Comments:
The paper by A. Gupta et al. addresses the representation of vegetation conditions and their effects on snow simulations based on high-resolution modeling over California. Several model experiments were conducted using different forcing datasets and vegetation schemes. While the paper contains several elements of interest to The Cryosphere readership, it currently does not sufficiently highlight its main contributions, and substantial revisions are required. Below, I outline my main concerns, followed by line-by-line comments.

Reply:

Thank you for your time and efforts to review our work. We have carefully considered all your comments and have addressed them to the best of our ability. We sincerely acknowledge that your suggestions have helped improve the clarity and readability of the manuscript. We appreciate your constructive feedback and welcome any further comments or suggestions.

Main Comments:

Comments:

My main concerns are related to the evaluation data. The authors rely heavily on gridded products for model evaluation, which they treat as ground truth. For instance, a 4 km daily gridded climate product is insufficient to evaluate the downscaled NLDAS-2 data, because it cannot resolve topographic gradients and local effects and introduces its own biases. Station observations (e.g., GHCN-d) are available and should be used to validate the downscaling and provide many evaluation points within the study region.

Reply:

We agree with your main concern about the difference in spatial resolution of the forcing data (NLDAS-2, AORC) and validation data (PRISM). PRISM observation over 4km grid is assumed as a standard product for mean values, our comparison was to show the difference in mean precipitation and intensity of forcing data. We believe that our statistical downscaling method produces better precipitation (NLDAS-2) compared to AORC precipitation over studied domain. We have further validated it by choosing some stational data (represented with * in Fig. 2 and Fig. 3) keeping in mind that these stations should also have the SWE datasets. However, we agree that those stational data validation were not descriptive (e.g., not corrected for under-catch). Keeping your suggestion, we have chosen 177 precipitation and 178 temperature station from The Global Historical Climatology Network (GHCN) network. The stations with high quality data (>95% available data) were chosen for the analysis. The minimal missing data were filled using the long-term average. However, these datasets are available for the low altitude stations only. We have updated the figure from evaluation of these new datasets and add the details to the supplementary material. We have also modified other details based on yours and additional reviewer comments.

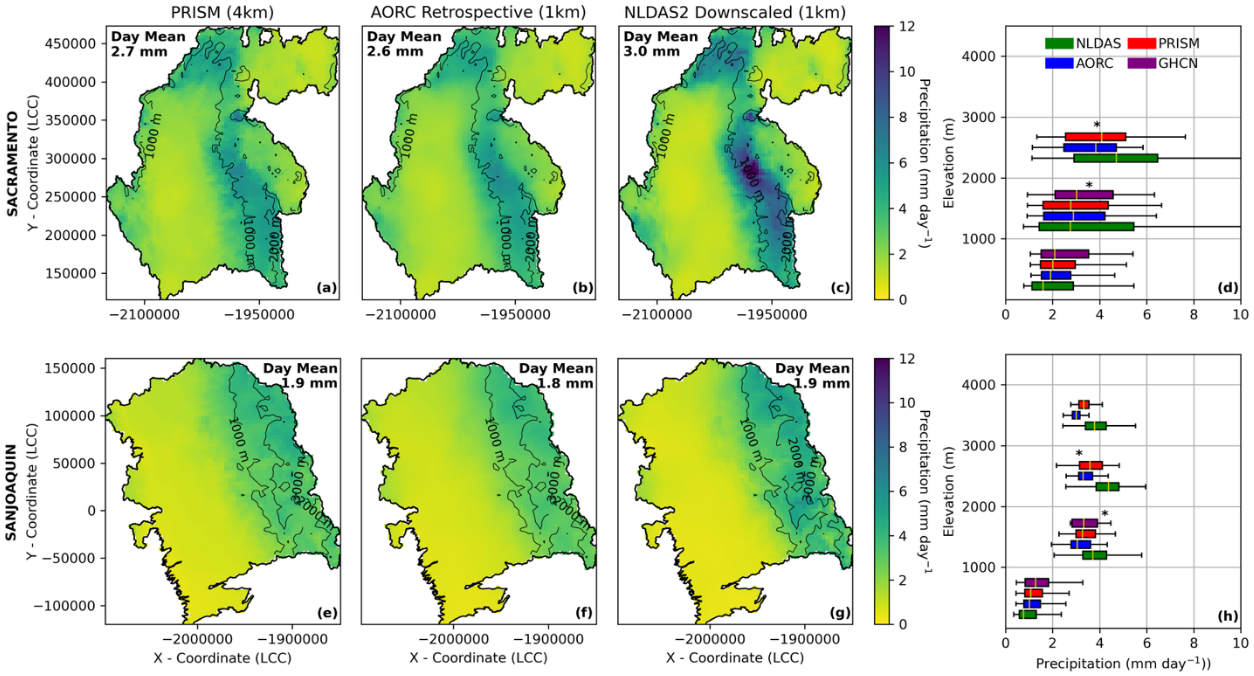


*Figure 2. Downscaled 1 km AORC and NLDAS-2 precipitation (mm day-1) compared to the 4 km PRISM data using WRF-Hydro MFE over the Sacramento (upper panel, a-c) and San Joaquin River Basins (lower panel, e-g). Spatial daily mean has been annotated in the plot. Boxplot of these gridded products and GHCN observation stations over the two River Basins are plotted in d and h, * represents the mean from the available CDWR stations in each River Basins (Fig. 1).*

