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: This work was supported by the Utah Water Research Laboratory for this research and National Science Foundation under collaborative grants OAC-1664061 and OAC-1664119. 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. We also thank the Natural Resources Conservation Service (NRCS) staff, Danny Tappa; Jeff Anderson; Mage Hultstrand; Scott Pattee; Brian Domonkos; Dave Eiriksson; and Jordan Clayton, for their help providing us information on which SNOTEL sites are open.

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. 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 as well as the binary comparison of snow cover presence that hampered useful interpretation of SCAF comparisons. This is in part due to the shortcomings associated with both model SCAF parameterization and MODIS observations, particularly in vegetated regions. 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 et al., 2016; Gergel et al., 2017; Li et al., 2017; Livneh & Badger, 2020; Mote, 2003; Mote et al., 2005; Regonda et al., 2005; Stewart et al., 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 (AHPS, https://water.weather.gov/ahps/rfc/rfc.php) modeling infrastructure (McEnery et al., 2005).

While Franz et al. (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 et al., 2006; Zalenski et al., 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, Barlage, Cabell, Casali, 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 countrywide 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 a specific configuration of 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.

The NWM (Gochis, Barlage, Cabell, Dugger, et al., 2020) has been running in NWS operations since 2016 to support operational flood forecasts. The latest operational version, version 2.0, was implemented in June of 2019. Prior to this operational deployment, the NWM version 2.0 retrospective analysis data were generated (by the NWM team) for investigations into the performance of the NWM. These are publicly available in Google Cloud Storage (National Weather Service, 2019).

These retrospective analysis results contain output from a 26-year simulation (January 1993 through December 2018), hereafter is referred to as NWM-R2. The meteorological forcing data used for the version-2 retrospective analysis configuration was drawn from the North American Land Data Assimilation System II (NLDAS2) datasets, a gridded product with spatial resolution of 1/8th-degree and hourly temporal resolution. The non-precipitation forcing fields in NLDAS2 are from the analysis fields of the National Centers for Environmental Prediction (NCEP)/North American Regional Reanalysis (NARR), i.e., a retrospective dataset, while the precipitation is from the gage-based NCEP/Climate Prediction Center (CPC). As a pre-processing step, the NWM team downscaled the NLDAS2 data and applied a mountain mapper (Hou et al., 2014) adjustment to the precipitation data to adjust the values for climatological variation due to topography and wind directions (RafieeiNasab et al., 2020). The result forcing dataset is a 1 km spatial resolution data layer for each hour which contains incoming short- and longwave radiation, specific humidity, air temperature, surface pressure, near surface wind, and precipitation rate. 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.

[Locate Figure 1 near here]

The following section—Model, Data, and Experimental Design—first presents a summary of the NWM-R2 snow parameterization. Then, it describes 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 NWM-R2 SWE, precipitation, air temperature, SCAF, and presence or absence of snow with observations from SNOTEL and MODIS. It also compares modeled and observed snowmelt timing. We conclude with a discussion of the uncertainties and limitations in our analysis and present ideas for future work.

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 NWM-R2 inputs and outputs, in-situ measurements, and remotely sensed data from MODIS for water years 2008-2018. NWM-R2 inputs that we used in our analysis were hourly NLDAS2-based precipitation, hourly NLDAS2-based air temperature, and elevation—derived from the 30 m Digital Elevation Model (Zhang et al., 2021)—with 1 km spatial resolution. We used NWM-R2 outputs of 3-

hourly SWE and SCAF with 1 km spatial resolution from the land surface module. We retrieved these inputs and outputs for NWM grid cells containing SNOTEL sites based on the nearest neighbour approach. In-situ measurements comprised daily precipitation, daily air temperature, elevation, and daily SWE from SNOTEL. Remotely sensed MODIS daily snow-covered areas with nominal 500 m spatial resolution were from the MODIS sensor. The model, in-situ, and remotely sensed datasets thus 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).

[Locate Figure 2 near here]

2.1 NWM-R2 Snow Parameterization (Noah-MP) and Snow Reanalysis Products

The NWM-R2 uses a particular configuration of Noah-MP (Table 1) as the land surface model to simulate snow processes as a 1-dimensional vertical column over 1 km spatial resolution grid cells with no representation of any lateral snow processes within a grid cell. Details of the NWM-R2 are given in WRF-Hydro version 5.1.1 documentation (Gochis, Barlage, Cabell, Casali, et al., 2020) and the code (Gochis, Barlage, Cabell, Dugger, et al., 2020). WRF-Hydro version 5.1.1 is the WRF-Hydro version used in NWM-R2. However, (Gochis, Barlage, Cabell, Casali, et al., 2020) does not describe details of the snow parameterization. Instead reference is made to the Noah-MP technical description (Yang et al., 2011) and associated paper (Niu et al., 2011). Here we have summarized key features of the snow parameterization that pertain to the interpretation of our results. The focus in this paper is on NWM-R2 results, practically, amounts to a large-scale test of Noah-MP as configured for use in the NWM.

[Locate Table 1 near here]

2.1.1 Snowfall

The separation of precipitation into rainfall or snowfall is based on Jordan's (1991) algorithm that uses near surface air temperature thresholds [Equations (1-2)].

$$f_{p,ice} = \begin{cases} 1.0 & T_{sfc} \le T_{frz} + 0.5 \\ 1.0 - (-\ 54.632\ +\ 0.2 \times T_{sfc}) & T_{frz} + 0.5 \le T_{sfc} \le T_{frz} + 2.0 \\ 0.6 & T_{frz} + 2.0 \le T_{sfc} \le T_{frz} + 2.5 \\ 0.0 & T_{sfc} > T_{frz} + 2.5 \end{cases} \quad (1)$$

$$rain = P \times \left(1 - f_{p,ice}\right) \quad (2)$$
$$snow = P \times f_{p,ice}$$

where $f_{p,ice}$ is the snow fraction in precipitation, T_{sfc} [K] is the surface air temperature, T_{frz} [273.16 K] is freezing/melting point, and P [mm s^{-1}] is the input precipitation. Freshly fallen snow density (ρ_{fs} [kg/m^3]) is calculated using Equation (3), based on Hedstrom and Pomeroy (1998).

$$\rho_{fs} = \min\left(120, 67.92 + 51.25e^{\left(\frac{T_{sfc}-T_{frz}}{2.59}\right)}\right) \quad (3)$$

2.1.2 Vegetation and Snow Interception

In Noah-MP, a single-layer vegetation canopy model characterizes the fraction covered by vegetation (FVEG) in each model grid cell. 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, the model uses this fraction to adjust the LAI and Stem Area Index (SAI), which are used in the snow interception model. The snow interception model allows for both liquid water and ice to be present on the vegetation canopy; and includes loading/unloading of snowfall, melting of intercepted snow and refreezing of the meltwater, frost/sublimation of canopy-intercepted snow, and dew/evaporation. The model solves the canopy liquid water balance [Equation (4)] and ice balance [Equation (5)] based on Niu and Yang (2004).

$$\frac{\partial M_{liq}}{\partial t} = R_{intr} + (R_{dew} - R_{eva}) + (R_{melt} - R_{frz}) \quad (4)$$

$$\frac{\partial M_{ice}}{\partial t} = (R_{load} - R_{unload}) + (R_{frost} - R_{sub}) + (R_{frz} - R_{melt}) \quad (5)$$

where M_{liq} [kg m^{-2}] is the storage of liquid water in the canopy, and R_{intr} [kg m^{-2} s^{-1}], R_{dew} [kg m^{-2} s^{-1}], and R_{eva} [kg m^{-2} s^{-1}] are interception rate for rain, dew rate, and evaporation rate, respectively. R_{melt} [kg m^{-2} s^{-1}] and R_{frz} [kg m^{-2} s^{-1}] are melting and refreezing rates. M_{ice} [kg m^{-2}] is the storage of ice in the canopy and R_{load} [kg m^{-2} s^{-1}] and R_{unload} [kg m^{-2} s^{-1}] are snow loading and unloading rates, respectively. R_{frost} [kg m^{-2} s^{-1}] and R_{sub} [kg m^{-2} s^{-1}] are frost and sublimation rates. Heat transported by snow and rain to the vegetation canopy layer, the vegetated ground, and non-vegetated ground is also computed; and is used later in the energy balance computation.

2.1.3 Snow-Covered Area and Snow Albedo

Noah-MP calculates SCAF based on snowpack density (ρ_{sno} [kg m^{-3}]), snow depth (h_{sno} [m]) from initial conditions or the previous time step, snow surface roughness length ($z_{0,g}$ [m]), density of fresh snow (ρ_{new} [kg m^{-3}]), and a dimensionless area-depth factor (m) that determines the curve relating SCAF and snow depth [Equation (6)] as developed by Niu and Yang (2007).

$$SCAF = tanh\left(\frac{h_{sno}}{2.5 z_{0,g}\left(\frac{\rho_{sno}}{\rho_{new}}\right)^{m}}\right), \qquad \rho_{sno} = \frac{SWE}{h_{sno}} \tag{6}$$

In NWM-R2 calculations of snow-covered area, ρ_{new} and $z_{0,g}$ are constants set equal to 100 kg m^{-3} and 0.002 m, respectively. However, the factor m is among the parameters that are adjusted during calibration to minimize differences between modeled and observed streamflow over calibration watersheds (Lahmers et al., 2019; 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. SCAF is used to weight the ground emissivity and ground surface resistance. It also affects the computed snow surface albedo that is modeled using the Biosphere-Atmosphere Transfer Scheme (BATS). BATS (Yang & Dickinson, 1996) models direct and diffusive radiation in visible and near-infrared bands separately accounting for fresh snow albedo, snow age, grain size growth, impurity, and solar zenith angle.

2.1.4 Surface Energy Balance, Radiation, and Momentum Fluxes

Shortwave radiation is modeled over the entire grid cell using a modified two-stream approximation (Niu & Yang, 2004) 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 (Niu et al., 2011). 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, Barlage, et al., 2014). Precipitation advected heat is also computed separately for the canopy vegetation, vegetated ground surface, and non-vegetated ground surface. The vegetation canopy temperature (T_v), the vegetated ground surface temperature ($T_{g,v}$), and the non-vegetated ground surface temperature ($T_{g,b}$) are estimated using the Newton-Raphson method with 20 iterations. If the snow depth is greater than a specified snow depth ($\geq$ 0.05 m) and the ground

surface temperature ($T_{g,v}$/$T_{g,b}$) is greater than the freezing point (273.16 K), the ground temperature is updated to $(1-SCAF) \times T_g + SCAF \times T_{frz}$, and all turbulent fluxes are reevaluated. Finally, these radiative and turbulent fluxes are then aggregated based on the vegetated fraction (FVEG) parameter.

2.1.5 Snowpack Vertical Discretization and Snow Thermal Properties

The Noah-MP snow module uses up to three snow layers, depending on depth (from initial conditions or the last time step). The state variables for each layer are the mass of liquid water, mass of ice, layer thickness, and layer temperature. Snow can also exist in the model without being represented by explicit snow layers. This occurs when the total snowpack thickness is less than a specified minimum snow depth (< 0.025 m). In this case, the only state variable is the mass of snow.

