Inter-Model Diagnostics for Two Snow Models and Implications for Management

by

Elizabeth S. Houle

B.S., University of Virginia, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master of Science

Department of Civil, Environmental and Architectural Engineering

2015

This thesis entitled:
Inter-Model Diagnostics for Two Snow Models and Implications for Management
written by Elizabeth S. Houle
has been approved for the Department of Civil, Environmental and Architectural Engineering

_____

Prof. Joseph Kasprzyk

_____

Dr. Ben Livneh

_____

Prof. Edie Zagona

_____

Prof. Balaji Rajagopalan

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Houle, Elizabeth S. (M.S., Civil Engineering)

Inter-Model Diagnostics for Two Snow Models and Implications for Management

Thesis directed by Prof. Joseph Kasprzyk

Water supply in the western United States is dominated by snowmelt, and as a result water management is increasingly reliant on numerical modeling of snowmelt processes, including snow accumulation and ablation. We seek to advance a framework for providing model diagnostics for such systems by combining an improved understanding of model structural differences (i.e., conceptual vs. physically based) and parameter sensitivities. The two snow models used in this study are SNOW-17, a conceptual degree-day model, and the Variable Infiltration Capacity (VIC) snow model, which is physically based and solves the full water and energy balances. To better understand the performance of these models, global sensitivity analysis and multi-objective calibration methods were applied to identify important parameters and show calibrated parameter values. For the physically based model, we contribute a novel exploration of some parameters that can be adjusted within the model, including the liquid water holding capacity, the density of newly fallen snow, snow roughness, and snow albedo decay parameters. For each model run, snow sensitivities and errors (i.e., snow water equivalent validation) are visualized to better understand the effect of changing parameters on model outputs. The sensitivity analyses and multi-objective calibrations resulted in model parameterizations that produced Nash-Sutcilffe Efficiency values up to 0.88 through 0.95 across all site locations. Additionally, a temperature change analysis was conducted for each model to explore how model parameterizations affect portrayals of climate change. Accurately predicting water yield from snowpack is essential for water management, and it is used here as a practical measure to determine the importance of model parameter sensitivity and calibration. The analysis was conducted across a range of snow-dominated locations representing a variety of climates across the western United States (e.g. continental, maritime, intermountain).

Dedication

To women in science and engineering.

Acknowledgements

I thank Dr. Joseph Kasprzyk for remaining flexible and open-minded in working with my interests and skills to define a project in which I am truly interested. I appreciate his support, enthusiasm, and mini lectures in which I have learned my most valuable graduate school lessons. I would also like to express gratitude to Dr. Ben Livneh, mentor and guitarist extraordinaire, who agreed to work with me and take time out of his busy days to answer my emails and attend weekly meetings regarding my work. Thank you Dr. Balaji and Dr. Zagona for serving on my committee and attending weekly group research meetings to provide support and feedback on students' research. Thank you to Dr. Jon Herman for assistance with the MATLAB Borg wrapper.

The support of my parents has been critical to my graduate school success and experience. Education is a privilege, and words cannot express my gratitude for the opportunities for which I have been granted. I also want to thank my friends, grad school and non-grad school, for keeping me sane throughout the graduate school experience. I would like to extend a special thanks to Abby Watson and Mark Schutte with whom I've spent countless hours working and socializing throughout the past two years. The two of you have become lifelong friends.

Lastly, I want to thank Janus, the supercomputer, which is supported by the NSF (CNS-0821794) and CU Boulder. The Janus supercomputer is a joint effort of CU Boulder, University of Colorado Denver, and the National Center for Atmospheric Research. I will keep my authenticator on my key-chain to remember the hours spent running code together.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

As much as 75 percent of the water supply in the western United States (U.S.) is derived from snowmelt (*USGS*, 2014), and snowmelt-dominated watersheds are especially sensitive to changes in temperature and precipitation (*Barnett et al.*, 2005). In these systems, reservoirs are used to store water coming from snowmelt, and management of the reservoirs is critically dependent on runoff timing. General Circulation Models (GCMs) used to predict temperature and precipitation changes on a global scale, generally agree that there will be rising temperatures under most climate change scenarios (*Barnett et al.*, 2005). However, the magnitude of the temperature increase and the extent of precipitation changes are variable across GCMs (*Knutti and Sedlacek*, 2012). Implications resulting from temperature change include large reduction in snowpack, insufficient reservoir storage to handle the projected shift in streamflows, and increased frequency and severity of droughts (*Barnett et al.*, 2005; *Mote et al.*, 2005).

Although there is consensus that temperature patterns will be modified under a changing climate, it is still an open question how these changes will propagate to impacts on water management. In order to adapt to these changes, managers must be able to model detailed hydrological processes under perturbed input conditions. In general, there are two broad categories of hydrological models used to make such predictions: conceptual and physically based hydrological models. A conceptual model represents the physical processes in a simplified and sometimes indirect form, whereas a physically based model represents physical processes by solving energy and mass balance equations. Regardless of model type, *Wagener* (2007) and *Mendoza et al.* (2015), among others,

highlight the need for an understanding of parameter sensitivity (i.e. how changes in the parameters lead to changes in the model performance), which is a step toward finding relationships between regional watershed characteristics and model parameters.

This study seeks to contribute to the body of work investigating the behavior of hydrological models. For example, *Koskela et al.* (2012) used Bayesian inference to study the effect of precipitation and model structural uncertainty on estimates of model parameters in a conceptual rainfall-runoff model coupled with a simple degree-day snow model over the western U.S. *Koskela et al.* (2012) hypothesized that informative priors for latent model variables could be replaced by additional data. However, the authors determined the additional data did not provide new information that could replace parameter uncertainty, implying that a more complex model is not necessarily more accurate. For example, a complex model could be overparameterized, which means some of the model's parameters do not significantly affect the model's output, and in affect those parameters are just adding unnecessary complexity to the model without making any improvements to the output. Parameter sensitivity analyses are useful in determining which parameters are the most sensitive and affect the output versus which parameters are not sensitive and could likely be ignored.

In addition to parameter sensitivity, model calibration is an important aspect of using hydrological models, finding the optimal fit between model output and observed data. By calibrating the model parameters, models can be used for predictive purposes. Additionally, model calibration can be complimented by sensitivity analysis (*van Werkhoven et al.*, 2009), since calibration gives the best values of parameters, whereas sensitivity analysis often only shows that a parameter is important but does not suggest its optimal value. Because there exists multiple, sometimes conflicting, measures of model fit, calibration is often performed with respect to multiple objectives, employing multi-objective evolutionary algorithms (MOEAs) to find multiple parameter values that balance the objectives (*Efstratiadis and Koutsoyiannis*, 2010). An example study of how calibration is used in snow models is *Manandhar et al.* (2013), which successfully calibrated and evaluated a hydrological model integrated with a simple degree-day snow accumulation and ablation model using the

Nash-Sutcliffe Efficiency (NSE) coefficient and the volume bias (VB), and comparing snow cover patterns in the model-generated snow cover map with a satellite-captured map.

Since snowmelt is the main source of water in the Western U.S., the ability to produce accurate projections of snow water equivalent (SWE), which is the amount of water contained within a snowpack, is critical to water resources management. This thesis explores two contrasting snow models, a simple conceptual model and a complex physically based model, that are used for operational and research purposes, respectively. The research includes three phases. The first phase uses a model parameter sensitivity analyses to address the question of whether complex physically based models perform as well or better than simple conceptual models. The second phase involves multi-objective model calibrations to address the question of how we identify well-performing parameter sets with respect to several, conflicting objectives. The third phase uses a climate change scenario to explore how errors in parameterizations propagate to portrayals of climate change.

1.1 Summary of Chapters

1.1.1 Chapter 2: Background

Chapter 2 provides the reader with background information on the principal topics in this thesis. The chapter reviews historical research related to model parameter uncertainty and multi-objective calibration and discusses the need for application of modern sophisticated diagnostic tools to hydrological models. Technical details on snow models and the snowpack energy balance are then given. Finally, the chapter defines the diagnostic tools used in this research, global sensitivity analysis and multi-objective calibration using a multi-objective evolutionary algorithm.

1.1.2 Chapter 3: Methods

Chapter 3 presents an analysis of two contrasting snow models at various locations across the western U.S. The chapter describes the selected site locations, data sources, and data manipulation

techniques. It also provides background on modeling with respect to climate change and explanation of the specific climate change methods used. The model setup and computational experiments are outlined to provide details regarding the specific experiments performed and the resources used for their execution.

1.1.3 Chapter 4: Results

Chapter 4 presents baseline model results, using default model parameter values, sensitivity analysis results, multi-objective calibration results, and temperature change scenario results. The first part of this study explores the ensemble of model outputs created by varying model parameters, and examines which parameters have the greatest effect on model output variance. This involves varying historically-fixed parameters within the physically-based model, which is a technique that has not been extensively researched. The second part of this study executes a multi-objective calibration of the two snow models to further examine the optimal parameter sets, and the values of those parameters. The third part of this study focuses on a temperature change scenario applied to both snow models in an attempt to gain insight on how model parameterizations affect climate change portrayals.

1.1.4 Chapter 5: Discussion, Conclusions, and Future Work

Chapter 5 discusses the findings of this research, implications of this study for water resources management, and paths for this research to be extended into the future.

Chapter 2

Background

2.1 Need for Improved Model Diagnostics and Understanding

Hydrologic processes have been modeled with a range of complexity, from a simple conceptual bucket model (*Manabe*, 1969), moving towards physically based models, which represent processes by solving energy and mass balance equations, in recent decades. Development of complex physically based models, which represent processes with physics-based equations, in theory, should result in improved model performance. However, research has demonstrated that model complexity is not directly correlated to the quality of simulation results (*Reed et al.*, 2004; *Smith et al.*, 2012; *Mendoza et al.*, 2015). In fact, research into poor model performance, improvements to hydrological modeling, and reduction in uncertainty within modeling has been largely studied over the past twenty-five years. *Bergstrom* (1991) emphasized that there should not be a conflict between the use of conceptual and physically based models because they were developed for different purposes, including operational use and research, respectively. However, despite the type of model, the user must be aware of the model's applicability, weaknesses, and uncertainties. *Bergstrom* (1991) also touched on the idea of parameter uncertainty and emphasized that good model performance does not necessarily mean a model is describing the system correctly. This conclusion launched a surge of studies that attempted to gain an understanding of model structure and model parameters and their effects on model outputs. *Beven* (1993, 1996a,b, 2000, 2002) expanded on Bergstrom's theory of the importance of model parameters. *Beven* argued that an optimal parameter set can always be found, in that the parameter set yields a model run with the best fit with respect to an objective

function. However, the optimal parameter set may not prove to be robust with respect to different datasets, and thus the parameter set is right for the wrong reasons. Additionally, *Beven* (1993) introduced the concept of equifinality, which is the idea that there could be two or more parameter sets that have the same performance, but are drastically different. In other words, for a given application, we must decide which model parameterizations adequately describe of the system under study. This can be aided with tools such as sensitivity analyses and multi-objective calibration methods. He also discussed the possibility of limiting the uncertainty of parameters, and further explored the idea of model parameter uncertainty by developing and discussing new methods of refining, limiting, and quantifying parameter uncertainty.

While *Bergstrom* (1991) and *Beven* (1993, 1996a,b, 2000, 2002) addressed the issue of eliminating parameter uncertainty in model performance, *Gupta* (1998) and *Boyle et al.* (2000) advanced the concept of improving model performance and diagnostics by viewing model calibration as a multi-objective problem. This means there is no one unique best parameter set, and rather, there are many optimal parameter sets that perform well with respect to different desired objectives. *Gupta* (1998) discusses how the classical approach to model calibration has limitations, especially with the development of more complex models, because they do not recognize that model identification and calibration are inherently multi-objective problems. Many automatic calibration studies have used simple conceptually based models (e.g., *Duan et al.*, 1994) due to computational limitations. However, *Gupta* (2008 and 2012) discussed how the introduction of complex physically based models resulted in increased complexity of calibration. *Gupta* (2008 and 2012) recognized the need for sophisticated approaches to model evaluation and introduced a consistent and systematic framework for assessment of model structural adequacy. *Wagener et al.* (2009) similarly stressed that the development of more complex models requires more complex model diagnostic tools, specifically global sensitivity analysis, which can indicate how the uncertainty in a model output can be allocated to different model characteristics, and multi-objective calibration methods, which result in a grouping of potential parameter sets for a model, thereby providing information regarding tradeoffs among different model parameters. Researchers have recognized the need for

more sophisticated diagnostic tools when using more sophisticated models, however, there has been limited work on the actual application of these advanced tools to sophisticated models.

Broadly, the discussion of parameter values is a subset of the general uncertainty that is associated with hydrological modeling. There is wide literature on uncertainty in hydrological modeling, which is briefly reviewed here. *Uhlenbrook et al.* (1999) investigated uncertainties that arise from identifying a model structure and model parameters for a conceptual model. They concluded good results can be obtained with unrealistic concepts, and model prediction should be provided as ranges rather than single values. *Vrugt et al.* (2005) recognized the considerable uncertainty in model predictions as a result of model structure errors and data uncertainty. They presented a global optimization and data assimilation method, simultaneous optimization and data assimilation (SODA), to simultaneously improve treatment of input, output, parameter, and model structural uncertainty in hydrological model calibration. *Kuczera et al.* (2006) developed a framework to address sources of error, including uncertainty in data, for a conceptual rainfall-runoff model. The authors performed a simple sensitivity analysis and concluded uncertainty in rainfall data dominated model uncertainty. *Thyer et al.* (2009) assessed the ability of the Bayesian total error analysis (BATEA) approach and other standard calibration approaches to address the quantification of predictive uncertainty and estimation of parameter uncertainty.

Despite the fact that researchers recognize the need for more sophisticated model diagnostic tools (*Clark et al.*, 2008, 2011), there nevertheless has been limited research in applying these tools primarily due to (1) lack of awareness of the scientific importance of multiple diagnostics and (2) limitations in computing power, such as lack of access to powerful supercomputers that have the ability to run complicated models and programs that take hours to compute. Advancements in computing power and resources allow us to sample from high dimensional spaces and relax limiting assumptions of models (*Wagener et al.*, 2009). Another reason for limited work on complex model parameter uncertainty is that the process of adjusting parameters is labor intensive because it typically requires manual alteration of lines of code and recompiling of the models (*Mendoza et al.*, 2015). Now that modern computing power is sufficient to run sophisticated models and model

evaluation methods, research regarding parameter uncertainty and sensitivity can advance at an accelerated pace (*Gupta*, 1998, 2008; *Kirchner*, 2006; *Wagener et al.*, 2009).

Some investigations have used sensitivity analyses to obtain information regarding model parameter sensitivity and uncertainty and model calibration. A range of models were involved in these investigations including a lumped hydrological model, a complex physically-based hydrological model, a conceptual distributed hydrological model, and a biosphere model. *Tang et al.* (2006) compared several local and global sensitivity analysis methods applied to the Sacramento Soil Moisture Accounting Model (SAC-SMA) coupled with SNOW-17 and determined the Sobol' sensitivity method provides robust and detailed sensitivity rankings. *Demaria et al.* (2007) applied a Monte Carlo sensitivity analysis to land surface parameters within the VIC model and concluded the base flow formulation within the model was overparameterized. These results demonstrated how complex models can be reduced to a more parsimonious form to improve modeling of ungauged basins. *Van Werkhoven et al.* (2008) used Sobol' sensitivity analysis to evaluate a conceptual distributed model, the National Weather Service's Hydrology Laboratory Distributed Hydrologic Modeling System (HL-DHMS). *Foglia* (2009) used error-based weighting of observation and prior information data, local sensitivity analysis, and single-objective nonlinear regression to evaluate the predictive ability of a distributed rainfall-runoff model. *Rosolem et al.* (2012) used Sobol' sensitivity analysis to evaluate parameter sensitivity within the Simple Biosphere 3 model and determine the controlling model parameters. Lack of sensitivity in some of the parameters indicated possible overparamaterization of some model processes. *Mendoza et al.* (2015) took initial steps to explore parameter sensitivity within complex models. They examined the impact of fixed model parameters in the Noah Multi-Physics (Noah-MP) Land Surface Model (Noah-MP) and concluded that some of the hard-coded parameters showed very high sensitivity, and calibrating these parameters resulted in improved model accuracy.

Physically based models, in theory, should result in improved performance because they are meant to accurately reflect processes using fundamental physics. However, studies have showed physically based models actually perform equally or only slightly better than their simpler con-

ceptual counterparts (*Reed et al.*, 2004; *Smith et al.*, 2012; *Mendoza et al.*, 2015). For example, the distributed model intercomparison project (DMIP) and DMIP2, which compared conceptual lumped and distributed watershed model simulations generated with uncalibrated and calibrated parameters, concluded the distributed models did not greatly outperform the conceptual models (*Reed et al.*, 2004; *Smith et al.*, 2012). As a result, model frameworks and competing model elements have grown into an area of increased research and interest (*Mendoza et al.*, 2015). For example, *Pomeroy et al.* (2007) developed the cold regions hydrological model (CRHM) to explore different parameterizations of a cold region's hydrological cycle. *Clark et al.* (2008) developed the Framework for Understanding Structural Errors (FUSE) to test different soil parameterizations found in widely used hydrological bucket models. *Niu et al.* (2011) developed the Noah-MP framework to test biophysical and hydrological processes in an effort to improve the Noah land surface model (LSM). These modeling frameworks lack experimentation with hard-coded parameters within the models, which provide an additional source of uncertainty (*Mendoza et al.*, 2015).

In summary, construction of more complex models must be accompanied by development of more powerful sophisticated identification and evaluation tools, such as combining sensitivity analysis and optimization methods to constrain uncertainty (*Wagener et al.*, 2009). Improving hydrological models requires developing effective methods to define and discriminate among competing modeling options within model structure and parameters.

2.2 Snow Models

Most snow models have been developed to simulate snow accumulation and ablation processes and to ultimately predict snow water equivalent (SWE) (*He et al.*, 2011). Simple conceptual snow models were originally developed for operational purposes to aid in runoff forecasts, but an increased need to analyze environmental issues and extreme events has prompted the creation of more complex physically-based snow models based on conservation of mass and energy equations. However, this evolution in snow modeling brings about additional uncertainties such as landscape heterogeneity, scale issues, and large numbers of model parameters (*Kirnbauer et al.*, 1994). There

exist a range of snow models with varying complexities, making it difficult to determine the most appropriate model for a specified application. For example, a simple temperature-index model only uses air temperature as an index for estimating snow processes, while a physically based model solves energy and mass conservation equations (*Kumar et al.*, 2013). As a result, snow model frameworks have been developed to study snow model components. *Essery et al.* (2013) developed the Joint UK Land Environment Simulator (JULES) Investigation Model (JIM) to test different simulations of snow processes. Snow Model Intercomparison Projects (SnowMIP) modeled specific snow accumulation and ablation processes using many snow models that differ in complexity and that have varying abilities. In comparing the snow models, SnowMIP concluded that there is no clear link between model performance and model complexity (*Essery et al.*, 2013). According to SnowMIP, model performance is highly dependent on model application (*Etchevers et al.*, 2004). Successful implementation of a selected snow model relies on model application and accuracy of model parameters (*Franz et al.*, 2008; *He et al.*, 2011).

2.2.1 Snowmelt and the Snowpack Energy Budget

A basic understanding of snowmelt and the snowpack energy balance is necessary to understand the different ways of modeling such processes. There are three basic phases of snowmelt: warming phase, ripening phase, and output phase. The snowmelt process requires absorption of energy of an amount equal to the sum of the energy absorption associated with the three phases. These phases occur via the absorption of energy from the sun, the earth and atmosphere, and precipitation.

The warming phase occurs when absorbed energy raises the temperature of the snowpack to a point at which the snowpack is isothermal. The amount of energy required for the snowpack to become isothermal is the cold content, which depends on the density of water, the temperature of the snowpack, and SWE. The ripening phase occurs when energy is used to melt snow, but the water is retained in the pore spaces of the snowpack. Once a snowpack cannot retain additional liquid, it is ripe. The output phase is when additional absorption of energy and snowmelt result

in the output of water from the snowpack. These processes do not occur in perfect sequence. For example, the snow surface can melt and percolate through the snowpack pores where it refreezes and creates a denser snowpack, which continues to melt and refreeze many times before output occurs.

In a temperature index model, temperature is used as the sole index to determine the energy exchange across the air-snow interface. Most important physical processes that take place within the snowpack are explicitly included, but in simplified forms, using temperature (*Anderson*, 2002, 2006). For example, Equation 2.1 describes melt during non-rain periods in the conceptual temperature-index SNOW-17 model. Melt occurs when the air temperature (T_a), is above a base value ($MBASE$); otherwise, melt (M) does not occur.

$$M = M_f(T_a - MBASE) \tag{2.1}$$

where M is the depth of melt (mm), M_f is the seasonally varying melt factor (mm/°C), T_a is the air temperature (°C), and $MBASE$ is the temperature above which melt will not occur (°C) (*Anderson*, 2006).

In a physically based model, these processes are accounted for via mass and energy balance equations. Below is the energy budget equation for a snowpack

$$Q_{net} = K + L + Q_h + Q_e + Q_g + Q_p + \Delta Q/\Delta t \tag{2.2}$$

where Q_{net} is the net energy available to the pack, K is net shortwave radiation at the surface of the pack, L is net longwave radiation at the surface, Q_h is the surface sensible heat flux, Q_e is the surface latent heat flux, Q_g is ground heat flux, Q_p is the energy advected to the pack from external sources (e.g. rain), and $\Delta Q/\Delta t$ is the change in internal energy within the pack, which is equivalent to the cold content divided by the time step. Fluxes are positive when directed into the pack. This equation is only solved for process-based models (e.g. VIC) (*Livneh et al.*, 2010). Figure 2.1[1] depicts this snowpack energy balance.

[1] `http://www.meted.ucar.edu/hydro/basic_int/snowmelt/`

Each component of the energy balance will now be briefly described. Shortwave radiation is energy from the sun and is affected by time of year, latitude, slope, aspect, cloud cover, and mainly surface albedo. Longwave radiation is energy emitted from the Earth and atmosphere and is affected by temperature and atmospheric emissivity, which increases with increasing water vapor. Outgoing longwave radiation usually exceeds incoming longwave radiation because the earth's surface is warmer than the atmosphere and has a daily cycle of warming in the afternoon (*Ahrens*, 2012). The change in internal energy within the pack is dependent on the cold content of the pack, which is the amount of energy required to bring the pack to a temperature of 0 °C (*Livneh et al.*, 2010). Sensible and latent heat fluxes are the exchange of energy between the earth's surface and atmosphere due to a temperature gradient and a water vapor gradient, respectively. These fluxes are affected by wind and radiation.

