Spring 1-1-2017

A Multi-Algorithm Approach to Suspended Sediment Modeling in the Colorado Front Range

Jenna Reed Stewart
University of Colorado at Boulder, jennareedstewart@gmail.com

A MULTI-ALGORITHM APPROACH TO SUSPENDED SEDIMENT MODELING IN THE COLORADO FRONT RANGE

by

JENNA REED STEWART

B.S., University of Oregon, 2013

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Master of Science

Department of Civil, Environmental and Architectural Engineering

2017

This thesis entitled:
A multi-algorithm approach to suspended sediment modeling in the Colorado Front Range
written by Jenna Reed Stewart
has been approved for the Department of Civil, Environmental and Architectural Engineering

Prof. Ben Livneh

Prof. Balaji Rajagopalan

Prof. Joseph Kasprzyk

Dr. J. Toby Minear

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.

Stewart, Jenna Reed (M.S., Civil Engineering)

A multi-algorithm approach to suspended sediment modeling in the Colorado Front Range

Thesis directed by Assistant Professor Ben Livneh

Climatic and land cover changes present important uncertainties into the rates of streamflow and soil erosion in mountainous watersheds. Soil erosion adds constituents to streams, altering water chemistry and streambed morphology, which can impact drinking water treatment and water resources infrastructure. We applied five erosion and suspended sediment load algorithms within a common hydrologic framework to quantify uncertainty and evaluate predictability in two steep, forested catchments (> 1,000 km^2). The algorithms were chosen from among widely used sediment models, including empirical models: monovariate rating curve (MRC), and the Modified Universal Soil Loss Equation (MUSLE), a stochastic model: the Load Estimator (LOADEST), a conceptual model: the Hydrological Simulation Program—Fortran (HSPF), and a physically based model: the Distributed Hydrology Soil Vegetation Model (DHSVM). We coupled the algorithms with the Variable Infiltration Capacity Model (VIC), using hydrologic and meteorological inputs and fluxes generated from VIC. A multi-objective calibration was applied to the algorithms. Performance of optimized parameter sets from the calibration were validated over an ancillary period, as well as in an inter-basin transfer to a separate catchment to explore parameter robustness. This work highlights the tradeoffs in sediment prediction across a range of algorithm structures and catchments. Model performance showed consistent decreases when parameter sets were applied to time periods with greatly differing SSL magnitudes than the calibration period. Solutions from a joint algorithm calibration favored simulated streamflow partitioning into runoff and baseflow that optimized SSL timing,

impacting the flexibility and robustness of the streamflow to adapt to different time periods. Transferability performance was highest in algorithms with lower dependence on streamflow performance, the HSPF and the DHSVM. We expect that these more flexible and robust algorithms would likely fair better in predicting future climate scenarios due to their inclusion of physical conditions, precipitation rates and vegetation coverage, rather than solely relying on streamflow as in the case of the MRC. Future work will include applying this multi-algorithm routine to the Western United States, covering a greater number of catchments across varying climate, topography and land use regimes.

Dedication

To women in science and engineering.

Acknowledgements

I thank Dr. Ben Livneh for his consistent guidance, wisdom, patience and light-hearted wit, and for helping me tremendously throughout this process. His encouragement and excitement for the project inspired me, and I greatly appreciate the countless hours of help he has provided me. I also thank Dr. Joseph Kasprzyk for taking extensive time out of his days to help me develop a parallelized program, and for providing helpful advice and proverbs. I would also like to express gratitude to Dr. Balaji Rajagopalan, who has taught me statistical tools that I will carry in my pocket, and who is always sharing his words of wisdom and excitement for research. Thank you as well to Dr. J. Toby Minear for being an expert in sediment processes and modeling, and for providing invaluable knowledge and insights.

I cannot express enough gratitude for my partner Wesley, my sister Claire, and my parents and loved ones, Peggy, Craig and Barb, for always being there with love, support, and laughter throughout this experience. I owe many thanks to my wonderful friends as well who have guided me through this period with warmth.

Lastly, I acknowledge the EPA for funding this research, and 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.

Contents

Tables

Figures

CHAPTER 1

INTRODUCTION

According to the 2015 Colorado Climate Change Vulnerability Study (Gordon and Ojima, 2015), temperatures in Colorado are projected to rise by between +1.4°C and +2.8°C by 2050, increasing vulnerability to droughts and wildfire, and potentially leading to higher rates of soil erosion (Murdoch, Baron, & Miller, 2000; Smith, Sheridan, Lane, Nyman, & Haydon, 2011; Whitehead, Wilby, Battabee, Kernan, & Wade, 2009) throughout the state. Soil erosion adds constituents to streams, altering water chemistry and streambed morphology, which can adversely affect aquatic life (Rice, Greenwood, & Joyce, 2001) and water resource infrastructure (Podolak & Doyle, 2015). This occurs through increased contaminant concentrations (Delpla, Jung, Baures, Clement, & Thomas, 2009) and sediment buildup above reservoirs (Podolak & Doyle, 2015).

Climate extremes and human alterations to land cover have intensified erosion and sediment loading in streams from local to global scales (Walling, 2006). Simulating suspended sediment loading within a catchment is therefore an important task for water resources management. Many sediment modeling initiatives have computed erosion and transport over individual hillslopes or within small channel reaches or basins (< 50 km^2). However, there is a gap for expanding these calculations to larger scales. In an analysis of global sediment fluxes between land surface and ocean interactions, Walling (2006) documented the importance of sediment flux connectivity in major rivers between upstream

and downstream locations, as well as across large, continental scopes. Furthermore, studies show that global erosion rates have increased due to climatic and human induced changes to riverine systems, and that sediment fluxes have decreased due to retention in reservoirs (Meybeck, Laroche, Dürr, & Syvitski, 2003; Syvitski, Peckham, Hilberman, & Mulder, 2003; Syvitski, Vörösmarty, Kettner, & Green, 2005).

This work provides a quantitative inter-comparison of erosion and sediment transport methods under a unified hydrologic framework and a critical evaluation with real catchments (>1,000 km^2), representing an important step in large-scale sediment modeling that will be integral in future water resources management.

1.1 Summary of Chapters

1.1.1 Chapter 2: Background

Chapter 2 provides background information on topics relevant to the thesis. Topics include the processes of soil erosion and suspended sediment loading in catchments, focusing on drivers and conditions that contribute to changes in erosion and loading rates. We discuss historical erosion and suspended sediment modeling initiatives and their limitations, and propose a method for applying a multi-algorithm routine inserted into a common hydrologic framework.

1.1.2 Chapter 3: Methods

Chapter 3 describes our proposed methodology. We include information on site locations, data sources, and algorithms. Further, we discuss the methods for implementing the multi-algorithm routine under a common large-scale framework, proposing a critical area approach. Finally, we provide information on the multi-objective optimization routine

for calibrating the algorithms, and discuss the processes of validation and transfer of parameter sets to a neighboring catchment.

1.1.3 Chapter 4: Results

Chapter 4 presents the results from the multi-algorithm routine. We provide statistical analyses of observed streamflow, climate, and suspended sediment loading data, discussing differences between calibration and validation periods. We then discuss the calibration performance of all experiments under individual and joint algorithm calibrations. Further, we present the differences in validation and transferability performance across the experiments, and highlight the best performing algorithms.

1.1.4 Chapter 5: Discussion, Conclusions, and Future Work

Chapter 5 discusses the outcomes of the multi-algorithm routine, offering explanations for algorithm performances, larger implications of the study, and theoretical and practical limitations. Finally, we present conclusions and paths for expanding this work to greater scales.

