SWE triangle metrics we uncovered that differences in data sets were partly driven by discrepancies in the accumulation season (SAR and AS), but larger differences were found in the melt season (SMR and MS), although both the accumulation and melt season share common sensitivities to the peak timing and maximum amount of snowpack (SPD and TWV).

In contrast, the atmospheric reanalysis-driven RCM simulations show a more consistent pattern of bias in the SWE triangle metrics. As a group, these simulations demonstrate low SAR, low TWV, early SPD, high SMR, and a correspondingly short AS and MS, indicating that the models are generally overly sensitive to late-winter to early-spring conditions resulting in abrupt snowmelt. The greatest variation across these metrics can be seen in the SMR, but it is not clear that SMR bias can fully explain variation in the DJF SWE skill across the models. For instance, The DJF SWE skill scores indicate that higher resolution models perform better in terms of spatial ratio of variance and root-mean square error (Figure 4), but the intermediate resolution WRF run (25 km) has comparable (or better) skill in all six SWE triangle metrics compared with the higher resolution 12-km simulations (Figure 8). This finding further indicates that a relatively coarse single metric evaluation based on

average DJF SWE accumulation can miss more subtle strengths and weaknesses across data sets including the role of compensating errors.

Somewhat surprisingly, the same RCMs forced with GCM boundary conditions rather than atmospheric reanalysis show several improvements in the DJF SWE skill scores (Figure 4). The SWE triangle metrics offer insight into why this is the case. For example, the RCMs forced by MPI_ESM and EC_EARTH had high-biased SWE total water (ASB) up to +4.49 (MPI_ESM_RegCM4) MAF and those forced by CanESM2 exhibited a low bias down to −3.20 MAF, although this was less than or within the range exhibited by the RCM-ERA Interim simulations. Many of the GCM-forced RCM simulations show high-biased TWV and SAR compared to the consistently low-biased values for these metrics in the atmospheric reanalysis forced RCM runs (Figure 8). The bias in simulated SWE could be due to model temperature biases, the covariability of precipitation and temperature biases, and model structural/parameter uncertainty in the snow dynamics such as differences in the land surface models in how they parameterize latent heat loss, snow layers and melt water transfer, and albedo. The additional snow accumulation in these simulations is overcome by an even greater model spread in SMR biases compared to reanalysis driven RCMs, which tends to improve the SPD, MS, and AS metrics. Because the mean DJF SWE metric focuses on a fixed period of time, an offset of SPD could yield a poor score, even if other aspects of the SWE triangle are accurate. Thus it is notable that the GCM-driven simulations show improvements in SPD despite large biases in other aspects of the SWE triangle.

Considering interannual variability in the SWE triangle metrics provides additional insight into model performance and sources of bias, several GCM-driven simulations produced too few large peak water volumes, with one simulation only producing a single water year above 7 MAF and another producing 13 water years at ≥15 MAF (4 times that of SNSR). Year-to-year variability in melt rates were much larger than either the observationally constrained snow products and/or RCM-ERA Interim simulations. Thus, an erratic and less constrained year-to-year melt rate was characteristic of many RCM-GCM simulations.

By characterizing mountains as natural reservoirs using the multimetric SWE triangle approach described in this paper, we offer more detailed information that may be more relevant to water management decision-makers. The seasonal timing and magnitude of snowmelt is one of most critical considerations for reservoir management (Hamlet & Lettenmaier, 1999; Steinschneider & Brown, 2012), and models are key tools for characterizing both the interannual variability in these quantities and their longer-term trajectories of change.