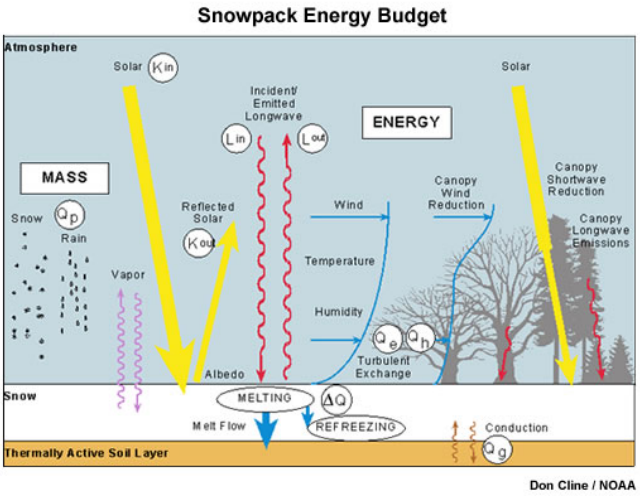


Figure 2.1: Mass and energy balance components of a snowpack. The components in the figure are defined in Equation 2.3.

Albedo has a large impact on the amount of incoming solar radiation, which is a main driver of snowmelt during the spring season. The impact of solar radiation on melt energy is generally of increasing importance with decreasing latitude, as a result of greater radiation moving equatorially (*Livneh et al.*, 2010). Albedo is the fraction of incident solar radiation reflected by the snowpack, and it usually dominates the snowpack energy balance during the melt season, which is the period

during which runoff from snowmelt is produced, typically between April and July in the western U.S. (*Warren and Wiscombe*, 1980; *Warren*, 1982; *Livneh et al.*, 2010). However, it is difficult to obtain representative values of snow albedo because physical properties of the snow such as grain size, shape, age, liquid water content, solar zenith angle, cloud cover, snowpack thickness, and snow density all affect albedo decay (*Livneh et al.*, 2010). As time passes, the albedo of the snow decreases, which increases the amount of solar radiation absorbed and thus the melt rate. Different models represent the energy balance, albedo, and albedo decay in various ways. Therefore, this research explores parameters associated both with albedo and its decay function, to determine those parameters' contribution to model performance. The two distinct snow models used in this research are described below.

2.2.2 A Conceptual Temperature-Index Model

SNOW-17 is a conceptual snow accumulation and ablation model first developed by *Anderson* (1973) as a component of the United States National Weather Service (NWS) River Forecast System for use in river forecasting. The NWS has used SNOW-17 for decades to model snow accumulation and ablation in snow-dominated regions to create short- and long-lead stream flow forecasts (*Anderson*, 1973; *He et al.*, 2011). SNOW-17 is an index model that uses average daily air temperature as the only index to simulate heat storage of the snowpack, snowmelt, liquid water retention, and transmission to determine the energy exchange across the snow-air interface with empirically based relationships (*He et al.*, 2011). It also requires temperature and precipitation as input variables, while it outputs SWE and runoff time series (*Anderson*, 1973, 2006).

SNOW-17, when applied at a point location, is primarily controlled by 10 parameters. As previously mentioned, conceptual model parameters often cannot be measured in the field, and this is the case for SNOW-17. This research focuses on eight of these ten parameters: SCF, NMF, TIPM, MBASE, MFMIN, MFMAX, UADJ, and PLWHC. Table 2.1 provides a description of each SNOW-17 parameter accounted for in this research. The SNOW-17 parameters, PXTEMP (°C), which is the temperature that separates snowfall from rainfall, and the DAYGM (mm/d), which

characterizes the ground heat flux, were set to their default values for the purposes of this research.

The snow correction factor (SCF) parameter, which accounts for gauge catch deficiencies and losses during the accumulation season, is of particular interest in this research due to its control of precipitation adjustment in the model. Snow measurement error due to gauge under-catch, resulting from wind speed, type of gauge, and site location, is well-documented (*Rasmussen et al.*, 2012). As a result, SNOW-17 includes SCF as a multiplying factor that adjusts all new snow amounts to account for gauging errors before they are added to the existing snow cover using the following equation

$$P_n = P \cdot f_s \cdot SCF \tag{2.3}$$

where P_n is the water equivalent of new snowfall (mm), P is the total precipitation input to the model (mm), and f_s is the fraction of precipitation in the form of snow. SCF is an average value over all of the accumulation periods is typically used to calibrate the model. The value is usually chosen to give best estimate of the amount of water in snow cover at beginning of the melt season (*Anderson*, 2006).

Significant research regarding parameter sensitivities and estimation methods, model structural uncertainty, and forecast improvements of SNOW-17 exists (*Hogue et al.*, 2000, 2003, 2006; *Smith et al.*, 2003; *Etchevers et al.*, 2004; *Slater et al.*, 2006; *Tang et al.*, 2007; *Leisenring and Moradkhani*, 2011; *He et al.*, 2011, *Werner and Yeager*, 2012). For example, *Hogue et al.* (2000, 2006) introduced a multi-step automatic calibration scheme (MACS) for SNOW-17. MACS combined strengths of automatic and manual calibration procedures to improve and facilitate the SNOW-17 calibration process. *Hogue et al.* recognized limitations of MACS, and realized a multi-objective calibration procedure that allows for tradeoffs among objective functions was a more powerful method. Multi-objective calibration tools have since been developed, and this research uses these tools to further the ideas of *Hogue et al.* by combining sensitivity analysis and multi-objective calibration.

2.2.3 A Physically Based Distributed Model

The Variable Infiltration Capacity (VIC) model is a physically based, distributed, hydrological model that requires minimum and maximum daily temperature, precipitation, and wind speed as inputs while providing a wide range of outputs, including daily SWE, flow, water balance terms and energy balance terms. VIC was originally designed to couple with GCMs to project water resources under temperature and precipitation change scenarios. *Cherkauer et al.* (2003) realized the importance of cold season processes to the hydrologic cycle over much of the northern hemisphere and discusses more recent improvements to the representation of snow cover within VIC. The snow model component of VIC considers snow in several forms: ground snow pack, snow in vegetation canopy, and snow on top of lake ice (*Cherkauer et al.*, 2003). Only ground snowpack is considered in this research because data were collected in open areas that do not have canopy cover. The VIC ground snowpack consists of two layers, a thin surface layer and a thick deep pack layer, for which VIC solves the full mass and energy balances.

Energy exchange between the atmosphere and snowpack, as described in Equation 2.3, occurs only with the surface layer. The total energy available for refreezing liquid water or melting snowpack over a given time step depends on the net energy exchange at the snow surface. Net radiation at the snow surface is calculated given incoming shortwave and longwave radiation, which are determined via the maximum and minimum daily temperature and daily precipitation values. The sensible heat flux is determined via air density, specific heat of air, air temperature, and aerodynamic resistance between the snow surface and near-surface reference height. The latent heat flux is determined via the latent heat of vaporization when liquid water is present or the latent heat of sublimation in the absence of it, atmospheric pressure, vapor, and saturation vapor pressure. Advected energy to the snowpack is determined based on the specific heat of water, the depth of rainfall, temperature of precipitation, and the water equivalent of snowfall. If Q_{net} is negative, energy is being lost by the snowpack, and the liquid water is refrozen. If Q_{net} is sufficiently negative all liquid water within the pack will refreeze. If Q_{net} is positive, then excess

energy available after the cold content has been satisfied will produce snowmelt (*Andreadis et al.*, 2009).

Mass balance within the VIC snow model is dictated by the turbulent fluxes, where latent heat exchanges water with the liquid phase if liquid water is present in the snowpack, and sensible heat exchanges water with the ice phase in the absence of liquid water. If the water equivalent of ice exceeds the maximum thickness of the surface layer, then the excess, along with its cold content is distributed to the deeper layer. If the liquid water exceeds the liquid water holding capacity of the surface layer, the excess is drained to the pack layer. If the temperature of the pack layer is below freezing, the liquid water can refreeze. Liquid water remaining in the pack above the holding capacity is routed to the soil as pack outflow (*Andreadis et al.*, 2009).

VIC assumes snow albedo decays with age based on *U.S Army Corps of Engineers* (1956) (*Livneh et al.*, 2010). It uses two separate sets of coefficients for the accumulation and melt seasons, which are defined based on the absence and presence of liquid in the snow surface layer. For each season, the snow albedo, α_{snow}, from the decay rates is parameterized as

$$\alpha_{snow} = \alpha_{max} A^{t^B} \tag{2.4}$$

where α_{max} is the albedo value of fresh fallen snow, t is time since last snowfall in days, and the default values for A and B are 0.94 and 0.58 (0.82 and 0.46), respectively. Figure 2.2 is an example of albedo decay curves as used in the VIC model (*Livneh et al.*, 2010).

The three main parameters that are adjusted for cells without overlying vegetation are maximum temperature at which snowfall occurs, minimum temperature at which rainfall occurs, and snow surface roughness. *Chen et al.* (2014) explored the ability of six land surface models, including VIC, to simulate snowpack, but did not explore the sensitivity of SWE to the specification of various snow parameters. As a result, eight parameters that were suspected to affect model output were selected for this research. These parameters, described in Table 2.2, include: new snow density, snow roughness, liquid water holding capacity, new snow albedo, two accumulation period

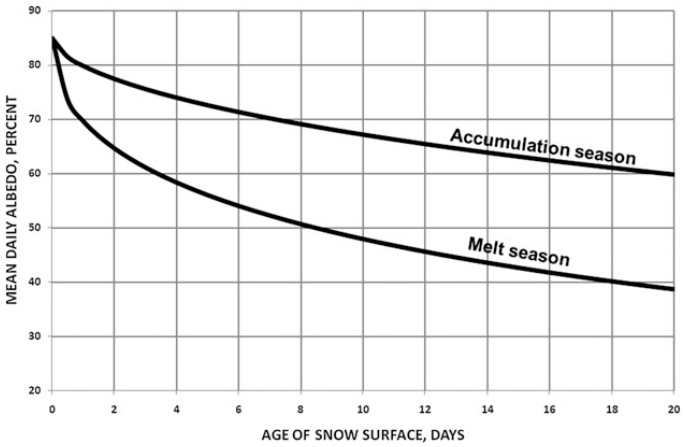


Figure 2.2: Snow albedo decay as a function of days since last snowfall (*Livneh et al.*, 2010). Equation 2.4 describes the curves shown here.

albedo decay curve parameters, and two thaw period albedo decay curve parameters.

Additional research involving VIC includes *Wood et al.* (1992), *Liang et al.* (1996, 2001 , 2003), *Parada et al.* (2003), *Kavetski et al.* (2006), *Demaria et al.* (2007), *Livneh et al.* (2010). There is limited research with respect to parameter sensitivity analyses on the VIC model, and research that does exist focuses on the soil parameters within the model. Notably, a recent study of VIC simulations of global flood risk used a Monte Carlo uncertainty analysis of some of VIC's hydrological parameters (*Reed et al.*, 2015).

2.3 Model Diagnostic Approaches

Mendoza et al. (2015) discusses the differences between *a priori* constraints on model parameter estimation for physically based models versus conceptual models. Physically based models have parameters that represent real-world values, and thus it is possible to impose *a priori* constraints on possible values for these parameters. Conceptual model parameters do not directly represent real-world values, and model equations are generalizations of processes, thus making it difficult to impose *a priori* constraints on these unknown values. As a result, conceptual model parameters are known to contain a large amount of uncertainty. However, physically based models use empirical functions in which the parameters may also contain uncertainty, which has not yet been extensively explored. *Mendoza et al.* (2015) emphasizes that exposing model parameter values

Table 2.1: Snow-17 parameter descriptions, default values, and ranges used in the Sobol' sensitivity analysis and multi-objective calibration. These parameter ranges were estimated from previous work including *Anderson* (1973), *Hogue et al.* (2000), *Franz et al.* (2008), and *He et al.* (2011).

				Range	
Parameter	Description	Units	Default Value	Min	Max
SCF	snow correction factor; accounts for gage catch deficiencies and losses during accumulation periods which controls the snowfall input into the model	--	1.0	**0.70**	**1.40**
UADJ	average wind function during rain-on-snow events; controls melt during rain-on-snow events	mm/mb/6hrs	0.04	**0.03**	**0.20**
MBASE	temperature above which melt typically occurs; controls snowmelt during non-rain periods	°C	1.00	**0.00**	**2.00**
MFMAX	maximum melt factor (June 21st); used to calculate melt factor which controls snowmelt during non-rain periods	mm/°C/6hrs	1.05	**0.50**	**2.00**
MFMIN	minimum melt factor (Dec. 21st); used to calculate melt factor which controls snowmelt during non-rain periods	mm/°C/6hrs	0.60	**0.05**	**0.50**
TIPM	antecedent temperature index parameter; controls heat exchange during the non-melt period	--	0.10	**0.05**	**1.00**
NMF	maximum negative melt factor; controls heat exchange during the non-melt period	mm/°C/6hrs	0.15	**0.05**	**0.50**
PLWHC	percent liquid water holding capacity of a well-aged snow cover; the amount of water a snowpack can hold before output from the snowpack occurs	%	0.04	**0.02**	**0.30**

Table 2.2: VIC parameter descriptions, default values, and ranges used in the Sobol' sensitivity analysis and multi-objective calibration. The new snow density, snow roughness, liquid water holding capacity, and new snow albedo values were estimated from a survey of literature. Albedo decay curve parameter ranges (A_{accum}, B_{accum}, A_{thaw}, B_{thaw}) were estimated by visual inspection of the sensitivity of the albedo decay curves with respect to the parameter values.

				Range	
Parameter	Description	Units	Default Value	Min	Max
Density	new snow density	kg/m^3	50	**25**	**150**
Roughness	snow roughness	m	0.03	**0.001**	**0.03**
LWHC	liquid water holding capacity	%	0.035	**0.02**	**0.15**
Albedo	new snow albedo	--	0.85	**0.80**	**0.90**
Accum. A	accumulation period albedo decay parameter	--	0.94	**0.30**	**0.99**
Accum. B	accumulation period albedo decay parameter	--	0.58	**0.30**	**0.99**
Thaw A	ablation period albedo decay parameter	--	0.82	**0.10**	**0.99**
Thaw B	ablation period albedo decay parameter	--	0.46	**0.10**	**0.99**

to users is crucial for future model development and improvement.

In the past several decades, a variety of diagnostic tools have been developed and applied within the hydrological community. These methods, summarized in *He et al.* (2011), include the maximum likelihood algorithm (*Restrepo and Bras*, 1985), the generalized likelihood uncertainty estimation (GLUE) (*Beven and Binley*, 1992), the shuffled complex evolution (SCE-UA) algorithm (*Duan et al.*, 1994), the multistep automatic calibration scheme (MACS) (*Hogue et al.*, 2000, 2006), the Bayesian recursive estimation (BARE) (*Thiemann et al.*, 2001), the Parameter Identification Methods Based on the Localization of Informatin (PIMLI) (*Vrugt et al.*, 2002), the Dynamic Identifiability Analysis (DYNIA) (*Wagener et al.*, 2003), and Markov chain Monte Carlo methods (*Kuczera and Parent*, 1998; *Vrugt et al.*, 2003, 2008a, 2008b).

2.3.1 Sensitivity Analysis

Sensitivity analyses have been applied to hydrological models to identify important parameters within a model and assist with parameter estimation and calibration (*Tang et al.*, 2007; *Foglia et al.*, 2009; *Wagener et al.*, 2009a, 2009b; *van Werkhoven et al.*, 2009; *Rosolem et al.*, 2012; *Rakovec et al.*, 2014). Two types of sensitivity methods are available, local and global. Local methods consist of the derivative-based approach in which the model is only executed a few times, compared to a global sensitivity method in which an array of derivatives must be computed. However, the local approach cannot be used when the model input is uncertain because, unlike a global sensitivity analysis, it does not have the capability to explore the space of the input factors (*Saltelli et al.*, 2008). Global sensitivity analysis methods are applied to the two snow models to identify parameters and explore the sensitivity of model outputs with respect to changes in parameters. Sensitivity analysis seeks to find which parameters are most important in controlling a model's response. The Sobol' sensitivity analysis method used in this research is a global method that varies all of the model's parameters in predefined regions to quantify the amount that each of the k parameters contributes to the variance of the output, $V(Y)$, where the output, Y, is the objective function matching the modeled SWE to the observed SWE. This method is based on variance

decomposition

$$V(Y) = \sum_{i=1}^{k} V_i + \sum_{i<j}^{k} V_{ij} + \ldots + V_{1,\ldots,k} \tag{2.5}$$

where V_i is the partial variance representing the first-order effect of parameter X_i on the model output Y; V_{ij} is the partial variance representing the second-order effect of interactions between parameters X_i and X_j; $V_{1,\ldots,k}$ is the interaction effect of all k parameters. Contributions to output variance can be caused by variations in one parameter (first-order effects) or interactions of parameters with one or more other parameters (second-order effects). The sensitivity of each parameter is assessed based on its contribution. The total order effect is the most comprehensive measure of a parameter's sensitivity because it represents the summation of all variance contribution involving that parameter (*Tang et al.*, 2007; *van Werkhoven et al.*, 2009; *Rosero et al.*, 2010). For more information on Sobol's methods see *Sobol'* (1993). The first order effects are calculated via

$$S_i = \frac{V_i}{V} \tag{2.6}$$

The second order effects are calculated via

$$S_{ij} = \frac{V_{ij}}{V} \tag{2.7}$$

The total order effects are calculated via

$$S_{Tj} = S_{ij} + \sum_{j \neq i} S_{ij} + \ldots \tag{2.8}$$

Saltelli (2002) introduced a new strategy for the computation of total order sensitivity indices that requires 50% fewer model evaluations than the previous method. This study uses an implementation within the MOEA Framework[2] , to execute Sobol', which employs the Saltelli strategy and has been used in prior studies (see *Hadka and Reed*, 2012). The default number of re-samples

[2] `http://www.moeaframework.org`

in the MOEA Framework for the bootstrapping is 1,000. The inputs (model parameters) are independently and uniformly distributed. Model simulations, each consisting of different parameter sets, described in Tables 2.1 and 2.2, were performed for each sensitivity analysis experiment.

2.3.2 Multi-Objective Model Calibration

Multi-objective evolutionary algorithms (MOEAs) are often used for multi-objective calibration of hydrological models (*Efstratiadis and Koustoyiannis*, 2010). The approach uses multiple measures of model performance to fit the model parameters. MOEAs use population-based searches to develop tradeoffs among multiple, often conflicting objectives, in a single algorithm run. The purpose of an MOEA is to approximate a Pareto-optimal set of solutions, which implies there cannot be an improvement in one objective without degradation in another. The optimization process does not use weights between the objectives, and instead allows the user to find his own preferences between objectives (e.g., different calibration measures). In order for an MOEA to be successful, an MOEA should be able to solve problems and allow for control of the precision of the solution set in order to generate a reasonable number of solutions. We use the Borg MOEA in this work, an adaptive MOEA framework that uses multiple variation operators to modify itself based on problem properties (*Hadka and Reed*, 2013). Recent diagnostic tests of several state-of-the-art algorithms have shown the superior performance of the Borg MOEA framework (*Hadka and Reed*, 2012). *Tang et al.* (2006) and *Kollat et al.* (2012) performed previous research using Borg's predecessor , epsilon-nondominated sorted genetic algorithm II ($\epsilon - NGSAII$). *Tang et al.* analyzed the computational efficiency, accuracy, and ease-of-use of several evolutionary multi-objective optimization algorithms, including $\epsilon - NGSAII$. They determined $\epsilon - NGSAII$ to be one of the superior algorithms for hydrologic model calibration, with its strength in ease-of-use due to rapid convergence with high quality solutions. *Kollat et al.* used $\epsilon - NGSAII$ to optimize hydrological model parameters based on four objectives. They concluded in order for modern multi-objective calibration to provide meaningful results, users must be able to control the precision of objective function tradeoffs, which can be achieved via epsilon-dominance, a feature included in Borg.

2.4 Objective Functions

Model performance in this research is quantified using several quantitative performance metrics, chosen to reflect typical practice in hydrological studies. The objective functions are used both in the sensitivity analyses and multi-objective calibrations. Each objective function uses observed and modeled SWE directly or indirectly to determine how well the quantity and timing of modeled SWE matches the observed SWE. Although other objective functions could be employed in a study such as this, objective functions should be chosen to meet the modeling goals of a study. In some cases, results of a Sobol' analysis can be used to choose objectives (*Kasprzyk et al.*, 2012). The first objective function is a commonly used statistical metric, root mean squared error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(x_{1,t} - x_{2,t})^2}{n}} \tag{2.9}$$

where n is the number of timesteps (t), $x_{1,t}$, is the modeled SWE in timestep, t, and $x_{2,t}$ is the observed SWE in timestep t. Both models were run on a daily timestep in this research. RMSE can range from 0 to ∞, and its values are dependent on the units of the values being compared. The closer the RMSE value is to 0, the better the model performance. Because it uses the squared difference between modeled and observed data, the RMSE assists with fitting high SWE portions of the graph (*van Werkhoven et al.*, 2009).

The second objective function is the Nash-Sutcliffe Efficiency (NSE) index, which is another statistical metric based on least squares method and is defined as

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{n}(x_{2,t} - x_{1,t})^2}{\sum_{t=1}^{n}(x_{2,t} - \overline{x_2})^2} \tag{2.10}$$

where $\overline{x_2}$ is the mean observed SWE, and other variables are as defined previously. NSE efficiencies can range from $-\infty$ to 1, and an efficiency of 1 corresponds to a perfect match of modeled to observed SWE. An NSE of less than 0 indicates that the model is no better than using

the observed mean as a predictor (*Gupta et al.*, 2009). NSE is often used as a calibration metric, and its values are helpful because they can be compared across sites. RMSE and NSE are similar, and model runs with optimal RMSE values will often be the model runs with optimal NSE values, resulting in minimal tradeoffs between the two objective functions. Additional, non-statistical, objective functions specifically related to SWE include measures of the differences in maximum SWE accumulation per season (peak), and number of days when snow is on the ground (length). A day with SWE greater than 1 millimeter is defined as a day with snow on the ground (*Mizukami et al.*, 2014). The closer the peak and length values are to 0, the better the model performance. The peak and length objectives are defined as

$$\text{peak} = |(\max(x_1) - \max(x_2)| \tag{2.11}$$

$$\text{length} = |(\text{length}(x_{1,t} > 1mm) - \text{length}(x_{2,t} > 1mm)| \tag{2.12}$$

where x_1 is a matrix of daily modeled SWE for a single water year; x_2 is a matrix of daily observed SWE for the same water year. Remaining variables are as defined previously. A variety of objective functions were used because they account for different traits of a SWE curve, and one would choose to focus on different objective functions depending on the goal of the modeling.

