Process-Based Calibration of WRF-Hydro Model in Unregulated Mountainous Basin in Central Arizona

by

Abdinur Hussein

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2020 by the
Graduate Supervisory Committee:

Giuseppe Mascaro, Co-Chair
Enrique Vivoni, Co-Chair
Tianfang Xu

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

The National Oceanic and Atmospheric Administration (NOAA)'s National Water Model (NWM) will provide the next generation of operational streamflow forecasts at different lead times across United States using the Weather Research and Forecasting (WRF)-Hydro hydrologic system. These forecasts are crucial for flood protection agencies and water utilities, including the Salt River Project (SRP). The main goal of this study is to calibrate WRF-Hydro in the Oak Creek Basin (OCB; ~820 km^2), an unregulated mountain sub-watershed of the Salt and Verde River basins in Central Arizona, whose water resources are managed by SRP and crucial for the Phoenix Metropolitan area. As in the NWM, WRF-Hydro was set up at 1-km (250-m) resolution for the computation of the rainfall-runoff (routing) processes. Model forcings were obtained by bias correcting meteorological data from the North American Land Data Assimilation System-2 (NLDAS-2). A manual calibration approach was designed that targets, in sequence, the sets of model parameters controlling four main processes responsible for streamflow and flood generation in the OCB. After a first calibration effort, it was found that WRF-Hydro is able to simulate runoff generated after snowmelt and baseflow, as well as magnitude and timing of flood peaks due to winter storms. However, the model underestimates the magnitude of flood peaks caused by summer thunderstorms, likely because these storms are not captured by NLDAS-2. To circumvent this, a seasonal modification of soil parameters was adopted. When doing so, acceptable model performances were obtained during calibration (2008-2011) and validation (2012-2017) periods (NSE > 0.62 and RMSE = ~2.5 m^3/s at the daily time scale).

The process-based calibration strategy utilized in this work provides a new approach to identify areas of structural improvement for WRF-Hydro and the NWM.

ACKNOWLEDGMENTS

This work was developed under the project “Calibration of WRF-Hydro in the Salt and Verde Rivers for Future Implementation in the National Water Model” funded by the Salt River Project (Award number AWD00032221). The views expressed in this work do not necessarily represent those of the supporting organization. I would like to express my utmost gratitude to National Center for Atmospheric Research (NCAR), especially Dave Gochis, for providing the hydro-meteorological forcing data.

I owe my utmost gratitude to my advisor Professor Giuseppe Mascaro, who gave me an opportunity to work on this project and helped me succeed in my pursuit of studying a master’s degree at Arizona State University. He has been instrumental in guiding me throughout this work, from the conception of the thesis topic to revisions of the work. This thesis would not have been complete without his advice, encouragement, and support. I would also like to extend my gratitude to other committee members; Dr. Vivoni and Dr. Xu for their time and feedback that culminated in enriching the scope of my thesis.

I would also like to thank my fellow colleagues and friends; Adil, Xin for their help. I would also like to thank my brother, Ibrahim and his wife Faiza for helping me acclimatize to living abroad and wonderful times we shared.

Finally, to my deceased father, I wish you were alive today to witness this momentous accomplishment. I am sure you would have been proud. I dedicate this milestone to you to honor your memory. To my mother, thank you for your unparalleled love, support, and prayers. My academic journey would not have been possible without both of you

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

1.1 Literature Review 8

2 STUDY AREA AND DATASETS 10

2.1 Study Area and its Characteristics 10

2.2 Datasets 16

2.2.1 Geospatial Datasets 16

2.2.2 Model Forcings 16

2.2.3 Ground Observations 17

3 METHODOLOGY 19

3.1 Flood Characterization 19

3.2 Overview of WRF-Hydro 23

3.2.1 Surface Energy Balance 25

3.2.2 Runoff and Infiltration 25

3.2.3 Aquifer Recharge 27

3.2.4 Snow Dynamics 27

3.2.5 Terrestrial Hydrologic Routing 28

3.2.6 Baseflow 29

CHAPTER Page

3.3 Model Setup in the Oak Creek Basin....30

3.4 Precipitation Bias Correction....31

3.5 Model Calibration Targeting Distinct Hydrological Processes....35

3.6 Metrics....40

4 RESULTS....42

4.1 Model Spin-up....42

4.2 Hydrologic Simulation with Default Parameter Values....42

4.3 Step 1: Calibration of Winter and Summer Floods....44

4.3.1 Effect of DKSAT and REFKDT....44

4.3.2 Effect of SMCMAX....49

4.4 Step 2: Calibration of Baseflow....51

4.5 Step 3: Shape of Hydrograph....54

4.6 Step 4: Snow melt....55

4.6.1 Effect of MFSNO....55

4.6.2 Vegetation Parameters....58

4.7 Step 5: Final Tuning....60

4.8 Step 6: Validation....64

5 DISCUSSION....71

5.1 Forcing Limitations....71

5.2 Summary of Parameter Sensitivties....72

5.3 Comparison with Existing Literature....74

5.4 Model Limitations....75

CHAPTER Page

5.5 Terrestrial Water Balance Model 76

6 CONCLUSIONS AND FUTURE WORK 79

REFERENCES 82

APPENDIX

A SUPPLEMENTAL FIGURES FOR MODEL SPIN-UP 88

B SNOW DYNAMICS AT POINT SCALE 91

C SUPPLEMENTAL FIGURES FOR BIAS-CORRECTION 93

LIST OF TABLES

Table Page

1. Noah-MP Selected Physics Options for this Study. The ID Numbers Refer to the Values that can be Specified in the Model Input File…….....24

2. Slope and Intercept of the Regression Line Between Monthly Precipitation and Elevation. Note that these Regressions Should be Applied only in the Range of Elevation from ~1100 to ~2500 m a.s.l. that was used to Estimate the Coefficients…….……..33

3. WRF-Hydro Parameters Considered for Calibration for this Study With the Corresponding Ranges Based on Previous Publications. For Some of the Parameters, the Governing Equation Reported in Section 3.1 has been Indicated. The Main Hydrological Processes Controlled by Each Parameter is also Indicated. Based on these, Calibration of Parameter can be Tailored for Each Flood Generation Mechanisms……..38

4. Metrics Quantifying Model Performance during Calibration and Validation Periods. All Results Except for the Validation at Sedona are Referred to the Discharge at the Basin Outlet at Cornville. The Metrics are Computed at Daily Resolution. The Asterisk (*) Shows the Simulations where Calibration Worsened the Metrics……..47

Table Page

5. Metrics Quantifying the Model Performance in the Simulation of Winter and Summer Flood, as well as of Runoff due to Snow Melt. (*) Indicates a Simulation where the Metric Worsened when Compared to the Previous Run. Note that the Simulation Runs that Does not Significantly Impact MRE, T_{lag}, and Snow melt Bias are not Shown in Table……..…….....48

6. Metrics Quantifying Model Performance in Simulating Snow during Calibration and ValidationPeriods……..70

7. Annual Values of the Water Balance Components; i.e, P, ET, and R Computed from Outputs of Calibrated WRF-Hydro Model for 2008-2016. The Storage Term is Given by $Ds/Dt_1 = P – (R_{obs}+ET)$ While $Ds/Dt_2 = P – (R_{sim}+ET)$. The Storage Term Shows both Negative and Positive Values Indicating Wet and Dry Years. The Terrestrial Water Balance Does not Zero-Out for All the Years Because of Model Limitations and or Limitations of Manual Inadequate Calibration…….……78

LIST OF FIGURES

Figure Page

1. (a) Salt and Verde Basins Within the State of Arizona. (b) Digital Elevation Model (DEM) of the Salt and Verde Basins Draining to the Phoenix Metro Region, along with Boundaries of the Oak Creek Sub-basin. (c) 250-m DEM and Boundaries of the Oak Creek Watershed Drained at the USGS Stream gage near Cornville, along with Stream Networks, Rain gages, SNOTEL Stations, and USGS Stream gages……..11

2. (a) Climatological Mean Monthly Precipitation (P), Temperature (T), and Runoff (Q) in the OCB. The Monthly P and T Means were Calculated by Averaging the Climatological Means of Observations of all Gages and SNOTEL Stations. The Monthly Q Means were Derived from the Instantaneous Discharge at the Cornville gage and Converted to Depth (mm). (b) Month and Day of Occurrence of the Annual Peak Discharges at Cornville Stream Gage Recorded from 1941 to 2019 Showing Bimodal Distribution of Flood Peaks……...……..13

3. Maps of Dominant (a) Soil Texture and (b) Land Cover Types in 1-km Pixels in the Oak Creek Basin Used for the Hydrologic Simulations……...….……...15

Figure Page

4. Characterization of Streamflow and Flood Regimes. (a) Time Series of Observed Discharge showing three Distinct Streamflow and Flood Regimes, i.e., Peaks due to Winter Rain, Summer Monsoons, and Snow melt. There is also Constant Baseflow of >0.48 m^3/s. (b) Zoom on Snowmelt Season showing Diurnal Variation of Discharge due to Effect of Temperature on Snow. (c) Mean Snow Depth Associated with the Streamflow and Flood Regimes. Note that the Depth of Snow Decreases during Late Spring due to Snow melt……..……...22

5. (a)-(b) Mean Monthly Precipitation for (a) January and (b) July 2008 from Gages and NLDAS Pixels. (c) Monthly Mean Areal Precipitation for 2008 Before and After Bias Correction……..34

6. Schematic of Calibration Steps showing the Runoff Generation Processes and Parameters that Control each Process. Note that Vegetation Parameters such as *HVT*, *MP*, *CWPVT*, and *VCMX25* also Affects Winter and Summer Peaks beside Snow melt……37

Figure Page

7. (a) Time Series of Observed Discharge and Simulated Discharge using the Default Parameters for the Period 2008-2011. Simulated Discharge using Default Parameters shows Significant Time Lag during Snow Melt Seasons, Poor Simulation of Winter Rain and Summer Monsoon Peaks, and Under-simulation of Baseflow. (b) Flow Duration Curve (FDC) of Observed and Simulated Discharge for the entire Time Series. (c) Zoom on the Portion of FDC Accounting for the Flood Peaks…….....43

8. (a) Time Series of Observed Discharge and Simulated Discharge for the Calibration Period. (b)-(c) Zoom on two Winter Rain Peaks in (b) January and (c) July 2008. (d) FDCs of Observed and Simulated Discharge for the Entire Time Series. (e) Zoom on the Portion of FDC Accounting for Floods…….…...46

Figure Page

9. (a) Time Series of Observed Discharge and Simulated Discharge for the Calibration Period. (b) Zoom on two Winter Rain Peaks in January of 2008, shows Better Simulation of Peaks using Small Multiplicative Values of *SMCMAX* = 0.75 (red line) while Higher Values of *SMCMAX* = 1.5 Results in Relatively Poor Simulation of The Peaks (blue line) (c) Zoom on a Summer Monsoon Peak in August of 2008, Simulation of Peaks using Small Multiplicative Values of *SMCMAX* = 0.75 (red line) while Higher Values of *SMCMAX* = 1.5 Results in Relatively Poor Simulation of the Peaks (blue line) (d) FDC of Observed and Simulated Discharge for the Entire Time Series. (e) Zoom on the Portion of FDC Accounting for the Floods……50

10. (a) Time Series of Observed and Simulated Discharge using the Default and Calibrated Parameters of the Groundwater (GW) Module, Parameters that affect Groundwater Recharge such as *SLOPE*, *BEXP* and *SMCMAX* for the Calibration Period. (b) Zoom on a Snow melt Period showing Significant Delay in Snow melt during Late Spring. (c) FDC of Observed and Simulated Discharge showing Significant Under- simulation of Baseflow with Original Groundwater Parameters and very Small Bias in Baseflow Simulation with Calibrated Baseflow Parameters. (c) Zoom on the Portion of FDC Accounting for the Highest Peaks……..…..53

Figure Page

11. (a) Time Series of Observed and Simulated Discharge for Simulation Period (2008-2011). (b) Zoom on Snow melt Season in Spring of 2008 showing Improved Lag Time during Snow melt Season by Increasing *MFSNO* from Default Value of 1 to 3.5. (c) FDC of Observed and Simulated Streamflow. (d) Zoom on the Portion of FDC Accounting for the Highest Peaks……..……..……57

12. (a) Time Series of Observed and Simulated Discharge for Simulation Period (2008-2011). (b) Zoom on Snow melt Season in Spring of 2008 showing Slight Increase in Discharge for *HVT* = 2 (red line) (c) Zoom on Summer Season in August of 2008 showing no Effect of Increase in *HVT* on Summer Peaks (red line). FDC of Observed and Simulated Streamflow. (d) Zoom on the Portion Of FDC Accounting for the Highest Peaks……...……..…59

13. (a) Time Series of Observed and Simulated Discharge for Calibration Period (2008-2011) Simulated using Tuned Value of *DKSAT* and *OVROUGHRTFAC* to Improve Shape of Hydrograph. Note that Winter Rain Peaks are Simulated well while Summer Monsoon Peaks are Under-Simulated (b) FDC of Observed and Simulated Streamflow. (c) Zoom on the Portion of FDC Accounting for the Highest Peaks…….61

Figure Page

14. (a) Time Series of Observed and Simulated Discharge for Calibration Period (2008-2011) Simulated using Dynamic Values of *DKSAT* And *REFKDT* to Improve Simulation of Summer Peaks. Note that Smaller Values of *DKSAT = 0.3* and *REFKDT = 0.1* Improves Summer Monsoon Peaks (b) Zoom on few Peaks in August of 2008 showing Marked Improvement in Simulation of Summer Peaks when Seasonal Dynamic Values of Parameters that Control Infiltration is Used. (c) FDC of Observed and Simulated Streamflow. (d) Zoom on the Portion of FDC Accounting for the Highest Peaks……63

15. Time Series of Validation Period (2012-2017). (a) Time Series showing Simulated and Observed Discharge at Basin Outlet. (b) FDC of Observed and Simulated Streamflow at the Outlet of the Basin. (c) Zoom on the Portion of FDC Accounting for the Highest Peaks……..67

16. Validation at Sedona Gage (2008-2017). (a) Time Series showing Simulated and Observed Discharge Evaluated at Oak Creek Near Sedona Gage. (b) FDC of Observed and Simulated Streamflow at Interior Gage. (c) Zoom on the Portion of FDC Accounting for the Highest Peaks……68

Figure Page

17. Comparison of Snow Simulation obtained Through Calibrated WRF-Hydro Model with SNODAS Model. (a) Simulated Snow Water Equivalent (SWE) (b) Simulated Snow Depth……...…….....69

CHAPTER 1

INTRODUCTION

Streamflow predictions are crucial for water resources management (Yeh et al. 1982) and civil protection from flooding (Anghileri et al. 2016), which are two fundamental tasks to support economic prosperity while protecting lives and property. The need of accurate streamflow predictions has become more pressing in light of the increasing occurrence and magnitude of hydrologic extremes, including floods and droughts (Endicott 2013). For example, from 1980 to 2019, extreme flood events experienced throughout the continental United States have led to more than 13,000 fatalities and damages for ~$1.75 trillion to properties, infrastructure, and crops (NOAA 2020). These damages have increased over the last decades, making floods one of the deadliest disasters in the world (Wijayarathne and Coulibaly 2020). Prolonged drought events have also caused significant damages to agriculture. This coupled with projected future climate change concerns have caused and continue to cause significant tension between water managers and users (Gautam et al. 2018). In Arizona, to reduce the effects of such prolonged droughts and floods, extensive system of reservoirs, dams, and aqueducts have been built for water distribution and storage purposes. In addition to surface water resources management, underground storage facilities have been developed to store excess water during wet years so that the stored groundwater can be pumped for use during dry years.

To support water managers and reduce the effects of hydrologic extreme events, a number of streamflow forecasting techniques have been developed over the past decades based on statistical approaches and hydrological models. Statistical methods use past

streamflow records and climate observations, indices, and simulations to predict streamflow over different time periods and lead times. For example, (Slater et al. 2019) presents dynamic statistical method, where statistical model is fitted based on seasonal precipitation observations, crop acreage, and other indices to dynamically improve probabilistic streamflow forecast by improving predictand and predictors interrelation annually. Historically, the National Weather Service (NWS) also conduct streamflow forecast based on observations from various agencies such as U.S. Geological Survey (USGS), the U.S. Army Corps of Engineers, and other local agencies and uses it as the basis for forecasting the potential damages of future floods (Williams 1975; Zalenski et al. 2017). An alternative strategy is based on the application of hydrologic models. Currently, the NWS River Forecasting Centers use conceptual, lumped hydrologic models to issue flood predictions. To do this, precipitation observations and or rainfall forecast is used to drive lumped rainfall-runoff model and the amount of runoff generated for each contributing sub-basin is used to determine the magnitude and time of the peak at basin outlet (Hydrologist and NOAA 2010). Previous studies such as (Zalenski et al. 2017) have explored forecasting errors that results from application of river stage forecasting models used by NWS and their findings points towards the need to use more sophisticated models for operational flood forecasting.

The next generation of operational streamflow forecasts issued by the NWS will be provided by the National Water Model (NWM), which is currently under development and testing by the Office of Water Prediction (OWP) National Water Center (NWC). The basis for NWM is community-based, distributed WRF-Hydro model. The model utilizes mathematical models to represent physical processes such as infiltration, water

movement through soil layers, and snow formation and ablation and can be used to forecast streamflow at different lead times across Continental United States (CONUS). Unlike currently operational forecasting model, NWM forecast all hydrologic processes which define energy and water balance, therefore, uncertainties in streamflow forecast is was found to reduce (Yucel et al. 2015; Senatore et al. 2015). Also, the model can expand streamflow forecast to locations where none previously existed. The forecast lead times include: short range forecast which is an 18-hour deterministic forecast used largely for forecasting flash floods, medium range forecast which is approximately 10-day forecast used to for flood control and long range forecast which is typically 30-day forecast used for reservoir management. Currently, the default mode of NWM provides adequate streamflow forecast only in certain regions of CONUS. For example, in humid southeastern regions, and eastern and north-eastern regions, the model provides high correlation between observation and forecasted simulation but with a positive volumetric bias (Dugger et al 2017). In northern plains and southwestern US, the model is subject to substantial streamflow forecast errors (i.e., very low correlation and negative bias) (Dugger et al 2017), therefore, calibration of the model of the model is needed in these regions.

One of the tasks currently undertaken by the NWC is to calibrate the NWM across the country using distributed streamflow observations of the United States Geological Survey (USGS). In addition, efforts have been devoted to by the international research community to calibrate WRF-Hydro in different basins of the world and in United States (US) e.g., Yucel et al. (2015) in Turkey, Kerandi et al. (2018) in Kenya, Senatore et al. (2015); Verri et al. (2017) in Italy, Silver et al. (2017) in Israel, Naabil et al. (2017) in

Ghana, and Lahmers et al. (2019); Somos-Valenzuela and Palmer (2018); Xue et al. (2018) in US.

A recent study by (Lahmers et al. 2015, 2019) has calibrated WRF-Hydro in two basins (Gila River and Babocomari River basins) in southern Arizona, finding that although the model presents structural limitations, it can be used for streamflow forecasts. No study has yet been conducted to apply the model in the central and northern parts of the state. These regions, which are mainly part of the Salt and Verde River basins, are located in a transition zone from arid and semiarid climate of the Sonoran Desert to alpine climate of the Rockies. As such, the hydrologic regime includes a large variety of processes including snow, North American Monsoon (NAM), rain-on-snow, and continuous springs. Calibrating a hydrologic model under these conditions is a challenging task and very little research has been devoted to this effort. A notable exception is the work of (Hawkins et al. 2015), who applied a physically-based distributed hydrologic model in the Beaver Creek watershed (area of ~1100 km^2) in central Arizona to investigate the impact of climate change on its hydrologic response during the summer monsoon. Research is needed to (1) assess the ability of existing hydrologic models to simulate the variety of hydrologic processes occurring in this region, including snowmelt and winter floods; (2) suggest areas where existing models need structural improvements; and (3) improve operational streamflow predictions.

Currently, operational streamflow predictions and forecasts in the Salt and Verde Rivers are issued by the Salt River Project (SRP), the largest water provider for Phoenix metropolitan area. This metropolitan region is the fifth most populous in US, with more than 4.5 million people, and is experiencing one the fastest population growths according

to 2019 US Census Bureau. SRP delivers ~984 million m^3 of water per year through a system of seven reservoirs in the basins with a total capacity of ~2.8 billion cubic meters or ~2.3 million acre-foot and canals. To support its water management operations and decisions, SRP started generating seasonal streamflow forecasts in 1998 using the Entropy Limited precipitation model described in (Christensen, R. A., and Eilbert 1985). This statistical model utilizes historical records of air temperature and precipitation data at many global sites to produce seasonal probabilistic estimates of precipitation and runoff (Christensen, R. A., and Eilbert 1985). The simplistic nature of this model is reflected in the low accuracy of its streamflow forecast (55%) (Christensen, R. A., and Eilbert 1985) and is prone to misinterpretation which may lead to bad forecast. To further improve streamflow forecast, in 1999, SRP developed a multilinear regression model to convert the probabilistic forecast of Entropy Limited precipitation model into a deterministic forecast. The accuracy of statistical approach of forecasting streamflow depends on (1) the quality of original probabilistic estimates, (2) user's interpretation, and (3) goodness of fit of the regression model. Due to climate changes, the use of non-physical based model to forecast streamflow is erroneous. In areas such as Phoenix, where water management policies are centered around the accuracy of streamflow forecast, operation of reservoirs requires accurate deterministic forecast that is simulated by physically based hydrologic model. Given the shortcomings of the existing statistical model utilized by SRP, it is important to use hydrological model to forecast soil moisture conditions, snow depth, and runoff for better reservoir operations and preparedness of the occurrence of flash floods. One of such hydrological model is WRF-Hydro model which

has been shown to adequately forecast streamflow and floods at small spatial resolution (Yucel et al. 2015; Senatore et al. 2015).

