# A Comparison of National Water Model Retrospective Analysis Snow Outputs at SNOTEL Sites Across the Western U.S.

**Running Title**: National Water Model versus Observed Snow

Irene Garousi-Nejad [1] and David G. Tarboton [1]

[1] Department of Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University, Logan, UT, USA

**Corresponding Author**: Irene Garousi-Nejad, Department of Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University, Logan, UT. Email: i.garousi@aggiemail.usu.edu

**Keywords:** Snow Water Equivalent, Snow-Covered Area Fraction, National Water Model, Noah-MP, Snow Parameterization, Remote Sensing, MODIS, SNOTEL

**Acknowledgement**: We are grateful for funding support from the Utah Water Research Laboratory for this research. This work used compute allocation TG-EAR190007 from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 (Towns et al., 2014). The support and resources from the Center for High Performance Computing (CHPC) at the University of Utah are gratefully acknowledged. Thanks to Arezoo RafieeiNasab of NCAR who supplied some of the National Water Model inputs for the study region and to David Gochis and the NCAR WRH-Hydro research team for encouragement and discussion.

**Conflict of Interest**: The authors declare that they have no conflict of interest.

**Abstract**

This study compares the U.S. National Water Model (NWM) reanalysis snow outputs to observed snow water equivalent (SWE) and snow-covered area fraction (SCAF) at SNOTEL sites across the Western U.S. This was done to evaluate and identify opportunities for improving the modeling of snow in the NWM. SWE was obtained from SNOTEL sites, while SCAF was obtained from MODIS observations at a nominal 500 m grid scale. Retrospective NWM results were at a 1000 m grid scale. We compared results for SNOTEL sites to gridded NWM and MODIS outputs for the grid cells encompassing each SNOTEL site. Differences between modeled and observed SWE were attributed to both model errors, as well as errors in inputs, notably precipitation and temperature. The NWM generally under-predicted SWE, partly due to precipitation input differences. There was also a slight general bias for model input temperature to be cooler than observed, counter to the direction expected to lead to under-modeling of SWE. There was also under-modeling of SWE for a subset of sites where precipitation inputs were good. Furthermore, the NWM generally tends to melt snow early. There was considerable variability between modeled and observed SCAF that hampered useful interpretation of these comparisons. This is in part due to the model grid SCAF essentially being binary (snow or no snow) while observations from MODIS are much more fractional. However, when SCAF was aggregated across all sites and years, modeled SCAF tended to be more than observed using MODIS. These differences are regional with generally better SWE and SCAF results in the Central Basin and Range and differences tending to become larger the further away regions are from this region. These findings identify areas where predictions from the NWM involving snow may be better or worse, and suggest opportunities for research directed towards model improvements.

## 1. INTRODUCTION

Accurate water supply forecasts will become increasingly crucial as western populations grow and demand more water, and as operational agencies have to manage water under global environmental change (Bhatti, Koike, & Shrestha, 2016; Gergel, Nijssen, Abatzoglou, Lettenmaier, & Stumbaugh, 2017; Li, Wrzesien, Durand, Adam, & Lettenmaier, 2017; Livneh & Badger, 2020; Mote, 2003; Mote, Hamlet, Clark, & Lettenmaier, 2005; Regonda, Rajagopalan, Clark, & Pitlick, 2005; Stewart, Cayan, & Dettinger, 2004, 2005). Many scientific challenges in understanding and preparing for global environmental change rest upon our ability to predict streamflow and snowmelt quantity, timing, and spatial patterns that are important for decision making in water-sensitive sectors. In the United States, the

National Weather Service (NWS) of the National Oceanic and Atmospheric Administration (NOAA) is responsible for short- and long-term streamflow predictions across the U.S. Prior to 2016, NWS operational forecasts were limited to forecasts from NWS River Forecast Centers (RFC) at about 4000 forecast points. These were produced predominantly using the Sacramento soil moisture accounting model (SAC-SMA) to simulate runoff production and SNOW-17 model to simulate snowpack and snowmelt, within the Advanced Hydrologic Prediction System (https://water.weather.gov/ahps/rfc/rfc.php) modeling infrastructure (McEnery, Ingram, Duan, Adams, & Anderson, 2005).

While Franz, Hogue, and Sorooshian (2008) showed that SNOW-17 performed well over the Reynolds Creek Experimental Watershed located in southwestern Idaho, other studies found limitations such as being unable to capture snowmelt timing precisely due to its simple conceptual framework, its inability to represent spatial variability of land properties, and its dependence on extensive calibration for each basin using historical data (Lundquist & Flint, 2006; Shamir, Carpenter, Fickenscher, & Georgakakos, 2006; Zalenski, Krajewski, Quintero, Restrepo, & Buan, 2017). Furthermore, a National Research Council committee identified a gap between what is now considered state-of-the-art modeling capabilities and those used in AHPS (National Research Council, 2006). It concluded that the NWS needs to incorporate more advanced hydrologic science into their hydrologic models.

The increasing availability of distributed geographic data and computer power has made it possible to develop national/continental scale, physically-based, and distributed models. In 2016, NOAA's Office of Water Prediction implemented the National Water Model (NWM) as a physically-based distributed model based on the Weather Research and Forecasting Model Hydrological modeling system (WRF-Hydro) framework (Gochis et al., 2020) to provide nationally consistent operational hydrologic forecasting capability. The main goals of the NWM were to provide forecast streamflow, produce spatially continuous national estimates of hydrologic states (soil moisture, snowpack, etc.), and to implement a modeling architecture that permits rapid infusion of new data and science.

The NWM provides hourly flow forecasts at about 2.7 million locations in the U.S. In addition to the increased number of forecast locations, another advantage of the NWM is that it utilizes the physically-based Noah-MultiParameterization (Noah-MP) land surface model to represent the land-atmosphere interactions including snow processes. There have been several studies evaluating results from the NWM. For instance, Viterbo et al. (2020) evaluated the prediction of flooding in NWM streamflow forecasts. They found that errors were due to both meteorological input errors as well as hydrologic process representation. In another study, Lahmers et al. (2019) improved the performance of WRF-Hydro configured as NWM version 1.1 by implementing a conceptual channel infiltration function into the model architecture. They concluded that accounting for channel infiltration loss in the semi-arid Western U.S. improves the streamflow behaviour simulated when the model is forced with high-resolution precipitation input. However, we are not aware of a systematic and thorough evaluation of the NWM snow outputs.

In 2019, the NWM retrospective analysis data used to evaluate version 2.0 prior to operational deployment, were published (https://registry.opendata.aws/nwm-archive/). This retrospective analysis contains output from a 26-year simulation (January 1993 through December 2018), where the model was driven by meteorological inputs produced from observed (for rainfall) and reanalysis (for other required meteorological inputs) datasets, and is referred to as NWM-R2. In terms of snow, outputs include gridded snow water equivalent (SWE), the amount of water stored in a snowpack, and the snow-covered area fraction (SCAF). Across the Western U.S., snow is observed at 808 snow telemetry (SNOTEL) sites that provide data intended to quantify snow and inform water supply forecasts. Illustrative comparisons of NWM-R2 SWE to SNOTEL SWE (Figure 1) indicate that SWE is well modeled at some locations (Figure 1a) while significantly different from observations at other locations (Figure 1b). Accurate modeling of SWE is a necessary condition for accurate physically-based modeling of runoff. This motivated the need, addressed in this study, to systematically evaluate the performance of NWM-R2 simulations of SWE and SCAF against available SNOTEL measurements and the moderate resolution imaging spectroradiometer (MODIS) satellite imagery to answer the following questions:

- How well does the NWM model simulate snowpack (in terms of SWE, SCAF, and snowmelt timing) compare to observations over the entire Western U.S.?
- What are the potential causes responsible for discrepancies in NWM-R2 SWE, SCAF, and snowmelt timing?
- Are these discrepancies associated with the model input errors or the snow parameterization in the model?