CHAPTER 2

LITERATURE REVIEW

2.1 Suspended Sediment Background

Sediment comprises the suspended and deposited soil particles within a stream. Sedimentation occurs through the transport of eroded soil into streams, which is driven by both physical and biogeochemical processes. Soil erosion is the result of drivers such as raindrop impact, overland flow, gullying and mass movement (Knighton, 2014; Wicks & Bathurst, 1996). Though these processes are hard to quantify on large scales, overall rates of soil erosion and sedimentation can be inferred across climate and soil regimes by predictors such as vegetation cover, soil type, precipitation rates, topography, and catchment area (Knighton, 2014). The relationships between sediment loads, soil type, and vegetation were determined and validated through in-situ data collection and empirical analysis (Goode, Luce, & Buffington, 2012; Wolman & Miller, 1960). Climate effects on sediment loading were determined through reviewing climate extreme models and studies across land-use types and climate regimes (Moore & Demuth, 2001; Whitehead et al., 2009). Both topography and catchment area have also been found to impact sediment loads through multiple regression analysis on topographic features in 47 basins located throughout the Andes (Aalto, Dunne, & Guyot, 2006), and through experimental watersheds used to determine the effect of catchment size on erosion patterns (Lane, Hernandez, & Nichols, 1997).

Sediment can be divided into categories of bedload and suspended load. Bedload is the portion of sediment within a stream that is dominantly comprised of large particles such as sands and gravels that require more shear stress to move (Knighton, 2014). Suspended sediment, which is commonly termed as wash load and is generated from hillslopes, is mostly comprised of fine silts and clays, and is dependent on particle grain size, streamflow velocity, shear stress, and the resulting mobilization threshold (Wilcock & McArdell, 1993). This section describes the processes of erosion and sediment transport in catchments, highlighting key contributing drivers and conditions.

2.2 Soil Erosion Drivers

Spatial and temporal patterns of soil erosion are complex. Soil erosion occurs through complicated interactions of earth's processes, affected by myriad climatic, physical, chemical and biological factors. The conditions and drivers for soil erosion have been studied at great length (Morgan, 2009; Toy, Foster, & Renard, 2002), and though many processes contribute to erosional patterns, key conditions and drivers frequently appear. Soil erosion depends greatly on conditions of climate, topography, soil texture, vegetation and land use. Broadly speaking, the key drivers for erosion are raindrop impact, overland flow and gullying, and mass movement (Toy et al., 2002).

Though gullying and mass movement contribute to sediment loading, the focus of this study is on hillslope processes consisting of raindrop impact and overland flow. If the conditions are favorable for soil particle detachment, particles can be removed from the soil surface by the kinetic energy from raindrop impact or by the shear stress of overland flow. After detachment, particles are either transported downhill as suspended material by overland flow or deposited. Here, we discuss the conditions that favor soil detachment from

hillslopes and contribute to erosion.

2.2.1 Topography

Topography is important to consider when estimating erosion patterns in a basin (Aalto et al., 2006). In a global analysis of erosion rates as a function of slope, Larsen, Montgomery, and Greenberg (2014) found that watersheds dominated by steep terrain tend to produce higher sediment yields caused by extensive relief eroding the landscape. However, steep terrains are often supply limited, as the sheer slopes are not conducive to deposition and much of the material gets washed away. Landscapes that produce the most soil therefore tend to be mid to high angled slopes that have a balance between topographic relief and supply. Along with slope angle, the length and shape of a slope also affect erosion rates. With a longer slope, the volume and velocity of overland flow increases, thereby generating more potential erosional power. However, the shape of the slope can either compound or mitigate the effects of slope length. In a laboratory study of slope shape on sediment yield, Rieke-Zapp and Nearing (2005) documented higher yields from uniform and convex slopes than from concave slopes. Uniform slopes rarely occur in nature, whereas convex and concave slopes are more common.

2.2.2 Soil type and texture

Whether a soil particle will be detached from the land surface depends on the properties of the soil. Key factors that affect soil erodibility are texture, aggregation, and shear strength (Bryan, 2000; Middleton, 1930). Soil texture, which is the proportion of clay, silt and sand, is an important measure for erodibility, as the level of compaction impacts overland flow (Middleton, 1930). Though texture is stable through time, aggregation and shear strength are highly variable. Studies have shown that these temporal variations

directly affect erodibility, and are important for calculating the resistance of soil to erosive forces (Bryan, 2000; Nearing, Foster, Lane, & Finkner, 1989). Aggregation, which is the combination of two or more soil particles that have a stronger bond than nearby particles (Kemper & Chepil, 1965), alters the permeability of the soil, which affects pore space and infiltration rates. Furthermore, break-down of aggregates by disruptive forces is highly variable depending on the chemical and structural bonds of the aggregated particles. Shear strength, which is the ability of soil particles to resist erosive forces, is highly dependent on climatic conditions. Shear strength curves are developed from using frictional forces, and soil cohesion as an intercept term.

2.2.3 Climate

Erosion is affected by climate through changes in precipitation and temperature regimes. Precipitation frequency and intensity affect raindrop impact, whereas temperature affects freeze-thaw cycles and snow-melt rates. For precipitation, the kinetic energy of a raindrop detaches soil particles from the land surface as a function of raindrop mass and velocity. Though the effects of precipitation on soil detachment largely depend on soil texture, slope and vegetation structure, for the majority of cases, increased precipitation causes increased erosion (Morgan, 2009). Soils that are exposed to more frequent or more intense precipitation events can become weaker and more susceptible to erosion through time, although wetter environments are typically associated with increased vegetation density that limit soil erosion over the long-term (Nearing, Pruski, & O'neal, 2004).

Temperature impacts erosion by affecting snow melt rates and freeze-thaw cycles. Frozen or snow-covered soils are largely impervious to erosion. However, thawing soils and snow-melt driven overland flow can greatly increase erosion rates and sediment concentrations (Glysson, 1987). Therefore, a large amount of erosion occurs in the spring in

mountainous regions as soils thaw and snow-melt driven overland flow increases. The effects of temperature on erosion rates are complex, though generally higher temperatures increase evaporation rates, and decrease plant biomass and soil residues, which could heighten erosion rates (Nearing et al., 2004).

Climate change has the potential to intensify sediment loading (Nearing et al., 2004; Whitehead et al., 2009). This can occur through extreme precipitation and drought events, or through changes in temperature that affect freeze-thaw cycles. Furthermore, consecutive extreme events create compounding effects. For example, extreme precipitation following a prolonged drought can cause a flush of soil into streams. Whitehead et al. (2009) present a synthesis on the role of extreme climate events on sediment loading. Focusing on the effects of heightened precipitation rates and temperature on soil processes, they found that large fluxes of suspended sediment loads occurred during storm events following droughts. Climate indirectly affects soil erosion as well though disturbances to vegetation and soil texture by heightened wildfires and flooding.

2.2.4 Vegetation and Wildfire

Generally, a denser vegetation canopy reduces erosion on hillslopes (Morgan, 2009). This occurs through the obstruction of precipitation and overland flow, and greater root stability for soils. With the presence of a canopy, precipitation is either intercepted and evaporated, intercepted and then falls through as leaf drip or stemflow, or is not intercepted and proceeds to the ground as direct throughfall. In the second scenario, erosion rates depend on the canopy height and the leaf drip diameter. In the third scenario, erosion rates directly depend on the precipitation intensity and length of the precipitation event (Morgan, 2009). Bare ground, therefore, is much more susceptible to rainfall erosion than a heavily forested area with a dense canopy.

Wildfires can cause extreme increases in erosion and sediment loading in large part from reduced vegetation coverage. In a review of wildfire impacts on water supply, Smith et al. (2011) analyzed burned and unburned sites within forest catchments of varying sizes, and documented an increase to suspended sediment loading of 1-1459 times in burned areas compared to unburned areas. The magnitude of erosion and sediment load increase is often correlated with burn severity. In addition to reducing vegetation density, wildfires affect the soil properties as well, causing a buildup of hydrophobic residue and increasing the soil's water repellency. Heightened water repellency decreases infiltration rates, which generates more overland flow and thereby produces more shear stress acting on the soil. Martin and Moody (2001) compared soil infiltration rates between two fire-impacted catchments in Colorado and New Mexico. They found that fires altered the composition and abundance of riparian vegetation and simultaneously increased the repellency of soil to water, thereby causing a flush of eroded soil and runoff into the system.

