5-2021

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

Conor Tyson
Utah State University

EFFECTS OF CLIMATE FORCING UNCERTAINTY ON HIGH-RESOLUTION SNOW MODELING AND STREAMFLOW PREDICTION IN A MOUNTAINOUS KARST WATERSHED

by

Conor Tyson

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

Approved:

_____	_____
Bethany Neilson, Ph.D. Co-Major Professor	Tianfang Xu, Ph.D. Co-Major Professor
_____	_____
David Tarboton, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Interim Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2021

ABSTRACT

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

by

Conor Tyson, Master of Science

Utah State University, 2021

Major Professors: Drs. Bethany Neilson and Tianfang Xu
Department: Civil and Environmental Engineering

In the mountainous Western U.S., a considerable portion of water supply originates as snowmelt passing through karst watersheds. Karst watersheds are basins containing fissures, fractures, and conduits in carbonate rocks, creating heterogeneous groundwater flow and storage dynamics. Accurately simulating streamflow in snow-dominated, karst basins is challenging. Meteorological forcings are affected by complex terrains and exhibit high spatial variability, while observation data are typically scarce. The spatially varying forcings and topography lead to high spatial variability in snow accumulation and melt rates. In addition, commonly used rainfall-runoff models may not perform well when faced with the hydrogeologic heterogeneity of karst watersheds. To overcome these challenges, we simulated snow processes at a fine resolution using a physically based snow model and used the model outputs to run a deep learning model to simulate streamflow. The snow model was forced with meteorological variables from a Weather Research and Forecasting (WRF) model and the North American Land Data Assimilation

System (NLDAS). The two datasets were used at base resolution and orographically adjusted according to topography. This study was performed to understand how the choice of climate data and downscaling techniques affect the simulation of snow and streamflow in a snow dominated, mountainous, karst watershed. We found the climate datasets and downscaling methods resulted in large differences in simulated snow water equivalent (SWE) and snowmelt rate. The differences in SWE between the datasets are noticeably larger than the differences in precipitation, showing the effects of other variables, particularly radiation and temperature. In addition, the WRF dataset resulted in snow melt occurring about a month later than the NLDAS dataset. Despite the differences, the resulting simulated streamflow from the deep learning model showed a close match to the measured streamflow during training and testing periods. This suggests that when deep learning models are trained on a particular dataset, it learns the rainfall-runoff response and performs well on testing datasets with similar patterns. Among all experiments, streamflow simulated using orographically adjusted WRF meteorological forcings resulted in the highest accuracy due in part to the relatively high base resolution, temporal consistency, and later melt of the WRF dataset.

(73 Pages)

PUBLIC ABSTRACT

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

Conor D. Tyson

Snow-dominated, karst watersheds present particular challenges to accurately modeling streamflow in response to differing climate conditions. This is due to the uneven distribution of snow within a basin, the varied melting rates due to terrain and climate, and the difficulty determining flow paths through karst conduits below ground. One possible solution to these challenges is to model snow at a fine scale, but climate variables are not available at these smaller spatial scales. The choices about which climate dataset to use, and how to downscale the data to a fine scale, will likely affect the accuracy of streamflow simulations. This comprises the primary goal of the thesis. For this project we simulated fine resolution snow processes using two climate datasets downscaled in two different ways and used the simulated snowmelt to feed a deep learning model that can learn patterns between the simulated snowmelt and observed streamflow to then simulate streamflow. The climate datasets differ significantly and resulted in highly different patterns of snow accumulation and melt. However, the deep learning model was able to learn the patterns with the different datasets and accurately generate streamflow for all climate datasets and downscaling methods.

ACKNOWLEDGMENTS

I would first like to thank Dr. Tianfang Xu for encouraging me to get a Master's Degree and presenting me with a project both challenging and interesting. She has worked closely with me on this work and writing with countless helpful pieces of guidance and direction. Even after she departed USU and became a professor at Arizona State University she continued to work with me and meet remotely so that I could complete this thesis.

I would also like to thank my committee members Dr. Bethany Neilson and Dr. David Tarboton. Dr. Neilson took over some responsibilities for me when Dr. Xu left and have enabled my progress towards a degree to continue without any disruption. Dr. Tarboton ensured I was still fully supported by the Utah Water Research Laboratory. Both have provided valuable insight and direction for this project.

I also need to thank Qianqiu Longyang who provided essential work for this project by setting up and running the deep learning model, without which, this project would be missing most of its results. She also helped explain the workings of the deep learning model so I could better understand and write about it in my thesis.

Finally I would like to thank my wife, Allison, who encouraged me to return to school to seek a second Bachelor's, who returned to work so I could return to school, who has fully supported me when my return to school ended up taking six years, and who has loved and supported me the entire time.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In many areas of the world, particularly northern latitudes, continental interiors, and near mountainous regions, much of the water for residential and agriculture use originates as snowpack (Adam et al., 2009). Many 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 critical for simulating streamflow and water supply availability.

Karst watersheds can contain three porosity types within the carbonate rock: micropores, small fissures, and large conduits (Hartmann et al., 2014). In snow-dominated Karst regions, portions of meltwater enter sinkholes directly connected to karst conduits, diffuse through small fissures into the karst conduits, and enter soil matrices and slowly diffuses into karst conduits or streams (Hartmann et al., 2014; Spangler, 2011; White, 2002). Therefore, karst watersheds contain high spatial heterogeneity in groundwater recharge and flow. Depending on the location within the watershed 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). 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).

Hydrologic models are commonly utilized to simulate streamflow at various spatial and temporal scales. However, these models are run at scales unable to simulate heterogeneity found in karst basins. Many of these models use a lumped conceptualization of the snow processes, where the spatial variability of the snowpack is not considered or simplified (e.g., snowmelt represented by a limited number of elevation bands). For instance, Najafi et al. (2012) combined SNOW-17, a lumped model with elevation bands that simulates snow accumulation and melt, with the Sacramento Soil Moisture Accounting model, a lumped catchment water balance model, for ensemble streamflow prediction (ESP) in a 754 km^2 watershed and found improved estimates of probable streamflow ranges. Hegdahl et al. (2019) used HBV, a lumped hydrologic model for snow and glacial melt, with ten elevation zones throughout Norway and found temperature accuracy had a large effect on streamflow simulation accuracy, particularly in the snowmelt season. It is clear the fine scale spatial variability of SWE cannot be represented in lumped and coarse grid snow models, yet is likely to have a significant effect on streamflow at mesoscales (areas from one hundred to one thousand square kilometers), particularly in karst basins.

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; Winstral et al., 2014) as has been demonstrated in field studies. For instance, Flerchinger

et al. (1992) studied a 0.26 km^2 watershed in Idaho and collected groundwater depth, precipitation, snow accumulation, snowmelt, and streamflow at a 30 m resolution for two years to characterize the response of groundwater and streamflow to SWE levels. They found the groundwater response time from snowmelt of isolated snowdrifts to streams was 2-3 days for a year with average precipitation, while was 32-46 days for a year with less than average precipitation levels. Clark et al. (2011) studied two watersheds of 30 km^2 and 12.5 km^2 in Colorado and collected over 2000 snow depth and density measurements to determine which factors control SWE variability. They found SWE variability at a watershed scale (100-10,000 m) is driven by elevation, temperature, and radiation but fine scale (<100 m) SWE variability is dominated by wind drift, slope, and canopy interception. Tarboton et al. (2000) studied a 9.5 km^2 watershed in the Austrian Alps. They combined aerial surveys of snow cover with field measurements of SWE to set initial snow cover conditions for a snow model with 25 m resolution, then compared simulated snow cover at later times to aerial surveys and field measurements observations. The difference between model simulated and observed snow cover was within 10% and arose primarily from combinations of incorrect albedo and the effects of complex topography. These *in situ* studies produced high resolution results, but the methods are impractical for mesoscale studies.

At the mesoscale, the spatial variability can be resolved by high resolution distributed snow modeling. Winstral et al. (2014) looked at a small watershed in Idaho covering 6.2 km^2 to determine the effects of different resolutions of climate data on modeled SWE levels. They downscaled climate variables to a 10 m resolution, averaged the 10 m resolution variables into coarser resolution grids ranging from 30 m to 1500 m,

and used these data in a distributed, two-layer snow model (ISNOBAL). The simulated SWE and snowmelt from the coarser grids were compared to results from the 10 m resolution as a benchmark. They found the 100 m resolution simulated SWE differed from the 10 m resolution simulated SWE for both snow accumulation and ablation periods by less than 4%. Resolutions of 250 m, 500 m, and larger differed in simulating SWE accumulation and melting by 12%, 18%, and >20%, respectively. Schlögl et al. (2016) studied the effect various snow model modifications (e.g., resolution, roughness, albedo) have on simulated snow cover for two watersheds in the alps with areas of 145 km^2 and 356 km^2 utilizing Alpine3D, a physics-based, spatially distributed, three dimensional snow model. For studying the effect of resolution, they ran the model using resolutions from 1000 m to 25 m and found coarser grids flattened slopes, causing overestimation of SWE for southern slopes and underestimation of SWE for northern slopes. They concluded there was an overall 10% positive bias in simulated SWE for the 1000 m resolution compared to the 25 m, while only an overall 2% negative bias in simulated SWE between the 200 m and 25 m resolutions.

High resolution distributed snow models require topography (e.g., elevation), land use (e.g., canopy coverage), and climate forcing data at scales appropriate for the model grid size. While topography and land use data are available at a fine resolution (e.g., < 100 m), climate forcing data (temperature, precipitation, radiation, humidity, and wind speed) are not generally available at a fine resolution (Mizukami et al., 2016; Shamir and Georgakakos, 2006) and need to be downscaled to the resolution of hydrologic models. Mendoza et al. (2016) studied the effect of different resolutions of climate data on hydrologic responses using data dynamically downscaled to resolutions of 36, 12, and 4

km by a Weather Research and Forecasting (WRF) regional climate model on three watersheds in the Colorado River basin. They found the 36 km and 12 km resolution simulation results underestimated basin-average precipitation compared to the 4 km resolution, resulting in lower runoff on all three tested hydrologic models, demonstrating climate data resolution has a larger effect on modeled runoff than the choice of hydrologic model. Mizukami et al. (2014) used a variety of climate forcing data and downscaling techniques over the continental United States to investigate the effects of different downscaling techniques on hydrologic model responses. They found up to 20% difference in simulated runoff at high elevations is due primarily to differences in temperature and radiation among the forcing datasets and the interpolation techniques. Although downscaled climate data improves simulation results, downscaling to a fine resolution (<1 km) brings additional uncertainty and challenges to the modeling process.

To overcome these challenges, several researchers have developed methods for downscaling climate variables to fine resolutions (< 1 km) for use in distributed hydrologic models. These methods generally use the available fine resolution topography data and adjust climate variables according to known topographical effects on climate. Hungerford et al. (1989) developed MTCLIM which extrapolates climate data from a ground station to nearby areas using elevation, slope, and aspect. Liston and Elder (2006) developed MicroMet, which interpolates climate data from ground stations to 30-1000 m resolution utilizing elevation, slope, aspect, and cloudiness. Thornton et al. (2012) created Daymet, which expands MTCLIM algorithms to produce national 1 km gridded output. Fiddes and Gruber (2014) developed TopoSCALE, which downscales climate model reanalysis data utilizing pressure derived from elevation in addition to topography. Sen

Gupta and Tarboton (2016) developed MSDH, a software tool furthering the methods of Liston and Elder (2006) to use topography-based interpolation to downscale climate variables down to a 30-100 m resolution. Daly (2006) reviewed strengths and weaknesses of multiple downscaling techniques such as kriging, Daymet (Thornton et al., 2012), and PRISM (Daly et al., 1994) and recommended using multiple datasets and downscaling methods because datasets and methods typically contains bias, and combining methods may cancel out these biases.

Fine resolution, downscaled climate data used to force a snow model will generate detailed snow accumulation and melt simulations. However, due to the complexities of karst geology, common hydrologic models fail to adequately model flow in karst basins. One possible solution to this problem is to feed the snow model results to a deep learning process or machine learning technique. Machine learning techniques are powerful tools for learning complex, nonlinear relations. For instance, Ren et al. (2018) combined HBV with a Bayesian neural network for modeling streamflow. They found the combined models had higher accuracy in simulating streamflow and streamflow uncertainty than either of the two model components alone. Machine learning models are not dependent on understanding hydrologic processes within a specific watershed, and thus less prone to structural error. Data driven, deep learning network architectures can potentially address the complexities of hydrologic processes within karst aquifers.

This study aims to understand how modeling fine-resolution snow accumulation and melt with different climate datasets and downscaling techniques affects our ability to simulate streamflow in a snow-dominated mountainous karst watershed. As a test, we focus on the Logan River watershed in northern Utah. The watershed is the main drinking

and agricultural water supply for the City of Logan, Utah. Using a high-resolution snow model and a deep learning rainfall-runoff model, our specific objectives include:

1. Quantify the spatial and temporal patterns of SWE and snowmelt rate and how these patterns are affected by the choice of climate inputs and downscaling methods.

2. Identify how streamflow responds to spatially and temporally varying snowmelt.

3. Determine the accuracy of streamflow predictions using simulated snowmelt rates from different climate datasets and downscaling techniques.

We will first discuss the study area where the study was performed. Next, we will discuss the methods used in this study including forcing data, snow model, and deep learning model. Then, the results of this study are discussed and how they relate to the three main objectives. We conclude with the main points covered and the engineering significance of this study.

CHAPTER II

STUDY AREA

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 km^2 and an elevation range from under 1400 m to over 3000 m (Figure 1b). The study area is heavily forested (Figure 1c) and contains both coniferous and deciduous forests. Average basin-average precipitation is about 0.9 m, with high variability due to the terrain, but more than 50% of the precipitation falls as snow as seen in the data used for this study (Section 3.1). Water 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 water flow direction.

The karst features in the watershed result in significant karst piracy, requiring the study area to include terrain outside the topographic watershed, particularly in areas with known karst geology (Spangler, 2011). The western border of the watershed includes terrain composed of formations known to contain karst springs in the watershed: Garden City Formation, Lake Town dolomite, Water canyon formation, and Beirdneau Formation (Spangler, 2011). Thus the western edge of the study area was expanded to include outcrop of these formations and potential recharge area (Figure 2), and also to include Green and Water canyons, where tracer studies have demonstrated water from the neighboring watershed enters the Logan watershed (Spangler, 2011). The northern border of the watershed was expanded to include all of the area near Franklin Basin, an area with relatively flat terrain that leads to uncertainty in topography-based watershed boundary

delineation. To the north of Franklin Basin and northeast of the study domain, groundwater flow appears to generally carry water away from the watershed. The eastern and southern boundaries of the topographical watershed have geology with less karst conduit development, so in these areas less expansion of the study area was required (Spangler, 2011).