Answers to these questions are needed to further improve the NWM snow components, and ultimately runoff and water supply forecasts in snowmelt-dominated regions. While U.S. based, the NWM is built using the WRF-Hydro modeling framework that has been applied worldwide, and the lessons learned from this comparison across the U.S.

have application to the representation of snow processes in national and continental scale models throughout the world.

The following section, Model, Data and Experimental Design, first presents a summary of the NWM-R2 snow parameterization. Then, it described the datasets used in this study, comprised of the NWM-R2 reanalysis products, SNOTEL snow observations, and MODIS imagery giving the snow-covered area fraction. Next, it presents the metrics that were used for evaluating the model results versus observations. The results section compares the amount of NWM-R2 SWE, precipitation, air temperature, and SCAF with observations from SNOTEL and MODIS. It also compares modeled and observed snowmelt timing. Then there is discussion, followed by conclusions. The key conclusions are, that as currently parameterized, the NWM under simulates snow accumulation and models melt too soon. Part of the under simulation is due to biases in precipitation inputs, but even where precipitation inputs are good, there is still some under simulation, suggesting that model structure and overall energy balance process representations need to be improved.

## 2. MODEL, DATA, and EXPERIMENTAL DESIGN

The study region comprises the SNOTEL sites across the Western U.S. (Figure 2a). The model is the NWM version 2.0 reanalysis (NWM-R2), that includes Noah-MP land surface components for snow. Data include: (1) NWM-R2 inputs (precipitation, air temperature, and elevation) and outputs (SWE and SCAF) from the land surface module, (2) In-situ measurements of precipitation and air temperature, elevation, and SWE from SNOTEL, and (3) remotely sensed snow-covered areas captured by the MODIS sensor for water years 2008-2018. These three datasets have different spatial resolutions (Figure 2b). The difference in scale is a potential source of uncertainty in our comparative analysis, and needs to be recognized in interpretation. There are small differences in elevation between SNOTEL (point elevations) and NWM-R2 (1 km grid elevations), that may impact temperature comparisons due to lapse rate effects, but there does not appear to be any significant bias (Figure 2c).

[Insert Figure 2]

### 2.1 NWM-R2 Snow Parameterization and Snow Reanalysis Products

The NWM uses Noah-MP as the land surface model to simulate snow processes through a 1-dimensional vertical column over 1 km spatial resolution grid cells. There are four main features of Noah-MP that are used in snow processes simulation. We describe these features briefly here.

#### *2.1.1 Snowfall*

The separation of precipitation into rainfall or snowfall is one of the most sensitive parameterizations in simulating cold−region hydrological processes (Loth, Graf, & Oberhuber, 1993). It is common for precipitation partitioning to be based on near surface air temperature. Noah-MP uses Jordan's (1991) algorithm with double thresholds to partition precipitation into rainfall and snowfall. This approach ignores some physical processes controlling precipitation phase by not incorporating humidity. This has been reported to lead to biases in SWE, snow depth, and snow cover fraction (Chen, Liu, Dudhia, & Chen, 2014; Harder & Pomeroy, 2014; Wang et al., 2019).

#### *2.1.2 Vegetation and Snow Interception*

A single-layer vegetation canopy model characterizes the fraction covered by vegetation (FVEG) in each model grid. Since the Noah-MP dynamic vegetation option is set off in NWM-R2, the model uses the maximum vegetation fraction from the Leaf area index (LAI) table as FVEG. If a model grid has a FVEG>0 and a snow depth greater than 0.025 m (from initial conditions or the last time step), the model computes the fraction of canopy buried by snow based on the snow depth and the canopy height. Then, it uses that fraction to adjust the LAI and stem area index (SAI) after burying by snow, which are used in the canopy interception (for both liquid water and ice mass) calculation. The vegetation canopy model also includes melt and refreeze processes.

#### *2.1.3 Energy Balance and Snow Albedo*

Shortwave radiation is modeled over the entire grid cell using a modified two-stream approximation treating the vegetation as evenly distributed with gaps. The result is canopy-absorbed and ground-absorbed solar radiation over the grid cell. Longwave radiation, latent heat, sensible heat, and ground heat fluxes are modeled, using a tile approach that treats vegetated and bare fractions of the cell separately. These fluxes are then aggregated based on the vegetated fraction (FVEG) parameter. Noah-MP treats turbulence fluxes between the snowpack, vegetation canopy, and air using Monin-Obukhov similarity theory to model atmospheric stability conditions. Stability corrections of under canopy turbulent transfer account for the strong stable condition of a warmer canopy overlying the snow surface during the melt season (Chen et al., 2014). Snow surface albedo is modeled using the Biosphere-Atmosphere Transfer Scheme (BATS) that has separate direct and diffusive radiation in visible and near-infrared bands accounting for fresh snow albedo, snow age, grain size growth, impurity, and especially solar zenith angle.

#### *2.1.4 Snowpack Treatment*

The Noah-MP snow module uses up to three snow layers, depending on depth, to simulate liquid water retention, refreezing, and snowpack densification including a thin surface layer to quantify snow surface temperature. Noah-MP calculates SCAF based on snow density, SWE, snow depth, density of fresh snow, and an area-depth snowmelt factor that determines the curve relating SCAF and snow depth in the melting season. In NWM-R2 simulations, snowmelt factor is a calibration parameter that was adjusted to match streamflow over calibration watersheds (RafieeiNasab et al., 2020). The functional relationship between SCAF and depth quantifies small-scale variability of snow within a computational grid element which plays an important role in the process governing snow accumulation and ablation. There are limitations with the current SCAF representation in Noah-MP as it lacks the distinct representation of some factors affecting SCAF such as vegetation (type and dynamic) and topography. These limitations affect the accurate simulation of SCAF and SWE (Helbig, Herwijnen, Magnusson, & Jonas, 2015; Magand, Ducharne, Le Moine, & Gascoin, 2014; Swenson & Lawrence, 2012; Wrzesien, Pavelsky, Kapnick, Durand, & Painter, 2015).

This study used the NWM-R2's land surface model outputs, which are geospatial gridded results with a spatial resolution of 1 km and temporal resolution of 3-hours. We obtained the NWM-R2 SWE (model code name: SNEQV) and SCAF (model code name: FSNO) from the NOAA Google Cloud archive using a Jupyter Notebook (Tarboton & Garousi-Nejad, 2020) that we developed. Then, we averaged 3-hourly results to daily values to have a similar temporal resolution when comparing the NWM-R2 results with SNOTEL and MODIS observations (because both these datasets produce daily data). We also obtained the precipitation, air temperature, and elevation input data used for NWM-R2 simulations. The WRF-Hydro team at NCAR prepared precipitation and air temperature values for us as those data were not available on the Google Cloud archive.

