

NM WRRRI Student Water Research Grant Final Report

Student Researcher: Christopher Cox

Faculty Advisor: Enrico Zorzetto

Project title:

Evaluation of Earth System Model snow simulations in the Southwestern US and implications for future snow projections

Description of research problem and research objectives:

Snow is a fundamental component of the global hydrologic system, and it plays a key role in storing water and modulating surface runoff and groundwater recharge across the Western US. Climate change is projected to influence both snow quantity and seasonality with predominantly negative implications for New Mexico (Dunbar et al., 2022). Given the extent to which NM communities and economies rely on snow-fed surface water resources, it is essential to quantify the degree to which snow in the southwest will change in the future and how this change will impact availability and seasonality of NM water resources.

Earth System Models (ESMs) are used to predict future changes in the global water budget. Recently, advances have been made in the representation of snow in ESMs (Hao et al., 2023, Zorzetto et al., 2024), as well as, in the representation of complex topography (Chaney et al., 2018). These advances have the potential to lead to regionally-relevant ESM snow predictions in mountainous areas like northern NM. Validation over this region is challenging, as complex terrain requires a combination of in-situ data and satellite observations to fully capture the spatial variability of snow water equivalent (SWE).

The objective of this study is to evaluate the performance of ESM snow simulations in the southwestern US. Specifically, we compare snow simulations generated by the NOAA Geophysical Fluid Dynamics Laboratory (GFDL) Land Model v4 (LM4) against NRCS SNOTEL snow observations. While in-situ snow observations are limited in space and time, they provide a reliable metric with which to identify model weaknesses. Furthermore, focusing our analysis on the southwestern US gives a better understanding of model performance in a region that is predominantly arid.

Methodology

Snow simulations are completed at 52 sites across New Mexico and Arizona using the NOAA Geophysical Fluid Dynamics Laboratory (GFDL) Land Model v4 (LM4). This land model is one component of the larger Earth System Model, which includes atmosphere, ocean, and sea-ice components (Shevliakova et al., 2024). Recent improvements to LM4 include sophisticated representation of snow processes including dynamic layering, snow grain metamorphism, and grain size dependent albedo (Zorzetto et al., 2024). This is a major upgrade from prior snow parameterizations that utilized a fixed density and a constant number of snow layers. Dynamic vegetation, including interactions with snow, is also simulated by LM4. Only a subset of land surface models include dynamic vegetation, making LM4 an important point of comparison.

Each of the 52 sites in the model simulations is a Natural Resource Conservation Service (NRCS) SNOTEL station. The SNOTEL network consists of snow monitoring stations across the Western US (Serreze et al., 1999). Each station is instrumented to record snow depth, snow water equivalent (SWE), air temperature, and precipitation. The SNOTEL SWE measurements are directly comparable to SWE output by LM4 providing daily level performance metrics. SNOTEL data (daily) typically spans a range of 10-15 years thru the present, but some stations have data going back to the early and middle 20th century. Although SNOTEL locations are selected to optimize runoff prediction in hydrologically important basins, the locations cover a wide range of elevations and land cover types. The 52 sites included in this work are all the SNOTEL stations in both NM and AZ (see figure 2), giving a broad sense of LM4 snow simulation performance in the southwestern US.

Running LM4 at a point location, uncoupled from an atmospheric model, requires input of surface meteorological forcing data for each model timestep. Inputs include air temperature, air pressure, precipitation, solar radiation, and wind speed. Ideally, in-situ measurements of each of the forcing variables would be used to run the model. Unfortunately, the SNOTEL stations have a limited sensor suite with only air temperature and precipitation measured at most sites. Furthermore, SNOTEL data, especially at the hourly timescale, requires extensive quality control prior to use (Yan et al., 2018). Reanalysis data, which combines satellite and surface observations with atmospheric modelling, is used as an alternative to in-situ measurements of forcing data. We use NLDAS-2, which provides all the needed forcings hourly starting in 1979 on a 1/8th degree grid covering central North America. NLDAS-2 is part of the NASA Land Data Assimilation system. Most of the variables in NLDAS are spatially interpolated and downscaled from the 32km NARR atmospheric model. Incoming shortwave radiation is also bias corrected using satellite observations. NLDAS-2 precipitation is not output by NARR, but is aggregated from surface observations.