Although a consensus on the best available snow models across spatial and temporal scales was not identified in the most recent Snow Model Intercomparison Project (Essery et al., 2009), this study, and others, were able to identify common parameter and structural uncertainties in the community of snowpack models. For example, Barlage et al. (2010) identified that the Noah land surface model (used in the seven WRF simulations in NA-CORDEX) can have excessive latent heat flux and melting during the accumulation phase of the snow season; van Kampenhout et al. (2017) identified in the Community Land Model (not used in NA-CORDEX) that parameterizations representing meltwater percolation, storage and refreezing caused significant temperature increases due to latent heat release in deeper snowpack whereas shallower snowpack tended to send more percolation water to runoff, which removed latent heat; and Chen et al. (2014) identified that across six widely used snow models, including the Noah and VIC model used in this study, parameterization of snow albedo and below-canopy turbulence and radiation transfer was critical to accurately simulate SWE. Further studies that build on the aforementioned are needed to isolate the relative contributions of the atmospheric forcing data set, topographic resolution, and land surface model parameter/structural uncertainties in shaping peak SWE timing and the processes that contribute to excessive spring snowmelt rates found in this study. One enticing approach for tackling this problem is through the use of a more modular approach to snowpack modeling such as the Structure For Unifying Multiple Modeling Alternatives (Clark et al., 2015a, 2015b) coupled with the use of multimetric model evaluation frameworks, such as the SWE triangle presented in this study.

Through explicit evaluation of model performance based on these quantities, we create an opportunity to break down model biases into different components and identify the aggregate result of these model biases. Our analysis reveals that many aspects of model bias are not evident using a simpler single-metric approach such as evaluating the skill of reproducing spatial patterns of average DJF SWE. Future work to attribute

and eliminate such biases in the snow models could ultimately improve the accuracy and credibility of future snowpack projections and better support decision-makers who are faced with unprecedented uncertainty in future snowpack conditions.

References

Anderson, E. A. (1976). A point of energy and mass balance model of snow cover (*NOAA Tech. Rep. NWS 19*). United States: Office of Hydrology.

Ashfaq, M., Ghosh, S., Kao, S.-C., Bowling, L. C., Mote, P., Touma, D., et al. (2013). Near-term acceleration of hydroclimatic change in the western U. S. *Journal of Geophysical Research: Atmospheres*, *118*, 10,676–10,693. https://doi.org/10.1002/jgrd.50816

Ashfaq, M., Rastogi, D., Mei, R., Kao, S.-C., Gangrade, S., Naz, B. S., & Touma, D. (2016). High-resolution ensemble projections of near-term regional climate over the continental United States. *Journal of Geophysical Research: Atmospheres*, *121*, 9943–9963. https://doi.org/10.1002%2F2016JD025285

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*, W08432. https://doi.org/10.1029/2005WR004387

Barlage, M., Chen, F., Tewari, M., Ikeda, K., Gochis, D., Dudhia, J., et al. (2010). Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *Journal of Geophysical Research*, *115*, D22101. https://doi.org/10.1029/2009JD013470

Bartos, M. D., & Chester, M. V. (2015). Impacts of climate change on electric power supply in the Western United States. *Nature Climate Change*, *5*(8), 748–752. https://doi.org/10.1038/nclimate2648

Belmecheri, S., Babst, F., Wahl, E. R., Stahle, D. W., & Trouet, V. (2016). Multi-century evaluation of Sierra Nevada snowpack. *Nature Climate Change*, *6*(1), 2–3. https://doi.org/10.1038/nclimate2809

Bohr, G. S., & Aguado, E. (2001). Use of April 1 SWE measurements as estimates of peak seasonal snowpack and total cold-season precipitation. *Water Resources Research*, *37*(1), 51–60. https://doi.org/10.1029/2000WR900256

CCTAG (2015). Perspectives and guidance for climate change analysis, California Department of Water Resources, Climate Change Technical Advisory Group, pp. 1-92. http://www.water.ca.gov/climatechange/docs/2015/Perspectives_Guidance_Climate_Change_Analysis.pdf

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., et al. (2014). Modeling seasonal snowpack evolution in the complex terrain and forested Colorado headwaters region: A model intercomparison study. *Journal of Geophysical Research: Atmospheres*, *119*, 13,795–13,819. https://doi.org/10.1002/2014JD022167

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015a). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*, 2498–2514. https://doi.org/10.1002/2015WR017200

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research*, *51*, 2515–2542. https://doi.org/10.1002/2015WR017200

Cristea, N. C., Breckheimer, I., Raleigh, M. S., HilleRisLambers, J., & Lundquist, J. D. (2017). An evaluation of terrain-based downscaling of fractional snow covered area datasets based on Lidar-derived snow data and orthoimagery. *Water Resources Research*, *53*, 6802–6820. https://doi.org/10.1002/2017WR020799