## 2.2 SNOTEL

SNOTEL stations, managed by the Natural Resources Conservation Service (NRCS), generally consist of a snow pillow, an air temperature sensor, and a storage precipitation gage. Our study used the daily precipitation, air temperature, SWE, and snow depth values measured at SNOTEL sites as a reference dataset to evaluate the NWM-R2 precipitation, air temperature, SWE and snow depth. We realize that SNOTEL data must be used with some caution because the sites are mostly located in small clearings within forests protected by forest canopies, leading to differences in exposure to wind and radiation (McCreight, Small, & Larson, 2014). Furthermore, SNOTEL data do not undergo a high correction level

(Swenson & Lawrence, 2012). In some instances, we found unrealistically high temperature values that needed to be filtered out. Nevertheless, SNOTEL data remain the only widespread in situ SWE observations available for model validation (Barlage et al., 2010; Clow, Nanus, Verdin, & Schmidt, 2012; Livneh, Xia, Mitchell, Ek, & Lettenmaier, 2010; Pan et al., 2003; Toure et al., 2016). We automated retrieval of the SNOTEL data by calling its Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI) web service from a Jupyter Notebook script (Garousi-Nejad & Tarboton, 2021a).

**2.3 MODIS**

NASA's MODIS instrument launched aboard the Terra satellite in late 1999 is designed to observe and monitor Earth changes, such as snow cover. MODIS has spectral bands in the visible and near-infrared regions, nominal 500 m spatial resolution, and near-daily global coverage. The daily snow-cover gridded tile product, MOD10A1, has been used and improved over time in multiple snow studies (Aalstad, Westermann, & Bertino, 2020; Bennett, Cherry, Balk, & Lindsey, 2019; Magand et al., 2014; Masson et al., 2018; Salomonson & Appel, 2006; Swenson & Lawrence, 2012). We used products from the current version of the MODIS snow-cover algorithm which is the collection 6 suite of MODIS (hereafter referred to as MODIS-C6, or just MODIS). We chose to use MODIS-C6 (Hall & Riggs, 2016) as a reference to evaluate NWM-R2 SCAF because the improvements/revisions to MODIS-C6 (i.e., accounting for the surface temperature and surface height) led to a notable increase in accuracy of snow cover detection on mountain ranges and low illumination conditions in the Northern Hemisphere during spring and summer (Riggs, Hall, & Román, 2017).

The MODIS-C6 snow algorithm is designed to detect snow cover based on the normalized ratio of the differences in reflectance in band 4 (centred at 0.56 µm, visible green) and band 6 (centred at 1.64 µm) of the MODIS instrument with revisions applied to alleviate snow detection commission errors (reported for previous versions) for which snow detection is uncertain. The MODIS-C6 supplies the Normalized Difference Snow Index (NDSI) rather than snow cover (product name: NDSI_Snow_Cover). This approach allows users to have the option to modify the NDSI using the global empirical model or develop region-specific models (Riggs, Hall, & Román, 2016; Aalstad et al., 2020). In this study, we first developed a script to retrieve average NDSI_Snow_Cover (500 m spatial resolution) for each NWM grid cell containing a SNOTEL site (1 km spatial resolution) from Google Earth Engine (Garousi-Nejad & Tarboton, 2021b). Valid NDSI_Snow_Cover values range between 0-100 with values above 100 indicating missing data, no decision, night, inland water, ocean, cloud, and detector saturated issues, masked out in Google Earth Engine. The returned MODIS images thus have spatial gaps due to the masking. We filled gaps in each image with NDSI_Snow_Cover from the most previous valid value (forward filling). Then, we applied the globally-determined linear model of Riggs et al. (2016) to compute MODIS-C6 SCAF from NDSI_Snow_Cover values [Equation (1)].

$$SCAF = \min[\max(-0.01 + 1.45 \times NDSI, 0), 1] \quad \text{where } NDSI \in [0,1] \qquad (1)$$

The resulting data set includes 2,504,102 site-days in the period of overlap between NWM-R2 and SNOTEL data. We organized the SNOTEL sites into subgroups using Omernik Ecoregions level III (Omernik & Griffith, 2014) available from the Commission for Environmental Corporation (http://www.cec.org/north-american-environmental-atlas/terrestrial-ecoregions-level-iii/) to identify regional differences in model results versus observations. The ecoregions are areas with general similarities in location, climate, vegetation, hydrology, terrain, wildlife, and land use and have been used in multiple prior studies (Sun et al., 2019; Trujillo & Molotch, 2014).

**2.4 Metrics**

We used several metrics to compare NWM-R2 snow water equivalent (SWE), snow covered area fraction (SCAF), precipitation (P), and snowmelt timing against SNOTEL SWE and MODIS-C6 SCAF.

- **First day of the month comparisons** were used for NWM-R2 SWE (modeled) versus SNOTEL SWE (observed) for months Nov-Jun. **Monthly precipitation and average air temperature** were also compared for these months. These monthly comparisons let us evaluate the seasonal variability of snow in both modeled and observed datasets for data in the period of overlap between NWM-R2 and SNOTEL data.
- We also compared SWE on the date of observed peak SWE (**same day comparison**) and observed and modeled peak SWE (**different day comparison**), and SCAF on these same and different dates. Total precipitation accumulated from the start of the water year, Oct 1, to the date of peak SWE was also computed to assess the degree to which differences may be attributable to precipitation differences. This was done for both same day (observed peak SWE) and different day (observed and modeled peak day) comparisons. The different peak day comparison addresses the possibility that peak modeled and observed

SWE may be close, but appear further apart in same day comparisons due to a timing mismatch.

- To compare the melt timing, we used the **half melt from peak SWE date** (Clow, 2010). This date, when half the snowpack has melted serves as a measure of melt timing somewhat robust to small fluctuations or a long period where SWE is flat near the peak. We categorized the differences between observed and modeled half melt dates as close (within 5 days), model early (the model is 6 to 19 days ahead of observed), model late (the model is 6 to 19 days after observed), and far apart (the modeled and observed differ by 20 days or more).

We also computed commonly used statistics:

- Coefficient of determination [$r^2$, Equation (2)] that ranges from -1 to 1 with 1 indicating a perfect positive linear relationship but insensitive to proportional differences between modeled and observed data;
- Spearman's rank correlation [Spearmanr, Equation (3)], a non-parametric measure of correlation used to measure the strength of association between modeled and observed values where value 1 means a perfect positive correlation;
- Root mean square error [RMSE, Equation (4)], a measure of how concentrated the data are around the line of best fit;
- Nash Sutcliffe efficiency [NSE, Equation (5)], a normalized statistic that determines the relative magnitude of the residual variance compared to observed values ranging from -∞ to 1 with 1 indicating observed and modeled data fits the 1:1 line; and
- Bias [Bias, Equation (6)], the average of the difference between modeled and observed.

$$r^2 = \left[ \frac{\sum_{t=1}^{N} (M_t - \overline{O}_t)(M_t - \overline{M}_t)}{\sqrt{\sum_{t=1}^{N} (O_t - \overline{O}_t)^2 \sum_{t=1}^{N} (M_t - \overline{M}_t)^2}} \right]^2 \tag{2}$$

$$\text{Spearmanr} = 1 - \frac{6 \sum_{t=1}^{N} d_t^2}{N(N^2-1)} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} (O_t - M_t)^2}{N}} \tag{4}$$

$$NSE = 1 - \frac{\sum_{t=1}^{N} (O_t - M_t)^2}{\sum_{t=1}^{N} (O_t - M_t)^2} \tag{5}$$

$$Bias = \frac{\sum_{t=1}^{N} (M_t - O_t)}{N} \tag{6}$$

where $M_t$ is model simulation, $O_t$ is observation, N is the total number of simulations or observations, $d_t$ is difference between observed and modeled rank, and the overbar indicates average.

