

Effects of Meteorological Forcing Uncertainty on High-Resolution Snow Modeling and Streamflow Prediction in a Mountainous Karst Watershed

Conor Tyson^{1,2}, Qianqiu Longyang³, Bethany T. Neilson¹, Ruijie Zeng³, Tianfang Xu^{3*}

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

² Jordan Valley Water Conservancy District, West Jordan, UT.

³ School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ.

* Corresponding author: Tianfang Xu (tianfang.xu@asu.edu)

Highlights:

- Hybrid snow and deep learning models simulate snowmelt and streamflow of a mountainous karst watershed
- Choice of meteorological forcing and downscaling led to different spatiotemporal patterns of snowmelt
- The deep learning model learned streamflow responses that vary across meteorological forcing sets
- Averaging streamflow simulated from all meteorological forcing yielded the highest accuracy

Abstract

In the mountainous Western U.S., a considerable portion of water supply originates as snowmelt passing through karst watersheds. Accurately simulating streamflow in snow-dominated, karst basins is important for water resources management. However, this has been challenging due to high spatiotemporal variability of meteorological and hydrogeological processes in these watersheds and scarcity of climate stations. To overcome these challenges, a physically based snow model is used to simulate snow processes at 100 m resolution, and the calculated snowmelt and potential evapotranspiration rates are fed into a deep learning model to simulate streamflow. The snow model was driven by meteorological variables from a regional scale Weather Research and Forecasting (WRF) model or from the North American Land Data Assimilation System (NLDAS-2). The two datasets were used both at the original resolution and downscaled to 100 m resolution based on orographic adjustments, leading to four sets of forcings. Snow model simulation results from the four sets of forcings showed large differences in simulated snow water equivalent (SWE) and snowmelt rate and timing. However, the differences were damped in simulated streamflow, as the deep learning model is partially immune to input bias and picked up different streamflow responses to snowmelt and rainfall when trained using snow model results. While the meteorological datasets considered yielded close streamflow simulation accuracy, averaging simulated streamflow from the four sets of forcings consistently achieved better performance, suggesting the value of including multiple meteorological datasets for modeling streamflow in mountainous watersheds.

Keywords

Snow hydrology; Karst; Rainfall-runoff modeling; Deep learning; Downscaling

1. Introduction

In many mountainous regions of the world, much of the water for residential and agriculture use originates as snowpack (Adam et al., 2009). Some of these snow-dominated watersheds include karst formations where carbonate rock has undergone dissolution resulting in fissures and conduits. Water is able to flow through karst conduits much faster than in porous matrices. Karst watersheds supply water to approximately one fourth of the world's population (Hartmann et al., 2014). Streamflow in karst, snow-dominated watersheds is controlled both by snow processes and karst hydrogeology. Accurate quantification of snow accumulation (often measured as snow water equivalence, SWE) and melt in these snow-dominated, karst regions are essential for simulating streamflow and predicting water supply availability.

SWE levels and melt rates in mountainous regions have high spatial variability due to topography (i.e., elevation, slope, aspect) and canopy coverage and their combined effects on precipitation, temperature, and radiation (Shamir and Georgakakos, 2006; Clark et al., 2011; Winstral et al., 2014). In mountainous karst watersheds, small scale (≤ 1 km) spatial variability of snow processes may have a significant effect on the timing and magnitude of streamflow. Due to hydrogeologic heterogeneity, portions of meltwater feed the karst aquifer by entering sinkholes directly connected to karst conduits, diffusing through small fissures or fractures, or infiltrating the soil matrix that slowly diffuses into karst conduits (Hartmann et al., 2014; Spangler, 2011; White, 2002). Depending on where the snowmelt occurs, recharge can take a wide range of travel times to reach the stream channel, varying from days to years (Spangler, 2011; Goldscheider and Drew, 2014). In addition, karst watersheds frequently display “piracy”, where water in one watershed flows across topographic watershed boundaries into neighboring watersheds (Hartmann et al., 2014; Spangler, 2011).

At the mesoscale (100 – 1000 km²), the spatial variability in snow processes can be resolved by high resolution snow modeling (Schlögl et al., 2016; Winstral et al., 2014). These models require topography (e.g., elevation), land use (e.g., canopy coverage), and meteorological forcing data at scales appropriate for the model grid size. Due to the scarcity of ground based observations, especially in mountainous areas, distributed hydrologic models typically rely on meteorological datasets derived by reanalysis, interpolation of ground stations, remote sensing techniques, or the combination of two or three methods. Uncertainty in meteorological variables, precipitation in particular, has been recognized as a key source of uncertainty in hydrologic modeling (Hong et al., 2006; Salamon and Feyen, 2009; Strauch et al., 2012; Eum et al., 2014; Fallah et al., 2020). In addition, meteorological data are generally provided at coarse resolutions (e.g., 4 km, 0.125°) (Mizukami et al., 2016; Shamir and Georgakakos, 2006). Methods have been developed for downscaling meteorological variables to the resolution of hydrologic models (e.g., ≤ 1 km) (Hungerford et al., 1989; Liston and Elder, 2006; Thornton et al., 2012; Fiddes and Gruber, 2014; Sen Gupta and Tarboton, 2016; Mital et al., 2022) by using fine resolution topography data and adjusting meteorological variables according to known topographical effects on climate. Unfortunately, this downscaling process brings additional uncertainty to the modeling process (Dibike et al., 2007; Behnke et al., 2016; Shuai et al., 2022).

While numerous studies have applied various hydrologic models to a range of watersheds, due to the complexities of karst geology traditional hydrologic models fail to adequately model flow in karst basins. Therefore, a hybrid modeling approach that combines a spatially distributed, physically based snow model and a deep learning model has been developed to overcome the lack of information about subsurface hydrologic connectivity that is common in these watersheds (Xu et al., 2022). In this hybrid model, the snow model simulates

snow accumulation and melt processes at 100 m resolution, which are fed into a deep learning model that simulates streamflow response to snowmelt. The deep learning model is based on the ConvLSTM architecture (Shi et al., 2015) capable of handling spatiotemporal dynamics such as the precipitation-discharge processes in the mountainous karst watersheds. Unlike physically based or conceptual karst models, machine learning and deep learning models do not require site specific knowledge of subsurface hydrologic connectivity (Li et al., 2017). In addition, they often do not impose mass balance and use scaled input data (e.g., precipitation or snowmelt). While various studies have investigated the effects of meteorological forcing uncertainty on hydrologic modeling using physically based models (Elsner et al., 2014; Eum et al., 2014; Maina et al., 2020; Shuai et al., 2022; Rasouli et al., 2022), such understanding has been lacking for deep learning models.

This study aims to understand how the choice of meteorological datasets and downscaling techniques affect simulating snow accumulation, melt and streamflow focused on a snow-dominated mountainous karst watershed on the Utah-Idaho border. We hypothesize that physically based simulation of spatially varying variables such as SWE and snowmelt is more sensitive to the uncertainties in meteorological forcing than deep learning-based simulation of streamflow. To test this hypothesis, this study employs modeling experiments to: (1) quantify the spatial and temporal patterns of SWE and snowmelt rate and how these patterns are affected by the choice of meteorological inputs and downscaling methods, and (2) simulate streamflow response to snowmelt, which is modeled from different meteorological datasets and downscaling techniques, using a deep learning model and determine its accuracy.

2. Study Site

Our study area is the canyon portion of the Logan River watershed located in northeastern Utah and southeastern Idaho (Figure 1a). This portion of the Logan River watershed has an area of 550 km² and an elevation range from 1366 m to 3037 m (Figure 1b). The study area contains both coniferous and deciduous forested areas (Figure 1c). Average basin-average precipitation is about 1080 mm, and more than 50% of the precipitation falls as snow, according to the Weather Research & Forecasting (WRF)-derived meteorological dataset used in this study (Section 3.1). The river flows primarily from the north and east to the south and west of the watershed. However, developed karst conduits and sinkholes in the watershed add complexity to subsurface water flow direction. The karst features in and around the topographically defined watershed result in significant karst piracy, requiring the study area to include terrain outside the topographic watershed, particularly in areas with known karst subsurface connections (Spangler, 2011).