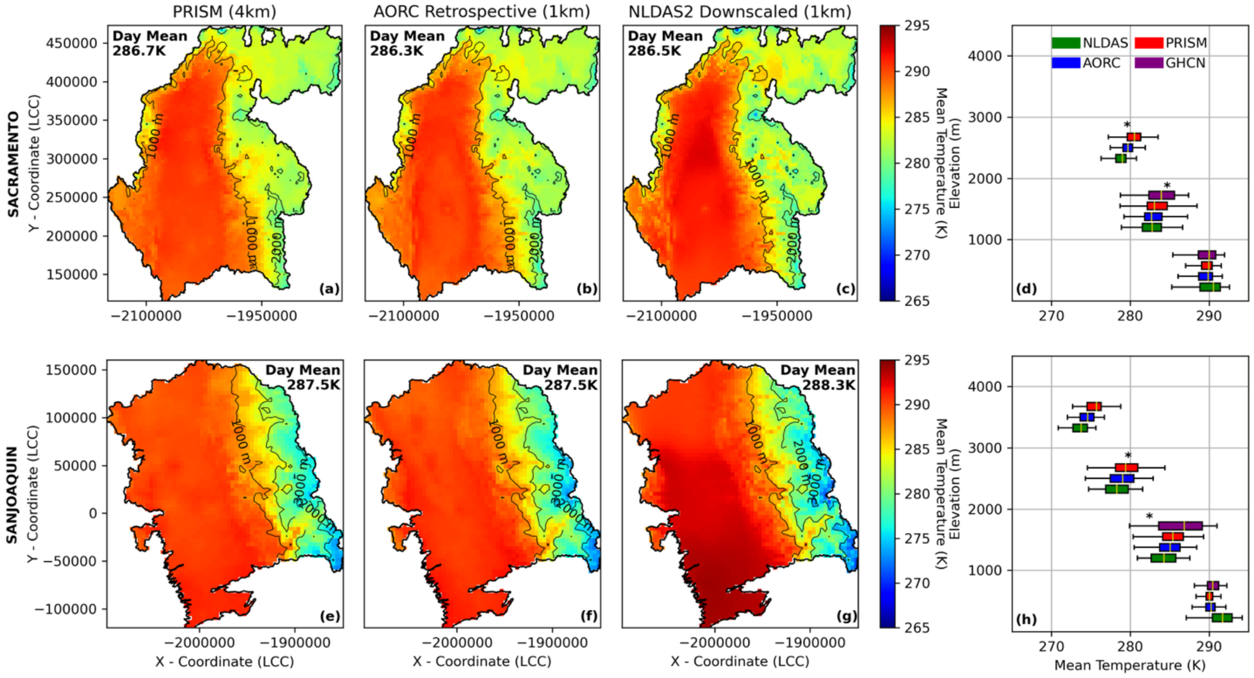
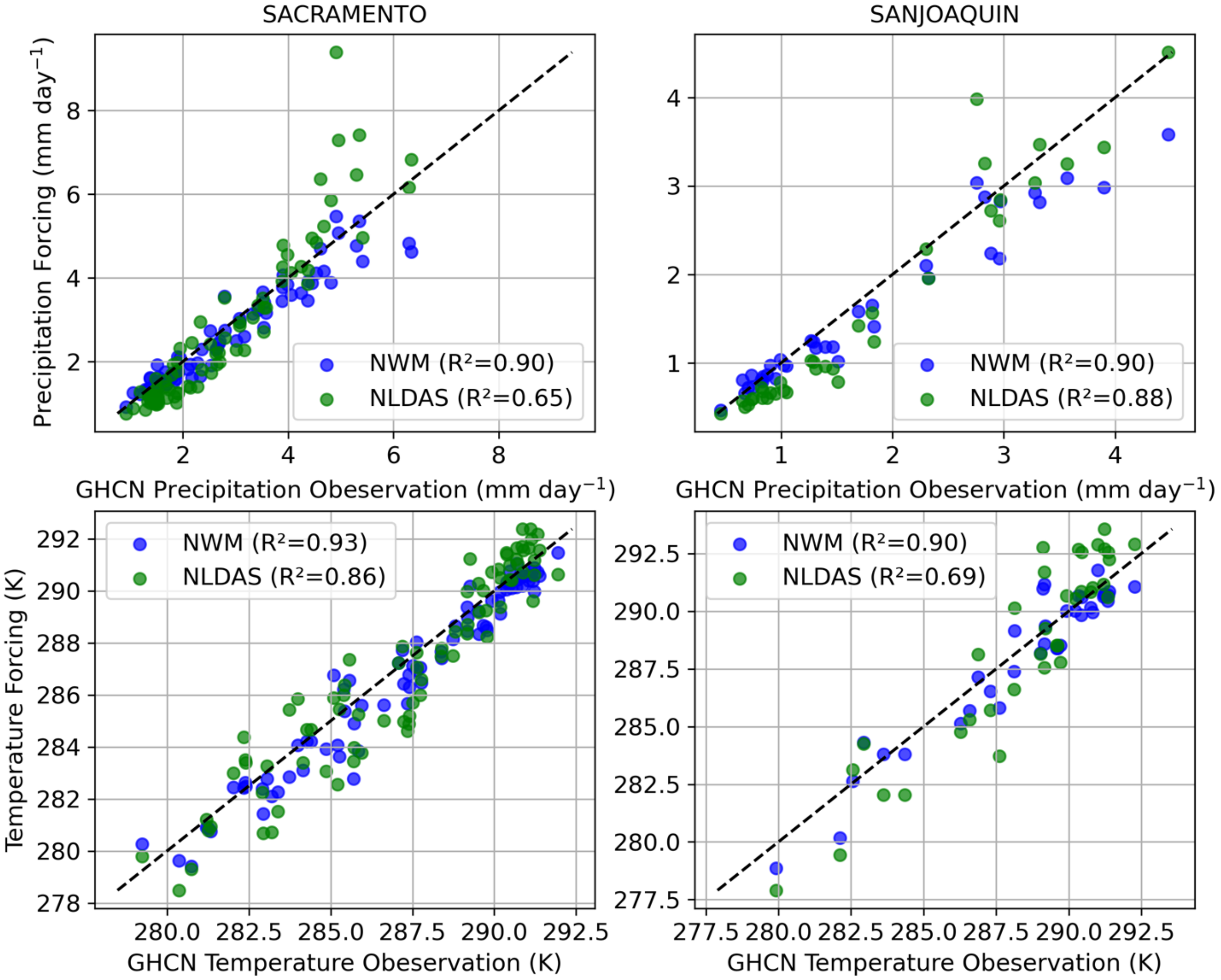


*Figure 3. Downscaled 1 km AORC and NLDAS-2 temperature (K) compared to the 4 km PRISM data using WRF-Hydro MFE over the Sacramento (upper panel, a-c) and San Joaquin River Basins (lower panel, e-g). Spatial daily mean has been annotated in the plot. Boxplot of these gridded products and GHCN observation stations over the two River Basins are plotted in d and h, * represents the mean from the available CDWR stations in each River Basins (Fig. 1).*

We have additionally plotted the scatter plot for all GHCN station to include in the supplementary figure with statistics:



Supplementary Figure: Downscaled 1 km NLDAS-2 and AORC precipitation (mm day-1) and temperature (K) compared to GHCN station data using scatter plot over the Sacramento (left panel) and San Joaquin River Basins (right panel).

Comments:

A similar issue applies to the UA-SWE product: please include an error analysis against actual SNOTEL sites within the study region, rather than validating against a gridded or assimilated product. Section 3.2 should be revised for clarity-it is currently not obvious which datasets are used for which evaluation. I recommend adding a summary table listing, for each variable, the datasets used for validation.

Regarding the evaluation of SWE at the selected CDWR station in the main manuscript: why was this specific station chosen? While I acknowledge that results from other stations appear in the supplementary material, it would be valuable to present more than a single station in the main text.