Multiple years were modeled, and for each objective function we considered each water year separately. Therefore, each year has four objective function values (RMSE, NSE, peak, and length). The worst performing value for each objective function out of the water years was selected as the representative objective value for all of the water years, and thus the entire model run. This means the maximum RMSE, peak, and length values and the minimum NSE value out of the water years were chosen as the objective values for a single model run. This method eliminates the possibility of obtaining a good objective value for a model run that performs very well for some water years, but poorly for others, and ensures a model run is as good or better than its objective value for all water years.

Chapter 3

Methods

3.1 Study Locations and Data

The snow models in this research were applied to five study sites located at the headwaters of major U.S. watersheds and representing a range of snow-driven climates in the Western United States. Hart's Pass (Hart) in Washington and CSS Lab (CSS) and Squaw Valley (Squaw) in California all represent maritime climates, which are characterized as having a deep high density snowpack, frequent storms, and a mild regime. Galena in Idaho represents an intermountain climate, which is characterized as having an intermediate snowpack depth and mild temperatures, but fewer storms and rain than maritime climates. Copper Mountain (Copper) in Colorado represents a continental climate, which is characterized as having a shallow less dense snowpack, less frequent storms with periods of drought, and colder temperatures. Figure 3.1 depicts a map of the site locations.

Harts Pass is part of the Methow Watershed, which flows into the Upper Columbia River. Galena is part of the Big Wood River Watershed, which flows into the Snake River and eventually the Columbia River. The Columbia River is important for the water supply of populated areas such as Portland, Oregon. CSS Lab is part of the Upper Yuba Watershed, which flows southwest into the Sacramento River. Squaw Valley is in the Truckee Watershed, which flows into the North Fork American River. Both the Sacramento River and American River provide water to the populated Sacramento, California area. Copper Mountain is in the Blue River Watershed, which flows into the Upper Colorado River, providing water for Denver, Colorado and surrounding states.

Daily precipitation, air temperature, and SWE data were adapted from the Natural Resources Conservation Service SNOw TELemetry (SNOTEL) sites[1] . Daily wind data were adapted from the North American Research Re-analysis winds re-sampled to a $\frac{1}{16}^{\circ}$ data set[2] (*Livneh et al.*, 2013). The grid cells closest to the selected SNOTEL sites were chosen.

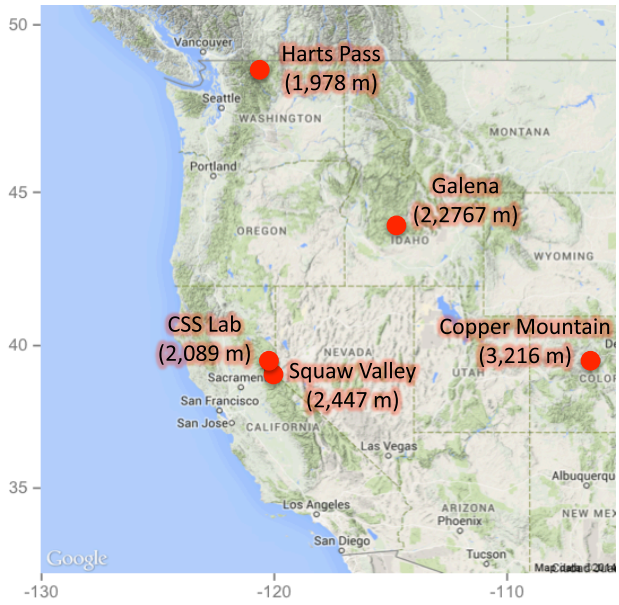


Figure 3.1: Map depicting the SNOTEL study site locations across the Western United States.

Inspection of data revealed missing or bad temperature data, which were filled using a linear regression model using valid data points surrounding the faulty data as predictors (*Livneh et al.*, 2014)[3] . Data for CSS Lab, Squaw Valley, Galena, and Copper Mountain were obtained for the 1998 through 2002 water years, and data for Hart's Pass were obtained for 2000 through 2003 water years. Different water years were selected for Hart's Pass due to large periods of missing data during the water years used for the other sites.

SWE and precipitation data were less problematic in terms of missing data; however, obtain-

[1] http://www.wcc.nrcs.usda.gov/snow/

[2] Wind data obtained here: `ftp://ftp.hydro.washington.edu/pub/blivneh/CONUS/asc.v.1.2.1915.2011.bz2/`

[3] This linear regression model code was implemented in MATLAB and provided by Dominik Schneider at the University of Colorado

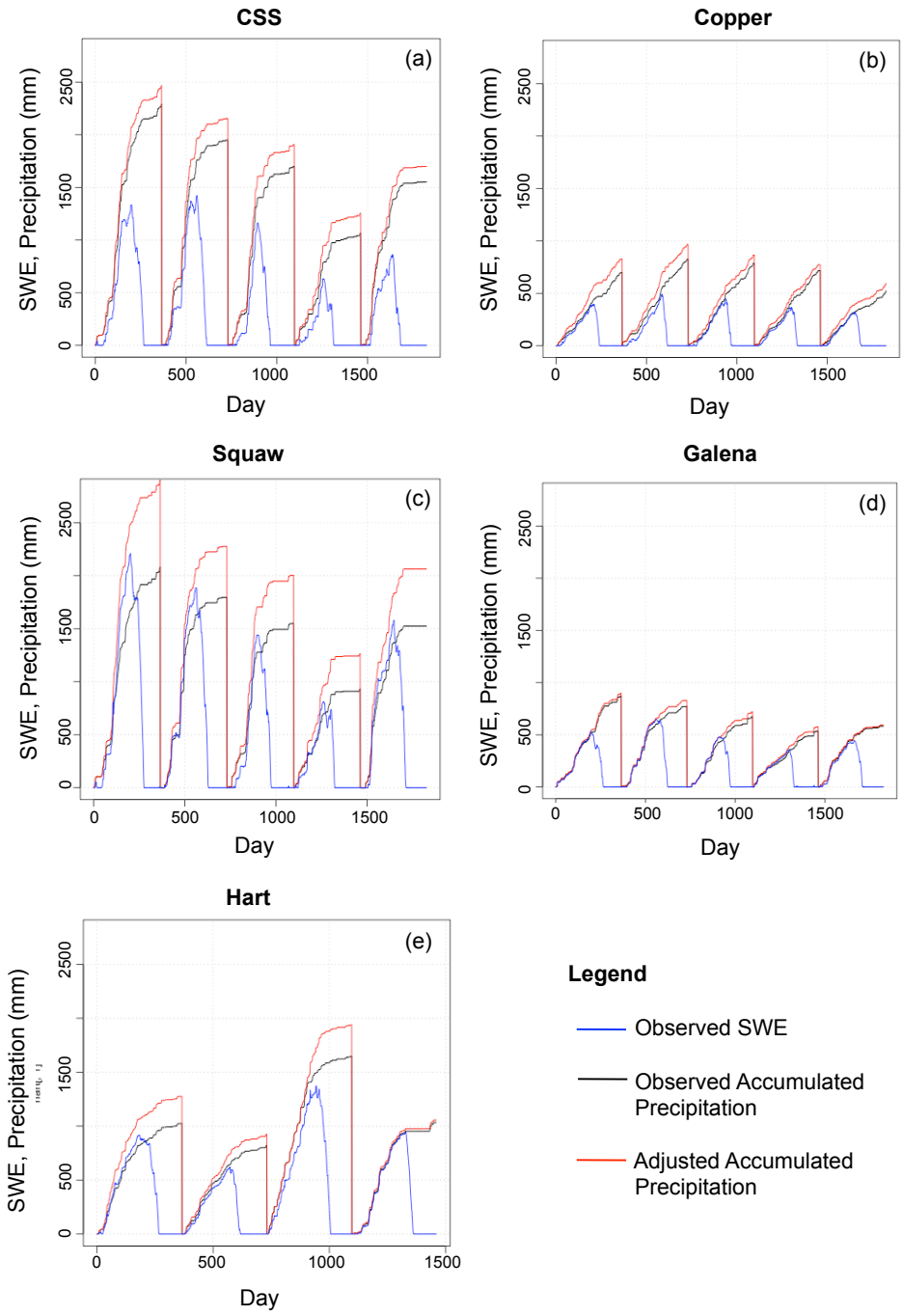


Figure 3.2: Adjusted precipitation for all of the site locations. Notice the observed SWE exceeded the observed accumulated precipitation for some time periods. As a result, it is impossible for a snow model to obey conservation of mass, producing inaccurate projections of SWE. SWE never exceeds the adjusted accumulated precipitation.

ing reliable precipitation estimates at high elevations can be difficult due to wind redistribution processes in snow-covered catchments, which affect snow accumulation. Under-catch of frozen precipitation can result in a dataset containing less precipitation than SWE, which is impossible. As a result, a model cannot provide accurate outputs because the conservation of mass cannot be preserved. Thus, a continuous time series of daily precipitation at the SNOTEL observing stations that account for under-catch of frozen precipitation was necessary in order to provide accurate model outputs (*Livneh et al.*, 2014). A procedure adopted from *Livneh et al.* (2014) was used to adjust the precipitation data. Daily change in SWE accumulation can augment observed precipitation information when undercatch occurs. The following are the steps to complete the precipitation adjustments. First, compute the accumulation of precipitation and SWE over non-overlapping 7-day blocks starting at day 1 of the data and continuing until the end of the data record. This results in one value for accumulated precipitation, $P7_n$, and one value for SWE, $SWE7_n$, for every block, n, which is 7 days. If there are not 7 days remaining at the end of the data record, calculate accumulated precipitation and SWE for the last record day.

$$P7_n = A(n+7) - A(n) \tag{3.1}$$

$$SWE7_n = SWE(n+7) - SWE(n) \tag{3.2}$$

For each day, i, in block n, the following daily precipitation (P_i^*) adjustments adjustments are then made

$$P_i^* = dSWE_i, \text{ if } SWE7_n > P7_n \text{ and } dSWE_i \geq 0 \tag{3.3}$$

$$P_i^* = \max(P_i, 0), \text{ if } SWE7_n > P7_n \text{ and } dSWE_i < 0 \tag{3.4}$$

$$P_i^* = \max(P_i, dSWE_i), \text{ if } SWE7_n \leq P7_n \text{ and } P_i > 0 \tag{3.5}$$

$$0, \text{ if } SWE7_n \leq P7_n \text{ and } P_i = 0 \tag{3.6}$$

where $dSWE_i$ is the daily change in SWE. Figure 3.2 shows the how precipitation adjustment

method affected the measured precipitation at all site locations. It also depicts the different magnitudes of precipitation and SWE that exist at each site, varying with climate regimes. Henceforth, site plots will not have the same y-axis scale and thus will appear to have more similar precipitation and SWE values.

3.2 Climate Change Experiment

Precipitation and temperature change scenarios are applied to each site for both models to determine model sensitivity with respect to climate change. Projected climate change temperature and precipitation data come from the outputs of Global Circulation Models (GCMs), which are numerical models representing physical processes in the atmosphere, ocean, cryosphere, and land surface, to simulate the response of the global climate system in increasing greenhouse gas concentrations[4] . Due to limitations in computational resources, GCM outputs are run on a coarse spatial scale, on the order of hundreds of kilometers, which is not an appropriate spatial scale for many regional or local climate impact studies (e.g., basin-level response to change). As a result, 'downscaling' techniques have been developed to relate meso-scale atmospheric predictors to regional hydrometerology, which allows for projection of regional temperature and precipitation with respect to climate change (*Wilby et al.*, 1997).

The World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM) (*Taylor et al.*, 2011) initiated a new set of coordinated climate model experiments, which comprise the fifth phase of the Coupled Model Intercomparison Project (CMIP5). CMIP5 produced outputs for four different climate scenarios or representative concentration pathways (RCP2.6, RCP4.5, RCP6.0, RCP8.5). RCPs are mitigation scenarios that assume policy actions will be taken to achieve emissions targets. The RCPs are based on a range of projections of future population growth, technological development, and societal responses. The labels for the RCPs provide a rough estimate of the radiative forcing in year 2100. RCP8.5 represents a high emissions scenario, RCP 6 and RCP4.5 represent mid-range emissions scenarios, and RCP2.6 represents a

[4] `http://www.ipcc-data.org/guidelines/pages/gcm_guide.html`

low peak-and-decay scenario (*Taylor et al.*, 2012). Downscaled CMIP3 and CMIP5 climate and hydrology projections are available at $\frac{1}{8}^{\circ}$ to 2° spatial resolution translations of climate projections, over the contiguous U.S. developed using two complex downscaling techniques: bias correction spatial disaggregation (BCSD), which provides monthly data, and bias-correction and constructed analog (BCCA), which provides daily data[5] . These data provide access to climate and hydrological projections at spatial and temporal scales relevant to some watershed and basin-scale decisions. This study uses downscaled monthly temperature and precipitation data for the RCP4.5 scenario, which is a mid-range mitigation emissions scenario, at each site location to perform the climate analysis portion of this research, discussed below.

According to *Wood et al.* (1997), *Hay et al.* (2000), and *Fowler et al.* (2007), a simple and common method of applying temperature and precipitation changes to historical data and dealing with GCM inadequacies is to compute the difference between the current and future simulations and add these changes to the observed time series. This is known as the delta-change approach, and this method assumes that GCMs more accurately simulate relative change than absolute values (i.e., there is a constant bias through time). This method ignores any change in the variability of temperature and precipitation. Some researchers, including *Prudhomme et al.* (2002), have suggested using other more complex and improved methods; however, *Hay et al.* (2000) compared several downscaling methods and determined that the delta-change results provided more conservative estimates of changes in future runoff. As a result, this research uses existing downscaled GCM data and applied the delta-change method to obtain a new data set for each site location. The following set of steps describes the procedure for obtaining perturbed forcing datasets for the climate change analysis used in this research. Assume that there exists a number of water years of existing field-measured data. The following steps will yield perturbed precipitation and temperature data as input to the models.

(1) Download historic and future downscaled monthly precipitation and temperature data from all model outputs (71 model runs) produced in CMIP5 for one RCP scenario (*Taylor et*

[5] `http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html`

al., 2012). The number of water years of historic and future data should be the same. The historic downscaled data should be for the same water years as the existing field-measured data.

(2) For each month within each water year, average the 71 temperature and precipitation monthly values obtained from 71 CMIP5 model runs. This will produce twelve precipitation and twelve temperature values per water year.

(3) Calculate monthly average precipitation and temperature values for the historic CMIP5 simulation period and for the future CMIP5 simulation period, resulting in twelve values for each period. Do this by averaging the monthly averages for each month obtained for each water year in step (2). For example, average all of the June monthly values for the future CMIP5 simulation period to get a single temperature for the month of June for the future period.

(4) For temperature data, obtain the differences (ie. delta change) in the averaged monthly temperatures between the historic CMIP5 simulation period and future period, which were calculated in step (3). Apply this monthly delta-change to the existing field-measured daily historic temperature data. For precipitation data, obtain the percent differences in monthly precipitation between the historic and future simulation periods. Apply this percent change to the existing field-measured daily historic precipitation data. This prevents days with no precipitation from having precipitation in the future scenario.

The precipitation and temperature change scenarios are not coupled, and instead are run as two separate scenarios for each model. Thus, the precipitation change scenario data is coupled with the original historical temperature data, and the temperature change scenario data is coupled with the original historical precipitation data as model inputs. The precipitation and temperature change scenarios are applied separately because delta-change values are obtained by averaging GCM outputs instead of focusing on specific GCMs and lets us isolate their relative impacts.

GCMs are consistent in that they all produce an increase in temperature, but differ in the magnitude of increase, whereas they vary on whether precipitation will increase or decrease, and averaging precipitation outputs results in these varying predictions canceling each other out. There is no globally-coherent precipitation change signal, and as a result precipitation is much more sensitive than temperature to regional conditions (*Henderson-Sellers and McGuffie*, 2012). Precipitation sensitivity and disagreement among GCMs produces precipitation scenario results that are not as intriguing as the temperature change scenario results. As a result, the climate change experiment results and discussion sections focus on the temperature change scenario.

3.3 Model Setup

In order to maintain consistency throughout the research, SNOW-17 and VIC were both run at point locations for the same site locations and water years. The vegetation setting for the VIC runs was set for bare soil because the data collection locations (SNOTEL sites) are always in forest clearings with no canopy. A MATLAB version of SNOW-17 was used for the analyses in this research[6] . The main files required to run SNOW-17 include a precipitation input file and a temperature input file. Since the SNOW-17 parameters are not usually fixed, no editing of source code was necessary to perform the sensitivity analyses and multi-objective calibrations in this research.

VIC.4.1.2.c source code, written in the C programming language[7] , was used for this research. The input files required to run VIC include (1) a global parameter file, which combines the file paths for all other input files, a file path for the output, and model options, (2) a meteorological forcing file, which includes daily precipitation, maximum and minimum temperatures, and wind, (3) a soil parameter file, which contains soil parameter values for each grid cell, and (4) vegetation parameter file, which contains information on the particular kind of vegetation for each grid cell. For this research, VIC was run for a single point for each SNOTEL site. No SNOTEL sites have

[6] MATLAB SNOW-17 obtained from Mark S. Raleigh at the University of Washington, `http://depts.washington.edu/mtnhydr/people/mraleig1/toolbox2.html`

[7] VIC.4.1.2.c source code obtained from http://github.com

vegetation covering the snow measurement equipment, and thus the grid cell was run as a bare surface, with no vegetation. The parameters explored in the VIC model are typically fixed within the source code, thus completing the sensitivity analyses and multi-objective calibrations required editing and re-compiling of the source code.

3.4 Computational Experiment

Below is an outline of the computational experiments conducted in this research. Due to the computational cost of the analyses, all of these experiments were completed on the University of Colorado Boulder supercomputer (JANUS[8]).

(1) Sobol' Sensitivity Analysis of eight parameters in SNOW-17.

(2) Sobol' Sensitivity Analysis of eight parameters in VIC.

(3) Sobol' Sensitivity Analysis of seven parameters in SNOW-17. The SCF parameter was fixed for this analysis.

(4) Multi-objective calibration of SNOW-17 for both the eight and seven parameter scenarios.

(5) Multi-objective calibration of VIC.

(6) SNOW-17 climate change scenarios.

(7) VIC climate change scenarios.

As previously mentioned, eight parameters were chosen for the sensitivity analyses for both SNOW-17 and VIC. Ranges for all of the SNOW-17 parameters were chosen based on calibration information provided in *Anderson* (1973), *Hogue et al.* (2000), *Franz et al.* (2008), and *He et al.* (2011). Ranges for new snow density, snow roughness, liquid water holding capacity, and new snow albedo were chosen based on literature values. Ranges for the albedo decay parameters were chosen

[8] The Janus supercomputer at CU-Boulder is comprised of 1368 compute nodes, each containing 12 cores (2.8 ghz Intel processors), for a total 16,416 available cores.

based on a visual exploration of the sensitivity of the albedo decay curves with respect to varying parameter values. Tables 2.1 and 2.2 summarize the parameters for each model, and contain the original fixed parameter values, ranges for each parameter, and citations for literature containing information regarding parameter ranges.

The Sobol' sensitivity analyses were completed using an MOEA Framework[9] , which is a Java library. For each analysis, a total of *2N(p+1)* parameter sets were used, where N is the sample size (1,000 for all analyses), and p is the number of parameters. Sensitivity analyses for both SNOW-17 and VIC involved eight parameters, which resulted in a total of 18,000 model runs. A second set of sensitivity analyses for SNOW-17 included seven parameters, resulting in a total of 16,000 model runs. By default, the MOEA Framework used to conduct the Sobol' analysis employs bootstrap analysis using confidence intervals within the MOEA Framework (*Sobol'*, (1993); *Saltelli*, 2002; *Tang et al.*, 2006).

The multi-objective calibrations were completed using the Borg MOEA (*Hadka and Reed*, 2012). The default parameterization of the Borg MOEA framework was used, which is justified based on comparative studies that show Borg's good performance over its parameter range (e.g. *Reed et al.*, 2013). In this study, the algorithm was run for 10 seeds and 100,000 function evaluations for SNOW-17 and 10 seeds and 150,000 function evaluations for VIC. Using numerous seeds ensures the calibration results are not an artifact of the MOEA starting point. Since Borg has proven to be robust across its parameter space, the default parameters, described in Table 3.1, were used in this study. Several epsilon values were tested, but the final epsilon resolutions for each objective are presented in Table 3.2.

[9] http://www.moeaframework.org/

Table 3.1: Default parameters used in the Borg MOEA. The following operators are included: simulated binary crossover (SBX), differential evolution (DE), parent centric mating (PCX), simplex crossover (SPX), uniform normally distributed crossover (UNDX), uniform mutation (UM), and polynomial mutation (PM) (*Woodruff et al.*, 2013).

Parameter	Value
Initial population size	100
Tournament selection size	2
SBX rate	1.0
SBX distribution index	15.0
DE crossover rate	1.0
DE step size	0.5
PCX parents	3
PCX offspring	2
PCX eta	0.1
PCX zeta	0.1
SPX parents	3
SPX offspring	2
SPX epsilon	2
UNDX parents	3
UNDX offspring	2
UNDX eta	0.5
UNDX zeta	0.5
UM rate	$\frac{1}{27} = \frac{1}{\textbf{number of DV}}$
UM rate	$\frac{1}{27} = \frac{1}{\textbf{number of DV}}$
PM distribution index	20

Table 3.2: ϵ resolutions used with the Borg MOEA for the objective functions. ϵ resolutions depend on the range of possible values for the objective functions.

Objective	Search Epsilon	Sort Epsilon
Peak	0.01	0.01
Length	0.1	0.1
NSE	0.001	0.001
RMSE	0.01	0.01

Chapter 4

Results

The purpose of this research is to use sophisticated diagnostic tools to address the following questions (1) Do complex physically based models perform as well or better than simple conceptual models? (2) How do we identify well-performing parameter sets with respect to several, conflicting objectives? (3) How do errors in parameterizations propagate to portrayals of climate change? Results addressing these questions are described below. Section 4.1 presents baseline model results as motivation for this research. Section 4.2 presents ensemble and Sobol' sensitivity analysis results to explore whether complex physically based models perform as well or better than simple conceptual models. Section 4.3 presents multi-objective model calibration results to identify well-performing parameter sets with respect to conflicting objectives. Section 4.4 presents temperature change scenario results to explore how model parameterizations propagate to portrayals of climate change.