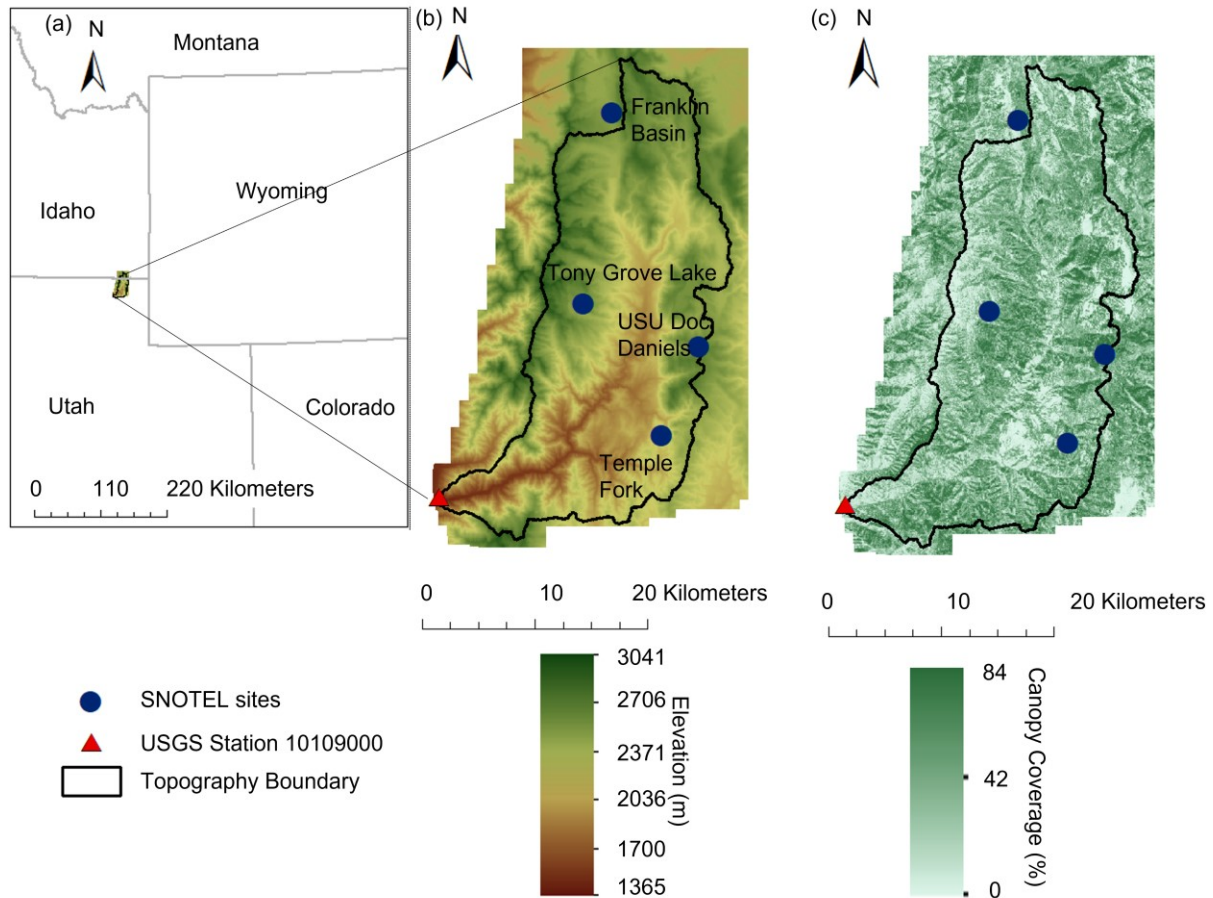


Figure 1. Maps showing (a) location, (b) elevation, and (c) canopy coverage (NLCD, 2019) of the study area. In addition, topography boundary of Logan River watershed and locations of SNOTEL sites and the USGS station are indicated.

Hydrometeorological data of the study area are available from seven SNOTEL stations. Among them, four stations have data of three or more years and will be used in this study (Table 1). Discharge records of the Logan River are provided by USGS station 10109000 (Fig. 1) located at the watershed outlet since 1954. Two diversions, via the Highline Canal and at the Dewitt Spring, exist upstream of the USGS gage for agricultural and municipal water uses. Daily diversion rates via the Highline Canal are obtained from USGS station 10108400. Monthly use

of Dewitt Spring discharge was obtained from Logan City and assumed to occur evenly over each month. Daily naturalized streamflow was determined by summing the observed streamflow at the USGS gage and the two diversion rates. The combined rates from the three sites are hereafter referred to as observed streamflow.

Table 1: SNOTEL stations in the study area with three or more years of records (USDA NRCS, 2022). Canopy coverage is calculated for the UEB grid where the station is located.

Site Name	Year Data Begins	Latitude	Longitude	Elevation (m)	Canopy Coverage (%)
Franklin Basin	1979	42.05	-111.6	2481	29
Temple Fork	2001	41.79	-111.55	2257	38
Tony Grove Lake	1978	41.9	-111.63	2583	34
USU Doc Daniels	2007	41.86	-111.51	2521	43

3. Methods

In order to assess how uncertainty in meteorological forcing would affect the simulation of snow processes and streamflow within the Logan River watershed, we performed modeling experiments using various meteorological forcing and compare the simulation results (Fig. 2). We used two forcing datasets, at their raw resolution and downscaled, respectively, leading to four sets of meteorological forcing. A 100-m spatial resolution and 1 hour temporal resolution Utah Energy Balance (UEB) model (Mahat and Tarboton, 2012; Tarboton and Luce, 1996) is used to simulate snowmelt and rainfall. The simulated snowmelt and rainfall as well as potential

evapotranspiration were fed into the deep learning model to simulate streamflow (Xu et al. 2022).

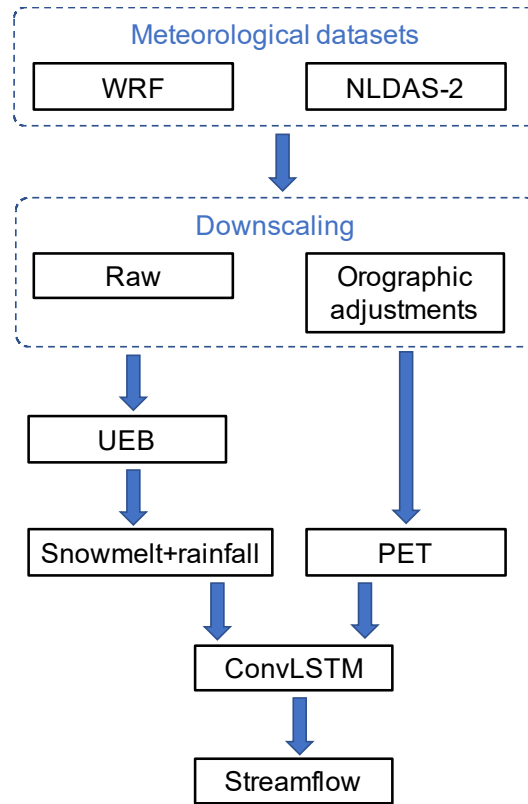


Figure 2. Flowchart of the modeling experiments using meteorological forcing generated from two meteorological datasets, with and without downscaling, to drive the Utah Energy Balance (UEB) and Convolutional Long Short-term Memory (ConvLSTM) models.

3.1. Meteorological forcing and orographic adjustments

Meteorological variables required to run the UEB model were obtained from two sources of meteorological data. The first is a dynamically downscaled dataset based on simulation results of a Weather Research & Forecasting (WRF) model centered on the Wasatch Mountains, Utah (Scalzitti et al., 2016). The WRF model used initial and boundary conditions derived from Climate Forecast System Reanalysis. The dataset is available at 4 km resolution and hourly time

steps from 1985 to 2010. The simulated precipitation was validated using SNOTEL and Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al., 2008) datasets. Because it is dynamically downscaled, the WRF dataset has an advantage of resolving processes over complex topography at a scale much finer than the resolution of Global Climate Models (~50 to hundreds of kilometers).

The second meteorological dataset is the NLDAS-2 (Xia et al., 2012) forcing, a national dataset commonly used in hydrologic studies. The NLDAS forcing dataset has a 0.125° (~12 km) spatial resolution and 1 hour temporal resolution from 1979 to the present. Non-precipitation variables were derived primarily from the NCEP North American Regional Reanalysis (NARR, Lin et al., 2001; Mesinger et al., 2006). The NARR meteorological variables at 32-km resolution and 3-hourly time steps were spatially interpolated to 0.125° and temporally disaggregated to hourly (Cosgrove et al., 2003). Precipitation was derived by temporally disaggregating the Climate Prediction Center (CPC) PRISM-adjusted daily CONUS gauge data using hourly Doppler Stage II radar (available since 1996, Lin and Mitchell, 2005). When the Stage II radar was not available, half-hourly CMORPH (available since 2002, Joyce et al., 2004), CPC Hourly Precipitation Dataset (Higgins et al., 1996), or NARR data were used in descending preference (Ferguson and Mocko, 2017).

To run the UEB model, meteorological variables (2 m humidity, surface downward shortwave surface radiation, surface downward longwave surface radiation, wind speed at 2 meters above the surface, temperature at 2 m above the surface, surface pressure, and precipitation) were derived from the WRF and NLDAS datasets both as raw data and with orographic adjustments. For raw data simulations, each UEB grid is assigned the value of meteorological variables from the location of the UEB grid. For orographic adjustment runs, the

raw meteorological variables were first bilinearly interpolated within the study area, and then adjusted to account for elevation and other topographical features using methods developed in Liston and Elder (2006) and Sen Gupta and Tarboton (2013), detailed in Text S1, Supplementary Material. Temperature was adjusted using a lapse rate derived from SNOTEL stations in the study area. Precipitation was adjusted based on a monthly factor and the elevation difference. Specific humidity was downscaled using elevation adjusted temperature calculated previously and the elevation difference and then converted to relative humidity for the UEB model. Incoming longwave radiation was altered first according to the adjusted temperature and then by calculating air emissivity for elevation. Shortwave radiation was altered according to atmospheric pressure that had been altered according to the elevation difference and monthly coefficients for the Bristow-Campbell parameterization of the shortwave radiation as a function of air temperature diurnal range. Wind was adjusted based on wind direction and the slope, aspect, and curvature of the grid cell and its neighboring cells.

3.2. Utah Energy Balance snow model

We simulated snow accumulation and melt at 100 m spatial resolution and 1 hour temporal resolution for the study area using the UEB model (Mahat and Tarboton, 2012; Tarboton and Luce, 1996). The UEB model is a physically based model utilizing mass and energy balances of the snowpack (both on land and intercepted by canopy) and can be run using a distributed (gridded) setup. We chose the 100 m resolution to capture the small scale variability in snow accumulation and melt. Based on preliminary results, changing resolution from 200 m to 100 m could cause SWE differing by more than 20% at model grids having the largest variabilities in slope and aspect due to the smoothing out of aspect and slope. However, changing resolution from 100 m to 50 m altered SWE levels by less than 5% at these grids, suggesting 100

m resolution adequately captured slope and aspect at reasonable computational cost. The UEB snow model was run for all years the WRF dataset was available (Oct. 1985 – Jun. 2010). UEB simulated hourly SWE and the combined snowmelt and rainfall were lumped to daily resolution by taking SWE from the last hour of the day, and summing up snowmelt and rainfall for each day.

Parameter values were set following recommendations in Tarboton and Luce (1996) and Mahat and Tarboton (2012). A digital elevation model (DEM) with 1/3 arc second (about 9 m) resolution (U.S. Geological Survey, 2017) was coarsened to the 100 m UEB grids and then used to calculate slope and aspect. Canopy coverage data was obtained from the NLCD 2011 tree canopy dataset (Coulston et al., 2012). Leaf area index and forest structure parameters were derived from NLCD land use dataset (Yang et al., 2018) following recommendations in Sen Gupta and Tarboton (2012). Canopy height data was obtained from NASA Landfire database (Nelson et al., 2013).