Reply:

We do not have any SNOTEL stations in over this region; hence we used the California Department of Water Resource (CDWR) stations. The station analysis has been included in supplementary material (S1 and S2). However, keeping in mind yours and additional reviewer's comments we will bring back those figures to the main text, and we will add more detailed statistics. We will also replace Figure 5 to include all the statistical analysis related to station datasets. In total we have 23 SWE stations from CDWR, these stations were chosen with highest quality of available datasets and their distribution across the study area. Yes, we agree that section 3.2 can we grouped further in gridded and station data. Following your suggestions, here is reorganization of validation data section:

3.2 Validation data

To evaluate the performance of the meteorological forcing, snow processes, and vegetation dynamics in this study, multiple independent observational and gridded datasets were used. These datasets are grouped below according to the variables they are used to validate precipitation and temperature, snow, and leaf area index (LAI).

3.2.1 Precipitation and temperature validation

PRISM

We used PRISM (Parameter-elevation Regressions on Independent Slopes Model) daily data to evaluate precipitation and temperature from AORC and the downscaled NLDAS-2 data. PRISM, which is developed and maintained by the PRISM Climate Group at Oregon State University, provides high-resolution (4 km) daily climate data for various meteorological variables (Daly et al., 2000a). Monthly normals are the baseline datasets which are modeled using digital elevation model as a predictor grid and the daily normals are derived using the "nudging and smoothing" technique. The daily normals include precipitation, minimum, maximum & mean temperature, and minimum & maximum vapor pressure deficit over the CONUS. In this study we have used the daily temporal resolution temperature and precipitation datasets from PRISM. Note that all 1 km datasets (AORC and downscaled NLDAS data) are averaged to 4 km resolution for the comparison.

CDWR stations

California Cooperative Snow Surveys (CCSS), which is part of the California Department of Water Resources (CDWR), conducts snow surveys in the mountains of California (https://water.ca.gov/Programs/Flood-Management/Flood-Data/Snow-Surveys). Established in 1929 by the California Legislature, this program is a partnership of more than 50 state, federal, and private agencies. CCSS maintains a total of 265 snow courses and 130 snow sensors located throughout the Sierra Nevada and Shasta-Trinity mountains. This study uses data from 23 CDWR stations across the study area to obtain the precipitation, temperature and SWE data (Fig. 1).

3.2.2 Snow validation

NWM v3.0 SWE

To analyze the improvement in snow estimation with respect to current operational NWM we have downloaded 5-year (2015 – 2019) retrospective SWE estimation from NWM v3.0. NWM uses the prescribed monthly LAI, the Jordan (1991) scheme for rain-snow partitioning, the Ball-Berry scheme for stomatal resistance and the BATS scheme (Dickinson et al., 1993) for snow surface albedo. The retrospective simulations from NWM v3.0 are available for a period between 1979 – 2023 and can be downloaded from the Amazon Web Service (AWS) storage (https://registry.opendata.aws/nwm-archive/). These datasets are available at 3-hourly, which we aggregated to daily timescale for daily comparison.

UA-SWE

UA snowpack data (UA-SWE), developed at the University of Arizona, is based on observed SWE and snow depth from Snow Telemetry (SNOTEL) and National Weather Service Cooperative Observer stations across CONUS (Broxton et al., 2024). Initially, UA-SWE was developed at 4 km daily resolution using the 4 km PRISM data, though they have recently been downscaled to 800 m resolution by using 800 m PRISM climate data, and accounting for the effects of smaller-scale topographic and forest cover variations (Broxton et al 2023). Generally, the 800 m and 4 km versions are similar, except the 800 m version shows more detail and has a little less SWE in forested areas. This study uses the 800 m version of the UA-SWE dataset.

SNODAS

Snow Data Assimilation System (SNODAS) is a data assimilation system developed by National Operational Hydrologic Remote Sensing Center (NOHRSC) at NOAA (Barrett, 2003). SNODAS provides snow cover estimates at 1 km and daily resolution to support the hydrological modeling and analysis. SNODAS assimilates data from satellite, airborne platforms, and ground stations. The main inputs to the SNODAS includes the downscaled outputs from numerical weather prediction (NWP) models to simulate the snow cover using a physically based, mass and energy balance snow model.

3.2.3 Vegetation (LAI) validation

MODIS and Sun Yat-Sen University LAI

This study uses the MOD15A2H Version 6.1 data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on-board the Terra satellite. MODIS captures data in 36 spectral bands (Myneni et al., 2021) and provides detailed information about the Earth's surface and atmosphere, including vegetation, land cover, and cloud properties. The MOD15A2H dataset is a combined LAI and Fraction of Photosynthetically Active Radiation (FPAR) product that is an

8-day composite with 500-meter spatial resolution. In this study, we aggregate the 8-day product into seasonal data for comparison with the modelled data. We also used the reprocessed MODIS Version 6.1 LAI data from Land-Atmospheric Interaction Research Group at Sun Yat-Sen University (Lin et al., 2023). Their product applies spatio-temporal filtering and smoothing to MODIS LAI data to reduce their spatial and temporal inconsistency. These datasets are also available at the 8-day temporal resolution and 500-meter spatial resolution which have been converted to seasonal data for comparison in this study.

Here is the new statistics table (comparison of simulated SWE against CDWR stations) which will replace the figure 5, and along with we will pull the S1 and S2 figure from the supplementary material to main text.

	NLDAS-2				AORC			
	DynVeg_Off		DynVeg_Pred		DynVeg_Off		DynVeg_Pred	
	R^2	KGE	R^2	KGE	R^2	KGE	R^2	KGE
NLS	0.20	0.25	0.04	0.14	0.63	0.52	0.39	0.35
BLC	0.85	0.72	0.42	0.39	0.69	0.56	0.35	0.37
SLT	0.84	0.64	0.98	0.88	0.46	0.46	0.37	0.40
SNM	0.83	0.70	0.72	0.61	0.03	0.23	-0.14	0.16
RTL	0.93	0.92	0.84	0.68	0.74	0.61	0.57	0.48
HYS	-1.22	-0.25	-0.14	0.13	0.91	0.80	0.71	0.63
GOL	0.83	0.68	0.97	0.93	0.43	0.44	0.13	0.30
PLP	-0.16	0.02	0.48	0.35	0.86	0.74	0.69	0.60
SIL	0.51	0.47	0.86	0.73	0.96	0.91	0.96	0.87
KTL	0.65	0.55	0.53	0.46	0.50	0.45	0.35	0.36
CAP	0.92	0.89	0.91	0.82	0.87	0.74	0.69	0.59
BLS	0.94	0.94	0.88	0.73	0.53	0.48	0.11	0.28
GRV	0.96	0.96	0.91	0.82	0.76	0.60	0.90	0.75
PSR	0.70	0.56	0.94	0.84	0.48	0.44	0.11	0.26
HNT	0.89	0.74	0.96	0.91	0.90	0.77	0.59	0.53
CHM	0.63	0.46	0.89	0.73	0.58	0.55	0.31	0.40
BLD	0.76	0.76	0.92	0.93	0.94	0.92	0.83	0.69
TMR	0.76	0.57	0.94	0.79	0.88	0.77	0.69	0.60
SLM	0.67	0.62	0.91	0.83	0.89	0.76	0.68	0.58
GRM	0.61	0.52	0.26	0.33	0.78	0.63	0.45	0.40
GNL	0.90	0.83	0.95	0.91	0.78	0.68	0.61	0.56
TUM	0.75	0.65	0.96	0.93	0.82	0.67	0.98	0.93
HHM	0.96	0.89	0.96	0.84	0.61	0.56	0.66	0.60