2.3 Sediment Loading Drivers

There are many factors that contribute to suspended sediment transport throughout a catchment. Primarily, suspended sediment transport is reliant upon the supply of eroded soil, catchment scale and contributing area, transport capacity of streamflow, grain size distribution, and connectivity of hillslopes to channels.

2.3.1 Catchment Scale

Total sediment loading is often correlated with catchment area, as larger drainage areas become sinks for more eroded material. However, analyses show that sediment yield normalized by catchment area often decreases after a threshold, as erosion inputs are

masked by in-stream processes of transport and deposition (De Vente & Poesen, 2005; Lane et al., 1997). Below this threshold, which is site dependent, normalized sediment yield is positively correlated with catchment area.

2.3.2 Transport Capacity, Shear Stress and Grain Size Distribution

Suspended sediment transport is largely dependent on the transport capacity of streamflow. As noted in a review of sediment transport models by Yang (2006), streamflows with higher velocities have a greater capacity to entrain particles. As streamflow velocity decreases, particles are deposited within a channel because the particle settling velocity exceeds that of the streamflow. Settling velocities can range from 0.198 cm/s for a 0.05 mm particle diameter to 46.1 cm/s for a 5mm particle diameter (Gibbs, Matthews, & Link, 1971). Sediment transport is also connected with the grain size distribution of in-stream particles. Shear stress increases at higher streamflow velocities, overcoming the thresholds required for larger particles to move (Wilcock & McArdell, 1993). Therefore, at high streamflow velocities sediment loading increases due to bedload motion and a greater entrainment capacity of the streamflow. Despite the contribution of bedload to total sediment loads, in a review of grain size partitioning of suspended particles, Turowski, Rickenmann, and Dadson (2010) found that in large catchments the rough proportion of bedload in suspension ranged from 1% for gravel and rock dominated bed material, to 30-50% in sand dominated bed material, with a general trend of 10% bed material being in suspension. However, the proportion of bedload in suspension also varies as a function of streamflow rate.

2.3.3 Connectivity

Though erosion occurs throughout a catchment, only a portion of the eroded soil

ends up in the stream channel due to hillslope and channel connectivity. Here we define connectivity as the movement of sediment sources to sinks, mitigated by the proximity of hillslopes to channels, and the contribution of tributaries to larger channels (Bracken, Turnbull, Wainwright, & Bogaart, 2015). Soil eroded from a hillslope adjacent to a stream channel will affect sediment concentrations more than soil eroded from a disconnected hillslope. In steep, mountainous, headwaters regions, hillslope erosion is highly correlated with sediment concentration, whereas in terrains with shallow relief and larger streams, the sediment concentration is more dependent on bank erosion and sediment transported from upstream (Benda, Hassan, Church, & May, 2005). Though connectivity is largely controlled by hydrologic processes, the buildup of in-stream sediments and woody debris can greatly affect sediment transport as well. For example, even in a highly connected system, sediment transport is hindered by landscape buffers such as vegetation or fallen trees, in-stream barriers such as large woody debris, or blankets such as floodplains and armored channel beds (Fryirs, Brierley, Preston, & Kasai, 2007).

2.4 Existing Soil Erosion and Sediment Transport Models

Existing erosion and sediment transport models vary in terms of spatial and temporal scales of application, parameter requirements, and model dimension such as one-dimensional averages of a cross section, two-dimensional averages of cross-section and depth, or three-dimensional resolving of both the vertical and horizontal components. Due to the reliance on hydrologic inputs, erosion and sediment transport models are frequently coupled with hydrologic or land surface models (Arnold, Srinivasan, Muttiah, & Williams, 1998; Doten, Bowling, Lanini, Maurer, & Lettenmaier, 2006; Nearing et al., 1989), requiring more extensive data inputs and computations. Numerous hydrologic models have

been developed to address the challenges of measurement limitations, and heterogeneities in stream characteristics (Pechlivanidis, Jackson, McIntyre, & Wheater, 2011; Singh & Woolhiser, 2002). Concurrently, efforts have been made to simulate soil erosion and sediment transport using a variety of empirical, conceptual, stochastic and physically-based methods.

Empirical methods, such as the Universal Soil Loss Equation (USLE) (Wischmeier, Smith, & others, 1960), are computationally efficient and are derived from relationships in available data. However, their major shortcoming is a homogenous representation of catchment systems, as well as estimating outputs based on a single event or over a large time-step (Pechlivanidis et al., 2011). Conversely, conceptual models represent a catchment or hillslope through storage systems, but are limited by parametric uncertainty due to their lack of in-situ observations (Aksoy & Kavvas, 2005). Stochastic methods can relate multiple predictors to a response using regression techniques, though these methods are also reliant on historical data (Helsel & Hirsch, 2002). Physically-based models tend to be the most representative of a system, since they attempt to reconcile the physical properties and processes through solving mass and energy conservation equations. However, the complexity and data requirements for these models often exceed available information (Ranzi, Le, & Rulli, 2012). To date, many modeling initiatives are hybrids of the three model categories, combined to increase accuracy and efficiency (Kabir, Dutta, & Hironaka, 2011; Pechlivanidis et al., 2011; Zuliziana, Tanuma, Yoshimura, & Saavedra, 2015).

2.4.1 Existing Erosion Models

A common model used to estimate soil loss in a catchment is the USLE. Developed for agricultural applications, the USLE is an empirical method that computes annual soil loss as a linear product of empirical factors of soil erodibility, topography, annual rainfall,

land cover, and land conservation practice. Because the USLE was limited to agricultural field scales, the equation was further developed to incorporate more rigorous parameter estimations in the Revised USLE (RUSLE) (Renard et al., 1997) and to exchange rainfall with catchment runoff in the Modified USLE (MUSLE) (Williams & Berndt, 1977). Though the USLE and its derivatives include local watershed characteristics, they suffer key limitations, most notably: the homogenization of attributes within the equation, the lack of spatial and temporal representation, and an assumption of linearity among predictor and response variables (Merritt, Letcher, & Jakeman, 2003). Because of the inherent spatial and temporal variability of soil erosion, models of increasing complexity have been developed to capture heterogeneities within a catchment. One such empirical model called the Sediment Delivery Distributed Model (SEDD) (Ferro & Porto, 2000), uses computations from USLE with an added spatial distribution component to estimate sediment delivery within specified morphological units. Empirical models succeed in areas where observed data is readily available, however, this is often cumbersome as continuously monitored erosion data are rare (Merritt et al., 2003). With increased complexity, the Watershed Erosion Prediction Project (WEPP) (Nearing et al., 1989) was developed to provide physical estimates of erosion from small hillslopes through rill and interill processes. Rills are defined as small channels or grooves within a hillslope, whereas interill areas are the segments of land between rills. Furthermore, a coupling scheme was developed to estimate WEPP hillslope erosion over larger areas by coupling the erosion computations with a large-scale hydrologic model (Mao, Cherkauer, & Flanagan, 2010).

2.4.2 Existing Sediment Transport Models

Annual soil loss estimates are useful for long-term land management practices. However, shorter term estimates can be more valuable for many applications and these

require routing the soil through a catchment to estimate in-stream sediment yield, involving additional computations. The relationship between soil loss and sediment yield is difficult to quantify, as the transport of eroded soil particles greatly depends on streamflow and land surface conditions. The simplest approach to numerical modeling of sediment loads is an empirically based sediment rating curve (Gray & Simões, 2008). This mono-variate rating curve (MRC) fits streamflow data with suspended sediment within a given catchment using scalars and exponents. Though MRC estimates are predictive for long-term loads and general trends, they are reliant upon consistent timing between sediment fluxes and peaks of streamflow. Therefore, the MRC is limited by its inability to capture heterogeneous hillslope erosion processes.