Snow thermal properties including partial volume of ice, partial volume of liquid water, effective porosity, bulk density [based on Lynch-Stieglitz (1994)], volumetric specific heat, and thermal conductivity are computed for each snow layer [Equations (7-12)]. Energy for phase change (melting/refreezing) is also computed for each layer.

$$\theta_{ice,i} = \frac{Mass_{ice,i}}{\Delta Z_i \times \rho_{ice}} \quad (7)$$

$$\theta_{e,i} = 1 - \theta_{ice,i} \quad (8)$$

$$\theta_{liquid,i} = \min\left(\theta_{e,i}, \frac{Mass_{liquid,i}}{\Delta Z_i \times \rho_{water}}\right) \quad (9)$$

$$\rho_{snow,i} = \frac{Mass_{ice,i} + Mass_{liquid,i}}{\Delta Z_i} \quad (10)$$

$$C_{v,i} = C_{ice} \times \theta_{ice,i} + C_{liquid} \times \theta_{liquid,i} \quad (11)$$

$$k_i = 3.2217 \times 10^{-6} \times {\rho_{snow,i}}^2 \quad (12)$$

where $\theta_{ice,i}$ [m^{-3}/m^{-3}] is partial volume ice of snow layer *i*, $Mass_{ice,i}$ [kg m^{-2}] is snow ice mass of snow layer *i*, ΔZ_i [m] is the snow layer thickness of snow layer *i*, ρ_{ice} [917 kg m^{-3}] is ice density, $\theta_{e,i}$ [m^{-3}/m^{-3}] is the effective porosity of snow layer *i*, $\theta_{liquid,i}$ [m^{-3}/m^{-3}] is partial volume of liquid water of snow layer *i*, $Mass_{liquid,i}$ [kg m^{-2}] is liquid water mass of snow layer *i*, ρ_{water} [1000 kg m^{-2}] is liquid water density, $\rho_{snow,i}$ [kg/m^{-3}] is bulk density of snow layer *i*, $C_{v,i}$ [J m^{-3} K^{-1}] is volumetric specific heat of snow layer *i*, C_{ice} [2.094×10^6 J m^{-3} K^{-1}] is specific heat capacity of ice, C_{liquid} [4.188×10^6 J m^{-3} K^{-1}] is specific heat capacity of liquid water, and k_i [W m^{-1} K^{-1}] is thermal conductivity of snow layer *i*.

Heat flux between layers is calculated based on temperature gradient and thermal conductivity, and then this is used to update layer temperatures using a semi-implicit

numerical scheme. When heat flux calculations result in temperatures of snow layers greater than freezing, the excess energy is used to adjust (melt or freeze) liquid water present. The change in the density of the snow with time due to destructive metamorphism, the weight of the overlying layers of snow, and melting (which dictates layer thickness) is modeled, following Anderson (1976) as a function of snow temperature (Niu et al., 2011).

2.1.6 Snow Water Equivalent and Snow Depth

The change in SWE is balanced by the input snowfall (Q_{snow}) reaching the surface in forms of drip and throughfall; and output snowmelt (M), snow sublimation, and snow frost [both expressed as E in Equation (13)].

$$\frac{dSWE}{dt} = Q_{snow} - M - E \tag{13}$$

When new snowfall occurs in a time step, the snow depth and snow ice are increased based on the snow depth increasing rate and the input snowfall rate (both outputs of the snow interception module), respectively. After the depth, phase change and compaction calculations, the number of snow layers is adjusted by either combining the neighbour layers or subdividing them following Jordan (1991). If rainfall (in terms of drip and throughfall) occurs, it is added to the liquid water of the snow layer. The liquid water movement within a snow layer is added to the underlying snow layer when the liquid water content within a snow layer exceeds the layer's liquid water-holding capacity for snowpack (0.03 m^3/m^3). Finally, the liquid water of the snow layer updates after the water flows out of the layer.

2.1.7 Post-processing NWM-R2 Snow Reanalysis Products

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) for grid cells containing SNOTEL sites based on the nearest neighbour approach [code available at Garousi-Nejad and Tarboton (2021d)] from the NOAA Google Cloud archive using a Jupyter Notebook [code available at Tarboton and Garousi-Nejad (2021)]. Then, we averaged 3-hourly results to daily values [code available at Garousi-Nejad and Tarboton (2021f)] 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 hourly precipitation, hourly air temperature, and elevation input data used for NWM-R2 simulations for the selected grid cells. The WRF-Hydro team at NCAR provided precipitation and air temperature values for us as those data were not

306 available on the Google Cloud archive. Then, we computed daily precipitation and the daily
307 average temperature [code available at Garousi-Nejad and Tarboton (2021f)].

308 **2.2 SNOTEL**

309 SNOTEL stations, managed by the Natural Resources Conservation Service (NRCS),
310 generally consist of a snow pillow, an air temperature sensor, and a storage precipitation
311 gage. Our study used the daily precipitation, air temperature, and SWE values measured at
312 SNOTEL sites as a reference dataset to evaluate the NWM-R2 precipitation, air temperature,
313 and SWE. We realize that SNOTEL data must be used with some caution because the sites
314 are mostly located in small clearings within forests protected by forest canopies, leading to
315 differences in exposure to wind and radiation (McCreight et al., 2014). Furthermore,
316 SNOTEL data do not undergo a high correction level (Swenson & Lawrence, 2012). In some
317 instances, we found unrealistically high temperature values that needed to be filtered out.
318 Nevertheless, SNOTEL data remain the only widespread in situ SWE observations available
319 for model validation in the Western U.S. (Barlage et al., 2010; Clow et al., 2012; Livneh et
320 al., 2010; Pan et al., 2003; Toure et al., 2016). We automated retrieval of the SNOTEL data
321 by calling its Consortium of Universities for the Advancement of Hydrologic Science, Inc
322 (CUAHSI) web service from a Jupyter Notebook script (Garousi-Nejad & Tarboton, 2021c).

323 **2.3 MODIS**

324 The National Aeronautics and Space Administration (NASA)'s MODIS instrument
325 launched aboard the Terra satellite in late 1999 is designed to observe and monitor Earth
326 changes, such as snow cover. MODIS has spectral bands in the visible and near-infrared
327 regions, nominal 500 m spatial resolution, and near-daily global coverage. The daily snow-
328 cover gridded tile product, MOD10A1, has been used and improved over time in multiple
329 snow studies (Aalstad et al., 2020; Bennett et al., 2019; Magand et al., 2014; Masson et al.,
330 2018; Salomonson & Appel, 2006; Swenson & Lawrence, 2012). We used products from the
331 current version of the MODIS snow-cover algorithm which is the collection 6 suite of
332 MODIS (hereafter referred to as MODIS-C6, or just MODIS). We chose to use MODIS-C6
333 (Hall & Riggs, 2016) as a reference to evaluate NWM-R2 SCAF because the
334 improvements/revisions to MODIS-C6 (i.e., accounting for the surface temperature and
335 surface height) led to a notable increase in accuracy of snow cover detection on mountain
336 ranges and low illumination conditions in the Northern Hemisphere during spring and
337 summer (Riggs et al., 2017).

338 The MODIS-C6 snow algorithm is designed to detect snow cover based on the
339 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 products include this ratio, the Normalized Difference Snow Index (NDSI, product name: NDSI_Snow_Cover) rather than snow cover. This approach allows users to have the option to estimate snow cover using the global empirical model [Equation (14)] or develop region-specific models (Riggs et al., 2016). In this study, we developed a script (Garousi-Nejad & Tarboton, 2021b) run in Google Earth Engine to retrieve NDSI_Snow_Cover for each NWM grid cell containing a SNOTEL site. Since MODIS output is available on a 500 m grid and NWM grid cells are 1 km in size, the script averaged NDSI_Snow_Cover over the four MODIS grid cells that have their centroid within the NWM grid cell (Figure 2). 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, which we masked out in Google Earth Engine. The returned MODIS images thus have spatial gaps due to this 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. (Riggs et al., 2016) to compute MODIS SCAF from NDSI_Snow_Cover values [Equation (14)].

$$\mathrm{SCAF} = \min[\max(-0.01 + 1.45 \times \mathrm{NDSI}, 0), 1] \qquad \text{where } \mathrm{NDSI} \in [0,1] \tag{14}$$

In Equation (14), the MODIS SCFA is always estimated as 1 for NDSI values equal or greater than 0.7, and it changes linearly for NDSI values between 0 to 0.7.

The resulting dataset includes 2,504,102 site-days in the period of overlap between NWM-R2 and SNOTEL data [data and code used to aggregate it are available at Garousi-Nejad and Tarboton, (2021e)]. 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.

Seasonal

- **First day of the month comparisons** were used for NWM-R2 SWE/SCAF (modeled) versus SNOTEL SWE and MODIS SCAF (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.

Snow Water Equivalent and Snow-Covered Area at peak SWE

- Modeled and observed SWE and SCAF were compared on the date of observed peak SWE (**same day comparison**).
- Modeled and observed peak SWE do not necessarily occur on the same date. We compared both SWE and SCAF on the separate dates where peak SWE was modeled and observed (**different day comparison**).
- Model input and SNOTEL observed total precipitation accumulated from the start of the water year, Oct 1, to the date of peak SWE were also compared.

Total precipitation was 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.

Direct (binary) comparison of snow presence or absence

- **Full snow cover**. Daily modeled SCAF taken as full snow if SCAF is ≥ 0.95. Daily MODIS inferred (observed) SCAF taken as full snow if NDSI is ≥ 0.7.
- **Some snow cover**. Daily SCAF taken as indicating some snow if modeled SCAF, or MODIS NDSI > 0.3.

First, we classified the snow presence or absence grid cells based on these thresholds. We then counted the number of classified grid cells for both observed and modeled datasets for each date. This was done only for grid cells locations where SNOTEL sites exist, because

our scripts extracting NWM output were only run at these locations and running for all grid cells across the Western U.S. was computationally prohibitive.

- **Presence Absence comparison metrics** were used to indicated the degree-of-overlap between modeled and observed datasets (Horritt & Bates, 2002; Sangwan & Merwade, 2015).

The correctness metric [Equation (15)] compares the total number of modeled and observed grid cells having some or full snow cover, while the fit metric [Equation (16)] quantifies whether modeled and observed locations match, scaled by the total area mapped with snow (either full or some).

$$C_t = \frac{Modeled_{snow}}{Observed_{snow}} \tag{15}$$

$$F_t = \frac{Modeled_{snow} \cap Observed_{snow}}{Modeled_{snow} \cup Observed_{snow}} \tag{16}$$

where C_t and F_t are correctness and fit metrics computed for date *t*, respectively, and $Modeled_{snow}$ and $Observed_{snow}$ are grid cells classified as snowy cells on that date. Correctness (C_t) and Fit (F_t) should both ideally be 1 (100%).

To account for the fact that MODIS may be interpreting vegetated grid cells as snow free and thus underestimating the snow cover (Steele et al., 2017; X. Wang et al., 2017), while NWM-R2 may have snow beneath the vegetation canopy, and that SNOTEL sites are often in openings much smaller than the cell size (1 km) in generally forested areas, we requested, and obtained from the NRCS (the agency that operates SNOTEL) a list of sites in generally open areas. We report separate metrics for these sites reported to be open. The NRCS indicated that SNOTEL sites may be open due to canopy disturbance caused by pine bark beetle damage and fire, which may have occurred during the study period, resulting in some uncertainty as to sites being open early on.

Melt timing

- **Half melt from peak SWE date** (Clow, 2010).

The 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).