UEB simulation results are compared with SWE observations at SNOTEL stations and MODIS fractional snow-covered area (SCA) product at 500 m resolution (MOD10A1, Hall and Riggs, 2016). SWE recorded at end of day is compared with simulated SWE at the UEB grid containing the SNOTEL site. The MOD10A1 SCA product is derived based on Normalized Difference Snow Index (NDSI). The product is known to have lower accuracy in forested areas (Hall and Riggs, 2016; Yang et al., 2015). However, it is the only publicly available spatially observed snow dataset with desired coverage and resolution. In this study, we use MOD10A1 as a secondary validation that is independent from SNOTEL stations and provides spatial information. To compare MOD10A1 SCA to UEB results, the percent of UEB grids in each MOD10A1 grid with SWE above 10 mm was calculated to determine the UEB simulated SCA.

Experiments showed that varying SWE threshold results in negligible difference in SCA during December to June.

3.3. Deep learning karst model

A deep learning model using the Convolutional Long Short-Term Memory (ConvLSTM) architecture is applied to simulate Logan River streamflow based on the UEB model outputs. The implementation of the architecture used in this work is the same as in Xu et al. (2022) and shown in Fig. 3. The ConvLSTM architecture (Shi et al., 2015) integrates convolutional networks into LSTM (Kratzert et al., 2018):

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + b_i),$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + b_f),$$

$$o_t = (W_{xo} * x_t + W_{ho} * h_{t-1} + b_o),$$

$$c_t = i_t \odot \tanh(W_{xg} * x_t + W_{hg} * h_{t-1} + b_g) + f_t \odot c_{t-1},$$

$$h_t = o_t \odot \tanh(c_t),$$

where σ represents the sigmoid function, $\tanh$ is the hyperbolic tangent function, $*$ is convolutional operation, and $\odot$ is the element-wise multiplication. At each time step, new inputs (x_t) are combined with hidden state from the previous time step h_{t-1} to determine the input gate (i_t), forget gate (f_t), and output gate (o_t). The cell memory c_t contains historical information and is analogous to the watershed storage. The input gate controls how new information is used to update the cell memory, and the forget gate determines how much of the old cell memory c_{t-1} will be forgotten. The output gate is used to calculate the hidden state (h_t) from the cell memory. The convolutional operation applies a filter that sweeps through the spatially distributed x_t and h_{t-1} (Fig. 3). In this way, ConvLSTM calculates cell memory of a given grid using information

of itself as well as its neighbor grids. We stack three layers of ConvLSTM, where the hidden state calculated by each layer is used as input by the next layer. The hidden state of the third layer at the last time step passes through a fully connected layer to output streamflow (Q_{t+1}).

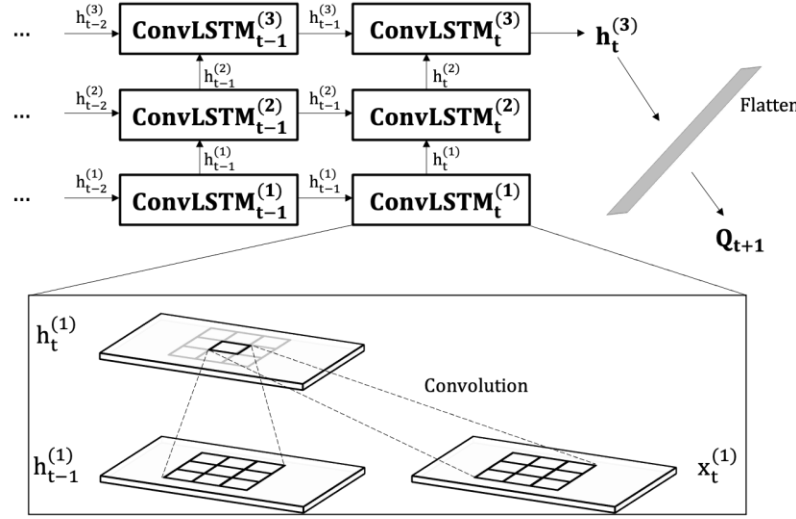


Figure 3. The architecture of the ConvLSTM model with three stacked ConvLSTM layers. At time step t , the ConvLSTM cell applies a convolutional operation and updates the hidden state of each grid using input and past hidden state of this grid and its neighbors. The input x_t is processed by layer 1, and the resulting $h_t^{(1)}$ is fed into layer 2. Similarly, the hidden state of layer 2, $h_t^{(2)}$, is used by layer 3 to calculate hidden state $h_t^{(3)}$. Lastly, a fully-connected layer outputs streamflow using flattened (vectorized) $h_t^{(3)}$.

UEB simulated hourly snowmelt and rainfall combined was lumped to daily to generate inputs to the ConvLSTM model. The inputs also include daily potential evapotranspiration (PET) rates calculated using the Hamon method (Hamon, 1960) from the raw and orographically adjusted meteorological forcing, respectively. The 100 m resolution PET and snowmelt and rainfall were aggregated to 1.6 km resolution grids before being used as inputs to the ConvLSTM

model. Given that the UEB simulation uses meteorological datasets downscaled to 100 m resolution, the process of scaling results back up seems counterintuitive. However, the high resolution snow predictions are necessary to capture the spatially varying meteorological forcing and snow accumulation and melt processes as controlled by topography in mountainous areas (Schlögl et al., 2016; Winstral et al., 2014). To keep the deep learning computational costs reasonable, the detailed snow predictions are coarsened, but the values representing the larger grid size now accounts for the within grid variability (Xu et al., 2022).

Because the distribution of streamflow is heavy-tailed, we performed a nonlinear transformation:

$$Y = \ln(\sqrt{\tilde{Y}} + 1),$$

where $\tilde{Y}$ is the original observed or simulated streamflow. Next, $\tilde{Y}$ and the inputs (snowmelt and rainfall, PET) were all normalized, i.e., linearly scaled to the range of [0, 1]. Data from water years (WY) 1986-2002 and 2003-2010 were used as training and test datasets for the ConvLSTM model. Dropout was implemented as described in Gal and Ghahramani (2016a, b) to mitigate overfitting. Training was performed using Adam optimizer (Kingma and Ba, 2014) for 500 epochs. From the last 100 epochs, we save the models at randomly selected 10 epochs. During the test period, we run each model for 20 times using random dropout masks (Gal and Ghahramani, 2016b). This results in 200 time series of simulated streamflow for each of the four meteorological forcing. The average of 200 realizations was calculated as the final simulated streamflow.

3.4. Evaluation of simulated streamflow

The simulated streamflow using two meteorological forcing datasets at their original resolution and downscaled is evaluated using three different statistics: percent bias (PBIAS, Gupta et al., 1999), Root Mean Square Error (RMSE), Nash-Sutcliffe efficiency (NSE), and Kling-Gupta efficiency (KGE, Gupta et al., 2009), as defined below:

$$PBIAS = \frac{\sum(Y_o - Y_m)}{\sum Y_o} \times 100\% \quad (4)$$

$$RMSE = \sqrt{\frac{\sum(Y_o - Y_m)^2}{n}} \quad (5)$$

$$NSE = \frac{\sum(Y_o - Y_m)^2}{\sum(Y_o - \bar{Y}_o)^2} \quad (6)$$

$$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma_m}{\sigma_o} - 1\right)^2 + \left(\frac{\bar{Y}_m}{\bar{Y}_o} - 1\right)^2} \quad (7)$$

In Eqn. (4-7), Y_o is the observed streamflow, Y_m is streamflow simulated by the ConvLSTM models, n is the number of data points, $\bar{Y}_o, \bar{Y}_m$ are the average of the observed and simulated streamflow values, respectively, r is the correlation coefficient between Y_o, Y_m , and σ_o, σ_m are the standard deviation of Y_o, Y_m . A positive PBIAS suggests the simulations are overall lower than the measurements, while a negative PBIAS shows the simulations are generally higher than measurements. Lower RMSE and NSE and KGE closer to one indicate better fit to measurements.

In addition to the four performance metrics, we calculated the time lag between UEB simulated rainfall plus snowmelt and ConvLSTM simulated streamflow as a measure of the timing of streamflow response to snowmelt and rainfall learned by the ConvLSTM model (Yilmaz et al., 2008). This was done by calculating the correlation coefficient between simulated streamflow and snowmelt plus rainfall with varying lags; the lag that maximizes the correlation was selected as the time lag signature. Following the recommendation in Yilmaz et al. (2008),

the calculation was performed using high flow with exceedance probability less than or equal to 0.2. The time lag was compared to lag calculated for observed streamflow using raw and orographically adjusted WRF and NLDAS forcing.

4. Results

4.1. Comparing meteorological variables from different datasets and downscaling methods

Noticeable differences are found among raw and orographically adjusted WRF and NLDAS meteorological forcing, especially in the amount of precipitation (Fig. 4). During the study period (WY 1986-2010), WRF annual cumulative precipitation is 15.5% higher than NLDAS on average and ranges from NLDAS being 35% higher than WRF precipitation in 1986 to WRF being 41% higher than NLDAS in 2006. Orographic adjustments increased annual average precipitation by 3.4% for both WRF and NLDAS, less than the difference between the two datasets and with little interannual variation. Besides precipitation, the WRF dataset has overall lower temperature, significantly higher incoming shortwave radiation, lower incoming longwave radiation, higher wind speeds, and higher relative humidity than NLDAS dataset. Orographic adjustment increased spatial variability, suggested by flatter distributions, particularly for NLDAS forcing (Fig. 4). For both datasets, orographic adjustment led to overall lower temperature and slightly higher incoming shortwave radiation, while its effects on incoming longwave radiation differ between the two datasets.