Comments:

For the first part of the results (meteorological downscaling): since the downscaling was not performed or developed by the authors, it should not be presented as a primary result. While such an evaluation is important context, it would be more appropriate in an appendix or supplementary section, serving as supporting material for the discussion. Moreover, the assessment itself is rather limited (mostly qualitative or daily means); please expand it to include more quantitative analysis.

Reply:

I agree with your comments as we failed to highlight this in the abstract that NLDAS-2 datasets were downscaled by us though not developed by us. We have used the WRF-Hydro Meteorological Forcing Engine (MFE) to downscale ~12 km NLDAS2 datasets to 1 km resolution. We will replace the 3rd sentence in abstract with the following sentence "To address this issue, we applied Noah-MP version 5.0 with a dynamic vegetation module over the Sacramento and San Joaquin River Basins in California at 1 km resolution. The model was driven by Analysis of Record for Calibration (AORC) forcing at native resolution and by NLDAS-2 forcing that was downscaled from ~12 km to 1 km using the WRF-Hydro Meteorological Forcing Engine." We have elaborated the section 4.2 with equations and highlight that we have used two forcing datasets, one at native resolution and one downscaled by MFE. You can find the descriptive method on L247 comment reply. We will keep the downscaling figure in main text to show betterment and update it with GHCN station analysis.

Comments:

The discussion is extremely brief and does not adequately engage with the caveats and implications of the study. It needs to be substantially expanded and rewritten to interpret results critically in light of known uncertainties.

Reply:

We have expanded the discussion based on our new analyses and add more details to explain why choosing high-quality precipitation and leaf area index datasets is important in snow modeling. More references and an outline for future work is also be added to bring a more insights for readers. Here is the revised discussion section:

We have shown the capability of dynamic vegetation module to simulate LAI and how it can impact the snowpack dynamics through processes of snow interception and canopy shading. Such impacts can result in longer snowmelt, altering soil moisture, and partitioning of hydrological fluxes. Snow-vegetation interaction is complex yet intriguing in nature. In our simulations we found that the canopy intercepted snow decreases the incoming radiation due to increased albedo and decreases the overall melting energy at the snow surface. Such phenomenon was well observed in many publications especially at the local observation scale (Strasser et al, 2011; Stahli et al, 2009; Niu et al, 2004). Chen et al. (2014) highlighted the importance of forcing and model physics in perturbing the below canopy radiation budget however, they cannot explain the canopy shading effect. The canopy shading decreases the solar energy reaching the ground snow, which results in less episodic snowmelt during the accumulation season and more SWE. This kind of impact can only be observed with the more accurate estimation of LAI, which is achieved with the dynamic vegetation module. With limited gridded estimates, we have shown that the dynamic vegetation produces better LAI values, which helps to better simulate the SWE dynamics. For the simulation purpose, we have used two kinds of forcing datasets AORC and NLDAS-2. AORC forcing is a well-known dataset for calibrating hydrological models over CONUS (Cosgrove et al., 2024; Kitzmiller et al., 2018). We have used the mountain mapper algorithm (one of method used in generating AORC datasets) to downscale the NLDAS-2 datasets. In our study regions in California, we found that AORC is underestimated compared to PRISM and the downscaled NLDAS-2. Hence, the modelled SWE with the AORC forcing is underestimated compared to the gridded estimates and more significantly when compared to the 23 CDWR stational data over the mountains. Many studies have suggested that the operational NWM, which uses the same forcing (AORC) generally underestimated the peak SWE in the Western United States (Garousi-Nejad and Tarboton, 2022; Yang et al., 2022, 2023). This study also indicates that using AORC, the model underestimates SWE in lower and mid-elevation stations (sections 5.2) in consistent with Yang et al. (2022).

Dynamic vegetation mainly simulates the LAI and GVF through the carbon budgets of plants' photosynthesis and respiration (Niu et al., 2011). Having more accurate LAI helps improve the snow on the ground, canopy interception of snow, radiation transfer and canopy evaporation/sublimation (Hedstrom and Pomeroy, 1998; Moeser et al., 2015). The chosen dynamic vegetation method outperform the prescribed LAI (section 6.3). This enhances the SWE simulations across elevation. We have shown that NLDAS-2 produces more precipitation (closer to PRISM) compared to AORC. However, when we compared the gridded precipitation to the station observation, we have found that AORC matches quite well CDWR stations (Fig. 10a, b). This comparison is contrary to the SWE datasets from same stations, which shows a more comparable SWE to NLDAS-2 forcing and severe underestimated SWE with AORC forcing (Fig. 10c, d). These contrasting results can be largely attributed to limitations in the observational datasets. In particular, the CDWR precipitation measurements are not corrected for gauge undercatch, which can lead to systematic overestimation of snowfall (Lundquist et al., 2015). In contrast, the gridded SWE products, SNODAS and UA-SWE, are derived using data assimilation and machine learning approaches, respectively, and incorporate multiple in situ SWE or snow depth observations. Relative to these gridded datasets, the NLDAS-2 driven simulations tend to overestimate SWE, whereas simulations forced by AORC exhibit a clear underestimation. Moreover, CDWR stations are often preferentially located in areas with deeper snowpacks, potentially influenced by wind driven snow redistribution, which introduces a positive bias when extrapolating point-scale observations to the 1-km model grid. Given these representativeness issues, simulations forced by NLDAS-2 with prescribed vegetation parameters (LAI and GVF) substantially overestimate SWE, while the inclusion of a dynamic vegetation module yields more realistic SWE estimates that better align with SNODAS. Nevertheless, even with dynamic vegetation, the NLDAS-2 forced simulations still produce higher SWE than SNODAS, likely because the modeled LAI remains larger than that derived from MODIS observations.