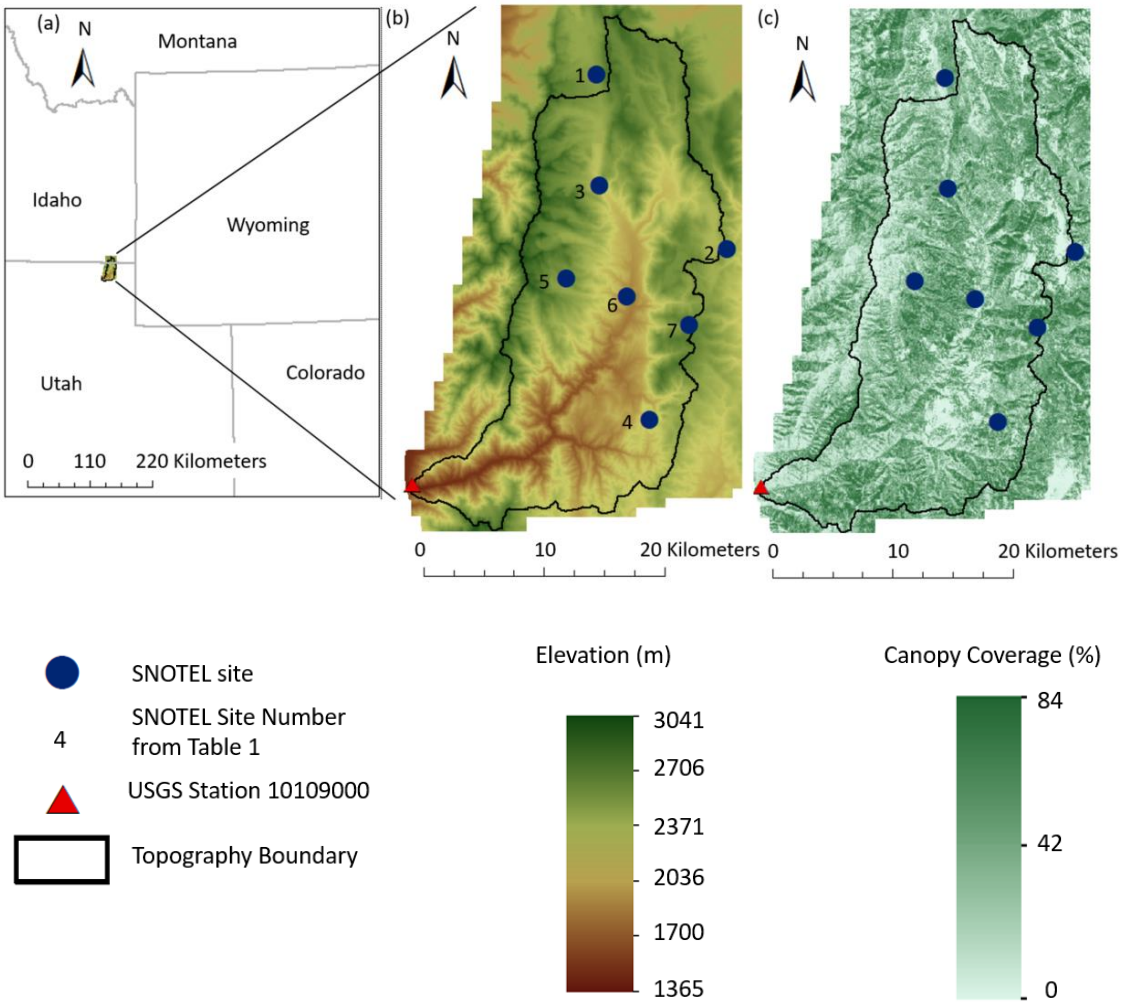

Figure 1: Maps showing (a) location (b) elevation and (c) canopy coverage (NLCD, 2019) of study area. In addition, topography boundary of Logan River watershed and locations of SNOTEL sites and USGS station are indicated.

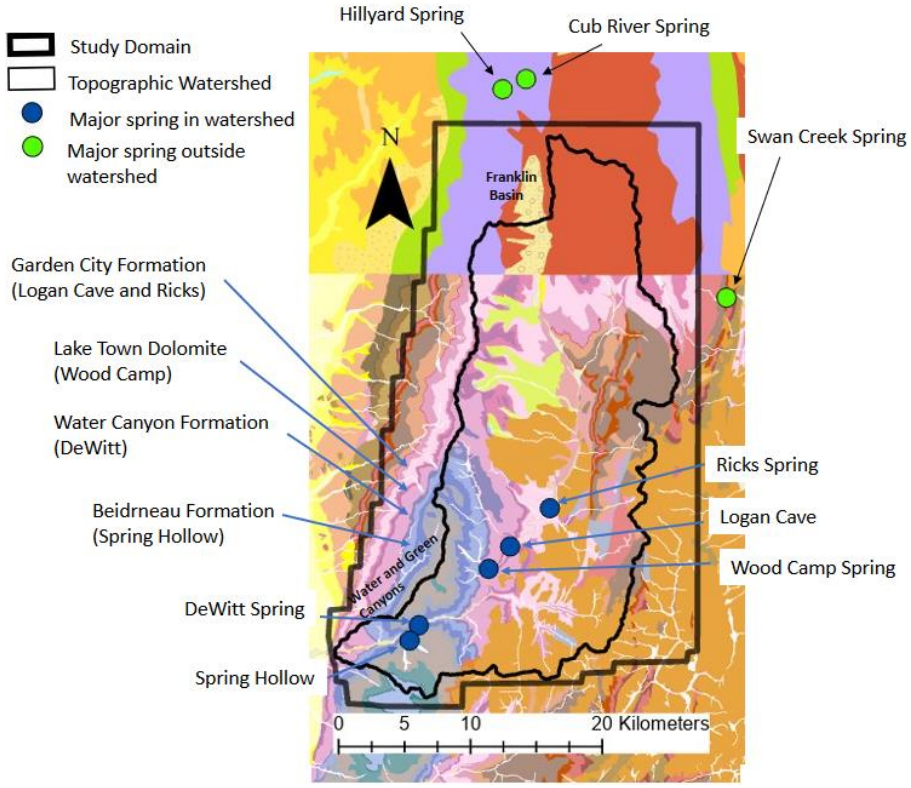

Figure 2: Study area geology, including major springs, areas of known piracy, and labeled formations with known karst features. Note: The Utah and Idaho geology maps have different color schemes and levels of detail (Dover, 2007; Lewis et al., 2012).

Hydrologic and climatic data of the study area are available from a USGS stream gage and seven SNOTEL sites. Discharge records of the Logan River are provided by USGS station 10109000 (Figure 1b, c) at a daily interval since 1954 and a 15-minute interval since 1984. For this study, daily streamflow was determined by summing the daily flows from the USGS gage, diversion rates of Highline Canal upstream of the USGS gage, and diversion rates at DeWitt springs within the study area. The combined rates from the three sites are hereafter referred to as measured streamflow.

There are seven SNOTEL (snow telemetry (USDA, 2016)) stations within the study area (Figure 1b, c, Table 1). Two of the SNOTEL sites have been collecting daily SWE and precipitation data since the late 1970s, one site began collecting data in 2001, and the remainder began collecting data between 2007 and 2009. All SNOTEL stations in the study area now collect SWE, precipitation, and minimum and maximum temperatures on a daily basis. SNOTEL data provides actual on-site data, and is a valuable source for verification of simulated data. However, ground based data still has problems of bias and inaccuracies, including the phenomenon of under-catch (Mizukami et al., 2014; Raleigh et al., 2016). In addition, SNOTEL stations are typically built in flat clearings, and are not representative of the entire landscape.

Table 1: SNOTEL stations in study area with first year of data. Elevation and canopy coverage are calculated for the UEB grid where the station is located as described in Section 3.1.

Site	Name	Year Data Begins	Elevation (m)	Canopy Coverage (%)
1	Franklin Basin	1979	2479	29
2	Garden City Summit	2009	2342	70
3	Klondike Narrows	2009	2219	37
4	Temple Fork	2001	2257	38
5	Tony Grove Lake	1978	2579	34
6	Tony Grove RS	2009	1934	23
7	USU Doc Daniels	2007	2623	43

Another available source of measured data in the study area is satellite derived fractional snow-covered area (SCA) from the MODIS datasets (Hall and Riggs, 2016). MODIS has a number of different SCA datasets available for the study area, but for this study the MOD10A1 dataset was selected. This dataset provides daily SCA with a 500 m resolution over the study area. Being completely separate from SNOTEL or other ground based information, MODIS provides a secondary source of validation for model results. However, MODIS data is known to have lower accuracy in areas of high vegetation or complex terrain, two factors very prevalent in the study area (Hall and Riggs, 2007; Huang et al., 2011).

CHAPTER III

METHODS

For this project, climate data was taken from two sources, a Weather Research and Forecasting (WRF) regional climate model and the North American Land Data Assimilation System (NLDAS-2) (Scalzitti et al., 2016; Xia et al., 2012). These datasets were orographically adjusted according to the study area topography. The raw and adjusted climate data were used to run a physically-based snow model, Utah Energy Balance (UEB) model (Mahat and Tarboton, 2012). The snow model results were then used to run a deep learning model using a new network structure (Xu et al., In Prep). Using measured data from the study area, including snow water equivalent (SWE), snow-covered area (SCA), and streamflow, simulated results were analyzed and compared to measured results.

3.1 Meteorological Forcing Data

Two sources of climate forcing data were used to run the UEB model. The first is a dynamically downscaled dataset derived using a WRF dataset centered on the Wasatch Mountains, Utah (Scalzitti et al., 2016). The WRF model was run with initial and boundary conditions from Climate Forecast System Reanalysis (CFSR). The CFSR data is downscaled based on elevation and ground cover down to a 4 km resolution. The resulting data was verified using SNOTEL data in the Utah region. The resulting dataset has results with hourly time steps from 1985 to 2010. This WRF dataset has an advantage of being downscaled to a relatively fine scale over Utah with extra steps taken to match local conditions. This provides global climate data better focused on this study area.

However, the base data of CFSR is very coarse (38 km) and has errors and biases transferred to the WRF data (Scalzitti et al., 2016). In addition, the verification used creating this data (SNOTEL), is also used as a verification in this study, and thus not a true independent verification source.

The second climate forcing utilized is the NLDAS (Xia et al., 2012) forcing dataset, a national dataset commonly used in hydrologic studies. Most NLDAS forcing data is derived from North American Regional Resolution (NAAR), a dataset with 32 km spatial resolution and 3 hour temporal resolution from 1979 to the present. The data is derived from ground station, satellite, and airborne data across the nation (Mesinger et al., 2006). NLDAS interpolates the NAAR data to a resolution of 1/8 degree (about 14 km in the study area) and hourly time steps. In addition, pressure, longwave radiation, temperature, and specific humidity are adjusted for the elevations of the NLDAS grids. NLDAS precipitation is not derived from NAAR, but from a separate national database of daily precipitation gages. Therefore, NLDAS-2 data is primarily from observational data, this has the advantage of being “on site” actual data, but has the potential disadvantages of unequal spatial distribution, measurement error, and gage bias.

3.2 Orographic adjustments of meteorological forcings

To run the UEB model, the two climate datasets described in section 2.1 were used as climate forcings both as raw data and with orographic adjustment. The following variables were obtained from the WRF dataset: 2 meter specific humidity (Q2), downward shortwave surface radiation (ACSWDNB), downward longwave surface radiation (ACLWDNB), directional 10 meter wind speeds (U10 and V10), 2 meter

temperature (T2), surface pressure (PSFC), and precipitation (RAINNC). The WRF cumulative variables of precipitation and radiation were first converted to hourly variables. Relative humidity (RH) was derived from specific humidity (Q2) and pressure (PSFC) using Equation 1 (Monteith and Unsworth, 2008; Sen Gupta and Tarboton, 2016). Equation (2) was used to convert 10 m wind (U10 and V10) to 2 m wind (U2 and V2), with z0 = 0.01 m and d = 0 m (Archer and Jacobson, 2003), while Equation 3 was used to combine the wind speeds in x (U2), y (V2) directions for total wind speed (WS) (Sen Gupta and Tarboton, 2016).

$$RH = \frac{\frac{PSFC * Q2}{.622 + Q2}}{611 * e^{\frac{17.27*T2}{T2+237.3}}} \quad (1)$$

$$U2 = U10 * \frac{\log\left(\frac{2m - d}{z0}\right)}{\log\left(\frac{10m - d}{z0}\right)} \quad (2)$$

$$WS = \sqrt{U2^2 + V2^2} \quad (3)$$

For NLDAS-2 data the following variables were obtained: specific humidity at 2 meters, surface downward longwave radiation, surface downward shortwave radiation, directional wind components at 10 meters, air temperature at 2 meters, surface pressure, and precipitation hourly total. Since the NLDAS dataset has the same variables with the same units as the WRF dataset, the same conversions and calculations were performed to prepare variables for the UEB model.

For raw data simulations, each UEB grid cell within the climate dataset is assigned the value of climate variables from the area. For orographic adjustment runs, the raw climate variables are first bilinearly interpolated within the study area. These interpolated climate variables are then adjusted to account for elevation and other topographical features using methods developed in Liston and Elder (2006) and Sen Gupta and Tarboton (2016). Here, temperature is adjusted using a lapse rate derived from SNOTEL stations in the study area. Precipitation is adjusted based on a monthly adjustment factor and the elevation difference. The relative humidity value derived from the dataset is utilized to recalculate specific humidity. The specific humidity is then adjusted using elevation adjusted temperature calculated previously and the elevation difference. The adjusted specific humidity is used to calculate adjusted relative humidity. Incoming longwave radiation is altered first according to the previously adjusted temperature and then by calculating air emissivity for elevation. Shortwave radiation is altered according to the monthly B coefficient for the Bristow-Campbell equation (also needed for UEB model, determined by monthly average diurnal temperature range) and atmospheric pressure that has been altered according to the elevation difference. Wind is adjusted based on wind direction and the slope, aspect, and curvature of the grid cell and its neighboring cells.

The UEB snow model was run for all years the WRF climate dataset was available (WY 1986-2010). 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.

3.3 UEB Snow Model

We simulated snow accumulation and melt at 100-m spatial resolution and 1 hour temporal resolution using the UEB model (Mahat and Tarboton, 2012; Tarboton and Luce, 1996) for the study area. The UEB model is a physically based model utilizing mass and energy balances of the snowpack and represents the snow as two layers, one layer above the forest canopy and the other on the surface. The model simulates three state variables to represent the snow (SWE, energy content, and surface age) and can be run using a distributed (gridded) setup. The model requires numerous basin-wide parameters (albedo, snow density, etc.), and a number of site conditions such as canopy coverage, canopy height, leaf area, slope, aspect, latitude, and longitude. The model is driven by climate inputs of temperature, precipitation, wind speed, radiation, and relative humidity at a temporal resolution sufficient to discern the diurnal pattern.