In order to more accurately represent the surface meteorology at the SNOTEL site, additional downscaling and correction of the NLDAS data is needed. Here the SNOTEL air temperature

and precipitation data is especially useful despite the daily timescale. Hourly NLDAS air temperatures are shifted and stretched to match the SNOTEL daily maximum and minimum temperatures. NLDAS precipitation is rescaled to match the total monthly SNOTEL precipitation. The temperature correction algorithm is verified at a SNOTEL site in Utah (Tony Grove Ranger Station) where there is a secondary independent temperature measurement available as part of the Logan River Observatory (<https://uwrl.usu.edu/lro/>) (Jones et al., 2017). Visual inspection of the corrected NLDAS temperatures compared to the secondary station indicate substantial improvement (figure 1). Since, temperature and precipitation are two of the most significant inputs for snow modeling, it is expected these corrections will result in substantial model performance improvement beyond what might be achieved from complex downscaling algorithms. Additionally, this study utilizes a quality controlled SNOTEL dataset available from PNNL. Quality control methods applied to the dataset includes elimination of erroneous data and outliers, consistency checks, and bias adjustment to both temperature and precipitation (Sun et al. 2019, Yan et al. 2016).

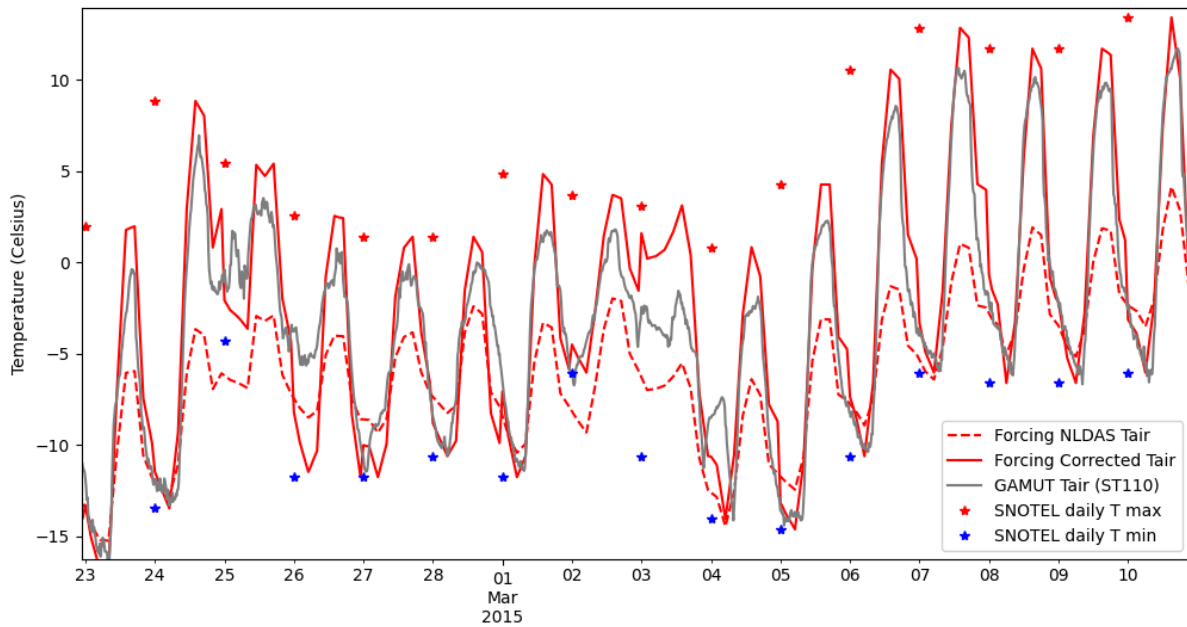


Figure 1: Comparison of air temperature from SNOTEL, NLDAS, and GAMUT (Logan River Observatory) stations to adjusted forcing temperature at a site near Logan, UT.

Prior to running the model at each site, model spinup is needed to properly initialize model parameters. Two of the most important parameters for snow simulation are surface vegetation and soil conditions (temperature, moisture). LM4 models vegetation dynamically and the model is initialized with bare ground. At this point, we have configured the model to only simulate grass. Future simulations may utilize observations of vegetation types at each site to determine the model configuration. Nonetheless, since SNOTEL stations are used for runoff prediction, sites are not typically located in dense canopies. So, grass is reasonable at the local scale. Soil temperature and moisture are initially set to an arbitrary initial value and gradually adjust to an

equilibrium state during spin-up. Unadjusted NLDAS-2 forcings spanning a ten year period provide model input during spin-up and the forcings are looped to cover the 100 year spinup period.