In this study, WRF-Hydro model was calibrated in Oak Creek basin. This study aims to provide a basis for process-based approach to calibrate WRF-Hydro model. This approach identifies the hydrological processes of the study site, and aggregates model parameters based on parameter effect on hydrological processes, thus adding a new dimension of simplifying typical trial and error model calibration technique while reducing uncertainty in model parameter estimates described in (Yatheendradas et al. 2008). To do this, the observed streamflow at the basin outlet is characterized into four main runoff and flood generation mechanisms, i.e., winter and summer floods, snow melt, and baseflow. The analysis of the model outputs focuses on the model's ability to simulate the hydrological processes that controls the four runoff generation mechanisms. Before calibration, the forcing dataset was evaluated for its application in the study site. Precipitation bias-correction was performed based on elevation as in (Liston and Elder 2006). The bias-corrected forcing was used to calibrate the model. To capture volume and time of the peak of the floods, metrices was used to determine the parameter estimates that results in acceptable hydrological response. Spatially distributed physical model parameters are calibrated by applying multipliers which ensures that spatial distribution of such parameters is preserved. The sets of the parameters selected for calibration and the methodology adopted here is consistent with (Dugger et al. 2017). Specifically, the aim of this study is to (1) provide sets of parameters that can be incorporated in the NWM to improve streamflow forecasts for SRP, (2) evaluate whether WRF-Hydro is able to simulate the four main runoff and flood generating mechanisms in

the Oak Creek basin, and (3) suggest areas of improvements for the WRF-Hydro model structure. The goal of this thesis is to meet these objectives. Besides individual process, other metrics to assess hydrologic models was utilized to quantify the model performance through calibration and validation periods.

It is important to note that work and financial investments are needed to increase the utility of the NWM in southwestern and other regions of the country. This primarily involve model calibration, enhancement of model structure, and improvement of model forcings. This work presents calibration strategy, suggestions for model structure improvements, and bias-correction methodology that can be applied in other regions where the model is currently not performing well. To improve operational limitations of the default mode of NMW, SRP funded this work so that the calibrated model can be utilized for accurate streamflow forecast.

The characteristics of study and datasets used are described in chapter 2. The methodology of flood characterization and adopted calibration strategy are presented in chapter 3. The results of process-based model calibration through sensitivity analysis are in chapter 4. The main findings of the study and model limitations are discussed in chapter 5, and the study is summarized in chapter 6.

1.1 LITERATURE REVIEW

Calibration of physically based distributed hydrological models requires optimization of model parameters for better simulation of the different physical processes. In addition to manual optimization technique, different parameter optimization methods such as Shuffled Complex Evolution method developed at University of Arizona (SCE-UA) (Duan et al. 1994) , Particle Swarm Optimization (PSO) method (Chen et al. 2016) , and parameter estimation (PEST) method (Gallagher and Doherty 2007) have been developed to automatically determine optimal parameter values. Although the algorithms of these automatic methods are inherently different, they generally tend to optimize model parameters by (1) generating sets of parameters, and (2) allowing the parameter values to move through the surface/domain until optimal value is determined. Manual calibration technique involves selection of sets of parameters to be calibrated, running of simulations, comparing simulations and observations both qualitatively and quantitively through metrices, and selection of model parameter value that results in best possible match between observation and simulation. Automatic calibration techniques require excessive computational time to optimize parameters and may result in poor representation of model physical processes. Manual calibration technique also requires significant amount of time to optimize parameter value through repetitive trial and error method and its effectiveness depend on the skill of the modeler. (Boyle et al. 2000) proposed multicriteria formulation method of combining manual and automatic calibration techniques to take advantage of the strengths of each method and reduce overall weakness of each method.

Since the emphasis of this work is better representation of physical processes in the study basin, manual calibration approach was used. However, this work goes beyond the typical, unguided trial and error method by characterizing and targeting the different physical processes, hence, solving the difficulties associated with manually calibrating multi-parameter model described in (Yucel et al. 2015). The “guided” calibration approach proposed in this study reduces complexity of manual calibration and increases computational efficiency.

CHAPTER 2

STUDY AREA AND DATASETS

2.1 STUDY AREA AND ITS CHARACTERISTICS

The Oak Creek basin (OCB) is a sub-watershed of the Verde River basin located in north central Arizona between the longitudes of 111.4°W and 112.0°W and the latitudes of 34.7°N to 35.2°N (Figs. 1a, b). The OCB has an area of approximately 820 km^2, which includes an upper portion with mountainous ridges and a downstream flatter region. The basin elevation ranges from 1059 m above sea level (asl). at the outlet to 2581 m asl. at the peak of Mormon mountain, the average height is 1844 m asl. The basin is drained by Oak Creek River (OCR). The OCR is a perennial river that receives the contribution of several ephemeral streams. In the upper portion of the basin, it flows at the bottom of the Oak Creek Canyon until it reaches the city of Sedona; downstream of Sedona, the river flows in a more open region until it eventually drains into the Verde River. Together with the Salt River, the Verde River is a key water source for the Phoenix metropolitan region (Fig. 1b; Guan et al. 2019). The basin is fairly instrumented; there are National Centers for Environmental Information (NCEI) rain gages, SNOpack TELemetry (SNOTEL) stations and United States Geological Survey (USGS) gages (Fig. 1c). OCB was chosen for this study because (1) the water resources in the basin is relatively less regulated and (2) the characteristics of the basin is such that it encompasses a wide range of different geographical and climatological features in central Arizona (Mascaro 2017) thus the basin is a good representative of other sub-watersheds in Salt and Verde River Basins.

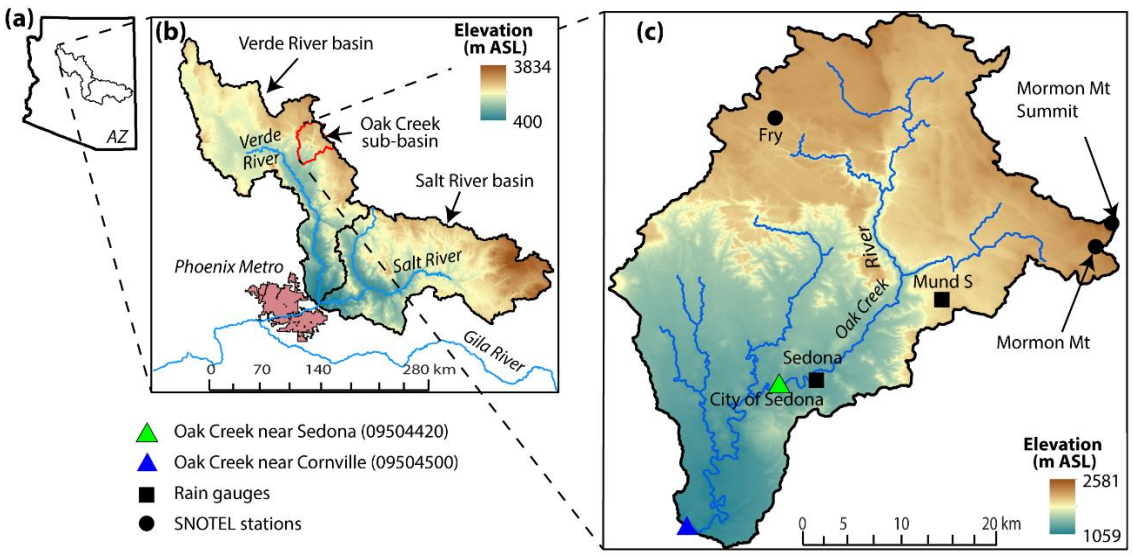

Figure 1: (a) Salt and Verde basins within the state of Arizona. (b) Digital elevation model (DEM) of the Salt and Verde basins draining to the Phoenix Metro region, along with boundaries of the Oak Creek sub-basin. (c) 250-m DEM and boundaries of the Oak Creek watershed drained at the USGS stream gage near Cornville, along with stream network, rain gages, SNOTEL stations, and USGS stream gages.

Climate in this region is characterized by two seasons with distinct storm types that lead to a bimodal precipitation distribution (Fig. 2a). From November to March, cold fronts mainly originated in the Pacific cause snowstorms that, combined with cold temperatures (averages approaching 0 ºC; Fig. 2a), sustain the existence of a snowpack at high elevations. Records at the SNOTEL stations indicate that the snowpack is present, on average, from mid-December to late April in this basin. The summer season from July to September is dominated by the North American monsoon that leads to diurnally modulated convective thunderstorms (Mascaro 2017). As a result, the rainfall is highly localized and characterized by high intensity and short rainfall duration. The mean annual temperature and precipitation are 8.2 °C and 565 mm, respectively. The seasonal precipitation regime impacts differently the river flow seasonality and flood regimes at the outlet of the basin. The mean monthly river flow is unimodal, with higher values in late winter and spring, due to snowmelt and rain-on-snow, and lower values throughout the rest of the year (Fig. 2a). The runoff continuously increases from late winter and peaks in early March, as a result of snowmelt occurring in the upper part of the basin (Fig. 2a). OCR is, therefore, a perennial river with a constant baseflow >~0.5 m^3/s (or > 2.5 mm). Note that summer convective storms do not impact significantly the mean flow because of high evapotranspiration rates and short storm events in this season (Fig 2a).

The flood regime is also bimodal, as revealed by the annual maxima recorded at the outlet of the basin from 1941 to 2019, with (1) larger peaks in winter originated by rain-on-snow and saturation-excess runoff generated by storms with mean duration of 10 hours based on NLDAS-2 forcing dataset, and (2) relatively smaller peaks in summer caused by high-intensity and short-duration of less than or one hour thunderstorms that

promote fast runoff generation, likely due to infiltration excess (Fig. 2b). The climatological means of the annual flood peaks for winter and summer seasons are 255 m^3/s and 100 $m^3/$, respectively.

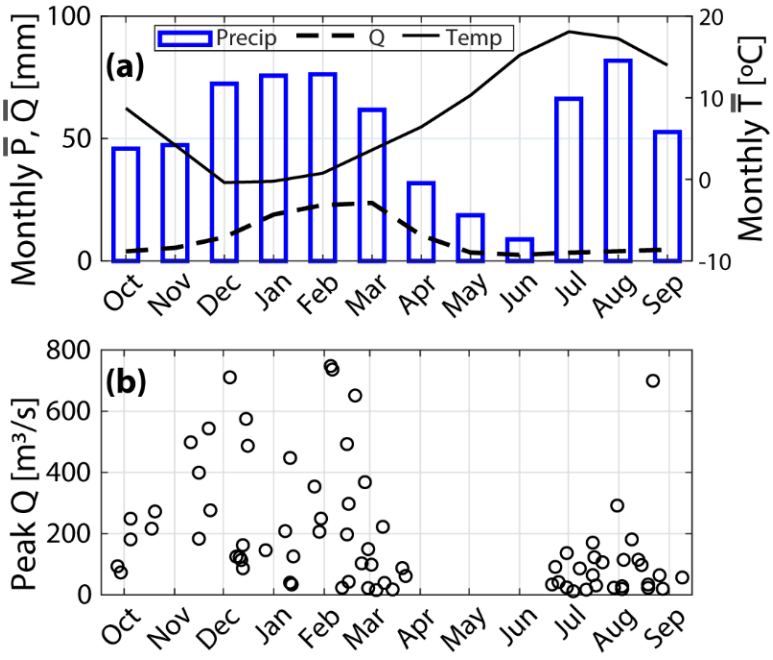

Figure 2: (a) Climatological mean monthly precipitation (P), temperature (T), and runoff (Q) in the OCB. The monthly P and T means were calculated by averaging the climatological means of observations of all gages and SNOTEL stations. The monthly Q means were derived from the instantaneous discharge at the Cornville gage and converted to depth (mm). (b) Month and day of occurrence of the annual peak discharges at Cornville stream gage recorded from 1941 to 2019 showing bimodal distribution of flood peaks.

Fig. 3 presents the maps of the dominant soil texture and land cover classes in 1-km pixels adopted for the hydrologic simulations derived from the Digital General Soil Map of the United States or STATSGO2 (Survey 1998), and the USGS National Land Cover (NLCD; Homer et al. 2015) database, respectively. The predominant soil type is loam, which accounts for approximately 60% of all soil types and is mainly found in the upper and lower parts of the basin (Fig. 3a). The second dominant soil type is bedrock, which accounts for ~21% of the soil type and is located in the middle part of the basin. The spatial distribution of these soil types indicates that physical characteristics of the spatially varying soil need to be preserved when modifying WRF-Hydro model soil parameters such as soil hydraulic conductivity. Turning the attention to land cover, the large majority (~85%) of the OCB is covered with evergreen needleleaf forest or pine trees, that occupy the northern and central parts of the basin (Fig. 3b). The lower parts of the basin are largely covered by shrubs accounting for 11% of all vegetation types. Other less prominent land cover types include grassland, irrigated cropland and urban vegetation, located in the southwestern lowland regions, mainly in and around city of Sedona. Unlike soil, the vegetation types in the basin are characterized by lower spatial heterogeneity.

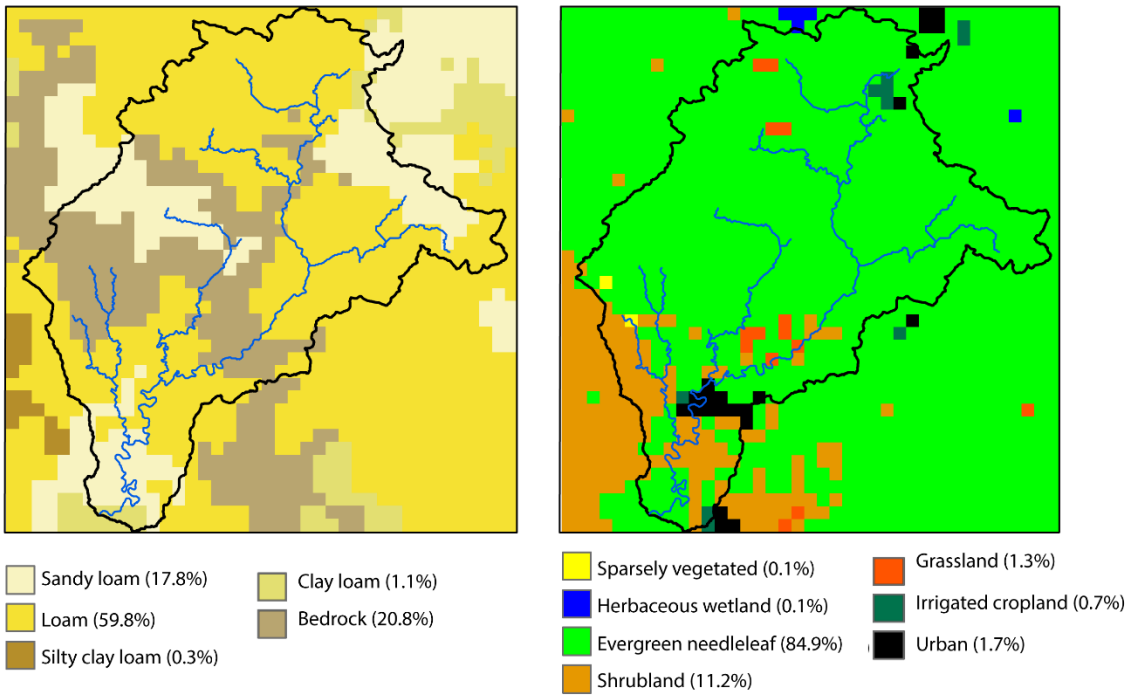

Figure 3: Maps of dominant (a) soil texture and (b) land cover types in 1-km pixels in the Oak Creek basin used for the hydrologic simulations.

2.2. DATASETS

2.2.1 GEOSPATIAL DATASETS

WRF-Hydro model was setup in OCB using the geospatial datasets adopted for the National Water Model (NWM). Terrain data were obtained from Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS) (Lehner et al., 2008), which is derived from digital elevation model (DEM) data of the Shuttle Radar Topography Mission (SRTM) with multiple resolutions including 250-m and 1 km (30 arc seconds). As mentioned, soil texture was obtained from STATSGO2 and land cover from NLCD. These datasets were compared with other sources such as Global Land One-kilometer Base Elevation (GLOBE) for terrain data and local soil survey from Soil Survey Geographical database (SSURGO) for soil texture and was determined to be appropriate data for the study. For the application WRF-Hydro: (1) 1-km grids were generated for terrain, soil texture, and land cover to apply Noah MP; and (2) a 250-m terrain grid was created to apply routing schemes.

2.2.2 MODEL FORCINGS

Spatially-distributed hydrometeorological forcings from 2008 to 2017 were obtained from the North American Land Data Assimilation System (NLDAS-2) (Mitchell et al. 2004; Kumar et al. 2014). NLDAS-2 data is generated from best available observations and model outputs. The data has a spatial and temporal resolution of 0.125-degree (~12-km) and 1 hour, respectively, and is available from 1979 to present in NetCDF format. Each file contains 16 different fields. Noah-MP LSM model coupled to WRF-Hydro model uses 8 out of the 16 fields. These fields include air temperature at 2 m, near surface (10-m) wind speed (horizontal and vertical components), incoming

longwave and shortwave radiation (bias corrected), specific humidity, surface pressure, and precipitation. As explained in (Mitchell et al. 2004), precipitation is derived from temporal disaggregation of gauge-only Climate Prediction Center (CPC) analysis of daily rainfall observations performed directly on the NLDAS grid (12 km by 12 km domain), adjusted for different orography based on the widely applied Parameter elevation Regression on Independent Slopes Model (PRISM) climatology. In this study, the NLDAS-2 meteorological data used to drive the model were spatially downscaled at 1 km resolution using techniques accounting for elevation effects (ADD REFERNCE TO WEBSITE WITH R SCRIPTS). In addition, as explained in section 3.4, precipitation data were further bias corrected using station observations.

2.2.3 GROUND OBSERVATIONS

Observations is necessary for calibration and validation of model simulations. Validation and bias-correction of the NLDAS-2 precipitation forcings was done using ground observations from: (1) two rain gages at daily resolution from the NOAA National Centers for Environmental Information (NCEI) database, including Mund's Park, available from 1986 to 2012, and Sedona, available from 1960 to 2012; and (2) three SNOTEL stations, including Fry (available from 1978 to present), Mormon Mt. (available from 1978 to present), and Mormon Mt. Summit (available from 2008 to present). In addition to precipitation, the SNOTEL stations report temperature, snow water equivalent (SWE), and snow depth. The locations of gages and stations are reported in Fig. 1c. The SNOTEL stations are designed such that it remains functional without maintenance for a year or more. Each SNOTEL sites are equipped with sensors measuring temperature, wind, solar radiation, snow depth and snow pillow which tracks

the amount of water in the snowpack, i.e., Snow Water Equivalent (SWE). Precipitation gages at SNOTEL stations are either tipping bucket or standard storage gages. The gages record all form of precipitation (both liquid and solid precipitation in form of snowfall) throughout the year. The tipping bucket gages are fitted with heating device to melt snowfall into liquid water and total precipitation is then recorded. Precipitation records during winter seasons are usually prone to error such as underestimation of snowfall caused by wind. However, during bias-correction, precipitation records at SNOTEL stations from 2008-2017 was compared against historical record. It was found that the distribution of subset data (2008-2017) is similar to historical data for all the 3 stations and therefore, the data is reliable for bias-correction. The NCEI rain gages are also tipping bucket gages. The procedure adopted to bias correct the NLDAS-2 precipitation forcings is described in section 3.4. Finally, to calibrate and validate the hydrologic simulations, the discharge records at 15-min resolution from two USGS stream gages were used. These are located at the basin outlet and close to the city of Sedona (Fig. 1c) and are available from 1987 and 1986 to present, respectively.

CHAPTER 3

METHODOLOGY

3.1 FLOOD CHARACTERIZATION

The methodology designed for model calibration targets the four main mechanisms of runoff and flood generation. In particular, two main flood regimes were considered, including (1) winter floods caused by rainfall, and (2) summer peaks caused by monsoonal convective thunderstorms. The third mechanism considered is runoff generated as a result of snowmelt in spring. The fourth mechanism accounts instead for the processes leading to baseflow. As an example, Fig. 4a presents the time series of observed discharge at the outlet of OCB in 2008 with the identification of the peaks that corresponds to three mechanisms: snow melt, summer monsoon, and winter rain. Baseflow is low flows with discharge value less than 1 m^3/s and the flow values correspond to percent exceedance > ~40% (Fig. 4c).