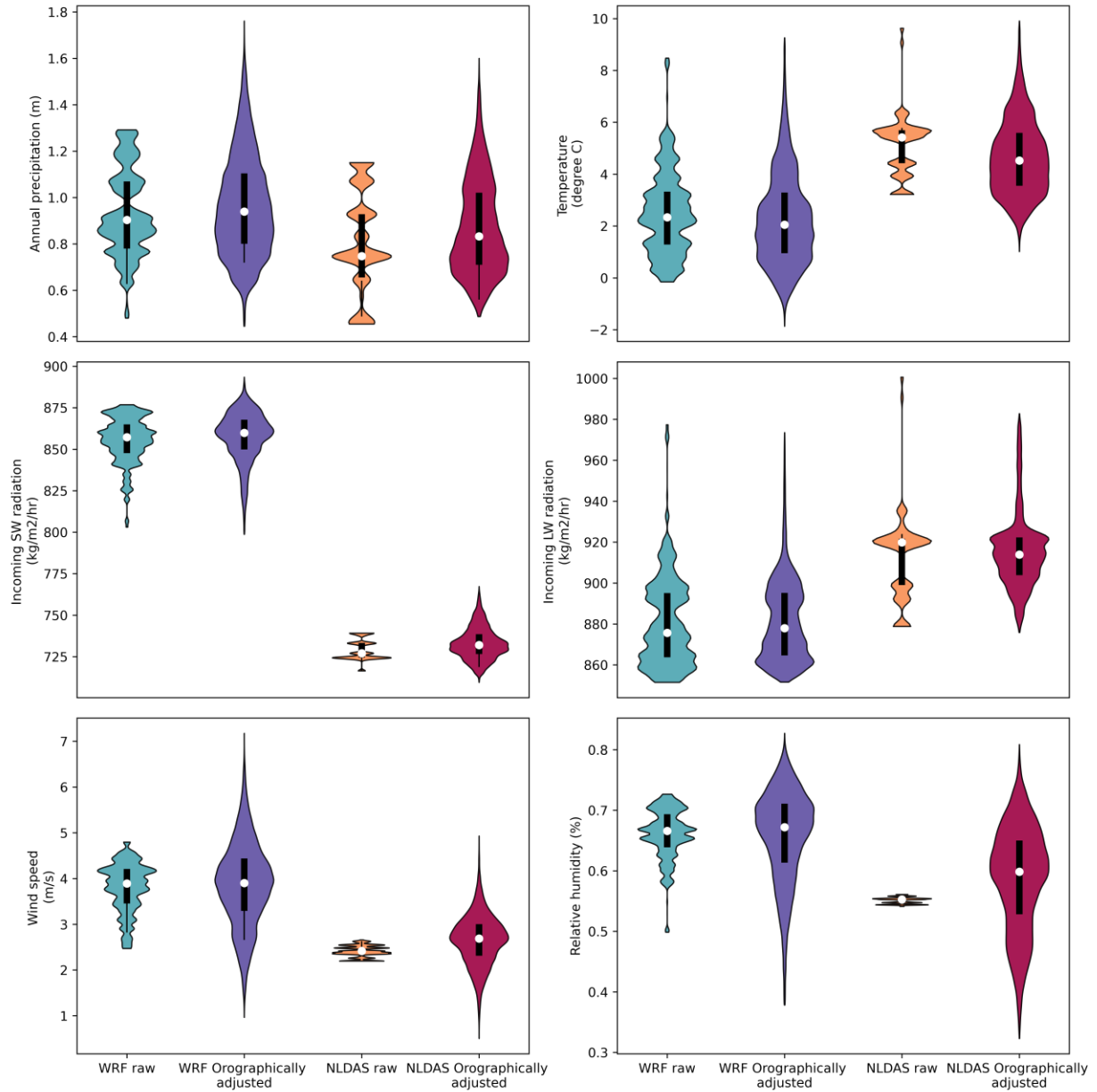


Figure 4. Violin plots showing probability density of annual total precipitation, temperature, incoming shortwave (SW) and longwave (LW) radiation, wind speed, and relative humidity from raw and orographically adjusted WRF and NLDAS forcing, averaged during WY 1986-2010. Colored shades, dark boxes, whiskers, and white dots show density, 1st and 3rd quantiles, inner fences, and median values.

4.2. SWE and snowmelt rates simulated using different datasets and downscaling methods

The differences in the meteorological forcing propagated to UEB simulated SWE (Fig. 5). On average during 1986-2010, annual peak SWE simulated using the raw WRF dataset is 90% higher than raw NLDAS and can vary between 19% in 1998 to 170% in 2003. The orographically adjusted WRF dataset resulted in SWE levels 9% higher than raw, ranging from 7% higher in 1989 to 15% higher in 2001, while orographic adjustment of NLDAS data results in SWE levels increased by an average of 12%, ranging from 9% in 1986 to 19% in 2010. The differences in SWE are higher than the differences in precipitation due to orographic adjustment and especially the choice of meteorological dataset (Fig. S2). In addition, even in years when the NLDAS dataset had higher precipitation (e.g., WY 1998), WRF still led to higher SWE. This highlights that the other meteorological variables affected SWE levels. Comparison with SNOTEL SWE measurements shows an underestimation of SWE using raw NLDAS forcing in most years especially at the Tony Grove Lake station, while overestimation is observed when using raw WRF (Fig. 5). Orographic adjustments increased SWE for both datasets, resulting in higher RMSE for WRF while lower RMSE for NLDAS (Table S2).

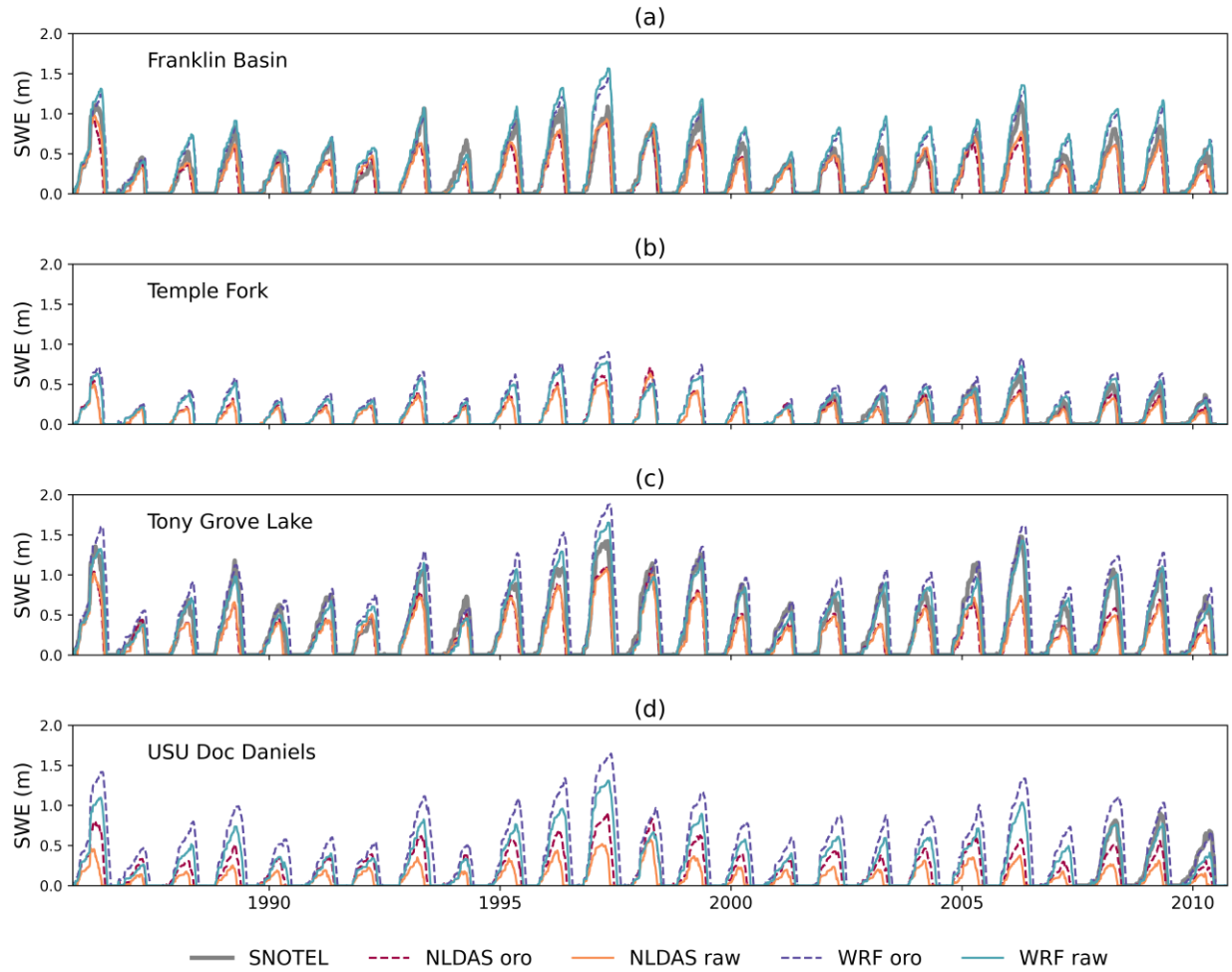


Figure 5. SWE measurements at SNOTEL stations (Fig. 1) and UEB simulated SWE in model grids where the SNOTEL stations are located. UEB simulations are driven by raw and orographically adjusted WRF and NLDAS meteorological datasets. SWE measurements are available during the study period for Franklin Basin and Tony Grove Lake stations, since WY2002 at Temple Fork, and since WY 2008 at Doc Daniels.

Simulation results from the four sets of meteorological forcing also differ in fractional SCA compared to MODIS SCA (MOD10A1 product, Hall and Riggs, 2016) (Fig. 6). The WRF dataset resulted in higher SCA than NLDAS; both datasets led to significantly higher SCA than MOD10A1. MOD10A1 is known to have difficulty detecting snow with canopy presence,

particularly in complex terrains (Hall and Riggs, 2016; Huang et al., 2011; Shamir and Georgakakos, 2006). In the study area, we found that the MOD10A1 SCA tends to be substantially lower than 100% (Fig. S3-S6), while we observed full snow coverage at higher elevations during winter months. Because of the likely underestimation of MOD10A1 SCA, we focus the comparison on the monthly change of SCA rather than the absolute values. The NLDAS-forced runs led to similar melt period as the MOD10A1 SCA product, while the WRF forcing yielded later melt by approximately one month. For both datasets, orographic adjustment slightly delayed snow melt.