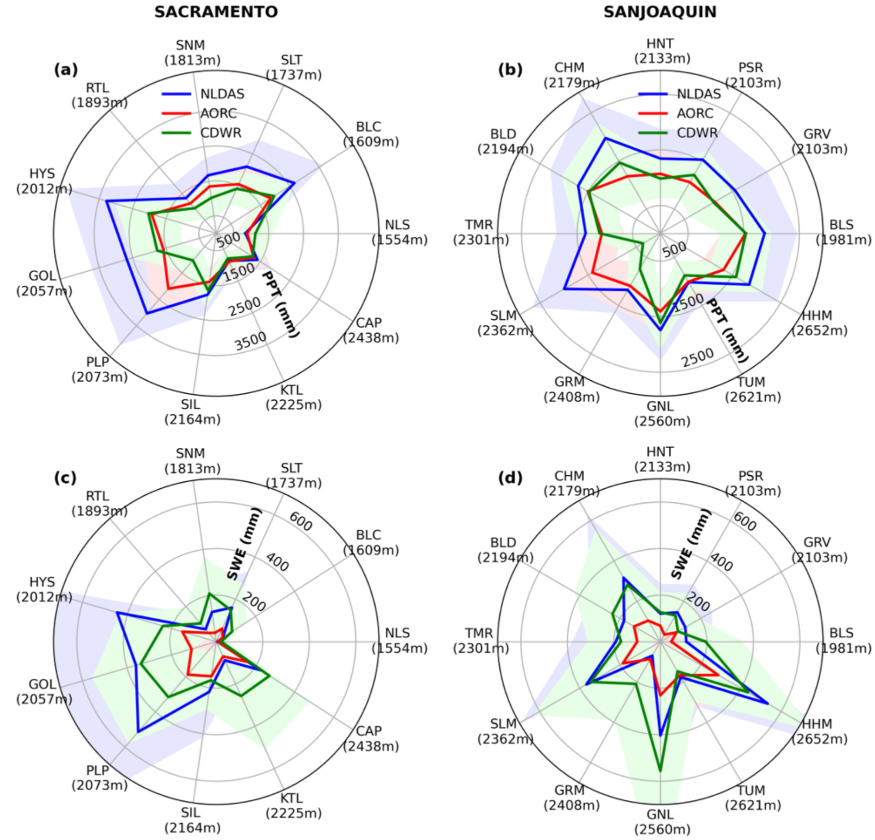


Figure 10 Comparison of yearly mean precipitation (a & b) and SWE (c & d) data from NLDAS-2 and AORC against the CDWR stational data (left panels: 11 stations in Sacramento River Basin; right panels: 12 stations in the San Joaquin River Basin). The modeled SWE data are taken from the DynVeg_Pred simulation. The shaded region represents the standard deviation of the datasets across 5 years at each station. The CDWR precipitation data are not corrected for gage undercatch.

This study highlights the importance of accurately representing fine-scale vegetation processes in large-scale hydrological and land surface models, demonstrating that subgrid vegetation heterogeneity can substantially alter simulated hydrological behavior. In particular, under- or over-representation of leaf area index (LAI) at the subgrid scale can significantly influence the diurnal dynamics of energy partitioning, snow accumulation and runoff generation by modulating canopy interception, radiation transfer, and snow–vegetation interactions. These findings suggest that commonly used grid-averaged vegetation parameterizations may obscure critical nonlinear responses in snow accumulation and melt processes. Future land surface modeling efforts should therefore explicitly investigate the role of subgrid vegetation structure and phenology to reduce uncertainties in regional- to global-scale snow and hydrologic simulations. Because LAI directly affects both snow accumulation and ablation through canopy shading, longwave enhancement, and interception processes, it also controls snow persistence and seasonal snow cover duration. Changes in snow cover duration can, in turn, modify surface albedo and land–atmosphere energy exchanges, potentially regional atmospheric feedback (Lee et al., 2024). Incorporating these mechanisms offers a promising pathway toward more physically consistent coupled land–atmosphere modeling frameworks.

Comments:

The abstract should also be updated to better reflect the findings actually supported by the results. For example, the statement: ‘Using the 1-km NLDAS-2 forcing, the default vegetation scheme with prescribed leaf area index (LAI) and vegetation cover fraction produces too much SWE on the ground due mainly to the strong canopy shading effect despite more snow intercepted by the canopy.’ is not properly supported by any quantitative analysis in the manuscript. Please ensure that claims in the abstract are clearly supported by the results.

Reply:

We have revised the sentence as: “Using the 1-km NLDAS-2 forcing, the default vegetation scheme, which prescribes leaf area index (LAI) and vegetation cover fraction, produces excessive SWE on the ground, mainly because strong canopy shading reduces snowmelt, even though more snowfall is intercepted by the canopy”. We agree that we haven’t included the detailed statistics in the table, and readers may get confused about our findings. In the simulations, we have shown that even with the underrepresented snowfall, the SWE can be higher if we overrepresent the LAI values. We have now included the detailed statistics table (in the previous reply) to represent the differences in simulation with respect to point observations.

Comments:

In general, the language of this manuscript requires improvement. There are numerous incomplete or unclear sentences. Acronyms are inconsistently introduced or redefined (e.g., CONUS is never introduced, LSM several times, ..). Some hyperlinks are broken, and several cited papers are missing from the bibliography. The figures also need attention: improve overall quality, adjust color maps for clarity and accessibility, and ensure captions are complete and precise.

Reply:

Thanks for the suggestions. We will improve the paper’s readability and overall quality.

Line by line comments:

Comments: L50: SNOTEL needs to be introduced and defined.

Reply: We will expand SNOTEL as “Natural Resources Conservation Service Snowpack Telemetry Network (SNOTEL)”

Comments: L50: NCAR CLM v4.0 is an outdated version; please contextualize results or comparisons using more recent model developments.

Reply: We will also include following reference "Eldardiry, H., Sun, N., Yan, H., Reed, P., Thurber, T., & Rice, J. (2025). Characterizing how meteorological forcing selection and parameter uncertainty influence Community Land Model version 5 hydrological applications in the United States. Journal of Advances in Modeling Earth Systems, 17(3), e2024MS004222." They also show that western US watersheds are forcing (snow) dominant and underrepresentation of forcing datasets brings high uncertainty in these regions.

Comments: L53–58: Add appropriate citations.