Flood peaks are characterized into the four categories based on (1) presence or absence of antecedent rainfall, (2) temperature, and (3) local maxima. Initially, local maximum peaks were identified based on 5% largest peaks. This threshold was chosen as it represents flows greater than 10 m^3/s. The identified peaks were further classified into winter and summer seasons based on the months. The summer months are July, August, September, and October while winter months are November to April. There were no peaks identified in May and June. This is historically true as there was no annual maximum peak recorded at basin outlet from 1941 to 2019 in May and June (Fig. 2b) due to very small rainfall (Fig. 2a). Since monsoon convective thunderstorms occur from July

to October, the identified peaks within these months are classified as monsoon peaks and are shown by red squares in Fig. 4a.

The identified peaks during winter months are further into (1) winter rain and (2) snow melt depending on absence or presence of antecedent rainfall and temperature. The time of an identified peak is used to determine the duration of antecedent rainfall with rainfall hiatus threshold of 2 hours. If the hiatus threshold is exceeded, it is deemed that only precipitation before hiatus threshold exceedance contributed to the flood peak. The total precipitation contributing to the flood peak is calculated as the product of rainfall intensity and duration (sum of hours from time of the peak to the start time of the contributing precipitation). If total precipitation is greater than 0 mm, the corresponding peak is as a result of winter rain. For instance, in Fig. 4a winter flood peaks are shown by red crosses. It is clear from Fig. 4a, that the winter flood peaks are all preceded by precipitation and hence were accurately classified. To further confirm that the peak is due to rainfall in winter, mean temperature 6 hours before the time of the peak was calculated. 6-hour duration was chosen because of daylight during winter is short. If the mean temperature is below 0 ^{o}C, snow melt is not occurring, and the peak is mainly due to winter rain. Note that the melting of snow due to rainfall is not accounted as the contribution of snow melt is assumed to be insignificant compared to runoff generated due to rain.

Run-off generated during late spring (late February to April) is mainly due to melting of snow as temperature increases (Fig. 2a). Similar analysis of winter rain peak was done for snow melt peaks. If precipitation is ~0 mm and mean temperature exceeds 0 ^{o}C, the peak is due to melting of snow. Fig. 4b shows a zoom on snow melt season. An

example of snow melt peaks is shown by blue circle in Fig. 4a. The classification of peaks due to snow melt is accurate because (1) there is no antecedent rainfall for the peaks (Fig. 4b), (2) snow depth corresponding to the identified peaks is decreasing (Fig. 4d), and (3) The discharge is characterized by fluctuations due to snow melt during the day and refreezing at night. This diurnal variation of discharge is shown in Fig. 4b.

Computation of metrics related to winter rain and summer monsoon floods is explained in section 3.6. The discharge values modelled with default WRF-Hydro model parameters shows significant time delay in snow melt (Fig. 7a). Therefore, the computation of bias during snowmelt seasons needs to be accurately computed. To do this, the same time window for observed and simulated discharge must be compared. This was done by determining the time lag that results in maximum correlation coefficient between observed and simulated discharge during snow melt seasons. The time lag is then used to correct the differential lag between observed and simulated discharge. Subsequently, bias is computed. The calibration technique adopted in this study, therefore, focuses on accurate representation of the aforementioned four flood generation mechanisms.

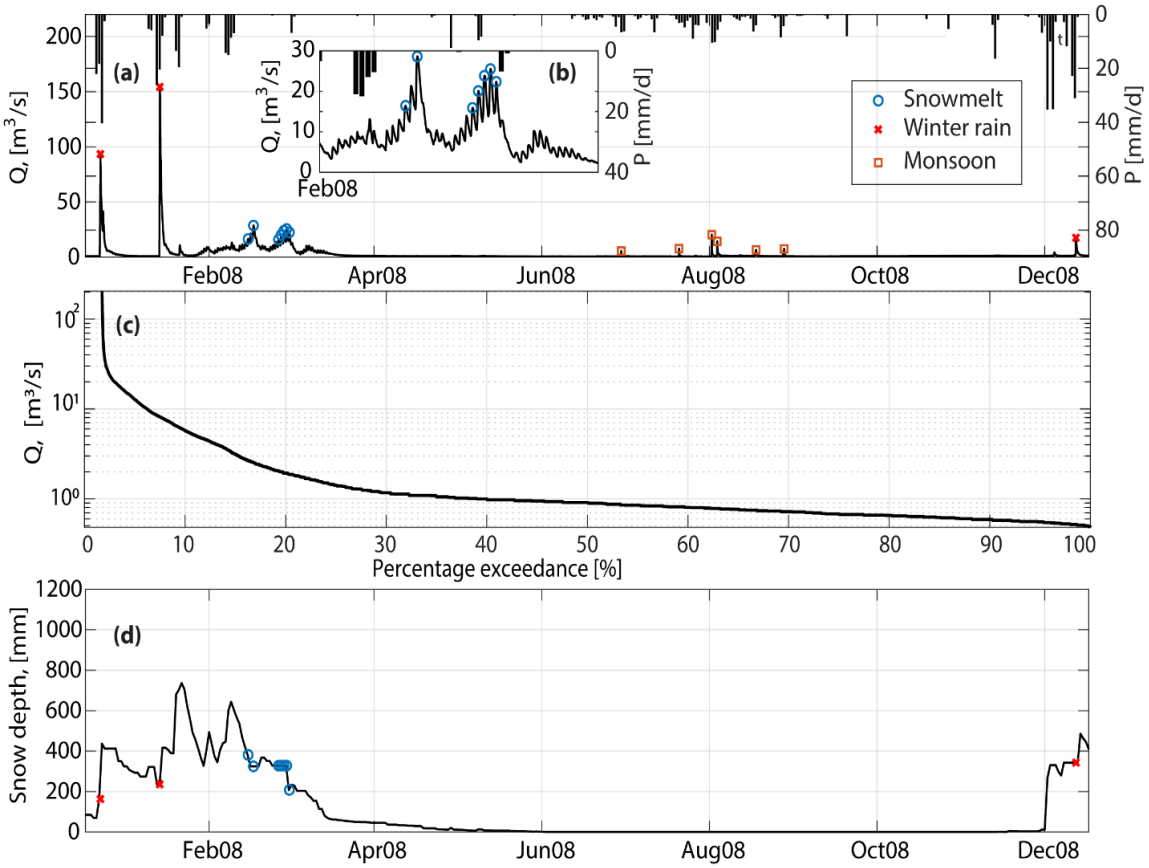

Figure 4: Characterization of streamflow and flood regimes. (a) Time series of observed discharge showing three distinct streamflow and flood regimes, i.e., peaks due to winter rain, summer monsoons, and snow melt. There is also constant baseflow of >0.48 m^3/s. (b) Zoom on snowmelt season showing diurnal variation of discharge due to effect of temperature on snow. Note that the high peaks are not preceded by rainfall showing that the peaks are due to snow melt. (c) Mean snow depth associated with the streamflow and flood regimes. Note that the depth of snow decreases during late spring due to snow melt.

3.2. OVERVIEW OF WRF-HYDRO

WRF-Hydro is a community-based, distributed hydrometeorological modeling system that couples the WRF atmospheric model with modules simulating lateral water movement and aquifer processes (Gochis et al. 2015). In WRF, land surface processes are simulated by a one-dimensional land surface model (LSM) that represents water and energy fluxes along the vertical direction. While multiple LSMs are available as an option in WRF, at the time of this writing, only Noah and its multi-parameterization extension Noah-MP (Niu et al. 2011a) could be selected in WRH-Hydro. Here, Noah-MP was adopted as in the NWM. The switch-activated modules available in WRF-Hydro allow its use in (1) “uncoupled” or “offline” mode as a stand-alone LSM with forcings provided externally, or (2) “fully-coupled” mode where WRF-Hydro model is coupled to weather and climate models. Given the goal of this work, the uncoupled or stand-alone option was used here.

In the following, the critical processes of Noah-MP LSM are first described, including: surface energy balance; runoff generation and infiltration; aquifer recharge; and snowmelt and accumulation. Next, an overview of the enhancements introduced by WRF-Hydro to simulate baseflow, lateral surface and subsurface flow routing, and channel routing is provided. In doing so, a number of governing equations that are useful to explain the calibration strategy adopted in this study were reported. Table 1 presents the physics options selected for the application of Noah-MP.

Table 1**:** Noah-MP selected physics options for this study. The ID numbers refer to the values that can be specified in the model input file.

Processes	**Option ID in WRF-Hydro**	**Description**	**References**
Dynamic vegetation	4	Monthly LAI from MPTABLE.TBL	(Niu et al. 2011a)
Canopy stomatal resistance	1	Ball-Berry	(Ball et al. 1987)
Soil moisture factor	1	Original Noah	(Chen and Dudhia 2001)
Runoff and groundwater	3	Free drainage	-
Supercooled liquid water	1	No iteration	(Niu and Yang 2006)
Frozen soil permeability	1	Hydraulic properties form total soil water & ice	(Niu and Yang 2006)
Surface drag coefficient	1	Original Noah model	(Chen et al. 1997)
Radiation transfer	3	Two-stream applied to vegetated fraction	(Yang and Friedl 2003)
Snow surface albedo	2	CLASS	(Verseghy et al. 1993)
Precipitation partition into rainfall and snowfall	2	BATS scheme	
Lower boundary condition for soil temperature	2	Lowest soil temperature from model inputs	-
Temperature time scheme	1	Semi-implicit	

3.2.1. SURFACE ENERGY BALANCE

Noah-MP uses a mosaicked sub-grid scheme to represent surface bare soil and vegetated areas, whose fraction can be dynamically updated based on remote-sensing observations. The surface energy balance is calculated separately for each part through a two-stream approximation scheme (Niu and Yang 2004). Canopy transpiration and evaporation from bare soil and canopy leaves are then estimated from the resulting latent heat fluxes. In this study, the dynamic vegetation option with properties derived from climatological monthly means of leaf area index (LAI) provided as lookup tables was used (Table 1).

3.2.2. RUNOFF AND INFILTRATION

The soil column is represented through four layers of different thickness extending up to 2 meters (0.1 m, 0.3 m, 0.6 m, and 1 m thickness for 1st, 2nd, 3rd, and 4th layers, respectively). Depending on soil moisture state and precipitation intensities, runoff can be generated and/or water can infiltrate through the soil layers and percolate to an underlying unconfined aquifer. Four options are available to simulate runoff generation and groundwater. Here, the infiltration-excess-based surface runoff scheme proposed by Schaake et al. (1996) and Chen and Dudhia (2001) was used. In this scheme, the control volume is the soil surface, and therefore, surface runoff, *R*, is equal to the difference between precipitation, *P*, and maximum infiltration, I_{max}:

$$R = P - I_{max}. \tag{1}$$

I_{max}, in turn, is given by:

$$I_{max} = P_x \cdot \left\{ \frac{D_{total}\left[1-\exp\left(-\frac{DKSAT \cdot REFKDT \cdot \Delta t_1}{REFDK}\right)\right]}{P_x + D_{total}\left[1-\exp\left(-\frac{DKSAT \cdot REFKDT \cdot \Delta t_1}{REFDK}\right)\right]} \right\}, \tag{2}$$

where Δt_1 is the model time step in days; *DKSAT* is the saturated hydraulic soil conductivity, depending on soil texture; *REFDK* is a reference saturated hydraulic conductivity of silty clay loam, assumed spatially constant and equal to 2×10^{-6} m/s; and *REFKDT* is the surface runoff parameter. Note that if silty clay loam soil type exists within a study area and during calibration, *DKSAT* value is changed for the soil classes, REFDK must be changed accordingly. Since there is no silty clay loam soil type within this study basin, default value of *REFDK* is appropriate. Being function of precipitation inputs and soil properties (see equation (2)), I_{max} is spatially varying. P_x is excess precipitation or throughfall from canopy and is given by:

$$P_x = \max(0, Q_{wat} \cdot \Delta t), \quad (3)$$

where Q_{wat} is the water input to the soil surface, and Δt is the model time step in hours (i.e., $\Delta t = \Delta t_1 \cdot 86400$). Q_{wat} is calculated differently depending on the existence of a snow layer (see below) and it accounts for rainwater, melting water from the bottom of the snowpack, soil surface dew rate adjusted for frost. Finally, the term D_{total} in equation (2) is the total soil moisture content that can potentially infiltrate, which depends on soil properties:

$$D_{total} = \sum_{k=1}^{N} \Delta z_k \cdot (SMCMAX - SMC_k), \quad (4)$$

where Δz_k and SMC_k are the thickness and volumetric soil moisture content of the k-th soil layer ($k = 1, \ldots, N = 4$), respectively; and *SMCMAX* is the saturated volumetric soil moisture content dependent on soil type. Equation (4) is modified to account for the presence of ice in the soil.

From equations (1)-(4), surface runoff, R_s, is computed as the maximum between zero and the difference between the available water rate and the maximum infiltration corrected for soil impermeability to frozen conditions:

$$R_s = \max\left(0, (Q_{wat} - E) - I_{max} \cdot F_{cr}\right) \quad (5)$$

where E is the total evaporation rate (sum of evaporation from bare soil and canopy leaves, canopy transpiration, and snow sublimation); F_{cr} is the impermeable fraction due to frozen soil; and I_{max} and Q_{wat} have been previously defined.

3.2.3. AQUIFER RECHARGE

The infiltration into the soil column recharges the underlying unconfined aquifer with a flow given by:

$$Q_{bot} = SLOPE \cdot DKSAT \cdot \left[\max\left(0.01, \frac{SMC_4}{SMCMAX}\right)\right]^{2 \cdot BEXP + 3} \cdot (1 - F_{cr}) \quad (6)$$

where *BEXP* is the Clapp-Hornberger exponent, SMC_4 is the soil moisture content in the bottom soil layer (4th layer), and *SLOPE* is the gravitational slope coefficient. The dynamics of water storage in the aquifer is the difference between aquifer recharge (equation (6)) and baseflow (equation (8)). Equation (6) combines the aquifer recharge and subsurface runoff; therefore, it governs the interflow processes between soil column and underlying aquifer.

3.2.4. SNOW DYNAMICS

On top of the soil, up to three layers of snowpack can accumulate. Noah-MP computes snow and soil skin temperatures based on mass and energy balance in the snowpack and uses them to determine the amount of snow melting or freezing in each snow layer. In doing so, an important variable affecting runoff generation is the snow

cover fraction, *SCF*, which is empirically computed based on snow depth, h_{snow}, ground roughness length, z_0, and snow density, r_{snow}:

$$SCF = \tanh\left(\frac{h_{snow}}{2.5z_o\left(\frac{\rho_{snow}}{\rho_{new}}\right)^{MFSNO}}\right), \quad (7)$$

where *MFSNO* is the snow melting factor parameter, r_{new} is the density of new snow ($\sim 100\,\frac{kg}{m^3}$). To calculate the liquid water from the snowpack bottom to soil, SWE and snow ice content are computed in each snowpack layer. Finally, snow canopy interception and subsequent throughfall, melting and sublimation are simulated as a function of meteorological forcings and vegetation properties, including the Leaf Area Index (LAI).

3.2.5. TERRESTRIAL HYDROLOGIC ROUTING

The main improvements introduced by WRF-Hydro are multiple physics options for subsurface, surface overland flow, channel, and lake/reservoir routing (Yucel et al. 2015; Gochis et al. 2018). To implement these new capabilities, WRF-Hydro operates on a multiresolution, gridded nested domain. A coarser grid (here, 1-km resolution) is used to simulate land surface fluxes with Noah-MP. Moisture fluxes (runoff and infiltrated water) are then routed laterally in a higher-resolution terrain grid (here, 250-m resolution) that allows capturing changes in elevation responsible for gravitational water redistribution. In WRF-Hydro, the subsurface flow module computes water table changes using Dupuit-Forcheimer assumptions, which calculates hydraulic gradient as the differences in groundwater table depth along the steepest direction in eight possible directions (D8) around each routing grid (Gochis et al. 2015). Routing is first computed for the subsurface saturated flow using a quasi-three-dimensional flow representation that

accounts for the effects of topography, saturated soil depth, and saturated hydraulic conductivity. Depending on local conditions (e.g., if subsurface flow causes a grid to become saturated), exfiltration may occur in soil columns and is added to infiltration excess runoff, if present. Overland flow is calculated utilizing a diffusive wave formulation similar to Julien et al. (1995) and Ogden (1997), which can be optionally applied in one or two dimensions; here, the one-dimensional option is chosen. Finally, routing in the channel network can be computed via a suite of possible algorithms that operate either on the high-resolution grid or on a vectorized network of channel reaches. In this study, an explicit, one-dimensional diffusive wave solution of the St. Venant equation applied on the gridded channel network was used.

3.2.6. BASEFLOW

Another enhancement available in WRF-Hydro is a conceptual groundwater model for the simulation of baseflow (Gochis et al. 2018). The scheme assigns a bucket to each sub-basin (defined by the user). Water percolating from the soil column recharges the bucket and the following equations control the outflow released to the channel at the sub-basin outlet (i.e., the baseflow), Q_{out}:

$$Q_{out} = \begin{cases} C_o \cdot \exp\left(\alpha \frac{z}{z_{max}} - 1\right), & z \leq z_{max} \\ \frac{z - z_{max}}{A \cdot \Delta t}, & z > z_{max} \end{cases}, \quad (8)$$

where C_o and a are the bucket parameters, z is the water depth in the bucket at a given time step, z_{max} is the maximum depth of the bucket before water is spilled, A is the sub-basin area, and Dt is the model time step.

3.3. MODEL SETUP IN THE OAK CREEK BASIN

The WRF-Hydro domain was set up by defining a rectangular region that entirely includes the OCB. For this region, the maps of elevation, soil properties and land cover classes were derived at 1-km spatial resolution. These are shown in Fig. 1c and Fig. 3a, b, while the sources of the geospatial data are described in section 2.2.1. A set of grids used by LSM with land surface properties such as soil and vegetation types, green vegetation fraction/LAI, latitude and longitude were created and stored in a NetCDF file named "geo_em.nc". A grid with spatially distributed land surface model parameters (with hardcoded name "soil_properties.nc") was also created and modified during the calibration process. This allows for modification of LSM parameters efficiently instead of changing parameters for each class in the corresponding look-up table. To apply the routing schemes, 250-m terrain grids with elevation, stream order, channel grid, and flow direction were obtained and stored in NetCDF file named "fulldom.nc". In addition, grids defining the groundwater sub-basins and associated properties (see section 3.1.6) were derived.

The model code is provided with a set of several text files and look-up tables containing model parameters with their default values; all these files have hardcoded names, as described in detail by (Gochis et al. 2018). Hydrometeorological forcings included 1-km grids matching the coarse-resolution domain where Noah-MP is applied of incoming long and short-wave radiation, specific humidity air temperature at 2 m, surface pressure, near-surface (10-m) wind speed, and precipitation (Senatore et al. 2015; Gochis et al. 2015). As mentioned in Section 2.2, forcings were obtained from downscaled

NLDAS-2 data, with precipitation bias corrected through rain gage observations. All these grids are stored in NetCDF files.

3.4. PRECIPITATION BIAS CORRECTION

Precipitation records observed at two rain gages and three SNOTEL stations (Fig. 1c) were used to bias correct the NLDAS forcings. For each station, the climatological monthly means were computed using the entire available record; in doing so, a monthly observation was included if at least 27 days were non-missing. This resulted in records of at least 10 years for each station. The existence of a linear relationship between climatological monthly precipitation means and elevation for the selected stations was then investigated. It was found that the slope of the trendline is (1) positive in all months, and (2) statistically significantly larger than zero (p-value<0.05) in all months except for March, April, June and September. The p-value of those four months are 0.33, 0.36,0.93, and 0.26 respectively and the slope of regression line in these months is flat compared to months with statistically significant p-value (Fig. C1). Results for these four months are conditioned by (1) the fact that they are relatively drier (Fig. 2a); and (2) the climatological means at the highest gage (in terms of elevation) at Mormon Mt. Summit are based on the shortest record length (10 years) relative to other stations. Based on these considerations, the existence of a linear trend between precipitation and elevation was assumed in all months.

As a next step, the same analyses were performed for the NLDAS pixels containing the five stations for the period 2008-2017. It was found that the slope of the trend line is positive in all months and statistically significantly different from zero, but the magnitude is much smaller than the values found for the gages. As an example, Figs.

5a,b present these relations for the months of January and July respectively, which are representative months for the two distinct storm types shown in Fig. 2a.