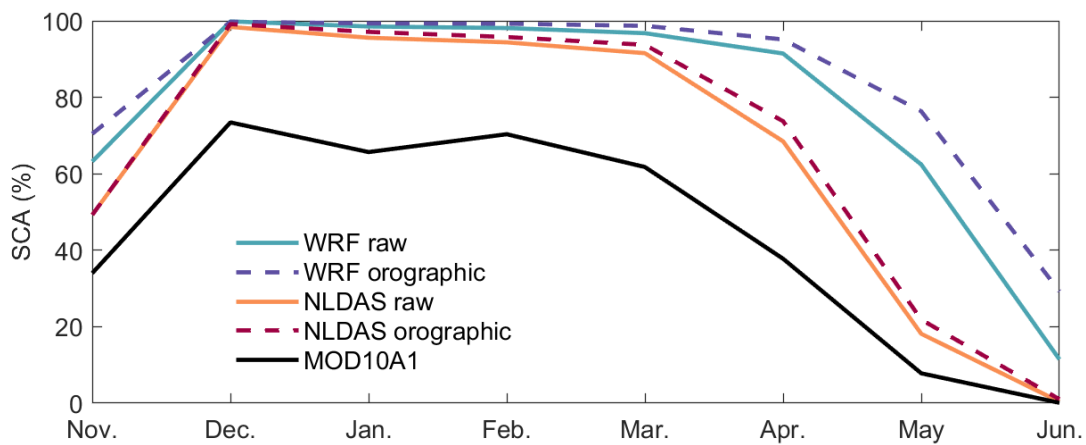


Figure 6. Multi-year and spatially average monthly snow cover area (SCA) from the MOD10A1 product and UEB simulations using different combinations of meteorological datasets and downscaling methods during WY 2001-2010.

Figure 7 further shows spatial distribution of multi-year average snowmelt and rainfall in April, May, and June, according to UEB runs forced by the raw and orographically adjusted WRF and NLDAS forcing. Results simulated with the raw data show discontinuity patterns due to the coarse resolutions of raw data (4 km for WRF, and 12 km for NLDAS). Compared to

NLDAS runs, the WRF dataset led to lower snowmelt occurring in April and higher snowmelt in June. At higher elevations, because orographic adjustment led to higher SWE levels, it yielded higher monthly snowmelt rates occurring at later times.

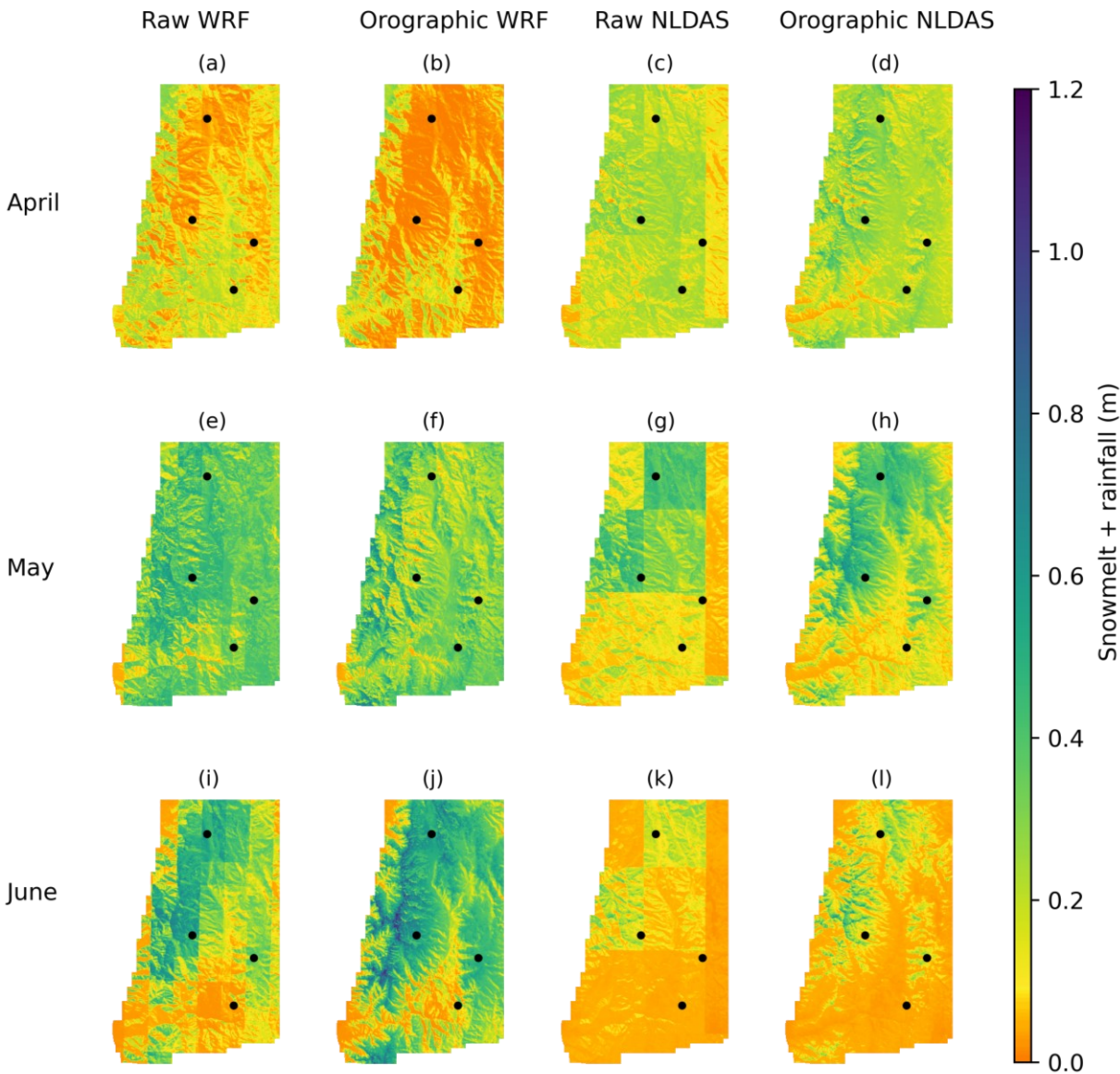


Figure 7. Monthly rainfall and snowmelt for April (top), May (middle), and June (bottom) averaged from 1986 to 2010 arising from the raw and orographically adjusted WRF and NLDAS meteorological datasets. Dots show the location of SNOTEL sites (Fig. 1).

4.3. Effects of different melt rate and timing on streamflow simulation

Testing the performance of the deep learning models using UEB simulated snowmelt driven by the WRF and NLDAS datasets with and without orographic adjustments illustrated that the WRF dataset led to overestimation of streamflow (negative PBIAS), while the NLDAS dataset led to underestimation (positive PBIAS). Other performance metrics (RMSE, NSE and KGE) show similar test accuracy among the four deep learning models. The WRF-driven simulations yielded lower RMSE, higher NSE, and higher KGE than simulations driven by NLDAS, although the differences in the performance metrics are small. Orographic adjustment improved streamflow accuracy at a slightly higher degree for NLDAS than WRF. This is likely because orographically adjusting the WRF dataset worsened the overestimation of SWE, thus deteriorating overestimation of streamflow. Because WRF and NLDAS forcing led to offsetting PBIAS, we additionally calculated an ensemble mean, as the average of streamflow simulated with raw and orographically adjusted WRF and NLDAS forcing. The ensemble mean yielded overall the best performance, suggested by lowest RMSE, highest NSE, slightly higher KGE, and the second lowest PBIAS (Table 2). The improvement in RMSE and NSE yielded by the ensemble mean appears more significant than the differences caused by orographic adjustment and the choice between WRF and NLDAS datasets.

The time lag that maximizes correlation between both simulated and observed streamflow and UEB simulated snowmelt and rainfall shows an overall trend where the time lag of simulated streamflow agrees with that of observed streamflow but tends to be slightly higher (Table 3). Driving the UEB model using the WRF dataset led to much shorter lag (i.e., fast

response) between streamflow and snowmelt than using the NLDAS dataset, and orographic adjustment led to slightly shorter lags.

Table 2. NSE, MSE, PBIAS, and KGE during the test period of simulated streamflow based on the WRF and NLDAS datasets at original resolution and downscaled.

Meteorological	Downscaling	PBIAS	RMSE	NSE	KGE
Data	Method	(%)	(mm/day)		
WRF	Raw	-6.37	0.400	0.827	0.876
	Adjusted	-7.22	0.385	0.840	0.887
NLDAS	Raw	4.61	0.434	0.797	0.864
	Adjusted	1.14	0.402	0.826	0.884
Ensemble Mean		-1.96	0.326	0.885	0.889

Table 3. Time lag in days corresponding to the maximum cross-correlation coefficient between observed and simulated streamflow and spatially averaged snowmelt and rainfall time series simulated using raw and orographically adjusted WRF and NLDAS forcing datasets.

Meteorological	WRF, raw	WRF, adjusted	NLDAS, raw	NLDAS, adjusted
forcing				
Observed streamflow	2	1	28	27
Simulated streamflow	4	1	32	28

Observed and simulated streamflow typically exhibits multiple peaks in response to several major snowmelt events (Fig. 8-10). Overall, WRF-driven simulations yielded higher streamflow peaks in May and June than NLDAS-driven simulations, resulting in a better fit to observed streamflow in WY 2006 and 2007 (Fig. 8, 9). However, WRF-driven simulations often

overestimate streamflow during recession period (Table 2, Fig. S7), such as in WY 2009 (Fig. 10), while sometimes underestimate in April and May (Fig. 8a). Orographic adjustments enhanced late-season peaks occurring in June and recession flow, but it also can dampen early-season peaks (Fig. 9a, 10a).