We utilized the UEB snow model with 100 m grids to capture the small scale variability in snow accumulation and melt. Based on preliminary results, changing resolution from 200 m to 100 m could affect SWE levels >20% at certain locations due to the smoothing out of aspect and slope. However, changing resolution from 100 m to 50 m alters SWE levels less than 5% suggesting 100 m resolution adequately captures slope and aspect. This model is run with the different climate forcings described in Section 3.1, WRF and NLDAS.

Parameter values were set following recommendations in Tarboton and Sen Gupta (2013). To obtain slope and aspect, a digital elevation model (DEM) with 1/3 arc second (about 9 m) resolution was used (U.S. Geological Survey, 2017). The average elevation

for each 100 m grid was calculated and ArcGIS used to calculate slope based on the elevation of each grid and its neighboring grids. Aspect was calculated using the elevation obtained for the neighboring grids. Atmospheric pressure was determined using the elevations and calculating average pressure at each elevation.

Canopy coverage data was obtained from the NLCD 2011 tree canopy dataset (Coulston et al., 2012). NLCD datasets prior to 2011 did not include the tree canopy coverage. The dataset has a 30 m resolution, and the average value within each 100 m UEB grid was obtained. For leaf area index (LAI) and forest structure (ycage) parameters, the NLCD land use dataset was utilized (Yang et al., 2018). Although the NLCD has multiple years with land use datasets, comparisons between the different datasets for the study area showed no significant changes in the years of the study. The mode of land cover classes for the 30 m cells within each 100 m UEB grid was obtained because land classification systems cannot be averaged. These land cover numbers for each grid were converted to LAI and ycage following recommendations in Sen Gupta and Tarboton (2012) and land cover classes not covered therein were converted to the most similar class. Canopy height data was obtained from NASA Landfire database (Nelson et al., 2013), which also has a 30 m resolution. This resolution was higher than other canopy height datasets available, and the dataset had more variability than was obtained using the recommendations of Sen Gupta and Tarboton (2012) on land use type. The height data was upscaled to UEB resolution by averaging. Finally, since UEB requires canopy coverage, height of canopy, and leaf area index to all be either zero or not zero at the same grid point, all grids with a canopy coverage of zero had height and leaf area set to

zero, and any grid with canopy coverage value but no height or LAI had the zero value(s) raised to one.

3.4 Deep Learning Karst Model

A deep learning model using a new network structure, Modified ConvLSTM (Xu et al., In Prep) is applied to simulate Logan River streamflow based on the UEB model outputs. The Modified ConvLSTM architecture is tailored from the Convolutional Long Short-Term Memory (ConvLSTM) which was proposed for precipitation nowcasting (Shi et al., 2015). By combining multiple convolutional layers with two layers of LSTM cells, the Modified ConvLSTM is capable of learning long-term memory effects of the watershed storage and extracting spatial patterns from UEB simulated snowmelt and watershed storage. In order to alleviate overfitting, a dropout at a rate of 0.5 was implemented.

UEB simulated combined snowmelt/rainfall and potential evapotranspiration (PET) calculated from temperature using the Hamon method (Hamon, 1960), are inputs to the deep learning model. Inputs are taken from the day for which streamflow is being simulated and the preceding 364 days to capture temporal dependency up to an annual scale. In order to save computational time and decrease computational complexity, the 100 m resolution grids of inputs are aggregated to 1.6 km resolution grids. At each time step, new inputs are combined with the previous time steps, analogous to the watershed storage, to determine the output (i.e., streamflow).

The aggregation of inputs to 1.6 km resolution results in data first being downscaled to run the UEB model, then upscaled to run the deep learning model. This

process appears odd, but is required for best results. High resolution is preferred for the snow model because both meteorological forcings and the snow accumulation and melt processes are controlled by topography, and topographic variables exhibit high spatial variability in mountainous areas (Schlögl et al., 2016; Winstral et al., 2014). However, this resolution results in a large amount of data, and it is necessary to upscale data to coarser resolution so that the computational cost for the data-driven model is affordable.

The model is trained on data from water years (WY) 1986-2002 and tested on data from WY 2003-2010, with the Mean Squared Error (MSE) (Equation 4 (Murphy, 1998)) measured to determine coefficients used for the training. During the training period, the deep learning model learns the relationship between snowmelt/rainfall, potential evapotranspiration, and streamflow. Then it uses the learned relationships to simulate streamflow for years during the testing period when the model does not know the measured streamflow value.

The results from the deep learning model were compared between the different datasets and downscaling methods using three different statistics: MSE , Nash-Sutcliff efficiency (NSE, Equation 5 (Knoben et al., 2019)), and mean percent bias (PBIAS, Equation 6 (Khair et al., 2017)). A positive PBIAS shows the simulations are generally lower than the measured, while a negative PBIAS shows the simulations are generally higher than measured. The statistics are calculated using all data and high flow periods. High flow periods were flows above the median streamflow.

$$MSE = \frac{\sum(Y_{mea} - Y_{mod})^2}{n} \qquad (4)$$

$$NSE = \frac{\sum(Y_{mea} - Y_{mod})^2}{\sum(Y_{mea} - \hat{Y}_{mea})^2} \quad (5)$$

$$PBIAS = \sum \frac{Y_{mea} - Y_{mod}}{Y_{mea}} * \frac{100}{n} \quad (6)$$

In equations (4-6), Y_{mea} are the measured streamflow values, Y_{mod} are the modeled streamflow values, n is the number of data points, and $\hat{Y}_{mea}$ is the average of the measured streamflow values.

3.5 Analyze Correlations between Snowmelt and Streamflow

To determine if there are particular UEB grid cells in the study area where the modeled output snowmelt and rainfall input correlates closely with measured streamflow, we performed a correlation analysis. The analysis results complement the deep learning karst model and help interpret the model results. The snowmelt and rainfall input values for each grid cell between March 1 and August 31 yearly was selected to focus only on times with high flow (defined by months with snowmelt and rainfall inputs higher in at least one dataset). A correlation analysis was then performed using MATLAB, correlating the daily snowmelt and rainfall input for each grid to summed recorded streamflow that occurred on specified lag periods after the input. The lag times tested were 0-12 days, 20-40 days with 2 day steps, and 60 days. This analysis allows both a determination of which grids have the highest correlation with streamflow, and what lag time results in highest correlations with each grid cell.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Spatial and Temporal Patterns of SWE and Snowmelt Rate and Effects of the Choice of Climate Inputs and Downscaling Methods

4.1.1 Comparing Weather Variables from Different Datasets and Downscaling Methods

The climate datasets and downscaling methods differ in the amount of precipitation both at the point and basin scales. Basin wide, WRF annual cumulative precipitation is 15.5% higher than NLDAS, averaged over 1986- 2010 WY, with high interannual variation (Figure 3a). This variation ranges from NLDAS cumulative precipitation being 35% higher than WRF precipitation in 1986 to WRF cumulative precipitation 41% higher than NLDAS in 2006. This pattern of the coarser climate dataset having lower precipitation matches the results of Mendoza et al. (2016), where precipitation data from WRF data in three watersheds in Colorado at 36 km and 12 km resolution, were lower than precipitation from the 4 km resolution WRF data due to the smoothing effect of averaging over mountainous areas. Interestingly, overall this difference between the two datasets appears to increase over time. Prior to 2002 there is high annual variability between the two datasets with the NLDAS dataset occasionally having higher precipitation, starting in 2002 the WRF dataset has noticeably higher precipitation values than the NLDAS dataset every year with less variation. This change in pattern suggests the relationship between the two datasets may not be consistent throughout the study period and that one of the datasets might have a temporal trend of

increasing bias. Orographic adjustments increase annual basin average precipitation by 3.4% for both WRF and NLDAS, less than the difference between the two datasets and with little interannual variation. The differences in precipitation and other climate variables influences SWE as described in section 4.1.2.

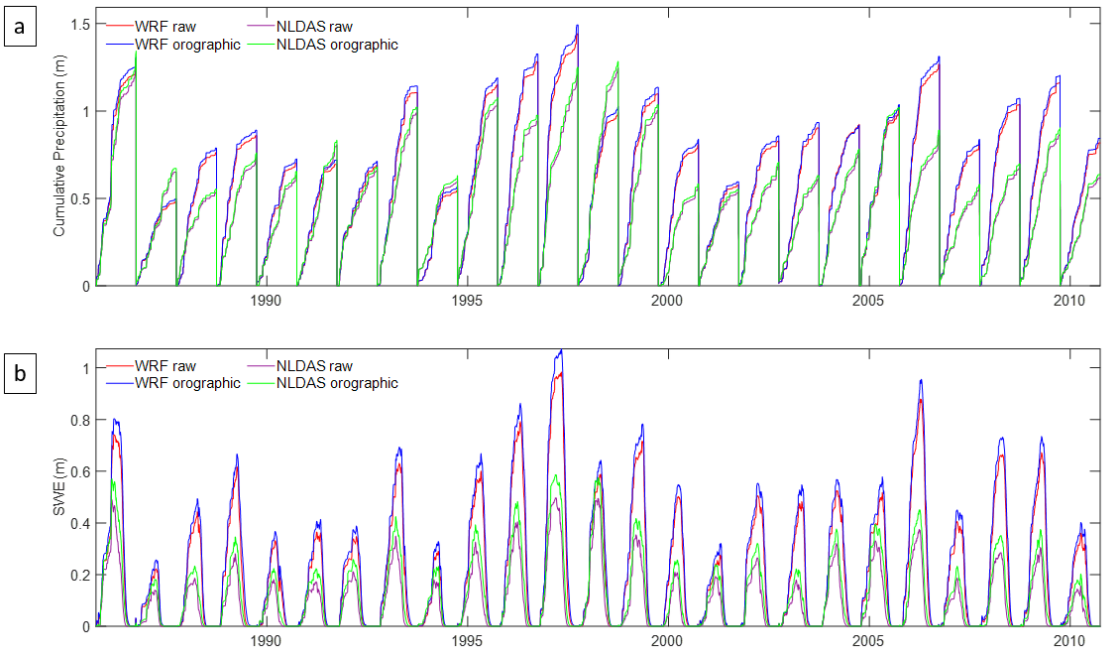

Figure 3: Comparing study area average (a) annual cumulative precipitation and (b) SWE between the different datasets and downscaling methods for the years both datasets had available.

At the point scale, the differences between climate datasets and downscaling methods can show spatial variability (Figure 4). For instance, between WY 1986 and 2010 WRF cumulative precipitation is 9.5% higher than NLDAS at the Franklin Basin SNOTEL site (FB) and 6.3% higher at the Tony Grove Lake SNOTEL site (TGL). Orographic adjustment of NLDAS data lowers precipitation at both sites by 4% and

3.9%, respectively, while for WRF data orographic adjustment lowers precipitation by 6.9% at FB and increases precipitation by 6.6% at TGL. The spatial variability of precipitation difference results in varied effects on snow accumulation and melt in different areas within the study area. This is, however, dependent on the spatial variability within the dataset and the direction of elevation difference between the datasets for each grid.

Figure 4 also compared the precipitation from the different datasets and downscaling methods to measured precipitation at the SNOTEL sites. It is not clear which dataset is an overall better match to the measured data. Figure 5 compares simulated and measured yearly precipitation for at FB and TGL, the two stations with data covering the entire study period. Table 2 further gives the NSE values for the simulated yearly precipitation values for these two sites.

Looking as the NSE values, no single dataset or downscaling method is the best match for both locations. At FB, orographically adjusted WRF has the best fit, yet at TGL it has the worst fit while raw WRF has the best fit. Raw WRF precipitation appears to have overall good fit for the two combined areas. However, given the complexities induced by comparing the readings at a SNOTEL clearing to a 100 m by 100 m grid as well as precipitation measurement error, it is difficult to determine if any dataset is closer to actual conditions throughout the study area.

The different climate datasets and downscaling methods also have differing values of temperature, incoming radiation, wind speed, and humidity, resulting in complex effects on the simulated snow. Table 3 shows the WRF dataset has lower temperature, lower incoming longwave radiation, higher incoming shortwave radiation,

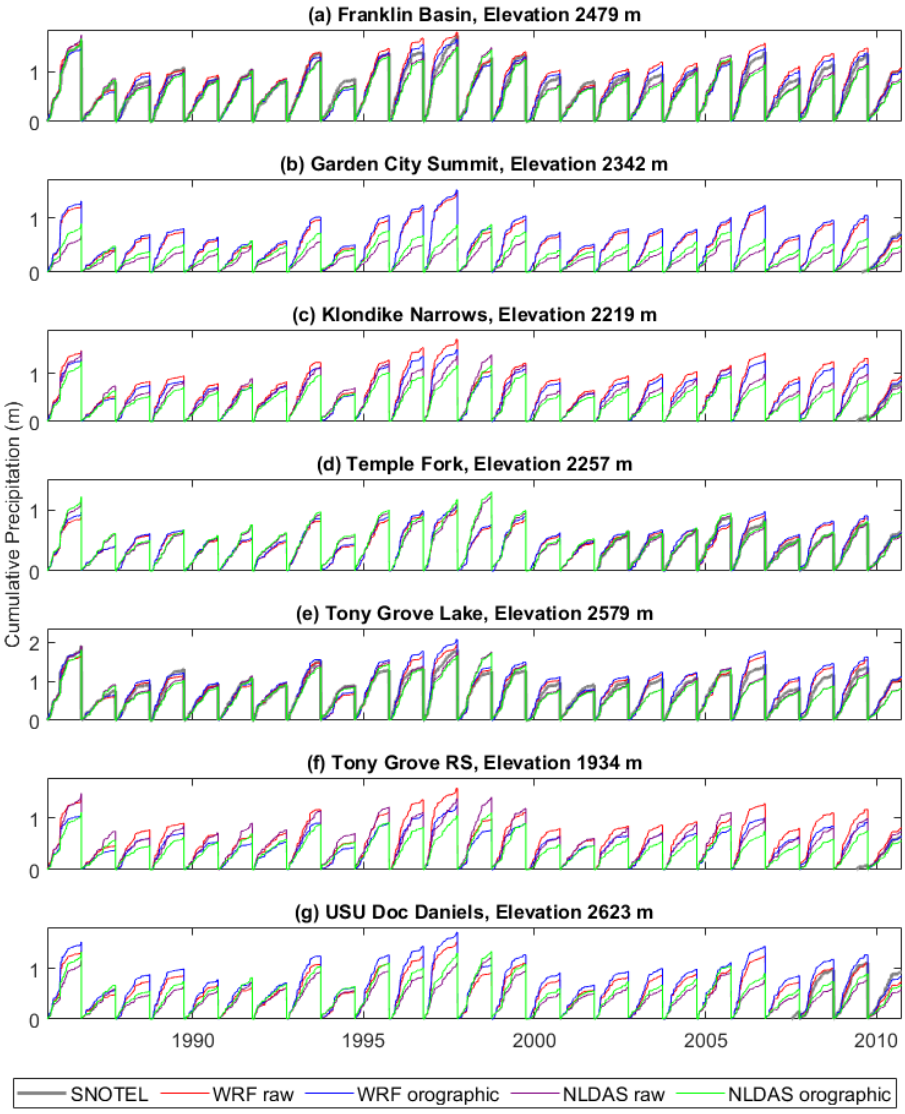

Figure 4: Comparison of annual cumulative precipitation from different datasets and downscaling methods. Measured precipitation at SNOTEL plotted when available

higher wind speeds, and higher relative humidity on average than NLDAS dataset. Although the higher incoming shortwave radiation would lower SWE, the lower incoming longwave radiation, lower temperatures, and higher precipitation all would

increase SWE levels. Wind speed, relative humidity, and their interactions with other climate variables could further affect SWE accumulation and snowmelt rate.