# 3. RESULTS

## 3.1 Seasonal (Monthly) Comparison

We compared the NWM-R2 SWE results with observations from SNOTEL and found a persistent bias in modeled SWE across most months (Figure 3). Results show that throughout the accumulation phase (Nov-Feb), the rank correlation between observed and modeled SWE increases (Spearmanr from 0.7 to 0.8). However, this does not necessarily indicate an acceptable model performance. The discrepancies between the observed and modeled SWE increase as snow accumulates (RMSE 21 to 135 mm). In the ablation phase (Mar-Jun), the rank correlation decreases, and discrepancies are highest in May (Bias -149 mm, RMSE 292 mm). The increasing scatter in later months (Figure 3) shows that the NWM generally performs well during the accumulation phase but simulates SWE less well during the ablation phase. Most points fall below the 1:1 line (red line). The points clustered into vertical and horizontal lines on the bottom and left axes of scatter plots in May and Jun indicate early and late modeling of complete melt out, respectively.

[Insert Figure 3]

The comparison between the NWM-R2 SCAF and estimates from MODIS-C6 revealed that the modeled SCAF is highly uncorrelated with what is detected by satellite imagery (Figure 4). Throughout the last three months of the accumulation phase (Dec-Feb), the NWM results show that more than 70% of points (each representing one NWM grid cell that includes a SNOTEL site and a water year) have SCAF 0.9-1, while less than 10% have SCAF 0-0.1 (histograms in Figure 4). In contrast to the binary behaviour of the NWM-R2 SCAF, MODIS SCAF exhibits gradual increases and decreases. At most, 30% of the observed data have SCAF values ranging from 0.9-1 during the accumulation phase. In December, 14% of the observed data have SCAF greater than 0.9, while about 70% of modeled points have SCAF greater than 0.9. During the ablation phase (Mar-Jun), both modeled and observed datasets have relatively a similar data percentage with SCAF less than 0.1. However, the portion of the points where modeled SCAF is above 0.9 is still much more significant (3-7 times depending on the month) than those in the observed dataset (histograms in Figure 4).

[Insert Figure 4]

The SCAF comparisons above are only at SNOTEL sites. We did not undertake the computation needed to compare NWM-R2 and MODIS-C6 for all grid cells and dates. However, as an illustration for locations beyond SNOTEL sites NWM-R2 and MODIS-C6 SCAF maps on Dec 1, 2011 (Figure 5) show that while patterns are generally the same, MODIS SCAF seems less than modeled. Note that the MODIS-C6 SCAF map (Figure 5a) has gaps and cloud areas (grey) that we did not fill in from the most recent previous image with data (as described in Section 3) for this visualization. NWM-R2 SCAF covers the entire region selected based on the MODIS tiles. The visual comparison of a zoomed-in map for the region where observed SCAF were available for more than 90% of the area reveals both similarities and differences between NWM-R2 and MODIS-C6 datasets (Figure 5c and 5d). The NWM-R2 SCAF map for the zoomed-in area shows more white regions (i.e., SCAF values greater than 0.9), suggesting that NWM tends to overestimate SCAF compared to observations from MODIS.

[Insert Figure 5]

Scatterplots of monthly precipitation (Figure 6) indicate model input precipitation generally less than measured at SNOTEL sites, possibly contributing to under-modeling of SWE (Figure 3). Spearmanr and NSE values show an acceptable correlation between modeled and observed monthly precipitation (on average, 0.8 for both statistics). However, the precipitation bias is larger during the accumulation phase than the ablation phase, suggesting that increased SWE scatter, in the ablation phase, is less associated with precipitation input errors than other factors during the ablation phase snowmelt.

[Insert Figure 6]

Elevation, through orographic effects, is often suspected as a contributor to precipitation bias. However, the comparison of model input elevation (1 km grid cell) with SNOTEL point elevation (Figure 2) indicated no bias and small scatter ($r^2$=0.98 in Figure 2c). There are, nevertheless, discrepancies between the NWM-R2 monthly averaged air temperature inputs and the monthly averages of the daily mean air temperature measured at SNOTEL sites (Figure 7), reported as the 24-hour average of a minimum four samples per hour (USDA, 2011). NWM-R2 air temperatures are generally slightly below observations. This is counter to the direction needed to explain discrepancies in SWE as colder model input air temperatures should result in (1) greater fractions of precipitation as snowfall and (2) slower rather than quicker snowmelt, both processes that increase rather than decrease SWE.

[Insert Figure 7]

The seasonal pattern of SWE and SCAF averaged across all SNOTEL site years for each specific day (Figure 8) further indicates the general under modeling of SWE and over modeling of SCAF relative to SNOTEL and MODIS observations, respectively.

[Insert Figure 8]

Discrepancies between the seasonal pattern of SWE and SCAF are regional and somewhat different for SWE than SCAF (Figure 9 and Figure 10, respectively). The NWM SWE was better in the Klamath Mountains, Blue Mountains, and Central Basin and Range (region 9, 2, and 5, respectively, in Figure 9) with SWE bias differences tending to become larger further to the north and east across the study region. However, the NWM SCAF are closer to the observations in the Northern Basin and Range, Sierra Nevada, and Central Basin and Range regions (regions 12, 13, and 5, respectively, in Figure 10), with SCAF differences tending to become larger the further away regions are from the Central Basin and Range region.

[Insert Figure 9]

[Insert Figure 10]

### 3.2 Observed Peak SWE (Same Day and Different Day) Comparison

The scatterplot of modeled versus observed SWE on the date of peak observed SWE (Figure 11a) indicates a general downward bias in modeled SWE. NWM SCAF clusters around 1 on this date (histograms in Figure 11b) while MODIS SCAF is more fractional, and similar to monthly SCAF the point comparisons are scattered and poor. Precipitation accumulated from Oct 1 to the date of observed peak SWE indicates model input precipitation generally less than SNOTEL observed (Figure 11c: Bias -111 mm, RMSE 212 mm). This suggests that under estimation of model precipitation inputs may be a contributor to under modeling of peak SWE. This comparison may also be influenced by the fact that observed SWE is at its peak, but modeled SWE is not.

[Insert Figure 11]

We also compared observed and modeled peak SWE, noting that these do not necessarily occur on the same date (Figure 12). Results are similar to the observed peak SWE date comparison. Here the accumulated observed and modeled precipitation (Figure 12c) are over the accumulation period, to their respective peak SWE dates, a possible reason for increased scatter and poorer error metrics in this figure.