The 100 year spinup period differs for each site depending on the quantity of SNOTEL data at the location. This is done to maximize the amount of in-situ snow observations that can be used for model evaluation. If, for example, a given site has SNOTEL observations that start in 2005, the model spinup starts in 1905. The looped NLDAS spinup forcing data for a site spans the 10 year period before the start of the SNOTEL data. So, the example spin up starting in 1905 would utilize NLDAS data from 1995 to 2005. Since NLDAS is only available back to 1980 and 10 years of data are used for spinup, any SNOTEL data from before 1990 is not utilized in this study. Overall, this workflow ensures that the model run at each site utilizes the same spinup procedure.

At present, three model experiments have been completed at each of the NM and AZ SNOTEL sites: CM, GL, GL-Tcorr. The first model run, CM, establishes a baseline performance level utilizing unadjusted NLDAS forcing and the unimproved snow simulation developed prior to improvements implemented by Zorzetto et al. 2024 (GLASS). This model configuration has fixed snow layering and density, and has been shown to exhibit a soil temperature bias due to insufficiently parameterized snow thermal conductivities (Zorzetto et al. 2024). The second model configuration (GL) also utilizes unadjusted NLDAS forcing, but snow is simulated using much improved GLASS implementation. Finally, GL-Tcorr, also uses the GLASS implementation, but NLDAS surface air temperatures have been corrected using the quality controlled SNOTEL temperature measurements. A model run including both temperature and precipitation corrections has not yet been performed. It is worth noting that, due to the significant differences between the CM and GLASS implementations, experiments utilizing the CM model must have a CM configured spinup (same with GL). Although the same forcing is used in both cases, the model temperatures after spinup will be different. During testing, it was found that at a high elevation SNOTEL site in Colorado, the CM model generated unrealistic permafrost during spinup due to the known cold bias in the CM implementation. Permafrost does not appear to be created by the CM model in NM and AZ site spinups, but initial soil temperatures are likely to be different between CM and GL model runs.

Results

Model performance across sites and different model configurations are evaluated using the bias and root mean square error (RMSE) of the daily model output SWE versus the SNOTEL daily SWE. Mean bias and RMSE for all 52 sites over each model run is shown in table 1.