Based on these results, the monthly NLDAS precipitation forcings was bias corrected through two multiplicative factors, accounting for (1) the difference between gage observations and co-located NLDAS estimates, and (2) the effect of elevation throughout the basin. If $P_{o,y,m}(z)$ is the original NLDAS precipitation for month m of year y in a pixel at elevation z, then the bias corrected value, $P_{y,m}(z)$, was calculated as:

$$P_{y,m}(z) = P_{o,y,m}(z) \cdot \omega_{y,m} \cdot \lambda_m(z). \quad (9)$$

In equation 9, $\omega_{y,m}$ is defined as:

$$\omega_{y,m} = \frac{1}{S}\sum_{i=1}^{S} \frac{P_{g,i,y,m}}{P_{NLDAS,i,y,m}}, \quad (10)$$

where $P_{g,i,y,m}$ is the monthly precipitation observed at the i-th station, $P_{NLDAS,i,y,m}$ is the monthly precipitation provided by NLDAS in the co-located pixel, and S is the number of available stations on month m of year y. The factor λ_m in equation (9) is defined as:

$$\lambda_m(z) = \frac{a_m + b_m z}{a_m + b_m \bar{z}}, \quad (11)$$

where a_m and b_m are the parameters of the linear regression between the climatological means of monthly precipitation observed at the stations and elevation for month m (m = 1, …, 12), and $\bar{z}$ is the mean basin elevation as provided by NLDAS. The values of a_m and b_m vary across the months and are reported in Table 2. As an example, the original and bias corrected mean areal precipitation values for all months of 2008 are presented in Fig. 5c. Our methodology provides a monthly adjustment factor for elevation of about 0.25 km^{-1}, which is included in the ranges reported in literature (see, e.g., Liston and

Elder, 2006). After computing $P_{y,m}(z)$, the bias corrected precipitation for a given day, d, and hour, h, $P_{y,m,d,h}(z)$, was calculated as:

$$P_{y,m,d,h}(z) = P_{0,y,m,d,h} + \frac{P_{y,m}(z) - P_{0,y,m}(z)}{N_{h,y,m}}, \quad (12)$$

where $N_{h,y,m}$ is the number of hours with nonzero precipitation occurred during that month.

Table 2**:** Slope and intercept of the regression line between monthly precipitation and elevation. Note that these regressions should be applied only in the range of elevation from ~1100 to ~2500 m a.s.l. that was used to estimate the coefficients.

Month,	a_m **(mm)**	b_m **(10^{-3} mm/m)**
Jan	-0.973	40.484
Feb	1.021	39.190
Mar	37.235	15.216
Apr	21.701	6.333
May	3.124	7.594
Jun	8.858	-0.198
Jul	12.933	23.060
Aug	5.899	35.008
Sep	42.331	6.387
Oct	21.709	12.329
Nov	7.804	20.302
Dec	-40.038	56.893

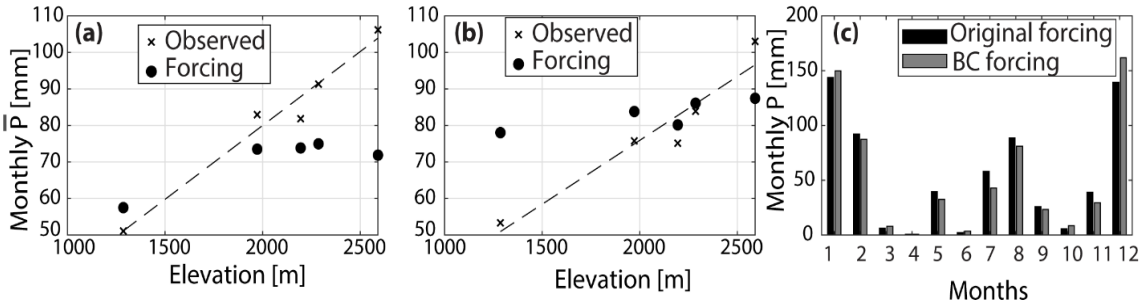

Figure 5**:** (a)-(b) Mean monthly precipitation for (a) January and (b) July 2008 from gages and NLDAS pixels. (c) Monthly mean areal precipitation for 2008 before and after bias correction.

3.5. MODEL CALIBRATION TARGETING DISTINCT HYDROLOGICAL PROCESSES

WRF-Hydro model was calibrated against discharge observations at the basin outlet during the period 2008-2011. The validation was conducted against multiple datasets, including: (1) discharge observations at the outlet from 2012 to 2017; (2) discharge observations at the interior stream gage in Sedona from 2008 to 2017; and (3) mean basin snow depth and SWE derived from the NOAA Snow Data Assimilation System (SNODAS) products (NOHRSC, 2004). To perform the calibration, a stepwise approach targeting the winter and summer floods, snow melt, and baseflow was implemented. This approach focuses on the key processes governing the hydrologic response of the study basin identified in section 3.1. Sensitivity tests was conducted and the results of previous studies (Cuntz et al. 2016a; Kerandi et al. 2018; Li et al. 2018; Lahmers et al. 2019a) was used to identify the model parameters that exert the major control on those processes. The parameters that were initially considered are listed in Table 3, where they are grouped into those related to soil, vegetation, groundwater, surface routing, and snow. For each parameter, the table presents the name used in WRF-Hydro, the description, the main hydrologic processes affected by the parameter, a representative equation where the parameter is used, and the range of values. The sensitivity tests were conducted by varying each parameter within a range that is physically plausible or has been proposed in previous studies. Depending on the corresponding changes in simulated discharge, the sensitivities of each parameter on the streamflow and flood generation mechanisms was quantified using the metrics described

in section 3.4. In this study, default values of the hard-corded parameters who sensitivities was described in (Cuntz et al. 2016b) were used. Some of the model parameters such as soil hydraulic conductivity (*DKSAT*) and porosity (*SMCMAX*) control more than one flood generation mechanisms (e.g., aquifer recharge and rainfall-runoff generation) and therefore, such parameters require special consideration during calibration. Vegetation parameters also affects water and energy balance hence requires close analysis. The calibration was focused on parameters with high and moderate sensitivities based on (1) effect of the parameter on hydrograph volume, and (2) effect of the parameter on the timing of the peaks. This is quantified through metrics such as mean relative error (MRE) and time lag (t_{lag}) of the largest winter and summer peaks. For those parameters with low sensitivity, default or calibrated values were used.

Once the dominant parameters were identified, calibration of their values through an approach based on five steps, described in the schematic diagram of Fig. 6 was performed. In order, the four steps target the calibration of parameters affecting: (1) winter and summer flood peaks; (2) baseflow component; (3) hydrograph shape; and (4) snow dynamics. The fifth step consists of a final minor tuning to further improve the performance metrics. The calibrated values of the parameters were then used to validate the model from 2012 to 2017. These sequential steps presented in Fig. 6 were chosen primarily based (1) the severity of the shortcomings of default model parameters described in section 4.2, and (2) the fact that soil parameters such as *DKSAT* and *SMCMAX* exerts control on various processes of energy and water fluxes.

The WRF-Hydro source code is distributed with input files that contain default values for each parameter (Gochis et al. 2018). These include a text file presenting global parameters of Noah-MP (GENPARM.TBL); look-up tables reporting soil and vegetation parameters for each class (SOILPARM.TBL and VEGPARM.TBL); and text files listing parameters controlling overland flow routing (HYDRO.TBL), channel routing (CHANPARM.TBL), and the groundwater bucket models (GWBUCKPARM.TBL). Parameters of Fig. 6 were modified changing their values in the corresponding input files. If their value is constant for the entire basin, their value is replaced with calibrated value and multipliers are applied to the default value stored in the NetCDF file "soil_properties.nc", if the parameter values are spatially distributed. The schematic diagram that summarizes the calibration steps is shown in Fig. 6.

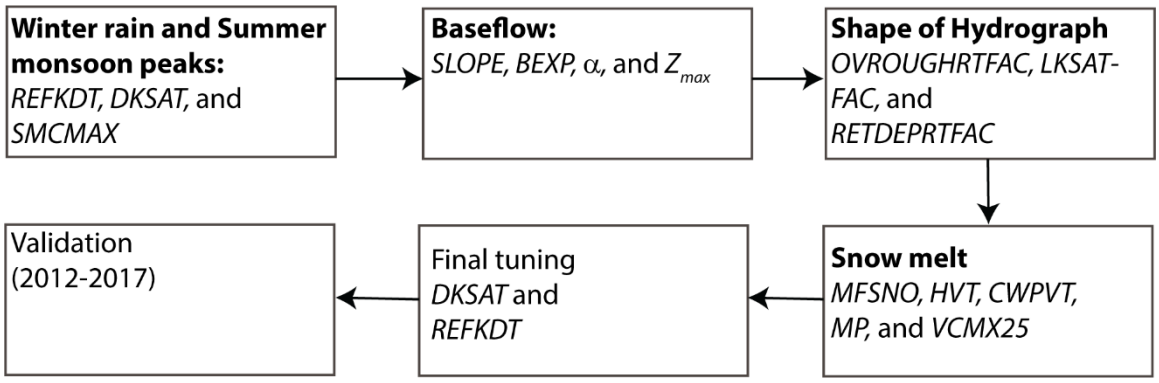

Figure 6: Schematic of calibration steps showing the runoff generation processes and parameters that control each process. Note that vegetation parameters such as *HVT*, *MP*, *CWPVT*, and *VCMX25* also affects winter and summer peaks beside snow melt.

Table 3: WRF-Hydro parameters considered for calibration for this study with the corresponding ranges based on previous publications. For some of the parameters, the governing equation reported in Section 3.1 has been indicated. The main hydrological processes controlled by each parameter is also indicated. Based on these, calibration of parameter can be tailored for each flood generation mechanisms.

Parameters (units)	Description	Main control on hydrologic response	Equation	Range
		Soil parameters		
BEXP (unitless)	Clapp-Hornberger B exponent	Soil water diffusivity	(6)	Dependent on soil texture
SMCMAX (m^3/m^3)	Saturated soil moisture	Infiltration	(4)-(5)	Dependent on soil texture
DKSAT (m/s)	Saturated soil hydraulic conductivity	Infiltration	(2)	Dependent on soil texture
		Runoff parameters		
REFKDT (unitless)	Parameter in surface run-off	Partitioning of total runoff into surface and subsurface runoff	(2)	0.1-10
LKSATFAC (unitless)	Multiplier on Saturated soil lateral conductivity	Routing/Interflow process	-	10-10,000
RETDEPRTFAC (unitless)	Maximum retention depth	Routing/Interflow process	-	0.1-10
SLOPE (unitless)	Slope index	Aquifer recharge	(6)	0.1-1

Continue on next page

		Vegetation parameters		
MP (unitless)	Slope of conductance to photosynthesis relationship	Transpiration	-	3-16
CWPVT (m^{-1})	Canopy wind parameter for canopy wind profile formulation	Snow/precipitation interception	-	
HVT (m)	Height of top vegetation canopy	Snow interception	-	0.25-1.5
VCMX25 ($\frac{\mu mol\ CO_2}{m^2/s}$)	Maximum rate of carboxylation at 25℃	Transpiration		10-100
		Groundwater parameters		
a (unitless)		Baseflow	(8)	1-8
z_{max} (mm)	Parameters of the GW model		(8)	Varies depending of depth of water table
C_0 (m^3/s)			(8)	varies
		Routing parameters		
OVROUG HRTFAC (unitless)	Manning's roughness coefficient	Routing/Interflow process	-	0-1
		Snow parameters		
MFSNO (unitless)	Snowmelt parameter	Snow ablation	(7)	0.5-4

Table 3 continued

3.6 METRICS

Model performances during the calibration and validation periods were visualized by comparing simulated and observed hydrographs and flow duration curves (FDCs). In addition, metrics were computed from the simulated and observed hydrographs at daily resolution, including Nash-Sutcliffe Efficiency coefficient (NSE), Root Mean Square Error (RMSE), bias, percentage bias, and correlation coefficient (CC). These metrics, which are widely used to assess hydrologic models (Biondi et al. 2012), are defined as:

$$NSE = 1 - \frac{\sum_{t=1}^{N}\left(Q_{s,t} - Q_{o,t}\right)^2}{\sum_{t=1}^{N}\left(Q_{o,t} - \overline{Q_o}\right)^2} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{t=1}^{N}\left(Q_{s,t} - Q_{o,t}\right)^2} \quad (14)$$

$$bias = \overline{Q_s} - \overline{Q_o} \quad (15)$$

$$CC = \frac{\frac{1}{N-1} \cdot \sum_{t=1}^{N}\left[\left(Q_{s,t} - \overline{Q_s}\right) \cdot \left(Q_{o,t} - \overline{Q_o}\right)\right]}{\left[\frac{1}{N-1} \cdot \sum_{t=1}^{N}\left(Q_{o,t} - \overline{Q_o}\right)^2\right]^{0.5} \cdot \left[\frac{1}{N-1} \cdot \sum_{t=1}^{N}\left(Q_{s,t} - \overline{Q_s}\right)^2\right]^{0.5}} \quad (16)$$

where N is the sample size; $Q_{s,t}$ and $Q_{o,t}$ are the simulated and observed discharge at a given time t; and $\overline{Q_s} = \sum_{t=1}^{N} Q_{s,t}$ and $\overline{Q_o} = \sum_{t=1}^{N} Q_{o,t}$ are the corresponding time averages.

To further quantify the model performance in the simulation of winter and summer flood peaks, the Mean Relative Error (MRE) and Time Lag (t_{lag}) between observed and simulated peaks were calculated.

These are defined as:

$$MRE = \frac{1}{M}\sum_{i=0}^{M}\left(\frac{Q_{s,i}-Q_{o,i}}{Q_{o,i}}\right) \times 100 \quad (17)$$

$$T_{lag} = \frac{1}{M}\sum_{i=0}^{M}(t_{sim,i} - t_{obs,i}), \quad (18)$$

where $Q_{s,i}$ is the simulated *i*-th peak, $Q_{o,i}$ is the corresponding *i*-th peak, $t_{sim,i}$ is the peak time of simulated *i*-th peak, and , $t_{obs,i}$ is the peak time of observed *i*-th peak.

CHAPTER 4

RESULTS

4.1 MODEL SPIN -UP

Prior to performing the calibration, a spin-up simulation of 10 years was conducted to achieve the numerical stability of the model outputs. A "restart" NetCDF files containing appropriate values of the state variables was generated after the spin-up and utilized to initialize the simulations on 1 January 2008 (Fig. A1). The state variables for snow related parameters such as SCF, SWE, and snow depth both with and without spin-up is shown in Fig A2. For the spin-up simulation, the default parameter values were used; since these are not the final parameters, the spin-up experiment was repeated with the calibrated parameters and it was found that the use of restart files based on the default parameters did not significantly affect the final results.

4.2. HYDROLOGIC SIMULATION WITH DEFAULT PARAMETER VALUES

Fig. 7a presents the time series of observed discharge at the basin outlet and the flow values simulated with the default parameters (labelled as Run 1). The corresponding FDC is reported in Figs. 7b, c; specifically, Fig. 7b emphasizes low and intermediate flow occurring in the dry and wet seasons, respectively, while Fig. 7c visualizes the portion of FDC accounting for the flood peaks. The use of default values for the model parameters leads to an underestimation of low flow values (flow values with percent exceedance > ~40%) , overestimation of intermediate flow values, poor simulation of winter rain and summer monsoon peaks (significantly underestimated, Table 5), and significant delay in snow melt (Figs. 6 a-c). As reported in Table 4, this is quantified by a very large positive bias (+2.52 Mm^3 or ~68 %), very low correlation coefficient (0.34), and low NSE (-0.64)

on the 4-year simulation. To improve these results, the first step of the calibration was aimed at calibrating the parameters that control surface runoff so as to improve simulation of winter rain and summer monsoon peaks.

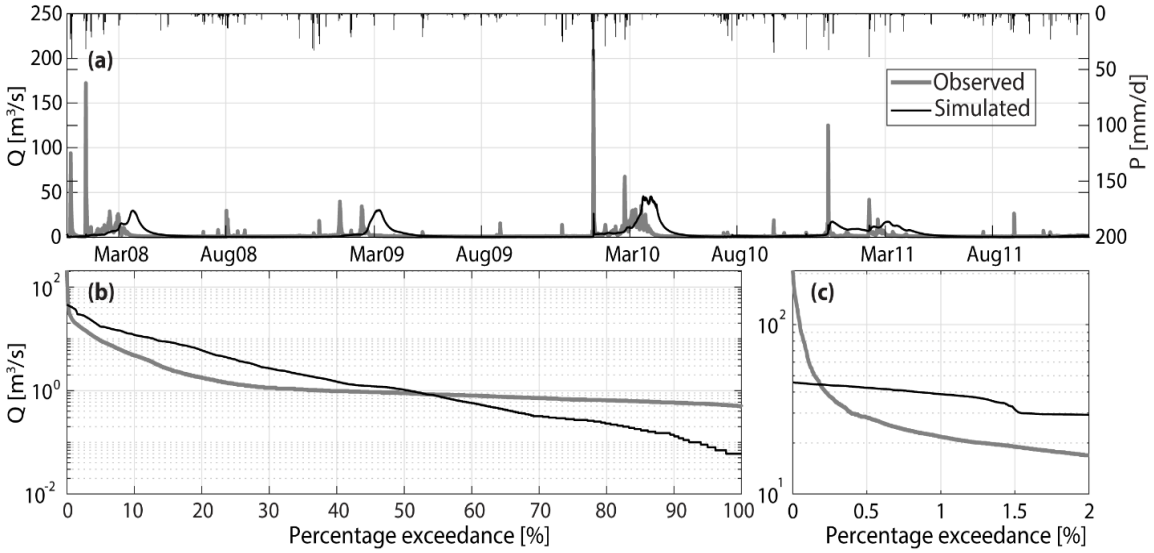

Figure 7**:** (a) Time series of observed discharge and simulated discharge using the default parameters for the period 2008-2011. Simulated discharge using default parameters shows significant time lag during snow melt seasons, poor simulation of winter rain and summer monsoon peaks, and under-simulation of baseflow. (b) Flow duration curve (FDC) of observed and simulated discharge for the entire time series. (c) Zoom on the portion of FDC accounting for the flood peaks.

4.3. STEP 1: CALIBRATION OF WINTER AND SUMMER FLOODS

The first step of the physics-based, manual calibration was focused on modifying the parameters affecting flood generation, timing and peaks due to winter and summer storms. This effort is described in the following sub-sections.

4.3.1 EFFECT OF *DKSAT* AND *REFKDT*

As shown by section 3.1.2, two key parameters affecting surface runoff generation are *DKSAT* and *REFKDT*. In particular, both parameters control the maximum infiltration and, in turn, surface runoff (see equations (1)-(2)): lower values for both parameters lead to decrease in infiltration capacity of the soil column, which in turn increases surface runoff. A sensitivity analysis for *DKSAT* and *REFKDT* was performed. *DKSAT* is spatially variable and multiplicative factors in the range from 0.3 to 2 (1 is the value used for the default case) were tested. This range of values was chosen based on physical range for the dominant soil type (loam) with the basin as reported in (Latorre and Moret-Fernández 2019; Rawls et al. 1983). *REFKDT* was assumed spatially constant as in previous applications, with a default value of 1.0 and a suggested range from 0.1 to 10 (Kerandi et al. 2018). Here, the calibration was performed with *REFKDT* ranging from 0.1 to 5 with 0.2 increment. Values above 5 were not tested, since they led to the negligible runoff generation.

Results of the combinations of *DKSAT* and *REFKDT* are summarized in Fig. 8, which shows time series and FDCs of observed and simulated discharge for the best and worst combinations of *REFKDT* and *DKSAT* values. *DKSAT* and *REFKDT* values of 0.3 results in: (1) improved representation of winter peaks (Figs. 8a,b, blue line) with MRE changing from changing from -83.1% in default (Run 1) to 23.2% and t_{lag} changing from

8.4 hours in default (Run 1) to 4.6 hours (Table 5; Run 2 best); and (2) slightly improved simulation of summer peaks (Figs. 8a,c, blue line) with MRE changing from -97.6% to -89.8% and t_{lag} changing from -10.4 hours to 3 hours (Table 5). Higher values of *REFKDT* and *DKSAT* of 5 and 2 respectively lead to significant underestimation of both summer and winter peaks, as also shown through the FDCs of Figs. 8d, e and the worse values of MRE and t_{lag} (Table 5; Run 2 worst). The optimal value of 0.3 was adopted for *REFKDT* and the multiplicative factor of *DKSAT* (henceforth labelled as Run 2). The metrics applied to the entire time series for this run are presented in Table 4. It is worth noting that this preliminary value of *DKSAT* will be further tuned for better model performance in case of over-simulation of flood peaks.

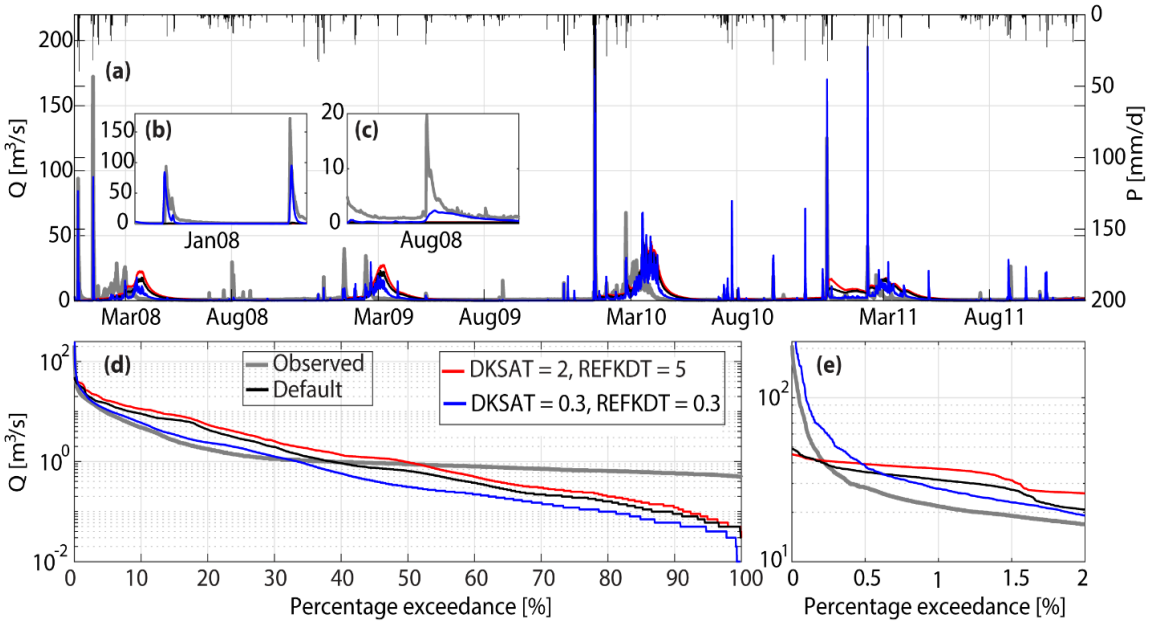

Figure 8**.** (a) Time series of observed discharge and simulated discharge for the calibration period. (b)-(c) Zoom on two winter rain peaks in (b) January and (c) July 2008. (d) FDCs of observed and simulated discharge for the entire time series. (e) Zoom on the portion of FDC accounting for floods.