4.1 Baseline Model Results

The diagnostic tools used in this research are presented with the context of a baseline run for each model, which refers to running each model with its default parameters. Baseline results are crucial for quantifying the effects of important model parameters identified in the sensitivity analysis and improvements in model runs due to calibration. Figure 4.1 presents baseline SWE results for all five study locations using default parameter values for the SNOW-17 and VIC models. The SNOW-17 SCF, snow correction factor, parameter was initially included in the model to account for gauge catch deficiencies, but it also implicitly includes losses during the snow accumulation period due to

sublimation and redistribution of blowing snow. The value of SCF is typically chosen to give the best estimate of SWE at the beginning of the melt season (*Anderson*, 2006). SCF does not have a default value, and will never have a single true value. In reality, this is a tuning parameter that should increase or decrease depending on precipitation. The baseline model runs use the default parameter sets and assume we have no prior information about the sites being modeled. Thus, an SCF value of 1, which indicates no gauge deficiencies or losses during the accumulation period, is used in the baseline runs. Figures 4.1 (a), (b), (c), (d), and (e) each have a unique y-axis scale, depending on the site's climate and amount of yearly accumulated precipitation, in order to better visualize the differences between modeled baseline SWE and observed SWE. Green lines on the plots represent SNOW-17 baseline SWE; blue lines represent VIC baseline SWE; black lines represent observed SWE; red lines represent the site-specific adjusted accumulated precipitation to ensure model results are realistic and do not defy conservation of mass by exceeding the accumulated precipitation. This concept was discussed in Chapter 3, Section 1.

Baseline SWE results for the SNOW-17 and VIC models, presented in Figure 4.1, demonstrate baseline modeled SWE has some obvious deviations from the observed SWE. For example, in Figure 4.1 (b), the VIC simulation in Copper water year two has an unusual dip in SWE that does not exist in the observed SWE. SNOW-17 is able to more accurately capture the shape of SWE in this water year. One hypothesis for this deviation is based on the fact that SNOW-17 uses daily mean temperature as an input while VIC uses daily minimum and maximum temperatures as inputs. VIC estimates solar radiation based on these minimum and maximum values, so a large range between these values results in VIC overestimating solar radiation for that particular time period. Another example is in Figure 4.1 (a) where neither model is able to accurately capture peak SWE for all water years. In Figure 4.1 (e), SNOW-17 and VIC both struggle to appropriately capture the SWE behavior in water year three.

Table 4.1 summarizes the objective function values for the baseline models for all results. The worst performing objective values (minimum NSE, maximum RMSE, peak, and length) out

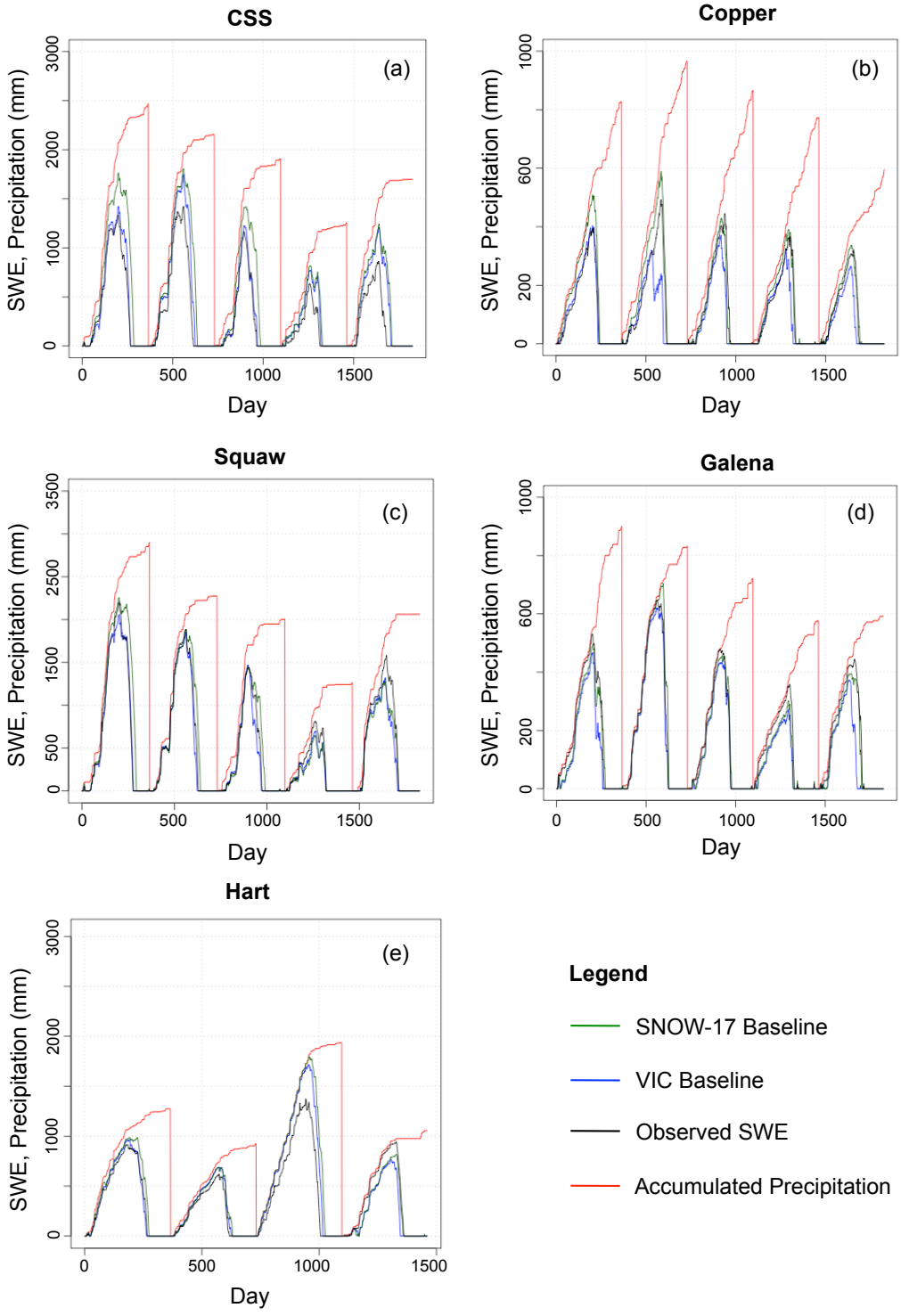


Figure 4.1: Baseline SWE outputs for SNOW-17 and VIC for all site locations. These outputs were produced using default model parameter sets (Tables 2.1 and 2.2). The only parameter without a default value is the SNOW-17 SCF parameter, which was fixed at a value of 1.

Table 4.1: Objective function values for each site for baseline SNOW-17 and VIC model runs using default parameter sets. The peak, length, and RMSE objective function are to be minimized, whereas the perfect value for the NSE objective function is 1. Each objective value represents the worst performing value for a particular water year for each site. The objective values for each site can represent different water years depending on which water year performed the worst for a particular objective function.

	Peak (mm)		Length (days)		NSE		RMSE (mm)	
	SNOW-17	VIC	SNOW-17	VIC	SNOW-17	VIC	SNOW-17	VIC
Copper	113.3	99.6	10	31	0.81	0.59	59.1	96.9
CSS	424.6	1381.4	28	24	-0.08	0.29	364.8	248.5
Squaw	302.7	1692.3	18	6	0.84	0.89	327.3	193.2
Galena	58.4	254.2	18	44	0.88	0.57	47.1	109.5
Hart	424.7	1351.7	20	11	0.52	0.77	335.4	231.7

of the modeled water years were chosen as the representative[1] objective values for a specific model run. Also, the water years for an objective function could vary between models, and the worst performing water year for one model may not be the same year for the other model. Visual inspection of Figure 4.1 can provide information regarding which water year the objective values represent for the baseline results. For example, the SNOW-17 NSE value for Hart is 0.52, and according to Figure 4.1 (e), this value is for water year three because this is the year in which the modeled SWE deviated the most from the observed SWE. The VIC NSE value for Hart is 0.77, which is logical because Figure 4.1 shows VIC performed slightly better than SNOW-17 for water year three. However, it is interesting that NSE increased by approximately 48% from SNOW-17 to VIC because visual inspection is not consistent with this great of an increase. This underscores the importance of combining visual inspection with quantitative analysis of model performance.

Figure 4.1 and Table 4.1 demonstrate there is room for improvement for both models and all site locations when modeling SWE. The Sobol' sensitivity analysis, multi-objective calibration, and climate change results presented below provide further insight into the differences between water

[1] This formulation means that reported objective function values could come from different water years. For example, in a given model parameterization, the worst NSE could be in the 3^{rd} water year and the worst peak performance comes in the 2^{nd} year. These values from different years represent that parameterization's performance overall.

years and site locations, and show that both models have the ability to overcome inadequacies in the baseline results.

4.2 Ensemble and Sobol' Sensitivity Analysis Results

4.2.1 SNOW-17 Results

This study uses Sobol' sensitivity analysis to determine individual parameter sensitivities of eight parameters used in SNOW-17. The Sobol' variance decomposition method requires an ensemble of model parameterizations to be created. This ensemble of model runs, shown in Figures 4.2 (i) through (j), can be investigated to show the range of model performance. These figures have the same y-axis scales as the baseline plots. The gray areas in Figures 4.2 (a) through (j) represent the ensemble of SWE outputs produced by varying parameter values within SNOW-17. The black lines represent observed SWE, and the red lines represent accumulated precipitation. The first column of plots, including Figures 4.2 (a), (c), (e), (g), and (i), represent the SWE ensembles produced when varying all eight of the SNOW-17 parameters. The second column of plots, including Figures 4.2 (b), (d), (f), (h), and (j), represent the SWE ensembles produced when the SCF parameter is fixed at a value of 1, and the remaining seven parameters are still varied. Figure 4.2 shows that including the SCF parameter in the ensemble produces modeled SWE that exceeds accumulated precipitation. The more accumulated precipitation at a site, the larger the exceedance. However, Figure 4.2 shows that when setting SCF to a fixed value and eliminating it from the parameter ensemble, SNOW-17 does not produce SWE that exceeds accumulated precipitation. Also, all plots show the greatest variation in model outputs during the the peak and melt seasons. These results indicate the SCF parameter can be responsible for unrealistic model results; however, they do not provide information regarding whether varying the SCF parameter could also produce better performing model runs than setting SCF to a fixed value, at a cost of risking the potential for the unrealistic runs.

Empirical models, such as SNOW-17, often use correction factors due to a large amount of

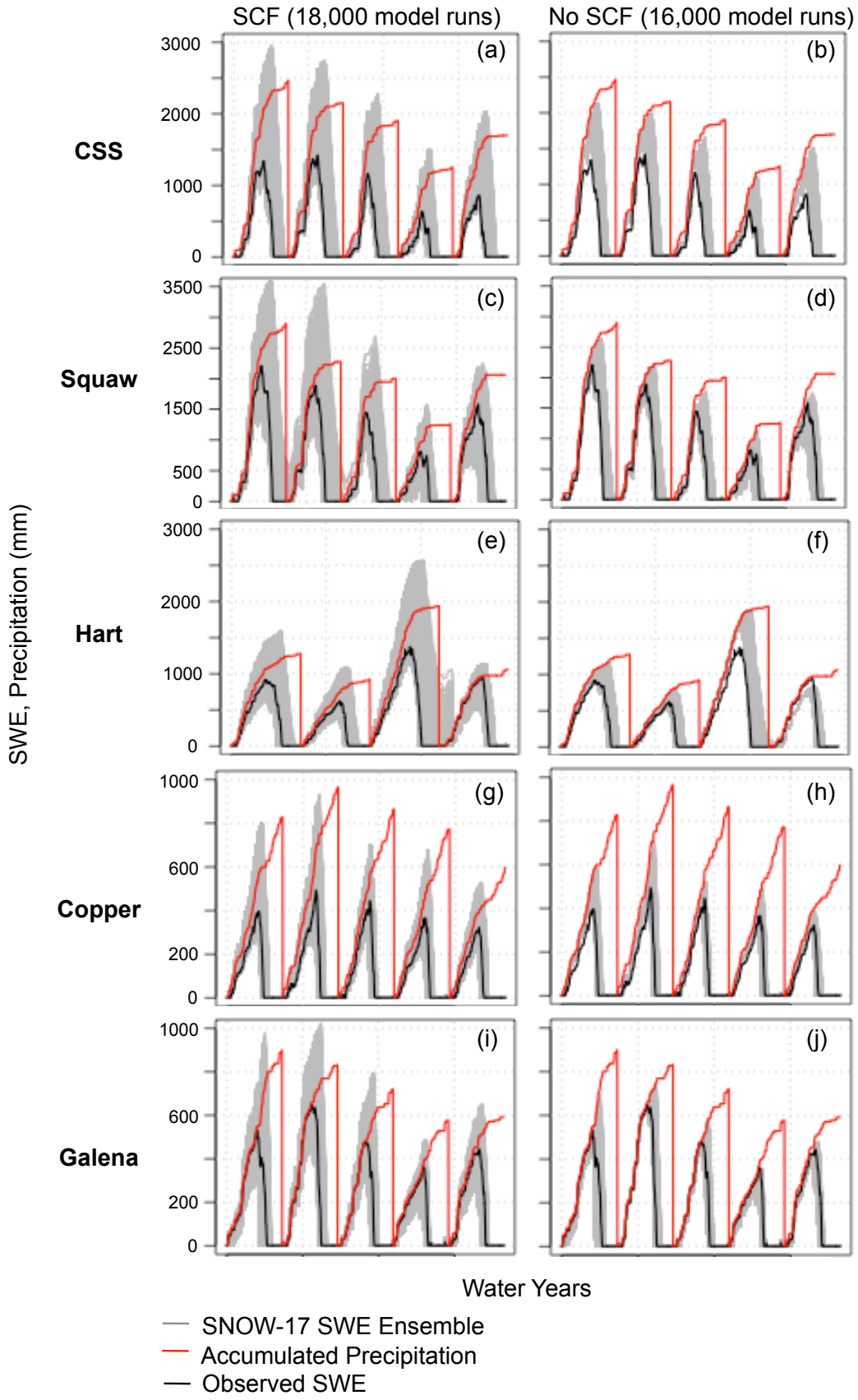


Figure 4.2: The ensemble of SNOW-17 SWE outputs resulting from the ensemble of parameter sets generated in the Sobol' analysis.

uncertainty in model forcings. For example, as discussed earlier, precipitation adjustment was required for this research due to issues with disagreement between observed accumulated precipitation and observed SWE. The SCF parameter represents this correction factor in the SNOW-17 model. The ensemble results demonstrated sensitivity of model output with respect to SCF. As a result, a Sobol' sensitivity analysis to explore individual parameter sensitivities of the eight SNOW-17 parameters was conducted. Results, which are presented as total order indices, provide a broad picture of model behavior since they estimate the effect of the individual parameters and their interactions with other parameters. A sensitivity threshold, t, is often set, such that parameters with sensitivity indices above t are considered sensitive. For example, *van Werkhoven et al.* (2009) explored several values of t including 0.05, 0.1, 0.2, 0.3 and the impact of selecting a particular sensitivity value and thus reducing the number of parameters involved in model calibration. This research uses sensitivity analysis to reveal sensitive parameters and explore the effects of changing parameter values, but does not eliminate parameters from calibration based on a sensitivity threshold.

Rasters presented in Figure 4.3 summarize the total order indices for SNOW-17 parameters for four objective functions for each study location. Figures 4.3 (a), (c), (e), (g), and (i) present total order indices for SNOW-17 parameters when SCF is included in the Sobol' sensitivity analysis. Figures 4.3 (b), (d), (f), (h), and (j) present total order indices for parameters when SCF is eliminated (fixed at a value of 1) from the Sobol' analysis. Numerical instabilities (truncation errors, and Monte Carlo approximation errors) led to some sensitivity indices of less than 0, but these values are shown as 0 for visualization purposes. The darker a box within each raster, the more sensitive a parameter with respect to a specific objective function. Sensitivity values greater than 0.8 indicate a parameter controls the model's response. SNOW-17 results depict that SCF is by far the most sensitive parameter within SNOW-17, which makes sense because the definition of SCF indicates it corrects for model errors that cannot be corrected via calibration of other parameters. The MFMAX parameter also demonstrates greater total order indices than other SNOW-17 parameters, especially with respect to the length objective function. MFMAX is the

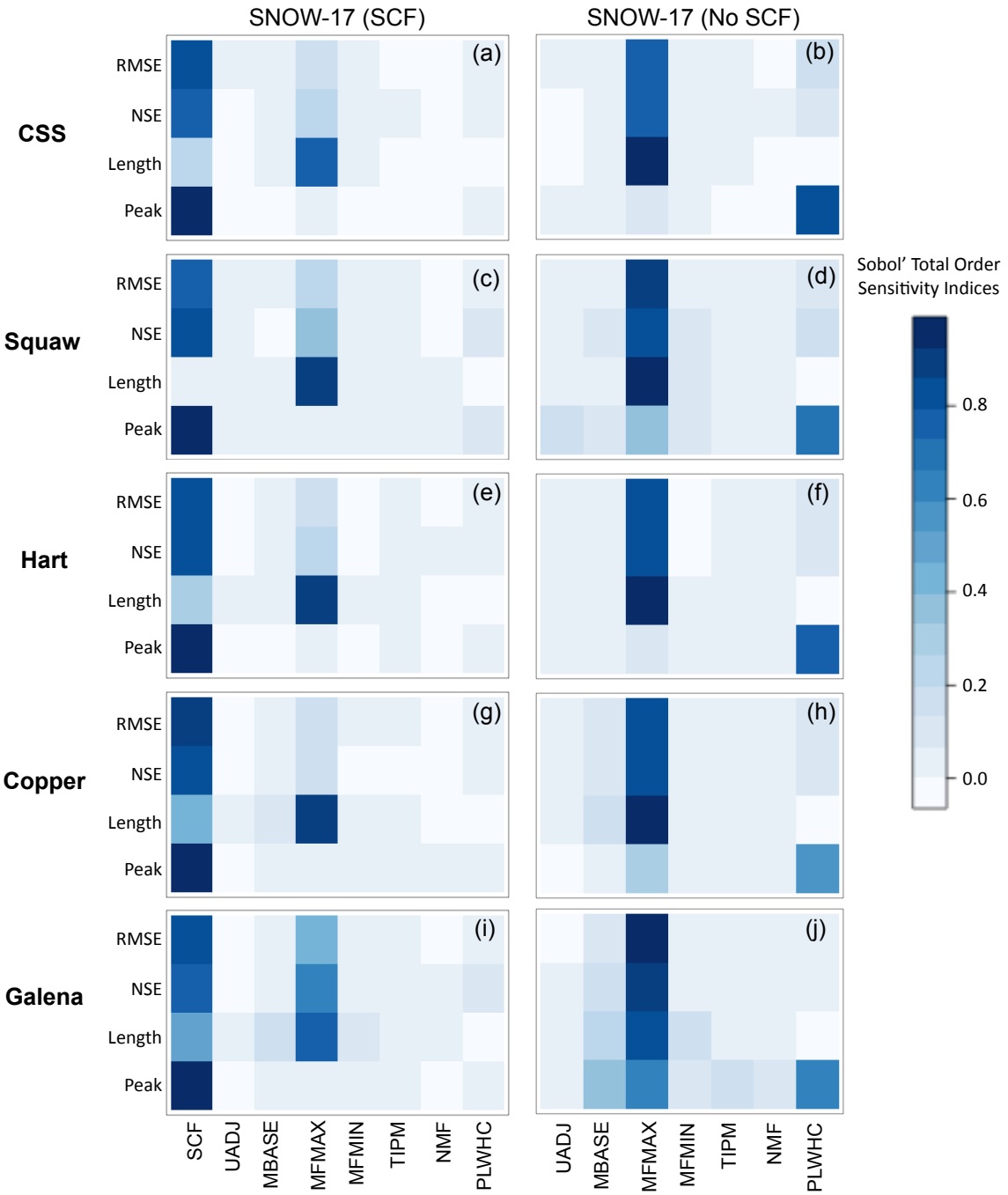
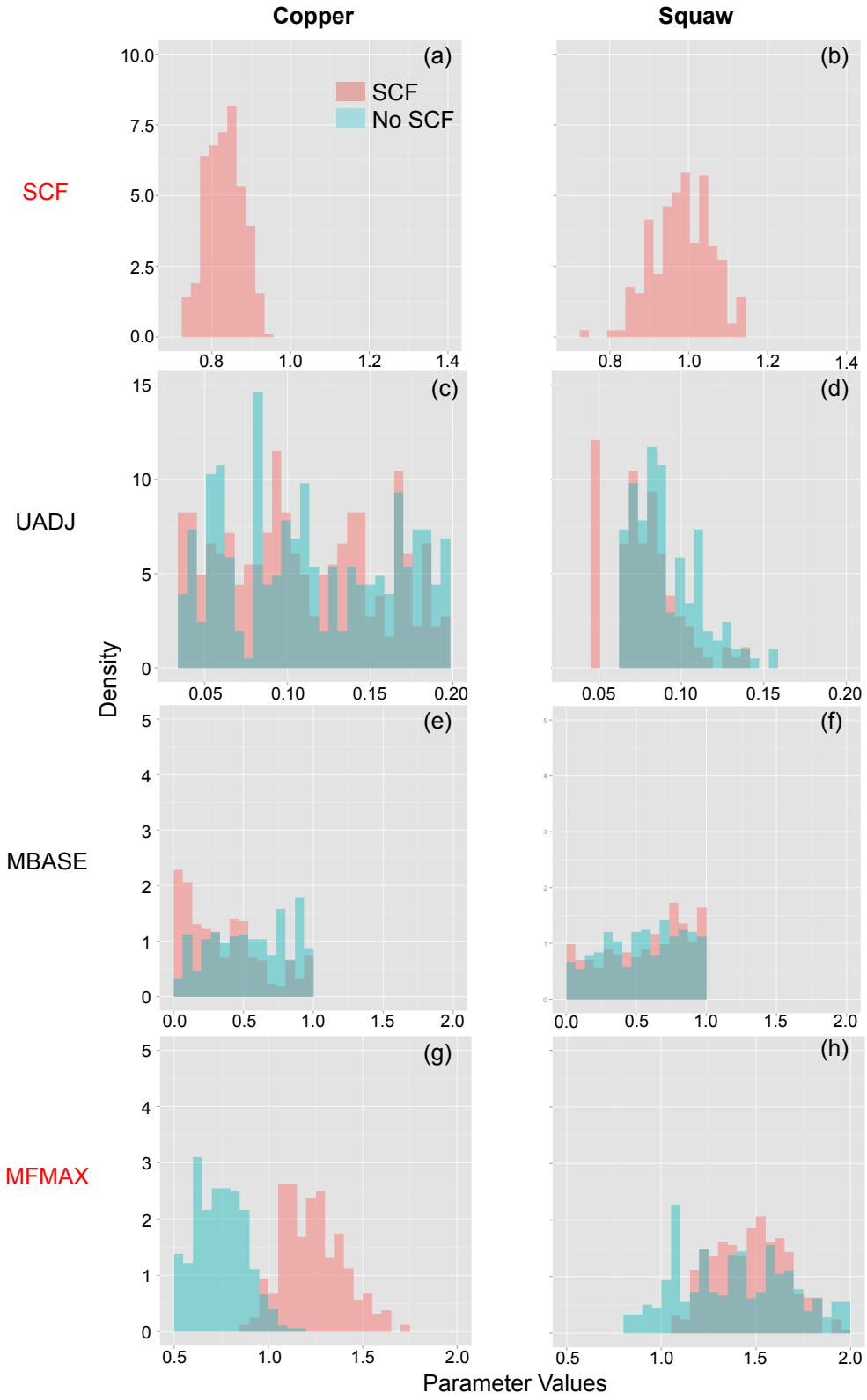


Figure 4.3: Total order sensitivity indices for the SNOW-17 parameters. The darker the box, the more sensitive a parameter with respect to a specific objective function.

maximum melt factor, which occurs on June 21^{st}, and is used to calculate the seasonal variation in the non-rain melt factor within SNOW-17. When SCF is removed from Sobol', the total order indices for MFMAX, across all objective functions and all sites, increase, along with the sensitivity of PLWHC with respect to the peak objective function. PLWHC is the percent liquid water holding capacity, which represents the amount of water a snowpack can hold before output from the snowpack occurs. Sobol' results indicate the SCF parameter controls the output when given the freedom to do so; however, when SCF is fixed, MFMAX has the most influence on the output. Since SCF is a correction factor and has the ability to control model output, perhaps it should be initially ignored in model calibration so SNOW-17 does not perform well for the wrong reason. This correction factor can differ greatly with time, and thus it seems appropriate to focus on other parameters during the calibration process.