	CM	GL	GL - Tcorr
Bias (mm SWE)	-38.9	-34.8	-34.3
RMSE (mm SWE)	80.6	75.3	73.6

Table 1

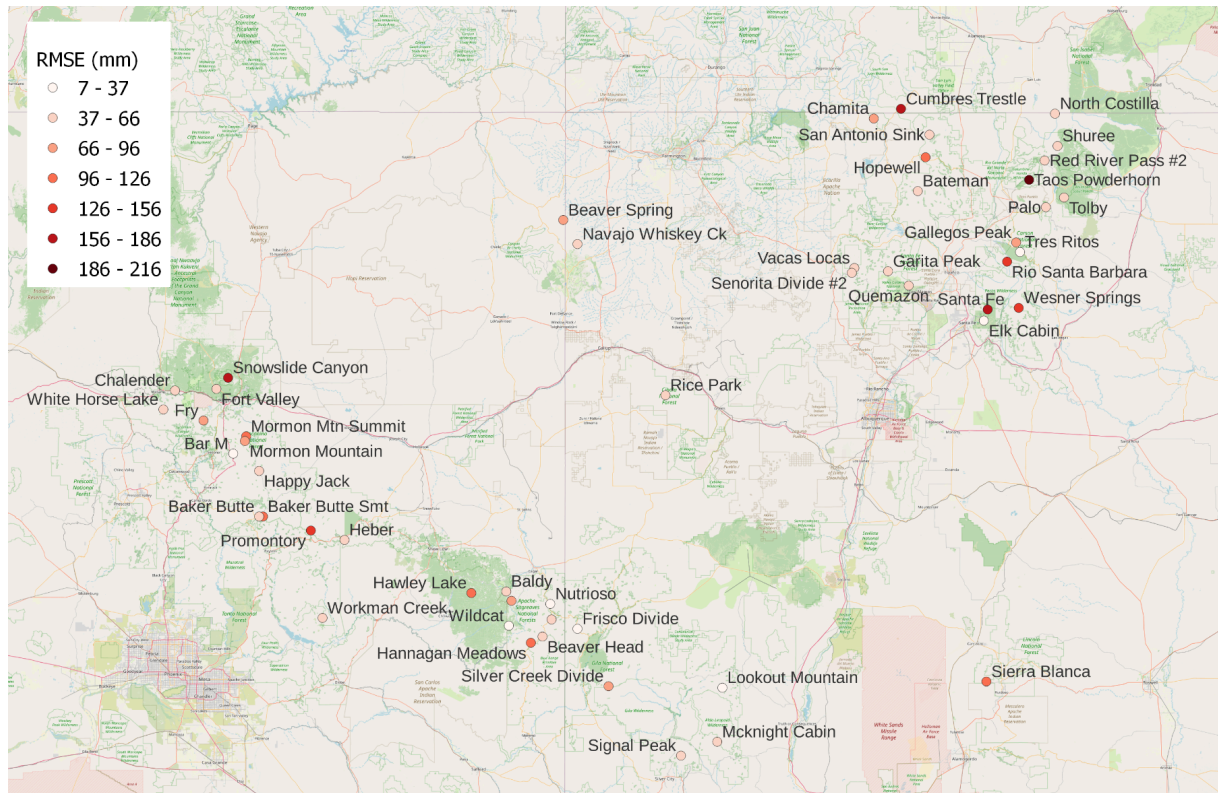


Figure 2: Sites with associated RMSE for GL model run.

The GL-Tcorr model run should be considered highly preliminary because the temperature correction algorithm needs improvement. Specifically, the current algorithm performs unlimited linear interpolation of any SNOTEL data gaps. It was thought that data gaps in the quality controlled SNOTEL dataset would be no more than a day or two. Instead, it is somewhat common to find sites with a month or more of missing data. Interpolated data does not play a direct role in model performance evaluation, but it does impact the corrected forcing temperatures which in turn influence the model output including time periods after the data gap when the forcing temperatures are valid.

In general, all models show a negative bias indicating that the models are on average outputting less snow than is observed. The more sophisticated GL model shows improvements in bias and RMSE relative to the CM model. Downscaling of temperature also seems to result in improved performance. Figure 2 shows a map view of the RMSE. There does not appear to be any regional patterns or influence due to latitude.