[Insert Figure 12]

Under modeling of SWE is also evident when comparing the observed and modeled peak SWE for a subset of SNOTEL sites where the model precipitation is relatively close to the observed (Figure 13b: Bias -96 mm, RMSE 168 mm). However, the errors are less than for the entire dataset SWE comparison. We chose this subset of sites based on the NSE measure between daily model input and observed precipitation being greater than or equal to 0.9 computed over the full study period. This subset shows a reduced bias (compared to the entire dataset) between the observed and modeled precipitation accumulated from Oct 1 to peak observed SWE date (Figure 13a).

[Insert Figure 13]

### 3.3 Melt Timing Comparison

For 68% of the site years analyzed, the modeled half melt date was earlier than observed. When further classified based on whether modeled half melt dates were close, ahead, behind or far apart from observed melt dates (Figure 14) we observe that the NWM half melt date was greater than 20 days from observed half melt date, for 34% of the site years, and off by 6 days or more for 75% of site years. For those site years where the difference was between 5 and 20 days, a greater percentage had the model melting ahead, than behind the observed. The site years that have modeled half melt date ahead of observed tend to have lower modeled half melt date SWE (which is by definition half the peak SWE) than observed (Figure 14b).

[Insert Figure 14]

## 4. DISCUSSION

The seasonal pattern of SWE and SCAF averaged across all SNOTEL site years shows that NWM generally under-estimates SWE and over-estimates SCAF relative to SNOTEL and MODIS observations, respectively. These discrepancies vary regionally with relatively better SWE results in the Arizona/New Mexico Mountains, Blue Mountains, and Central Basin and Range ecoregions; and better SCAF results in the Central Basin and Range and Sierra Nevada ecoregions tending to become larger the further away regions are from the Central Basin and Range. There are several sources of uncertainties in our comparisons that need to be pointed out. We compared SNOTEL site values to the single NWM grid cell values for the grid cell containing each SNOTEL site. We realize that using other approaches, such as bilinear or cubic interpolation of NWM grid values would give different values at each SNOTEL site, a question we did not explore.

Precipitation discrepancies suggest that SWE differences are partly due to discrepancies between observed precipitation and model input precipitation. There are multiple possible sources of uncertainty that may lead to this difference. First, SNOTEL latitude and longitude locations may not be precise in the geographic information from SNOTEL, as, for site security, exact site locations may not be reported. This may result in selecting a non-representative 1 km NWM grid cell. Second, there may be systematic bias for gage precipitation, particularly with snowfall measurements being subject to "under-catch" (Mote, 2003; Sun et al., 2019). However, we note that model input precipitation was typically less than measured at SNOTEL sites, indicating that if under−catch is an issue, it may be larger in the data used to produce model inputs. In NWM version 2.0, a mountain mapper adjustment has been applied to obtain input precipitation from NLDAS-2 (RafieeiNasab et al., 2020); nevertheless, there are still differences and biases compared to SNOTEL measurements that may be impacting model results. Third, there are also errors in SWE measurements due to factors such as wind causing snowdrifts on the snow pillow, or the small clearing SNOTEL site location not being representative of larger scale snowpack. It was not uncommon to see SWE greater than accumulated precipitation measured at SNOTEL sites, which could be due to either precipitation under-catch, or inflated SWE.

Our results show a cold (downward) bias for the model input air temperature compared to SNOTEL sites' observations. This is different from Naple et al. (2020), who reported a warm (upward) bias for the NWM retrospective runs compared to the New York State Mesonet observations. The cold bias in the model temperature input is counter to the direction expected to lead to the under-modeling of SWE, which needs more investigation. The discrepancies in model inputs appear to be not the only sources responsible for SWE differences. For sites with statistically highly correlated precipitation input (NSE>0.9), the results indicate that some SWE bias, potentially due to other factors, still remains. This opens up the question as to whether there are other deficiencies in the model that lead to SWE under-modeling. The partitioning of precipitation into rainfall and snowfall has been identified as one problem area within the NWM land surface model [i.e., Noah-MP] (Wang et al., 2019; Naple et al., 2020). Wang et al. (2019) suggest that using a snow-rain partitioning scheme based on the wet-bulb temperature within Noah-MP produces more snowfall and snow mass on the ground that agrees better with ground-based snow observations, particularly over mountainous regions in the Western U.S. Recently, Naple et al. (2020) shows that using the precipitation phase partition from the high-resolution rapid refresh (HRRR), in lieu of the operational method (Jordan, 1991), leads to improved snow results for the NWM version 2.0 configuration.

Our results show that, on average, the NWM tends to melt snow early (6-19 days) compared to SNOTEL observation. For 75% of the site years, the modeled date of half melt from peak SWE was off by 6 days or more from the observed half melt dates, sometimes being as far apart as 2 months (for example, Magic Mountain SNOTEL site, ID: 610 in Idaho, at water year 2010). This suggests that the modeling of melt timing is somewhat problematic and there is a need to further investigate overall energy balance and snow surface temperature, possibly drawing on ideas from the Utah Energy Balance model (Mahat & Tarboton, 2014; You, Tarboton, & Luce, 2014).

Overall, NWM-R2 SCAF was difficult to compare to MODIS-C6 SCAF using single SNOTEL sites and days. Some of this difficulty, manifested in the scatter in Figures 4, 11 and 12, may reflect the fact that the MODIS and NWM SCAF quantities are not really the same thing. MODIS may be interpreting vegetation as snow free, while NWM has snow beneath vegetation. In NWM-R2 results, the persistent low and high SCAF (<0.1 and >0.9, respectively) reflects that NWM treats SCAF as a binary metric in mountainous regions. NWM-R2 SCAF values stay near 1 with less variability between Dec-Apr for more than 70% of cases. This suggests that once the NWM grid cell (1 km spatial resolution) is more than 90% snow-covered, it is implausible for it to diverge from 1 for the rest of the accumulation phase and early ablation phase. This has also been noted by others (Helbig et al., 2015; Magand et al., 2014; Swenson & Lawrence, 2012; Wrzesien et al., 2015).

We recognize that the SCAF mapped from MODIS in this study also has uncertainties and limitations. First, the temporal forward filling approach that we used to fill gaps associated with clouds may miss some of the daily variability of snow cover, particularly in mountainous regions. Second, the parameters of Equation (1), which estimates SCAF from MODIS-C6 NDSI_Snow_Cover product, were those from Salomonson and Appel (2006) and were constant for our entire study region. Adjusting these parameters to improve the snow cover products from MODIS regionally has been suggested (Riggs, Hall, & Román, 2017). Third, MODIS NDSI_Snow_Cover grids (nominally 500 m) were averaged for 1 km NWM grid cells, using an unweighted approach in the Google Earth Engine platform. This approach selects MODIS grids whose centers fall within the target area (i.e., NWM grid cells). These scale differences may be a further source of uncertainty, compounded by the nonlinearity in Equation (1) [plateau at NDSI > 0.7] having an impact on SCAF from averaged NDSI.