Table 4. Metrics quantifying model performance during calibration and validation periods. All results except for the validation at Sedona are referred to the discharge at the basin outlet at Cornville. The metrics are computed at daily resolution. The asterisk (*) shows the simulations where calibration worsened the metrics

Runs	**Description**	**NSE**	**RMSE [m^3/s]**	**Bias [Mm^3] (%)**	**CC**	**Reference**
Run 1	Default	-0.64	6.06	2.52 (67.53)	0.34	Fig. 7
Run 2	*REFKDT* = 0.3; *DKSAT* = 0.3; rest default	-0.26	5.92	1.08 (23.27)	0.62	Fig. 8 black line
Run 3	*REFKDT* = 0.3, *DKSAT* = 0.3, *SMCMAX* = 0.75; rest Default	-2.03*	7.78*	2.39* (73.36)	0.65	Fig. 9 red line
Run 4	As Run 3; *BEXP* = 1.89; *RETDEPRTFAC* = 0, *SLOPE* = 0.75; Original baseflow parameters; rest default	-0.18	4.52	3.15* (96.74)	0.66	Fig. 10 black line
Run 5	As Run 3; C_o = 1.30, α = 3.18, z_{max} = 2000 mm, rest Default	0.21	4.46	0.56 (16.68)	0.67	Fig. 10 red line
Run 6	As Run 5; *MFSNO* = 3.5	0.41	2.54	0.54 (11.52)	0.70	Fig. 11 Blue line
Run 7	As Run 5; DKSAT = 0.48; *OVROUGHRTFAC* = 0.9,	0.56	2.36	0.35 (8.10)	0.75	Fig. 13
Run 8	As Run 7 for winter seasons; *DKSAT* = 0.3; *REFKDT* = 0.1 for summer seasons,	0.62	2.28	0.27 (7.29)	0.86	Fig 14
	Validation steps					
Run 9	Validation at the outlet	0.62	2.55	0.86 (9.82)	0.77	Fig. 15
Run 9	Validation at Sedona	0.46	3.17	1.02 (15.97)	0.64	Fig 16

Table 5: Metrics quantifying the model performance in the simulation of winter and summer flood, as well as of runoff due to snow melt. (*) indicates a simulation where the metric worsened when compared to the previous Run. Note that the simulation runs that does not significantly impact MRE, t_{lag}, and snow melt bias are not shown in this table.

		Winter peaks		**Summer peaks**		**Snow melt**
Runs	**Parameters**	**MRE [%]**	**t_{lag} [hrs]**	**MRE [%]**	**t_{lag} [hrs]**	**Bias [$10^3 m^3$]**
Run 1	All default	-83.1	8.4	-97.6	-10.4	9.3
Run2 (best)	*REFKDT* = 0.3, *DKSAT* = 0.3, rest default	23.2	4.6	-89.8	3.0	3.4
Run 2 (worst)	*REFKDT* = 5, *DKSAT* =2, rest default	-93.0	7.5	-98.0	-12	14.3
Run 3 (best)	As of Run 2 (best), *SMCMAX* = 0.75	18.3	3.0	-83.8	2.4	11.6*
Run 3 (worst)	As of Run 2 (best), *SMCMAX* = 1.5	-93.0	4.4	-96.9	-1.3	7.8
Run 7	*REFKDT* = 0.3, *DKSAT* = 0.48 for both summer and winter	12.2	2.6	-84.0	2.4	0.86
Run 8	As of Run 7 for winter*;* *REFKDT* = 0.1, *DKSAT* = 0.3 for summer	12.2	2.6	-12.0	1.8	0.86
Run 9 (outlet)	As of Run 8 (2012-2017)	12.3	3.0	-13.4	2.9	1.02

4.3.2. EFFECT OF *SMCMAX*

As a next step, a sensitivity analysis on *SMCMAX* (soil porosity) was performed. By regulating the volume of water that can be stored in the soil, this parameter influences *Dtotal* in equation (4), and, in turn, the soil infiltration capacity and the generated runoff (equations (1) and (2)). The sensitivity simulations were conducted using optimal multiplicative values of *DKSAT* and *REFKDT* from Run 2. Fig. 9a shows observed and simulated discharge for three representative multipliers applied to *SMCMAX*. A smaller (higher) value of *SMCMAX* of 0.75 (1.5) results in better (worse) simulation of summer and winter peaks, as showed by the red (blue) lines in Figs 9a-c). Using *SMCMAX* of 0.75 (hereafter, Run 3), both MRE and t_{lag} of winter and summer peaks slightly improved (i.e., their absolve value slightly decreased, as reported in Table 5). The improvement in MRE and t_{lag} due to the decrease in *SMCMAX* can be explained considering that runoff is generated and routed quickly because of the decrease in the "wetness" capacity of the soil required to generate runoff.

Despite a better simulation of flood peaks, the bias during snow melt increased from 3.4×10^3 to 11.6×10^3 m^3 (Table 5). Time lag also improved (Table 5). This result can be explained considering that a lower *SMCMAX* reduces infiltration of water from the snowpack; this, in turn, results in larger runoff volume.

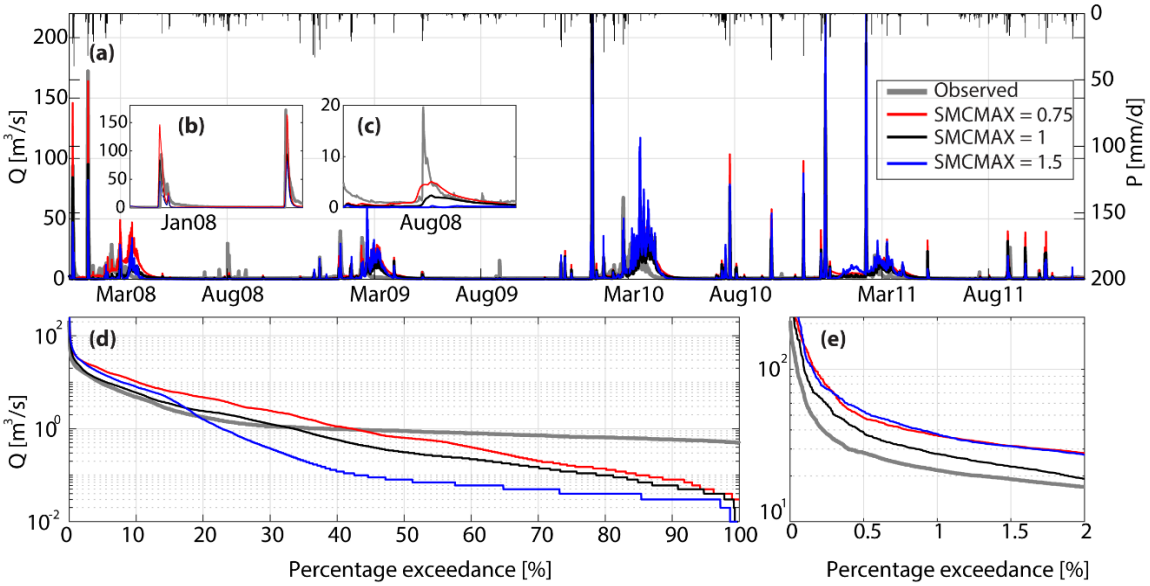

Figure 9: (a) Time series of observed discharge and simulated discharge for the calibration period. (b) Zoom on two winter rain peaks in January of 2008, shows better simulation of peaks using small multiplicative values of *SMCMAX* = 0.75 (red line) while higher values of *SMCMAX* = 1.5 results in relatively poor simulation of the peaks (blue line) (c) Zoom on a summer monsoon peak in August of 2008, simulation of peaks using small multiplicative values of *SMCMAX* = 0.75 (red line) while higher values of *SMCMAX* = 1.5 results in relatively poor simulation of the peaks (blue line) (d) FDC of observed and simulated discharge for the entire time series. (e) Zoom on the portion of FDC accounting for the floods**.**

4.4 STEP 2: CALIBRATION OF BASEFLOW

Results presented so far indicate a significant underestimation of the baseflow contribution. The next calibration step was to modify the parameters affecting this process. The attention was initially focused on *BEXP* and *SLOPE*, which influence percolation from the 4th layer of soil to the underlying aquifer, as shown in equation (6). *SLOPE* parameter is assumed spatially constant – a *SLOPE* value of 0 indicates no aquifer recharge, while higher values up to 1 (Niu et al. 2011b) increase the recharge into the aquifer and, ultimately, the baseflow contribution. Similarly, higher values of *BEXP* lead to higher drainage. Sensitivity tests were conducted for both parameters, finding the value of 0.75 and 1.89 to be optimal for *SLOPE* and *BEXP*, respectively. The higher calibrated value of *BEXP* compared to default value is justifiable because the basin is characterized by continuous baseflow > 0.48 m^3/s.

The time series and FDC of simulated discharge with the chosen parameter values of *REFKDT*, *DKSAT*, *SMCMAX*, *SLOPE*, and *BEXP* (hereafter, Run 4) are presented in Fig. 10 (black lines). Fig. 10b shows a zoom on a snow melt period in spring of 2008. The distorted shape of the hydrograph (Fig. 10b) implies that the parameters controlling the shape hydrograph must be calibrated and all the routing schemes be turned on. The corresponding FDC reported in Fig. 10c black line; shows significant an underestimation of baseflow. This can be quantified by very large negative bias value of 9852 m^3 when bias is computed for flows less than or equal to 1 m^3/s. Note that *BEXP* and *SLOPE* parameters control aquifer recharge and baseflow parameters control the amount of water in the aquifer that can be returned to stream network as baseflow. Therefore, the calibrated value of *BEXP* and *SLOPE* is the optimal values of groundwater drainage.

To improve simulation of baseflow, the baseflow/bucket model parameters were refined. By inspecting the observed FDC, the flow values remain fairly constant (~1 m^3/s) for percent exceedances ≥ 40. Since the bucket model of equation (8) provides a constant outflow even when $z = 0$, i.e. $Q_{out} = C_0 \cdot \exp(-1)$. This relation was used with $Q_{out} = 1\ m^3/s$ to derive the value of C_0 which was set equal to 1.30. Subsequently, different values of z_{max}, and a value of 2000 mm was found to be optimal for z_{max}. The drainage parameter a of the bucket model of equation (8) was found to be the most crucial parameter for the calibration of this process. In general, the higher a, the larger the groundwater contribution to the streamflow in the channel and vice versa. After proper tests, a value of 3.18 was found to be optimal for a. The resulting simulated hydrograph and FDC for calibrated baseflow parameters is shown in Fig. 10 (red line) (hereafter, Run 5). These changes results in significant improvements of the performance metrics applied to the entire time series (Table 4). The bias and CC considering the low flow values (i.e., discharge below the 40% exceedance) was also computed and found that the bias reduced to only +488 m^3 and the CC increases to 0.95 (calibrated baseflow parameters).

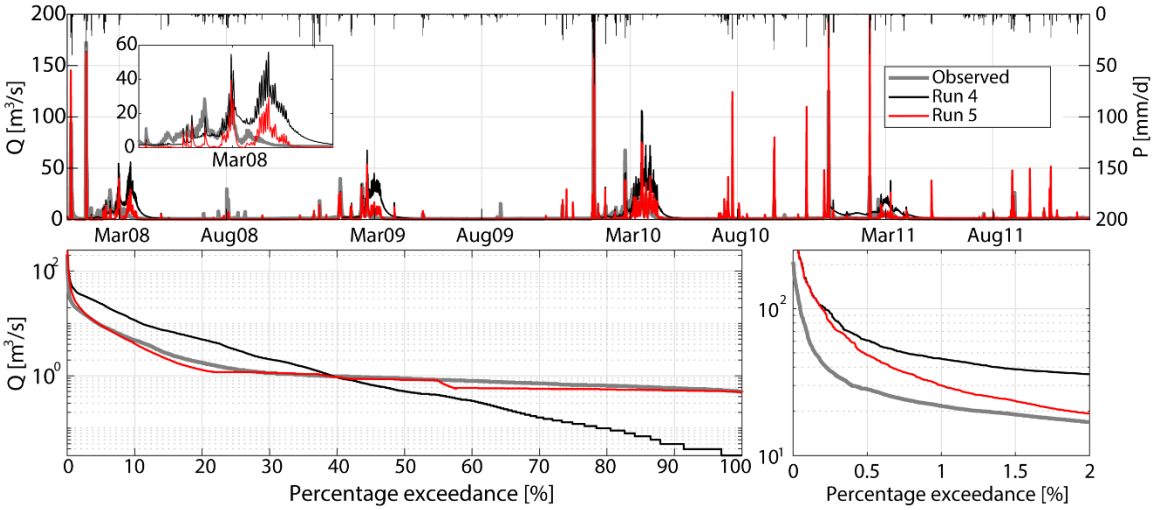

Figure 10: (a) Time series of observed and simulated discharge using the default and calibrated parameters of the groundwater (GW) module, parameters that affect groundwater recharge such as *SLOPE*, *BEXP* and *SMCMAX* for the calibration period. (b) Zoom on a snow melt period showing significant delay in snow melt during late spring. (c) Flow duration curve (FDC) of observed and simulated discharge showing significant under simulation of baseflow with original groundwater parameters and very small bias in baseflow simulation with calibrated baseflow parameters. (c) Zoom on the portion of FDC accounting for the highest peaks.

4.5 STEP 3: SHAPE OF HYDROGRAPH

To improve the shape of the hydrographs for both winter and summer peaks, we focused on parameters that control overland and subsurface flow, including *LKSATFAC, RETDEPRTFAC,* and *OVROUGHRTFAC.* The default values of these parameters are 1000, 1, and 1 respectively. *LKSATFACT* affects lateral redistribution of infiltrated water. Based on the sensitivity test, the default value of *LKSATFAC* was determined to be optimal. *RETDEPRTFAC* is a multiplier to the maximum retention depth before flow is routed as overland flow, whose default values are in turn assigned based on soil classes. The range of *RETDEPRTFAC* used in previous study is 0–10 (Yucel et al. 2015). Sensitivity analysis of three different values of *RETDEPRTFAC* (0, 1, 5) was done and *RETDEPRTFAC* = 0 was determined to be optimal. This result is consistent with (Yucel et al. 2015), who found that *RETDEPRTFAC* = 0 is optimal for steep surface slopes. *RETDEPRTFAC* = 0 means that runoff generated in each pixel is immediately routed downstream and there is less chance of local infiltration. This, in turn, leads to a steeper rise of the hydrograph and an increase in the hydrograph volume.

OVROUGHRTFAC is the multiplier on Manning's roughness for overland flow, whose default values are, again, assigned based on soil classes. (Yucel et al. 2015) suggests a range of *OVROUGHRTFAC* between 0 to 1. Higher values of *OVROUGHRTFAC* increase the flow resistance in the routing grids and, in turn, tend to delay the simulated peak in the channel. Based on sensitivity analysis of *OVROUGHRTFAC,* an optimal value of 0.9 is adequate for this parameter.

4.6. STEP 4: SNOW MELT

4.6.1 EFFECT OF *MFSNO*

The results of the calibration obtained so far are shown with red line in Fig. 10. In particular, Fig. 10b shows a zoom on the time series of observed and simulated discharge during the snowmelt season in 2008. The result clearly indicates a significant lag between observed and simulated hydrographs due to delayed in simulated snowmelt. *MFSNO* is the main parameter that controls this process. To understand the role of *MFSNO* and how its value was identified, it is worth to provide some explanations on snow accumulation and melt processes:

- During snowfall in early winter, the height of the snow increases (see, e.g., Fig. 4d) and so does snow cover fraction (SCF), i.e., the fraction of the grid cell covered by the snow. Since temperatures of snow and soil layers are below freezing temperature (< 0^{o}C) in early winter (Fig. 2a), snow accumulation occurs. As snow accumulates, the bulk density of the snowpack and snow depth increases.
- In Noah-MP, snow depth is modelled from snow depth increasing rate, which is calculated as the ratio of snowfall on the ground surface and bulk density of the snow fall. Snow fall on the ground surface is the sum of through snow fall and snow drip rate. Snow layers is then created from snow depth based on the thresholds described in (Niu et al. 2011)
- When temperature increases, snowmelt occurs. This process is energy dependent and controlled by air, soil, and snow temperature. If the temperature of the *i-th* layer of snowpack or soil skin temperature exceeds 0 ^{o}C, especially during midday hours, snow melting occurs; in contrast, if the temperature falls below 0 ^{o}C during night,

snow refreezing occurs. This cyclical change is evident in increase and decrease of the observed (and simulated) discharge throughout the snow melt season (see Fig. 4b and Fig 11b).

- During snow accumulation and snowmelt, SCF is modeled in Noah-MP based on equation (7), where *MFSNO* is one of the parameters. The effect of *MFSNO* is negligible in early winter, because the ratio of bulk density of snowpack and density of fresh snow is ~1 as the snow mass is relatively new. During the snowmelt season, occurring after snow has accumulated and its density increased, the role of *MFSNO* becomes important. The higher *MFSNO*, the lower SCF, thus, speeding up the transition of snow accumulation and ablation processes (Niu and Yang 2007a).

The simulations presented so far used *MFNSO* = 1. To fix the delay in simulated snowmelt, *MFSNO* was increased to an optimal value of 3.5 (hereafter, Run 6). Results are presented in Fig. 11. This change allowed decreasing the lag time between observed and simulated discharge from ~300 h to ~1 h, with a resulting in CC of 0.84 during snowmelt. Snow melt is also affected by topography and land cover. The effect of vegetation parameters on snow melt was investigated.

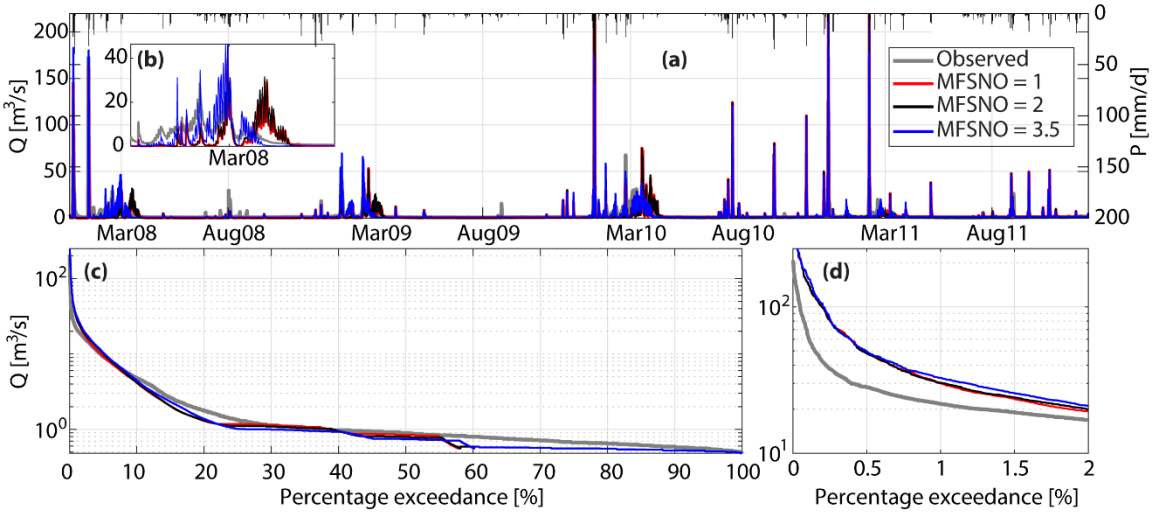

Figure 11: (a) Time series of observed and simulated discharge for simulation period (2008-2011). (b) Zoom on snow melt season in Spring of 2008 showing improved lag time during snow melt season by increasing *MFSNO* from default value of 1 to 3.5. (c) Flow Duration Curve of observed and simulated streamflow. (d) Zoom on the portion of FDC accounting for the highest peaks.