Another weakness of the MRC is that it generates a single sediment load for multiple streamflow values. Therefore, adding more variables into the equation can produce higher explanatory power, as different variables contribute to runoff and sediment loading in different ways (Gray & Simões, 2008; Helsel & Hirsch, 2002). For example, the rising limb of a hydrograph is controlled by rainfall, whereas the falling limb is controlled by antecedent moisture and runoff (Park et al., 1982). Incorporating both streamflow and rainfall intensity into multi-variate regression can therefore further separate the sediment values in relation to rainfall-runoff events (Syvitski, Morehead, Bahr, & Mulder, 2000). As seen in a United States Geological Survey (USGS) study, Gartner, Cannon, Helsel and Bandurraga (2009) applied multivariate regression across catchments to calculate sediment yield as a function of rainfall, catchment area, topographical features, and fire burn area. Furthermore, the USGS developed the Load Estimator (LOADEST) (Runkel, Crawford, & Cohn, 2004), a multi-variate regression algorithm that estimates sediment and nutrient loads from streamflow magnitude and data collection times.

With increasing complexity, numerous models have been applied for predicting both

soil loss and sediment loads in forested, mountainous regions. Such models include the Soil Water Assessment Tool (SWAT) (Arnold et al., 1998), which covers large ungauged drainage areas, provides continuous long-term sediment yields on a daily time-step, and closes the water-balance equation, albeit the sediment yield is calculated using the MUSLE. Similar to SWAT, the Hydrological Simulation Program—Fortran (HSPF) (Bicknell, Imhoff, Kittle Jr, Donigian Jr, & Johanson, 1996) estimates erosion and sediment transport within hydrologic units across large spatial scales. However, these erosion estimates are based on conceptual representations of rainfall impact and overland flow. Another more complex sediment delivery model is the System Hydrologique European-Sediment (SHESED) model (Wicks & Bathurst, 1996), which is physically-based, spatially distributed, used at the catchment scale, and models hillslope erosion on a grid scale through raindrop and leaf drip impacts and overland flow. SHESED also provides the framework for another spatially distributed sediment model developed for the Distributed Hydrology-Soil-Vegetation Model (DHSVM) (Doten et al., 2006; Wigmosta, Vail, & Lettenmaier, 1994). DHSVM is a spatially distributed, continuous temporal model designed for large catchment areas, which also incorporates mass wasting and erosion from forest roads. Numerous other physical erosion and sediment transport models exist, each developed to reconcile physical processes differently using hillslope and in-channel routing mechanisms such as in WEPP, and complex channel geometry and the kinematic wave equation as seen in the Kinetic Runoff and Erosion Model (KINEROS2) (Smith, Goodrich, Woolhiser, & Unkrich, 1995).

Erosion and sediment transport models have been developed to estimate soil loss and sediment yields over a range of hillslope and catchment sizes. As each method simulates processes differently, qualitative and quantitative comparisons are helpful to depict the most representative methods.

2.4.3 Model Comparisons

In a review of soil erosion and sediment transport models, Aksoy and Kavvas (2005) categorized the most common physically-based sediment transport schemes into multiple groups: lumped versus distributed, event based versus continuous, hillslope versus catchment scale, 1-dimensional versus 2-dimensional, and single-size versus multi-size grain scale. Merritt et al. (2003) also performed a review of erosion and transport models, but categorized the models based on input demands, outputs and catchment scale. The categories defined in these reviews provide a framework for which sediment transport models can be evaluated.

Though Merritt et al. (2003) and Aksoy and Kavvas (2005) reviewed sediment models according to different criteria, very few studies have quantitatively compared and analyzed the output of multiple erosion and sediment transport models over the same time period and catchment scale. One of the few, Jetten, De Roo, and Favis-Mortlock (1999) ran a soil erosion model comparison using seven field scale (0.01 ha to 10 ha) and seven catchment scale (40 ha) models, using empirical, conceptual and physical models in their study. Jetten et al. (1999) concluded that for the field scale approach, WEPP produced the best results, with a correlation coefficient of 0.75 for monthly soil loss and 0.84 for daily soil loss when compared with observational data. For the catchment scale, KINEROS2 produced the highest correlation coefficient of 1.0 when compared with observational data. Overall, the study indicated better model performance over longer periods across many of the models, and increased skill from relative results (e.g., the ratio of simulated to observed) than absolute results (e.g., the absolute difference between simulated and observed).

2.4.4 Limitations to Existing Modeling Initiatives

Though the interactions of bedload and suspended load have been accurately estimated on a point scale, across a channel cross section, or in a small reach (Walling, 1983), expanding those calculations to a watershed scale requires extensive fine resolution data inputs, and high computational capacity (Yang, 2006). Therefore, while the interactions between bedload and suspended load are important, large-scale erosion and sediment models focus less on the bedload component. Instead, the models emphasize hillslope erosion, particle size distribution of the eroded soil, and the mechanism for how the eroded soil is either suspended or deposited in a channel by using the continuity equation and transport capacity of overland flow and streamflow (Arnold, Williams, & Maidment, 1995; Doten et al., 2006; Nearing et al., 1989). The routing of the eroded soil from hillslopes through a large basin remains an open issue due to a high degree of heterogeneity in meteorology, soil and hillslope micro-topography, and the dependency of sediment loads on grain size and channel geometry (Garcia & Parker, 1991; Wilcock & McArdell, 1993; Yang, 2006). Furthermore, the simplification of physical processes and input parameters often results in underestimation of sediment yields in large basins (Yang, 2006).

2.5 Research Objectives

The montane forested catchments of the Colorado Front Range are susceptible to increased drought and wildfire occurrence by 2050 (Gordon and Ojima, 2015), potentially increasing rates of erosion and sediment loading throughout the area (Martin & Moody, 2001; Smith et al., 2011). Greater sediment loads impact water resources infrastructure and reservoir design life, and pose threats to water quality and treatment. Modeling

sediment loading in the Front Range is therefore critical to inform future mitigation efforts; for example, knowledge of future sediment supply to water treatment plants could aid in decisions such as whether to increase filtration measures, add more disinfection chemicals, vary the water supply, or prescribe forest thinning (Khan et al., 2015). However, large-scale sediment modeling initiatives often contain high levels of uncertainty (Kabir et al., 2011; Zuliziana et al., 2015). Therefore, our study sought to advance current erosion and sediment transport science by:

1. Coupling hillslope erosion computations within a single large-scale hydrologic modeling framework, to explore strengths and weaknesses of each algorithm within a consistent structure
2. Calibrating each algorithm using multi-objective optimization, and applying the best performing parameter sets to a neighboring catchment to assess transferability
3. Providing a quantitative model inter-comparison at much larger scales than previous investigations and characterizing the structural uncertainty among model estimates

To complete our objectives, we modeled soil loss and sediment loads in three large sub-basins of the Colorado Front Range using two empirical methods, one stochastic method, one conceptual model and one physically-based model inserted into a common hydrologic framework. The framework was the Variable Infiltration Capacity (VIC) model, which is a land surface model (LSM) that resolves both the water balance and energy balance (Liang, Lettenmaier, Wood, & Burges, 1994). The empirical methods were the MRC and an implementation of the MUSLE adapted by SWAT; the stochastic method was LOADEST; the conceptual model was HSPF; and the physically-based model was DHSVM. We hypothesized that:

1. The different methods of soil erosion and sediment transport would provide a range of uncertainty in sediment estimates at the catchment outlet
2. The most predictive sediment calculations would result from the most complex, physically-based model
3. Algorithms would perform better in the calibration than the validation, and both would perform better than the transfer to a neighboring catchment

CHAPTER 3

METHODS

3.1 Overview

We tested a diverse set of model algorithms for estimating suspended sediment loads (SSL) within a course 1/16˚ resolution framework consisting of (i) two empirical models, (ii) one stochastic model, (iii) one conceptual model, and (iv) one physically-based model. The sediment modules from these models were embedded within a large-scale hydrologic model, the Variable Infiltration Capacity (VIC) model (Liang et al., 1994) to provide consistency and comparability between the models under a common hydrologic framework. We applied the algorithms to three steep, mountainous catchments in the Colorado Front Range. Model outputs from the first two catchments were calibrated to daily streamflow and instantaneous SSL using multi-objective optimization. The best performing parameter sets were applied to a different period within the same catchment to test validation and applied to the third catchment to test parameter and model transferability. The transferability test reflected the flexibility and robustness of parameter sets.