## 5. CONCLUSIONS

A cell by cell comparison for sites and dates in the period of overlap between SNOTEL SWE with modeled SWE from NWM-R2 simulations, in general, shows that there is a tendency of NWM to under-estimate SWE early in the season and become progressively more biased late in the season compared to in situ observations of SWE. When aggregated across all sites and years, seasonal variations show an overall downward bias of about 55 mm with NSE 0.75 which varies regionally over Omernik ecoregions. SWE discrepancies were attributed to errors in inputs, notably precipitation and air temperature. The downward bias in precipitation input contributes to the downward biases in SWE and the SWE bias is persistent even when the model precipitation input is relatively close to the observed precipitation at SNOTEL sites with daily precipitation NSE higher than 0.9. However, the cold bias in the model temperature input is counter to the direction expected to lead to under-modeling of SWE. This needs further exploration. There was a significant variability between the MODIS SCAF and NWM SCAF in the cell by cell comparison for sites and dates in the period of overlap between model results and observations which hindered useful interpretation of these comparisons. The challenge in simulating SCAF is in part due to the model SCAF essentially being binary while observations are much more fractional. They may not reflect the same physical quantity. However, when aggregated across all sites and years, seasonal variations show an overall upward bias of 0.12 with NSE 0.76 which vary regionally for ecoregions. Our investigation opens some new questions for future research. First, it emphasizes the importance of having a more accurate (bias corrected) precipitation and air temperature input for the NWM. Second, there is a question as to whether, in circumstances where NWM struggles to accurately simulate SCAF, the SCAF parameterization should be improved or can be inferred from satellites. Using satellite-based snow-covered maps may potentially provide an approach or an opportunity for estimating SCAF as a way to overcome limitations associated with parameterization of SCAF in the snow model. However, there would need to be resolution of differences in definition of the physical quantity being compared. Overall, our evaluation effort identifies some challenges in the current snow parameterization within the NWM and suggests where potential development effort should be directed in the future. It would also be helpful, for future work, to have and use a more comprehensive observation data set that includes snowfall/rainfall measurements, canopy snow interception, turbulence and radiation fluxes above and below the canopy, and high-resolution snow-covered area information, to assess any model improvements.

## DATA AVAILABILITY

All data sources used in this research are publicly available.

- The NWM-R2 are available at the NOAA Google Cloud archive at https://console.cloud.google.com/storage/browser/national-water-model-v2?pli=1. The precipitation and air temperature inputs were prepared by the WRF-Hydro NCAR but are available on HydroShare for reproducibility purposes (Garousi-Nejad & Tarboton, 2021c). The NWM elevation dataset are available at https://www.nco.ncep.noaa.gov/pmb/codes/nwprod/nwm.v2.0.4/parm/domain/
- The NRCS SNOTEL data are available at https://www.wcc.nrcs.usda.gov/snow/
- The NASA MODIS data are available at https://nsidc.org/data/MOD10A1/versions/6
- The Omernik ecoregions are available at http://www.cec.org/north-american-environmental-atlas/terrestrial-ecoregions-level-iii/

All codes developed for this research are shared and publicly available on HydroShare and will be published for the reproducibility of the results after the review process (Garousi-Nejad & Tarboton, 2021d).

## REFERENCES

Aalstad, K., Westermann, S., & Bertino, L. (2020). Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography. *Remote Sensing of Environment*, *239*, 111618. https://doi.org/10.1016/j.rse.2019.111618

Barlage, M., Chen, F., Tewari, M., Ikeda, K., Gochis, D., Dudhia, J., … Mitchell, K. (2010). Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *Journal of Geophysical Research*, *115*(D22), D22101. https://doi.org/10.1029/2009JD013470

Bennett, K. E., Cherry, J. E., Balk, B., & Lindsey, S. (2019). Using MODIS estimates of fractional snow cover area to improve streamflow forecasts in interior Alaska.

*Hydrology and Earth System Sciences*, *23*(5), 2439–2459. https://doi.org/10.5194/hess-23-2439-2019

Bhatti, A. M., Koike, T., & Shrestha, M. (2016). Climate change impact assessment on mountain snow hydrology by water and energy budget-based distributed hydrological model. *Journal of Hydrology*, *543*, 523–541. https://doi.org/10.1016/j.jhydrol.2016.10.025

Chen, Fei, Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., … Yang, Z.-L. (2014). Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study. *Journal of Geophysical Research: Atmospheres*, *119*(24), 13,795-13,819. https://doi.org/10.1002/2014JD022167

Chen, Feng, Liu, C., Dudhia, J., & Chen, M. (2014). A sensitivity study of high-resolution regional climate simulations to three land surface models over the western United States: SENSITIVITY STUDY OF LSMS IN WRF. *Journal of Geophysical Research: Atmospheres*, *119*(12), 7271–7291. https://doi.org/10.1002/2014JD021827

Clow, D. W. (2010). Changes in the timing of snowmelt and streamflow in Colorado: A response to recent warming. *Journal of Climate*, *23*(9), 2293–2306. USGS Publications Warehouse. https://doi.org/10.1175/2009JCLI2951.1

Clow, D. W., Nanus, L., Verdin, K. L., & Schmidt, J. (2012). Evaluation of SNODAS snow depth and snow water equivalent estimates for the Colorado Rocky Mountains, USA: EVALUATION OF SNODAS. *Hydrological Processes*, *26*(17), 2583–2591. https://doi.org/10.1002/hyp.9385

Franz, K. J., Hogue, T. S., & Sorooshian, S. (2008). Operational snow modeling: Addressing the challenges of an energy balance model for National Weather Service forecasts. *Journal of Hydrology*, *360*(1–4), 48–66. https://doi.org/10.1016/j.jhydrol.2008.07.013

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., & Stumbaugh, M. R. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, *141*(2), 287–299. https://doi.org/10.1007/s10584-017-1899-y

Gochis, D. J., Barlage, M., Cabell, R., Casali, M., Dugger, A., FitzGerald, K., … Zhang, Y. (2020). The WRF-Hydro® modeling system technical description, (Version 5.1.1). NCAR Technical Note. https://ral.ucar.edu/sites/default/files/public/WRFHydroV511TechnicalDescription.pdf

Garousi-Nejad, I., D. Tarboton (2021a). Notebook for retrieval of precipitation, air temperature, and snow water equivalent measurements at SNOTEL sites, HydroShare, http://www.hydroshare.org/resource/d1fe0668734e4892b066f198c4015b06

Garousi-Nejad, I., D. Tarboton (2021b). JavaScript code for retrieval of MODIS Collection 6 NDSI snow cover at SNOTEL sites and a Jupyter Notebook to merge/reprocess data, HydroShare, http://www.hydroshare.org/resource/d287f010b2dd48edb0573415a56d47f8

Garousi-Nejad, I., D. Tarboton (2021c). Notebooks for pre-processing the retrieved National Water Model V2.0 Retrospective run results and inputs at SNOTEL sites, HydroShare, http://www.hydroshare.org/resource/1b66a752b0cc467eb0f46bda5fdc4b34

Garousi-Nejad, I., D. Tarboton (2021d). Data for A Comparison of National Water Model Retrospective Analysis Snow Outputs at SNOTEL Sites Across the Western U.S., HydroShare, http://www.hydroshare.org/resource/7a51f56c2cf24ae78012ac6a6d4815a6

Hall, D. K., & Riggs, G. A. (2016). MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid [Dataset: NDSI_Snow_Cover]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/MODIS/MOD10A1.006

Harder, P., & Pomeroy, J. W. (2014). Hydrological model uncertainty due to precipitation-phase partitioning methods: HYDROLOGIC MODEL UNCERTAINTY OF PRECIPITATION-PHASE METHODS. *Hydrological Processes*, *28*(14), 4311–4327. https://doi.org/10.1002/hyp.10214