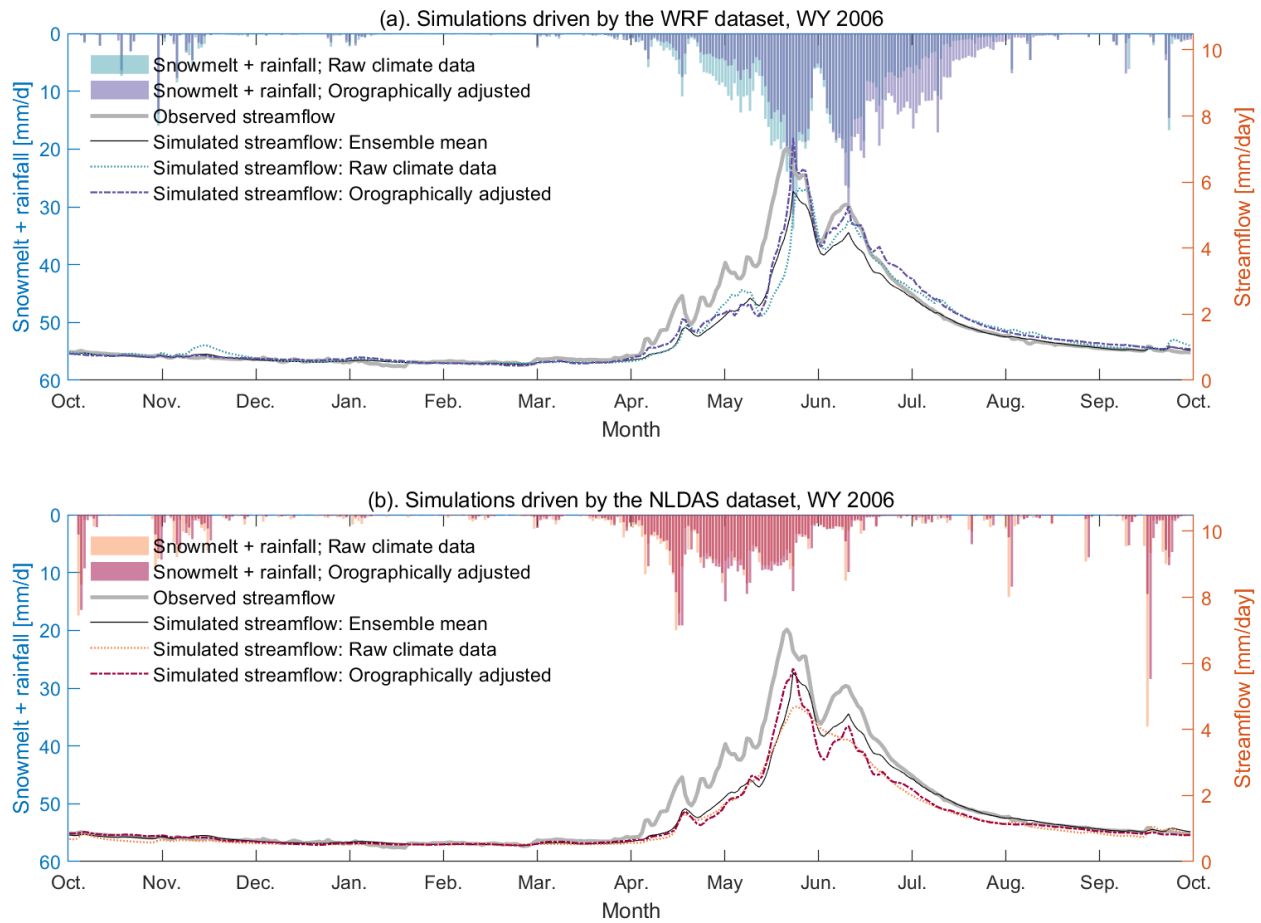


Figure 8. Observed (grey solid thick lines) and simulated streamflow in a wet year (WY 2006) using the WRF (a) and NLDAS (b) meteorological datasets in the raw form (dashed lines), orographically adjusted (dotted lines), and the ensemble mean (dark solid lines), respectively. Also shown are spatially averaged UEB simulated snowmelt and rainfall rates shown as daily sums.

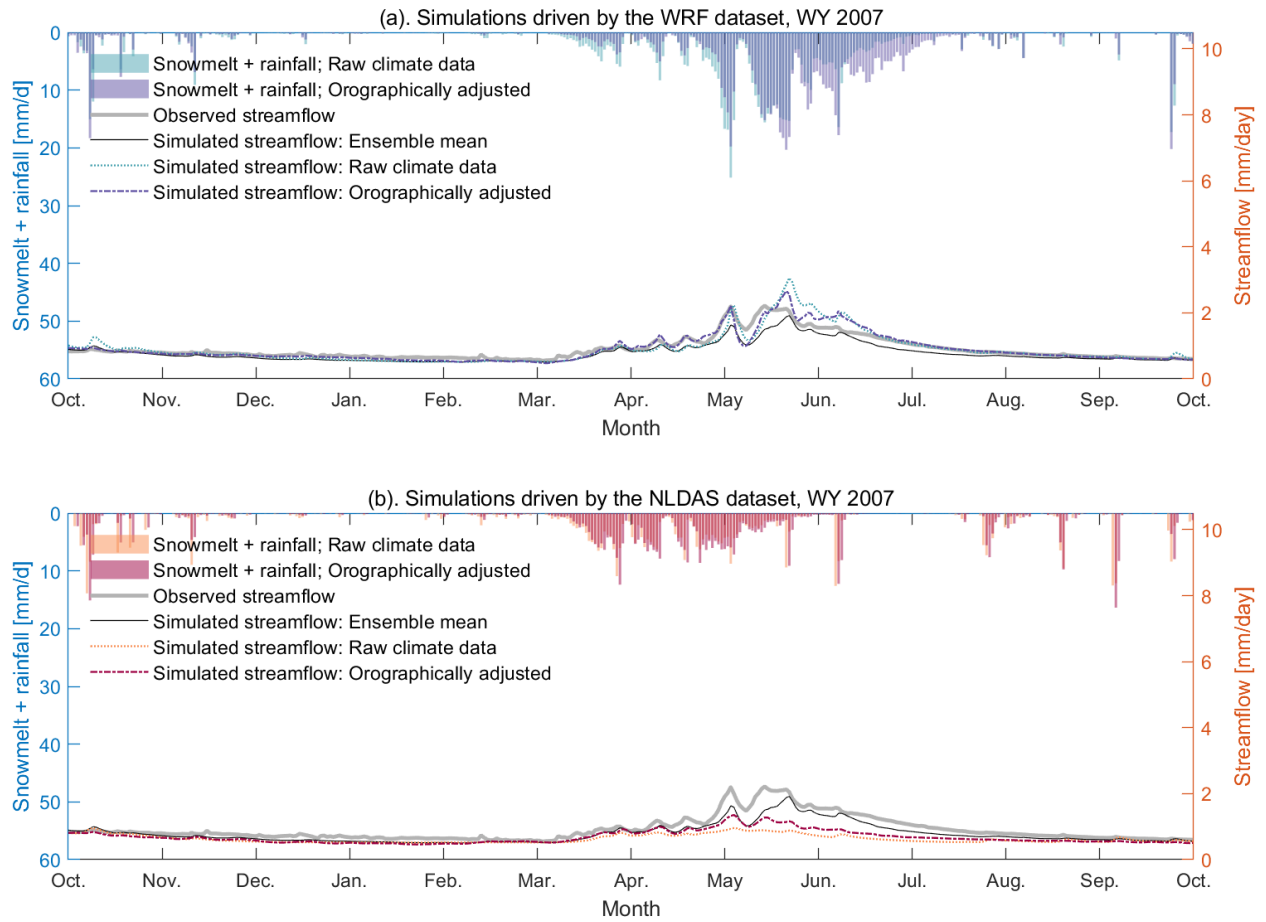


Figure 9. Observed (grey solid thick lines) and simulated streamflow in a dry year (WY 2007) using the WRF (a) and NLDAS (b) meteorological datasets in the raw form (dashed lines), orographically adjusted (dotted lines), and the ensemble mean (dark solid lines), respectively. Also shown are spatially averaged UEB simulated snowmelt and rainfall rates shown as daily sums.

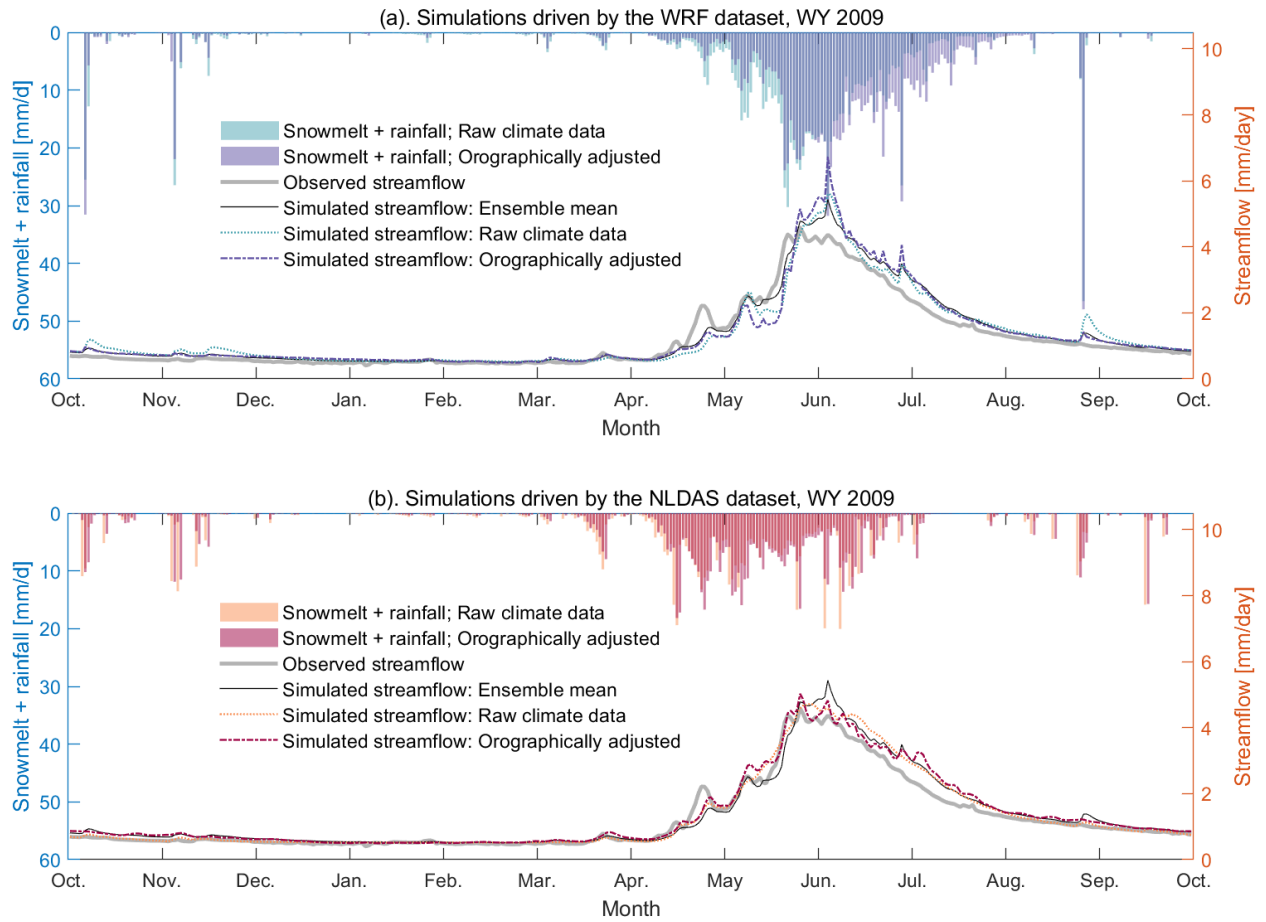


Figure 10. Observed (grey solid thick lines) and simulated streamflow in a normal year (WY 2009) using the WRF (a) and NLDAS (b) meteorological datasets in the raw form (dashed lines), orographically adjusted (dotted lines), and the ensemble mean (dark solid lines), respectively. Also shown are spatially averaged UEB simulated snowmelt and rainfall rates shown as daily sums.