Commonly used statistics

- Coefficient of determination [r^2, Equation (17)] 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 (18)], 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 (19)], a measure of how concentrated the data are around the line of best fit;
- Nash Sutcliffe efficiency [*NSE*, Equation (20)], 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 (21)], the average of the difference between modeled and observed.

$$r^2 = \left[\frac{\sum_{t=1}^{N} (M_t - \overline{O}_t)(M_t - \overline{M}_t)}{\sqrt{\sum_{t=1}^{N} (O_t - \overline{O}_t)^2 \sum_{t=1}^{N} (M_t - \overline{M}_t)^2}} \right]^2 \quad (17)$$

$$Spearmanr = 1 - \frac{6 \sum_{t=1}^{N} d_t^2}{N(N^2 - 1)} \quad (18)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} (O_t - M_t)^2}{N}} \quad (19)$$

$$NSE = 1 - \frac{\sum_{t=1}^{N} (O_t - M_t)^2}{\sum_{t=1}^{N} (O_t - M_t)^2} \quad (20)$$

$$Bias = \frac{\sum_{t=1}^{N} (M_t - O_t)}{N} \quad (21)$$

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.

[Locate Figure 3 near here]

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).

[Locate Figure 4 near here]

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.

[Locate Figure 5 near here]

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.

[Locate Figure 6 near here]

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 (U.S. Department of Agriculture, 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.

[Locate Figure 7 near here]

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.

[Locate Figure 8 near here]

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.

[Locate Figure 9 near here]

[Locate Figure 10 near here]

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.

[Locate Figure 11 near here]

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.

[Locate Figure 12 near here]

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

550 0.9 computed over the full study period. This subset shows a reduced bias (compared to the
551 entire dataset) between the observed and modeled precipitation accumulated from Oct 1 to
552 peak observed SWE date (Figure 13a).
553 [Locate Figure 13 near here]

554 **3.3 Direct (Binary) Comparison of Snow Presence or Absence**

555 The cell by cell binary comparison of snowy grid cells at SNOTEL sites shows that
556 this comparison does not work well for the all-snow-present condition, i.e., when the
557 observed and modeled SCAF thresholds were 0.7 and 0.95, respectively (Figure 14a). We
558 observed that the average *C* for the entire period of study was 9.4 and average F, 0.11. These
559 are poor degree of overlap statistics, and are due to the fact that MODIS never reports more
560 than about 30% of the area as having full snow.
561 [Locate Figure 14 near here]

562 However, the cell by cell binary evaluation for some snow present resulted in better
563 degree of overlap statistics (Figure 14b, $\bar{C}$=1.47 and $\bar{F}$=0.50). Discrepancies between the
564 modeled and observed snowy grid cells as implied by average *C* (=1.20) and *F* (=0.64) were
565 even less when we only focused on the 62 SNOTEL sites (about 8% of all sites) reported as
566 open (Figure 14c). Table 2 summarizes fit metrics for the snow cover binary comparison.
567 [Locate Table 2 near here]

568 **3.4 Melt Timing Comparison**

569 For 68% of the site years analyzed, the modeled half melt date was earlier than
570 observed. When further classified based on whether modeled half melt dates were close,
571 ahead, behind or far apart from observed melt dates (Figure 15a) we observe that the NWM
572 half melt date was greater than 20 days from observed half melt date, for 34% of the site
573 years, and off by 6 days or more for 75% of site years. For those site years where the
574 difference was between 5 and 20 days, a greater percentage had the model melting ahead,
575 than behind the observed. The site years that have modeled half melt date ahead of observed
576 tend to have lower modeled half melt date SWE (which is by definition half the peak SWE)
577 than observed (Figure 15b).
578 [Locate Figure 15 near here]

579 **4. DISCUSSION**

580 The seasonal pattern of SWE and SCAF averaged across all SNOTEL site-years
581 shows that NWM generally under-estimates SWE and over-estimates SCAF relative to
582 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. The spatial scale differences in different datasets is a source of uncertainty in this analysis. A point-scale measurement of SWE cannot with confidence validate the NWM-R2 grid cell value with nearest center, particularly in forest regions (McCreight et al., 2014). 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. In the cell by cell comparison between NWM-R2 and MODIS-C6 datasets, the mean value of MODIS grid cells would be different if using a different number of cells, e.g. nine grid cells instead of four.

Precipitation discrepancies suggest that SWE differences are partly due to discrepancies between observed precipitation (SNOTEL) and model input precipitation [adjusted NLDAS-2 (RafieeiNasab et al., 2020)]. 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, SNOTEL data do not undergo a high correction level (Swenson & Lawrence, 2012). It was not uncommon to see accumulated precipitation less than SWE at SNOTEL sites (notably for stations at higher elevations), which could be due to either precipitation under-catch, or inflated SWE (Meyer et al., 2012). This makes using this information for model comparison challenging, as the model cannot accumulate more snow than its precipitation input. This is an unresolvable difference and should be recognized as a source of uncertainty associated with the in-situ measurements used in this study.

Our results show a cold (downward) bias for the model input air temperature (based on NLDAS-2) 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, a point which needs more investigation.

The discrepancies in model inputs (precipitation and air temperature in this study) are not the only potential sources for SWE differences. Even at 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 that lead to SWE under-modeling, both due to observation and model errors. Errors in SWE measurements may occur, due to factors such as wind causing snowdrifts on the snow pillow (Meyer et al., 2012), or the small clearing SNOTEL site location not being representative of larger scale snowpack (McCreight et al., 2014). In the NWM land surface model (Noah-MP), the partitioning of precipitation into rainfall and snowfall, which is one of the most sensitive parameterizations in simulating cold-region hydrological processes (Loth et al., 1993), is based on Jordan's (1991) algorithm, which ignores some physical processes controlling precipitation phase by not incorporating humidity. This may lead to biases in SWE, snow depth, and snow cover fraction (Chen, Liu, et al., 2014; Harder & Pomeroy, 2014; Y. Wang et al., 2019). Y. 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 et al., 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,

651 and 12—may reflect the fact that the MODIS and NWM SCAF quantities are not really the
652 same thing. MODIS may be interpreting vegetation as snow free (Steele et al., 2017; X.
653 Wang et al., 2017), while NWM has snow beneath vegetation. In NWM-R2 results, the
654 persistent low and high SCAF (<0.1 and >0.9, respectively) reflects that NWM treats SCAF
655 as a binary metric in mountainous regions. NWM-R2 SCAF values stay near 1 with less
656 variability between Dec-Apr for more than 70% of cases. This suggests that once the NWM
657 grid cell (1 km spatial resolution) is more than 90% snow-covered, it is implausible for it to
658 diverge from 1 for the rest of the accumulation phase and early ablation phase. One possible
659 reason for this behaviour is the lack of representation of some factors affecting SCAF such as
660 vegetation type and seasonal change, and topography. These limitations affect the accurate
661 simulation of SCAF and SWE (Helbig et al., 2015; Magand et al., 2014; Swenson &
662 Lawrence, 2012; Wrzesien et al., 2015). Another possible reason for some of the differences
663 is the lack of any representation of snow drifting processes (i.e., wind-driven redistribution of
664 snow) in the snow model. Snow drifting increases the variability of snow depth within a grid
665 cell, which then, when melting starts leads to intervening (non-binary 0 or 1) snow covered
666 area fractions. This may be a factor contributing to differences in regions with modeled
667 SCAF less than 10% while the observed SCAF are more than 50% (points along the
668 horizontal axis of SCAF on March 1, April 1, and May 1 in Figure 4).
669 We recognize that the SCAF mapped from MODIS in this study also has uncertainties
670 and limitations. First, the temporal forward filling approach that we used to fill gaps
671 associated with clouds may miss some of the daily variability of snow cover, particularly in
672 mountainous regions. Second, the parameters of Equation (14), which estimates SCAF from
673 MODIS-C6 NDSI_Snow_Cover product, were those from Salomonson and Appel (2006) and
674 were constant for our entire study region. Adjusting these parameters to improve the snow
675 cover products from MODIS regionally has been suggested (Riggs et al., 2017). Third,
676 MODIS NDSI_Snow_Cover grids (nominally 500 m) were averaged for 1 km NWM grid
677 cells, using an unweighted approach in the Google Earth Engine platform. This approach
678 selects MODIS grids whose centers fall within the target area (i.e., NWM grid cells). These
679 scale differences may be a further source of uncertainty, compounded by the nonlinearity in
680 Equation (14) [plateau at NDSI > 0.7] having an impact on SCAF from averaged NDSI.
681 Results for the direct (binary) comparison of full snow cover were poor as MODIS
682 never reports more than about 30% of the area as having full snow, while the degree-of-
683 overlap between the modeled and observed results, in terms of average *C* and *F*, improved
684 considerably when comparing cells having some snow present. We interpret this as a

shortcoming of MODIS for this sort of comparison, perhaps due to the presence of vegetation. MODIS SCAF estimates may not account for snow beneath the canopy due to incapability of the sensor to see forest gaps (the snow-covered ground) through the vegetation canopy (Steele et al., 2017; X. Wang et al., 2017), while the NWM-R2 land surface model (Noah-MP) may estimate snow under the vegetation canopy in these locations. Our results show that discrepancies between modeled and observed snowy grid cells reduce when we focus only on the SNOTEL sites reported as open. For full snow present average *C* improves from 9.41 to 6.18 while average *F* improves from 0.11 to 0.16. These are still poor, but less poor. For some snow present average *C* improves from 1.47 to 1.2 and average *F* improves from 0.5 to 0.64, making them reasonably respectable, in comparison to the ideal values of 1. This suggests that forest vegetation is a dominant contributor to the disagreement between model and MODIS observed snow cover.

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 for the NWM-R2 configuration 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 are 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 is 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 as it lacks representation of vegetation and topography while observations are much more fractional. They may not reflect the same physical quantity. The binary comparison of full snow presence reveals that the degree-of-overlap between the modeled and observed results still remains poor, which is possibly due to uncertainties associated with MODIS observations in

vegetated areas. Results of the binary comparison of some snow presence improves when we focus only on the SNOTEL sites reported as open (average C=1.2 and average F=0.64). Also, 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 there is disagreement between the NWM SCAF (estimated by the Noah-MP module) and MODIS observations in the binary comparison, the SCAF parameterization should be improved or can be inferred from satellites while considering the uncertainties associated with these products. 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 specific settings of the Noah-MP as implemented in the NWM-R2 configuration and suggests where potential development effort should be directed in the future. It would also be helpful, for future work, to have a more comprehensive observation data set, beyond the SNOTEL sites, such as possibly Critical Zone Observatory (CZO) or experimental forest sites, that include snowfall/rainfall measurements, canopy snow interception, turbulence and radiation fluxes above and below the canopy. Another opportunity is to run the model at higher resolution which would involve downscaling the forcing inputs to higher resolution. Higher-resolution remotely sensed snow-covered area (e.g., from LANDSAT satellite) could then be used for model evaluation.