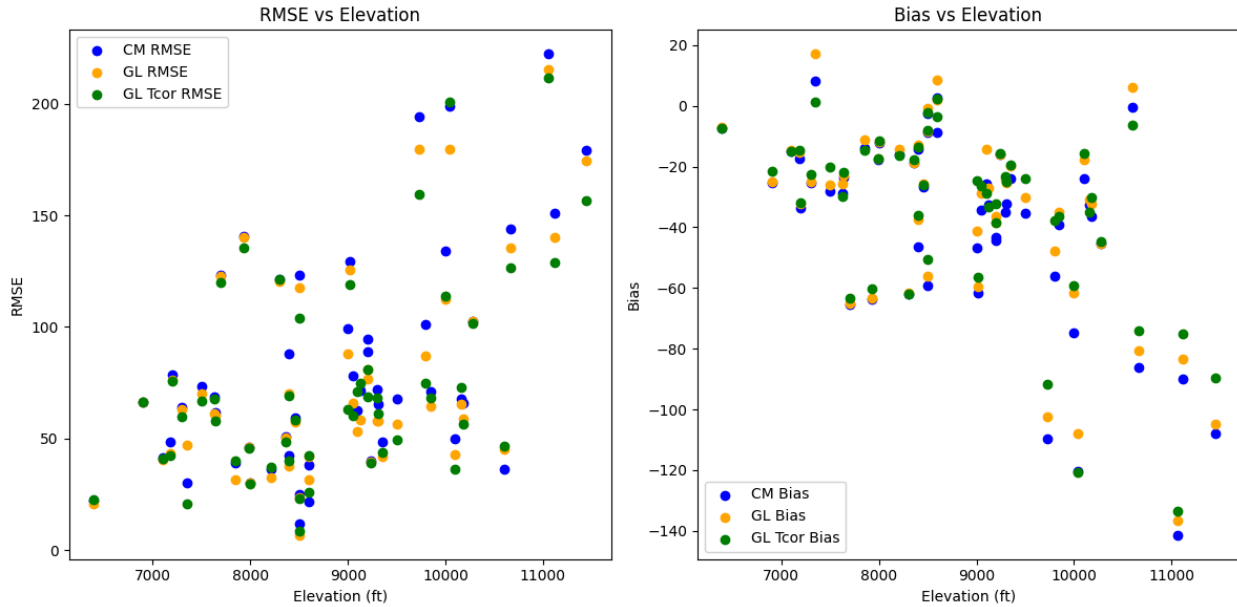


Figure 3 - Model RMSE and Bias with site elevation

There is a correlation between elevation and model performance with a general decrease in performance at higher elevations. This could be due, at least in part, to model handling of higher elevation site characteristics such as forests or colder temperatures. However, for this initial analysis, it seems more likely that the error is due to insufficient downscaling of forcing data. The elevation difference between the SNOTEL site and the NLDAS-2 grid elevation is often more significant at higher elevations where rougher topography is not well represented in the ~12km NLDAS grid. Figure 4 shows RMSE and Bias as a function of the difference between NLDAS elevation and SNOTEL elevation. Performance is worse where the elevation difference is negative (NLDAS lower than SNOTEL) because forcing temperatures are too warm and less precipitation falls as snow. The GL-Tcorr model run shows improvement, but does not eliminate the overall trend. Some sites, particularly Taos Powderhorn, show poor performance even when the NLDAS elevation is fairly close to the SNOTEL elevation.

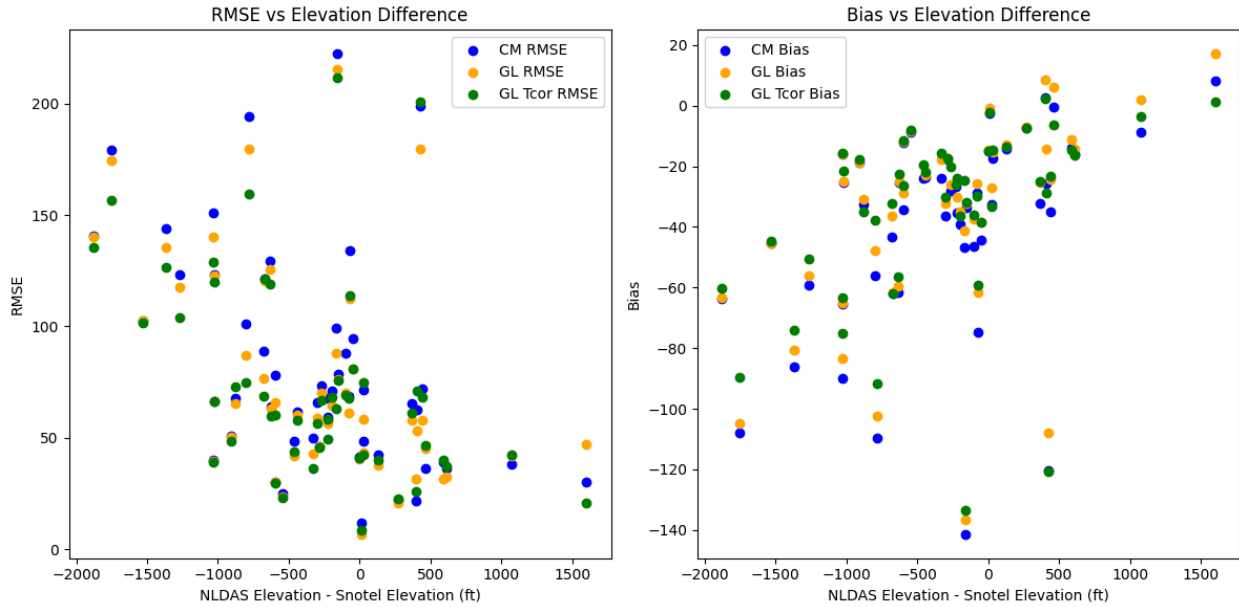


Figure 4 - RMSE and Bias as a function of elevation difference between NLDAS and SNOTEL.