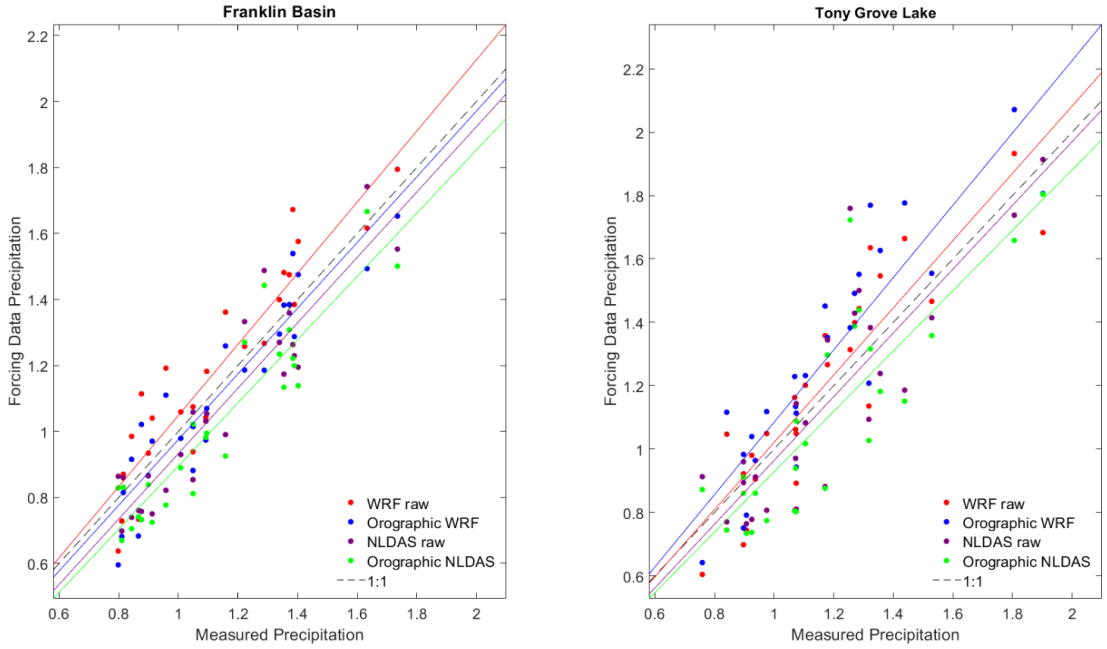

Figure 5: Relationship between measured precipitation at two SNOTEL sites and forcing data precipitation at area of SNOTEL site.

Table 2: NSE values between forcing data precipitation from different datasets and downscaling methods and measured precipitation at two SNOTEL sites in study area.

	Raw WRF	Orographic WRF	Raw NLDAS	Orographic NLDAS
Franklin Basin (FB)	0.75	0.83	0.77	0.66
Tony Grove Lake (TGL)	0.70	0.51	0.56	0.52

Table 3: Spatial and annual average of climate variables from different datasets and downscaling methods.

	WRF raw	WRF orographic	NLDAS raw	NLDAS orographic
Annual Precipitation (m)	0.93	0.96	0.80	0.83
Temperature (°C)	2.4	2.2	5.0	4.9
Incoming SW radiation (kg/m^2/hr)	856	858	727	729
Incoming LW radiation (kg/m^2/hr)	880	875	912	909
Wind Speed (m/s)	3.7	3.7	2.4	2.4
Relative Humidity (%)	66	66	55	57

4.1.2 Temporal Patterns of SWE simulated from different datasets and downscaling methods

Basin-average UEB simulated SWE arising from different climate datasets and downscaling techniques differ. The WRF dataset leads to higher SWE than the NLDAS dataset and orographic adjustment results in higher SWE than raw climate data for both datasets (Fig. 3b). The average difference in peak SWE between WRF and NLDAS forced runs is 90%, and can vary between 19% in 1998 to 170% in 2003. Interestingly, even in years where the NLDAS dataset had higher precipitation (e.g., 1998) WRF still simulated higher levels of SWE showing other climate variables beyond precipitation are affecting SWE levels. Orographic adjustment of WRF data results in SWE levels 9% higher than raw, with interannual variation ranging from 7% in 1989 to 15% in 2001,

while orographic adjustment of NLDAS data results in SWE levels rising an average of 12%, with interannual variation ranging from 9% in 1986 to 19% in 2010.

The differences in basin-average precipitation explains some, but not all, of the differences in simulated SWE resulting from the different datasets. As shown in Section 4.1.1, the basin-average precipitation levels between WRF dataset and NLDAS dataset differ by 15%, yet the resulting peak SWE simulations differ by 90%. In addition, orographic adjustment raises basin-average precipitation by 3.4%, yet raises peak SWE by 9% for WRF and 12% for NLDAS. This suggests the climate forcing variables are interactively affecting SWE.

The differences in temperature and incoming longwave radiation among the different datasets and downscaling methods might explain much of the remaining differences in basin-average SWE levels. Lower temperatures would result in more of the precipitation simulated as snow, increasing SWE, while lower incoming longwave radiation and temperature would delay snow melt and further increase SWE levels. Since WRF generally has higher precipitation, lower temperatures, and lower incoming longwave radiation than the NLDAS dataset, these climate forcings can explain the higher SWE simulation results. Likewise, orographic adjustment increases basin average precipitation and lowers basin average temperature and incoming longwave radiation, leading to higher simulated SWE.

WRF data does have higher shortwave radiation, which would reduce SWE. However, the high albedo of snow results in much of this radiation being reflected back into the atmosphere. Longwave radiation is absorbed at high rates by snow with snow's

high emissivity. Higher wind and lower RH would decrease SWE, but these factors are likely to have a smaller effect than the other weather variables (Mizukami et al. 2014).

Simulated SWE resulting from different climate forcing datasets and downscaling methods differs from measured SWE at SNOTEL sites, but the magnitude and direction of the difference varies among sites. As seen in Figure 6, WRF climate variables generally result in higher SWE at each SNOTEL site than was measured, while NLDAS climate variables generally results in lower SWE than measured. This occurs even when NLDAS had higher precipitation than was measured such as seen in Water Year 1998 at TGL (Figure 4). Meanwhile, orographic adjustment can both raise and lower SWE from raw input simulations based on whether the NLDAS and WRF datasets have a higher elevation than the actual elevation at the site, such as at Tony Grove RS, or a lower elevation than the actual elevation, such as at USU Doc Daniels. Also, orographic adjustment leads to large differences at some sites (Tony Grove RS for both WRF and NLDAS datasets), while almost indistinguishable difference (FB for the NLDAS dataset).

The climate dataset and downscaling method that yields SWE that best matches the measurements varies site to site (Table 4). Looking at the two sites with the longest recorded data at FB raw NLDAS has the best fit with a NSE value of 0.88, while at TGL WRF raw data result in the best fit to measured data with a NSE value of 0.94, no one dataset or downscaling method appears to be the best for all SNOTEL sites. The NSE values for change of SWE (Δ SWE) are low, suggesting neither dataset simulated the daily changes in snow accurately. All of these results support the suggestions of Daly (2006), that using multiple datasets from different sources are recommended to cancel out the biases in each individual dataset. Looking at Figures 4 and 5, the WRF data generally

over-predicts precipitation and SWE, while NLDAS data generally under-predicts both. Therefore, using both sets of data together can provide a more accurate picture of the actual weather and snow patterns in the basin.

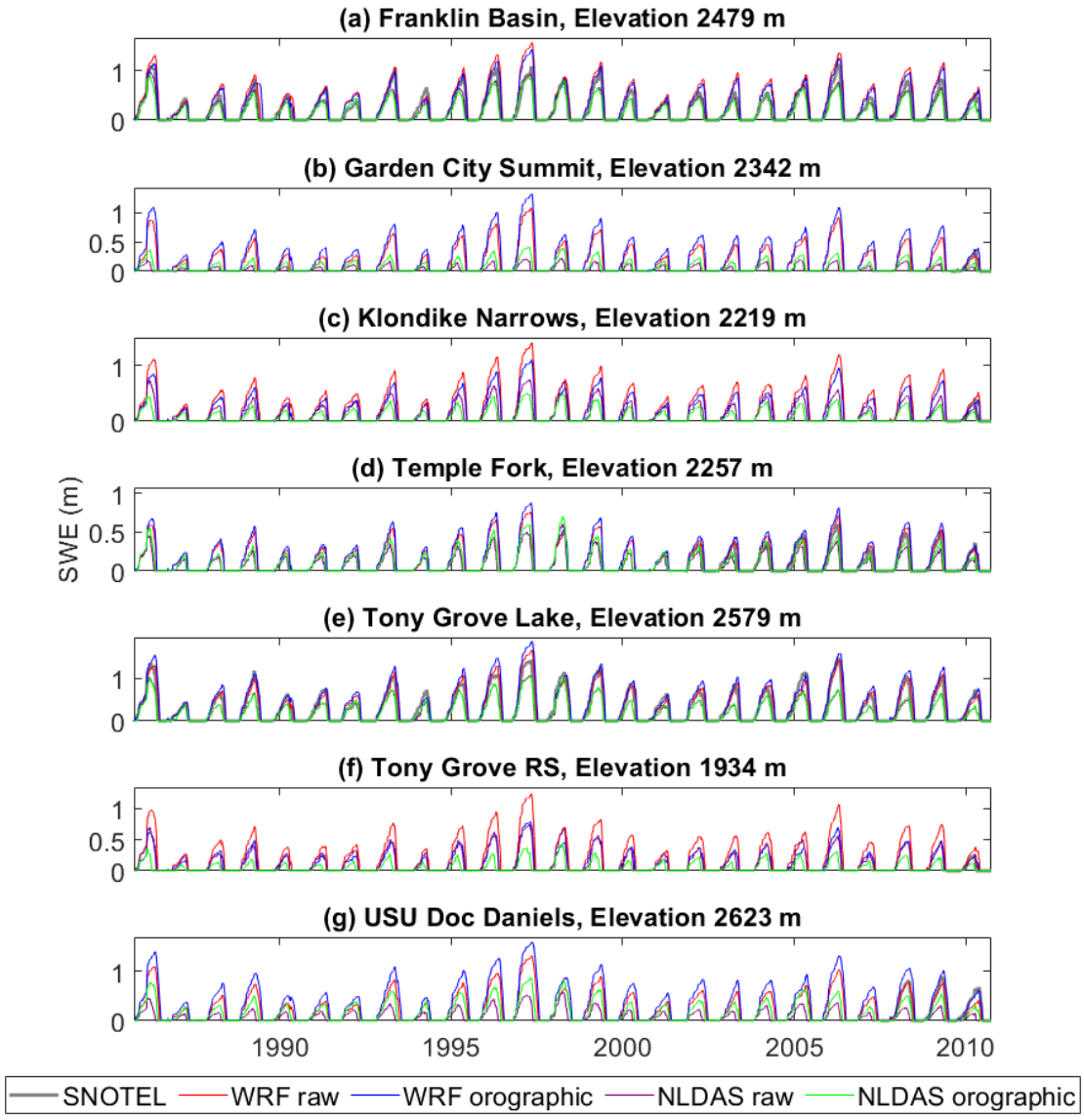

Figure 6: Comparison of SWE from different datasets and downscaling methods. Measured SWE at SNOTEL plotted when available

Table 4: NSE of UEB simulated SWE and change of SWE (Δ SWE) using different climate datasets and downscaling methods compared to SNOTEL change of SWE measurements during WY 1986 - 2010.

SNOTEL Stations	**Years of data WY 1986 - 2010**	**WRF Raw**		**WRF Orographic**		**NLDAS Raw**		**NLDAS Orographic**	
		SWE	Δ SWE	SWE	Δ SWE	SWE	Δ SWE	SWE	Δ SWE
Franklin Basin	25	0.60	0.03	0.76	0.17	0.88	0.48	0.79	0.35
Garden City Summit	1	0.91	0.28	0.56	-0.26	-0.023	0.03	0.39	0.19
Klondike Narrows	1	0.10	-1.25	0.62	-0.51	0.82	0.58	0.51	0.31
Temple Fork	9	0.81	0.03	0.55	-0.22	0.69	0.36	0.87	0.44
Tony Grove Lake	25	0.94	0.34	0.85	0.14	0.80	0.42	0.81	0.43
Tony Grove RS	1	0.88	-0.81	0.98	0.02	0.99	0.15	0.96	0.32
USU Doc Daniels	3	0.88	0.41	0.70	0.13	-0.082	0.12	0.56	0.43

The differences between modeled and measured SWE at SNOTEL sites are consistent with the differences between precipitation input data and measured precipitation at SNOTEL sites. The WRF dataset generally contained higher precipitation than recorded at SNOTEL sites and resulted in higher SWE simulations than measured at the sites. Likewise, the NLDAS dataset contained lower precipitation than SNOTEL measurements, and resulted in lower SWE simulations. In addition, sites or years which showed larger discrepancies between modeled and measured precipitation (e.g., TGL orographically adjusted WRF 1996) also showed larger differences between simulated and measured SWE. However, as with basin average, temperature and incoming

longwave radiation are also driving SWE, as seen when NLDAS has higher precipitation than measured (e.g., TGL in 1998) yet still simulated lower SWE than was measured.