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 prepared by the WRF-Hydro NCAR team, we have made available on HydroShare for reproducibility (Garousi-Nejad & Tarboton, 2021f). The NWM elevation dataset is 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 as a collection on HydroShare (Garousi-Nejad & Tarboton, 2021a) comprised of:

- Input data and code to get the indices of the NWM grid cells containing SNOTEL sites (Garousi-Nejad & Tarboton, 2021d)
- Input data, code to retrieve the NWM-R2 inputs and outputs at SNOTEL sites (Tarboton & Garousi-Nejad, 2021)
- Input data, code and output from post-processing the retrieved NWM-R2 inputs and outputs at SNOTEL sites (Garousi-Nejad & Tarboton, 2021f)
- Input data and code to retrieve precipitation, air temperature, and SWE measurements at SNOTEL sites (Garousi-Nejad & Tarboton, 2021c)
- Input data and Google Earth Engine code to retrieve averaged MODIS-C6 NDSI snow cover at SNOTEL sites (Garousi-Nejad & Tarboton, 2021b)
- Input data, code and output from combining the NWM inputs and outputs with observations form SNOTEL and MODIS at SNOTEL sites (Garousi-Nejad & Tarboton, 2021e)
- Input data, code and output used to produce Figures 1-4 and Figures 6-15 (Garousi-Nejad & Tarboton, 2021g)

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

Anderson, E. A. (1976). *A point energy and mass balance model of a snow cover* (NOAA Technical Report NWS 29). National Weather Service. https://repository.library.noaa.gov/view/noaa/6392

Barlage, M., Chen, F., Tewari, M., Ikeda, K., Gochis, D., Dudhia, J., Rasmussen, R., Livneh, B., Ek, M., & Mitchell, K. (2010). Noah land surface model modifications to improve

779 snowpack prediction in the Colorado Rocky Mountains. *Journal of Geophysical*
780 *Research*, *115*(D22), D22101. https://doi.org/10.1029/2009JD013470
781 Bennett, K. E., Cherry, J. E., Balk, B., & Lindsey, S. (2019). Using MODIS estimates of
782 fractional snow cover area to improve streamflow forecasts in interior Alaska.
783 *Hydrology and Earth System Sciences*, *23*(5), 2439–2459.
784 https://doi.org/10.5194/hess-23-2439-2019
785 Bhatti, A. M., Koike, T., & Shrestha, M. (2016). Climate change impact assessment on
786 mountain snow hydrology by water and energy budget-based distributed hydrological
787 model. *Journal of Hydrology*, *543*, 523–541.
788 https://doi.org/10.1016/j.jhydrol.2016.10.025
789 Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin,
790 C., Miguez-Macho, G., Niu, G.-Y., Wen, L., & Yang, Z.-L. (2014). Modeling
791 seasonal snowpack evolution in the complex terrain and forested Colorado
792 Headwaters region: A model intercomparison study. *Journal of Geophysical*
793 *Research: Atmospheres*, *119*(24), 13,795-13,819.
794 https://doi.org/10.1002/2014JD022167
795 Chen, F., Liu, C., Dudhia, J., & Chen, M. (2014). A sensitivity study of high-resolution
796 regional climate simulations to three land surface models over the western United
797 States: SENSITIVITY STUDY OF LSMS IN WRF. *Journal of Geophysical*
798 *Research: Atmospheres*, *119*(12), 7271–7291. https://doi.org/10.1002/2014JD021827
799 Clow, D. W. (2010). Changes in the timing of snowmelt and streamflow in Colorado: A
800 response to recent warming. *Journal of Climate*, *23*(9), 2293–2306. USGS
801 Publications Warehouse. https://doi.org/10.1175/2009JCLI2951.1
802 Clow, D. W., Nanus, L., Verdin, K. L., & Schmidt, J. (2012). Evaluation of SNODAS snow
803 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

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

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

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

Garousi-Nejad, I., & Tarboton, D. (2021d). *Notebook to get the indices of National Water Model V2.0 grid cells containing SNOTL sites*. HydroShare. https://www.hydroshare.org/resource/7839e3f3b4f54940bd3591b24803cacf/

Garousi-Nejad, I., & Tarboton, D. (2021e). *Notebooks for combining the National Water Model results/inputs with observations from SNOTEL and MODIS at SNOTEL sites.* HydroShare. https://www.hydroshare.org/resource/493e0ad05c2d45199427cc41a6c76de0/

Garousi-Nejad, I., & Tarboton, D. (2021f). *Notebooks for post-processing the retrieved National Water Model V2.0 Retrospective run results and inputs at SNOTEL sites*.

HydroShare.
https://www.hydroshare.org/resource/1b66a752b0cc467eb0f46bda5fdc4b34/

Garousi-Nejad, I., & Tarboton, D. (2021g). *Notebooks for visualizations reported at A Comparison of National Water Model Retrospective Analysis Snow Outputs at SNOTEL Sites Across the Western U.S.* HydroShare.
https://www.hydroshare.org/resource/8507aa41130e45bfb0752026cf2253ab/

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., Barlage, M., Cabell, R., Casali, M., Dugger, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., & 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

Gochis, D., Barlage, M., Cabell, R., Dugger, A., Fanfarillo, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Frazier, N., Johnson, D., Mattern, J. D., Karsten, L., Mills, T. J., & Fersch, B. (2020). *WRF-Hydro® v5.1.1* (v5.1.1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.3625238

Hall, D. K., & Riggs, G. A. (2016). *MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid* [Data set]. 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

Hedstrom, N. R., & Pomeroy, J. W. (1998). Measurements and modelling of snow interception in the boreal forest. *Hydrological Processes*, *12*, 1611–1625. https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11%3C1611::AID-HYP684%3E3.0.CO;2-4

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

Horritt, M. S., & Bates, P. D. (2002). Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology*, *268*(1–4), 87–99. https://doi.org/10.1016/S0022-1694(02)00121-X

Hou, D., charles, M., Luo, Y., Toth, Z., Zhu, Y., Krzysztofowicz, R., Lin, Y., Xie, P., Seo, D.-J., Pena, M., & Cui, B. (2014). Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage IV toward CPC Gauge-Based Analysis. *Journal of Hydrometeorology*, *15*(6), 2542–2557. https://doi.org/10.1175/JHM-D-11-0140.1

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., Goodrich, D., & 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

Lynch-Stieglitz, M. (1994). The Development and Validation of a Simple Snow Model for the GISS GCM. *Journal of Climate*, *7*(12), 1842–1855. http://www.jstor.org/stable/26198671

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

Meyer, J. D. D., Jin, J., & Wang, S.-Y. (2012). Systematic Patterns of the Inconsistency between Snow Water Equivalent and Accumulated Precipitation as Reported by the Snowpack Telemetry Network. *Journal of Hydrometeorology*, *13*(6), 1970–1976. https://doi.org/10.1175/JHM-D-12-066.1

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

926 Naple, P., Letcher, T., Minder, J. R., Gochis, D., Dugger, A., & RafieeiNasab, A. (2020).
927 *Improving parameterizations of snow in the National Water Model with observations*
928 *from the New York State Mesonet to better simulate snow and streamflow in the*
929 *northeastern United States*. AGU Fall Meeting, Virtual.
930 https://ui.adsabs.harvard.edu/abs/2020AGUFMC063.0006N/abstract
931 National Research Council. (2006). *Toward a New Advanced Hydrologic Prediction Service*
932 *(AHPS)* (p. 11598). National Academies Press. https://doi.org/10.17226/11598
933 National Weather Service. (2019). *National-Water-Model-V2*. Google Cloud Bucket.
934 https://console.cloud.google.com/storage/browser/national-water-model-v2
935 Niu, G.-Y., & Yang, Z.-L. (2004). Effects of vegetation canopy processes on snow surface
936 energy and mass balances: CANOPY EFFECTS ON SNOW PROCESSES. *Journal*
937 *of Geophysical Research: Atmospheres*, *109*(D23).
938 https://doi.org/10.1029/2004JD004884
939 Niu, G.-Y., & Yang, Z.-L. (2007). An observation-based formulation of snow cover fraction
940 and its evaluation over large North American river basins. *Journal of Geophysical*
941 *Research*, *112*(D21), D21101. https://doi.org/10.1029/2007JD008674
942 Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A.,
943 Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community
944 Noah land surface model with multiparameterization options (Noah-MP): 1. Model
945 description and evaluation with local-scale measurements. *Journal of Geophysical*
946 *Research*, *116*(D12), D12109. https://doi.org/10.1029/2010JD015139
947 Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the Conterminous United States:
948 Evolution of a Hierarchical Spatial Framework. *Environmental Management*, *54*(6),
949 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., Robock, A., Lohmann, D., Cosgrove, B., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., & 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., Yates, D., Sampson, K., McCreight, J., Read, L., Zhang, Y., & McAllister, M. (2020). *Overview of National Water Model calibration general strategy & optimization, NCAR Community WRF-Hydro Modeling System training workshop*. NCAR Community WRF-Hydro Modeling System training workshop. 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

975 Sangwan, N., & Merwade, V. (2015). A Faster and Economical Approach to Floodplain
976 Mapping Using Soil Information. *JAWRA Journal of the American Water Resources*
977 *Association*, *51*(5), 1286–1304. https://doi.org/10.1111/1752-1688.12306
978 Shamir, E., Carpenter, T. M., Fickenscher, P., & Georgakakos, K. P. (2006). Evaluation of
979 the National Weather Service Operational Hydrologic Model and Forecasts for the
980 American River Basin. *Journal of Hydrologic Engineering*, *11*(5), 392–407.
981 https://doi.org/10.1061/(ASCE)1084-0699(2006)11:5(392)
982 Steele, C., Dialesandro, J., James, D., Elias, E., Rango, A., & Bleiweiss, M. (2017).
983 Evaluating MODIS snow products for modelling snowmelt runoff: Case study of the
984 Rio Grande headwaters. *International Journal of Applied Earth Observation and*
985 *Geoinformation*, *63*, 234–243. https://doi.org/10.1016/j.jag.2017.08.007
986 Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2004). Changes in Snowmelt Runoff Timing
987 in Western North America under a `Business as Usual' Climate Change Scenario.
988 *Climatic Change*, *62*(1–3), 217–232.
989 https://doi.org/10.1023/B:CLIM.0000013702.22656.e8
990 Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward Earlier Streamflow
991 Timing across Western North America. *Journal of Climate*, *18*(8), 1136–1155.
992 https://doi.org/10.1175/JCLI3321.1
993 Sun, N., Yan, H., Wigmosta, M. S., Leung, L. R., Skaggs, R., & Hou, Z. (2019). Regional
994 Snow Parameters Estimation for Large-Domain Hydrological Applications in the
995 Western United States. *Journal of Geophysical Research: Atmospheres*, *124*(10),
996 5296–5313. https://doi.org/10.1029/2018JD030140
997 Swenson, S. C., & Lawrence, D. M. (2012). A new fractional snow-covered area
998 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. (2021). *Notebook for retrieval of National Water Model V2.0 Retrospective run results at SNOTEL sites*. HydroShare. https://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., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., & 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

U.S. Department of Agriculture. (2011). *Snow Survey and Water Supply Forecasting. National Engineering Handbook Part 622*. Water and Climate Center, Natural Resources Conservation Service. directives.sc.egov.usda.gov/viewerFS.aspx?hid=32040

Viterbo, F., Mahoney, K., Read, L., Salas, F., Bates, B., Elliott, J., Cosgrove, B., Dugger, A., Gochis, D., & 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, X., Zhu, Y., Chen, Y., Zheng, H., Liu, H., Huang, H., Liu, K., & Liu, L. (2017). Influences of forest on MODIS snow cover mapping and snow variations in the Amur River basin in Northeast Asia during 2000-2014. *Hydrological Processes*, *31*(18), 3225–3241. https://doi.org/10.1002/hyp.11249