Figure 4.3 shows histograms of SNOW-17 parameter values for the top 2% performing model runs with respect to NSE, for both the SCF and non-SCF Sobol' analyses. Figures 4.3 (a), (c), (e), (g), (i), (k), (m) and (o) are histograms for Copper, and Figures 4.3 (b), (d), (f), (h), (j), (l), (n) and (p) are histograms for Squaw. These two sites are selected for comparison because they represent two different climate regimes. Sensitive parameters identified in the sensitivity analysis are labeled in red. Figures 4.3 (a) and (b) show the SCF parameter values resulting in the best model runs have different distributions between the two sites. This seems appropriate because different climate regimes with different snow densities and snow depths likely have different sources of error, and would require different SCF values to correct for that error. For both SNOW-17 Sobol' scenarios, the two sites generally have different distributions of parameter values. Distributions with a narrower range and larger peak (i.e., higher likelihood) indicate that a parameter is identifiable. Figures 4.3 (g) and (h) show the MFMAX parameter is identifiable for Copper, but not identifiable for Squaw. This means MFMAX has a set range of values that produce good results, and this range is dependent on SCF. Figure 4.3 (o) shows Copper's distribution of PLWHC values is drastically different for the SCF and non-SCF Sobol' analyses, whereas Figure 4.3 (p) shows Squaw's PLWHC distributions for the two scenarios are similar to each other. Perhaps when SCF is fixed, the PLWHC


Copper
Squaw
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
SCF
No SCF
SCF
UADJ
MBASE
MFMAX
Density
Parameter Values

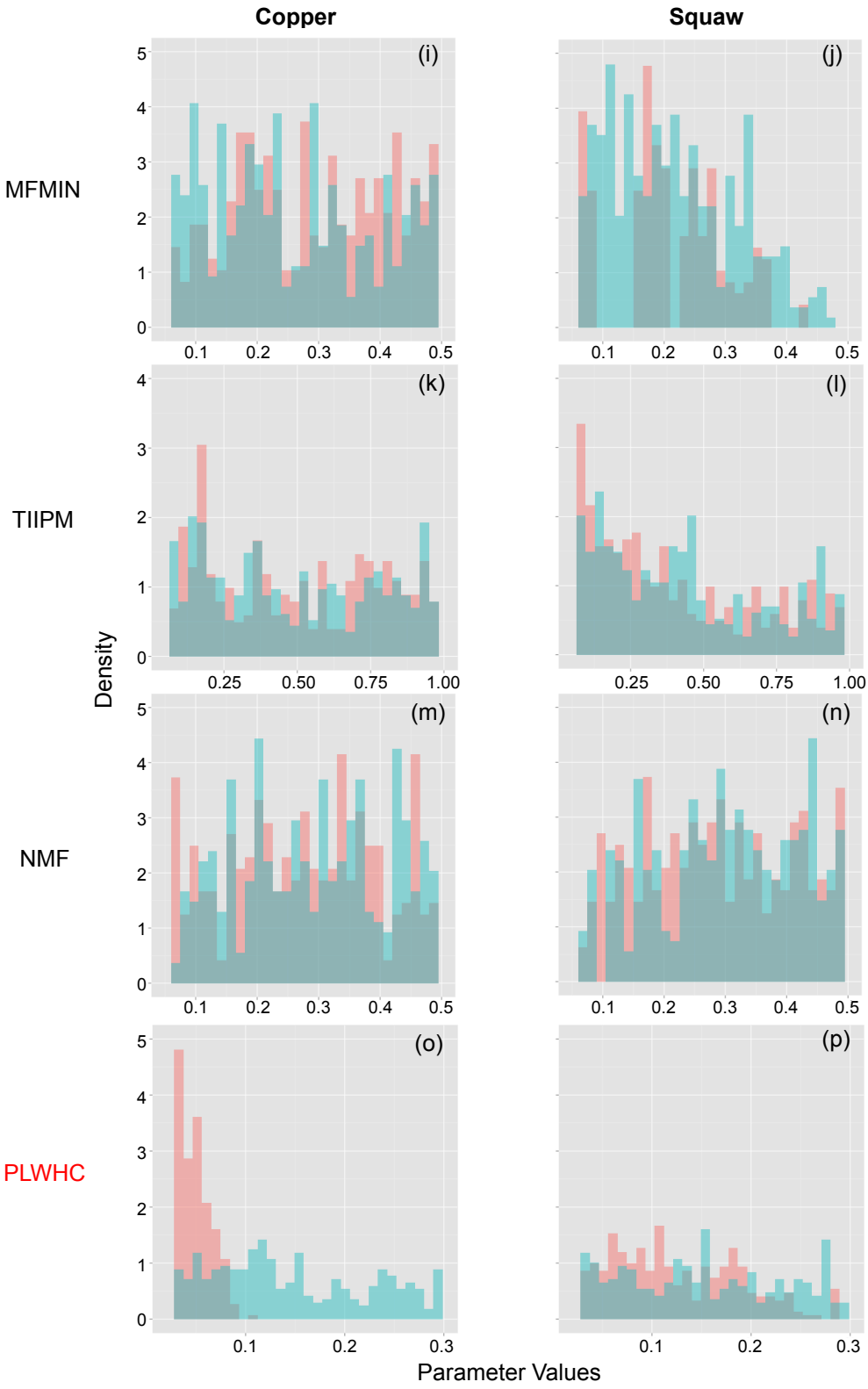


Figure 4.3: Histograms of the parameter values for the top 2% performing SNOW-17 model runs with respect to NSE. Copper and Squaw were selected to compare different climate regimes.

parameter has to make up for error that is accounted for by SCF when it's included, resulting in a larger range of PLWHC values. PLWHC represents the overall liquid water holding capacity of a well aged snow cover (*Anderson*, 2006). Including SCF in Sobol' causes the range of PLWHC values in top performing model runs to be much smaller at Copper than the values in the Sobol' runs without SCF. The variation in SCF values allows the model to have similar snow accumulation throughout the model runs, resulting in similar PLWHCs among all of the runs. Fixing the value of SCF does not allow the model to correct the snow accumulation via the SCF parameter, and thus differing snow accumulation scenarios will have very different water holding capacities, resulting in the larger range of PLWHC values. The remaining distributions of the non-sensitive parameters appear to not have any patterns. Adjusting bin sizes could change the appearance of some of these distributions. However, these distributions indicate the parameters are not identifiable, meaning their values do affect the performance of the model output with respect to the NSE objective.

4.2.2 VIC Results

The next several figures perform a similar analysis for the physically based VIC model. Ensembles of VIC SWE outputs, depicted in Figure 4.4, show large ranges of SWE produced by varying historically fixed parameters within the VIC snow model. Like in the SNOW-17 results, some SWE results in the Squaw and Hart ensembles, depicted in Figures 4.4 (b) and (c), respectively, exceed accumulated precipitation. These are also the two sites with the most accumulated precipitation. Also, some of the Squaw and Hart results indicate snow never fully melts, and there is existing snow on the ground when the new snow season begins. This is known as glaciation and can occur in VIC if the albedo is too high and/or there is no canopy. Glaciation is the cause of SWE exceeding the accumulated precipitation. The ranges of outputs are significant because they demonstrate that the SWE output produced by VIC is very dependent on fixed parameter values, and is not necessarily the best output or could potentially be good for the wrong reasons. Further exploration of model runs also reveals VIC is able to overcome the strange dip that occurred in Copper's second water year in the baseline results.

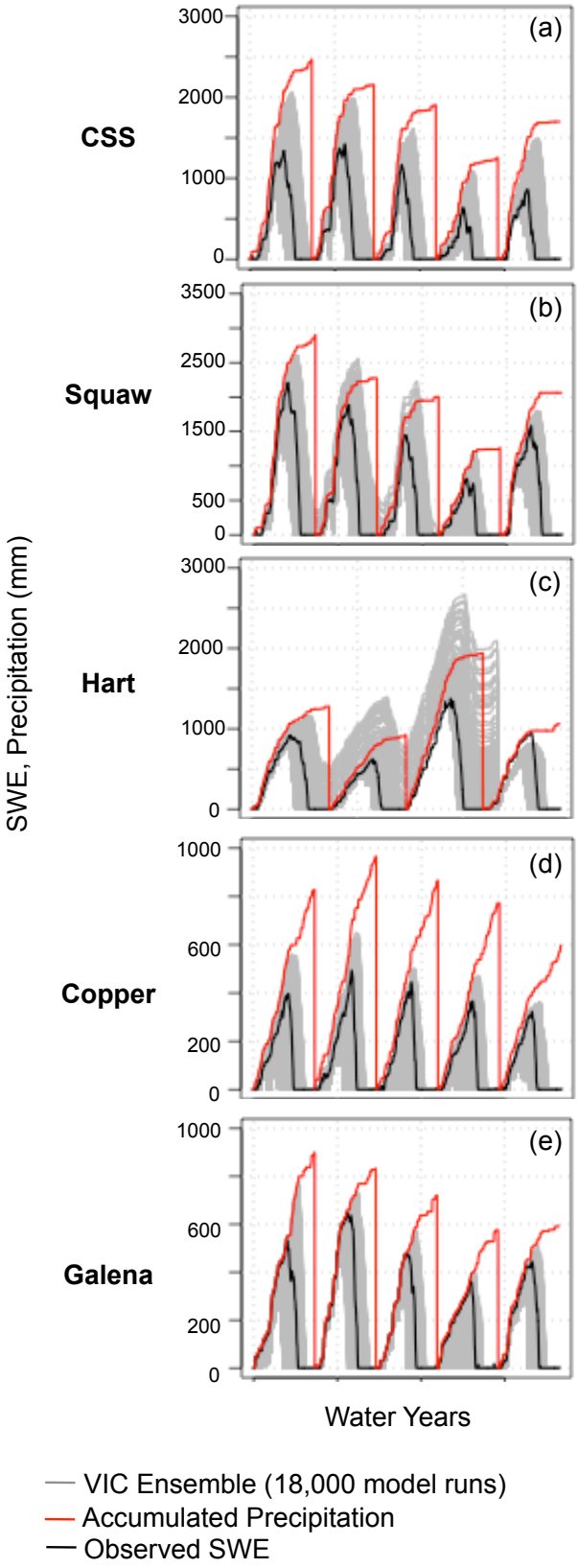


Figure 4.4: The ensemble of VIC SWE outputs resulting from the ensemble of parameter sets generated in the Sobol' analysis.

Since varying historically fixed parameters within VIC produces large ensembles of SWE results, further investigation into specifically which parameters cause these variations is achieved with the Sobol' sensitivity analysis. Figures 4.5 (a), (b), (c), (d), and (e) illustrate the Sobol' sensitivity analysis total order indices, which show that accumulation albedo decay parameter A and thaw albedo decay parameter A are the most sensitive parameters. These parameters are part of the accumulation and thaw albedo decay curve equations discussed in Chapter 2. The default values for these parameters were originally assumed based on one study location in 1972 (*Bras*, 1990). The result that any parameter has a high sensitivity within the VIC snow model is important because it means these normally fixed parameters have the power to greatly impact the variance of the VIC output. As discussed in *Mendoza et al.* (2015), the uncertainty of fixed parameters within complex physically based models is typically ignored, but these results indicate that perhaps this uncertainty should be explored. The baseline VIC results were not exceptionally good, and calibrating some of the sensitive parameters, such as the accumulation and thaw albedo decay parameters, could greatly improve model performance. The task of calibrating model parameters within a physically-based model seems daunting because of its complexity and large number of parameters. However, as demonstrated in this sensitivity analysis, not all of the parameters necessarily significantly contribute to the output variance. Sensitivity methods, such as Sobol' can be applied to other parameterizations within VIC or other complex models in order to uncover which parameters are most sensitive. Then, model sensitivity and calibration efforts can focus on these sensitive parameters, making such tasks less daunting. In addition, parameter sensitivity information can be incorporated into future developments and improvements of complex models. For example, model developers can recognize these sensitive parameters and create user-friendly methods for adjusting these parameters, which would make complex models more flexible, as called for in *Mendoza et al.* (2015).

Figures 4.5 show histograms of VIC parameter values for the top 2% performing model runs with respect to NSE. Figures 4.5 (a), (c), (e), (g), (i), (k), (m) and (o) are the histograms for Copper, and Figures 4.5 (b), (d), (f), (h), (j), (l), (n) and (p) are the histograms for Squaw.

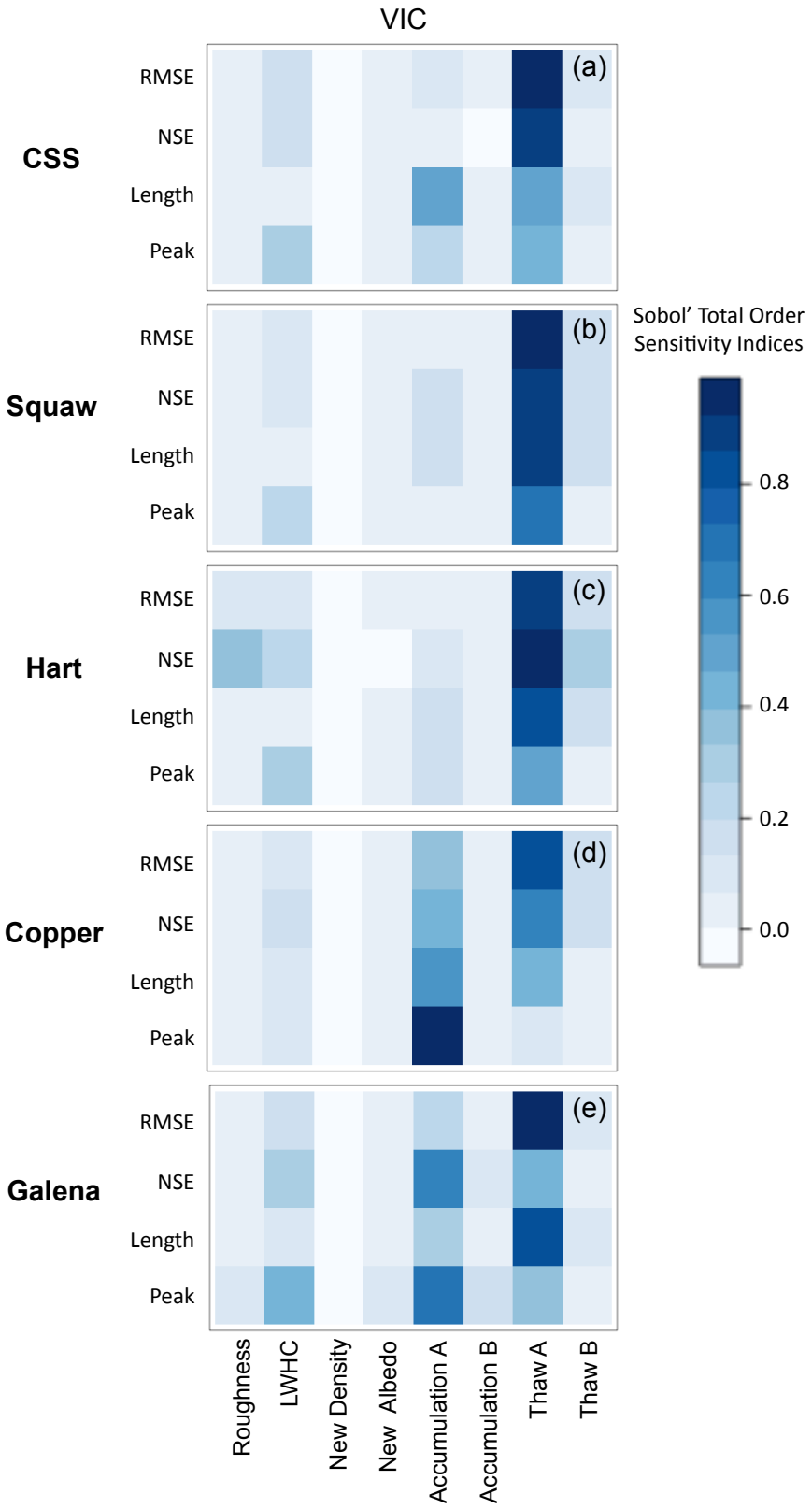
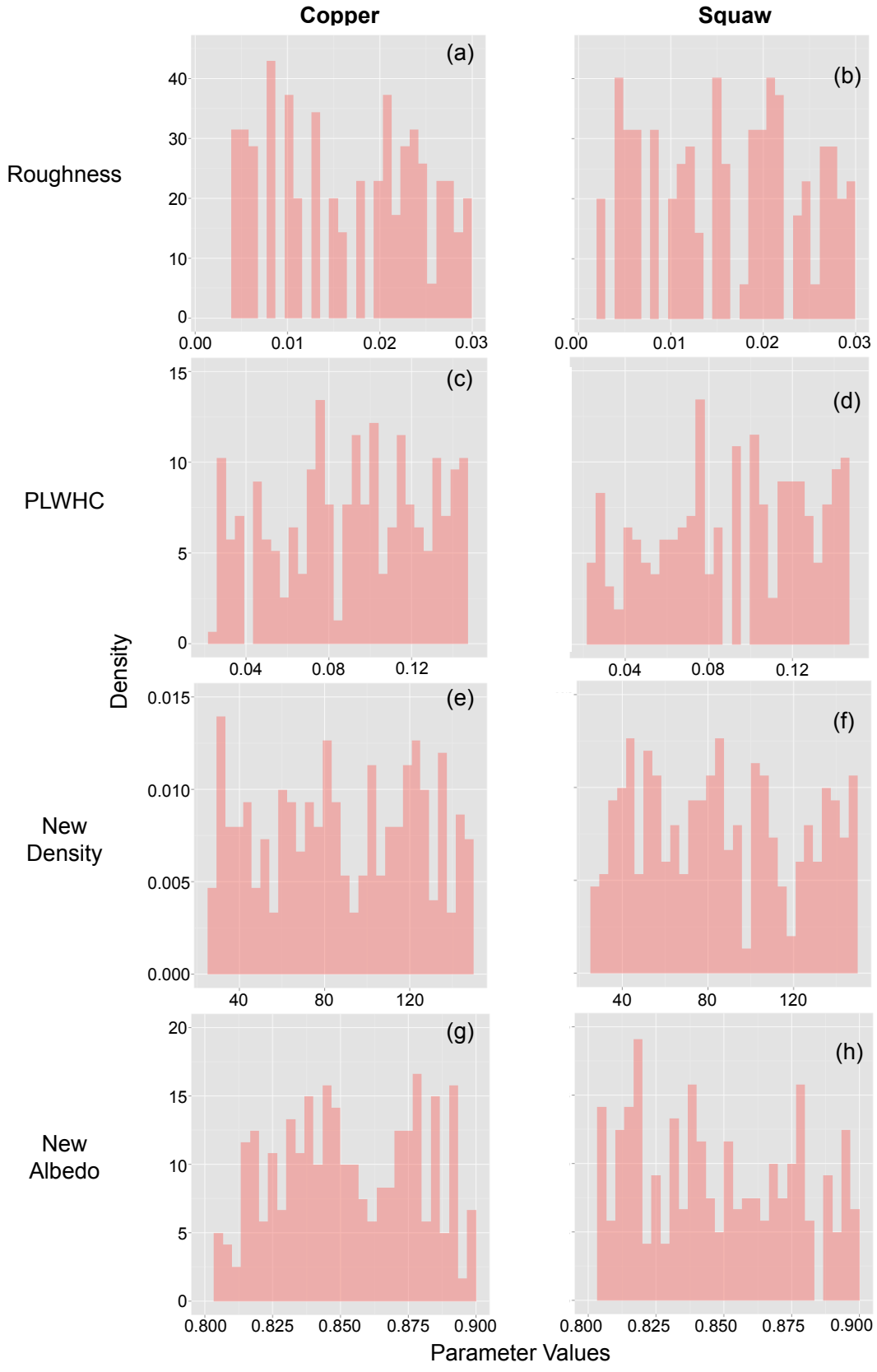


Figure 4.5: Total order sensitivity indices for the VIC parameters. The darker the box, the more sensitive the parameter with respect to a specific objective function. Sensitive parameters identified in the Sobol' analysis are highlighted in red.