Reply: Yes, we will add the following citation,

1. Rundel, P. W., & Millar, C. I. (2016). Alpine ecosystems. In: Zavaleta, E.; Mooney, H., eds. Ecosystems of California. Berkeley, California: University of California Press: 613-634. Chapter 29., 613-634.

Comments: L66–68: Add citations.

Reply: Dear reviewer, we will add the following citations,

1. Pomeroy, J. W., Bewley, D. S., Essery, R. L., Hedstrom, N. R., Link, T., Granger, R. J., ... & Janowicz, J. R. (2006). Shrub tundra snowmelt. Hydrological Processes: An International Journal, 20(4), 923-941.
2. Sturm, M., Douglas, T., Racine, C., & Liston, G. E. (2005). Changing snow and shrub conditions affect albedo with global implications. Journal of Geophysical Research: Biogeosciences, 110(G1).

Comments: L88: The acronym LSM was already defined in line 45.

Reply: We will only keep the LSMs acronym here and remove the expanded form.

Comments: L92: Clarify: "best by what metric?" Specify performance criteria.

Reply: We modify sentence as, "The snow model of Noah-MP represents one of the best snow models compared to other LSMs like JULES, Catchment LSM, and Noah when compared for annual maximum SWE"

Comments: Cho et al., 2022 is missing from the references.

Reply: Thank you for highlighting this issue we will add following citation to the bibliography: "Cho, E., Vuyovich, C. M., Kumar, S. V., Wrzesien, M. L., Kim, R. S., & Jacobs, J. M. (2022).

Precipitation biases and snow physics limitations drive the uncertainties in macroscale modeled snow water equivalent. Hydrology and Earth System Sciences Discussions, 2022, 1-22."

Comments: Clarify the aim of the study. Define what "high resolution" means in this context and specify the dynamic vegetation model used.

Reply: Dear reviewer we will modify it as follows:

This study aims to 1) improve the accuracy of SWE simulation over California at 1 km resolution by enhancing precipitation estimates using statistical downscaling approach 2) advance understanding of how vegetation processes, including canopy snow interception, sublimation, and radiative effects (shading and scattering), influence snowpack accumulation and ablation; and 3) identify the optimal combination of precipitation forcing and vegetation representation for improving SWE estimation across California. To address these objectives, we employ Noah-MP version 5 coupled with a dynamic vegetation model from Dickinson et al. (1998) that prognostically simulates leaf area index (LAI) and green vegetation fraction, enabling a detailed assessment of vegetation impacts on high-resolution snowpack simulations.

Comments: Revise - "significant snow cover" is vague; use a more quantitative or physically meaningful term.

Reply: We will revise the sentence as:

"The Sacramento River Basin covers ~70,000 km2 in northern California and includes vast snow cover and forests in the southern Cascade Mountains, Klamath Mountains and Sierra Nevada (Domagalski, 1998)".

Comments: L114: Since one forcing is considered superior, briefly justify why.

Reply: Here we add a clarification sentence:

"These forcing datasets differ in their mean precipitation, with NLDAS-2 exhibiting systematically higher precipitation amounts than AORC".

Comments: L116: CONUS acronym not introduced.

Reply: Here is the modified sentence:

AORC version 1.1 was developed at National Weather Service as a part of development of the contiguous United States (CONUS) scale hydrologic modeling capability and establishment of the National Water Model (Kitzmiller et al., 2018).

Comments: L126: The section structure is difficult to follow. Consider reorganizing by variable evaluated (e.g., temperature, precipitation, SWE) rather than by dataset name/acronym.

Reply: Dear Reviewer, thank you for this this comment, we have reorganized the section by type of datasets (temperature, precipitation, SWE, LAI) rather than based on how they have used in the study. You can find the revised version in main comment reply.

Comments: L127: Why are gridded products used for validation? Comparison outside of measurement sites lacks meaning; please justify or reconsider.

Reply: As stated above we have used the GHCN station to validate our forcing datasets. We have also cliped the data to the study region in the figure 2 and 3.

Comments: L135: "Averaged to 4 km resolution" — elaborate on the method used for upscaling.

Reply: We have used bilinear interpolation for upscaling. We will modify the sentence as: "Note that all 1 km datasets (AORC and downscaled NLDAS data) are upscaled to 4 km resolution using bilinear interpolation for the comparison".

Comments: L138: Why are actual SWE measurements not used here?

Reply: One of the motivations for this study was that NWM produced underestimated SWE, so to show improvement against NWM we have used SWE from NWM. We have also used same forcing and LAI as NWM and showed that they have underestimated precipitation and overestimated LAI. We have further used 23 SWE stations from CDWR to compare SWE at point scale.

Comments: L144 / L152: Additional gridded products are introduced—justify their relevance or quality.

Reply: As our SWE product from modeling is at gridded scale, so we wanted to compare our product with best available gridded SWE datasets in the USA. SNODAS and UASWE are best available gridded product in the USA. We will further include the quality statement from these products based on literature in the section.

Comments: L223 ff: Add more detail on the radiation transfer model, including its assumptions and parameterizations.

Reply: Here is the modified explanation:

Noah-MP represents vegetation shading and scattering effects on shortwave radiation using a modified two-stream radiation-transfer approximation that explicitly accounts for canopy structural heterogeneity through between-canopy and within-canopy gap probabilities (Niu et al., 2004; Yang and Friedl, 2003). In this framework, radiative transfer through the canopy is governed by the total canopy gap probability, P_c, defined as the sum of the between-crown gap probability P_{bc}, determined by canopy geometry and solar zenith angle, and the within-crown gap probability P_{wc}, parameterized using a modified Beer's law,

$$P_c = P_{bc} + P_{wc}; \qquad P_{wc} = exp(-\frac{F_a H_d}{cos\theta'})$$

where $F_a$is the foliage area volume density derived from the effective leaf and stem area index (LAI+SAI), $H_d$is the crown depth, and $\theta'$is the effective solar zenith angle adjusted for crown geometry. These gap probabilities regulate the fraction of incoming shortwave radiation that penetrates the canopy without interaction, as well as the portion that is scattered by vegetation elements. The direct beam shortwave radiation reaching the snow surface beneath the canopy is therefore given by,

$$Q_{b,b} = P_c$$

while diffuse shortwave radiation includes contributions from canopy-scattered direct and diffuse fluxes as well as diffuse radiation transmitted through canopy gaps. Noah-MP computes both below-canopy radiative fluxes that reach the ground snowpack and above-canopy upward fluxes per unit incident direct and diffuse radiation, which together determine the effective surface albedo across visible and near-infrared wavebands (Table 2). The canopy gap probabilities are dynamically linked to vegetation structural parameters, including canopy thickness, crown radius, tree density, and LAI and SAI, enabling realistic modulation of snow–radiation interactions across different forest types. In addition to shortwave processes, the model accounts for longwave radiation received by the snow surface, including atmospheric longwave radiation transmitted through canopy gaps and longwave emission from the canopy itself. These longwave contributions may either enhance or dampen snow surface energy inputs depending on vegetation density and structure, thereby influencing snow accumulation and melt processes at the grid scale.