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

Yang, Z.-L., & Dickinson, R. E. (1996). Description of the Biosphere-Atmosphere Transfer Scheme (BATS) for the Soil Moisture Workshop and evaluation of its performance. *Global and Planetary Change*, *13*(1–4), 117–134. https://doi.org/10.1016/0921-8181(95)00041-0

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research*, *116*(D12), D12110. https://doi.org/10.1029/2010JD015140

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

Zhang, J., Condon, L. E., Tran, H., & Maxwell, R. M. (2021). A national topographic dataset for hydrological modeling over the contiguous United States. *Earth System Science Data*, *13*(7), 3263–3279. https://doi.org/10.5194/essd-13-3263-2021

TABLE

Table 1. The Noah-MP land surface model options as defined in the National Water Model version 2.0 retrospective analysis configuration †

Code Name	Long Name	Physics Option Used
DYNAMIC_VEG_OPTION	Dynamic vegetation	4: Using monthly LAI is prescribed for various vegetation types
CANOPY_STOMATAL_RESISTANCE_OPTION	Canopy stomatal resistance	1: Ball-Berry
BTR_OPTION	Soil moisture factor for stomatal resistance	1: Noah type using soil moisture
RUNOFF_OPTION	Runoff and groundwater	3: Noah type surface and subsurface runoff (free drainage)
SURFACE_DRAG_OPTION	Surface layer drag coefficients	1: Monin-Obukhov
FROZEN_SOIL_OPTION	Frozen soil permeability	1: Using the total soil moisture to compute hydraulic properties
SUPERCOOLED_WATER_OPTION	Supercooled liquid water (or ice fraction)	1: No iteration (Form of the freezing-point depression equation)
RADIATIVE_TRANSFER_OPTION	Radiation transfer	3: Two-stream applied to vegetated fraction
SNOW_ALBEDO_OPTION	Ground snow surface albedo	2: BATS
PCP_PARTITION_OPTION	Partitioning precipitation into rainfall & snowfall	1: Jordan (1991)

†Based on Gochis, Barlage, Cabell, Casali, et al. (2020) and Gochis, Barlage, Cabell, Dugger, et. (2020).

1061 Table 1. (continued).

Code Name	Long Name	Physics Option Used
TBOT_OPTION	Lower boundary condition of soil temperature	2: TBOT at ZBOT (8m) read from a file
TEMP_TIME_SCHEME_OPTION	Snow/soil temperature time scheme (only layer 1)	3: Semi-implicit; flux top boundary condition, but FSNO for TS calculation
GLACIER_OPTION	Glacier treatment	2: Ice treatment more like original Noah
SURFACE_RESISTANCE_OPTION	Surface resistant to evaporation and sublimation	4: For non-snow; rsurf = rsurf_snow for snow (set in MPTABLE)

1062 †Based on Gochis, Barlage, Cabell, Casali, et al. (2020) and Gochis, Barlage, Cabell, Dugger,
1063 et. (2020).

Table 2. Summary of average correctness ($\bar{C}$) and average fit ($\bar{F}$) metrics evaluated for the binary comparison of snow presence or absence when considering (a) all SNOTEL sites and (b) sites reported as open approaches.

Average metrics	**(a) All 734 SNOTEL sites**		**(b) The 62 SNOTEL sites reported as open**	
	Snow Presence Condition		**Snow Presence Condition**	
	Full‡	Some§	Full‡	Some§
$\bar{C}$	9.41	1.47	6.18	1.20
$\bar{F}$	0.11	0.50	0.16	0.64

‡Daily modeled snow-covered area fraction (SCAF) taken as full snow if SCAF is ≥ 0.95. Daily MODIS SCAF taken as full snow if NDSI is ≥ 0.7.

§Daily modeled SCAF taken as some snow if SCAF is ≥ 0.3. Daily MODIS SCAF taken as some snow if NDSI is ≥ 0.3.

FIGURE

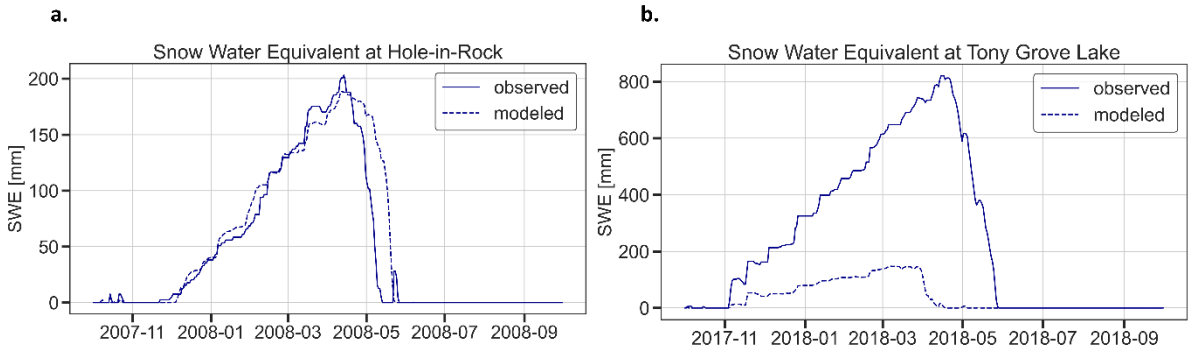

Figure 1. Snow water equivalent from the NWM version 2.0 reanalysis (NWM-R2) dataset compared to in-situ observations at two SNOTEL sites in Utah. (a) Hole-in-Rock site (ID: 528) located at 2794 m elevation for the water year 2008. (b) Tony Grove Lake site (ID: 823) located at 2582 m elevation for the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

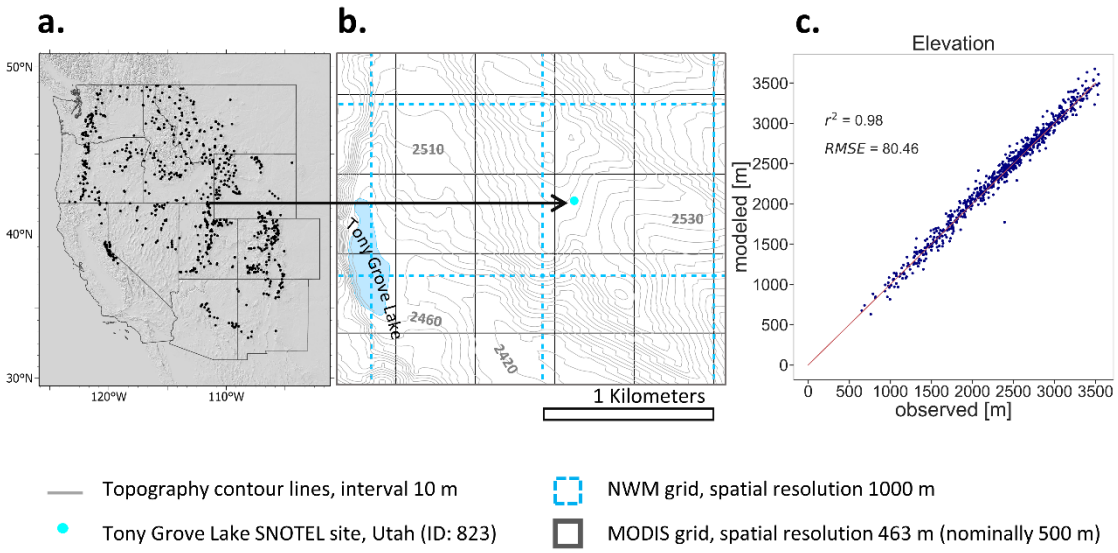

Figure 2. (a) SNOTEL sites (734 black dots) across the Western United States. (b) Illustrative relationship of Tony Grove Lake, Utah SNOTEL site (ID: 823), within NWM grid cells with a spatial resolution of 1 km and MODIS grid cells with a spatial resolution of 463 m (nominally 500 m). (c) NWM grid cell elevation vs. elevation reported for SNOTEL sites (observed). Note that there are four MODIS grid cells that have their centroid within each single NWM grid cell.

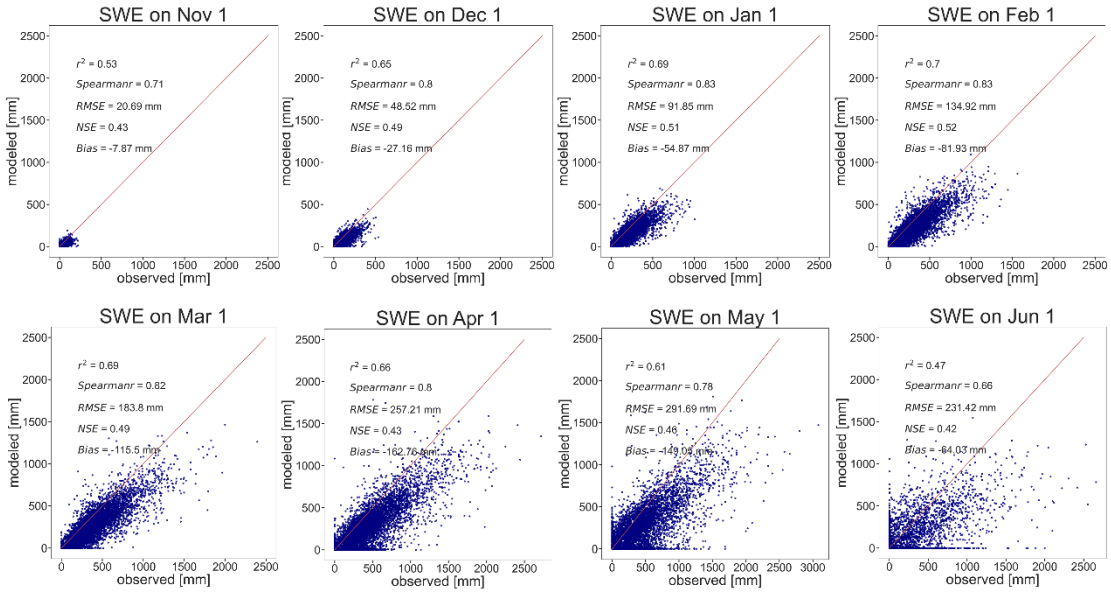

Figure 3. First day of month modeled (NWM-R2) vs. observed (SNOTEL) SWE. Each point is a site and date in the period of overlap between NWM-R2 and SNOTEL data.