Copper
Squaw
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Roughness
PLWHC
New Density
New Albedo
Density
Parameter Values
0.00
0.01
0.02
0.03
0
10
20
30
40
0.04
0.08
0.12
0
5
10
15
0.000
0.005
0.010
0.015
40
80
120
0
5
10
15
20
0.800
0.825
0.850
0.875
0.900

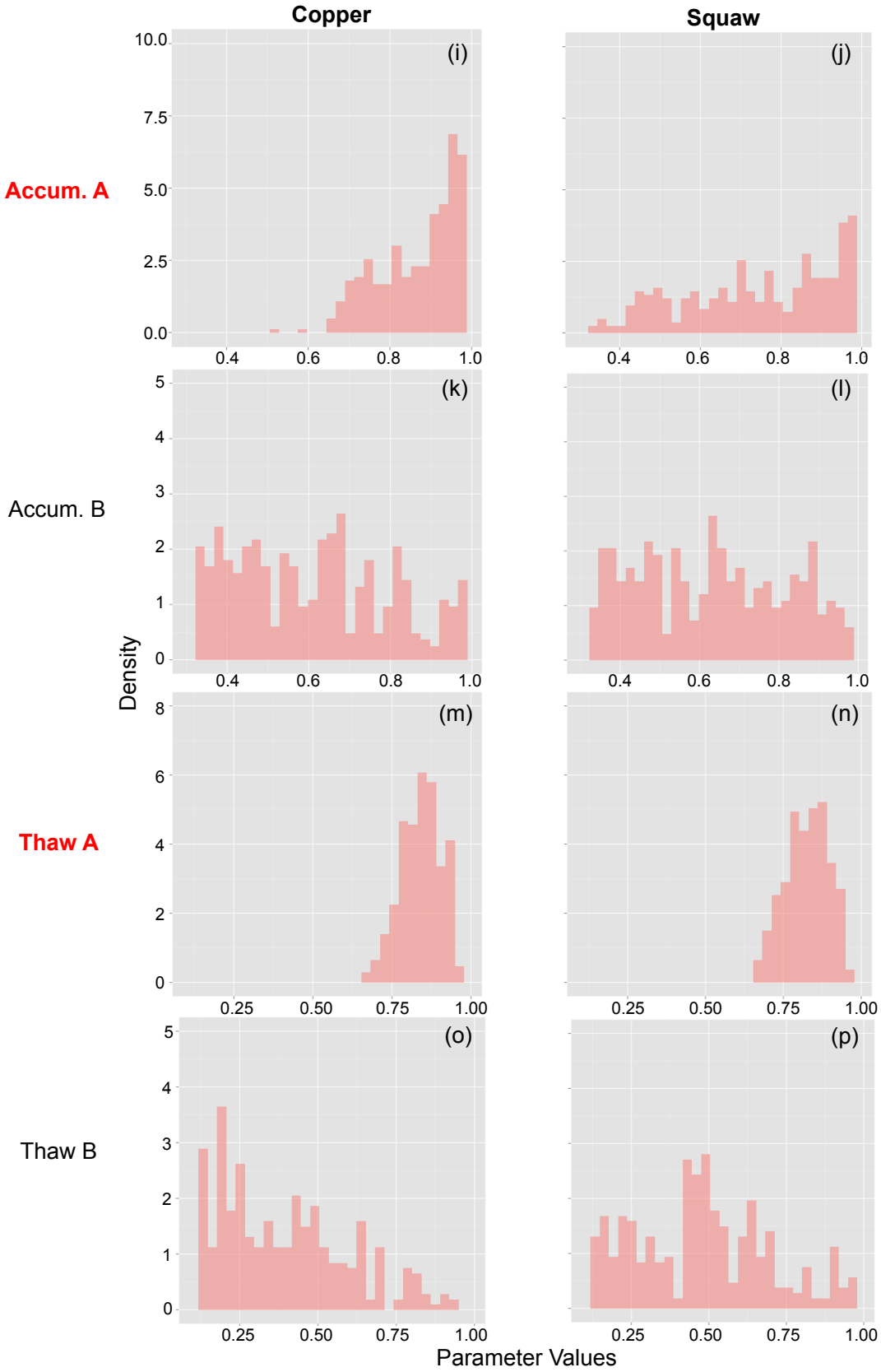


Figure 4.5: Histograms of the parameter values for the top 2 % performing VIC model runs with respect to NSE. Copper and Squaw were selected to compare different climate regimes. Sensitive parameters identified in the Sobol' analysis are highlighted in red.

Sensitive parameters identified in the sensitivity analysis are labeled in red. There is some variation in distributions between sites, but not as much as SNOW-17. Figures 4.5 (i), (m) and (n) stand out because they indicate that the accumulation A parameter for Copper and the thaw A parameter for both sites are identifiable. These parameters were also the most sensitive parameters in the Sobol' analysis. Since these parameters seem to control VIC output, it makes sense that they have a specific range of values that produce the good VIC results. Again, the distributions with no patterns indicate the parameters are not identifiable, meaning their values do affect the performance of the model output with respect to the NSE objective.

4.2.3 Comparison of SNOW-17 and VIC Results

Figure 4.6 presents bar graphs of the best objective values achieved for each model (SNOW-17 with SCF, SNOW-17 without SCF, VIC) and each site location. These objective values come from a random sample from Sobol', and objective values achieved via a multi-objective calibration may be different. The red bars represent values associated with SNOW-17 and a varying SCF, the green bars represent objective values associated with SNOW-17 and a fixed SCF, and the blue bars represent objective values associated with VIC. The aim is to minimize the peak, length, and RMSE objective functions and maximize the NSE objective function. Sobol' results produced parameter sets for all models and all sites that have improved objective function values from the baseline model results. Although all objective values are better than the baseline, there is a large variation in objective function values. The VIC peak objective values indicate VIC has trouble modeling the SWE peaks for the maritime climate sites. SNOW-17 with a fixed SCF value has more trouble modeling these peaks than SNOW-17 with a varying SCF value. VIC greatly outperforms the SNOW-17 models with respect to the length objective function at Copper and Squaw, and SNOW-17 outperforms VIC at CSS. VIC resulted in the overall best length objective value at Squaw, while it resulted in the overall worst peak objective value also at Squaw. This indicates there is a tradeoff between the two objective functions. The NSE and RMSE values are generally similar across all three model scenarios. However, SNOW-17 with a varying SCF outperforms both SNOW-

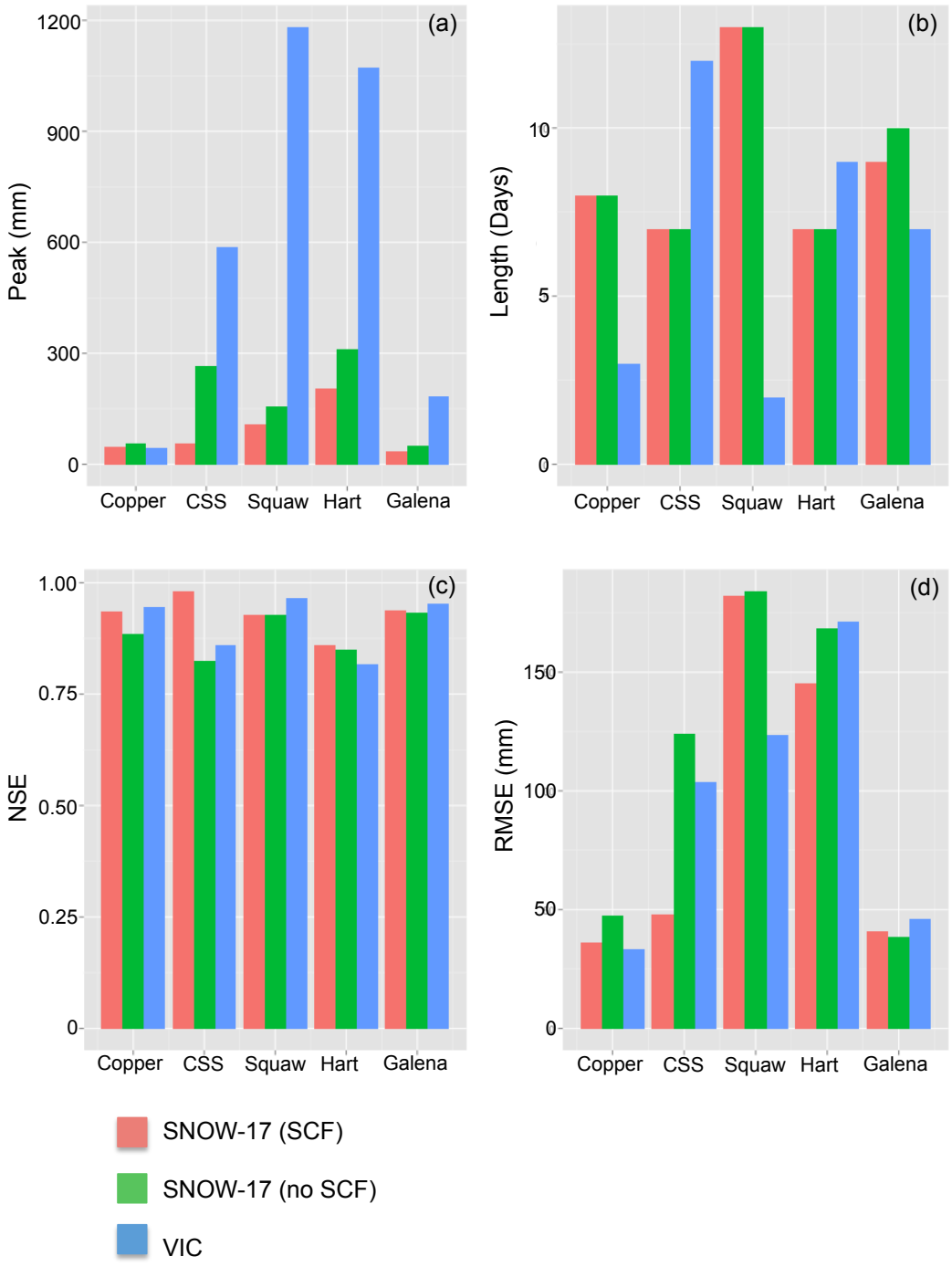


Figure 4.6: Bar graphs depicting the best objective values resulting from each Sobol' analysis for each site location. The aim is to minimize peak, length, and RMSE, and maximize NSE. The best objective values are not necessarily from the same model runs or same water years for each site. These objective values are from a random sample within Sobol', and the calibration results may have different objective values.

17 with a fixed SCF and VIC for NSE and RMSE at the CSS location, and VIC outperforms the SNOW-17 models with respect to RMSE at the Squaw location. Particular models seem to perform consistently better for some sites. For example, SNOW-17 with a varying SCF seems to perform better than the other models for CSS, whereas VIC seems to perform better for Copper.

4.3 Multi-Objective Model Calibration Results

Multi-objective calibrations of SNOW-17, including and not including SCF, and VIC are performed to further analyze the tradeoffs among objective functions and identify Pareto-approximate parameter sets, which in theory, should perform better than the parameterizations within the Sobol' sample that had the best objective function performance. Figures 4.7 (a) and (b) illustrate multi-objective calibration results for SNOW-17 including SCF and not including SCF, respectively. The axes of the three-dimensional (3-D) plots in Figure 4.7 represent values of the NSE, length and peak objective functions. The colorbar represents the value of RMSE for each solution. The arrows show direction of increasing preference (i.e. maximizing NSE, and minimizing length and peak), the color also shows increasing preference for the RMSE objective function, and the orange stars indicate the location of what is considered the optimal solution on the plots.

There seem to be two distinct clusters of solutions on the plot for SNOW-17 with SCF results (Figure 4.7 (a)). The cluster labeled (i) has good NSE, RMSE, and peak, but bad length values, while the cluster labeled (ii) has better peak values, but worse NSE, RMSE, and length values. Figure 4.7 (b) shows eliminating SCF from calibration eliminates the set of best performing solutions with respect to NSE and RMSE, and increases the performance of results with respect to the peak objective function. This indicates that including SCF can produce better values than not including it. It also demonstrates the importance of using several objective functions and deciding which ones are most important for a specific modeling goal.

Figure 4.8 is the same plot as Figure 4.7 (a) except the RMSE values are replaced by SCF values. The entire range of SCF values specified in Borg is not represented in the Borg results, which indicate that SCF values outside the range of 0.77 to 0.97 do not produce good results for

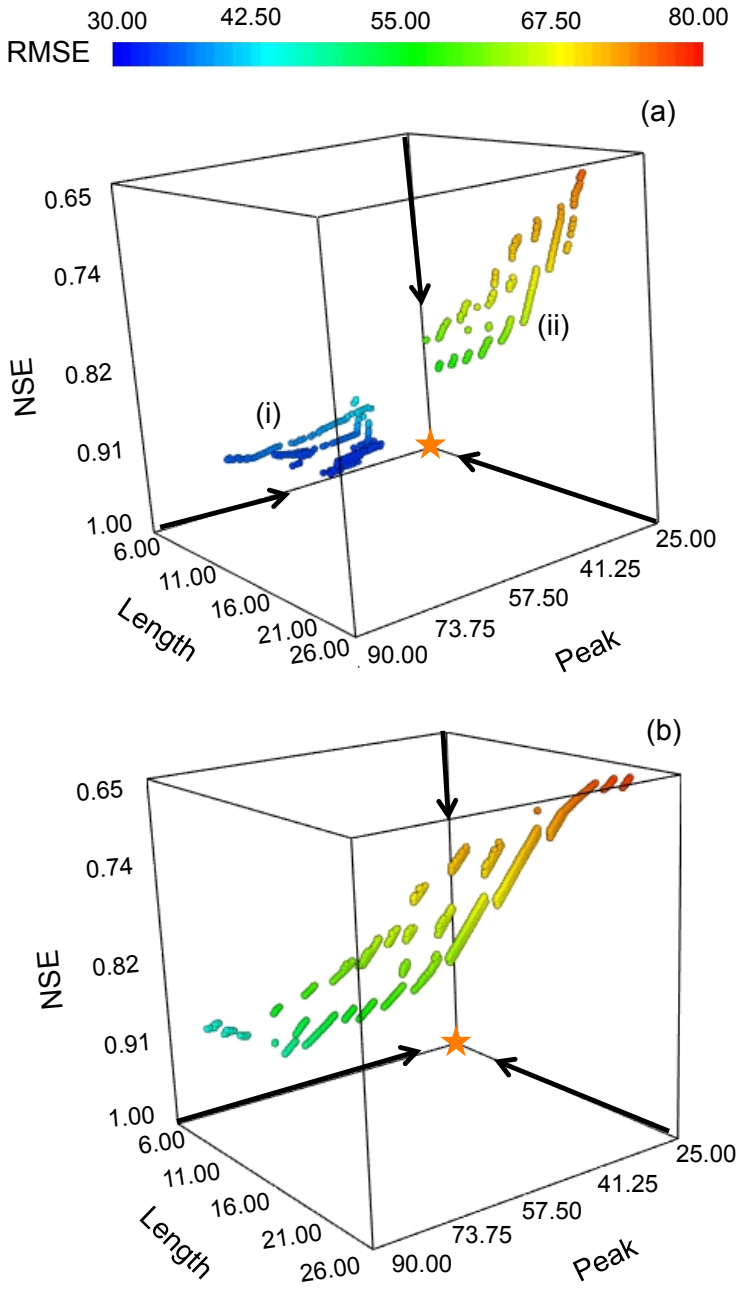


Figure 4.7: 3-D plots of SNOW-17 calibration results for Copper. Plot (a) depicts results when SCF was included in the calibration, and plot(b) depicts results when SCF was eliminated from calibration. The arrows point in the direction of preference, and the orange star indicates the optimal solution.

Table 4.2: Parameter and objective function values for three solutions resulting from the SNOW-17 multi-objective calibration.

	Solution 1	Solution 2	Solution 3
SCF	0.973	0.772	0.882
UADJ	0.030	0.030	0.057
MBASE	0.032	0.153	0.003
MFMAX	1.895	0.859	0.601
MFMIN	0.427	0.500	0.499
TIPM	0.175	0.050	0.254
NMF	0.500	0.500	0.499
PLWHC	0.020	0.268	0.023
Peak	26.89	45.45	64.12
Length	22	7	13
NSE	0.67	0.92	0.94
RMSE	75.55	41.66	32.55

Copper. This is consistent with Figure 4.3 which shows the SCF values for the best performing Sobol' are also within this same range. Also, there are groupings of SCF values in Figure 4.8, which indicate that certain SCF thresholds can really push a solution toward a specific desired objective function. The SCF values are all below the value of 1, which indicates over-catch may be an issue at this particular site location.

Figure 4.9 depicts the multi-objective calibration results for VIC at Copper, where arrows indicate direction of increasing preference. The VIC objective function values for the parameter sets have a much larger range than the SNOW-17 results. However, VIC also seems to have solutions with better NSE, RMSE, and length than SNOW-17, but not better peak. This indicates VIC may have issues with consistently underestimating peak SWE. The Pareto-approximate optimal set of solutions is generally concentrated in the same area on the plot, with some outliers that performed particularly well in peak or length, but not very well with respect to NSE and RMSE. This demonstrates the importance of selecting appropriate objective functions and defining the goal of modeling because if a modeler is particularly interested in only peak, the outlier parameter sets could be the best choice. However, there are parameter sets that performed almost as well in the peak objective function, but also have much better NSE, RMSE, and length values.

Tables 4.2, 4.3, and 4.4 display model parameter and objective function values for three solutions obtained in each multi-objective calibration (SNOW-17, SNOW-17 with a fixed SCF,

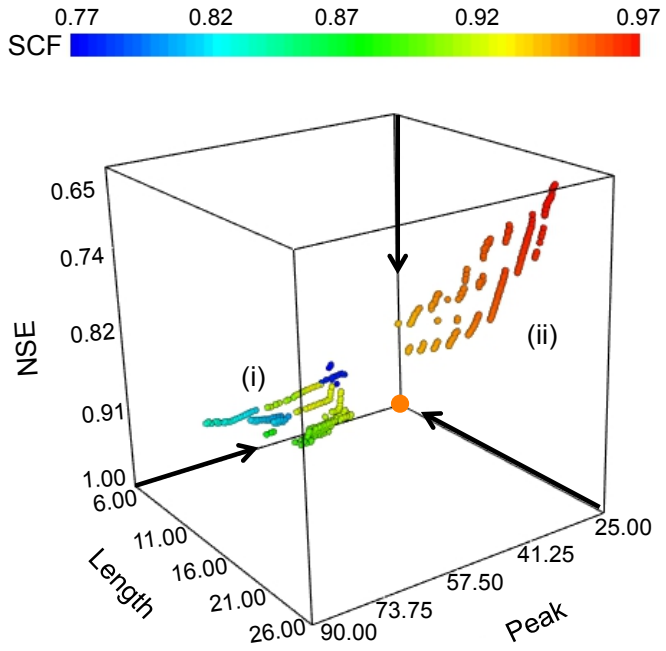


Figure 4.8: 3-D plot of SNOW-17 calibration results, including SCF, for Copper. This is the same plot as Figure 4.7 (a) except the RMSE values are replaced by SCF values. The arrows point in the direction of preference, and the orange star indicates the optimal solution.

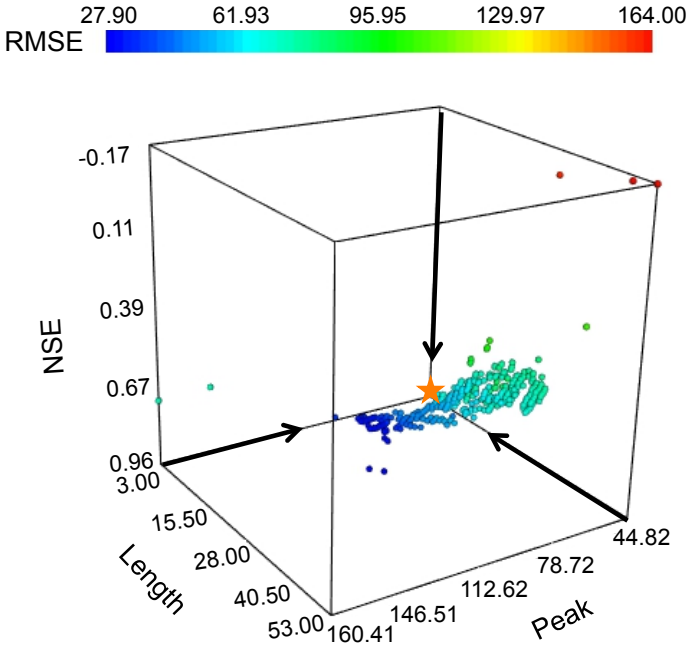


Figure 4.9: 3-D plots of VIC calibration results for Copper. The arrows point in the direction of preference, and the orange star indicates the optimal solution.

Table 4.3: Parameter and objective function values for three solutions resulting from the SNOW-17 (no SCF) multi-objective calibration.

	Solution 1	Solution 2	Solution 3
SCF	1.000	1.000	1.000
UADJ	0.030	0.106	0.056
MBASE	0.000	0.000	0.000
MFMAX	2.000	0.887	0.893
MFMIN	0.500	0.500	0.499
TIPM	0.179	0.180	0.118
NMF	0.500	0.314	0.216
PLWHC	0.020	0.020	0.020
Peak	30.36	84.03	85.64
Length	24	8	8
NSE	0.65	0.89	0.90
RMSE	77.69	44.10	43.61

Table 4.4: Parameter and objective function values for three solutions resulting from the VIC multi-objective calibration.

	Solution 1	Solution 2	Solution 3
Density	0.024	0.006	0.001
Roughness	0.124	0.118	0.150
LWHC	109.379	87.904	96.528
Albedo	0.817	0.866	0.882
Accum. A	0.990	0.444	0.645
Accum. B	0.945	0.457	0.376
Thaw A	0.961	0.768	0.756
Thaw B	0.926	0.727	0.100
Peak	44.82	156.90	109.19
Length	53	3	21
NSE	-0.17	0.71	0.95
RMSE	164.00	74.41	27.90

and VIC). The tables demonstrate the ability of an MOEA to find high-performing parameter sets with respect to different objectives. The VIC parameter sets also demonstrate a variety of historically-fixed parameter values that are able to produce better performing results than the baseline results with respect to specific objectives. Although the sensitivity analysis allows for analysis of parameterizations with respect to objectives, it does not identify a set of approximate optimal-performing parameterizations with respect to several objectives.

4.4 Comparison of Sensitivity Analysis and Multi-Objective Calibration

Table 4.5 compares the best objective values achieved in the Sobol' analyses and multi-objective calibrations for Copper. The goal is to minimize the peak, length, and RMSE functions, and maximize NSE to a value of 1. For all objective functions, the multi-objective calibration produces values equal to or better than the Sobol' results, except for the VIC peak function, in which Sobol' performs only slightly better than the calibration. A larger set of function evaluations for the multi-objective calibration may eventually result in a calibration peak value that is better than the Sobol' peak value. Another interesting observation is that VIC results in the best objective function values for all objective functions for both the Sobol' analysis and multi-objective calibration, except for the peak calibration value, where SNOW-17 producs a better value in both the SCF and non-SCF calibrations. This is consistent with the observations of the multi-objective calibration plots. Figure 4.8 showed the best-performing SNOW-17 parameter sets with respect to the peak objective function had a very specific range of SCF values, which happen to be close to the fixed value of 1 used for SCF in the second SNOW-17 calibration. This indicates SNOW-17 performs well in terms of peak with little to no tweaking of the SCF parameter at the Copper site.

4.5 Climate Change Results

A temperature change scenario and a precipitation scenario are applied to each site, as discussed in Chapter 2, Section 2.5. The precipitation scenario does not produce intriguing results, likely due to the fact that some GCMs say precipitation will increase while other GCMs say precip-

Table 4.5: Best performing objective function values resulting from the Sobol' and multi-objective calibration for Copper. The values are not necessarily from the same model run or water year because the worst value out of all five water years for each model run was chosen to represent the value for that particular model run.