Comments: L234 ff: Was there a model spin-up? If so, describe procedure and duration.

Reply: Yes, we have spin-up the model for 5 years, we have used year 2014 – 2015 for spinup and run 5 cycles, then end of the year restart file was used to simulate the model from 2015-2019. We have added the details in section 4.3 (255-256). We will add more details to it as explained above.

Comments: L247: Add a reference and brief explanation for the mountain mapper algorithm.

Reply: We will add the following details by replacing the old paragraph:

The meteorological variables from NLDAS-2 at 1/8 degree have been downscaled to 1 km resolution using the WRF Hydro Meteorological Forcing Engine (MFE), which uses the statistical downscaling approach using the mountain mapper algorithm to improve the quantitative estimate of precipitation (QPE). The Mountain Mapper precipitation product is generated by blending monthly precipitation climatology from the PRISM (800 m) with real-time hourly rain gauge observations. The PRISM monthly fields are first temporally disaggregated to an hourly scale, and differences between the climatological estimates and observed hourly gauge data are quantified using site-specific bias ratios (e_k) as follows:

$$e_k = g_k / p_k$$

where, g_k is hourly gauge observation and p_k is the normalized PRISM hourly rainfall.

These bias ratios are interpolated to modeling grid using inverse distance weighting (IDW) scheme. Finally, these interpolated bias ratios are multiplied by NLDAS-2 hourly datasets to calculate the downscaled precipitation product. Temperature was downscaled using the National Center for Atmospheric Research (NCAR) lapse rate over the CONUS region. Shortwave radiation, pressure and specific humidity were downscaled using the topographic adjustment according to elevation, while wind and longwave radiation were bilinearly interpolated (https://github.com/NCAR/WrfHydroForcing).

Comments: L248: Elaborate on how the correction factor is applied and justified.

Reply: We have added a detailed paragraph in above reply. We will add this paragraph to elaborate on the downscaling technique.

Tables & Figures

Comments: Table 1: Do not reuse LAI both as an acronym for Leaf Area Index and as a simulation label - rename simulation identifiers.

Reply: Dear reviewer, we agree that the table seems to be a bit confusing however, LAI and GVF are meant for Leaf Area Index and Green Vegetation Fraction, respectively. The simulation acronyms are DynVeg_Off and DynVeg_Pred. LAI and GVF are kept there to highlight the key difference. Here is the modified table:

ACRONYMS	DynVeg_Off	DynVeg_Pred
LAI Description	Prescribed monthly LAI climatology	Dynamic LAI (predicted from carbon storage)
GVF Description	Prescribed yearly maximum GVF	Dynamic GVF (predicted from LAI)

Comments: Tables 2 & 3: Consider moving to the appendix or expand explanations in-text.

Reply: As explained in the beginning of this document, we will elaborate the tables and keep it in the main text. We will also expand the explanations for these tables.

Comments: Figure 2: Improve colormap - differences in row 2 are hard to see. Since comparisons depend on elevation, include elevation contours.

Reply: Dear Reviewer, thanks for highlighting this issue. We have changed the colorbar to new scheme so that the small differences are visible. We have also included the elevation contour to complement horizontal bar plot. Please see the main comment reply.

Comments: Figure 3: Update colormap for clarity.

Reply: We have updated the colormap with viridis and jet color scheme.

Comments: Figure 4: Clarify: what does "independent SWE estimates, SNODAS" mean? What are the spatial units?

Reply: We call it independent as it doesn't directly involve any of our modeling approach to estimate SWE. SNODAS is 1 km SWE product over northern America. We will include these details to the caption as below:

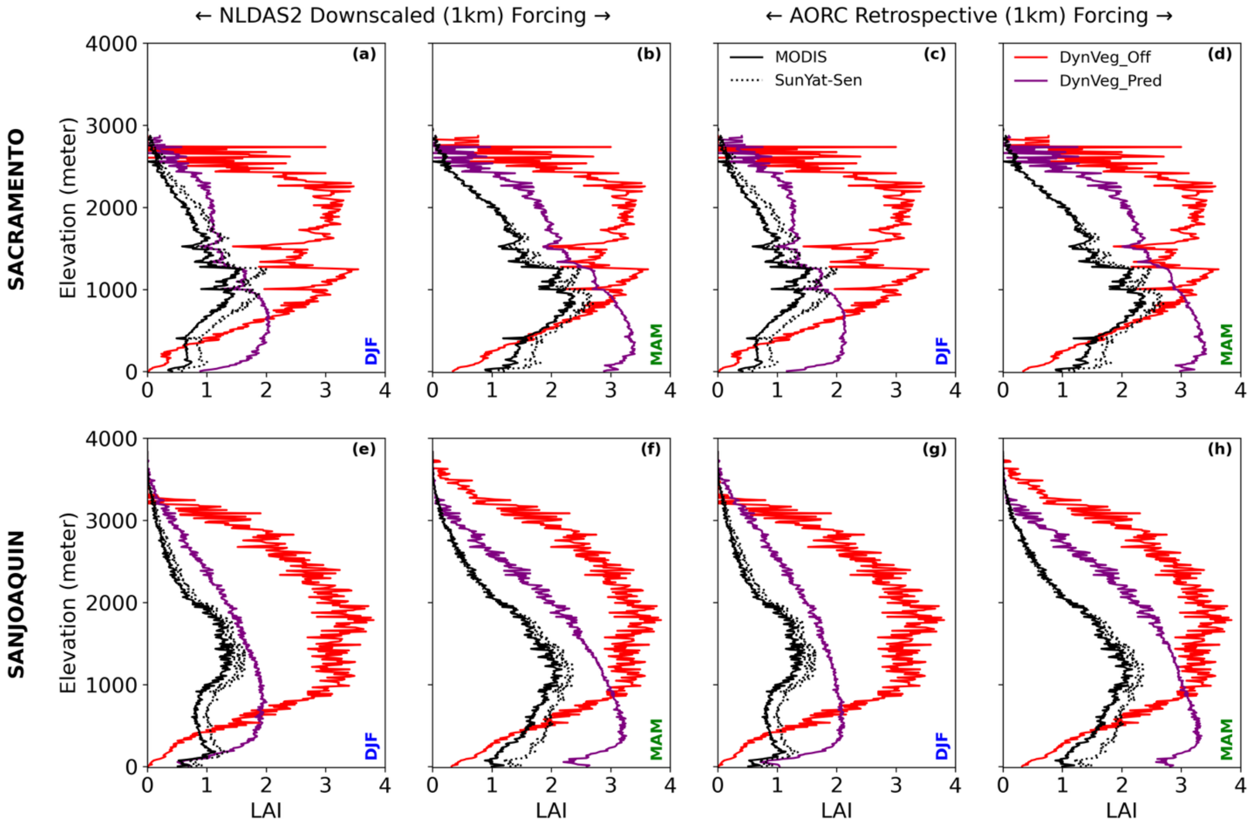
"Modeled snow water equivalent (SWE) driven by different atmospheric forcing data (left panels: NLDAS-2; right panels: AORC) under the prescribed vegetation (DynVeg_Off) and dynamic vegetation (DynVeg_Pred) in the Sacramento (upper two panels) and San Joaquin (lower two panels) River Basins. Independent SWE estimates, SNODAS (1 km daily, section 3.2), is also included."