4.6.2 VEGETATION PARAMETERS

To further improve the simulation of snowmelt the effect of one of the vegetation parameters, canopy top height (*HVT*) was investigated. This is multiplicative parameter of the spatially variable vegetation height whose default values are determined by the land cover map. The other vegetation parameters reported in Table 3 (*MP, CWPVT, VCMX25)* were found not to have a significant effect. Figure 12 presents results of the simulations for *HVT* = 0.75, 1, 2. The range of analysis was selected based on the physical height of a pine tree (the dominant land cover type in OCB). Increase in vegetation height slightly increases streamflow volume during the snowmelt season and slightly decreases winter rain peaks (Fig. 12b). This is because of (1) melting of extra intercepted snow on vegetation leaves and trunk, (2) the extra intercepted precipitation causes the melting of intercepted snow, and (3) the extra intercepted precipitation decreases water input on the soil surface, which decreases surface runoff, hence decrease in winter rain peaks. In contrast, changes in *HVT* do not have an appreciable effect on summer peaks (Fig. 12c). This may be due to combined effects of increased interception of rain and evaporation from the canopy. Overall, the effect of *HVT* on the physical processes controlling runoff generation is minimal. This is reflected in the metrics as none of the metrics registered notable improvement or deterioration except for bias in snow melt which worsened by ~1.2%. Therefore, default values of *HVT* and all other vegetation parameters was determined to be optimal.

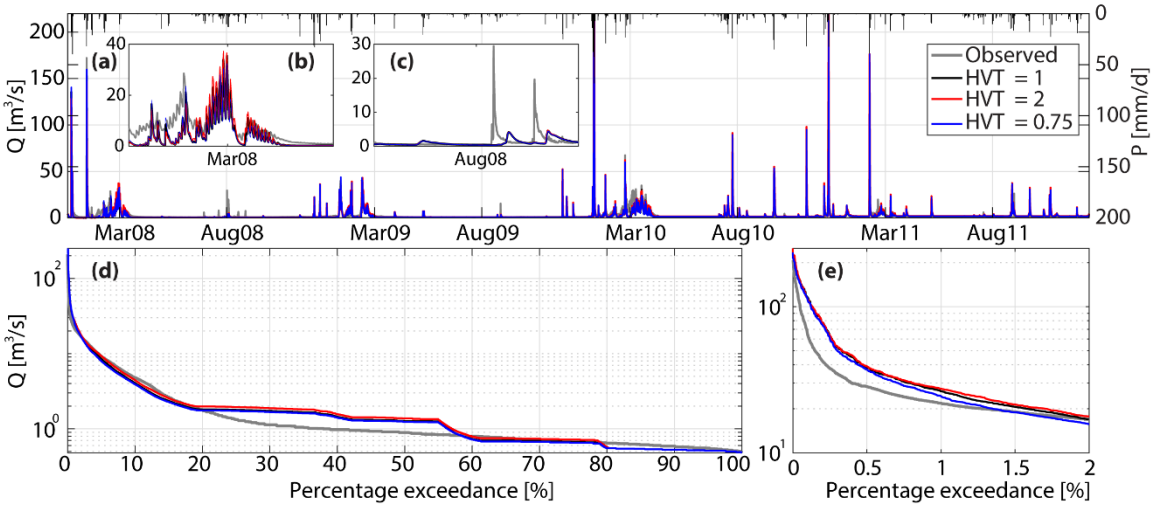

Figure 12**:** (a) Time series of observed and simulated discharge for simulation period (2008-2011). (b) Zoom on snow melt season in spring of 2008 showing slight increases in discharge for *HVT* = 2 (red line) (c) Zoom on summer season in August of 2008 showing no effect of increase in *HVT* on summer peaks with (red line). FDC of observed and simulated streamflow. (d) Zoom on the portion of FDC accounting for the highest peaks.

4.7 STEP 5: FINAL TUNING

The results of the calibration obtained so far are shown with black line in Figure 12. Fig. 12e shows over simulation of 2% of the largest peaks (Table 5, Run 3 best). This, therefore, requires tuning of parameters such as *DKSAT*. Over-simulation of largest peaks can be improved by increasing the preliminary value of *DKSAT* incrementally from 0.3 to an optimal value. As explained in section 4.2.1, increase in *DKSAT* leads to an increase in infiltration capacity of soil thus less run-off is generated, hence decreasing flood peaks. It was found that *DKSAT* value of 0.48 is optimal. Further increase of *DKSAT* value above 0.48 results in improved high peaks ($Q > 50$ m^3/s) but under simulate mid-to-high peaks ($Q > 25$ m^3/s). This may be due to further re-infiltration of water in the channel grid networks (channel losses). The calibrated model registered an acceptable result for all the metrics on daily scale with NSE = 0.56, RMSE = 2.36 m^3/s, and CC = 0.75. The total bias computed at the outlet of the basin is 0.35 Mm^3 (~8 % overestimation). However, the summer peaks are still under-simulated. This is reflected in MRE of summer peaks of -84 (significant under-simulation of summer peaks). This is perhaps due to the limitation of precipitation forcing as explained in section 5.1.

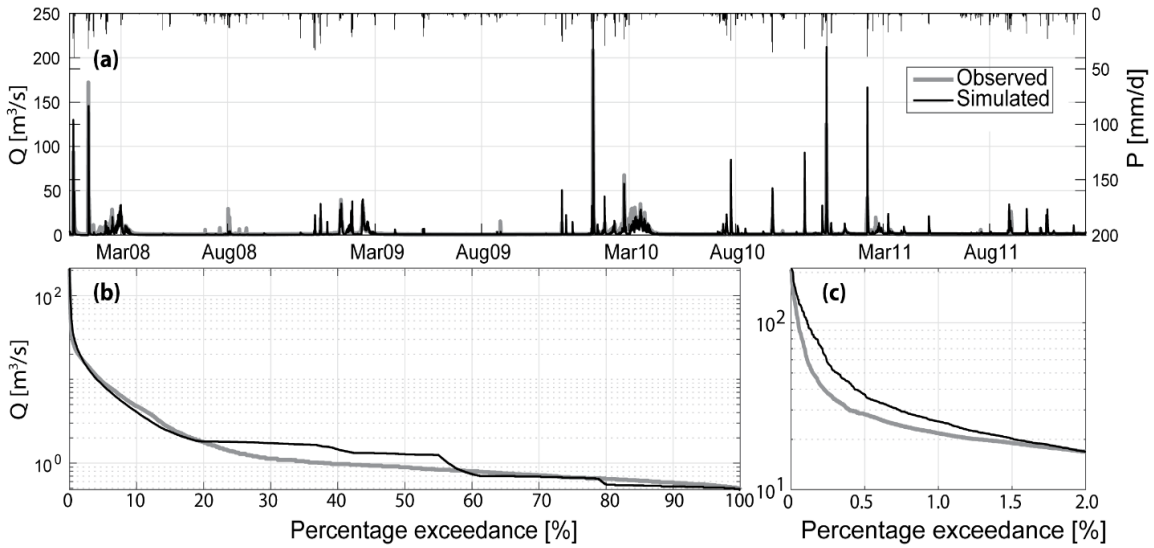

Figure 13: (a) Time series of observed and simulated discharge for calibration period (2008-2011) simulated using tuned value of *DKSAT* and *OVROUGHRTFAC* to improve shape of hydrograph. Note that winter rain peaks are simulated well while summer monsoon peaks are under-simulated (b) FDC of observed and simulated streamflow. (c) Zoom on the portion of FDC accounting for the highest peaks.

To improve under-simulation of summer peaks, different values of *DKSAT* and *REFKDT* was used for winter and summer seasons. These two parameters were selected because as they significantly impact surface infiltration and hence runoff compared to other parameters. To do this, the model was run with *DKSAT* and *REFKDT* multiplicative values of 0.48 and 0.3 respectively and other calibrated parameters for winter seasons and different values of *REFKDT* and *DKSAT* for summer seasons. It was found that *REFKDT* and *DKSAT* multiplicative values of 0.1 and 0.3 were optimal. Fig 14a shows the result of using dynamic values of *DKSAT* and *REFKDT* for winter and summer seasons. Fig 14b shows a zoom on 5 summer peaks in August of 2008 showing improved simulation of summer peaks. Average volumetric bias in summer peaks decreased from -15,428 m^3 (under-simulation) to -356 m^3 (under-simulation) while MRE and time lag improved by ~86 % and ~25% respectively from -84.0 % to -12% for MRE and from 2.4 hours to 1.8 hours for time lag. The calibrated model with dynamic parameters registered an improved result of NSE (0.62) and CC (0.86) for the entire 4-year simulation period at daily scale. The value of NSE is within acceptable range of 0 and 1 (Moriasi et al. 2007) . The RMSE value also improved slightly from 2.36 m^3/s to 2.28 m^3/s. The final value of RMSE is considered acceptable as per (Singh et al. 2004) who suggested that RMSE values less than half of the standard deviation of the observations may be considered low.

Overall, the result of the simulation with final calibrated parameters shows better correlation with observations at both interior and outlet gages in comparison with other similar studies such as (Lahmers et al. 2019a; Kerandi et al. 2018; Yucel et al. 2015; Silver et al. 2017).

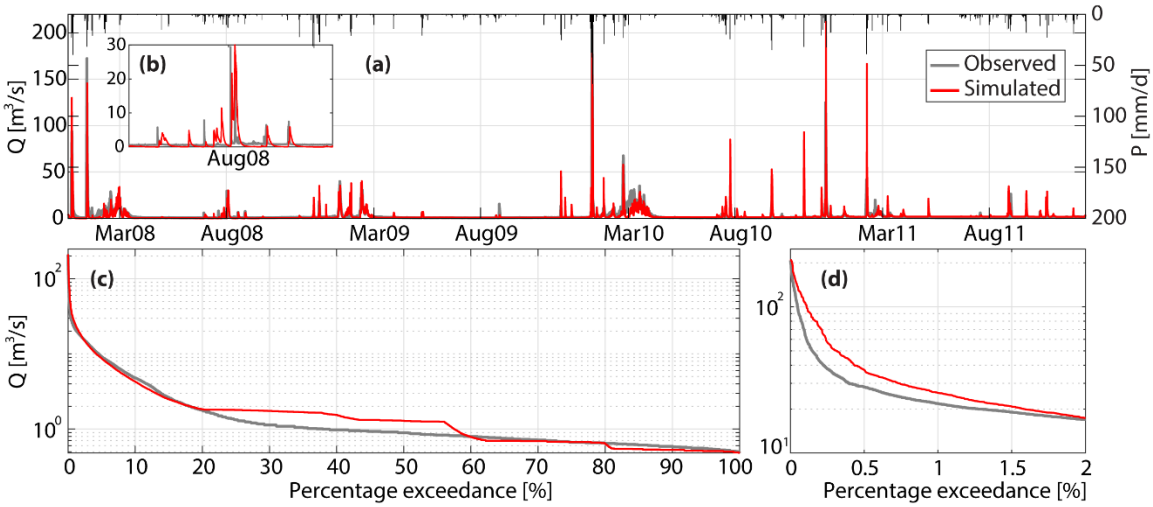

Figure 14**:** (a) Time series of observed and simulated discharge for calibration period (2008-2011) simulated using dynamic values of *DKSAT* and *REFKDT* to improve simulation of summer peaks. Note that smaller values of *DKSAT = 0.3* and *REFKDT = 0.1* improves summer monsoon peaks (b) Zoom on few peaks in August of 2008 showing marked improvement in simulation of summer peaks when seasonal dynamic values of parameters that control infiltration is used. (c) FDC of observed and simulated streamflow. (d) Zoom on the portion of FDC accounting for the highest peaks.

4.8 STEP 6: VALIDATION

The model calibration was validated from 2012 to 2017. The observed discharge at interior gage and outlet gage was used to validate the simulated discharge modelled by calibrated WRF-Hydro model. Similar to calibration, different values of *DKSAT* and *REFKDT* was used for winter and summer seasons. The validation result of the simulated discharge and observations at the basin outlet from 2012 to 2017 is shown in Figs. 15. The model performance during validation period is summarized in Table 4. The FDC of the time series of the validation period at basin outlet (Fig. 15c) shows an overestimation in simulated discharge, especially for high peaks. The validation of simulated discharge at the interior gage in Sedona from 2008 to 2017 is shown in Figs. 16. The result also shows an overestimation in simulated discharge for flood peaks. The model performance at the two gage locations are comparatively similar for most of the metrics as summarized in Table 4. The bias at the interior gage is +1.02 Mm^3 (~16% overestimation) while at outlet is 0.86 Mm^3 (~10% overestimation). The net reduction in bias between interior gage and outlet gage of 0.16 m^3 can be attributed to channel losses in ephemeral streams contributing to Oak Creek River. Channel losses is a significant component of water balance especially in this study basin which WRF-Hydro model does not take into account. The drainage area of the interior gage is significantly smaller than that of outlet gage, therefore, the total effect of channel/transmission loss is more significant at outlet gage than interior gage.

It is important to note that over-simulation of flood peaks noted at both interior and outlet gages can be attributed to model limitations, forcing limitation, and limitation of manual calibration. These limitations are reflected in overall model performance

although the metrics results for both calibration and validation periods are within acceptable range.

The quality of calibrated WRF-hydro model in OCB was also tested by the model's ability to simulate snow depth and SWE. To this goal, mean basin snow depth and SWE simulated by WRF-Hydro model was compared against mean SWE and snow depth derived from the NOAA Snow Data Assimilation System (SNODAS) products (NOHRSC (National Operational Hydrologic Remote Sensing Center) 2004). Although observations of snow depth and SWE at the SNOTEL stations exist, they were not used to validate the model because (1) possible localized accumulation of snow due to snow drifts caused by wind that may be recorded at the stations, and (2) unfair comparison of point-scale measurement to 1 km^2 domain. SNODAS data was used to validate the model because (1) grids have same spatial resolution (2) physically based, spatially distributed energy and mass balance snow model forced with similar meteorological forcing to WRF-hydro model is used to simulate snow. SNODAS dataset was downloaded and clipped to the study basin. The procedure followed in projecting the clipped data to basin grids is described in (Barrett 2003).

Fig 17 shows the comparison of SWE and snow depth simulated by calibrated WRF-Hydro model and SNODAS from 2008 to 2017. The model simulation exhibits high correlation of 0.85 and 0.89 for snow depth and SWE respectively with SNODAS. However, WRF-Hydro model under-simulates both SWE and snow depth compared to SNODAS. This may be due to the fact that SNODAS manually adjust their values daily based on observations (Carroll et al. 2001), and hence short duration snowfall can be captured. Close inspection of the time series of snow depth shows faster ablation of snow

during snow melt season in most of the years. This may be attributed to (1) high value of snow melt factor, *MFSNO* used to bridge the lag time during snow melt season, (2) external factors such as loss of vegetation cover due wild fire that exposes snowpack to excessive sunlight causing faster ablation of snow, and (3) temperature changes in which some years snow melt quickly compared others.

Observed snow depth and SWE recorded at the three SNOTEL stations were used to further validate the effectiveness of the model in simulating snow dynamics at point scale. To this goal, co-located snow depth and SWE values for both SNODAS and calibrated WRF-Hydro model were extracted and compared to the observations. The result shows appreciable correlation between observation and simulations for both SWE and snow depth (Fig. B1). However, the model under simulate both SWE and snow depth compared to observations due to possible localized accumulation of snow due to snow drifts caused by wind that may be recorded at the stations. Note that for Mormon Mountain Summit station, the records of both SWE and snow depth starts late in 2008, therefore, the values of snow depth and SWE in early winter season of 2008 is missing.

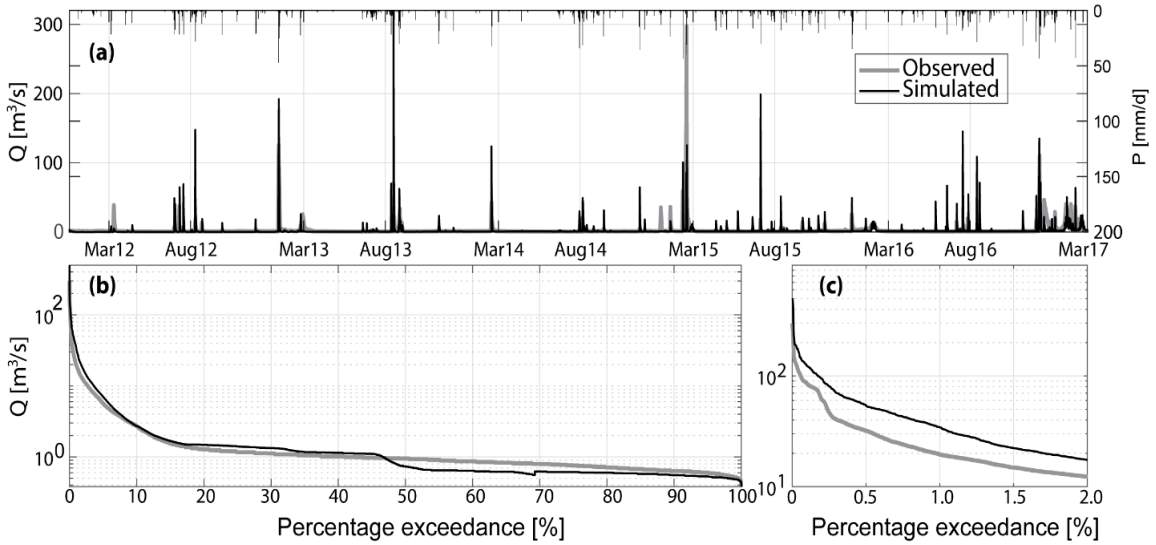

Figure 15: Time series of validation period (2012-2017). (a) Time series showing simulated and observed discharge at basin outlet. (b) FDC of observed and simulated streamflow at the outlet of the basin. (c) Zoom on the portion of FDC accounting for the highest peaks

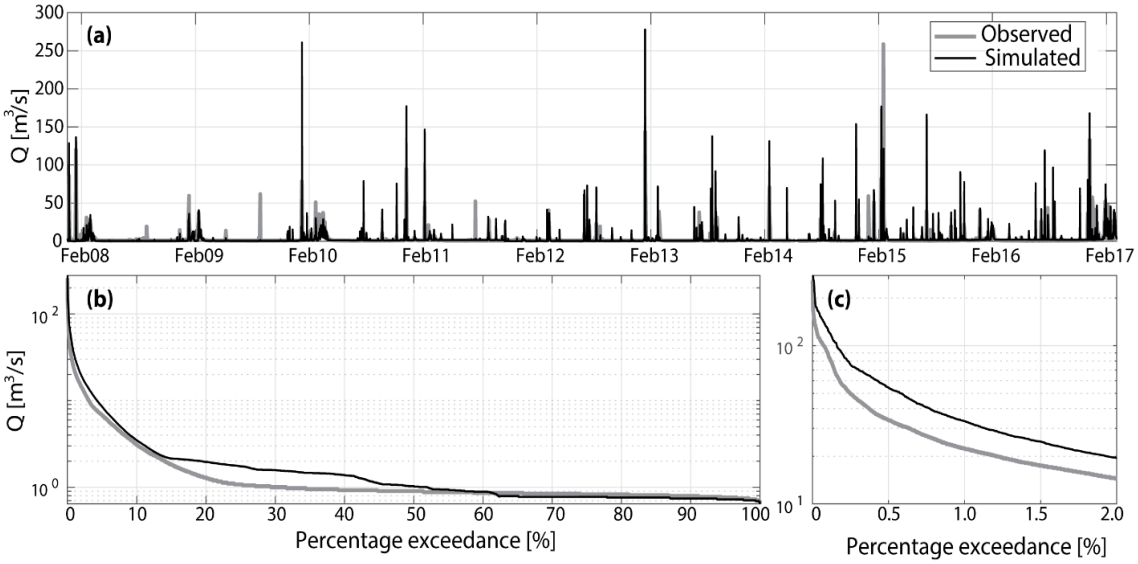

Figure 16: Validation at Sedona gage (2008-2017). (a) Time series showing simulated and observed discharge evaluated at Oak Creek near Sedona gage. (b) FDC of observed and simulated streamflow at interior gage. (c) Zoom on the portion of FDC accounting for the highest peaks

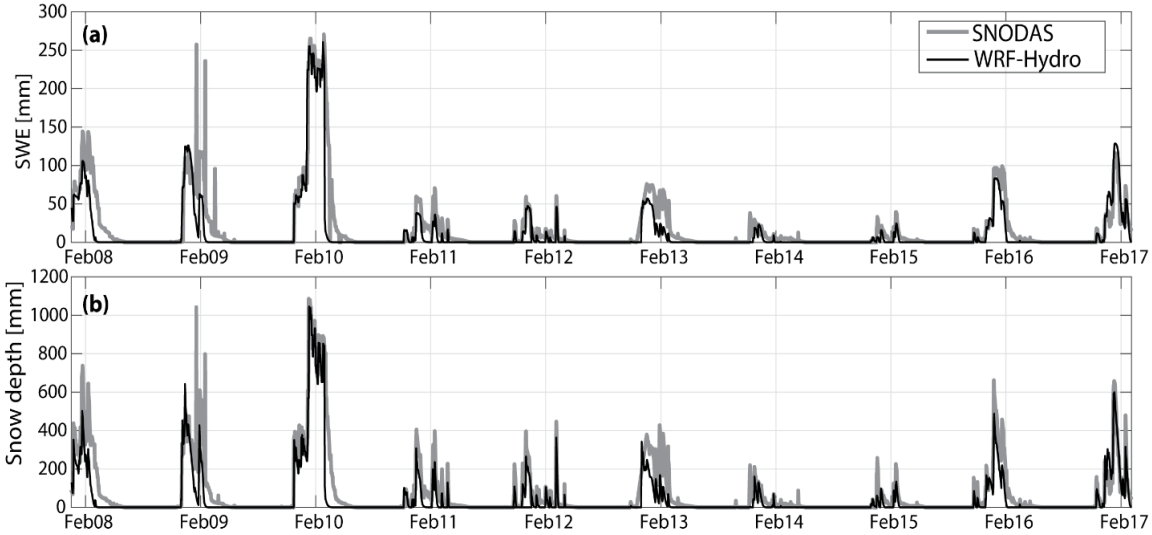

Figure 17**:** Comparison of snow simulation obtained through calibrated WRF-Hydro model with SNODAS model. (a) Simulated Snow Water Equivalent (SWE) (b) Simulated Snow depth.