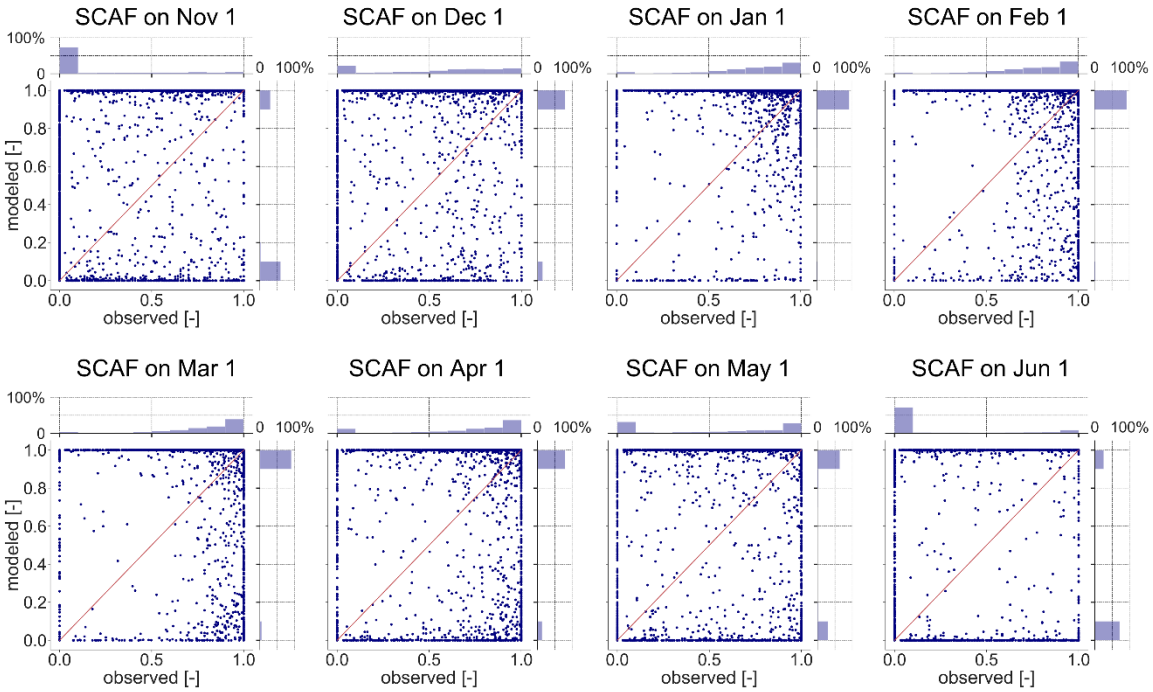

Figure 4. First day of month modeled (NWM-R2) vs. observed (MODIS-C6) SCAF for NWM grid cells and MODIS grid cells containing SNOTEL sites. Each point is a site and a date within the period of overlap between NWM and MODIS data. Axis histograms depict the SCAF distributions.

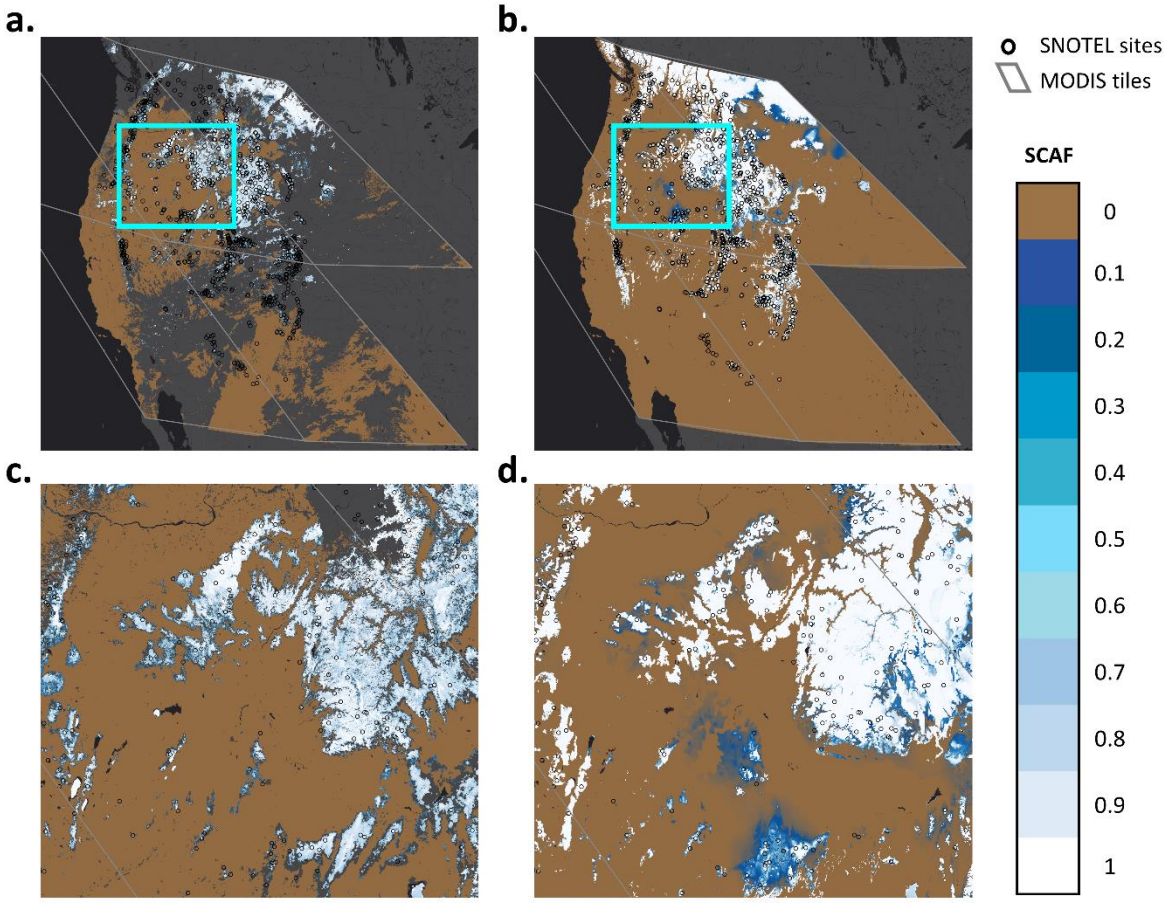

Figure 5. Comparison of NWM-R2 and MODIS-C6 SCAF maps over the study region on Dec 1, 2011. (a) MODIS-C6 SCAF estimated from NDSI_Snow_Cover values of five tiles (in grey). (b) NWM-R2 SCAF outputs at 00:00 UTC masked for the MODIS-C6 tiles. (c) The zoomed-in map of MODIC-C6 SCAF for the blue box in (a).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

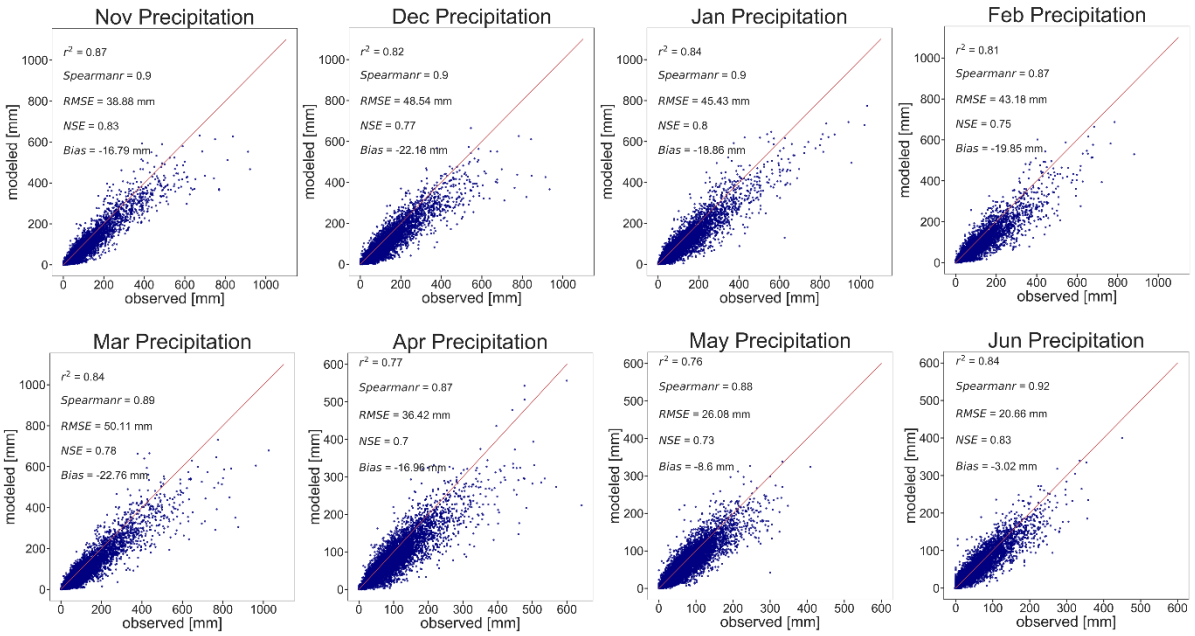

Figure 6. Comparison between NWM-R2 monthly precipitation input (labeled as modeled) and SNOTEL monthly precipitation (labeled as observed). Each point is a site and month in the period of overlap between NWM-R2 and SNOTEL data.

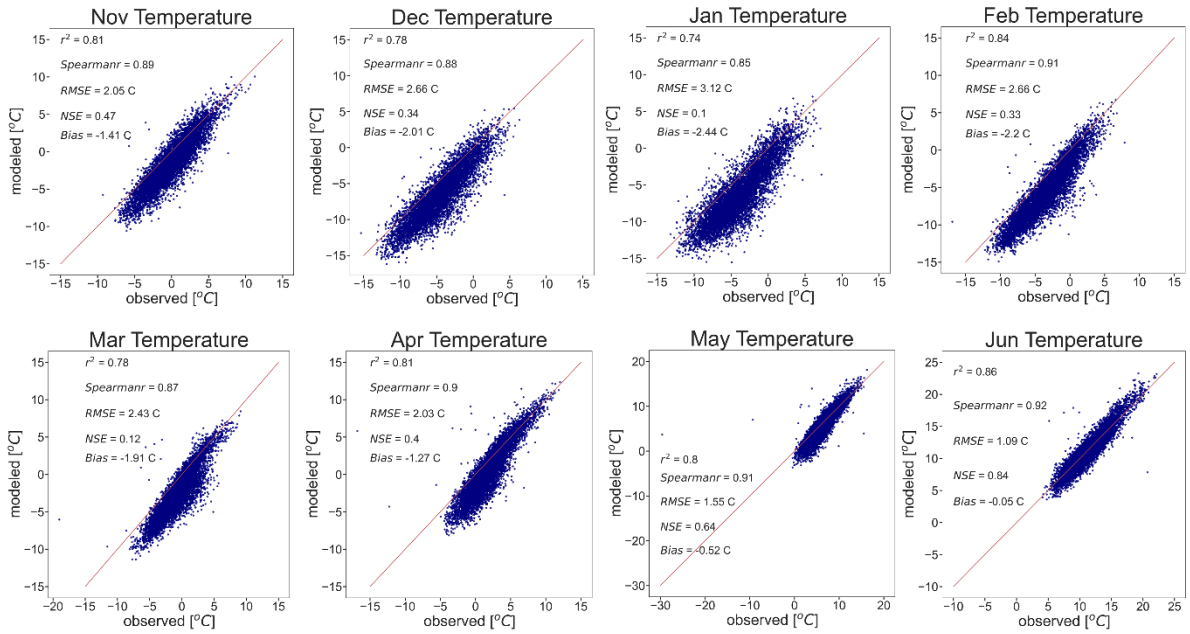

Figure 7. Comparison between NWM-R2 monthly average of hourly air temperature input (labeled as modeled) and SNOTEL monthly average of mean daily air temperature (labeled as observed). Each point is a site and month in the period of overlap between NWM-R2 and SNOTEL data.