5. Discussion

5.1. Value of representing topographic control on meteorological variables and snow processes

The substantially higher precipitation and SWE resulting from the WRF dataset than NLDAS forcing are consistent with findings in other studies (He et al., 2019; Wrzesien et al.,

2019) that compared the convection-permitting regional climate model simulations with statistically downscaled products. The differences in precipitation explain only a portion of the gaps in simulated SWE resulting from meteorological dataset and downscaling methods (Fig. S2) due to the effects of other meteorological variables, primarily air temperature. Higher air temperatures would result in less precipitation falling as snow as well as accelerate snow melt, leading to lower SWE. Similar underestimation of SWE and early snowmelt biases have been found in the NLDAS forcing for other North American mountain areas (Wrzesien et al., 2019; Shuai et al., 2022). Orographic adjustment increased watershed average precipitation and lowered temperature, leading to higher simulated SWE. On the other hand, compared to NLDAS the WRF dataset has higher incoming shortwave radiation and lower longwave radiation, which have offsetting effects on SWE. Higher wind speed and lower RH would decrease SWE, but these factors are likely to have a smaller effect than the other meteorological variables (Mizukami et al. 2014).

While the 100 m resolution UEB simulation already accounts for terrain controls on snow processes, when the coarse resolution (12 km) NLDAS meteorological forcing is used, the UEB model did not capture the snowmelt derived from deep snowpack southwest and east of the watershed (Fig. 7 c, g, k). Orographic adjustment overall improved the representation of SWE and snowmelt spatial variability (Fig. 5, 7 d, h, l, Table S2). On the other hand, as a dynamically downscaled dataset, the 4-km WRF forcing captures the terrain effects on precipitation reasonably well (Scalzitti et al., 2016; He et al., 2019), leading to more realistic spatial patterns of snowmelt and rainfall simulated by the UEB model than the raw NLDAS dataset (Fig. 7 a, e, i). However, orographic adjustment increased SWE RMSE at three of the four SNOTEL stations for WRF (Table S2), suggesting that it may be unnecessary to further downscale the 4-km

resolution WRF forcing. This supports previous finding in Rasmussen et al. (2011), while appearing to differ from Maina et al. (2020). The latter study found that meteorological forcing generated by WRF simulations at 0.5 km resolution yielded more accurate SWE than coarser resolution forcing. The different findings indicate that dynamic downscaling (as opposed to orographic adjustment) may be needed to adequately resolve spatial variabilities of forcing and SWE within 4 km resolution.

Despite mixed results with respect to SWE simulation accuracy (Table S2), orographically adjusted WRF and NLDAS driven UEB runs were able to capture snowmelt from deep snowpacks in high elevations (Fig. 7). Representing the spatial variability realistically is important for simulating streamflow, as a deep snowpack would sustain streamflow later into summer than a shallow snowpack. An additional factor is that due to the karst geology, the Logan River streamflow may be more sensitive to snowmelt at certain locations than others (Xu et al., 2022). For example, in the high elevation areas west of the study area, sinkholes and faults developed in outcropped carbonate units form pathways for fast recharge and conduit flow (Spangler et al., 2001; 2011).

5.2. Effects of differences in simulated snowmelt on the inference of streamflow response to snowmelt and rainfall

The overall trend of overestimating streamflow using the WRF dataset and underestimating streamflow using the NLDAS dataset particularly during spring runoff reflects the differences in simulated snowmelt (Figs. 5, 8-10, Table 2). Because WRF yields overall deeper snowpack that melts later than NLDAS, the resulting simulated streamflow could underestimate in early-season while overestimate in late-season. Similarly, orographic adjustments led to slightly higher late-season streamflow (Fig. 8-10). Late-season streamflow

peaks mostly result from snowmelt at higher elevations. This is where the snow melts last, and usually has significant enough volume of snowmelt to create major streamflow peaks. Orographic adjustment increases precipitation while lowering temperature at higher elevations (Text S1), leading to higher SWE and snowmelt volume in June and July (Fig. 7), and therefore larger streamflow peaks.

Despite the large difference in simulated snowmelt timing and rate, the ConvLSTM models simulated streamflow with similar accuracy among the runs using raw and orographically adjusted WRF and NLDAS datasets. This observation is consistent with findings in previous studies that streamflow simulated by physically based hydrologic models is less sensitive to the differences in meteorological forcing and resolution than spatially distributed variables such as SWE (Elsner et al., 2014; Shuai et al., 2022; Rasouli et al., 2022). In this study, the differences in snowmelt rate and timing were dampened as they propagated through the ConvLSTM models primarily due to two reasons. First, the ConvLSTM models used linearly scaled inputs. By doing so, they rely on the relative information as opposed to absolute values to infer streamflow response to snowmelt and rainfall. In addition, the learnable weights of ConvLSTM can “absorb” input biases when the model is trained and compensate for the bias during inference. In other words, when trained using snowmelt simulated by different meteorological forcing, the ConvLSTM models learned different streamflow response to snowmelt and rainfall. The ConvLSTM models trained using WRF-driven UEB simulation results learned fast streamflow response with 1~4 days of time lag, while the ConvLSTM models trained using NLDAS-driven UEB learned a time lag of approximately one month (Table 3). This is consistent with the one-month delay in snowmelt (Fig. 6).

The results suggest that the ConvLSTM models are at least partially immune from input bias when such bias is consistent during training and test periods. For coarse resolution meteorological dataset such as NLDAS, downscaling to a finer resolution using orographic adjustment overall improved SWE and streamflow simulation accuracy. Although orographically adjusting WRF led to deleterious overestimation of SWE, it slightly improved streamflow simulation accuracy (except for PBIAS) likely due to better representation of deep snowpack and late melt at high elevations. It is also noteworthy that the ensemble mean streamflow as the average of four simulations yielded the best performance during the test period (Table 2). This observation supports the inclusion of multiple meteorological datasets for streamflow simulation in mountainous watersheds (Wong et al., 2021). This is particularly important when utilizing deep learning models for streamflow prediction under projected climate scenarios.

6. Conclusions

We investigated the effects of different meteorological datasets and downscaling methods on snow and streamflow modeling in a snow-dominated mountainous karst watershed. Notable differences were found in the meteorological variables, precipitation in particular, from two meteorological datasets (NLDAS and WRF). Orographic adjustment effects are much smaller than the differences between the two datasets. The raw and orographically adjusted NLDAS and WRF datasets were used to drive the UEB model at 100 m resolution. The differences in precipitation led to amplified differences in SWE and snowmelt, highlighting the effects of temperature. Comparison with observed SWE at SNOTEL stations shows overestimation of WRF results and underestimation of NLDAS results. Orographic adjustment decreased SWE accuracy for WRF, while increasing SWE accuracy for NLDAS. The results suggest that orographic adjustment is necessary when using coarse resolution meteorological forcing for a

mesoscale watershed to capture the spatial distribution of snow processes, but may be unnecessary when dynamically downscaled forcing is available at 4 km or a finer resolution.

Snowmelt simulated by the UEB model were fed into a deep learning model to simulate streamflow response to spatiotemporally varied snowmelt. Despite large differences in SWE and snowmelt, much smaller differences were observed in simulated streamflow, suggesting the deep learning model is at least partially immune to systematic input bias due to input scaling and model flexibility to compensate for biases. Among all experiments, streamflow simulated using orographically adjusted WRF dataset resulted in overall higher accuracy, although the improvement in performance metrics contributed by orographic adjustment is smaller for the WRF dataset than for NLDAS. The improvement is likely due to a better representation of snow accumulation and melt at high elevation areas that overlap with concentrated recharge locations. Our results underline the value of dynamically downscaled meteorological datasets as well as orographically adjusting coarse resolution datasets when the former is unavailable for modeling streamflow in mountainous watersheds. In addition, ensemble analyses using multiple meteorological datasets can potentially improve streamflow simulation accuracy.

Acknowledgments

This work was supported by the US National Science Foundation Hydrologic Sciences program grants 2044051/2043363, the USGS 104b, the Utah Water Research Laboratory, and the Logan River Observatory. TX is also supported by NSF grant OAC-1931297. RZ and QL are supported by NOAA CPO-COM grant NA20OAR4310341.

Data Availability Statement

All data used in this research are publicly available. The code used for deep learning models is available at https://github.com/pseudoszechwaniens/UEB_ConvLSTM_model.

References

- Adam, J.C., Hamlet, A.F., and Lettenmaier, D.P. (2009). Implications of global climate change for snowmelt hydrology in the twenty-first century. *Hydrological Processes* 23(7), 962–972.
- Behnke, R., Vavrus, S., Allstadt, A., Albright, T., Thogmartin, W. E., & Radeloff, V. C. (2016). Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecological applications*, 26(5), 1338-1351.
- Bolton, D. (1980). The computation of equivalent potential temperature. *Monthly weather review*, 108(7), 1046-1053.
- Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., ... & Woods, R. A. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, 47(7).
- Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., ... & Meng, J. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, 108(D22).
- Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., and Brewer, C.K. (2012). Modeling Percent Tree Canopy Cover: A Pilot Study. *Photogrammetric Engineering & Remote Sensing* 78(7), 715–727.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. Pasteris (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: a Journal of the Royal Meteorological Society*, 28(15), 2031-2064.

Dibike, Y. B., Gachon, P., St-Hilaire, A., Ouarda, T. B., & Nguyen, V. T. V. (2008). Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada. *Theoretical and Applied Climatology*, 91(1), 149-170.