Helbig, N., van Herwijnen, A., Magnusson, J., & Jonas, T. (2015). Fractional snow-covered area parameterization over complex topography. *Hydrology and Earth System Sciences*, *19*(3), 1339–1351. https://doi.org/10.5194/hess-19-1339-2015

Jordan, R. E. (1991). A One-dimensional temperature model for a snow cover: Technical documentation for SNTHERM.89. Cold Regions Research and Engineering Laboratory (U.S.). http://hdl.handle.net/11681/11677

Lahmers, T. M., Gupta, H., Castro, C. L., Gochis, D. J., Yates, D., Dugger, A., … Hazenberg, P. (2019). Enhancing the Structure of the WRF-Hydro Hydrologic Model for Semiarid Environments. *Journal of Hydrometeorology*, *20*(4), 691–714. https://doi.org/10.1175/JHM-D-18-0064.1

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

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. *Nature Climate Change*, *10*(5), 452–458. https://doi.org/10.1038/s41558-020-0754-8

Livneh, B., Xia, Y., Mitchell, K. E., Ek, M. B., & Lettenmaier, D. P. (2010). Noah LSM Snow Model Diagnostics and Enhancements. *Journal of Hydrometeorology*, *11*(3), 721–738. https://doi.org/10.1175/2009JHM1174.1

Loth, B., Graf, H.-F., & Oberhuber, J. M. (1993). Snow cover model for global climate simulations. *Journal of Geophysical Research*, *98*(D6), 10451. https://doi.org/10.1029/93JD00324

Lundquist, J. D., & Flint, A. L. (2006). Onset of Snowmelt and Streamflow in 2004 in the Western United States: How Shading May Affect Spring Streamflow Timing in a Warmer World. *Journal of Hydrometeorology*, *7*(6), 1199–1217. https://doi.org/10.1175/JHM539.1

Magand, C., Ducharne, A., Le Moine, N., & Gascoin, S. (2014). Introducing Hysteresis in Snow Depletion Curves to Improve the Water Budget of a Land Surface Model in an Alpine Catchment. *Journal of Hydrometeorology*, *15*(2), 631–649. https://doi.org/10.1175/JHM-D-13-091.1

Mahat, V., & Tarboton, D. G. (2014). Representation of canopy snow interception, unloading and melt in a parsimonious snowmelt model: CANOPY SNOW INTERCEPTION, UNLOADING AND MELT. *Hydrological Processes*, *28*(26), 6320–6336. https://doi.org/10.1002/hyp.10116

Masson, T., Dumont, M., Mura, M., Sirguey, P., Gascoin, S., Dedieu, J.-P., & Chanussot, J. (2018). An Assessment of Existing Methodologies to Retrieve Snow Cover Fraction from MODIS Data. *Remote Sensing*, *10*(4), 619. https://doi.org/10.3390/rs10040619

McCreight, J. L., Small, E. E., & Larson, K. M. (2014). Snow depth, density, and SWE estimates derived from GPS reflection data: Validation in the western U. S. *Water Resources Research*, *50*(8), 6892–6909. https://doi.org/10.1002/2014WR015561

McEnery, J., Ingram, J., Duan, Q., Adams, T., & Anderson, L. (2005). NOAA'S ADVANCED HYDROLOGIC PREDICTION SERVICE: Building Pathways for Better Science in Water Forecasting. *Bulletin of the American Meteorological Society*, *86*(3), 375–386. https://doi.org/10.1175/BAMS-86-3-375

Mote, P. W. (2003). Trends in snow water equivalent in the Pacific Northwest and their climatic causes: TRENDS IN SNOW WATER EQUIVALENT. *Geophysical Research Letters*, *30*(12). https://doi.org/10.1029/2003GL017258

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), 39–50. https://doi.org/10.1175/BAMS-86-1-39

Naple, P., Letcher, T., Minder, J. R., Gochis. D., Dugger, A. L., & RafieeiNasab, A. (2020), Improving parameterizations of snow in the National Water Model with observations from the New York State Mesonet to better simulate snow and streamflow in the northeastern United States, Abstract [C063-0006] presented at 2020 Fall Meeting, AGU, Virtual, 1-17 Dec.

National Research Council. (2006). Towards advanced hydrologic prediction service (AHPS). National Research Council Committee to Assess the National Weather Service Advanced Hydrologic Prediction Service Initiative. Washington, DC: National Academies Press.

Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environmental Management*, *54*(6), 1249–1266. https://doi.org/10.1007/s00267-014-0364-1

Pan, M., Sheffield, J., Wood, E. F., Mitchell, K. E., Houser, P. R., Schaake, J. C., … Tarpley, J. D. (2003). Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 2003JD003994. https://doi.org/10.1029/2003JD003994

RafieeiNasab, A., Karsten, L., Dugger, A., FitzGerald, K., Cabell, R., Gochis, D., … McAllister, M. (2020). Overview of National Water Model calibration general strategy & optimization, NCAR Community WRF-Hydro Modeling System training workshop, November, https://ral.ucar.edu/projects/wrf_hydro/training-materials

Regonda, S. K., Rajagopalan, B., Clark, M., & Pitlick, J. (2005). Seasonal Cycle Shifts in Hydroclimatology over the Western United States. *Journal of Climate*, *18*(2), 372–384. https://doi.org/10.1175/JCLI-3272.1

Riggs, G.A., Hall, D. K., & Román, M. O. (2016). MODIS Snow Products Collection 6 User Guide. http://modis-snow-ice.gsfc.nasa.gov/uploads/C6_MODIS_Snow_User_Guide.pdf

Riggs, G.A., Hall, D. K., & Román, M. O. (2017). Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. *Earth System Science Data*, *9*(2), 765–777. https://doi.org/10.5194/essd-9-765-2017

Salomonson, V. V., & Appel, I. (2006). Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1747–1756. https://doi.org/10.1109/TGRS.2006.876029

Shamir, E., Carpenter, T. M., Fickenscher, P., & Georgakakos, K. P. (2006). Evaluation of the National Weather Service Operational Hydrologic Model and Forecasts for the American River Basin. *Journal of Hydrologic Engineering*, *11*(5), 392–407. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:5(392)

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2004). Changes in Snowmelt Runoff Timing in Western North America under a `Business as Usual' Climate Change Scenario. *Climatic Change*, *62*(1–3), 217–232. https://doi.org/10.1023/B:CLIM.0000013702.22656.e8

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward Earlier Streamflow Timing across Western North America. *Journal of Climate*, *18*(8), 1136–1155. https://doi.org/10.1175/JCLI3321.1

Sun, N., Yan, H., Wigmosta, M. S., Leung, L. R., Skaggs, R., & Hou, Z. (2019). Regional Snow Parameters Estimation for Large−Domain Hydrological Applications in the Western United States. *Journal of Geophysical Research: Atmospheres*, *124*(10), 5296–5313. https://doi.org/10.1029/2018JD030140

Swenson, S. C., & Lawrence, D. M. (2012). A new fractional snow-covered area parameterization for the Community Land Model and its effect on the surface energy balance: CLM SNOW COVER FRACTION. *Journal of Geophysical Research: Atmospheres*, *117*(D21), n/a-n/a. https://doi.org/10.1029/2012JD018178