3.2 Catchment Descriptions and Data Sources

We ran this study over three catchments on the Colorado Front Range: a sub-basin of the Cache La Poudre catchment delineated to USGS gage 06751490 North Fork Cache

La Poudre River at Livermore, CO[1] (CLP-L), a larger sub-basin of the Cache La Poudre catchment delineated to USGS gage 06752000 Cache La Poudre River at Mouth of Canyon near Fort Collins, CO[2] (CLP-F), and a sub-basin of the Clear Creek catchment delineated to USGS gage 06719505 Clear Creek at Golden, CO[3] (CC) (Figure 1). We selected these three catchments based on the availability of streamflow and suspended sediment data, and for minimal reservoir storage upstream of the gages. Reservoir storage was calculated as a percentage of mean annual streamflow using the GAGES-II database (Falcone, 2011). The USGS National Water Information System (NWIS) provides water quality data, including field-based measurements of depth-integrated suspended sediment concentrations (SSC) and SSL. NWIS suspended sediment data is scarce at each catchment, sampled at irregular intervals, except for 1981 in CC when daily SSC and SSL were inferred from a turbidimeter (Table 1). Turbidimeter estimates for SSC and SSL are calculated through regression from turbidity measurements, and therefore the data comes with inherent uncertainty. We obtained daily streamflow estimates from USGS for each of the suspended sediment periods for CLP-L, CLP-F, and CC.

The three catchments were selected for their similar climate, land use and soil regimes to test transferability of the suspended sediment model algorithms between catchments (CC and CLP-F) and from a smaller catchment located within a larger catchment (CLP-L and CLP-F). The catchments are dominated by snow-melt hydrology, with high elevation, continental dry climate prone to high snowfall in the winter, and peak snow-melt during the spring. The steep mountainous terrain in the area is dominantly alpine forests, shrubs, and grasslands, whereas the geology comprises intrusive igneous and sedimentary rocks that are largely granitic.

[1] https://waterdata.usgs.gov/co/nwis/uv/?site_no=06751490&agency_cd=USGS
[2] https://waterdata.usgs.gov/co/nwis/dv?referred_module=sw&site_no=06752000
[3] https://nwis.waterdata.usgs.gov/nwis/inventory/?site_no=06719505&agency_cd=USGS

Table 1. USGS sediment gage locations and the number of samples with available suspended sediment data

USGS Gage	Latitude	Longitude	Catchment	Sample Count	Sample Range	Sampling Method
06751490*	40.788	-105.525	CLP-L	134	1987 - 1999	Physical Measurement
06752000*	40.664	-105.224	CLP-F	78	1963 - 2002	Physical Measurement
06719505*	39.753	-105.235	CC	39	1981 - 1995	Physical Measurement
06719505**	39.451	-105.141	CC	177	1981 - 1981	Turbidimeter

* Irregular Intervals
** Daily

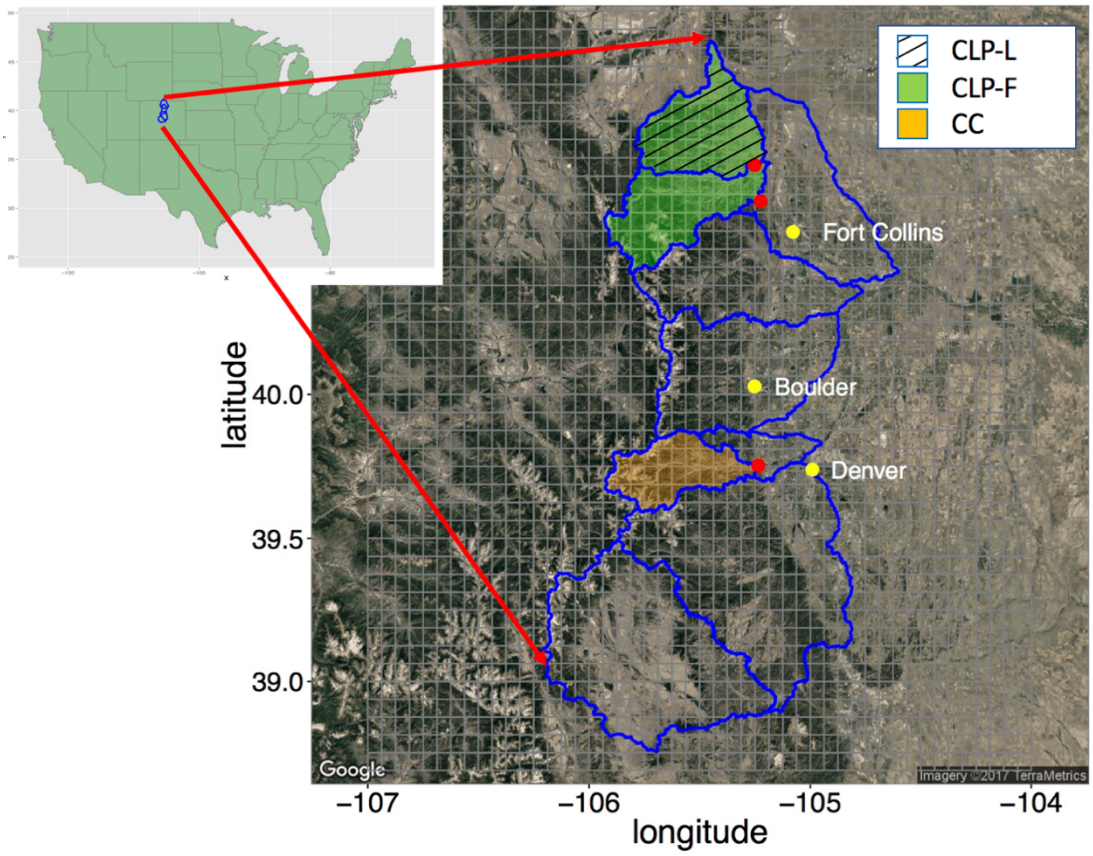

Figure 1. North Fork of Cache La Poudre at Livermore (CLP-L, USGS 06751490, 1,393 km^2), Cache La Poudre at Mouth of Canyon near Fort Collins (CLP-F, USGS 06752000, 2,017 km^2) and Clear Creek at Golden, CO (CC, USGS 06719505, 1,024 km^2) catchments located within the Colorado Front Range overlaid by 1/16˚ resolution VIC grid cells. Calculated upstream reservoir storage was 35%, 25% and 8% of mean annual streamflow, respectively. Shaded regions indicate catchment areas, red dots indicate USGS streamflow and SSL gauges, and yellow dots indicate nearby cities.

3.3 Description of Model Algorithms

3.3.1 The Variable Infiltration Capacity model

For the common hydrologic framework, we selected VIC, which is a physically-based land surface model that simulates hydrologic processes including baseflow, runoff, evapotranspiration, soil moisture, and snow water equivalent. VIC is spatially distributed, solves both the water and energy balance equations, and is typically applied at resolutions of 1/20˚ (~4 km) to 1˚ (~100km). By integrating sub-grid cell variation in vegetation, soil and snow elevation bands, VIC resolves a high
degree of sub-grid variability and physical complexity.

VIC was driven with a dataset for the conterminous United States developed at 1/16˚ spatial resolution on a continuous daily interval (Livneh et al., 2015). This dataset includes meteorological forcing files for precipitation, maximum temperature, minimum temperature and wind speed. Livneh et al. (2015) gridded the meteorology using observations obtained from the National Climatic Data Center Cooperative Observer stations coupled with the synergraphic mapping system algorithm, and scaled the precipitation to match the monthly output from the Parameter-Elevation Regressions on Independent Slopes Model.

Lateral transfer between grid cells is not incorporated into VIC, therefore we used the RVIC routing model developed for VIC simulations. RVIC solves the linearized Saint-Venant equations and uses a unit-hydrograph approach to route VIC model outputs to a user-defined location (Lohmann, Nolte-Holube, & Raschke, 1996). For each sediment module, we assumed constituent loads were advected with the streamflow, though future studies will incorporate further estimates of transport capacity and deposition.