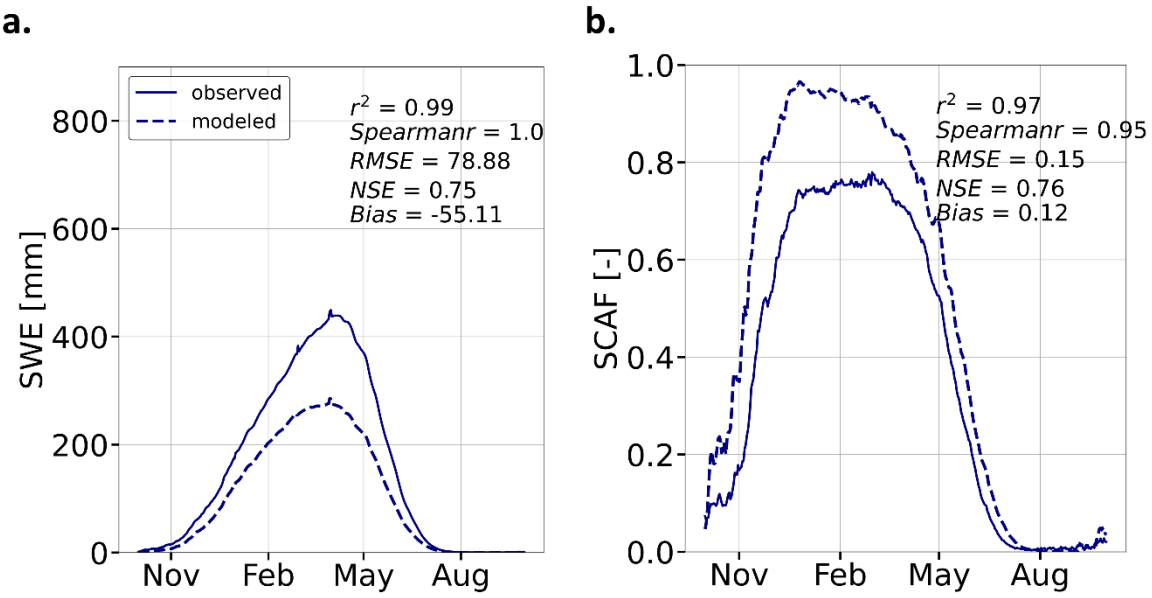

Figure 8. Modeled and observed (a) SWE and (b) SCAF averaged across all SNOTEL sites and years for each specific day of the (water) year.

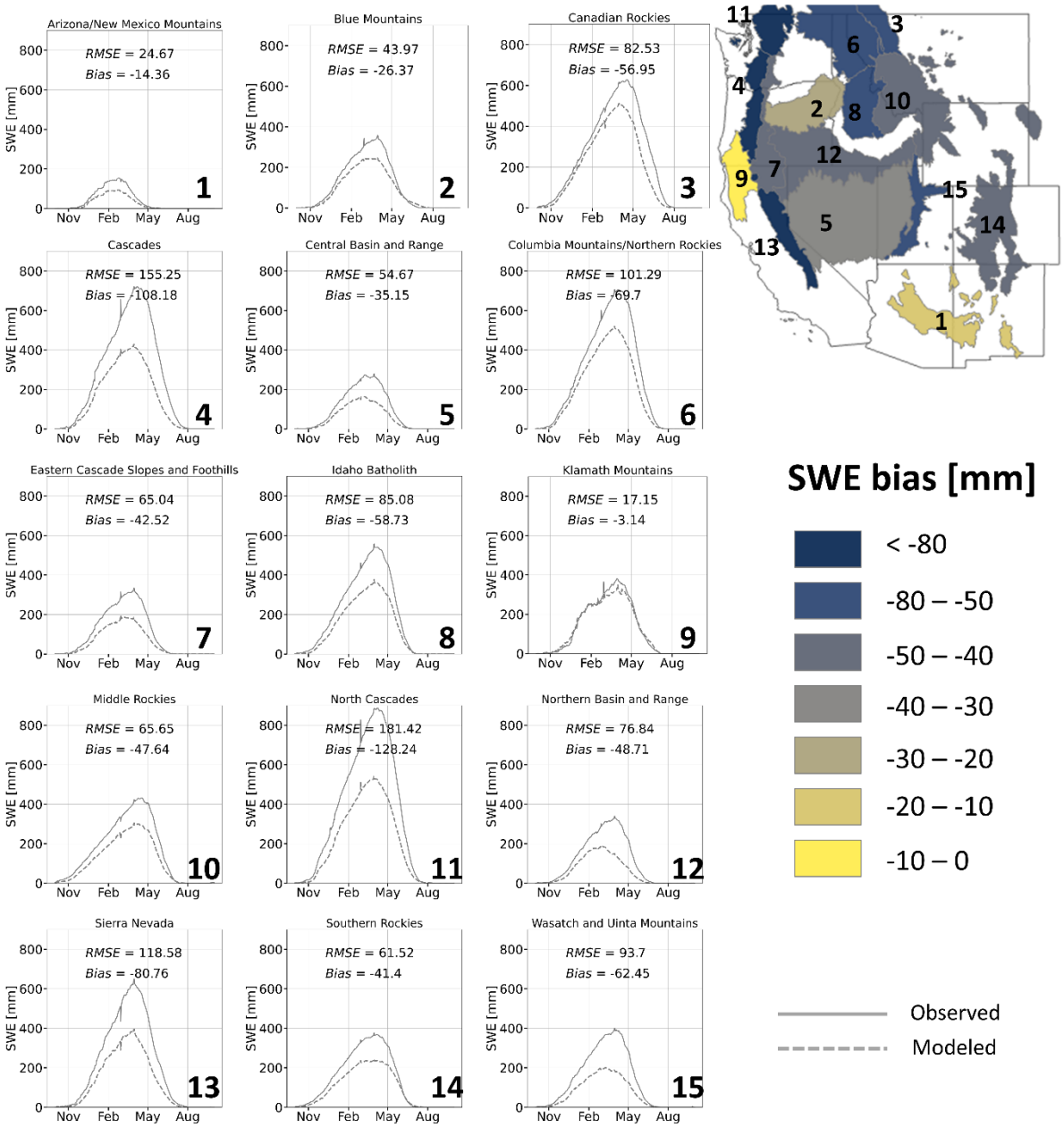

Figure 9. Modeled and observed SWE averaged across all SNOTEL sites and years for each specific day of the (water) year grouped by ecoregion. The map shows 15 Omernik ecoregions where colours represent the bias.

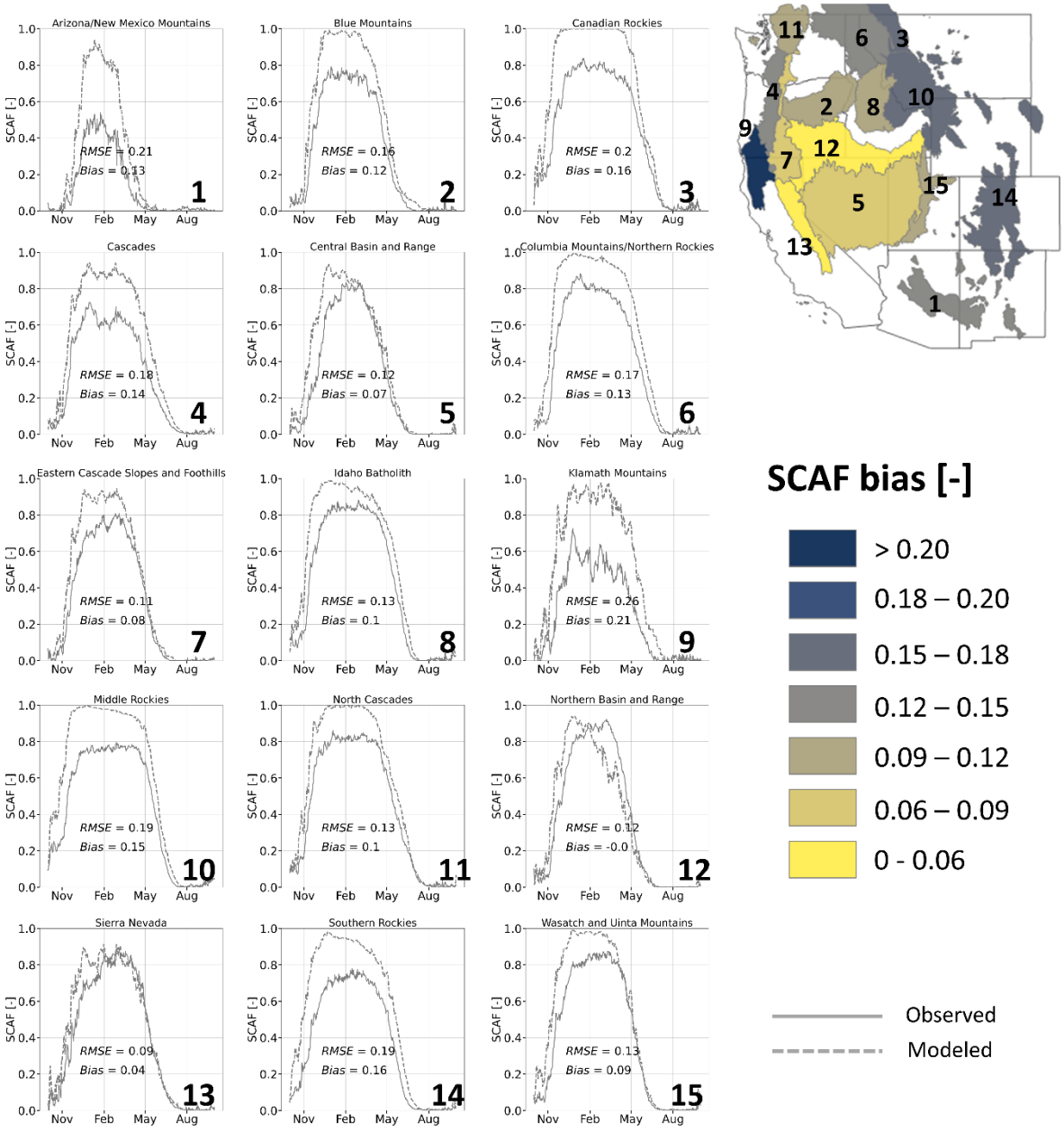

Figure 10. Modeled and observed SCAF averaged across all SNOTEL sites and years for each specific day of the (water) year grouped by ecoregion. The map shows 15 Omernik ecoregions where colours represent the bias.