Comments: Figure 5: The location of the label "GOL 2057 m" is confusing. Improve colormap and rewrite caption to make the number of stations and variables shown clear.

Reply: We have replaced this plot with the table as replied in the main comment.

Comments: Figure 6: Revise colormap for readability.

Reply: We will change the text color for more visibility. However, we would like to keep the line color as it is, because we have used same color schemes for lines throughout the manuscript to distinguish the simulations. Here is the modified figure:



Comments: Figure 7: Clarify whether values are averaged across the basin.

Reply: Yes, the values are averaged across the basin. We will include these details to the figure caption. Here is the updated version:

Modeled canopy intercepted snow rate using different atmospheric forcing data (left panels: NLDAS-2; right panels: AORC) under prescribed vegetation (DynVeg_Off) and dynamic vegetation (DynVeg_Pred) in the Sacramento (upper two panels) and San Joaquin (lower two panels) River Basins. A SWE threshold of 0.1 mm has been applied to mask the region and perform the spatial average.

Comments: L337: Why was this single station chosen? Provide justification or include multiple stations for robustness.

Reply: As discussed in previous paragraph we have included the multiple stations analysis for robustness. We have moved the Figure S1 and S2 to main text and added a detailed statistics table.

Comments: L343: "AORC simulations are underestimated by ~500 mm" — be more specific (relative to what? annual mean? which period?).

Reply: We will replace the sentence as follows:

"Here AORC simulations are underestimated by ~500 mm during the peak SWE in March month over simulation period (Fig. S1, Fig. S2)".

Comments: L345: Explain why the GRM station performs poorly.

Reply: This location shows low precipitation in PRISM and as well as downscaled NLDAS-2 (Fig. 2). It's highly likely that the low precipitation was inherited from the PRISM monthly climatology, this can we further investigated in future work. We will add this explanation to the line 345.

Comments: L388–389: The described "substantial modifications" are not visually evident in Figure 8. Consider including quantitative comparisons to support this claim.

Reply: We will separate this figure into accumulation and melting phase to distinguish how the vegetation can show separate behavior in accumulation and melting phase (similar to Fig. 9). We will add the average melting/accumulation snowmelt statistics to quantify the difference in the two simulations.

Discussion

Comments: L410: Sentence incomplete ("more than" missing?).

Reply: We will modify the sentence as:
"Over higher elevations (> 2,000 m), AORC underestimates precipitation than PRISM and

NLDAS-2 (Fig. 2) but overestimate temperature compared to NLDAS-2, thereby producing less snowfall".

Comments: L416: "This is favourable when compared to…" — rephrase in a more objective scientific tone.

Reply: We will rephrase it as:

"This performance is favorable when evaluated against CDWR station observations; however, it exhibits a positive bias compared to the gridded SWE products (SNODAS and UA-SWE)."

Comments: L421: Revise for clarity.

Reply: We have completely revised this paragraph as:

"These contrasting results can be largely attributed to limitations in the observational datasets. In particular, the CDWR precipitation measurements are not corrected for gauge undercatch, which can lead to systematic overestimation of snowfall (Lundquist et al., 2015). In contrast, the gridded SWE products, SNODAS and UA-SWE, are derived using data assimilation and machine learning approaches, respectively, and incorporate multiple in situ SWE or snow depth observations. Relative to these gridded datasets, the NLDAS-2 driven simulations tend to overestimate SWE, whereas simulations forced by AORC exhibit a clear underestimation. Moreover, CDWR stations are often preferentially located in areas with deeper snowpacks, potentially influenced by wind driven snow redistribution, which introduces a positive bias when extrapolating point-scale observations to the 1-km model grid. Given these representativeness issues, simulations forced by NLDAS-2 with prescribed vegetation parameters (LAI and GVF) substantially overestimate SWE, while the inclusion of a dynamic vegetation module yields more realistic SWE estimates that better align with SNODAS. Nevertheless, even with dynamic vegetation, the NLDAS-2 forced simulations still produce higher SWE than SNODAS, likely because the modeled LAI remains larger than that derived from MODIS observations".

Comments: L424: Omit "apparently"; use formal language.

Reply: We have revised this whole paragraph as explained in the previous comment.

Summary / Conclusions

Comments: L437: Specify the exact model version (e.g., CLM vX.Y) used.

Reply: This study uses CLM version 4, we will add the details to the text including the citation in the bibliography as included below.

Comments: L437: Toure et al., 2016 missing from references.

Reply: We will add the following citation to bibliography:

“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.”

Comments: L442 ff: “The Noah-MP predicted SWE is most sensitive to the high-resolution…” — support this with a robust statistical test.

Reply: This will be revised to “The Noah-MP predicted SWE is most sensitive to the high-resolution (1 km) precipitation inputs compared to other process representations, e.g., the vegetation dynamics.”

Comments: L459–460: Revise sentence for clarity and conciseness.

Reply: We will revise the sentence as follows:

“The forthcoming dynamically downscaled NLDAS-2 dataset at 1-km resolution (NLDAS-3) has the potential to improve snow simulations in the western United States and merits further evaluation.”

Comments: L470: The link currently directs to the generic ArcGIS StoryMap site - please provide the actual map or dataset link. If possible, also provide SWE data for reproducibility.

Reply: Dear reviewer, thank you so much for highlighting this issue. The data has been moved from ArcGIS Story maps to the university website for public access. Now the data is publicly available at this link: https://climate.arizona.edu/data/UA_SWE/DailyData_800m/, we will update this in the revised version.