	SNOW-17 (SCF)		SNOW-17 (no SCF)		VIC	
	Sobol'	Calibration	Sobol'	Calibration	Sobol'	Calibration
Peak (mm)	48.7	26.9	56.1	30.4	44.0	44.89
Length (days)	8	7	8	8	3	3
NSE	0.936	0.951	0.884	0.899	0.946	0.955
RMSE (mm)	36.0	32.5	47.7	43.6	33.3	27.9

itation will decrease, whereas GCMs agree on an increase in temperature, but differ in magnitude of this increase. The contrasting GCM precipitation outputs likely cancel each other out, and thus result in minimal differences between control model outputs and precipitation change scenario outputs. Results focus on the temperature change scenario. Figures 4.10 (a) and (b) and 4.11 (a) and (b) depict the daily differences in SWE between the control SNOW-17 and VIC model outputs, generated during the Sobol' analysis, and the temperature change scenarios for the same model parameterizations. Only the model ensemble with a fixed SCF is used for the SNOW-17 climate scenarios due to nonphysical effects of SCF on model outputs. The timeseries plot in each figure shows the site-specific observed SWE for the water years used in this research. The top raster in each figure shows daily SWE differences for the top 2 % performing control model runs (320 model runs for SNOW-17 with a fixed SCF and 360 model runs for VIC) with respect to NSE, and the bottom rasters show daily SWE differences for the the worst 2% performing control model runs with respect to NSE. Each column in the rasters represents a day, and each row in the rasters represents a model parameterization, and the color represents the daily SWE difference between the control modeled SWE and temperature change modeled SWE. The two extreme sets of model runs are chosen to examine how errors in parameterizations propagate to portrayals of climate change.

Since there are no future data to compare climate scenario model outputs, it is difficult to determine the quality of a modeled climate change scenario. However, if we trust model parame-

terizations, then we can gain some insight into how SWE will change. Figures 4.10 and 4.11 show the largest difference in daily SWE occurs during peak season (June), while minimal to no daily differences in SWE occur when there is little to no snow. Figures 4.10 (a) and (b) show SNOW-17 produces a similar signal for the best and worst model parameterizations with respect to NSE. Squaw has much larger daily SWE differences in terms of magnitude and length of the number of days the differences occur. This indicates the more snow a site receives, the larger the impact of a temperature change scenario.

The VIC temperature change results for Copper, presented in Figure 4.11 (a), show the top 2% model runs maintain a similar pattern to the SNOW-17 results, where the large differences are during the peak, and there are very minimal daily differences when there is very little to no snow. However, the bottom 2% model runs for Copper essentially lose any pattern and have a constant daily difference throughout all water years. Inspection of the worst performing VIC parameterizations reveals these parameterizations produce minimal SWE and do not maintain the observed SWE pattern, and instead have more variations in the accumulation and ablation of SWE. These parameterizations produce such minimal daily SWE that VIC is not able to maintain the original SWE pattern, and thus also does not maintain the temperature change results pattern that exists for the good parameterizations. This does not happen at the Squaw site for the VIC temperature change scenarios, presented in Figure 4.11 (b). However, there are differences between the Squaw SNOW-17 results and VIC results. The largest daily differences in SWE occur in the top performing model runs versus the worst performing for SNOW-17. Also, the daily SWE differences become very small during water year four, which seems to be an abnormally dry year in terms of SWE accumulation. This seems to be similar to what is happening in the bottom 2% Copper runs. This indicates that VIC has trouble projecting climate change for very low SWE accumulation.

The small amount of variation in daily differences in SWE between the top and bottom model runs for SNOW-17 suggests the SNOW-17 parameter values do not greatly affect the delta-change in daily SWE. Regardless of control model performance, SNOW-17 will produce a similar delta-change in SWE. The variation between the top and bottom model runs for VIC suggests that

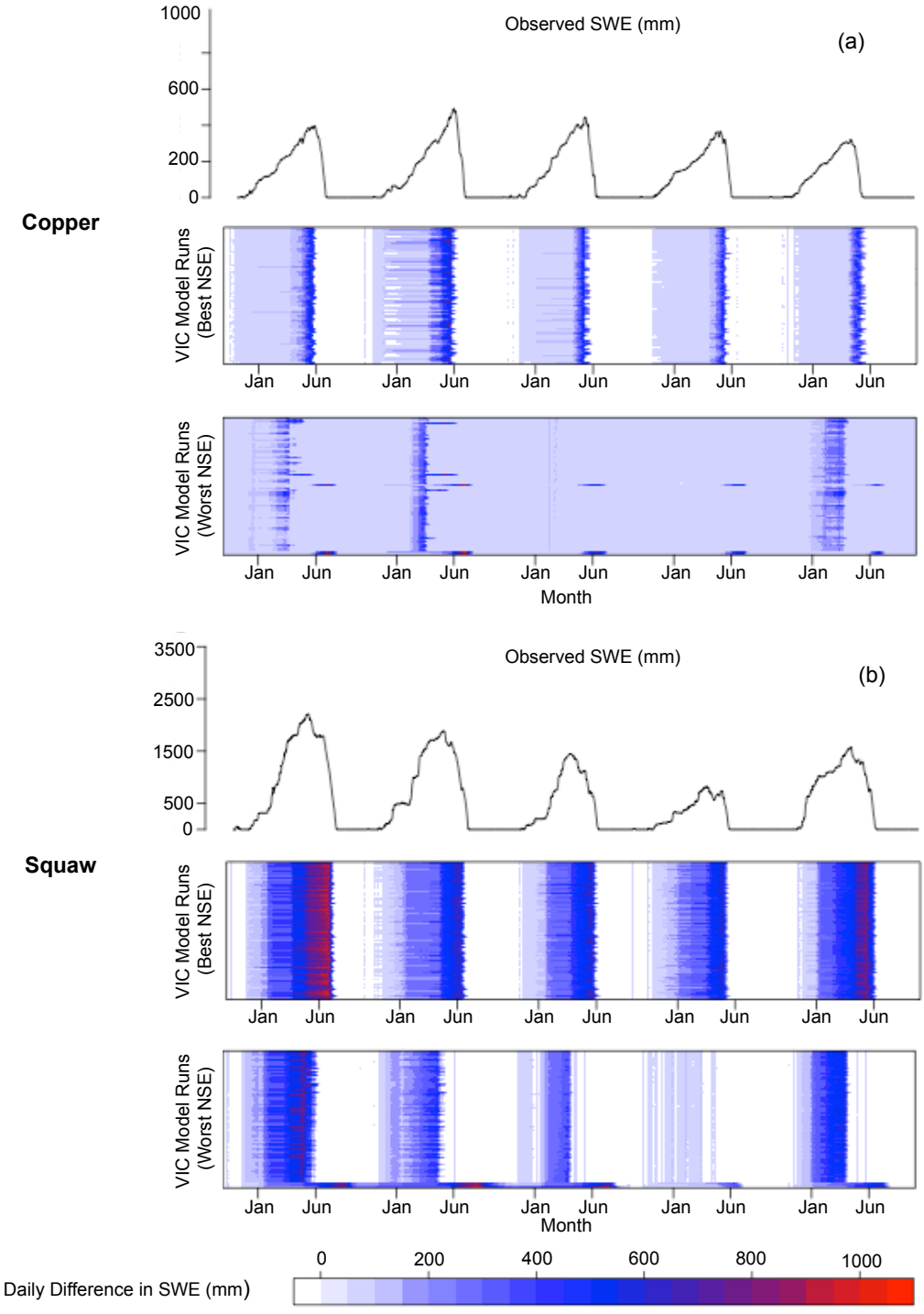


Figure 4.10: Rasters showing the daily difference in SWE between the original SNOW-17 ensembles and the temperature change results. Copper and Squaw were chosen to compare different climate regimes. The top raster for each site depicts differences for the top 2% of model runs with respect to NSE from the Sobol' analysis. The bottom rasters depict differences for the worst 2% of model runs with respect to NSE.

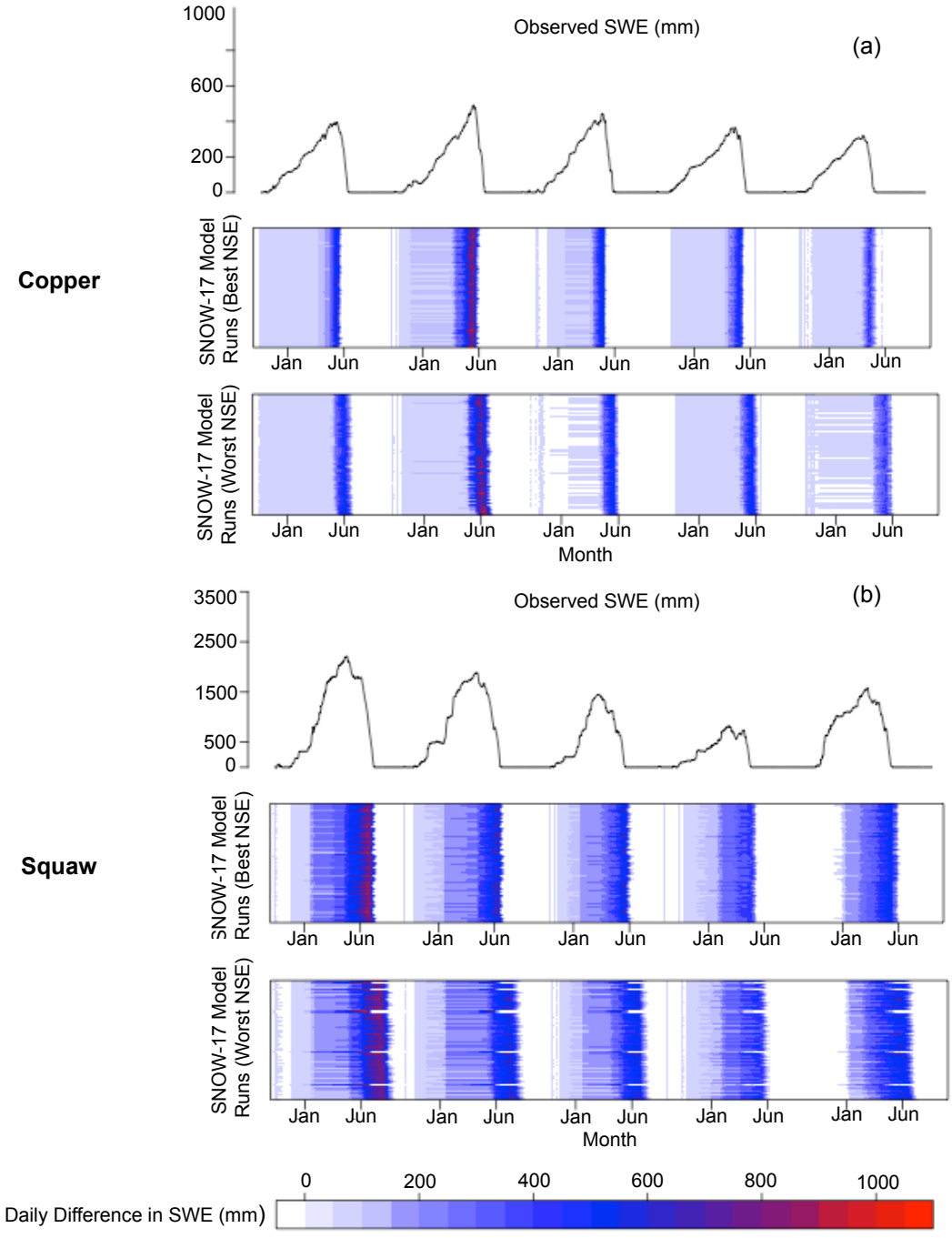


Figure 4.11: Rasters showing the daily difference in SWE between the original VIC SWE ensembles and the temperature change results. Copper and Squaw were chosen to compare different climate regimes. The top raster for each site depicts differences for the top 2% of model runs with respect to NSE from the Sobol' analysis. The bottom rasters depict differences for the worst 2% of model runs with respect to NSE.

VIC parameter values can have a large impact on the SWE delta-change. The parameters in the bottom VIC model runs likely have values that disturb the physics in the model and force it to produce an unlikely temperature change result. The consistent pattern of SWE daily differences that occurs throughout the SNOW-17 model runs seems unrealistic, and indicates SNOW-17 may not model real-world processes on a fine enough scale to accurately capture climate change at a specific site. Lack of sensitivity of SNOW-17 temperature change scenario results, could be due to other hard-coded parameters within the model or incomplete physics due to the nature of the conceptual model.

Chapter 5

Discussion, Conclusions, and Future Work

5.1 Discussion

Hydrological models are used to project future water resources conditions and make important management decisions based at least partially on these projections. However, hydrological modeling uncertainty ultimately propagates into model output, and as a result, there are significant implications of this uncertainty in water management decisions. Minimizing uncertainty in decision-making, which requires minimizing uncertainty in monitoring and modeling, is one of the main drivers for advancing hydrological models. This research addressed the following questions (1) Do complex physically based models perform as well or better than simple conceptual models? (2) How do we identify well-performing parameter sets with respect to several, conflicting objectives? (3) How do errors in parameterizations propagate to portrayals of climate change? This research explored the impact of using global model parameter sensitivity and calibration versus local model parameter sensitivity and calibration.

Development of complex physically based hydrological models, in theory, should result in improved model outputs compared to simple conceptual models, however, this is not always the case. Parameter values within physically-based models are based on *a priori* constraints because they are assumed to represent real-world values, whereas conceptual model parameters do not have these constraints because they cannot be measured in the real world (*Mendoza et al.*, 2015). The limited flexibility in physically based models results in model outputs that can only perform so well. For example, the baseline results presented in Chapter 4 showed that the physically based

model outputs had some obvious deviations from the observed SWE. Since the model parameters within the physically based model have historically been fixed, this baseline result would have been the best the model could have performed. As a result, the errors in SWE would propagate through the flow results and eventually research conclusions and management decisions. Conceptual model parameters are known to contain a large amount of uncertainty, and as a result model users spend time calibrating conceptual models and adjusting parameters to be site-specific. However, as demonstrated in *Mendoza et al.* (2015) and this research, physically based model parameters can also contain a large amount of uncertainty. Exposing these parameters within physically based models is crucial for future model development and improvement (*Mendoza et al.*, 2015).

Ensembles of SNOW-17 and VIC outputs demonstrated that changing model parameters, some historically calibrated and some historically fixed, can produce a large ensemble of model outputs, some good and some unrealistic. The Sobol' sensitivity analysis took this analysis a step further by revealing sensitive parameters, which have the largest impact on model output variance. In both models, several parameters had very high sensitivities and seemed to be controlling the model outputs. For example in SNOW-17, the SCF parameter was by far the most sensitive parameter, which is logical because SCF is a correction factor. When the SCF parameter was eliminated from the ensembles and set to a fixed value of 1, the Sobol' sensitivities of other parameters, particularly MFMAX, increased, and previously unrealistic model runs that exceeded accumulated precipitation were eliminated. MFMAX represents the maximum melt factor that will occur (on June 21^{st}) during the melt season, which has a large effect on how fast the snow melts during non-rain periods, thus affecting model output during the melt season. The MFMAX parameter also affects the energy exchange when there is no surface melt. The larger MFMAX, the smaller the change in heat deficit within the snowpack due to a temperature gradient (*Anderson*, 2006). The SNOW-17 PLWHC parameter also became more sensitive with respect to the peak objective function when SCF was set at a a fixed value. This parameter represents the percent liquid water holding capacity of the snowpack, or the maximum amount of water a snowpack can hold. After SCF was removed, PLWHC became the dominant parameter with respect to peak, likely due to

the fact that PLWHC dictates the amount of water a snowpack can hold before runoff will occur, and as a result dictates the accumulation of SWE.

VIC ensemble and Sobol' analysis results were particularly significant because this was a novel exploration of historically fixed VIC snow model parameters. Varying these parameters resulted in a large range of model outputs, and Sobol' revealed that two parameters in particular were very sensitive. This reveals that even in complex physically-based models, parameter uncertainty exists, which is to be expected, and this uncertainty can largely impact the model output. Fixed parameters within complex models can constrain the models, as discussed in *Mendoza et al.* (2015). This proves there is large uncertainty in model parameter values, which can propagate into the model output. It is important to recognize uncertainty of different model parameters and to maintain flexibility to model different regimes (*Mendoza et al.*, 2015).

Sobol' results for both SNOW-17 and VIC were different across objectives, showing the value of including multiple objectives in the study. Parameters have different sensitivities with respect to different objective functions, and thus it is important for a modeler to realize the goal of a modeling project and use appropriate objective functions. Results were also different across site locations. VIC results especially showed similar total order sensitivity patterns among sites with similar climates (e.g. Copper and Galena or CSS, Squaw and Hart). This indicates parameters affect model outputs differently depending on site location. For example, a site with more accurate precipitation data may be less dependent on the SCF value than a site with poor data, and thus the SCF parameter may have a lower sensitivity for that site. These results demonstrate that all parameters within a model are not necessarily important. Our study therefore underscores the fact that objective functions should be chosen carefully to capture the modeling aspects that are desired by the user.

Increasing model complexity does not necessarily lead to better results due to parameter uncertainty and traditional model calibration techniques. *Gupta* (1998) discussed how the classical approach to model calibration has limitations, especially with the development of more complex models, because they do not recognize that model identification and calibration are inherently

multi-objective problems. Conceptual models are easier to understand and manipulate, and as a result, users are able to gain a sense of how the model works using traditional diagnostic techniques, although more sophisticated modern techniques, such as multi-objective calibration, provide additional information regarding model operation.

Variations in results among the three climate regimes indicate site climate characteristics can be found in the signal of model parameterizations. For example, the SCF parameter for the top performing Copper SNOW-17 model runs with respect to NSE had a distribution of values that was different form the Squaw SCF parameter distribution (Figure 4.3). It seems a maritime climate requires a larger value of SCF than a continental climate. Another example is the MFMAX parameter, where the maritime and continental climates had different parameter values and distributions. Also, for Galena, the inter-mountain climate, MFMAX sensitivities with respect to RMSE, NSE and length exceeded the sensitivities of the other non-intermountain sites. VIC Sobol' results also indicate the three climate regimes have different parameter sensitivity patterns. The three maritime climates have similar patterns, where the continental and inter-mountain climate patterns differed from he maritime patterns, but were somewhat similar to each other. These results indicate model diagnostics and parameter values can provide valuable insight when either attempting to calibrate a model for an ungauged site location with a known climatic regime (*Sivapalan et al.*, 2003).

Multi-objective calibration of SNOW-17 produced an approximation of the Pareto-optimal set of parameters for two calibration scenarios, one including SCF and one not including SCF. The scenario including SCF had two distinct areas, one with good RMSE, NSE, and length, but bad peak, and one with good peak, but bad NSE, RMSE and length. Eliminating SCF from the calibration, eliminated the parameter sets that resulted in the best performing RMSE and NSE values. Instead, there is a linear tradeoff between peak and the other objectives (NSE, RMSE, and length), making it impossible to minimize peak while also minimizing, RMSE and length, and maximizing NSE. Although removing SCF from the ensemble earlier in the research removed the unrealistic results, it seems that removing SCF also eliminated the best performing results in terms of length, RMSE, and NSE. The calibration results also showed that the two distinct areas in the

calibration including SCF, contain groupings of SCF values. The results with good length, RMSE, and NSE contained SCF values of approximately 0.82 to 0.92, where the other set of results had SCF values of approximately 0.92 to 0.97. Lastly, all calibration solutions contain SCF values less than 1, which indicates there is an over-catch issue at Copper site location.

Multi-objective calibration of VIC produced a Pareto-optimal set with larger ranges of objective function values than SNOW-17, but included solutions with improved values of NSE, RMSE, and length. The solutions were generally concentrated in one area, with good NSE, RMSE, and length. There were some outliers that performed particularly well in length or peak, but poorly in the other objectives. Although SNOW-17 had more solutions with better peak values, there was a tradeoff between the peak objective functions and the remaining objective functions. The SCF parameter seemed to dictate the performance of the peak objective function, and the larger the deviation from a value of 1, the worse the peak value. The contrast between the two model's approximate Pareto-optimal solutions illustrates the importance of specifying modeling goals. If the only goal is to minimize the difference in peak at all costs, then perhaps SNOW-17 would be a better choice, but if the modeler is interested in minimizing multiple objectives, then perhaps VIC would be better. VIC also provides a physical basis for results.

Creating hydrological projections with respect to climate change is challenging due to disagreement among GCMs and the inability to validate model outputs. The climate portion of this research aimed to reveal information on how model parameterization errors propagate to portrayals of temperature change. Results were presented as daily differences in SWE between the control model runs and the temperature change model runs. The top performing control model runs maintained a consistent pattern, in which large SWE differences occurred during the peak accumulation season and minimal to no SWE differences occurred when there was little to no SWE, throughout sites and models. SNOW-17 produced similar signals between the best and worst performing control model runs, but we do not know if these signals faithfully represent future conditions in a perturbed climate. VIC demonstrated variation in the daily SWE delta-change throughout different model runs. For poor performing control model runs with lower SWE accumulation, VIC

did not maintain the original pattern and instead projected a uniform daily difference in SWE throughout all of the water years. When the physics are parameterized correctly for VIC, the patterns in daily SWE differences seem more realistic than when parameterized poorly. Further investigation revealed the poor parameterizations in VIC greatly underestimated SWE and did not maintain the typical SWE accumulation and ablation pattern, resulting in issues with maintaining the signals produced by the well-performing control results. Whereas poor performing SNOW-17 parameterizations overestimated SWE, maintaining the same SWE behavior as the observed, and resulting in a consistent daily change in SWE pattern throughout all parameterizations. We do not know what is going to happen in the future, but if we trust model parameterizations than we can gain some insight into how SWE will change.

5.2 Conclusions

Without the ability to continuously produce accurate SWE projections, it is difficult to create reliable long-term water management solutions. Continued research on understanding and improving hydrological model performance, especially with respect to applying sophisticated diagnostic tools and integrating them into the operational world, is necessary to improve modeling in this changing world. Current modeling practices do not always deliver the relevant information required for water management decisions. One example of this is discussed in *Staudinger et al.* (2014), which explores the ability of the Standardized Precipitation Index (SPI), the most widely used index to characterize droughts, to differentiate between meteorological and hydrological droughts. In snow-dominated regions, precipitation is temporarily stored as snow, and as a result, there is a significant difference between meteorological and hydrological droughts because of the delayed release of melt water to rivers. Similarly, snow modeling practices, particularly associated with complex models, do not use diagnostic tools that provide a clear understanding of how data, model parameter, and model structural uncertainties propagate through the model in to outputs. Like the SPI, which incorrectly characterized droughts in snow-dominated regions, complex snow models are incorrectly modeling SWE due to uncharacterized uncertainty.