3.3.2 Monovariate Rating Curve

The MRC, also known as a sediment rating curve, is an empirical method developed to estimate sediment loading from streamflow. The most common estimate for MRC is in the form of a power relationship, where SSL is calculated according to

$$\mathrm{SSL} = aQ^b \tag{1}$$

where Q is streamflow, a is a coefficient for the intercept, and b is an exponent for slope (Gray & Simões, 2008). This method is commonly used by the USGS as a course representation of sediment loading throughout a catchment. To develop the coefficients, we minimized nonlinear least squares in the base R package.

Despite being simple to execute, the MRC method is limited for several reasons. The MRC method relies on historical observations of SSL and streamflow to develop and fit to observations, limiting its application to ungauged catchments. Furthermore, the MRC predicts SSL from a single streamflow value, assuming the SSL and streamflow timing align. In systems where hysteresis occurs, e.g. where sediment peaks prior to streamflow, the MRC approach is inadequate for capturing SSL timing (Gray & Simões, 2008). Lastly, studies show that the MRC method tends to underestimate SSL, as the method ignores important contributions from additional drivers, such as precipitation, topography, and soil type among other factors (Gray & Simões, 2008).

To integrate the MRC into VIC, we ran VIC and RVIC, and then applied the MRC to the routed streamflow values at the catchment outlet. For this exercise, streamflow was considered the sum of VIC runoff and baseflow.

3.3.3 Modified Universal Soil Loss Equation

The MUSLE is an empirical method used to estimate soil loss from catchment

characteristics and land management factors. Adapted from the USLE, the MUSLE incorporates six predictors in a linear relationship as

$$A = RKLSCP \tag{2}$$

where A is average soil loss, R is a runoff factor, K is a soil erodibility index, LS is a topographical index for the length and steepness of a slope, C is a crop management factor representing vegetation, and P is a land conservation factor (Arnold et al., 1998; Wischmeier et al., 1960). To apply the MUSLE within VIC, we adopted a similar methodology to that used in SWAT. To calculate R for a given streamflow event, the SWAT method uses the total volume of runoff in an event (Q) and the peak streamflow rate of the event (q_p) as

$$R = 11.8\left(Q * q_p\right)^{0.56} \tag{3}$$

The LS factor is determined in SWAT through using slope steepness (S, m/m) and slope length (λ, m) as

$$LS = \left(\frac{\lambda}{22.1}\right)^{\xi} (65.41(S)^2 + 4.565(S) + 0.65 \tag{4}$$

where ξ is a function of slope:

$$\xi = 0.6\left(1 - \exp\left(-35.835(S)\right)\right) \tag{5}$$

Soil loss is calculated through the MUSLE for each spatially explicit Hydrologic Response Unit using the inputs listed above and runoff generated by the SWAT model. Total SSL is obtained for the outlet of the catchment by routing the soil using a scheme developed by Arnold et al. (1995) that simulates sediment deposition based on particle fall velocity.

Though MUSLE is mathematically efficient, the method is limited due to the linear relationship among predictors. Furthermore, the predictors are indices that are not directly observable and hence are often hard to estimate and constrain.

To apply the MUSLE in VIC, we incorporated the equations into the daily

computations within each VIC grid cell. Within each grid cell, we estimated Q as the sum of the daily streamflow, and q_p as the daily peak streamflow.

3.3.4 Load Estimator

Developed by the USGS, the LOADEST is a stochastic method that uses single and multi-variate regression to predict constituent loads such as suspended sediment and nutrients in streams from historical data (Runkel et al., 2004). The LOADEST incorporates methods from Helsel and Hirsch (2002) to predict SSL using a variety of linear regression models incorporating streamflow, collection time, and periodicity. Because SSL data are often not normally distributed, the SSL and streamflow data are logarithmically transformed to adhere to the linear regression requirements of normally distributed data. In the most generalized form, the linear regression is computed as

$$\ln S\hat{S}L = a_0 + \sum_{j=1}^{n} a_j X_j \tag{6}$$

where $S\hat{S}L$ is a vector of instantaneous loads, a_0 and a_j are model coefficients, X_j is an explanatory variable, and n is the number of explanatory variables (Runkel et al., 2004). To convert from log space to the instantaneous load, the base e can be raised to the power of the regression equation, which becomes a more complex, multi-variate form of the rating curve. Within the LOADEST, there are eleven pre-defined models, ranging from the simplest mono-variate estimation using streamflow (Q):

$$\ln S\hat{S}L = a_0 + a_1 \ln Q \tag{7}$$

to a more complex, multi-variate regression using an adjusted collection time (t)

$$\ln S\hat{S}L = a_0 + a_1 lnQ + a_2 lnQ^2 + a_3 \sin(2\pi t) + a_4 \cos(2\pi t) + a_5 t + a_6 t^2 \tag{8}$$

In Equations 7 and 8, $\ln Q$ is calculated as

$$\ln Q = \ln(streamflow) - 'center' of \ln(streamflow) \quad (9)$$

In Equation 8, t is calculated as

$$t = decimal\ time - 'center' of\ decimal\ time \quad (10)$$

The *'center' of ln(streamflow)* and *'center' of decimal time* are defined through LOADEST using the distributions of the data.

The models are validated by three error estimate methods over a calibration period: for normally distributed residuals, the methods are Maximum Likelihood Estimation (MLE) and Adjusted Maximum Likelihood Estimation (AMLE), and for not normally distributed residuals the method is Least Absolute Deviation (LAD). Selection of the best model occurs through analyzing the AMLE outputs based on Akaike Information Criterion (AIC) (Akaike, 1974). For complex systems, the user can preprocess the historical data by splitting it into seasons, rising or falling hydrograph limbs, or through using a threshold approach. In such a way, different models can be developed for various separations of the hydrograph, and combined to create a more robust fit.

Though the LOADEST is frequently used by USGS and tailored to estimate instantaneous SSL in streams, the method is limited by its reliance on historical data. Simultaneously, the method is unable to predict SSL from outside of the historical data calibration period, making it difficult to forecast scenarios of changing climates or varying seasons.

Like the MRC, to integrate the LOADEST into VIC, we ran VIC and RVIC, and then applied the LOADEST to the routed streamflow values at the catchment outlet. For this exercise, streamflow was considered the sum of VIC runoff and baseflow.

3.3.5 Hydrological Simulation Program—Fortran

The HSPF is a conceptual model developed by the Environmental Protection Agency

to simulate hydrologic processes and the transport of contaminants in catchments for water quality purposes (Bicknell et al., 1996; Johanson & Davis, 1980). The HSPF is based on the Stanford Watershed Model, and incorporates conceptualized hydrologic storage by including snow, surface, upper soil, lower soil and ground-water zones. The model is spatially distributed, using homogenous areas characterized by pervious and impervious land, as well as a free-flowing reach or mixed reservoir module. The HSPF contains both a hillslope erosion module and a bedload transport module. However, for the purposes of this study we applied only the hillslope component.

The hillslope erosion component is divided into two main categories: detachment by rainfall, and wash off and scour by overland flow. The module is applied over a pervious land segment, as impervious segments are assumed to not generate sediment. Rainfall detachment (*DET,* tons/ac/interval) is estimated by using the kinetic energy of raindrops as

$$DET = \left(dt\ (1 - CR)(P)(K)\left(\frac{I}{dt}\right)\right)^{JR} \tag{11}$$

where *dt* is the number of hours in the time interval, *CR* is the fraction of snow and vegetation cover, *P* is the practice management factor adopted from the USLE, *K* is the detachment coefficient adopted from the USLE, *I* is the rainfall intensity (in/interval), and *JR* is the detachment exponent. HSPF further simulates the effect of rainfall by decreasing *DET* following a day without rainfall using a parameter *AFFIX* as

$$DET = DET(1.0 - AFFIX) \tag{12}$$

Once the soil is detached by rainfall, it can either be re-deposited or transported by overland flow. To simulate this process, HSPF uses a conceptualized method for estimating transport capacity (*TC*, tons/ac/interval) as

$$TC = \ dt(KS)\left(\frac{SU + SO}{dt}\right)^{JS} \tag{13}$$

where *KS* is the transport coefficient, *SU* is the surface water storage (in), *SO* is the surface water outflow (in/interval), and *JS* is the transport exponent. The HSPF has two methods for simulating transport capacity, with this being the primary method despite the dimensional non-homogeneity. The transport capacity is then related to the detached sediment to estimate sediment washoff (*WSSD*, tons/ac/interval)

$$WSSD = \frac{DETS(SU)}{SU + SO} \qquad if\ TC > DET \tag{14}$$

$$WSSD = \frac{TC(SU)}{SU + SO} \qquad if\ TC < DET$$

where *DETS* is detached sediment storage (tons/ac).