Table 6: Metrics quantifying model performance in simulating snow during calibration and validation periods

SNODAS - WRF-Hydro comparison	**NSE**	**RMSE [mm]**	**Bias [mm]**	**CC**	**Reference**
SWE	0.65	2.74	76	0.89	Fig 17a
Snow depth	0.69	2.69	120	0.85	Fig 17b

CHAPTER 5

DISCUSSION

5.1 FORCING LIMITATIONS

The original NLDAS-2 forcing was downscaled and regridded to 1 km by 1 km pixels from coarse ~12 km resolution which spatially spreads precipitation over a large domain. The bias-corrected NLDAS-2 precipitation forcings that was used to drive the model does not account for the existence of highly localized and intense summer monsoon events. The climatological monthly mean precipitation for summer seasons recorded at SNOTEL stations and NCEI gages during calibration and validation periods was compared against co-located bias-corrected NLDAS-2 forcings pixels. It was found that forcings values are less than observations for all the 5 gages. This is because of the localized nature of summer monsoon events. Rain gages records precipitation at point scale and, therefore, localized precipitation can be captured while NLDAS-2 forcings spreads precipitation over 1 km^2 domain. The forcing limitation is evident in poor model response during summer seasons (Fig. 13a). The dominant runoff generation mechanisms during summer seasons in the study area is infiltration excess. From eqn. 2, surface runoff is generated when precipitation exceeds maximum infiltration capacity of the soil. Therefore, the spread of precipitation forcings over large domain reduces simulated surface runoff during summer months. To compensate for the limitations of the forcings, different multiplicative values of *REFKDT* and *DKSAT* was used for winter and summer seasons. The reduced values of *REFKDT* and *DKSAT* during summer seasons decreases infiltration and increases surface runoff, which results in better model response (Fig. 14b). Summer rainfall events are also in time (i.e., short duration). To reduce streamflow

simulation errors during summer seasons for southwestern United States, different precipitation forcing dataset need to be utilized. For instance, precipitation records of finely spaced gages in a square kilometer domain should be used to force the model.

5.2 SUMMARY OF PARAMETER SENSITIVTIES

In this study, sensitivity analysis of various parameters (both hard-corded and standard parameters) controlling different runoff generation mechanisms was performed. The hard-corded parameters tested in this study are roughness length of snow surface, *Z0SNO* or z_0 in eqn. 7, liquid water holding capacity of snowpack, *SSI*, and surface resistance for snow, *RSURF_SNOW* which are all listed in look-up table called "MPTABLE.TBL". These parameters affect turbulent energy exchanges (sublimation and energy flux to the snowpack). Since *MFSNO* have significant effect on snow sublimation rates, the global default value of *Z0SNO* (0.002 mm), *SSI* (0.03 m^3/m^3), and *RSURF_SNOW* (50 s/m) was determined to be adequate for the study site. The significant findings of the study are that:

(1) The soil parameters are more sensitive to all runoff generation mechanisms than vegetation parameters.
(2) Snow melt factor, *MFSNO* has significant effect on snow melt
(3) Forcing limitations, as explained in section 5.1, greatly impacts model's ability to simulate land surface processes such as rainfall-runoff generation.

Sensitivity analysis of soil parameters such as *DKSAT*, *BEXP*, and *SMCMAX* was performed. These parameters mainly affect winter rain and summer monsoon floods, and baseflow. In this study, sensitivity of a parameter is determined by the parameter effect on (1) volume of the hydrograph peak, and (2) timing of the peak. This is quantified

through metrices such as MRE, t_{lag}, and average bias. Of the soil parameters, it was found that *DKSAT* is the most sensitive parameter to runoff generation for winter rain and summer monsoon peaks, followed by *SMCMAX*. Both of these parameters control infiltration rate and thus affects surface runoff (eqn. 1-2). *BEXP* parameter mainly controls deep drainage and, therefore, its effect on winter rain and summer monsoon peaks is minimal. *REFKDT* is the most sensitive parameter of all runoff parameters solely based on its effect on hydrograph volume of the flood peaks. Higher *REFKDT* value (>1) reduces the simulated hydrograph volume as more water is partitioned as underground runoff instead of surface runoff. *HVT* is the most sensitive parameter of all vegetation parameters, followed by *CWPVT*. *HVT* parameter mainly affects winter rain floods and snow melt. This is because height of the canopy influences snow and rainfall interceptions, and transpiration and limits below-canopy evaporation. *HVT* determines the height of wind and humidity measurements which are used to calculate aerodynamic resistance, an important parameter of Penman-Monteith equation. Note that the effect of *HVT* on summer floods may be masked by forcing limitation. *CWPVT* is canopy wind parameter used for formulation of wind speed. *CWPVT* mainly affects unloading rate of snowfall from canopy leaves which further impacts the amount of snow mass on the ground surface (i.e., high *CWPVT* leads to an increase in snow depth, which further increases snow melt discharge) (Niu et al. 2011). The four runoff generation mechanisms are least sensitive to vegetation parameters such as *MP* and *VCMX25*. The diminished effect of vegetation parameters on the runoff generation mechanisms may be due to lack of vegetation heterogeneity in the study area compared to more heterogenous soil texture (Figs. 3a, b).

The second finding of the study was obtained through improvement of lag time during snow melt. As explained in section 4.6.1, increase in *MFSNO* increases snow ablation rates. The high optimal value of *MFSNO* is consistent with (Niu and Yang 2007) who found that *MFSNO* is larger for mountainous basins compared to plain river basins.

The third finding of the study was obtained through lack of improvement of simulated summer peaks as winter rain peaks are greatly improved. Both processes are controlled by same model parameters, therefore, if one process is improving while the other is not, it means that parameter optimization is adequate, and the issue may be limitations of forcing or model structural errors. As explained in section 5.1, the forcing limitations significantly affects simulations of summer floods.

5.3 COMPARISON WITH EXISTING LITERATURE

The maximum surface slope of the study basin is 42^{o} and the basin is dominated by steep slope especially in mountainous upstream areas, hence calibrated zero value of *RETDEPRTFAC* was determined to be adequate. This is consistent with (Yucel et al. 2015; Kerandi et al. 2018) who suggested that in areas with surface slopes higher than $30\text{-}45^{o}$, retention depth shows little to no accumulation. (Verri et al. 2017) noted that there is seasonality associated with soil physics and assumed different values of soil parameters for different seasons to achieve more appreciable model performance, a result which consisted with the findings presented in section 4.7. The soil parameters such as *DKSAT* and *SMCMAX* were found to be critical parameters for surface water fluxes for all the hydrological processes, finding which is consistent (Cuntz et al. 2016). However, Contrary to (Cuntz et al. 2016), vegetation parameters such as *MP* and *VCMX25* exhibits least sensitivities to runoff generation mechanisms. This perhaps might be due to

vegetation characteristics, as the dominant vegetation type (~85%, Fig. 3b) is non-deciduous plant, hence transpiration is relatively constant throughout the year. In terms of metrics, the model performance is better or similar to all studies that used same precipitation forcing such as (Lahmers et al. 2019).

5.4 MODEL LIMITATIONS

During calibration of the model in OCB, the following model limitations were noted:

- Baseflow model is highly conceptualized and requires longer simulation period to effectively simulate underground runoff. As described in other studies such as (Niu et al. 2007; Gochis et al. 2018), long simulation period is required for the model parameters to reach equilibrium. Indeed, (Niu et al. 2007) noted that it might take more 250 years of model spin-up for baseflow model to equilbriate in semi-arid environemnts. In semi-arid region, such as the study area, the simple "bucket" function is inadequate. This is due to (1) the water table being too deep, and (2) the assumption of baseflow model that there is only one-way connection between soil column and underlying aquifer is not physically true. This is because the water from aquifer can rise up due to capillary action and contribute to the soil moisture, hence the "bucket" model should be modified to account for two-way connection especially in semi-arid areas such as the study area. The impact of added moisture due to capillary action will induce errors in simulated soil moisture, which further leads to poor representation of processes such as evapotranspiration (ET) (as groundwater provides water for ET through root uptake), and discharge simulations. The importance of groundwater dynamics in modulation of energy and water fluxes cannot be overstated. The conceptual bucket model can be improved by developing a

more detailed conceptual baseflow model that can simulate the specific characteristics of regional aquifer. Model such as MODFLOW groundwater model can be used in this region because surface-subsurface water coupling is not very strong in northern and central Arizona because the aquifers are very deep (Pool et al 2011) . Studies such as (Aghlmand and Abbasi 2019), though at local scale, have also used MODFOW groundwater model to improve baseflow simulations. In areas of southwestern US where baseflow is not always present, the bucket model should be turned off to prevent unrealistic simulations of baseflow.

- Most the peaks are characterized by sudden recession as opposed to expected long recession limb. This perhaps is due to slow movement of water in soil column, which decreases the contribution of infiltrated water to total runoff.
- Some physical processes that are important in semi-arid regions such as Arizona are not represented in the model. One of these processes is transmission losses in ephemeral channels. As discussed in (Lahmers et al. 2019), accounting for channel infiltration will result in reduced streamflow errors.
- Uniform 2m deep soil column to bedrock everywhere is assumed to be representative of soil structure. The model does not account for the complexity of variance in soil thickness due to terrain.

5.5 TERRESTRIAL WATER BALANCE MODEL

The output of calibrated WRF-Hydro model was used to verify whether the water balance model closes out to zero or not. The water balance was evaluated on an annual scale. The terrestrial water balance components derived from model outputs are

precipitation (P), evapotranspiration (ET), and run-off (Q). The mean annual P, ET, and Q are computed for each year during calibration and validation. The storage term of water balance is computed as the difference between P and sum of ET and Q (both observed and simulated). ET is high throughout calibration and validation period because (1) high average temperature throughout the year (Fig. 2a), and (2) The dominant land cover type is evergreen and hence high rate of transpiration. The result of water balance during calibration and validation period is shown in Table 7. Note that the storage term is negative in some years and positive in other years corresponding to dry and wet years.

Table 7: Annual values of the water balance components; i.e, P, ET, and R computed from outputs of calibrated WRF-Hydro model for 2008-2016. The storage term is given by $dS/dt_1 = P - (R_{obs}+ET)$ while $dS/dt_2 = P - (R_{sim}+ET)$. The storage term shows both negative and positive values indicating wet and dry years. The terrestrial water balance does not zero-out for all the years because of model limitations and or limitations of manual inadequate calibration.

Year	**P (mm)**	**ET (mm)**	**Robs (mm)**	***Rsim (mm)***	***dS/dt₁* (mm)**	***dS/dt₂* (mm)**
2008	641	504	134	140	3	-3
2009	388	314	79	75	-5	-1
2010	857	668	182	187	7	2
2011	515	438	73	77	4	0
2012	422	367	58	59	-3	-4
2013	555	454	97	99	4	2
2014	366	312	60	57	-6	-3
2015	707	581	121	123	5	3
2016	511	414	101	98	-4	-1

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This work presents characterization of streamflow and floods into four runoff generation mechanisms: the winter and summer floods, snow melt, and baseflow. A stepwise approach targeting each mechanism was implemented to calibrate WRF-Hydro model in Oak Creek basin, a sub-basin of Verde River basin, in central Arizona. This study provides initial implementation of WRF-Hydro model in the study basin. Meteorological forcings obtained from NLDAS-2 was used to drive the model. The uncoupled mode of the model was used to analyze the model's ability to represent the physical processes of runoff generation.

The novelty of this study is the identification of key hydrological processes and model parameters that exert control on representation of those processes. This process-based manual calibration approach allows for (1) increased computational efficiency compared to automatic calibration techniques (i.e., scale-up to a larger basin is computationally expensive); and (2) identification of physically plausible values of the model parameters. The calibration approach presented in this work can support the calibration of the NWM in other basins. As result, this work can potentially reduce financial investments needed for the implementation of the operational NWM to forecast seasonal streamflow by water agencies such as SRP.

The model parameters were calibrated sequentially using streamflow observation data at the outlet of the basin, first by targeting winter and summer flood peaks, then baseflow component, hydrograph shape, and finally, snow dynamics. Sensitivity analysis was performed for each parameter and the range of analysis is constrained by either

plausible physical limits or limits specified in other studies. The discharge simulated by calibrated model was validated from 2012 to 2017 against two streamflow observations (at interior and outlet of the basin) and simulated mean snow depth and SWE was compared against those derived from SNODAS. The model performance during calibration and validation period presents promising results both quantitively and qualitatively. The main conclusions of this study are:

- NLDAS-2 precipitation forcings presents considerable limitations for local scale application. Prior to calibration, the precipitation forcing was compared against observations recorded at NCEI and SNOTEL gages. The result shows the existence of a linear relationship between climatological monthly precipitation means and elevation for the gages and no linear trend for NLDAS-2 forcings. Precipitation forcings was bias corrected for elevation. The bias-corrected forcings also presents errors especially during summer seasons, the findings that is consistent with (Lahmers et al. 2019).
- The soil, groundwater module, and snow parameters exert major control on hydrological processes presented in this work compared to vegetation parameters.
- Also as noted in (Senatore et al. 2015), model structural errors such as highly conceptualized exponential model results in sudden recession of peaks.

The results summarized in Tables 4, 5 and shown in figs. 14, 15, and 17 indicate that WRF-Hydro model is capable of simulating the four runoff generation mechanisms. Despite the limitations of the forcings and model's structural errors, the model performance for final calibration step and validation is acceptable. Therefore, this calibration effort at local scale will provide a basis for further evaluating the model's

capability to forecast streamflow. However, as noted in (Moreno et al. 2013), the accuracy of the model in forecasting streamflow at local scale will depends on the quality of precipitation input, especially during summer seasons.

The city of Sedona covers ~1.7% of the study basin. During calibration, the built environment of the basin was not treated in any special way in regard to (1) consideration of the imperviousness of city infrastructure, (2) urban land cover, and (3) possibility of withdrawal of surface water from Oak Creek River and injection of treated waste water into groundwater aquifer. As the population of the city is expected to grow, proper consideration of this built environment in a natural watershed is necessary. This is an important area of future model improvement.

REFERENCES

Aghlmand, and Abbasi, 2019: Application of MODFLOW with Boundary Conditions Analyses Based on Limited Available Observations: A Case Study of Birjand Plain in East Iran. *Water*, **11**, 1904, https://doi.org/10.3390/w11091904.

Anghileri, D., N. Voisin, A. Castelletti, F. Pianosi, B. Nijssen, and D. P. Lettenmaier, 2016: Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resour. Res.*, **52**, 4209–4225, https://doi.org/10.1002/2015WR017864.

Ball, J. T., I. E. Woodrow, and J. A. Berry, 1987: A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions. *Progress in Photosynthesis Research*, Springer Netherlands, 221–224.

Barrett, A. P., 2003: *National Operational Hydrologic Remote Sensing Center SNOw Data Assimilation System (SNODAS) Products at NSIDC*.

Biondi, D., G. Freni, V. Iacobellis, G. Mascaro, and A. Montanari, 2012: Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Phys. Chem. Earth*, **42**–**44**, https://doi.org/10.1016/j.pce.2011.07.037.

Boyle, D. P., H. V Gupta, and S. Sorooshian, 2000: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resour. Res.*, **36**, 3663–3674, https://doi.org/10.1029/2000WR900207.

Carroll, T., D. Cline, G. Fall, A. Nilsson, L. Li, and A. Rost, 2001: *NOHRSC OPERATIONS AND THE SIMULATION OF SNOW COVER PROPERTIES FOR THE COTERMINOUS U.S.*

Chen, F., and J. Dudhia, 2001: Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Weather Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, F., Z. Janjić, and K. Mitchell, 1997: Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary-Layer Meteorol.*, **85**, 391–421, https://doi.org/10.1023/A:1000531001463.

Chen, Y., J. Li, and H. Xu, 2016: Improving flood forecasting capability of physically based distributed hydrological models by parameter optimization. *Hydrol. Earth Syst. Sci*, **20**, 375–392, https://doi.org/10.5194/hess-20-375-2016.

Christensen, R. A., and Eilbert, R. F., 1985: Seasonal precipitation forecasting with a 6-7 month lead time in the Pacific Northwest using information theoretic model. *Mon. Wea. Rev*, 502–518.

Cuntz, M., J. Mai, L. Samaniego, M. Clark, V. Wulfmeyer, O. Branch, S. Attinger, and S. Thober, 2016a: The impact of standard and hard-coded parameters on the hydrologic

fluxes in the Noah-MP land surface model. *J. Geophys. Res.*, **121**, 10,676-10,700, https://doi.org/10.1002/2016JD025097.

——, ——, ——, ——, ——, ——, ——, and ——, 2016b: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *J. Geophys. Res.*, **121**, 10,676-10,700, https://doi.org/10.1002/2016JD025097.

D.R. Pool, Kyle W. Blasch, James B. Callegary, Stanley A. Leake, and L. F. G., 2011: Regional Groundwater-Flow Model of the Redwall-Muav, Coconino, and Alluvial Basin Aquifer Systems of Northern and Central Arizona. *Sci. Investig. Rep.*,.

Duan, Q., S. Sorooshian, and V. K. Gupta, 1994: Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.*, **158**, 265–284, https://doi.org/10.1016/0022-1694(94)90057-4.

Dugger, A., and A. Dugger, 2017: Learning from the National Water Model: Regional Improvements in Streamflow Prediction through Experimental Parameter and Physics Updates to the WRF-Hydro Community Model.

Endicott, E., 2013: *Special Supplement to the EXPLAINING EXTREME EVENTS OF 2013 From A Climate Perspective*.

Gallagher, M., and J. Doherty, 2007: Parameter estimation and uncertainty analysis for a watershed model. *Environ. Model. Softw.*, **22**, 1000–1020, https://doi.org/10.1016/j.envsoft.2006.06.007.

Gautam, J., G. Mascaro, E. Vivoni, and Z. Wang, 2018: *Evaluation of CMIP5 Historical Simulations in the Colorado River Basin*.

Gochis, D. J., W. Yu, and D. N. Yates, 2015: The NCAR WRF-Hydro Technical Description and User's Guide, version 3.0. NCAR Technical Document. 120.

——, and Coauthors, 2018: *The WRF-Hydro modeling system technical description, (Version 5.0)*. 107 pp.

Guan, X., G. Mascaro, D. Sampson, and R. Maciejewski, 2019: A metropolitan scale water management analysis of the food-energy-water nexus. *Sci. Total Environ.*, **In Press**.

Hawkins, G. A., E. R. Vivoni, A. Robles-Morua, G. Mascaro, E. Rivera, and F. Dominguez, 2015: A climate change projection for summer hydrologic conditions in a semiarid watershed of central Arizona. *J. Arid Environ.*, **118**, 9–20, https://doi.org/10.1016/j.jaridenv.2015.02.022.

Homer, C. G., and Coauthors, 2015: Completion of the 2011 National Land Cover Database for the conterminous United States – Representing a decade of land cover change information. *Photogramm. Eng. Remote Sensing*, **81**, 345–354.

Hydrologist, and NOAA, 2010: *Middle Atlantic River Forecast Center Middle Atlantic River Forecast Center Region Figure 1: Hourly Rainfall Totals for MARFC Region*.

Julien, P. Y., B. Saghafian, and F. L. Ogden, 1995: RASTER-BASED HYDROLOGIC MODELING OF SPATIALLY-VARIED SURFACE RUNOFF. *JAWRA J. Am. Water Resour. Assoc.*, **31**, 523–536, https://doi.org/10.1111/j.1752-1688.1995.tb04039.x.

Kerandi, N., J. Arnault, P. Laux, S. Wagner, J. Kitheka, and H. Kunstmann, 2018: Joint atmospheric-terrestrial water balances for East Africa: a WRF-Hydro case study for the upper Tana River basin. *Theor. Appl. Climatol.*, **131**, 1337–1355, https://doi.org/10.1007/s00704-017-2050-8.

Kumar, S. V., and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeorol.*, **15**, 2446–2469, https://doi.org/10.1175/JHM-D-13-0132.1.

Lahmers, T., and Coauthors, 2015: Optimization of precipitation and streamflow forecasts in the southwest Contiguous US for warm season convection. *AGUFM*, **2015**, H53A-1650.