Elsner, M. M., Gangopadhyay, S., Pruitt, T., Brekke, L. D., Mizukami, N., & Clark, M. P. (2014). How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations?. *Journal of Hydrometeorology*, 15(4), 1384-1403.

Eum, H. I., Dibike, Y., Prowse, T., & Bonsal, B. (2014). Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada. *Hydrological Processes*, 28(14), 4250-4271.

Fallah, A., O, S., & Orth, R. (2020). Climate-dependent propagation of precipitation uncertainty into the water cycle. *Hydrology and Earth System Sciences*, 24(7), 3725-3735.

Ferguson, C. R., & Mocko, D. M. (2017). Diagnosing an artificial trend in NLDAS-2 afternoon precipitation. *Journal of Hydrometeorology*, 18(4), 1051-1070.

Fiddes, J., and Gruber, S. (2014). TopoSCALE v.1.0: downscaling gridded climate data in complex terrain. *Geoscientific Model Development* 7, 387–405.

Flerchinger, G.N., Cooley, K.R., and Ralston, D.R. (1992). Groundwater response to snowmelt in a mountainous watershed. *Journal of Hydrology* 133, 293–311.

- Gal, Y., & Ghahramani, Z. (2016a). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050-1059). PMLR.
- Gal, Y., & Ghahramani, Z. (2016b). A theoretically grounded application of dropout in recurrent neural networks. *Advances in neural information processing systems*, 29, 1019-1027.
- Goldscheider, N., & Drew, D. (Eds.). (2014). *Methods in Karst Hydrogeology: IAH: International Contributions to Hydrogeology*, 26. Crc Press.
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2), 135-143.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1-2), 80-91.
- Hall, D.K., and Riggs, G.A. (2016). MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid (Boulder CO, USA: NASA Snow and Ice Data Center). Accessed on Aug. 15, 2020.
- Hamon, W.R. (1960). *Estimating potential evapotranspiration*, Doctoral dissertation, Massachusetts Institute of Technology.
- Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics* 52(3), 218–242.
- He, C., Chen, F., Barlage, M., Liu, C., Newman, A., Tang, W., ... & Rasmussen, R. (2019). Can convection-permitting modeling provide decent precipitation for offline high-resolution

snowpack simulations over mountains? *Journal of Geophysical Research*:

Atmospheres, 124(23), 12631-12654.

Hong, Y., Hsu, K. L., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water resources research*, 42(8).

Hungerford, R.D., Nemani, R.R., Running, S.W., and Coughlan, J.C. (1989). MTCLIM: a mountain microclimate simulation model (Ogden, UT). U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. P. Xie, (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, 5(3), 487–503.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005-6022.

Li, Z., Xu, X., Liu, M., Li, X., Zhang, R., Wang, K., & Xu, C. (2017). State-space prediction of spring discharge in a karst catchment in southwest China. *Journal of Hydrology*, 549, 264–276. <https://doi.org/10.1016/j.jhydrol.2017.04.001>

Lin, Y., Baldwin, M. E., Mitchell, Rogers, K. E., E., and DiMego, G. (2001). Spring 2001 changes to the NCEP Eta Analysis and Forecast System: Assimilation of observed precipitation. *18th Conf. Weather Analysis Forecasting/14th Conf.on Numerical Weather Prediction*, Fort Lauderdale, FL, Amer. Meteor. Soc., J92–J95.

Lin, Y., & Mitchell, K. E. (2005, January). 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications. In *Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA (Vol. 10)*.

Liston, G.E., and Elder, K. (2006). A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet). *Journal of Hydrometeorology* 7, 217–234.

Mahat, V., and Tarboton, D.G. (2012). Canopy radiation transmission for an energy balance snowmelt model: Canopy radiation for snowmelt. *Water Resources Research*, 48(1).

Maina, F. Z., Siirila-Woodburn, E. R., & Vahmani, P. (2020). Sensitivity of meteorological-forcing resolution on hydrologic variables. *Hydrology and Earth System Sciences*, 24(7), 3451-3474.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., ... & Shi, W. (2006). North American regional reanalysis. *Bulletin of the American Meteorological Society*, 87(3), 343-360.

Mital, U., Dwivedi, D., Brown, J. B., & Steefel, C. I. (2022). Downscaled hyper-resolution (400 m) gridded datasets of daily precipitation and temperature (2008–2019) for East Taylor subbasin (western United States). *Earth System Science Data Discussions*, 1-26.

Mizukami, N., P. Clark, M., G. Slater, A., D. Brekke, L., M. Elsner, M., R. Arnold, J., and Gangopadhyay, S. (2014). Hydrologic Implications of Different Large-Scale Meteorological Model Forcing Datasets in Mountainous Regions. *Journal of Hydrometeorology*, 15, 474–488.

Mizukami, N., Clark, M.P., Gutmann, E.D., Mendoza, P.A., Newman, A.J., Nijssen, B., Livneh, B., Hay, L.E., Arnold, J.R., and Brekke, L.D. (2016). Implications of the Methodological Choices for Hydrologic Portrayals of Climate Change over the Contiguous United States:

Statistically Downscaled Forcing Data and Hydrologic Models. *Journal of Hydrometeorology*, 17, 73–98.

Nelson, K.J., Connot, J., Peterson, B., and Martin, C. (2013). The Landfire Refresh Strategy: Updating the National Dataset. *Fire Ecology*, 9(2), 80–101.

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

Rasouli, K., Pomeroy, J. W., & Whitfield, P. H. (2022). The sensitivity of snow hydrology to changes in air temperature and precipitation in three North American headwater basins. *Journal of Hydrology*, 606, 127460.

Salamon, P., & Feyen, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, 376(3-4), 428-442.

Scalzitti, J., Strong, C., and Kochanski, A.K. (2016). A 26 year high-resolution dynamical downscaling over the Wasatch Mountains: Synoptic effects on winter precipitation performance: DYNAMICAL DOWNSCALING WASATCH. *Journal of Geophysical Research: Atmospheres*, 121, 3224–3240.

Schlögl, S., Marty, C., Bavay, M., and Lehning, M. (2016). Sensitivity of Alpine3D modeled snow cover to modifications in DEM resolution, station coverage and meteorological input quantities. *Environmental Modelling & Software*, 83, 387–396.

Sen Gupta, A., and Tarboton, D. (2012). Definitions of parameters and variables in UEBGrid snow and glacier melting model. Utah Water Research Laboratory. Available at

<http://hydrology.usu.edu/snow/uebgrid/UEBGridInterfaceDesign.docx>. Accessed on 10/23/2017.

Sen Gupta, A., and Tarboton, D.G. (2016). A tool for downscaling weather data from large-grid reanalysis products to finer spatial scales for distributed hydrological applications.

Environmental Modelling & Software, 84, 50–69.

Shamir, E., and Georgakakos, K.P. (2006). Distributed snow accumulation and ablation modeling in the American River basin. *Advances in Water Resources*, 29, 558–570.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., and Woo, W. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in neural information processing systems*, 28.

Spangler, L. E. (2001). Delineation of recharge areas for karst springs in logan canyon, Bear River range, northern Utah (pp. 01–4011). United States Geological Survey Karst Interest Group Proceedings, Water-Resources Investigations Report.

Spangler, L. E. (2011). Karst hydrogeology of the Bear River range in the vicinity of the Logan River, northern Utah. Geological Survey. Paper presented at Geological Society of America Rocky Mountain - Cordilleran section meeting, U.S.

Strauch, M., Bernhofer, C., Koide, S., Volk, M., Lorz, C., & Makeschin, F. (2012). Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. *Journal of Hydrology*, 414, 413-424.

Shuai, P., Chen, X., Mital, U., Coon, E. T., & Dwivedi, D. (2022). The effects of spatial and temporal resolution of gridded meteorological forcing on watershed hydrological responses. *Hydrology and Earth System Sciences*, 26(8), 2245-2276.

Tarboton, D.G., and Luce, C.H. (1996). Utah Energy Balance Snow Accumulation and Melt Model (UEB). 64.

Thornton, P.E., Thornton, M.M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., and Cook, R. (2012). Daymet: Daily surface weather on a 1 km grid for North America, 1980-2008. *Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)*.

USDA Natural Resources Conservation Service (NRCS) (2022). SNOwpack TELemetry Network (SNOTEL). NRCS. <https://data.nal.usda.gov/dataset/snowpack-telemetry-network-snotel>. Accessed 2022-07-09.

U.S. Geological Survey (2017). 1/3rd arc-second Digital Elevation Models (DEMs). USGS National Map 3DEP Downloadable Data Collection: U.S. Geological Survey.

White, W.B. (2002). Karst hydrology: recent developments and open questions. *Engineering Geology*, 65(2–3), 85–105.

Winstral, A., Marks, D., and Gurney, R. (2014). Assessing the Sensitivities of a Distributed Snow Model to Forcing Data Resolution. *Journal of Hydrometeorology*, 15(4), 1366–1383.

Wrzesien, M. L., Pavelsky, T. M., Durand, M. T., Dozier, J., & Lundquist, J. D. (2019). Characterizing biases in mountain snow accumulation from global datasets. *Water Resources Research*, 55(11), 9873-9891. <https://doi.org/10.1029/2019WR025350>

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products: WATER AND ENERGY FLUX ANALYSIS. *Journal of Geophysical Research: Atmospheres* 117(D3).

- Xu, T., Longyang, Q., Tyson, C., Zeng, R., & Neilson, B. T. (2022). Hybrid Physically Based and Deep Learning Modeling of a Snow Dominated, Mountainous, Karst Watershed. *Water Resources Research*, 58(3), e2021WR030993.
- Yang, J., Jiang, L., Ménard, C. B., Luo, J., Lemmetyinen, J., & Pulliainen, J. (2015). Evaluation of snow products over the Tibetan Plateau. *Hydrological Processes*, 29(15), 3247-3260.
- Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., et al. (2018). A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 108–123.
- Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, 44, W09417.
- Wong, J. S., Yassin, F., Famiglietti, J. S., & Pomeroy, J. W. (2021). A streamflow-oriented ranking-based methodological framework to combine multiple precipitation datasets across large river basins. *Journal of Hydrology*, 603, 127174.