4.1.3 Spatial Patterns of snow coverage from different datasets and downscaling methods

UEB simulated SWE arising from different climate datasets and downscaling techniques results in significant differences of spatial patterns of snow accumulation and melt throughout the study area. Table 5 shows the fractional snow-covered area (SCA) by month from the different input datasets and downscaling methods, averaged 1986-2010. For every month, WRF dataset results in higher SCA than NLDAS; orographic adjustment raises average SCA for both datasets. These differences are minor during peak snow cover (January and February) but become more significant during the snowmelt period (See also Figure 8b-e and 9b-e). In particular, the NLDAS dataset tend to result in the melt process occurring one month earlier on average. This is likely to have a significant effect on streamflow simulation.

Table 5: UEB simulated Fractional Snow Cover Area (%) by month with different climate datasets and downscaling methods (WY 1986-2010)

	Oct	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**
WRF raw	13	66	95	98	98	95	88	64	24	2.6	0.1	1.0
WRF orographic	15	69	95	98	98	96	89	67	29	4.9	0.4	1.1
NLDAS raw	6.1	48	86	95	94	86	60	25	4.5	0.1	0	0.2
NLDAS orographic	9.3	53	88	96	95	87	64	33	9.7	1.0	0	0.5

Simulated SWE resulting from different climate forcing datasets and downscaling methods also differs in fractional SCA compared to MODIS SCA (MOD10A1) (Hall and Riggs, 2016). Figure 7 compares SCA from the different datasets to MODIS SCA for WY 2001-2010. MODIS results in significantly lower SCA during the peak snow period (December through March). This can be partly explained by MODIS inability to detect snow under trees and other vegetation, particularly in complex terrain (Hall and Riggs, 2007; Huang et al., 2011; Shamir and Georgakakos, 2006). Figure 8 suggests this problem is occurring here as MODIS results in a SCA value for every grid in the study area during winter, but at lower fractional values than is simulated.

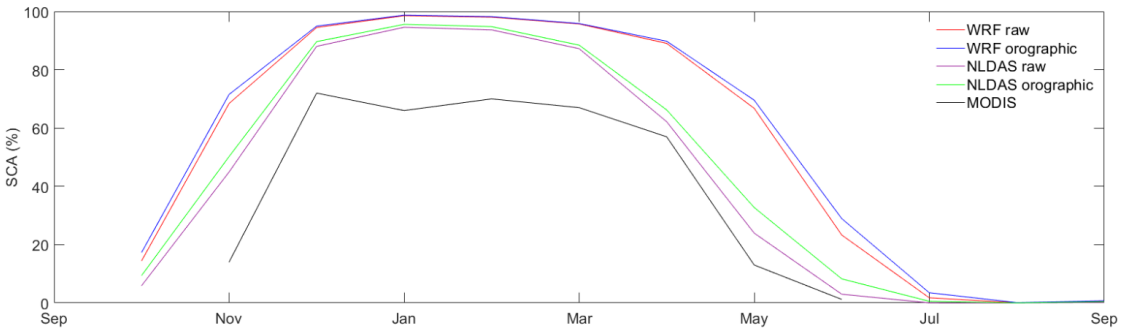

Figure 7: SCA by month from the different datasets and downscaling methods compared to the measured SCA from MOD10A1. Taken from WY 2001-2010.

The differences between simulated SWE and MOD10A1 SCA continue during the snow-melting periods. In May, the WRF raw input simulations result in 67% SCA, WRF orographically adjusted results in 70% SCA, NLDAS raw results in 24% SCA,

NLDAS orographically adjusted results in 33% SCA, and MODIS only has 13% SCA (Figure 7). Figure 9 shows the SCA difference between simulation results and MOD10A1 is from actual differing areas with no remaining snow rather than lower SCA on the same areas. Thus, WRF climate simulations still show widespread snow cover, NLDAS climate simulations results in less simulated snow coverage, while MODIS data shows even smaller areas of snow coverage.

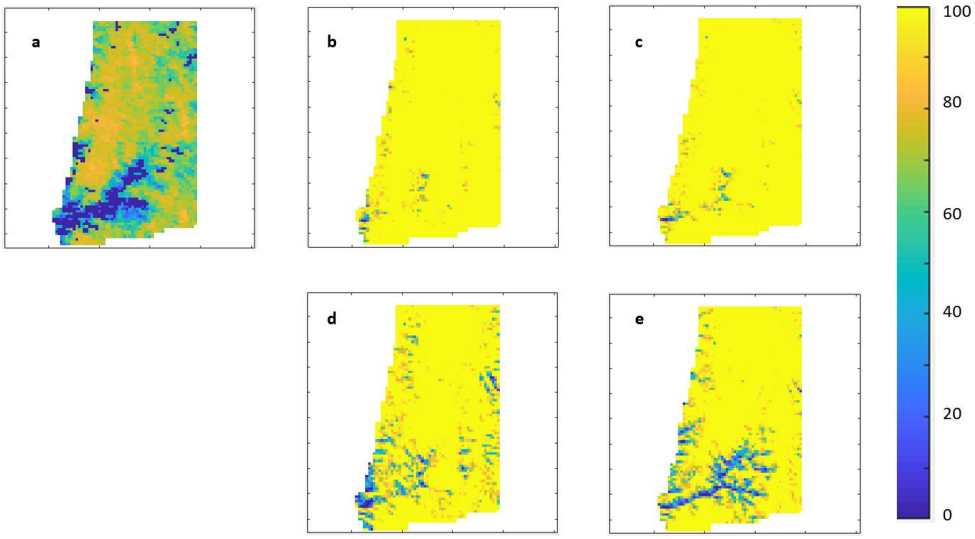

Figure 8: Spatially distributed Fractional Snow-Covered Area in the study area on January 29, 2007 from (a) MOD10A1 and UEB simulation with (b) WRF raw, (c) orographically adjusted WRF, (d) NLDAS raw, and (e) orographically adjusted NLDAS.

Because MOD10A1 SCA may suffer from degraded accuracy due to the presence of vegetation and complex terrain in the study area, we examined the monthly change of SCA in addition to the absolute values. Figure 7 shows that from April to May both MOD10A1 and NLDAS input simulations show a steep decline in SCA, while WRF input simulations also show an increased decline rate, but less steep than for the other

datasets. From May to June, only WRF input simulations show a steep decline in SCA, while both MOD10A1 and NLDAS input simulations have much less decrease in SCA. These results suggest that overall, UEB simulation driven by WRF meteorological variables seems to capture the early phase of the melt period (March to April) better than simulation driven by NLDAS, while the latter better captures the late phase of melt (April to May.)

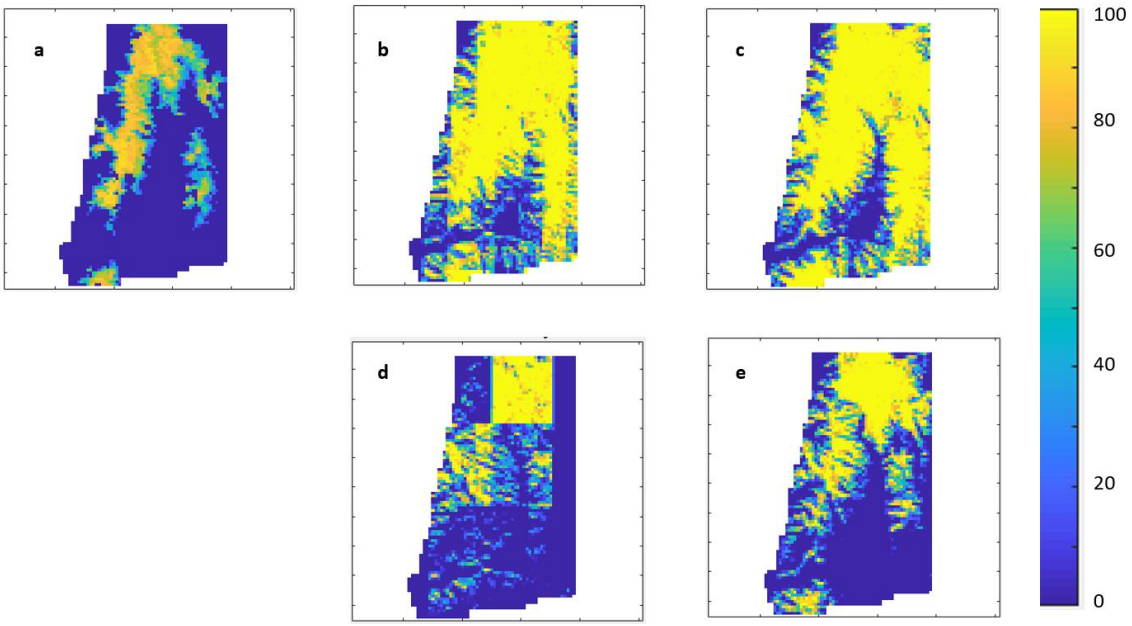

Figure 9: Spatially distributed Fractional Snow-Covered Area in the study area on May 26, 2006 from (a) MOD10A1 and UEB simulation with (b) WRF raw dataset, (c) orographically adjusted WRF dataset, (d) NLDAS raw dataset, and (e) orographically adjusted NLDAS dataset.

4.2 Streamflow Response to Spatially and Temporally Varying Snowmelt

Combined snowmelt and rainfall simulated using the different climate datasets and downscaling methods were correlated to Logan River streamflow using the methods described in 3.5. The results showed highest correlation coefficients at grids at higher

elevations, with a max correlation slightly above 0.72 for all UEB simulations (Figure 10). The low elevation area near the watershed outlet had lower coefficients, often below 0.5. The high correlation coefficients of the higher elevations generally occurred at small lag times (0 to 5 days), while the grid cells with lower coefficients generally also had longer lag times corresponding to their max coefficients (30 to 60 days) (Figure 11). These lag times of max coefficient for each grid are comparable in length to the time between snowmelt at the same grid and Logan River streamflow.

To further understand the response time of streamflow to snowmelt, Table 6 summarizes the time between the centroid of snowmelt/rainfall between April and August at six selected grids and the centroid of streamflow in the same period. The location of the six grids were marked with black dots in Figures 10 and 11; the points are at different elevations and scattered over the study area. The average time between centroids and the lag times from the correlation analysis for each dataset and downscaling

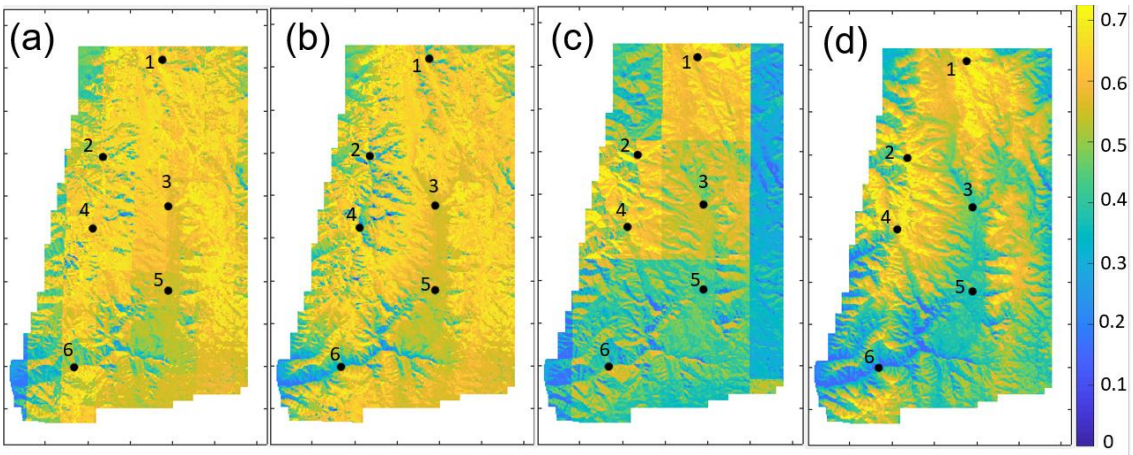

Figure 10: Maximum correlation coefficient in study area between lagged simulated snowmelt and rainfall inputs and measured streamflow time for (a) raw WRF, (b) orographically adjusted WRF, (c) raw NLDAS, and (d) orographically adjusted NLDAS. Numbered points refer to grids further analyzed in Figure 12 and Table 5.

method at each point are generally similar (Table 6). A negative time indicates the centroid of snowmelt and rainfall occurs after the centroid of streamflow, however the correlation analysis did not test for lag times less than zero.

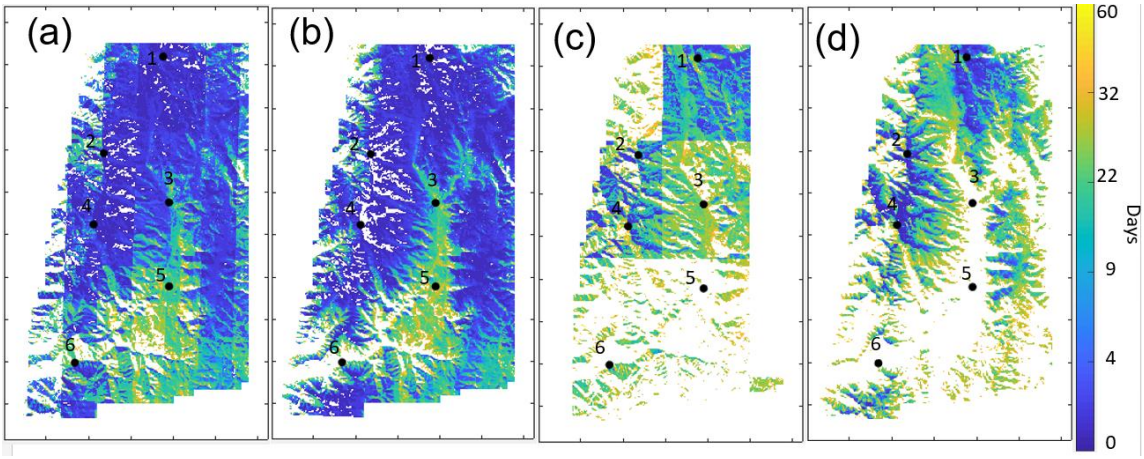

Figure 11: Time lag in days between simulated snowmelt and rainfall and measured streamflow corresponding to maximum correlation coefficients that were above 0.5 in the study area for (a) raw WRF, (b) orographically adjusted WRF, (c) raw NLDAS, and (d) orographically adjusted NLDAS. Numbered points refer to grids further analyzed in Figure 12 and Table 5.

Table 6: The average time between centroid of snowmelt/rainfall and centroid of streamflow during peak flow times (Apr. – Aug.) and the lag time of maximum correlation coefficient, from 4 UEB simulations at selected points (locations shown in Figures 10 and 11).