Dettinger, M. D., & Anderson, M. L. (2015). Storage in California's reservoirs and snowpack in this time of drought. *San Francisco Estuary and Watershed Science*, *13*(2), 1–5. http://escholarship.org/uc/item/8m26d692

Dilling, L., & Lemos, M. C. (2011). Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy. *Global Environmental Change*, *21*(2), 680–689. http://www.sciencedirect.com/science/article/pii/S0959378010001093

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stähli, M., Gustafsson, D., et al. (2009). SNOWMIP2: An evaluation of forest snow process simulations. *Bulletin of the American Meteorological Society*, *90*(8), 1120–1136. https://doi.org/10.1175/2009BAMS2629.1

Fayad, A., Gascoin, S., Faour, G., López-Moreno, J. I., Drapeau, L., Page, M. L., & Escadafal, R. (2017). Snow hydrology in Mediterranean mountain regions: A review. *Journal of Hydrology*, *551*, 374–396. https://doi.org/10.1016/j.jhydrol.2017.05.063

Galford, G. L., Nash, J., Betts, A. K., Carlson, S., Ford, S., Hoogenboom, A., et al. (2016). Bridging the climate information gap: A framework for engaging knowledge brokers and decision makers in state climate assessments. *Climatic Change*, *138*(3), 383–395. https://doi.org/10.1007/s10584-016-1756-4

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: The CORDEX framework. *World Meteorological Organization (WMO) Bulletin*, *58*(3), 175–183. https://www.cordex.org/images/pdf/cordex_giorgi_wmo.pdf

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*, D06104. https://doi.org/10.1029/2007JD008972

Hall, A. (2014). Projecting regional change. *Science*, *346*(6216), 1461–1462. http://science.sciencemag.org/content/346/6216/1461

Hamlet, A. F., & Lettenmaier, D. P. (1999). Effects of climate change on hydrology and water resources in the Columbia river basin. *JAWRA Journal of the American Water Resources Association*, *35*(6), 1597–1623.

Harpold, A. A., Dettinger, M., & Rajagopal, S. (2017). Defining snow drought and why it matters. *EOS. Transactions of the American Geophysical Union*, *98*.

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107.

Henn, B., Clark, M. P., Kavetski, D., McGurk, B., Painter, T. H., & Lundquist, J. D. (2016). Combining snow, streamflow, and precipitation gauge observations to infer basin-mean precipitation. *Water Resources Research*, *52*(11), 8700–8723.

Hossain, F., Arnold, J., Beighley, E., Brown, C., Burian, S., Chen, J., et al. (2015). What do experienced water managers think of water resources of our nation and its management infrastructure? *PLOS ONE*, *10*(11), 1–10. https://doi.org/10.1371/journal.pone.0142073

Huang, X., Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2016). An evaluation of the variable-resolution CESM for modeling California's climate. *Journal of Advances in Modeling Earth Systems*, *8*(1), 345–369.

Jones, A., Calvin, K., & Lamarque, J.-F. (2016). Climate modeling with decision makers in mind. *Eos*, *97*. https://doi.org/10.1029/2016EO051111

Kapnick, S., & Hall, A. (2010). Observed climate–snowpack relationships in California and their implications for the future. *Journal of Climate*, *23*(13), 3446–3456.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., et al. (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics*, *49*(1), 71–95.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., et al. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950s–2013. *Scientific Data*, *2*(150042).

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P., & Lettenmaier, D. P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states: Update and extensions. *Journal of Climate*, *26*(23), 9384–9392.

López-Moreno, J. I., Gascoin, S., Herrero, J., Sproles, E. A., Pons, M., Alonso-González, E., et al. (2017). Different sensitivities of snowpacks to warming in Mediterranean climate mountain areas. *Environmental Research Letters*, *12*(7), 74006.

Lowrey, J. L., Ray, A. J., & Webb, R. S. (2009). Factors influencing the use of climate information by Colorado municipal water managers. *Climate Research*, *40*(1), 103–119. http://www.jstor.org/stable/24870453

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A landsat-era sierra nevada snow reanalysis (1985–2015). *Journal of Hydrometeorology*, *17*(4), 1203–1221.