Scour from the soil matrix (*SCR,* tons/ac/interval) incorporates similar processes, using the overland flow as a metric for detachment:

$$SCR = \frac{SU}{SU + SO} dt(KG) \left(\frac{SU + SO}{dt}\right)^{JG} \tag{15}$$

where *KG* is the scour coefficient and *JG* is the scour exponent.

The hillslope erosion component of the HSPF is simple and efficient, using conceptualizations of physical mechanisms to predict SSL. However, the module is dependent on numerous coefficients and exponents that are not directly observable and estimating them can result in overfitting. Furthermore, HSPF has many parameters requiring calibration from historical observations, which makes the model difficult to apply in ungauged catchments that do not have available data.

To integrate HSPF into VIC, we computed erosion estimates within each VIC grid cell for the daily timestep, similar to the MUSLE. To obtain *SO* estimates, we used the sub-daily VIC runoff calculations for each grid cell, as this represents overland flow.

3.3.6 Distributed Hydrology Soil Vegetation Model

The DHSVM is a physically-based, spatially distributed LSM that resolves the energy and water balances (Wigmosta et al., 1994). The physics routines in the DHSVM are largely similar to VIC, though typically applied on a much finer resolution from 10 m to 250 m. Unlike VIC, the DHSVM incorporates a dynamic routing mechanism by redistributing runoff downslope based upon a DEM. Initially developed to simulate hydrology in forested mountainous regions, the DHSVM incorporates hydrologic components of canopy interception, evaporation, transpiration, subsurface saturation excess, and snow accumulation and melt.

Doten et al. (2006) adapted the model to include soil loss and sediment transport, using components from the hydrologic model as inputs. To generate and transport sediment, DHSVM applies four major processes: hillslope erosion, forest road erosion, mass wasting, and channel routing using a discrete approximation to the kinematic wave equation. The hillslope erosion model was based on the System Hydrologique European sediment (SHESED) model (Burton & Bathurst, 1998; Wicks & Bathurst, 1996) and incorporates overland flow and raindrop impact. For the purposes of this study, we focused on the hillslope erosion processes.

The hillslope erosion component of the model is deterministic and estimates detachment from three primary mechanisms: overland flow, raindrop energy, and leaf drip impact. Overland flow detachment (D_{of}, $m^3/s/m$) is calculated using a detachment coefficient (β_{de}), horizontal hillslope length (dy, m), settling velocity (v_s, m/s) and transport capacity (TC, m^3/m^3) as

$$D_{of} = \beta_{de}(dy)(v_s)(TC) \tag{16}$$

where β_{de} is estimated from soil cohesion (C_s, kPa) as

$$\beta_{de} = 0.79e^{-0.6C_s} \tag{17}$$

Settling velocity is dependent on median particle grain size (*D50*, m) and is calculated according to

$$v_s = \sqrt{\frac{4}{3} G \left(\frac{PDensity}{WDensity} - 1\right) D50} \tag{18}$$

where *PDensity* is particle density (kg/m^3), *WDensity* is water density (kg/m^3), and *G* is gravitational acceleration (m/s^2). Transport capacity is based on a unit stream power approach from the KINEROS model (Woolhiser, Smith, & Goodrich, 1990) using stream power (*SP*, m/s) calculated by

$$SP = \frac{streamflow}{(dy)(h)} slope \tag{19}$$

where *streamflow* is overland flow (m^3/s), *h* is water depth (m), and slope is a proportion (m/m). *SP* is then used to calculate transport capacity (*TC*, m^3 sediment / m^3 water) of the overland flow:

$$TC = \frac{0.05}{D50\left(\frac{PDensity}{WDensity} - 1\right)^2} \sqrt{\frac{slope(h)}{G}} (SP - critical\ threshold) \tag{20}$$

where *critical threshold* is a critical stream power threshold (m/s). For this study, we used a critical threshold of 0.004 m/s according to Doten et al. (2006). *DOF* is calculated as sheet flow, however, the dependency of β_{de} on soil cohesion serves to represent an empirical parameter for rill erosion as well.

Rainfall detachment is computed from the momentum squared of direct throughfall (M_R) and of leaf drip from vegetation (M_D) according to Wicks and Bathurst (1996). Momentum squared of direct throughfall is determined by

$$M_R = \alpha I^{\beta} \tag{21}$$

where *I* is rainfall intensity (mm/hr), and α and β are coefficients established from

regression studies by Wicks (1988). Momentum squared of leaf drip is determined by

$$M_D = \frac{\left(\frac{V\rho\pi D^3}{6}\right)^2 Drip\ Drain}{\left(\frac{\pi D^3}{6}\right)} \quad (22)$$

where V is leaf drip fall velocity (m/s) (adapted from Epema and Riezebos (1987), ρ is water density (kg/m^3), D is leaf drip diameter (m), *Drip* is the percentage of canopy drainage that falls from leaves, and *Drain* is drainage from the canopy (m/s). For this study, *Drip* is set to 5 mm per Wicks (1988).

The momentum squared of direct throughfall and leaf drip is combined to estimate overall raindrop detachment (D_R, kg/m^2/s) through

$$D_R = Kindex(f)(1 - C_G)((1 - C_c)M_R + M_D) \quad (23)$$

where *Kindex* is an erodibility parameter (1/J), f is a water depth correction factor, C_G is ground cover proportion, and C_C is canopy cover proportion. To determine f, the raindrop diameter (D_m) is related to water depth, as raindrop impact increases if raindrop diameter is larger than the water depth (Park, Mitchell, & Scarborough, 1982):

$$f = \exp\left(1 - \frac{h}{D_m}\right) \quad if\ h > D_m \quad (24)$$

$$f = 1 \quad if\ h \le D_m$$

Raindrop diameter is calculated from the Laws and Parsons equation (1943) as

$$D_m = 0.00124I^{0.182} \quad (25)$$

To estimate SSC, Doten et al. (2006) integrate overland flow detachment and raindrop detachment in DHSVM using a modified version of the SHESED finite difference equations for uniform sheet flow (Wicks & Bathurst, 1996) based on the two-dimensional conservation of mass equation:

$$\frac{\delta(QC)}{\delta x} + \frac{\delta(AC)}{\delta t} = e(x,t) \quad (26)$$

where Q is water volume, A is cross-sectional area of the streamflow, C is the concentration of sediment, x is horizontal distance, t is time, and e is erosion (area per unit time). The DHSVM approach assumes lateral transfer between model grid cells, and therefore accounts for both temporal and spatial sediment transport in the finite difference solution. Because VIC requires an external routing model, we adapted the equation in our integration to solely account for sediment concentration (SSC, m^3/m^3) at a single grid cell during the current time step as:

$$\mathrm{SSC} = \frac{D_r + D_{of}}{\frac{\alpha}{2dt}Q^{\beta} + \frac{\theta}{dx}Q + \beta_{de}(dy)(v_s)} \tag{27}$$

where Q is runoff (m^3/s), dx is hillslope dimension (m), dt is timestep (s), α is channel area determined by Manning's equation, β is set to 2/3, and θ is a time weighting factor set to 0.55 per Doten et al. (2006). The time weighting factor ranges from 0.55 to 1.0, and improves the stability of the finite difference equations (Wicks & Bathurst, 1996).

After SSC is calculated, TC is then treated as a threshold for how much sediment can be transported for a given streamflow. If SSC exceeds TC, SSC is set to TC and excess sediment is deposited in the grid cell. Therefore, final sediment outputs rely heavily on TC.