Two of the worst performing sites: Taos Powderhorn and Cumbres Trestle have elevations that are relatively close to the NLDAS elevation and appear as outliers in figure 4. Output from model spinup at the sites does not appear to be unusual with stable soil moisture and surface vegetation at the end of the 100 year spin up period. The GL model output was reviewed for both sites to determine what might be the cause of the low performance. At Taos Powderhorn, the model significantly underpredicts snow (figure 5). Downscaling of temperature and precipitation are not performed in this model run, but the NLDAS temperature and precipitation reasonably matches the SNOTEL data. Temperature and precipitation forcing at Cumbres Trestle also reasonably matches SNOTEL observations (note the significant quantity of missing data). Further investigation is needed, but it is speculated that the model may not be adequately partitioning precipitation between rain and snow at these sites.

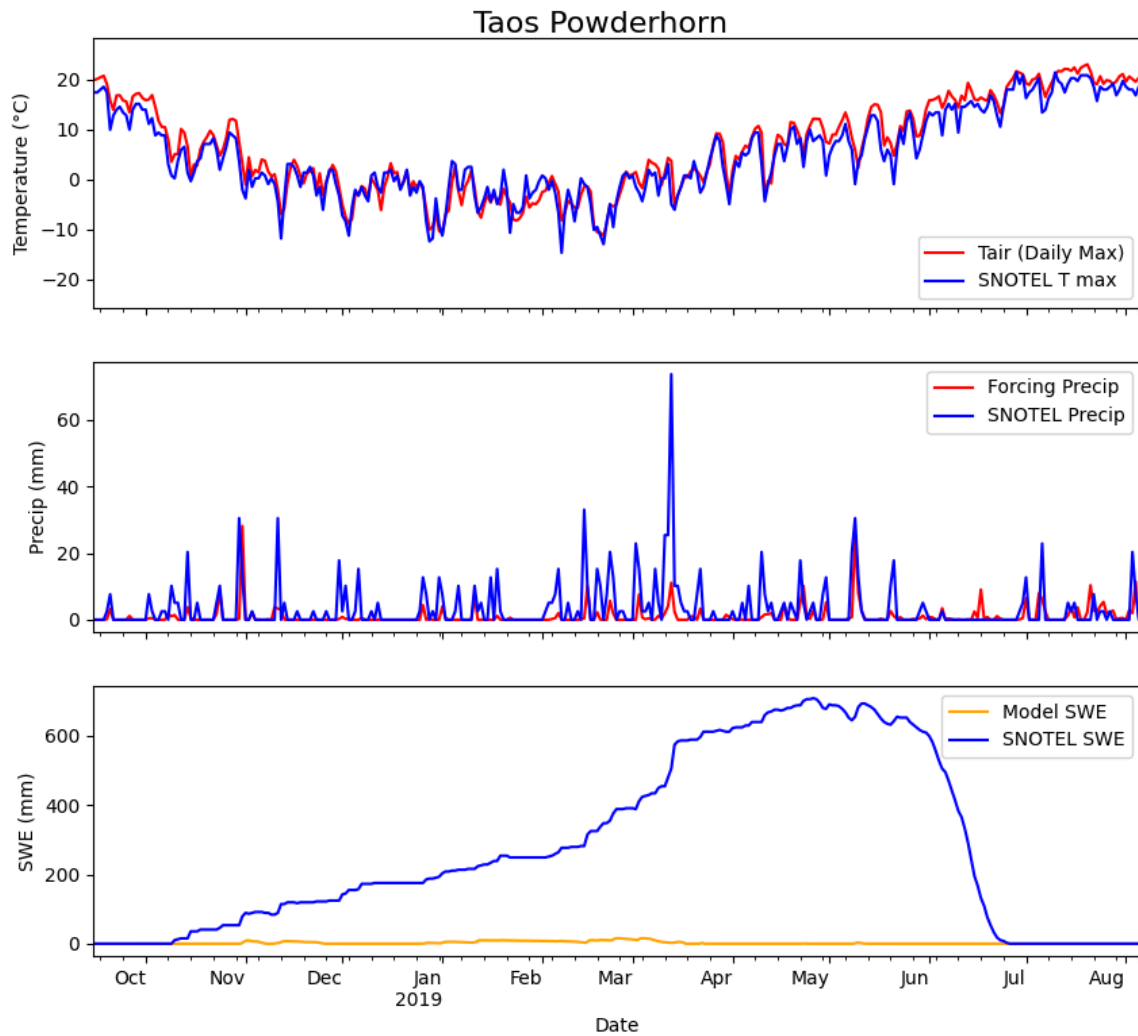


Figure 5: Model output (GL) compared to SNOTEL data at Taos Powderhorn focusing on the 2018/2019 winter season.