Tarboton, D., & Garousi-Nejad, I. (2020). Notebook for retrieval of National Water Model V2.0 Retrospective run results at SNOTEL sites, HydroShare, http://www.hydroshare.org/resource/3d4976bf6eb84dfbbe11446ab0e31a0a

Toure, A. M., Rodell, M., Yang, Z.-L., Beaudoing, H., Kim, E., Zhang, Y., & Kwon, Y. (2016). Evaluation of the Snow Simulations from the Community Land Model, Version 4 (CLM4). *Journal of Hydrometeorology*, *17*(1), 153–170. https://doi.org/10.1175/JHM-D-14-0165.1

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., … Wilkins-Diehr, N. (2014). XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering*, *16*(5), 62–74. https://doi.org/10.1109/MCSE.2014.80

Trujillo, E., & Molotch, N. P. (2014). Snowpack regimes of the Western United States. *Water Resources Research*, *50*(7), 5611–5623. https://doi.org/10.1002/2013WR014753

Viterbo, F., Mahoney, K., Read, L., Salas, F., Bates, B., Elliott, J., … Cifelli, R. (2020). A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood. *Journal of Hydrometeorology*, *21*(3), 475–499. https://doi.org/10.1175/JHM-D-19-0125.1

Wang, Y., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., & Niu, G. (2019). A Wet−Bulb Temperature−Based Rain−Snow Partitioning Scheme Improves Snowpack Prediction Over the Drier Western United States. *Geophysical Research Letters*, *46*(23), 13825–13835. https://doi.org/10.1029/2019GL085722

Wrzesien, M. L., Pavelsky, T. M., Kapnick, S. B., Durand, M. T., & Painter, T. H. (2015). Evaluation of snow cover fraction for regional climate simulations in the Sierra Nevada: EVALUATION OF SNOW COVER FOR REGIONAL SIMULATIONS IN THE SIERRA NEVADA. *International Journal of Climatology*, *35*(9), 2472–2484. https://doi.org/10.1002/joc.4136

You, J., Tarboton, D. G., & Luce, C. H. (2014). Modeling the snow surface temperature with a one-layer energy balance snowmelt model. *Hydrology and Earth System Sciences*, *18*(12), 5061–5076. https://doi.org/10.5194/hess-18-5061-2014

Zalenski, G., Krajewski, W. F., Quintero, F., Restrepo, P., & Buan, S. (2017). Analysis of National Weather Service Stage Forecast Errors. *Weather and Forecasting*, *32*(4), 1441–1465. https://doi.org/10.1175/WAF-D-16-0219.1

## FIGURE LEGENDS

Figure 1. Snow water equivalent from the NWM version 2.0 reanalysis (NWM-R2) dataset compared to in-situ observations at two SNOTEL sites in Utah. (a) Hole-in-Rock site (ID: 528) located at 2794 m elevation for the water year 2008. (b) Tony Grove Lake site (ID: 823) located at 2582 m elevation for the water year 2018.

Figure 2. (a) SNOTEL sites (734 black dots) across the Western United States. (b) Illustrative relationship of Tony Grove Lake, Utah SNOTEL site (ID: 823), within NWM grid cells with a spatial resolution of 1 km and MODIS grid cells with a spatial resolution

of 463 m (nominally 500 m). (c) NWM grid cell elevation vs. elevation reported for SNOTEL sites (observed).

Figure 3. First day of month modeled (NWM-R2) vs. observed (SNOTEL) SWE. Each point is a site and date in the period of overlap between NWM-R2 and SNOTEL data.

Figure 4. First day of month modeled (NWM-R2) vs. observed (MODIS-C6) SCAF for NWM grid cells and MODIS grid cells containing SNOTEL sites. Each point is a site and a date within the period of overlap between NWM and MODIS data. Axis histograms depict the SCAF distributions.

Figure 5. Comparison of NWM-R2 and MODIS-C6 SCAF maps over the study region on Dec 1, 2011. (a) MODIS-C6 SCAF estimated from NDSI_Snow_Cover values of five tiles (in grey). (b) NWM-R2 SCAF outputs at 00:00 UTC masked for the MODIS-C6 tiles. (c) The zoomed-in map of MODIC-C6 SCAF for the blue box in (a). (d) The zoomed-in map of NWM-R2 SCAF for the blue box in (b).

Figure 6. Comparison between NWM-R2 monthly precipitation input (labeled as modeled) and SNOTEL monthly precipitation (labeled as observed). Each point is a site and month in the period of overlap between NWM-R2 and SNOTEL data.

Figure 7. Comparison between NWM-R2 monthly average of hourly air temperature input (labeled as modeled) and SNOTEL monthly average of mean daily air temperature (labeled as observed). Each point is a site and month in the period of overlap between NWM-R2 and SNOTEL data.

Figure 8. Modeled and observed (a) SWE and (b) SCAF averaged across all SNOTEL sites and years for each specific day of the (water) year.

Figure 9. Modeled and observed SWE averaged across all SNOTEL sites and years for each specific day of the (water) year grouped by ecoregion. The map shows 15 Omernik ecoregions where colours represent the bias.

Figure 10. Modeled and observed SCAF averaged across all SNOTEL sites and years for each specific day of the (water) year grouped by ecoregion. The map shows 15 Omernik ecoregions where colours represent the bias.

Figure 11. Comparisons on date of observed peak SWE. (a) NWM-R2 vs. SNOTEL SWE, (b) NWM-R2 vs. MODIS-C6 SCAF, and (c) NWM-R2 vs. SNOTEL precipitation accumulated from Oct 1 to observed peak SWE date. Each point is a site and a water year (that starts Oct 1) in the period of overlap between NWM-R2 and SNOTEL data.

Figure 12. Different date comparison on dates of observed and modeled peak SWE (a) NWM-R2 vs. SNOTEL peak SWE, (b) NWM-R2 vs. MODIS-C6 SCAF, and (c)

NWM-R2 vs. SNOTEL precipitation accumulated from Oct 1 to observed and modeled peak SWE dates. Each point is a site and a water year (that starts Oct 1) in the period of overlap between NWM-R2 and SNOTEL data.

Figure 13. (a) NWM-R2 vs. SNOTEL precipitation accumulated from Oct 1 to observed and modeled peak SWE dates. This figure is similar to Figure 10 (a) but with colours separating points into two groups. The first group (dark blue) contains points where Nash Sutcliffe Efficiency (NSE) values for daily modeled vs. observed precipitation are equal to or greater than 0.9. The second group (light blue) includes points where NSE values for daily modeled vs. observed precipitation are less than 0.9. Statistics are reported separately for the NSE >= 0.9 and NSE < 0.9 subsets. (b) NWM-R2 peak SWE vs. SNOTEL peak SWE for points from (a) that have daily precipitation NSE equal to or greater than 0.9 (dark blue class).

Figure 14. Analysis of melt timing. (a) Classification of differences between observed and modeled dates of half melt from peak SWE. Close: modeled and observed within 5 days of each other; Behind: modeled 6 to 19 days after observed; Ahead: modeled 6 to 19 days before observed; Far apart: Modeled and observed more than 20 days apart. (b) NWM-R2 SWE vs. SNOTEL SWE date of half melt from peak.