Point Number from Figures 10 and 11	**Elevation (m)**	**WRF raw**		**WRF orographic**		**NLDAS Raw**		**NLDAS orographic**	
		Avg. time (days)	Lag Time (days)	Avg. time (days)	Lag Time (days)	Avg. time (days)	Lag Time (days)	Avg. time (days)	Lag Time (days)
1	2820	-3.6	0	-17	0	11	3	-1.1	0
2	2875	21	12	10	0	41	60	36	34
3	2036	18	12	24	20	26	26	34	40
4	2999	4.1	1	-15	0	23	22	6.3	0
5	1993	29	24	35	30	40	60	39	60
6	1542	4.3	1	5.8	1	19	20	24	22

Correlation coefficients arising from raw data and orographically adjusted data result in slightly different results. Although max correlation coefficients and overall patterns were similar, there were some distinct differences at locations of high coefficients. The raw data resulted in high coefficients appearing to occur in regions with similar coordinates regardless of elevation in the study area. In the orographically adjusted input results, the high coefficients occur at the high elevations and the low max coefficients at low elevations.

This correlation analysis was used to an attempt to identify sink holes or other areas in the study area where meltwater quickly recharges the karst aquifer and discharges to the river. To do this, grid cells with high correlation (<0.7) and lag times between 3 to 10 days were identified (Figure 12). This lag time range was selected as being typical flow times for water in karst conduits (Spangler, 2011). While many of the identified locations are aligned with outcrop areas of karstic geologic formations, they also coincide with higher elevation and north-facing slopes where snowpack is deep and snowmelt rates during high flow periods are large. In addition, snowpack may be stored in sinkholes and melt later than simulated by UEB, contributing to summer baseflow. Such delayed effects confounds the correlation analyses. Therefore, and no clear conclusions were made.

Figure 13a plots the yearly simulated onset of snowmelt (determined by the first 5 consecutive days in the water year with basin average snowmelt and rainfall over 2 mm) with the same water year's simulated centroid of snowmelt and rainfall. This figure confirms WRF dataset generally results in later melt (both onset and centroid) than the NLDAS dataset, as was suggested by Table 5. Interestingly, orographic adjustment

generally does not affect the onset of melt timing, but does result in a longer melting period evidenced by a later centroid of melt time, as seen by blue dots generally being directly above red dots in Figure 14a. This lengthened melt period could result from the higher peak SWE of orographic adjustment. However, the difference between orographically adjusted and raw climate variables appears to be smaller than the difference between the two datasets.

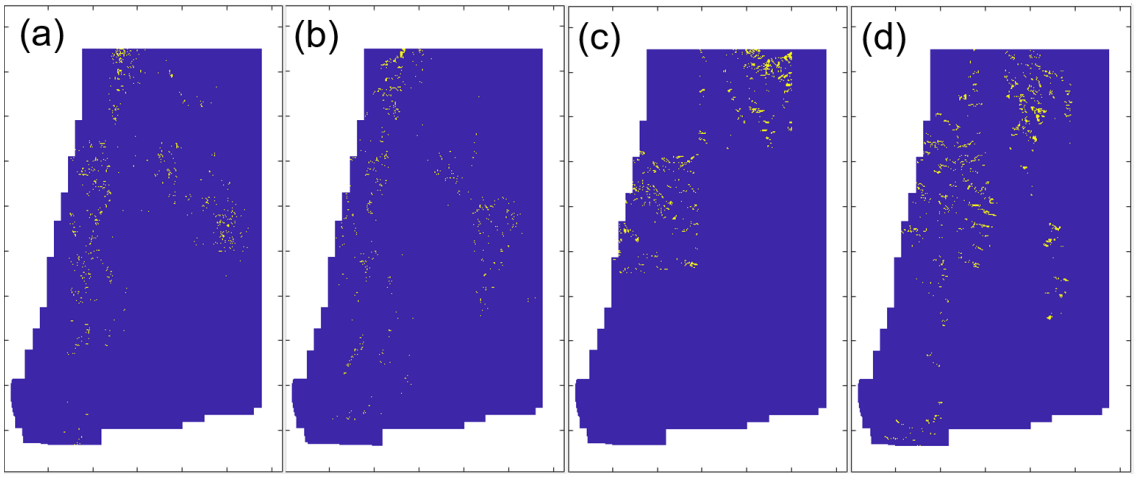

Figure 12: Locations where snowmelt and rainfall maximum correlation coefficients with Streamflow above 0.7 and lag times between 3 and 10 days- using (a) WRF raw data, (b) orographically adjusted WRF data, (c) NLDAS raw data, and (d) orographically adjusted NLDAS data.

Despite the differences in simulated snow accumulation and melt arising from the different climate datasets, the overall patterns across the different years are the same. For instance, all the datasets show 1986 and 1997 as years with higher precipitation and SWE and 1987 and 1994 as years with lower precipitation and SWE (Figure 3). For melt patterns, Figure 14a highlights two years corresponding to simulated early (2001) and late (2008) melt, respectively. Figure 13b shows measured streamflow for those two

years, confirming spring runoff occurred earlier in 2001 than in 2008. Therefore, major climate patterns between years (e.g., hot year, wet years, and early melt) are adequately captured by all of the datasets and downscaling methods, which should aid in accurate streamflow simulations.

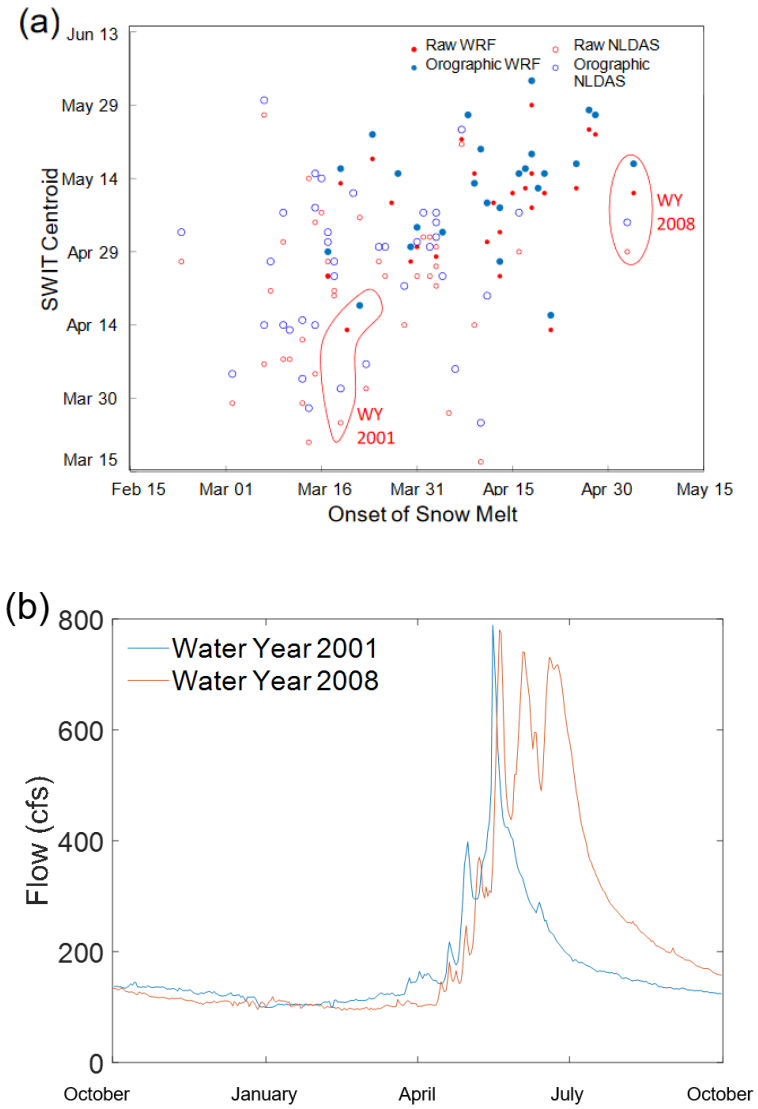

Figure 13: (a) Onset and centroid of the melting periods simulated using different datasets and downscaling methods. (b) Measured streamflow for an early melt year (2001) and a later melt year (2008).

Peak SWE is another simulated variable that can be compared to measured streamflow. Figure 14 shows the relationship between simulated peak SWE from the different datasets and downscaled methods to the annual average, peak, and average summer streamflows. These plots also show the relationship obtained by comparing peak SWE measured at SNOTEL sited (FB and TGL) to the measured streamflow. These plots show SWE simulated from WRF datasets results in regression slopes much closer to the SNOTEL data than the SWE simulated from NLDAS datasets. This result is consistent with the finding that the peak SWE from WRF simulations are closer to the measured SWE levels at these two SNOTEL sites (Table 2). The lower SWE in NLDAS simulations results in a flatter slope. However, the SWE simulated from NLDAS datasets does have a higher correlation to the streamflows than the SWE simulated from WRF. This suggests, despite NLDAS simulating lower SWE than measured, it is better at simulating the interannual variability. These mixed results again make it difficult to determine a single dataset resulting in the most accurate SWE simulations.

Figure 15 compares the snowmelt and rainfall in one year to the streamflow of Logan River at the grids marked in Figures 10 and 11. All of the higher elevation cells (a, b and d) show orographically adjusted data results in simulated snowmelt continuing later than was simulated from the raw datasets. On the contrary, for lower elevations (c, e, and f) the differences between orographically adjusted and raw datasets appears smaller. The smaller summertime peaks are rain events and not snow melt. Therefore, the lengthened melt in the study area with orographic adjustment seen in Figure 13a is primarily an effect in higher elevations. This is reasonable because high elevations are less likely to be captured by the coarse resolution of WRF and NLDAS due to smoothing. Therefore,

orographic adjustment for these areas is more likely to adjust for higher elevations, which increases precipitation and decreases temperature and longwave radiation.

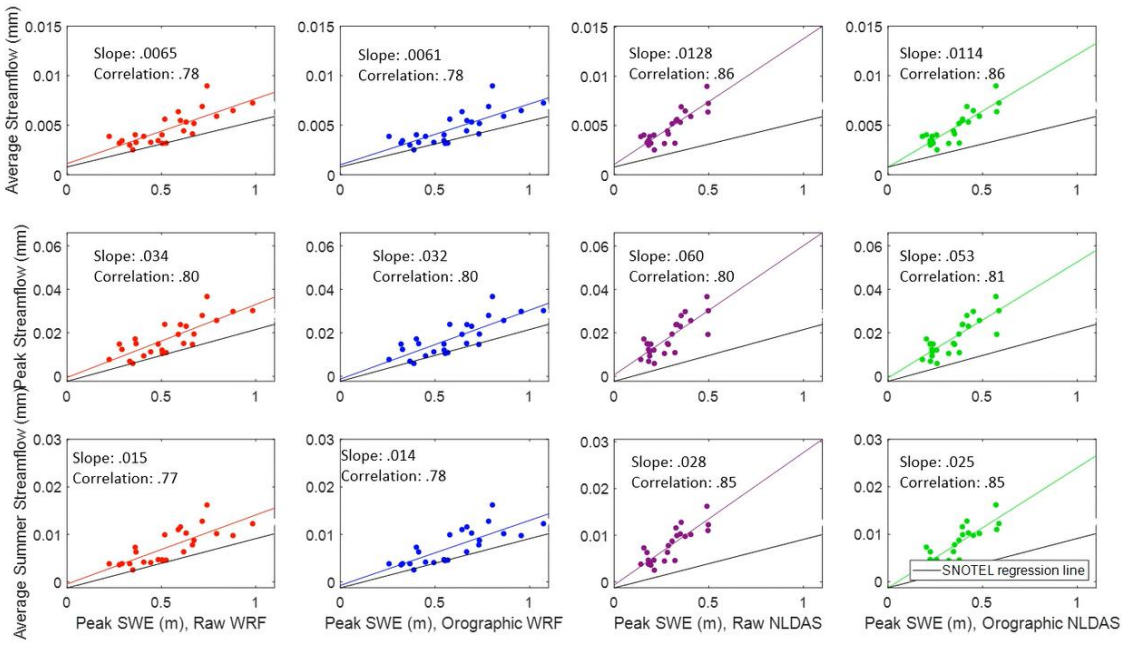

Figure 14: Relationship between peak SWE from different datasets and downscaling methods to annual average, peak, and average summer Logan River streamflow. Linear regression lines, slope, and correlation coefficients are also shown. SNOTEL regression line represents linear regression between peak SWE at combined FB and TGL SNOTEL sites to annual average, peak and average summer Logan River streamflow.

4.3 Effects of Choice of Climate Datasets and Downscaling Techniques on Simulated Streamflow

Simulated streamflow during the training and testing periods from the deep learning karst model are shown in Figure 16. Overall, the simulations appear to match closely. No single dataset or downscaling method shows a clear bias or high variance from the measured streamflow values. The differences between the measured and simulated

streamflow values from all of the datasets are less than what was seen between simulated and measured SWE at SNOTEL sites.

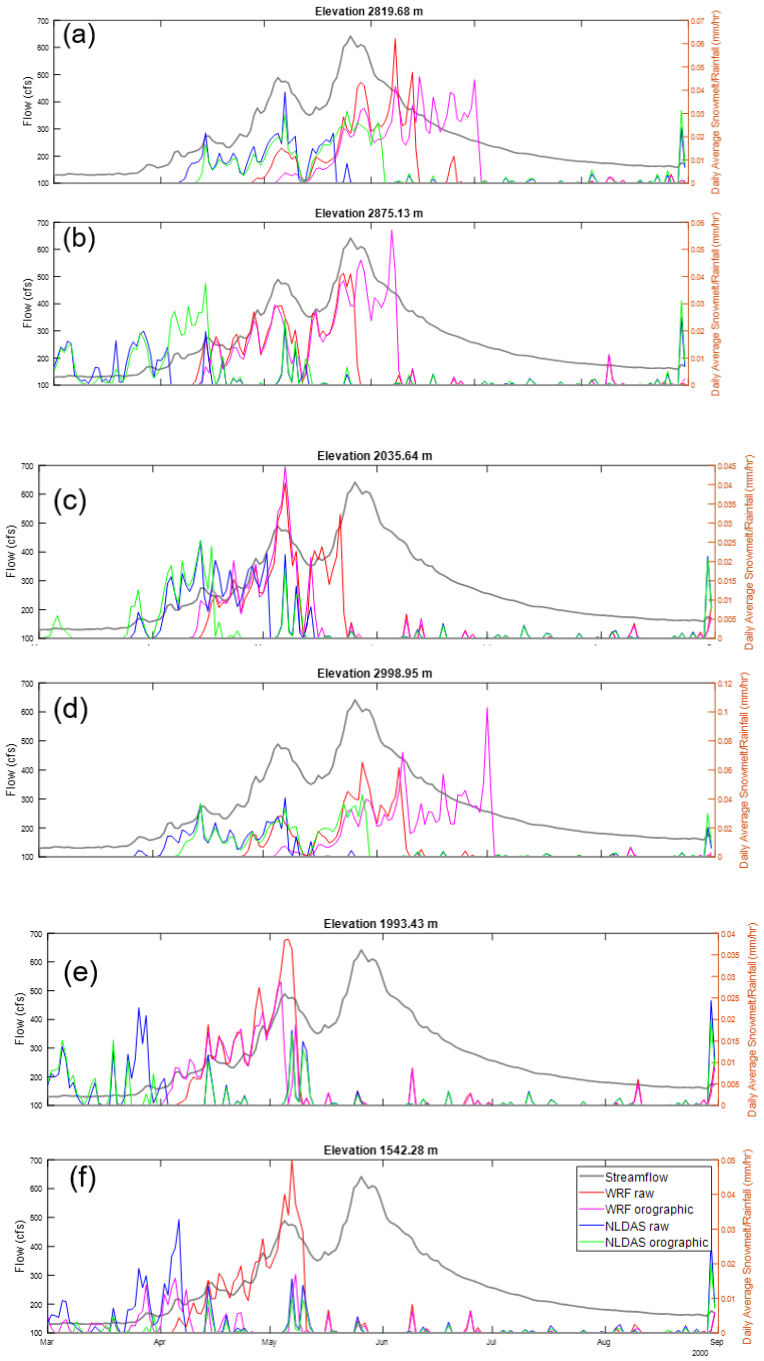

Figure 15: Combined snowmelt and rainfall at six selected grids and measured streamflow during the high flow period (Mar-Aug) in 2000.