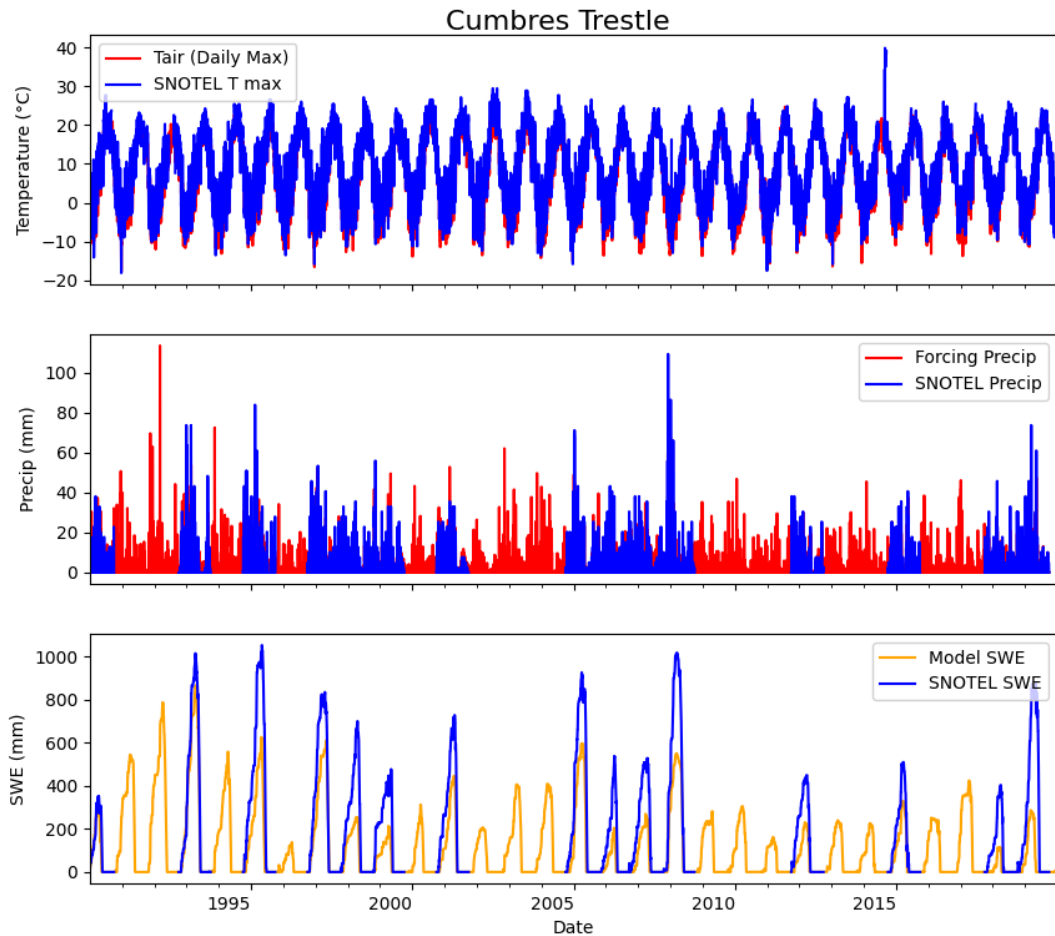


Figure 6: Model output (GL) at Cumbres Trestle compared to SNOTEL observations. A significant amount of SNOTEL data is missing at this site.

Much more work is needed to fully evaluate the model. First, downscaling of the forcing data will be applied to more variables with improved handling of missing SNOTEL data. Next, the modelling will be scaled up to all sites in the Western US (~700). Unfortunately, Alaska will not be included due to NLDAS-2 coverage. Finally, analysis of results will consider many more dimensions including slope, aspect, vegetation.

Provide a paragraph on who will benefit from your research results. Include any water agency that could use your results:

The results presented here are the first step in a larger effort to evaluate LM4 at all SNOTEL sites in the western US. Model evaluation using fully downscaled forcing data will enable detailed characterization of model behavior across a wide variety of geographies. Once model behavior is adequately characterized, future predictions for the southwest and NM made by the GFDL ESM can be interpreted with less uncertainty. Water resources in the southwest are highly dependent on winter snowfall, so a better understanding of how snow will change in the future also improves our ability to characterize stream flow and groundwater recharge. Furthermore, it may be possible to identify any model weaknesses specific to the unique climate and geography of the southwest, which could then be addressed in future versions of the model leading to ever more accurate snow projections. Results will be useful for all water agencies with headwaters dependent on snowmelt (Rio Grande, northern/central NM, Pecos Valley).