Throughout the research process, it was observed that modeled SWE performed poorly for model runs with poor parameter calibration. Using a poorly calibrated model for water management decisions could result in incorrect estimates of available water resources. As climate changes and population increases, water resources planning becomes more important to ensure communities have a sustainable water supply. Accurate snow projections will provide insight regarding the possibility of extreme circumstances, and will improve managers decision-making options for long-term water resources management. This research demonstrated that fixed parameters within a complex model have uncertainty that can affect the model output, and showed how sophisticated tools can tell us how uncertainty in model parameters propagates to model outputs. These insights can be used to develop and improve models that allow for user-friendly exploration of parameter uncertainties and improve modeling practices to create more robust management decisions.

5.3 Future Work

This research lays the foundation for future work to explore the effects of conceptual and complex snow model parameterizations on streamflow outputs of hydrological models for snowmelt-dominated basins. A major motivation for modeling SWE and streamflows is to project future water resources and make important management decisions based on these projections. Understanding parameter uncertainty and how it propagates to model outputs is important, and this information should make it to water managers so they are aware of model output uncertainty. An additional step to explore how different parameterizations affect approximate Pareto-optimal sets of water resources management solutions would provide additional insight into how parameter uncertainties can propagate to management decisions. However, it seems reasonable that reservoir operation and management decisions in a snowmelt-dominated basin would be affected by SWE projections. Additionally, it would be useful to explore the sensitivity of management solutions to the quality of model outputs to answer the question of: how accurate does a snow model really need to be in order to change a management decision? Does a difference of several millimeters in peak SWE across an entire basin, greatly affect water resources and reservoir operations? Such an analysis

could start simple by using only a point location and applying the error over the entire basin. If results seem promising, the research could be expanded to a more realistic scenario in which the snow and hydrological models are applied across areas instead of point locations.

A similar extension of this research could be applied to the temperature change scenarios and results. The ensemble of climate change scenarios could be coupled with hydrological models to quantify how particular changes in magnitude and timing of snowmelt affect streamflows. The SNOW-17 temperature change results in this study had similar daily delta SWE changes from the control runs throughout all model parameterizations, where VIC daily delta SWE results varied throughout the different model parameterizations. This is due to the fact that the worst-performing paramterizations for VIC are much worse than those for SNOW-17, and thus VIC is not able to simulate observed SWE behavior in these poorly performing model runs. An interesting analysis would compare streamflows resulting from the control model runs and from the temperature change model runs to see if the delta changes in streamflows have similar patterns to the delta changes in SWE for both models. The ability to quantify some of the uncertainty regarding how future snowpacks will respond to temperature change can provide a lot of utility in the research and management world.

Lastly, further exploration of VIC model parameters, or fixed parameters in other complex models, using sophisticated diagnostic tools, is necessary for improved performance of complex models. Uncertain and sensitive parameters within these models must be recognized, and future model improvements must incorporate this information in order to create user-friendly methods to adjust sensitive parameter values. This will give more flexibility to complex models versus the current state of such models which requires adjustment of hard-coded parameters to achieve such flexibility. *Mendoza et al.* (2015) also discussed the importance of integrating mechanisms for uncertainty quantification and analysis to facilitate diagnosis of model problems, refine model representations of natural processes, and understand major sources of uncertainty. The diagnostic tools used in this research could somehow be integrated, in a more user-friendly manner, into models and common modeling practices. Lastly, these tools could be used to effectively identify default

parameter values for complex models.

Bibliography

[1] D. Aherns. Meteorology Today: An Introduction to Weather, Climate and the Environment. Cengage Learning, 2012.

[2] E. Anderson. National weather service river forecast system: Snow accumulation and ablation model. NOAA Technical Memorandum NWS Hydro-17. U.S. National Weather Service, 1973.

[3] E. Anderson. Calibration of conceptual hydrologic models for use in river forecasting. Office of Hydrologic Development. National Weather Service, Silver Springs, MD., 2002.

[4] E. Anderson. National weather service river forecast system: Snow accumulation and ablation model - snow-17. NOAA Technical Memorandum NWS Hydro-17, 2006.

[5] K. Andreadis, P. Storck, and D. Lettenmaier. Modeling snow accumulation and ablation processes in forested environments. Water Resources Research, 45:W05429, 2009.

[6] T. P. Barnett, J.C. Adam, and D. P. Lettenmaier. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature, 438:303–309, 2005.

[7] S. Bergstrom. Principles and confidence in hydrological modelling. Nordic Hydrology, 22:123–136, 1991.

[8] K. J. Beven. Prophecy, reality, and uncertainty in distributed hydrological modelling. Advances in Water Resources, 16:41–51, 1993.

[9] K. J. Beven. Equifinality and uncertainty in geomorphological modelling. Advances in Water Resources, 16:41–51, 1996a.

[10] K. J. Beven. A discussion of distributed hydrological modelling. Advances in Water Resources, pages 255–278, 1996b.

[11] K. J. Beven. Uniqueness of place and the representation of hydrological processes. Hydrology and Earth Systems Sciences, 4(2):203–213, 2000.

[12] K. J. Beven. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrological Processes, 16:189–206, 2002.

[13] K. J. Beven and A. Binley. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6:279–298, 1992.

[14] D. P. Boyle, V. Gupta, and S. Sorooshian. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research, 36:3663–3674, 2000.

[15] R. Bras. Hydrology: An Introduction to Hydrological Science. Addison-Wesley Publishing Company, 1990.

[16] F. Chen, M. Barlage, M. Tewari, R. Rasmussen, J. Jin, D. Lettenmaier, B. Livneh, C. Lin, G. Miguez-Macho, G. Niu, L. Wen, and Z. Yang. Modeling seasonal snowpack evolution in the complex terrain and forested colorado headwaters region: A model intercomparison study. Journal of Geophysical Research: Atmospheres, 119:13,795–13,819, 2014.

[17] K. Cherkauer, L. Bowling, and D. Lettenmaier. Variable infiltration capacity cold land process model updates. Global and Planetary Change, 38:151–159, 2003.

[18] M. Clark, D. Kavetski, and F. Fenicia. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47:W09301, 2011.

[19] M. Clark, A. Slater, D. Rupp, R. Woods, J. Vrugt, H. Gupta, T. Wagener, and L. Hay. Framework for understanding structural errors (fuse): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44:W00B02, 2008.

[20] E. Demaria, B. Nijssen, and T. Wagener. Monte carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model. Journal of Geophysical Research, 112:D11113, 2007.

[21] Q. Duan, S. Sorooshian, and V. Gupta. Optimal use of the sce-ua global optimization method for calibrating watershed models. Journal of Hydrology, 158:265–284, 1994.

[22] A. Efstratiadis and D. Koutsoyiannis. One decade of multiobjective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal, 55(1):58–78, 2010.

[23] R. Essery, S. Morin, Y. Lejunne, and C. Menard. A comparison of 1701 snow models using observations from an alpine site. Advances in Water Resources, 55:131–148, 2013.

[24] P. Etchevers, E. Martin, R. Brown, C. Fierz, Y. Lejeune, E. Bazile, A. Boone, Y. Dai, R. Essery, A. Fernandez, Y. Gusev, R. Jordan, V. Koren, E. Kowalczyk, N. Nasonova, R. Pyles, A. Schilosser, A. Shmakin, T. Smirnova, U. Strasser, D. Verseghy, T. Yamazaki, and Z. Yang. Validation of the energy budget of an alpine snowpack simulated by several snow models (snowmip project). Annals of Glaciology, 38:150–158, 2004.

[25] X. Feng, A. Sahoo, K. Arsenault, P. Houser, Y. Luo, and T. Troy. The impact of snow model complexity at three clpx sites. Journal of Hydrometeorology, 9:1464–1481, 2008.

[26] L. Foglia, M. Hill, S. Mehl, and P. Burlando. Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. Water Resources Research, 45:W06427, 2009.

[27] H. Fowler, S. Blenkinsop, and C. Tibaldi. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Water Resources Research, 45:W06427, 2009.

[28] K. Franz, T. Hoduge, and S. Sorooshian. Operational snow modeling: addressing the challenges of an energy balance model for national weather service forecasts. Journal of Hydrology, 360:48–66, 2008.

[29] H. Gupta, M. Clark, J. Vrugt, J. Abramowitz, and M. Ye. Towards a comprehensive assessment of model structural adequacy. Water Resources Research, 48:W08301, 2012.

[30] H. Gupta, H. Kling, K. Yilmaz, and G. Martinez. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modeling. Journal of Hydrology, 377:80–91, 2009.

[31] H. Gupta, S. Sorooshian, and P. Yapo. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research, 34:751–763, 1998.

[32] H. Gupta, T. Wagener, and Y. Liu. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes, 22:3802–3813, 2008.

[33] D. Hadka and P. M. Reed. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. Evolutionary Computation, 20:423–452, 2012.

[34] D. Hadka and P. M. Reed. Borg: An auto-adaptive many-objective framework. Evolutionary Computation, 5212:213–259, 2013.

[35] L. Hay, R. Wilby, and G. Leavesley. A comparison of delta change and downscaled gcm scenarios for three mountainous basins in the united states. Journal of the American Water Resources Association, 36(2):387–397, 2000.

[36] M. He, T. Hogue, K. Franz, S. Margulis, and J. Vrugt. Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. Advances in Water Resources, 34:114–127, 2011a.

[37] M. He, T. Hogue, K. Franz, S. Margulis, and J. Vrugt. Corruption of parameter behavior and regionalization by model and forcing data errors: A bayesian example using the snow17 model. Water Resources Research, 47:W07546, 2011b.

[38] A. Henderson-Sellers and K. McGuffie. The Future of the World's Climate. Elsevier Science, 2012.

[39] J. Herman. What matters in a model? In BEE Seminar. Cornell University, 2014.

[40] T. Hogue, H. Gupta an dS. Sorooshian, and C. Tomkins. A multistep automatic calibration scheme for watershed models calibration of wathershed models. Journal of Hydrometeorology, 1:524–542, 2000.

[41] T. Hogue, L. Bastidas, H. Gupta, and S. Sorooshian. Evaluating model performance and parameter behavior for varying levels os land surface model complexity. Water Resources Research, 42:W08430, 2006a.

[42] T. Hogue, H. Gupta, and S. Sorooshian. A 'user-friendly' approach to parameter estimation in hydrologic models. Journal of Hydrology, 320:202–217, 2006b.

[43] T. Hogue, H. Gupta, S. Sorooshian, and C. Tomkins. A multi-step automatic calibration scheme for watershed models. In: Duan, Q., Gupta, H., Sorooshian, S. Rousseau, S., Turcotte, A., (Eds.) Advances in Calibration of Watershed models, Water Science and Application 6, American Geophysical Union, Washington, D.C., pp. 165–174.

[44] T. Hogue, S. Sorooshian, and H. Gupta. A multistep automatic calibration scheme for river forecasting models. Journal of Hydrometeorology, 1:524–542, 2000.

[45] J. Kasprzyk, P. Reed, B. Kirsch, and G. Characklis. Many-objective de novo water supply portfolio planning under deep uncertainty. Environmental Modelling and Software, 34:87–104, 2012.

[46] D. Kavetski, G. Kuczera, and S. Franks. Calibration of conceptual hydrological models revisited: 1. overcoming numerical artefacts. Journal of Hydrology, 320:173–186, 2006a.

[47] D. Kavetski, G. Kuczera, and S. Franks. Calibration of conceptual hydrological models revisited: 1. improving optimisation and analysis. Journal of Hydrology, 320:187–201, 2006b.

[48] J. W. Kirchner. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42:W03S04, 2006.

[49] R. Kirnbauer, G. Bloschl, and D. Gutkenecht. Entering the era of distributed snow models. Nordic Hydrology, 25:1–24, 1994.

[50] J. Kollat, P. Reed, and T. Wagener. When are multiobjective calibration trade-offs in hydrologic models meaningful? Water Resources Research, 48:W03520, 2012.

[51] J. J. Koskela, B. W. Croke, H. Koivusalo, A. J. Jakeman, and T. Kokkonen. Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment. Water Resources Research, 48:W11513, 2012.

[52] G. Kuczera, D. Kavetski, S. Franks, and M. Thyer. Towards a bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. Journal of Hydrology, 331:161–177, 2012.

[53] G. Kuczera and E. Parent. Monte carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. Journal of Hydrology, 211:69–85, 1998.

[54] M. Kumar, D. Marks, J. Dozier, M. Reba, and A. Winstral. Evaluation of distribute hydrologic impacts of temperature-index and energy-based snow models. Advances in Water Resources, 56:77–89, 2013.

[55] M Leisenring and H. Moradkhani. Snow water equivalent prediction using bayesian data assimilation methods. Stochastic Environmental Research Risk Assessment, 25:253–270, 2011.

[56] X. Liang, D. Lettenmaier, and E. Wood. One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. Journal of Geophysical Research, 101:21,403–21,422, 1996.

[57] X. Liang, Z. Xie, and M. Huang. A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (vic) land surface model. Journal of Geophysical Research, 108:721–738, 2003.

[58] B. Livneh, J. Deems, D. Schneider, J. Barsugli, and N. Molotch. Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. Water Resources Research, 50:8589–8610, 2014.

[59] B. Livneh, E. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. Andreadis, E. Maurer, and D. Lettenmaier. Long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states: Update and extensions. Water Resources Research, 26:9384–9392, 2013.

[60] B. Livneh and X. Xie. A new surface runoff parameterization with sub-grid scale soil heterogeneity for land surface models. Advances in Water Resources, 24:1173–1193, 2001.

[61] B. Livneh, X. Youlong, K. E. Mitchell, M. B. Ek, and D. P. Lettenmaier. Noah lsm snow model diagnostics and enhancements. Journal of Hydrometeorology, 11:721–738, 2010.

[62] S. Manabe. Climate and the ocean circulation. Monthly Weather Review, 97:775–805, 1969.

[63] S. Manandhar, V. P. Pandey, H. Ishidaira, and F. Kazama. Perturbation study of climate change impacts in a snow-fed river basin. Hydrological Processes, 27:3461–3474, 2013.

[64] P. A. Mendoza, M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta. Are we unnecessarily constraining the agility of complex process-based models? Water Resources Research, 51, 2014.

[65] N. Mizukami, M. P. Clark, A. G. Slater, L. D. Brekke, M. M. Elsner, J. R. Arnold, and S. Gangopadhyay. Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. Journal of Hydrometeorology, 15:474–488, 2014.

[66] P. W. Mote, A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier. Declining mountain snow pack in western north america. Bulletin of the American Meteorological Society, 86:39–49, 2005.

[67] G. Niu, Z. Yang, K. Mitchell, F. Chen, M. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia. The community noah land surface model with multiparameterization options (noah-mp): 1. model description and evaluation with local-scale measurements. Journal of Geophysical Research, 166:D12109, 2011.

[68] U.S. Army Corps of Engineers. Snow hydrology: Summary report of the snow investigations. U.S. Army of Engineers North Pacific Division, 437pp, 1956.

[69] L. Parada, J. Fram, and X. Liang. Multi-resolution calibration methodology for hydrologic models: Application to a sub-humid catchment. In: Duan, Q., Gupta, H., Sorooshian, S. Rousseau, S., Turcotte, A., (Eds.) Advances in Calibration of Watershed models, Water Science and Application 6, American Geophysical Union, Washington, D.C., pp. 197–211.

[70] J. Pomeroy, D. Gray, T. Brown, N. Hedstrom, W. Quinton, R. Granger, and S. Carey. The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence. Hydrological Processes, 21:2650–2667, 2007.

[71] C. Prudhomme, N. Reynard, and S. Crooks. Downscaling of global climate models for flood frequency analysis: where are we now? Hydrological Processes, 16:1137–1150, 2002.

[72] O. Rakovec, M. Hill, M. Clark, A. Weerts, A. Teuling, and R. Uijlenhoet. Distributed evaluation of local sensitivity analysis (delsa), with application to hydrologic models. Water Resources Research, 50:409–426, 2014.

[73] R. Rasmussen, B. Baker, J. Kochendorfer, T. Meyers, S. Landolt, A. Fischer, J. Black, J. Theriault, P. Kucera, D. Gochis, C. Smith, R. Nitu, M. Hall, K. Ikeda, and E. Gutmann. How well are we measuring snow: The noaa/faa/ncar winter precipitation test bed. American Meteorological Society, 93(6):811–829, 2012.

[74] P. Reed, N. Chaney, J. Herman, M. Ferringer, and E. Wood. Internationally coordinated multi-mission planning is now critical to sustain the space-based rainfall observations needed for managing floods globally. Environmental Research Letters, 10(2):024010, 2015.

[75] P. Reed, D. Hadka, J. Herman, J. Kasprzyk, and J. Kollat. Evolutionary multiobjective optimization in water resources: The past, present, and future. Advances in Water Resources, 51:D07103, 2012.

[76] S. Reed, V. Koren, M. Smith, Z. Zhang, F. Moreda, D. Seo, and DMIP Participants. Overall distributed model intercomparison project results. 298:27–60, 2004.

[77] J. Restrepo-Rosado and R. Bras. Automatic parameter estimation of a large conceptual rainfall-runoff model: a maximum likelihood approach. Applied Mathematics and Computation, 17:375–403, 1985.

[78] E. Rosero, Z. Yang, T. Wagener, L. Gulden, S. Yatheendradas, and G. Niu. Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the noah land surface model over transition zones during the warm season. Journal of Geophysical Research, 115:D07106, 2010.

[79] R. Rosolem, H. Gupta, W. Shuttleworth, X. Zeng, and L. Goncalves. A fully multiple-criteria implementation of the sobol' method for parameter sensitivity analysis. Journal of Geophysical Research, 117:438–456, 2013.

[80] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications, 145:438–456, 2002.

[81] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. Global Sensitivity Analysis: The Primer. John Wiley and Sons, New York, 2008.

[82] M. Sivapalan, K. Takeuchi, S.W. Franks, V.K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J.J. McDonnell, E.M. Mendiondo, P.E. O'Connell, T. Oki, J.W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe. Iahs decade on predictions in ungauged basins (pub), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48:857–880, 2003.

[83] A. Slater and M. Clark. Snow data assimilation via an ensemble kalman filter. Journal of Hydrometeorology, 7:478–493, 2005.

[84] M. Smith, V. Koren, Z Zhang, S. Reed, Z. Cui, B. Cosgrover, N. Mizukami, E. Anderson, and DMIP 2 Participants. Results of dmip 2 oklahoma experiments. Journal of Hydrology, 418-419:17–48, 2012.

[85] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiment, 1:407–417, 1993.

[86] M. Staudinger, K. Stahl, and J. Seibert. A drought index accounting for snow. Water Resources Research, 50:7861–7872, 2014.

[87] Y. Tang, P. Reed, K. van Werkhoven, and T. Wagener. Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. Water Resources Research, 43:W06415, 2007.

[88] Y. Tang, P. Reed, and T. Wagener. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration. Hydrology and Earth Systems Sciences, 10:289–307, 2006b.

[89] Y. Tang, P. Reed, T. Wagener, and K. van Werkhoven. Comparing sensitivity analysis methods to advance watershed model identification and evaluation. Hydrology and Earth System Sciences, 11:793–817, 2006a.

[90] K. Taylor, R. Stouffer, and G. Meehl. An overview of cmip5 and the experiment design. American Meteorological Society, April:485–498, 2012.

[91] M. Thiemann, M. Trosset, H. Gupta, and S. Sorooshian. Bayesian recursive parameter estimation for hydrologic models. Water Resources Research, 37(10):2521–2535, 2001.

[92] M. Thyer, B. Renard, D. Kavetski, and G. Kuczera. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using bayesian total error analysis. Water Resources Research, 45:W00B14, 2009.

[93] S. Uhlenbrook, J. Seibert, C. Leibundgut, and A. Rodhe. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. Hydrological Sciences Journal, 44:5:779–797, 1999.

[94] USGS. The water cycle: Snowmelt runoff. download from vendor site, 2014. version 21.

[95] K. van Werkhoven, T. Wagener, P. Reed, and Y. Tang. Rainfall characteristics define the value of streamflow observations for distributed watershed model identification. Geophysical Research letters, 35:L11403, 2008.

[96] K. van Werkhoven, T. Wagener, P. Reed, and Y. Tang. Sensitivity-guided reduction of parameteric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources, 32:1154–1169, 2009.

[97] J. Vrugt, W. Bouten, H. Gupta, and S. Sorooshian. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. Water Resources Research, 38(12):1312, 2002.

[98] J. Vrugt, C. Braak, M. Clark, J. Hyman, and B. Robinson. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. Water Resources Research, 44:W00B09, 2008b.

[99] J. Vrugt, C. Braak, H. Gupta, and D. Robertson. Equifinality of formal (dream) and informal (glue) bayesian approaches in hydrologic modeling? Stochastic Environmental Research Risk Assessment, 23:1011–1026, 2008a.

[100] J. Vrugt, C. Diks, H. Gupta, W. Bouten, and J. Verstraten. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41:W01017, 2005.

[101] J. Vrugt, H. Gupta, W. Bouten, and S. Sorooshian. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. Water Resources Research, 39(8):1201, 2003.

[102] T. Wagener. Can we model the hydrological impacts of environmental change. Hydrological Processes, 21:3233–3236, 2007.

[103] T. Wagener and H. Gupta. Model identification for hydrological forecasting under uncertainty. Stochastic Environmental Research and Risk Assessment, 19:378–387, 2007.

[104] T. Wagener, N. McIntyre, M. Lees, H. Wheater, and H. Gupta. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrological Processes, 17:455–476, 2003.

[105] T. Wagener, P. Reed, K. van Werkhoven, and Y. Tang. Advances in the identification and evaluation of complex environmental systems models. Journal of Hydroinformatics, 11.3-4:266–281, 2009a.

[106] T. Wagener, K. van Werkhoven, and Y. Tang. Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling. Water Resources Research, 45:W02501, 2009b.

[107] S. Warren. Optical properties of snow. Reviews of Geophysics, 2:67–89, 1982.

[108] K. Werner and K. Yeager. Challenges in forecasting the 2011 runoff season in the colorado basin. Journal of Hydrometeorology, 14:1364–1372, 2012.

[109] R. Wilby and T. Wigly. Downscaling general circulation output: a review of methods and limitations. Progress in Physical Geography, 21(4):530–548, 1997.

[110] W. Wiscombe and S. Warren. Model for the spectral albedo of snow. i:pure snow. Journal of the Atmospheric Sciences, 37:2712–2734, 1980.

[111] A. Wood, D. Lettenmaier, and R. Palmer. Assessing climate change implications for water resources planning. Journal of the Atmospheric Sciences, 37:2712–2734, 1997.

[112] A. Wood, P. Reed, and T. Simpson. Many objective visual analytics: rethinking the design of complex engineered systems. Structural Multidisciplinary Optimization, 48(1):201–219, 2013.

[113] E. Wood, D. Lettenmaier, and V. Zartarian. A land surface hydrology parameterization with subgrid variability for general circulation models. Journal of Geophysical Research, 97:2717–2728, 1997.