Margulis, S. A., Cortés, G., Girotto, M., Huning, L. S., Li, D., & Durand, M. (2016). Characterizing the extreme 2015 snowpack deficit in the Sierra Nevada (USA) and the implications for drought recovery. *Geophysical Research Letters*, *43*, 6341–6349. https://doi.org/10.1002/2016GL068520

Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., & Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrology and Earth System Sciences*, *14*(6), 1125–1138.

McCrary, R. R., McGinnis, S., & Mearns, L. O. (2017). Evaluation of snow water equivalent in NARCCAP simulations, including measures of observational uncertainty. *Journal of Hydrometeorology*, *18*(9), 2425–2452.

McNeeley, S. M., Tessendorf, S. A., Lazrus, H., Heikkila, T., Ferguson, I. M., Arrigo, J. S., et al. (2012). Catalyzing frontiers in water-climate-society research: A view from early career scientists and junior faculty. *Bulletin of the American Meteorological Society*, *93*(4), 477–484.

Mearns, L. O. (2017). The NA-CORDEX dataset, version 1.0, *NCAR Climate Data Gateway, date accessed: 07-2017 to 12-2017*. https://doi.org/10.5065/D6SJ1JCH

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, *319*(5863), 573–574.

Minder, J. R., Letcher, T. W., & Skiles, S. M. (2016). An evaluation of high-resolution regional climate model simulations of snow cover and albedo over the Rocky Mountains, with implications for the simulated snow-albedo feedback. *Journal of Geophysical Research: Atmospheres*, *121*, 9069–9088. https://doi.org/10.1002/2016JD024995

Moss, R. H., Meehl, G. A., Lemos, M. C., Smith, J. B., Arnold, J. R., Arnott, J. C., et al. (2013). Hell and high water: Practice-relevant adaptation science. *Science*, *342*(6159), 696–698.

Mote, P. W. (2006). Climate-driven variability and trends in mountain snowpack in Western North America. *Journal of Climate*, *19*(23), 6209–6220.

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in Western North America. *Bulletin of the American Meteorological Society*, *86*(1), 39–49.

Mount, J., & Hanak, E. (2015). Water Use in California, *Public Policy Institute of California*. http://www.ppic.org/publication/water-use-in-california/

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*(3), 214–219.

Naz, B. S., Kao, S.-C., Ashfaq, M., Rastogi, D., Mei, R., & Bowling, L. C. (2016). Regional hydrologic response to climate change in the conterminous United States using high-resolution hydroclimate simulations. *Global and Planetary Change*, *143*, 100–117.

Pagán, B. R., Ashfaq, M., Rastogi, D., Kendall, D. R., Kao, S.-C., Naz, B. S., et al. (2016). Extreme hydrological changes in the southwestern US drive reductions in water supply to Southern California by mid century. *Environmental Research Letters*, *11*(9), 94026. http://stacks.iop.org/1748-9326/11/i=9/a=094026

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (2016). The airborne snow observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, *184*, 139–152. https://doi.org/10.1016/j.rse.2016.06.018

Palmer, P. L. (1988). The SCS snow survey water supply forecasting program: Current operations and future directions. In *Western Snow Conference, Proc. No. 56*, Montana, pp. 43–51.

Pavelsky, T. M., Kapnick, S., & Hall, A. (2011). Accumulation and melt dynamics of snowpack from a multiresolution regional climate model in the central Sierra Nevada, California. *Journal of Geophysical Research*, *116*, D16115. https://doi.org/10.1029/2010JD015479

Pederson, G. T., Gray, S. T., Woodhouse, C. A., Betancourt, J. L., Fagre, D. B., Littell, J. S., et al. (2011). The unusual nature of recent snowpack declines in the North American Cordillera. *Science*, *333*(6040), 332–335.

Pepin, N., Bradley, R. S., Diaz, H. F., Baraër, M., Caceres, E. B., Forsythe, N., et al. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*, *5*(5), 424–430.