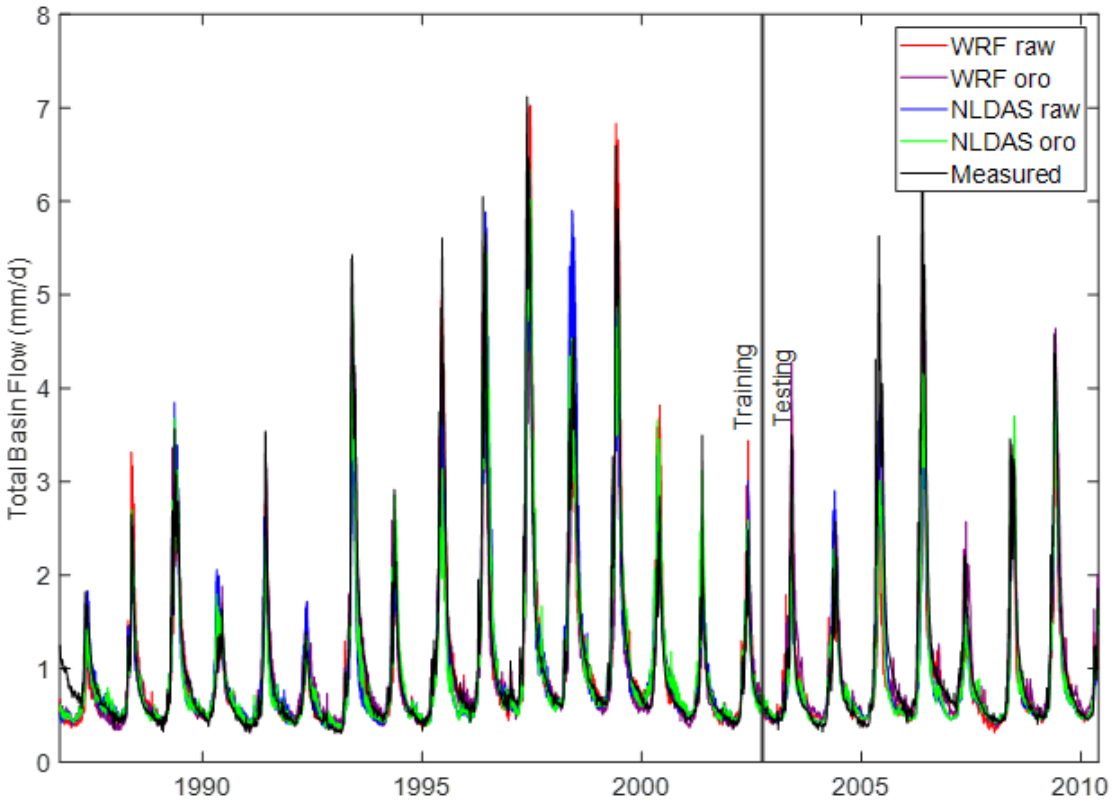

Figure 16: Simulated streamflow from deep learning model using combined snowmelt and rainfall results from UEB model runs with different datasets and downscaling methods. The training and test periods are separated with the vertical line.

Table 7 shows simulated streamflow performance by the deep learning model during both training and testing periods. First, the training period always has better values than the testing period. This is expected since the model was presented with streamflow observations and the weights of the deep network were adjusted to match the observations. Second, in general, the WRF dataset results in better performance than the NLDAS dataset. Third, orographic adjustment to climate variables results in higher NSE and lower MSE during the testing period for both WRF and NLDAS datasets. However, this improvement in values is greater for the WRF dataset than the NLDAS dataset.

Table 7: Nash-Sutcliffe efficiency (NSE), mean square error (MSE), and percent bias (PBIAS) of deep learning model simulated streamflow during training (1986-2002) and testing (2003-2010) periods

Climate Data	**Downscaling Method**	**NSE**		**MSE ((mm/day)^2)**		**PBIAS (%)**	
		Training	Testing	Training	Testing	Training	Testing
WRF	Raw	0.91	0.78	0.082	0.20	0.84	4.9
	Orographically Adjusted	0.90	0.86	0.089	0.13	0.54	-6.5
NLDAS	Raw	0.87	0.79	0.13	0.19	-0.15	7.9
	Orographically Adjusted	0.88	0.80	0.11	0.18	0.29	8.6

Table 8 shows the same statistics as in Table 7, but calculated for high flow only. Similarly as in Table 7 the WRF dataset leads to higher NSE and lower MSE. In addition, the WRF-forced model runs also yielded less biased streamflow for high flow periods. This is likely because the WRF dataset provides climate variables at a finer resolution. Noteworthy, all simulations yield positive PBIAS, suggesting the deep learning model underestimates flows especially during high flow periods. NLDAS-forced models are noticeably higher in PBIAS, possibly suggesting that the model has trouble with the lower snowmelt and rainfall values generated with NLDAS data even after orographic adjustment. Another possibility is the trend observed in Section 4.1.1 where the WRF and NLDAS datasets have a larger difference in precipitation in later years. This is likely the result of one dataset (possibly the NLDAS) not having consistent precipitation values. Therefore, the training and testing period would have differing patterns between snowmelt/rainfall and streamflow, thus leading to bias in simulated streamflow during the testing period.

Table 8: Nash-Sutcliffe efficiency (NSE), mean square error (MSE), and percent bias (PBIAS) of deep learning model simulated high flow (exceeding median flow) streamflow during training (1986-2002) and testing (2003-2010) periods

Climate Data	**Downscaling Method**	**NSE**		**MSE ((mm/day)^2)**		**PBIAS (%)**	
		Training	Testing	Training	Testing	Training	Testing
WRF	Raw	0.88	0.67	0.079	0.20	2.1	6.5
	Orographically Adjusted	0.87	0.80	0.086	0.12	2.5	0.21
NLDAS	Raw	0.82	0.68	0.12	0.19	2.0	8.4
	Orographically Adjusted	0.84	0.69	0.11	0.18	2.6	9.1

Both Tables 7 and 8 show NSE and MSE values for the testing period are very similar between the raw WRF and NLDAS datasets. Orographic adjustment of the NLDAS dataset yields similar values, however, orographic adjustment of the WRF dataset yields noticeably improved values. The reason for this is not completely clear. In Section 4.1.2, the orographically adjusted WRF simulations did not always result in the closest match to SNOTEL SWE values, but it is the only dataset that led to NSE values all greater than 0.5 (Table 4). Section 4.2 showed the WRF dataset and orographic adjustment resulting in later melting or a lengthened melt period. Another possible reason for better streamflow simulations is the orographically adjusted WRF dataset has the latest and longest melt period, which is easier for the deep learning model to simulate later streamflow accurately.

Figures 17 and 18 shows streamflow and spatially averaged UEB simulated snowmelt and rainfall from three years from the training and testing periods, respectively.

As expected from results seen in Section 4.4, the WRF dataset leads to later snowmelt than the NLDAS dataset.

Figures 17a, 17c, 18a, and 18c show two or three peaks in streamflow (bi-modal and tri-modal patterns) within a year, which is common in the study area. These figures show the orographically adjusted datasets are sometimes able to capture later peaks better

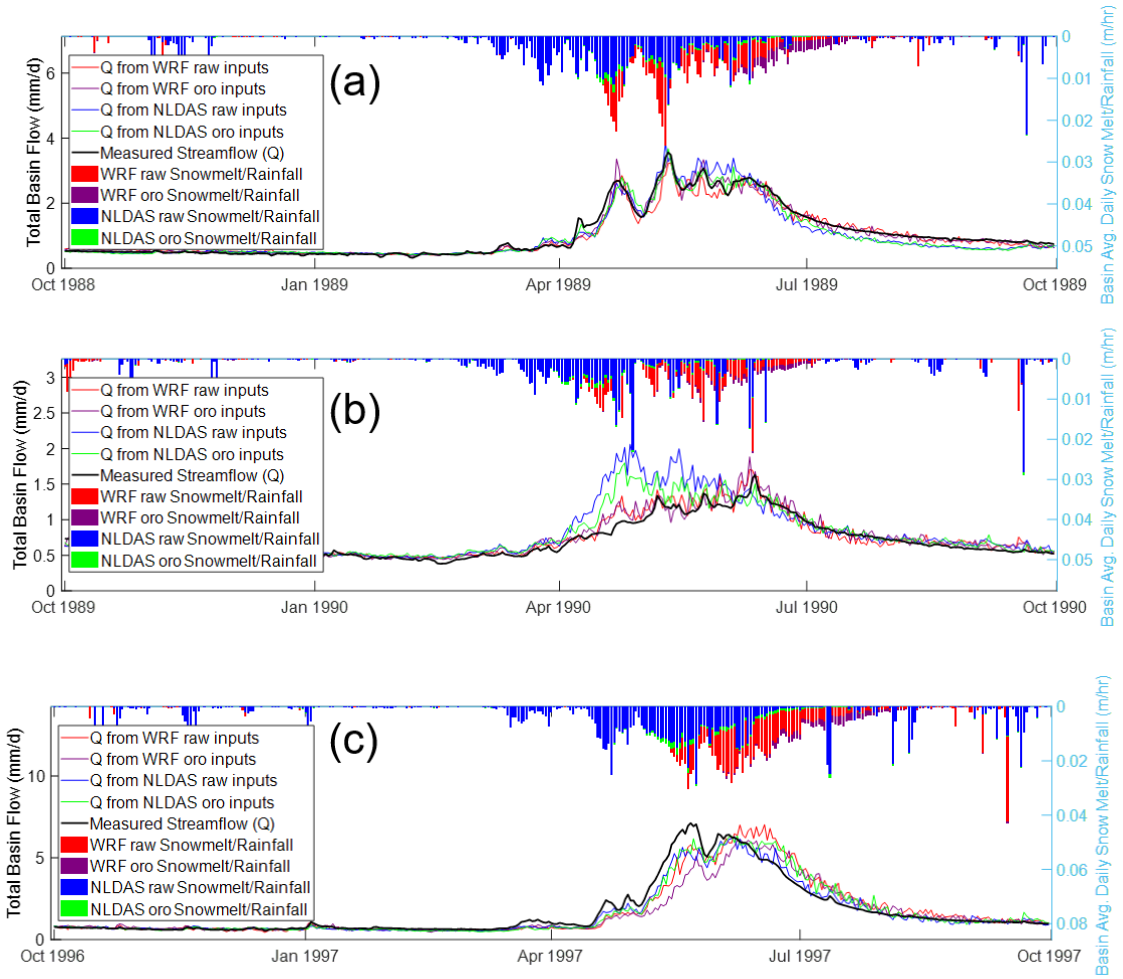

Figure 17: Measured and simulated streamflow compared to combined snowmelt and rainfall from the different datasets and downscaling methods for three years from the training period: (a) WY 1989 (an average year), (b) WY 1990 (a dry year), and (c) WY 1997 (a wet year).

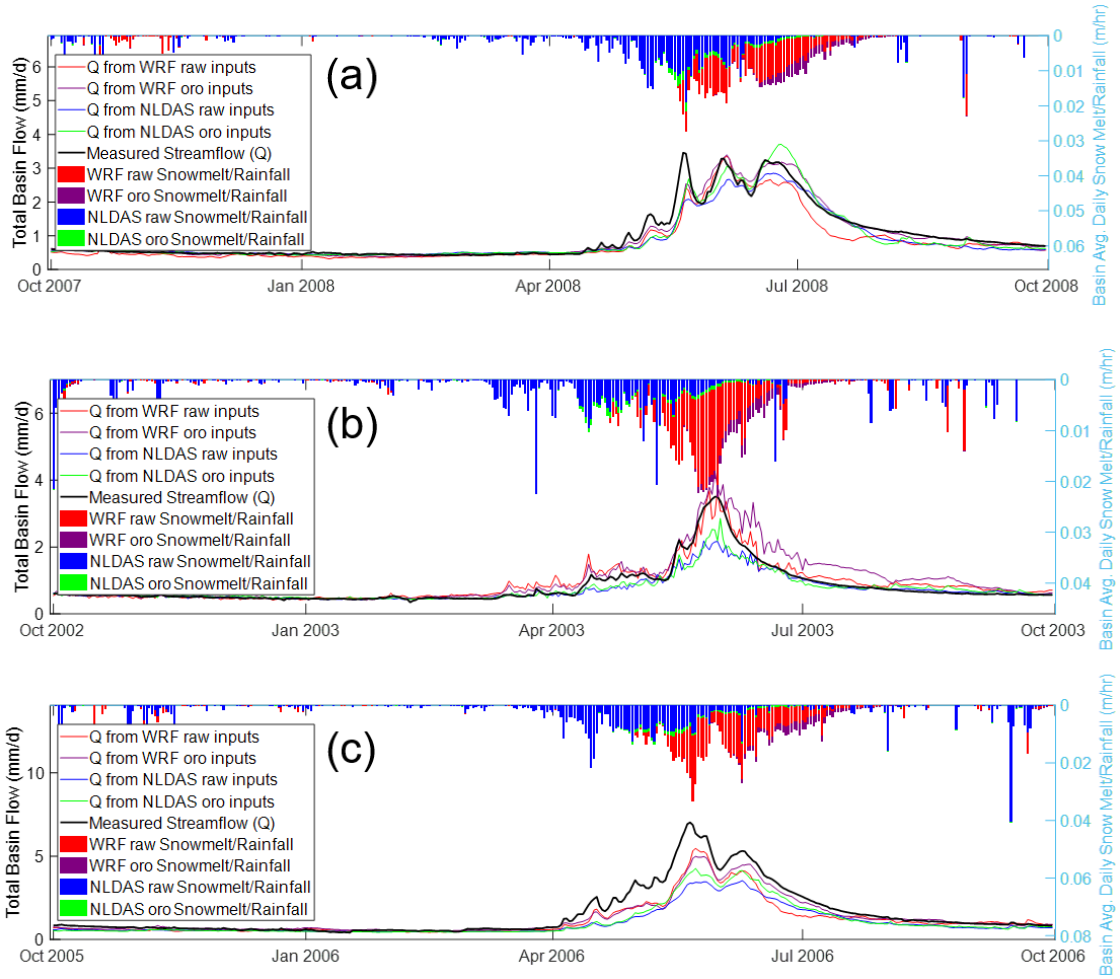

Figure 18: Measured and simulated streamflow compared to combined snowmelt and rainfall from the different datasets and downscaling methods for three years from the testing period (*a)* WY 2008 (an average year), (*b*) WY 2003 (a dry year), and (c) WY 2006 (a wet year).

than the raw datasets. For example, as shown in Figure 18a, the first two streamflow peaks are captured by simulations driven by the raw datasets, with the raw WRF dataset simulating the closest match to the second peak, but orographically adjusted data results in a higher distinct third peak. This occurs despite the snowmelt and rainfall having only small differences between raw and orographically adjusted data. Figure 16f shows a similar result. The results of Section 4.2 can possibly explain this ability to capture a later

peak due to snowmelt at higher elevations. Orographically adjusted data results in the same starting timing of snowmelt, but a longer melt period. Figure 15 in Section 4.2 further showed this lengthened melt occurred primarily at higher elevations, which melt last. Therefore, orographically adjusted data would result in more simulated snowmelt when these later peaks occur.