Describe how you have spent your grant funds:

As per the budget in the proposal, almost all funding has been used for a 3 month graduate research assistant stipend. A majority of the work was completed in June, July, and August 2025. A small amount of funding for data storage was not utilized as existing storage turned out to be sufficient.

List presentations you have made related to the project:

- WRRRI Fall Meeting Poster
- NMGS Spring Meeting Poster
- NMT E&ES graduate student seminar
- AGU Annual meeting 2025 (Abstract Submitted)

List publications or reports, if any, that you are preparing:

We anticipate developing a publication after applying our methodology to all SNOTEL sites across the Western US.

List any other students or faculty members who have assisted you with your project:

Enrico Zorzetto. I also received some assistance from Sergey Malyshev of NOAA GFDL.

Provide special recognition awards or notable achievements as a result of the research including any publicity such as newspaper articles, or similar:

None

Provide information on degree completion and future career plans:

Anticipated graduation is 2028 with plans for continued scientific research in academia or private settings.

References

- Chaney, N.W., et al., "Harnessing Big Data to Rethink Land Heterogeneity in Earth System Models." *Hydrology and Earth System Sciences* 22, no. 6 (June 14, 2018): 3311–30. <https://doi.org/10.5194/hess-22-3311-2018>.
- Dunbar, N.W., et al. *Climate Change in New Mexico Over the Next 50 Years: Impacts on Water Resources*. New Mexico Bureau of Geology and Mineral Resources, 2022. <https://doi.org/10.58799/B-164>.
- Hao, Dalei, et al. "Evaluation of E3SM land model snow simulations over the western United States." *The Cryosphere* 17.2 (2023): 673-697.
- Jones, A.S., Aanderud, Z.T., Horsburgh, J.S., Eiriksson, D.P., Dastrup, D., Cox, C., Jones, S.B., Bowling, D.R., Carlisle, J., Carling, G.T., Baker, M.A., 2017. Designing and Implementing a Network for Sensing Water Quality and Hydrology across Mountain to Urban Transitions. *JAWRA Journal of the American Water Resources Association* 53, 1095–1120. <https://doi.org/10.1111/1752-1688.12557>
- Shevliakova, E., Malyshev, S., Martinez-Cano, I., Milly, P.C.D., Pacala, S.W., Ginoux, P., Dunne, K.A., Dunne, J.P., Dupuis, C., Findell, K.L., Ghannam, K., Horowitz, L.W., Knutson, T.R., Krasting, J.P., Naik, V., Phillipps, P., Zadeh, N., Yu, Y., Zeng, F., Zeng, Y., 2024. The Land Component LM4.1 of the GFDL Earth System Model ESM4.1: Model Description and Characteristics of Land Surface Climate and Carbon Cycling in the Historical Simulation. *J Adv Model Earth Syst* 16, e2023MS003922. <https://doi.org/10.1029/2023MS003922>
- Serreze, M.C., Clark, M.P., Armstrong, R.L., McGinnis, D.A., Pulwarty, R.S., 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research* 35, 2145–2160. <https://doi.org/10.1029/1999WR900090>
- Sun, N., Yan, H., Wigmosta, M.S., Leung, L.R., Skaggs, R., Hou, Z., 2019. Regional Snow Parameters Estimation for Large-Domain Hydrological Applications in the Western United States. *Journal of Geophysical Research: Atmospheres* 124, 5296–5313. <https://doi.org/10.1029/2018JD030140>
- Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., Leung, R., 2018. Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design in Snow-Dominated Environments. *Water Resources Research* 54, 1093–1108. <https://doi.org/10.1002/2017WR021290>

Zorzetto, E., Malyshev, S., Ginoux, P., Shevliakova, E., 2024. A Global land snow scheme (GLASS) v1.0 for the GFDL Earth System Model: Formulation and evaluation at instrumented sites. EGU sphere 1–41. <https://doi.org/10.5194/egusphere-2024-506>