Pierce, D. W., Barnett, T. P., Hidalgo, H. G., Das, T., Bonfils, C., Santer, B. D., et al. (2008). Attribution of declining Western U.S. Snowpack to human effects. *Journal of Climate*, *21*(23), 6425–6444.

Pierce, D. W., Barnett, T. P., Santer, B. D., & Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, *106*(21), 8441–8446.

Pierce, D. W., & Cayan, D. R. (2013). The uneven response of different snow measures to human-induced climate warming. *Journal of Climate*, *26*(12), 4148–4167.

Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., et al. (2014). Climate change impacts on the water balance of the Colorado headwaters: High-resolution regional climate model simulations. *Journal of Hydrometeorology*, *15*(3), 1091–1116.

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., et al. (2011). High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *Journal of Climate*, *24*(12), 3015–3048.

Rhoades, A. M., Huang, X., Ullrich, P. A., & Zarzycki, C. M. (2016). Characterizing sierra Nevada snowpack using variable-resolution CESM. *Journal of Applied Meteorology and Climatology*, *55*(1), 173–196.

Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2017). Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM. *Climate Dynamics*, *50*(1–2), 261–288. https://doi.org/10.1007/s00382-017-3606-0

Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Johansen, H., Margulis, S. A., Morrison, H., et al. (2018). Sensitivity of mountain hydroclimate simulations in variable–resolution CESM to microphysics and horizontal resolution. *Journal of Advances in Modeling Earth Systems*, *10*(9). https://doi.org/10.1029/2018MS001326

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C., & Mote, P. W. (2013). Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *Journal of Geophysical Research: Atmospheres*, *118*(19), 10,884–10,906. https:doi.org/10.1002/jgrd.50843

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, *35*(7), 2145–2160. https:doi.org/10.1029/1999WR900090

Steinschneider, S., & Brown, C. (2012). Dynamic reservoir management with real-option risk hedging as a robust adaptation to nonstationary climate. *Water Resources Research*, *48*(5), W05524. https://doi.org/10.1029/2011WR011540

Sun, F., Hall, A., Schwartz, M., Walton, D. B., & Berg, N. (2016). Twenty-first-century snowfall and snowpack changes over the Southern California Mountains. *Journal of Climate*, *29*(1), 91–110. https://doi.org/10.1175/JCLI-D-15-0199.1

Tarroja, B., AghaKouchak, A., & Samuelsen, S. (2016). Quantifying climate change impacts on hydropower generation and implications on electric grid greenhouse gas emissions and operation. *Energy*, *111*, 295–305. https://doi.org/10.1016/j.energy.2016.05.131

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Tesfa, T. K., Tarboton, D. G., Watson, D. W., Schreuders, K. A. T., Baker, M. E., & Wallace, R. M. (2011). Extraction of hydrological proximity measures from DEMs using parallel processing. *Environmental Modelling and Software*, *26*(12), 1696–1709. https://doi.org/10.1016/j.envsoft.2011.07.018

Thompson, S. E., Sivapalan, M., Harman, C. J., Srinivasan, V., Hipsey, M. R., Reed, P., et al. (2013). Developing predictive insight into changing water systems: Use-inspired hydrologic science for the Anthropocene. *Hydrology and Earth System Sciences*, *17*(12), 5013–5039. https://doi.org/10.5194/hess-17-5013-2013

Trujillo, E., & Molotch, N. P. (2014). Snowpack regimes of the Western United States. *Water Resources Research*, *50*, 5611–5623. https://doi.org/10.1002/2013WR014753

van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G., & van den Broeke, M. R. (2017). Improving the representation of polar snow and firn in the community Earth System Model. *Journal of Advances in Modeling Earth Systems*, *9*(7), 2583–2600.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed.). Academic Press.

Wu, C., Liu, X., Lin, Z., Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., et al. (2017). Exploring a variable-resolution approach for simulating regional climate in the Rocky Mountain region using the VR-CESM. *Journal of Geophysical Research: Atmospheres*, *122*, 10,939–10,965. https://doi.org/10.1002/2017JD027008

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, *117*, D03109. https://doi.org/10.1029/2011JD016048

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*, *117*, D03109. https://doi.org/10.1029/2011JD016048