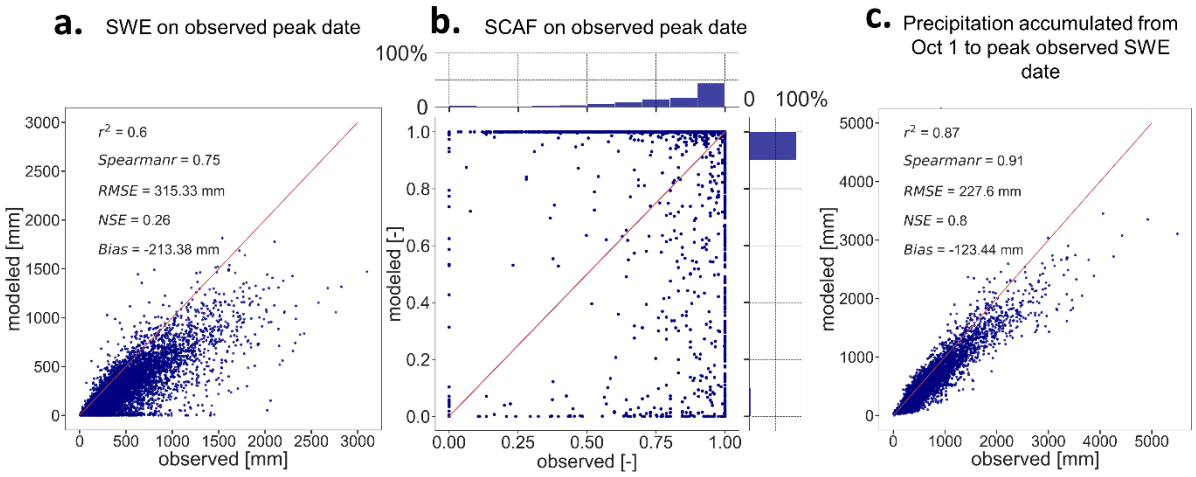

Figure 11. Comparisons on date of observed peak SWE. (a) NWM-R2 vs. SNOTEL SWE, (b) NWM-R2 vs. MODIS-C6 SCAF, and (c) NWM-R2 vs. SNOTEL precipitation accumulated from Oct 1 to observed peak SWE date. Each point is a site and a water year (that starts Oct 1) in the period of overlap between NWM-R2 and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

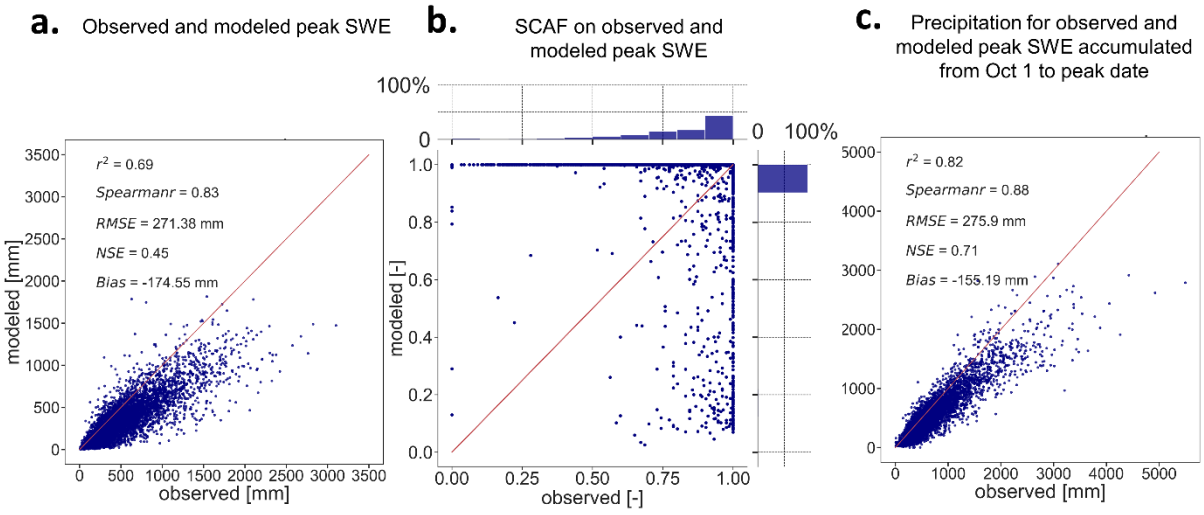

Figure 12. Different date comparison on dates of observed and modeled peak SWE (a) NWM-R2 vs. SNOTEL peak SWE, (b) NWM-R2 vs. MODIS-C6 SCAF, and (c) NWM-R2 vs. SNOTEL precipitation accumulated from Oct 1 to observed and modeled peak SWE dates. Each point is a site and a water year (that starts Oct 1) in the period of overlap between NWM-R2 and SNOTEL data.

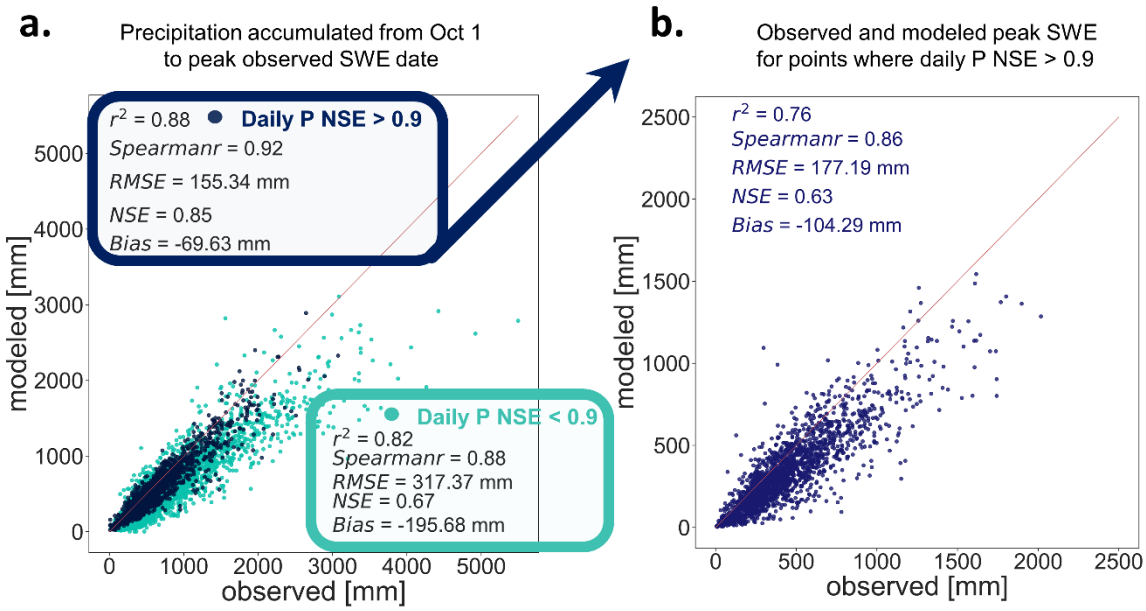

Figure 13. (a) NWM-R2 vs. SNOTEL precipitation accumulated from Oct 1 to observed and modeled peak SWE dates. This figure is similar to Figure 10 (a) but with colours separating points into two groups. The first group (dark blue) contains points where Nash Sutcliffe Efficiency (NSE) values for daily 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).

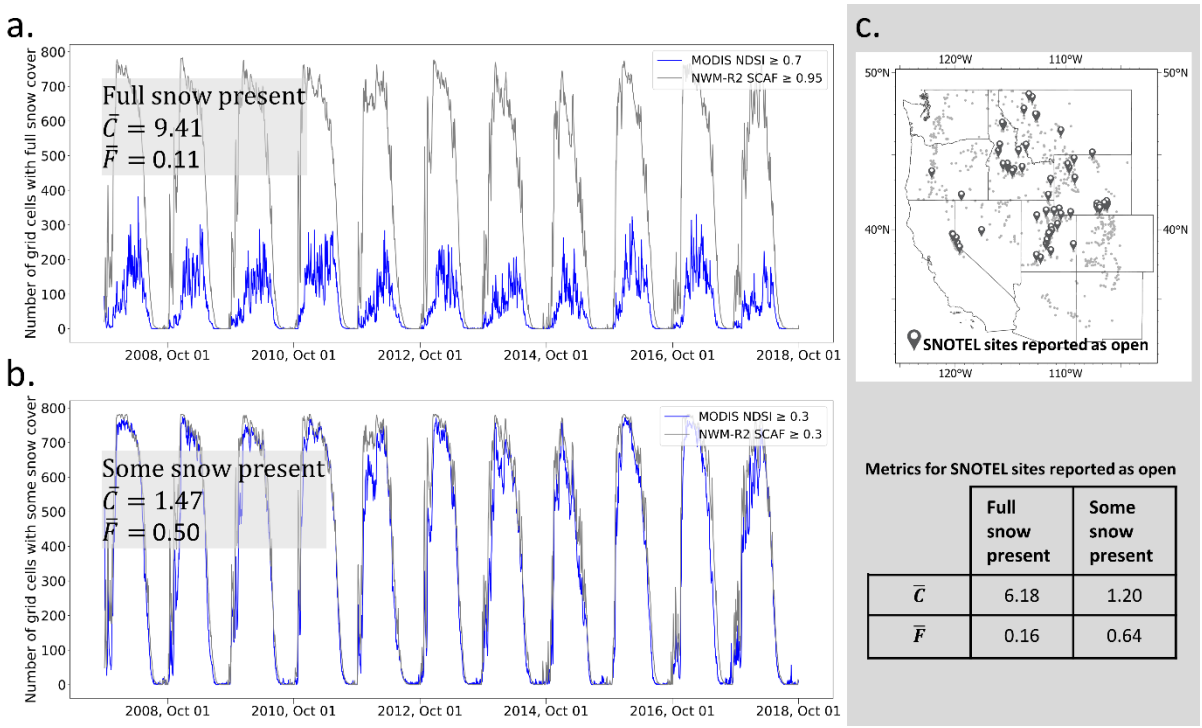

Figure 14. Direct (binary) comparison of snow presence considering all 734 SNOTEL sites with (a) full snow cover and (b) some snow cover. The modeled and observed thresholds for full snow cover were NWM-R2 SCAF≥ 0.95 and MODIS NDSI≥0.7, respectively. Lower thresholds were used for some snow cover (i.e., NWM-R2 SCAF>0.3 and MODIS NDSI>0.3). (c) Locations of the 62 SNOTEL sites reported as open. Average fit metrics (i.e., $\bar{C}$ and $\bar{F}$), presented here, quantitatively evaluate the degree-of-overlap between the modeled and observed snow presence.

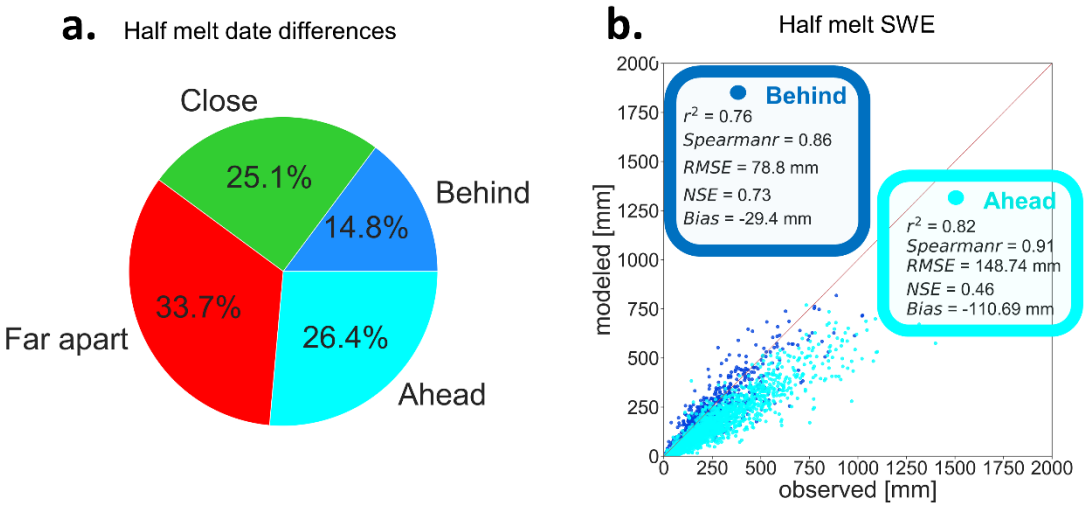

Figure 15. Analysis of melt timing. (a) Classification of differences between observed and modeled dates of half melt from peak SWE. Close: modeled and observed within 5 days of each other; Behind: modeled 6 to 19 days after observed; Ahead: modeled 6 to 19 days before observed; Far apart: Modeled and observed more than 20 days apart. (b) NWM-R2 SWE vs. SNOTEL SWE date of half melt from peak.