Lahmers, T. M., H. Gupta, C. L. C. L. Castro, D. J. D. J. Gochis, D. Yates, A. Dugger, D. Goodrich, and P. Hazenberg, 2019a: Enhancing the structure of the WRF-hydro hydrologic model for semiarid environments. *J. Hydrometeorol.*, **20**, 691–714, https://doi.org/10.1175/JHM-D-18-0064.1.

Lahmers, T. M., H. Gupta, C. L. Castro, D. J. Gochis, D. Yates, A. Dugger, D. Goodrich, and P. Hazenberg, 2019b: Enhancing the structure of the WRF-hydro hydrologic model for semiarid environments. *J. Hydrometeorol.*, **20**, 691–714, https://doi.org/10.1175/JHM-D-18-0064.1.

Latorre, B., and D. Moret-Fernández, 2019: Simultaneous estimation of the soil hydraulic conductivity and the van Genuchten water retention parameters from an upward infiltration experiment. *J. Hydrol.*, https://doi.org/10.1016/j.jhydrol.2019.03.011.

Lehner, B., Verdin, K., Jarvis, A., 2008: New global hydrography derived from spaceborne elevation data. Eos, Transactions, AGU, 89(10): 93-94.

Li, J., F. Chen, G. Zhang, M. Barlage, Y. Gan, Y. Xin, and C. Wang, 2018: Impacts of Land Cover and Soil Texture Uncertainty on Land Model Simulations Over the Central Tibetan Plateau. *J. Adv. Model. Earth Syst.*, **10**, 2121–2146, https://doi.org/10.1029/2018MS001377.

Liston, G. E., and K. Elder, 2006: A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). *J. Hydrometeorol.*, **7**, 217–234.

Mascaro, G., 2017: Multiscale Spatial and Temporal Statistical Properties of Rainfall in Central Arizona. *J. Hydrometeorol.*, **18**, 227–245, https://doi.org/10.1175/jhm-d-16-0167.1.

Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res. D Atmos.*, **109**, 1–32, https://doi.org/10.1029/2003jd003823.

Moreno, H. A., E. R. Vivoni, and D. J. Gochis, 2013: Limits to Flood Forecasting in the Colorado Front Range for Two Summer Convection Periods Using Radar Nowcasting and a Distributed Hydrologic Model. *J. Hydrometeorol.*, **14**, https://doi.org/10.1175/JHM-D-12-0129.1.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, 2007: *MODEL EVALUATION GUIDELINES FOR SYSTEMATIC QUANTIFICATION OF ACCURACY IN WATERSHED SIMULATIONS*.

Naabil, E., B. L. Lamptey, J. Arnault, H. Kunstmann, and A. Olufayo, 2017: Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa. *J. Hydrol. Reg. Stud.*, **12**, 196–209, https://doi.org/10.1016/j.ejrh.2017.05.010.

Niu, G.-Y., and Z.-L. Yang, 2004: Effects of vegetation canopy processes on snow surface energy and mass balances. *J. Geophys. Res. Atmos.*, **109**, https://doi.org/10.1029/2004JD004884.

——, ——, R. E. Dickinson, L. E. Gulden, and H. Su, 2007: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *J. Geophys. Res.*, **112**, D07103, https://doi.org/10.1029/2006JD007522.

Niu, G. Y., and Z. L. Yang, 2006: Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *J. Hydrometeorol.*, **7**, 937–952, https://doi.org/10.1175/JHM538.1.

Niu, G. Y., and Z. L. Yang, 2007a: An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *J. Geophys. Res. Atmos.*, **112**, https://doi.org/10.1029/2007JD008674.

——, and ——, 2007b: An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *J. Geophys. Res. Atmos.*, **112**, 1–14, https://doi.org/10.1029/2007JD008674.

——, and Coauthors, 2011a: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res. Atmos.*, **116**, https://doi.org/10.1029/2010JD015139.

——, and Coauthors, 2011b: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res. Atmos.*, **116**, 1–19, https://doi.org/10.1029/2010JD015139.

——, and Coauthors, 2011c: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res. Atmos.*, **116**, https://doi.org/10.1029/2010JD015139.

NOHRSC (National Operational Hydrologic Remote Sensing Center), 2004: Snow Data Assimilation System (SNODAS) data products at National Snow and Ice Data Center (NSIDC), Version 1. *Snow Data Assim. Syst. Data Prod. NSIDC, Version 1*, 1–13, https://doi.org/10.7265/N5TB14TC.

Ogden, F. L., 1997: Mathematical Models of Small Watershed Hydrology and Applications - Google Books.

Rawls, W. J., D. L. Brakensiek, and N. Miller, 1983: Green-ampt Infiltration Parameters from Soils Data. *J. Hydraul. Eng.*, **109**, 62–70, https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(62).

Schaake, J. C., V. I. Koren, Q.-Y. Duan, K. Mitchell, and F. Chen, 1996: Simple water balance model for estimating runoff at different spatial and temporal scales. *J. Geophys. Res. Atmos.*, **101**, 7461–7475, https://doi.org/10.1029/95JD02892.

Senatore, A., G. Mendicino, D. J. Gochis, W. Yu, D. N. Yates, and H. Kunstmann, 2015: Fully coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *J. Adv. Model. Earth Syst.*, **7**, 1693–1715, https://doi.org/10.1002/2015MS000510.

Silver, M., A. Karnieli, H. Ginat, E. Meiri, and E. Fredj, 2017: An innovative method for determining hydrological calibration parameters for the WRF-Hydro model in arid regions. *Environ. Model. Softw.*, **91**, 47–69, https://doi.org/10.1016/j.envsoft.2017.01.010.

Singh, J., H. V. Knapp, and M. Demissie, 2004: *Hydrologic Modeling of the Iroquois River Watershed Using HSPF and SWAT*.

Slater, L. J., G. Villarini, A. A. Bradley, and G. A. Vecchi, 2019: A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed. *Clim. Dyn.*, **53**, 7429–7445, https://doi.org/10.1007/s00382-017-3794-7.

Somos-Valenzuela, M., and R. Palmer, 2018: Use of WRF-Hydro over the Northeast of the US to Estimate Water Budget Tendencies in Small Watersheds. *Water*, **10**, 1709, https://doi.org/10.3390/w10121709.

Staff, N., National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2020). https://doi.org/10.25921/stkw-7w73.

Survey, staff S., 1998: Description of STATSGO2 Database | NRCS Soils. *Nat. Resour. Conserv. Serv. United States Dep. Agric.*,.

Verri, G., N. Pinardi, D. Gochis, J. Tribbia, A. Navarra, G. Coppini, and T. Vukicevic, 2017: A meteo-hydrological modelling system for the reconstruction of river runoff: the case of the Ofanto river catchment. *Nat. Hazards Earth Syst. Sci.*, **17**, 1741–1761, https://doi.org/10.5194/nhess-17-1741-2017.

Verseghy, D. L., N. A. McFarlane, and M. Lazare, 1993: Class—A Canadian land surface scheme for GCMS, II. Vegetation model and coupled runs. *Int. J. Climatol.*, **13**, 347–370, https://doi.org/10.1002/joc.3370130402.

Wijayarathne, D. B., and P. Coulibaly, 2020: Identification of hydrological models for operational flood forecasting in St. John's, Newfoundland, Canada. *J. Hydrol. Reg. Stud.*, **27**, 100646, https://doi.org/10.1016/j.ejrh.2019.100646.

Williams, G., 1975: A study of flash-flood susceptibility--a basin in southern Arizona.

Xue, Z., and Coauthors, 2018: Modeling Hydroclimatic Change in Southwest Louisiana Rivers. *Water*, **10**, 596, https://doi.org/10.3390/w10050596.

Yang, R., and M. A. Friedl, 2003: Modeling the effects of three-dimensional vegetation structure on surface radiation and energy balance in boreal forests. *J. Geophys. Res. D Atmos.*, **108**, https://doi.org/10.1029/2002jd003109.

Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner, and A. Stewart, 2008: Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resour. Res.*, **44**, https://doi.org/10.1029/2007WR005940.

Yeh, W. W.-G., L. Becker, and R. Zettlemoyer, 1982: Worth of Inflow Forecast for Reservoir Operation. *J. Water Resour. Plan. Manag. Div.*, 257–269.

Yucel, I., A. Onen, K. K. Yilmaz, and D. J. Gochis, 2015: Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *J. Hydrol.*, **523**, 49–66, https://doi.org/10.1016/j.jhydrol.2015.01.042.

Zalenski, G., W. F. Krajewski, F. Quintero, P. Restrepo, and S. Buan, 2017: Analysis of National Weather Service Stage Forecast Errors. https://doi.org/10.1175/WAF-D-16-0219.1.

APPENDIX A

SUPPLEMENTAL FIGURES FOR MODEL SPIN-UP

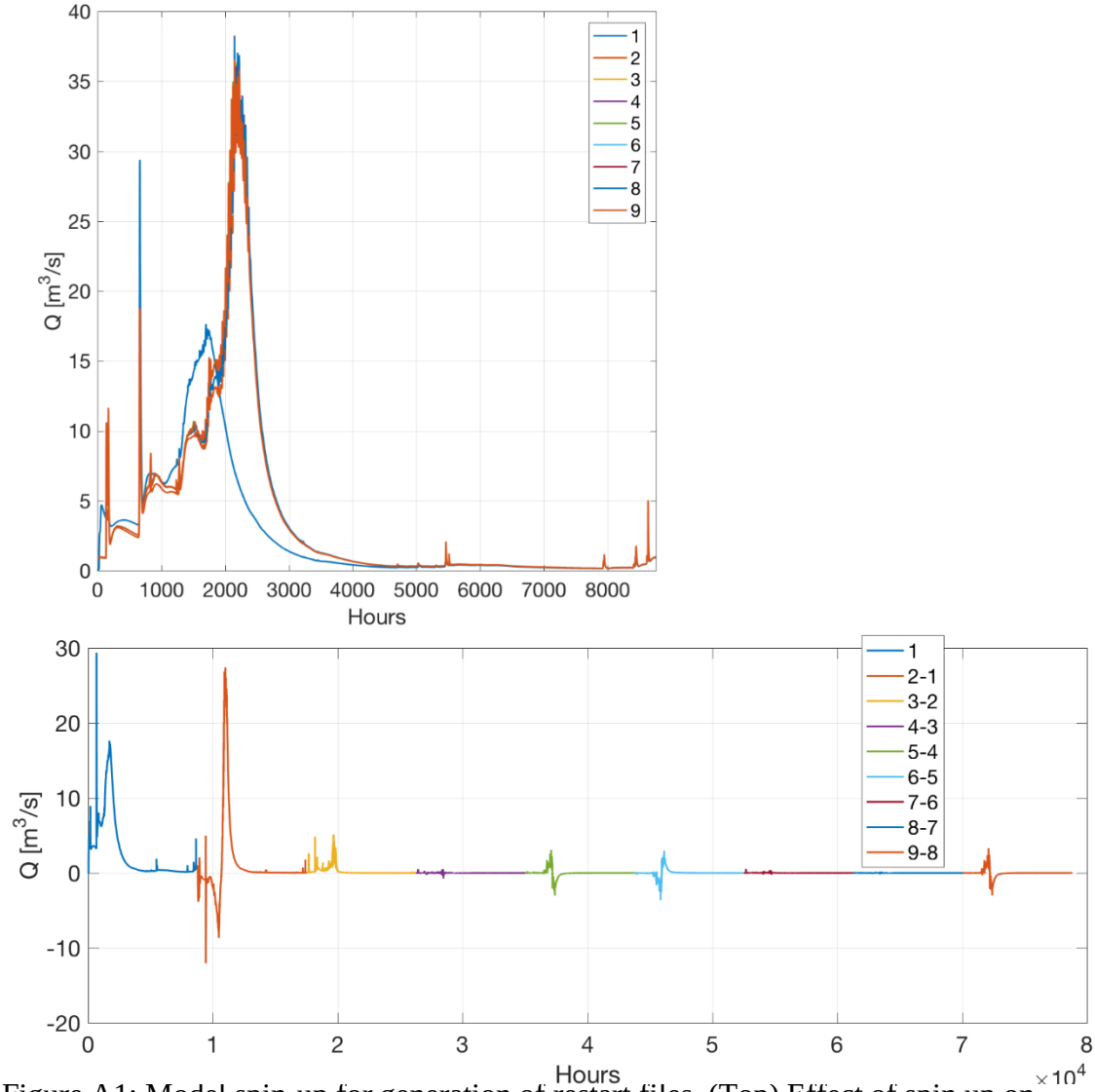

Figure A1: Model spin-up for generation of restart files. (Top) Effect of spin up on discharge in 2008 and (bottom) the relative bias between consecutive simulation. Note that the numerical stability of the model increases with every simulation till the stability is reach in the 10th year of simulation.

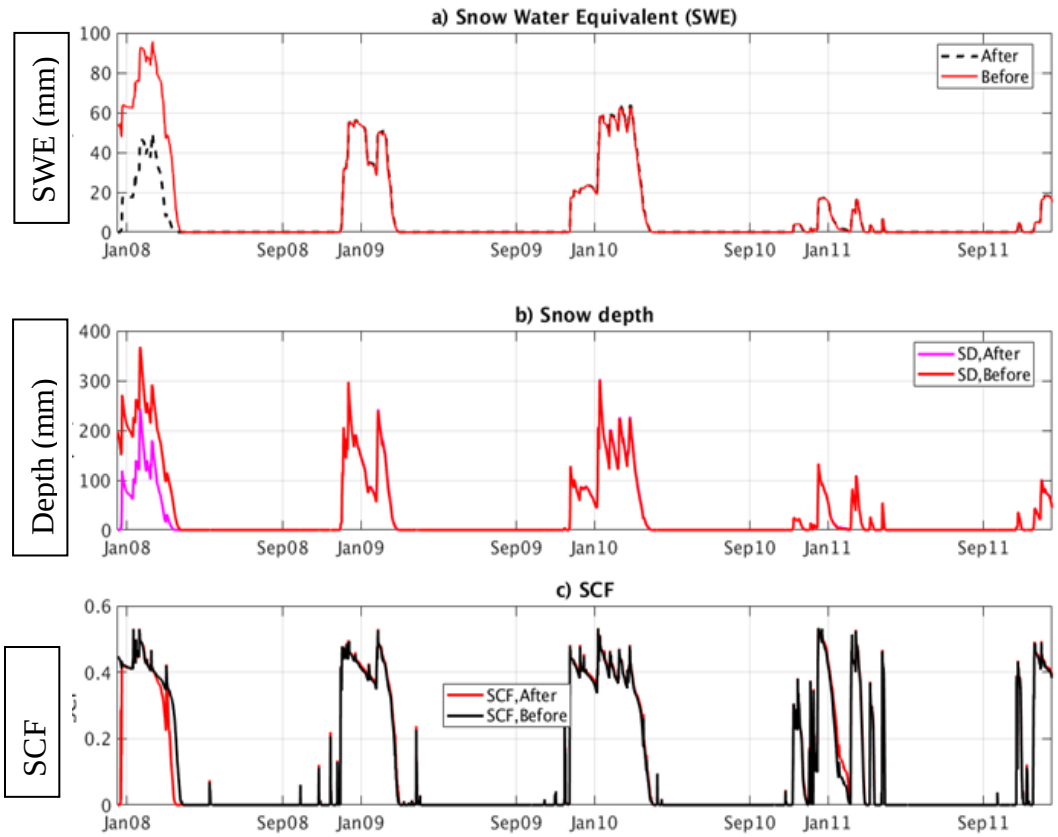

Figure A2: Effect of spin up on snow state variables. Note that both SWE and snow depth is severely under simulated without the spin-up (dashed black line).

APPENDIX B

SNOW DYNAMICS AT POINT SCALE

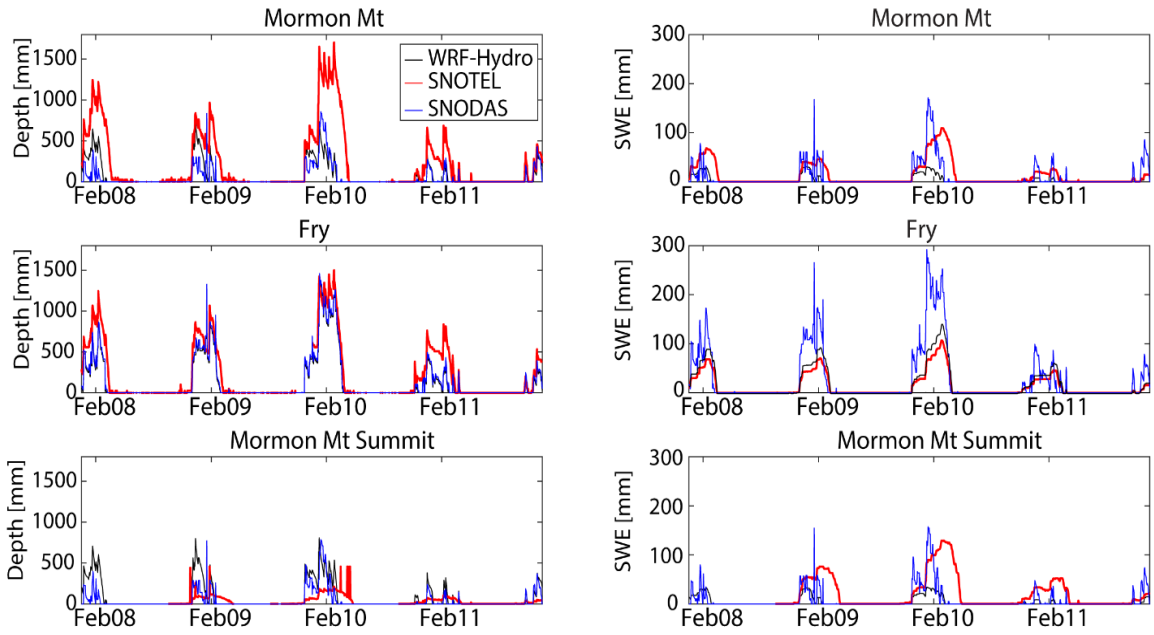

Figure B1: Time series of observed snow depth and SWE for the three SNOTEL stations, corresponding co-located SNODAS values, and simulated snow depth and SWE for the period 2008-2011. Note that WRF-Hydro model under-simulates snow depth and SWE at point scale compared to SNODAS and SNOTEL observations.

APPENDIX C

SUPPLEMENTAL FIGURES FOR BIAS-CORRECTION

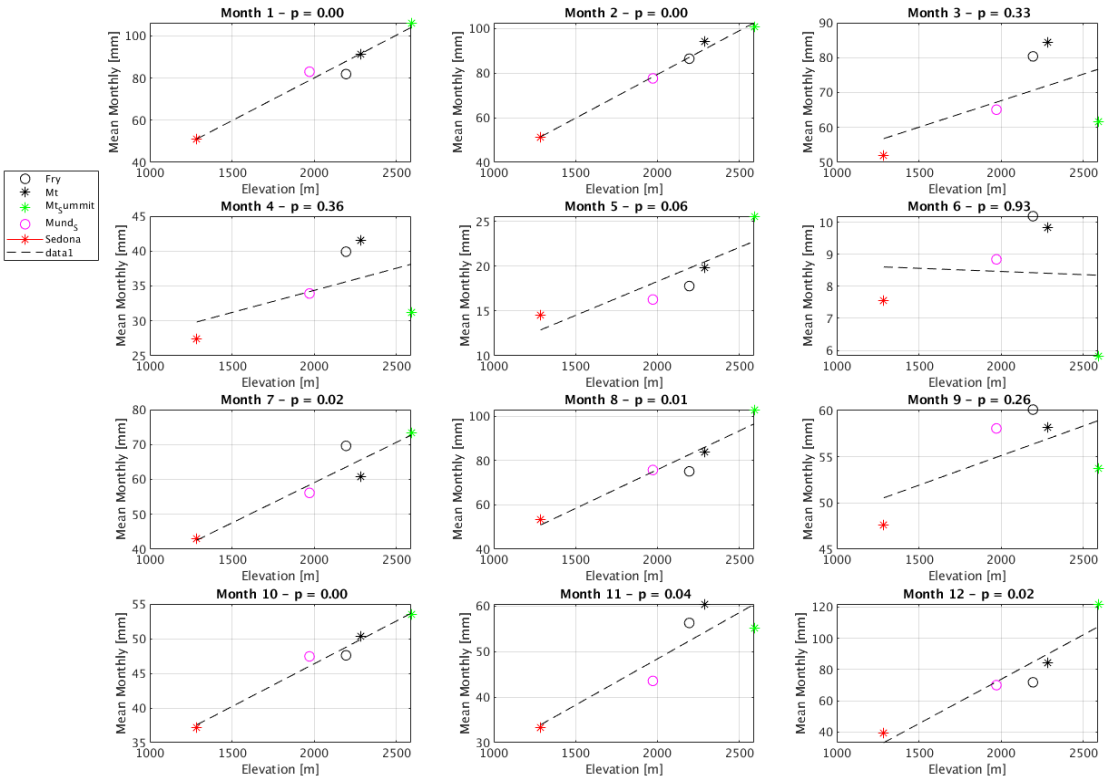

Figure C1: Climatological monthly mean precipitation for all stations showing the existence of linearity with elevation in all months except June. Statistically significant p-value was found April, May, June, and September. The climatological monthly mean precipitation values in all these months is comparatively small compared to other values showing that the rainfall is less in these months.

94