UEB simulations with the WRF dataset tend to have peak snowmelt very close to the time of peak streamflow, while NLDAS-forced simulations yield earlier peak snowmelt, matching the lag time values from Section 4.2. Since NLDAS-forced UEB consistently simulates peak snowmelt earlier than WRF-forced UEB, the deep learning model, is able to learn the different streamflow responses to snowmelt from the training period and yield satisfactory performance during the testing period.

Overall, the results suggest the WRF orographic dataset produces more accurate streamflow than NLDAS or WRF raw data. Given the NLDAS dataset has a resolution of about 14 km, whereas the WRF dataset has a 4 km resolution and orographically adjustment downscaled to 100 m, these results suggest higher resolution climate data can improve streamflow predictions as was seen in other studies (Mendoza et al., 2016; Schlögl et al., 2016; Winstral et al., 2014). However, the added benefit from high resolution only justifies the extra time and work to downscale climate variables if the base dataset has consistent data patterns. Looking at Table 7, orographic adjustment of NLDAS data improves the NSE value for the testing period by only 0.01. However, orographic adjustment of WRF data improved the NSE by 0.08 and results in the best fit to the data. Overall, the deep learning model is able to compensate for many deficiencies in climate data resolution, provided the model is given sufficient data in the training

period to learn the patterns and responses of the specific dataset, but the model is best able to simulate streamflow when the dataset has consistent patterns between the training and testing period.

CHAPTER V

CONCLUSIONS

This project is focused on determining how the different climate datasets and downscaling methods affect snow and streamflow modeling of the canyon portion of Logan River, a snow-dominated karst watershed. Climate data from two datasets (WRF and NLDAS) as raw and orographically adjusted, respectively, were input to a UEB snow model to simulate snow accumulation and melt at a fine resolution (100 m). The results from the snow model were fed into a deep learning karst model to simulate streamflow.

In order to accomplish our first objective, we compared the meteorological forcings between the datasets, compared the resulting SWE levels temporally and spatially, and compared these simulated SWE levels to measured SWE in the study area. The datasets and downscaling methods led to different values of meteorological forcings. In particular, the WRF dataset has higher precipitation, lower temperature, and lower incoming longwave radiation than the NLDAS dataset. Overall, orographic adjustment raised average precipitation and lowered average temperature and longwave radiation for both datasets, but to a lesser degree than between the two datasets. Comparing the precipitation values to measured data at SNOTEL sites leads to mixed results. Generally WRF data has precipitation values higher than was measured, while NLDAS precipitation values were lower than measured. Therefore, neither dataset matched the SNOTEL measurements substantially better than the other.

The different datasets and downscaling methods resulted in substantial differences in simulated SWE and timing and of snowmelt. The WRF dataset resulted in higher SWE

levels across all of the years simulated even in years when NLDAS data had higher accumulative precipitation. This suggested multiple climate variables interacted to influence SWE. In particular, the lower temperature and incoming longwave radiation of WRF dataset likely resulted in higher SWE. The WRF dataset resulted in snowmelt occurring one month later than the NLDAS dataset. This is likely the result of higher SWE and lower incoming longwave radiation. Orographic adjustment of data led to increased peak SWE for both datasets, but to a lesser degree than was observed between the two datasets. In addition, orographic adjustment increased the melting period length, resulting in more snowmelt occurring at a later date than was seen with raw datasets. This is likely because orographic adjustment increased precipitation and decreased incoming longwave radiation and temperature at higher elevation.

Comparing the simulated SWE to SNOTEL measurements revealed similar patterns as observed when comparing the precipitation values. Overall, the WRF dataset resulted in SWE values higher than measured, while the NLDAS dataset resulted in SWE values lower than measured. Neither dataset had the best match to measured data at every SNOTEL site. However, comparing simulated SCA to a remote sensing product (MOD10A1) showed SCA simulated from the NLDAS dataset resulted in a closer match. Compared to MOD10A1, the WRF dataset resulted in similar onset of melt but the higher SWE levels result in snowmelt taking one month longer to complete. However, both SNOTEL data and MOD10A1 have known biases, and it is difficult to determine if any dataset results in simulated values closer to actual snow levels.

In order to accomplish our second objective, we performed correlation analysis between simulated snowmelt and streamflow. We also compared some of these

correlations to correlations between measured values in the study area, and we observed how and where different datasets and downscaling methods affected the snowmelt simulations. Although the complexities of a large area, mountainous terrain, and geologic heterogeneity made correlation analysis inconclusive, multiple patterns between snowmelt rate and the streamflow were discovered. High and low flow years were found consistent with years with high and low precipitation, respectively, suggesting the interannual variability was correctly captured by all of the datasets. However, the higher peak SWE levels generated with WRF data better matched the patterns between peak SWE and streamflow seen in SNOTEL data. Interestingly, although NLDAS generated relationships between peak SWE and streamflow differed from the patterns observed at SNOTEL stations, they showed comparable levels of correlation. Therefore, as was seen in the simulated snow, neither dataset led to clearly higher correlation between streamflow and simulated snowmelt rates.

Looking at the spatially varying snowmelt rates, orographic adjustment had a greater effect on snowmelt timing at areas of high elevation. This showed that higher elevations were more likely to have differences between dataset elevation and actual elevation, resulting in more pronounced orographic affects. Therefore the snowmelt differences between raw and orographically adjusted data occur mostly at high elevations, resulting in different spatial patterns between the two downscaling method weather data.

In order to accomplish our third objective, we used the snow model outputs to run the deep learning karst model and analyzed the deep learning model's results. Although the different climate datasets and downscaling methods resulting in highly different

snowmelt rate, when these results are fed into a deep learning karst model, the model is able to simulate streamflow with acceptable accuracy for all cases. Because the overall patterns of simulated snowmelt rate of each UEB runs are overall consistent through the training and testing periods, the deep learning model is able to learn the response for a specific dataset and accurately simulate streamflow for this dataset. However, for all datasets except for the orographically adjusted WRF, an overall positive bias is found during testing period mainly due to underestimating streamflow during high flow periods.

The orographically adjusted WRF dataset results in the most accurate streamflow simulations. This is likely because the WRF dataset was generated by a high resolution regional climate model (4-km) and orographic adjustment led to higher SWE levels at high elevations and/or longer melt periods that more accurately represent actual conditions. However, for the NLDAS dataset the orographic adjustment only resulted in very small improvements in streamflow simulation accuracy, which may not justify the time and effort to downscale climate variables. Using longer periods of data to better constrain the deep learning model may be more useful in improving streamflow simulation accuracy.

The methods used in this project can be applied to other watersheds, both karst and non-karst, to determine whether the results obtained are specific to the Logan River watershed. In addition, the methods could be applied to sub-watersheds which have measured streamflow or spring discharge. A smaller test area would result in individual cells accounting for a larger proportion of the flow, which is advantageous for the correlation analyses.

Additionally, sensitivity analyses could be performed on the trained deep learning model. For example, the model can be run repeatedly with modified snowmelt inputs, each time with one cell masked, to determine how sensitive the streamflow is to each cell. A map could then be generated of the watershed showing which areas have the largest effect on streamflow. Similarly, a temporal sensitivity analysis could be performed to identify the time period of snowmelt and ET that has the greatest effect on streamflow.

CHAPTER VI

ENGINEERING SIGNIFICANCE

This study compared simulated precipitation and SWE from a dynamically downscaled dataset and an interpolated reanalysis dataset to measured values in the study area. Results show neither dataset capture field conditions exactly, however the dynamically downscaled data generally had higher precipitation and SWE than was measured, while the reanalysis dataset generally had lower precipitation and SWE. Researchers need to factor these biases when utilizing forcing datasets. However, comparing the simulated snow areas to satellite data, all simulations resulted in higher winter SCA than was detected. These results confirmed a shortcoming in satellite data, satellite SCA fails to accurately measure peak SCA in complex, highly vegetated landscape. Researchers need to be aware of this shortcoming when using the data to validate results.

The results from this study confirm that orographic adjustment of climate variables down to a fine scale can help improve streamflow simulations; however, the improvement gained from orographic adjustment depends on the raw dataset. The lack of improvement to results from orographically adjusted NLDAS data show the importance of ensuring the base dataset does not contain temporal trends of bias. Further, results suggest that when utilizing a deep learning model, ensuring consistency in the data is more critical than adjusting the data to a fine scale. This finding can help future researchers determine where to focus their time for best results. The modeling method with a physically based snow model feeding a data-driven streamflow model utilized in

this paper provides a potentially useful means of simulating streamflow and providing information to support water resources management for areas with snow dominated karst watersheds.

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*, 962–972.

Archer, C.L., and Jacobson, M.Z. (2003). Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements: FEASIBILITY OF U.S. WIND POWER. Journal of Geophysical Research: Atmospheres *108*, n/a-n/a.

Clark, M.P., Hendrikx, J., Slater, A.G., Kavetski, D., Anderson, B., Cullen, N.J., Kerr, T., Örn Hreinsson, E., and Woods, R.A. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review: REPRESENTING SPATIAL VARIABILITY OF SWE IN MODELS. Water Resources Research *47*.

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 *78*, 13.

Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology *26*, 707–721.

Daly, C., Neilson, R.P., and Phillips, D.L. (1994). A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. Journal of Applied Meteorology *33*, 140–158.

Dover, J.H. (2007). Geologic map of the Logan 30’ x 60’ quadrangle, Cache and Rich Counties, Utah, and Lincoln and Uinta Counties, Wyoming.

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.

Hall, D.K., and Riggs, G.A. (2007). Accuracy assessment of the MODIS snow products. Hydrological Processes *21*, 1534–1547.

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).

Hamon, W.R. (1960). ESTIMATING POTENTIAL EVAPOTRANSPIRATION. 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: KARST WATER RESOURCES PREDICTION. Reviews of Geophysics *52*, 218–242.

Hegdahl, T.J., Engeland, K., Steinsland, I., and Tallaksen, L.M. (2019). Streamflow forecast sensitivity to air temperature forecast calibration for 139 Norwegian catchments. Hydrology and Earth System Sciences Discussions 1–28.

Huang, X., Liang, T., Zhang, X., and Guo, Z. (2011). Validation of MODIS snow cover products using Landsat and ground measurements during the 2001–2005 snow seasons over northern Xinjiang, China. International Journal of Remote Sensing *32*, 133–152.

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).

Khair, U., Fahmi, H., Hakim, S.A., and Rahim, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. Journal of Physics: Conference Series *930*, 012002.

Knoben, W.J.M., Freer, J.E., and Woods, R.A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences *23*, 4323–4331.

Lewis, R.S., Link, P.K., Stanford, L.R., and Long, S.P. (2012). Geologic Map of Idaho.

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*.

Mendoza, P.A., Mizukami, N., Ikeda, K., Clark, M.P., Gutmann, E.D., Arnold, J.R., Brekke, L.D., and Rajagopalan, B. (2016). Effects of different regional climate model resolution and forcing scales on projected hydrologic changes. Journal of Hydrology *541*, 1003–1019.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., et al. (2006). North American Regional Reanalysis. Bulletin of the American Meteorological Society *87*, 343–360.

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.

Monteith, J., and Unsworth, M. (2008). Principles of Environmental Physics (USA: Academic Press).

Murphy, A. (1998). Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. Monthly Weather Review *116*, 8.

Najafi, M.R., Moradkhani, H., and Piechota, T.C. (2012). Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. Journal of Hydrology *442–443*, 105–116.

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

Raleigh, M.S., Livneh, B., Lapo, K., and Lundquist, J.D. (2016). How Does Availability of Meteorological Forcing Data Impact Physically Based Snowpack Simulations? Journal of Hydrometeorology *17*, 99–120.

Ren, W.W., Yang, T., Huang, C.S., Xu, C.Y., and Shao, Q.X. (2018). Improving monthly streamflow prediction in alpine regions: integrating HBV model with Bayesian neural network. Stochastic Environmental Research and Risk Assessment *32*, 3381–3396.

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.

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.

Spangler, L. (2011). Karst hydrogeology of the Bear River Range in the vicinity of the Logan River, northern Utah. Geological Society of America Rocky Mountain – Cordilleran Section Meeting.

Tarboton, D., and Sen Gupta, A. (2013). UEBGridInterfaceDesign.docx.

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

Tarboton, D., Bloschl, G., Cooley, K., Kirnbauer, R., and Luce, C. (2000). Spatial Snow Cover Processes at Ku¨ htai and Reynolds Creek. In Spatial Patterns in Catchment Hydrology: Observations and Modelling, (Cambridge: Cambridge University Press), pp. 158–186.

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).

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

USDA (2016). Snow Telemetry (SNOTEL) Data Collection Network.

White, W.B. (2002). Karst hydrology: recent developments and open questions. Engineering Geology *65*, 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*, 1366–1383.

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*, n/a-n/a.

Xu, T., Qianqiu, L., Tyson, C., Zeng, R., and Neilson, B. (In Prep). Hybrid physically-based and deep learning modeling of a snow dominated mountainous karst watershed.

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.