DHSVM is the most complex model in our inter-comparison, as it simulates erosion and sediment yield through empirically determined and physically based methods. However, our adaptation of the model is limited due to the spatial upscaling of the VIC integration. DHSVM relies heavily on fine resolution vegetation and topography, and our current large-scale implementation limits the amount of detail in these physical processes. Furthermore, the finite difference solution to the conservation of mass equation accounts for changes in space and time, and our adaptation only accounts for a single timestep and a single distance, potentially introducing bias into our estimates by underrepresenting the sediment contribution from upslope grid cells and previous time steps.

Like the HSPF, we integrated DHSVM into VIC by computing erosion estimates within each VIC grid cell for the daily timestep. To obtain Q, we used the sub-daily VIC runoff calculations for each grid cell, as this represents overland flow.

3.4 Proposed Methodology and Rationale

3.4.1 Coupling of Erosion and Sediment Algorithms with VIC

We coupled the individual algorithms from each of the models listed above with VIC, thereby developing a diverse estimate of erosion processes and SSL applicable to large-scale catchments. Like VIC-WEPP (Mao et al., 2010), we applied the hillslope erosion computations using the hydrologic, soil, and vegetation components from VIC. Because the MRC and the LOADEST require streamflow at the catchment outlet, we used calibrated streamflow values to run the two modules and we calculated SSL after routing simulated streamflow to the catchment outlet. For the MUSLE, HSPF and DHSVM, we inserted the erosion computations into the VIC algorithm using a critical area approach (described in the next section) to calculate sediment on fine resolution hillslopes in productive regions of the catchment. Using the percentage of critical area in the catchment, we then scaled up to the native VIC resolution. Using RVIC and a multi-objective calibration routine, we routed the eroded soil from the MUSLE, HSPF, and DHSVM through the catchment and produced calibrated SSL outputs at the catchment outlet.

3.4.2 Hillslope calculations and critical area approach

A difficulty in estimating catchment-scale erosion and sediment transport lies in spatial discretization. As erosion processes often occur on the scale of meters, implementing erosion equations on the scale of kilometers be computationally cumbersome and

inefficient. Therefore, in our study we used a critical area approach to isolate model computations over productive regions where soil erosion is most likely to occur. Applying this approach, for each VIC grid cell (~6 km resolution) we calculated erosion from the MUSLE, HSPF and DHSVM using the metrics of an individual hillslope (e.g., calculated from a 10 m resolution DEM) incorporating hydrologic inputs from the VIC grid cell but scaled down to the hillslope dimensions. After computing erosion estimates on the hillslope from each method, we then used critical area as a metric for scaling the erosion up to the VIC grid cell:

$$\mathrm{SSL} = HE\left(\frac{Grid\ Area}{HE\ Area}\right)(Critical\ Area) \tag{28}$$

where SSL is in (kg/m^2/s), *Grid Area* is the area of an individual VIC grid cell, *HE Area* is the hillslope area, and *Critical Area* is the proportion of the grid cell that is deemed to contribute sediment supply.

We identified the critical area as those areas within each catchment exceeding a slope steepness threshold together with the intersection between specified vegetation types and stream proximities. For slope steepness, we used the threshold developed by Larsen et al. (2014), that defines 15˚ slopes and above in mountainous regions (~10% of global topography) to contribute >50% of the estimated global sediment (19 Gt/yr) (Figure 2a). Because erosion occurs most frequently on shrubs, grasslands or bare ground (SGB), we incorporated these land cover types into our calculations. However, to ensure adequate coverage, the inclusion of forested coverage into our estimates was needed (Figure 2b) because our catchments are largely forest-dominated. To develop an uncertainty range in our estimates, we applied two stream proximities into our filtering: 100 m and 500 m (Figure 2c) and compared them with both individual and combined land cover types.

Critical area was determined using the same method across all catchments, and for

demonstration we use CC as an example below. For CC the critical area ranged from 0.21% applying SGB within a 100 m stream proximity, to 3.8% applying combined forest and SGB within a 500 m stream proximity (Figure 2d, Figure 3). Results were similar for CLP-L with ranges from 0.25% to 2.3%, whereas results for CLP-F were higher with ranges from 3.2% to 36%.

3.4.3 DEM and Slope Distribution Selection

Distributed soil loss and sediment transport models account for topography on varying degrees depending on the grid cell resolution. However, studies show that a coarse grid cell resolution leads to decreased variability in slope steepness estimates and ultimately causes substantial underestimations in soil erosion (Aalto et al., 2006; Larsen et al., 2014). For example, in a study on the Back Creek catchment in the Upper Roanoke River Basin, Wu, Li and Huang (2005) estimated soil loss using the Universal Soil Loss Equation (USLE) across a range of 10 m to 250 m grid cells. Their study analyzed the differences in the slope length and steepness factor (LS) between grid cell sizes, and computed a decrease in soil erosion of 60% from the 10m grid to the 250m grid. The decrease in soil erosion was attributed to reduced LS values in the larger cells. Zhang and Montgomery (1994) completed a similar study on grid cell resolution effects at Mettman Ridge near Coos Bay, Oregon and Tennessee Valley in Marin County, California. Using digital elevation models (DEMs) gridded to 2, 4, 10, 30, and 90 m, Zhang and Montgomery (1994) found the distribution in mean topographical slope to vary substantially among grid cell resolutions. In their analysis, they documented that a 10 m resolution is adequate for capturing the topographical characteristics while also minimizing computation. Therefore, we elected to use a 10 m DEM as the resolution in our hillslope erosion estimates.

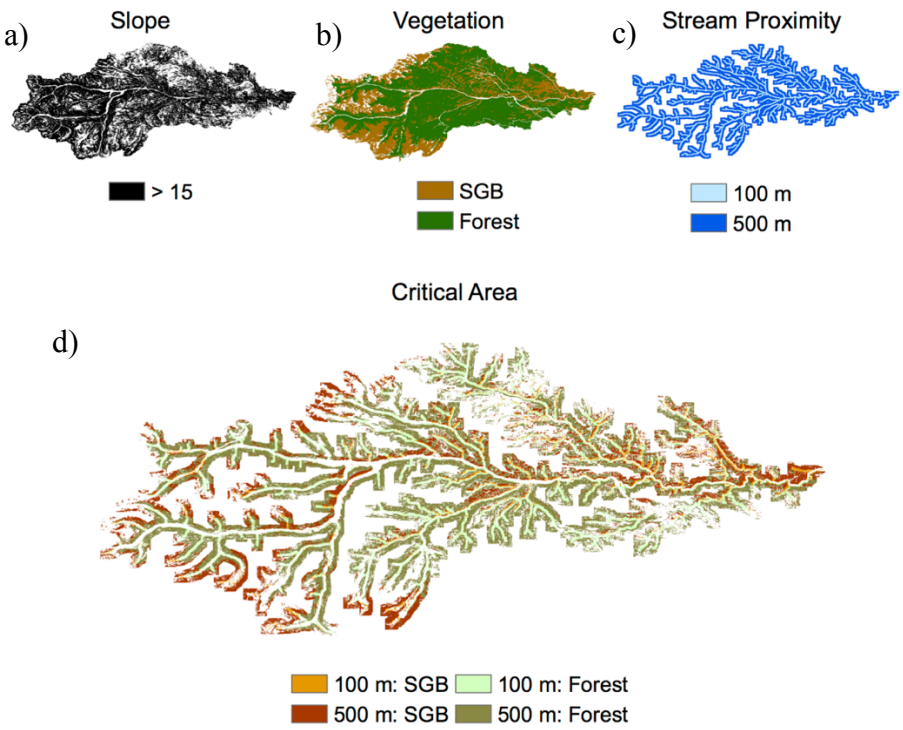

Figure 2. Critical areas computed for CC using (a) slope threshold of 15˚ and above, (b) vegetation types including shrub, grassland, bare ground (SGB), and forest, (c) and stream proximities of 100 m and 500 m. Critical area included four estimated regions (d).

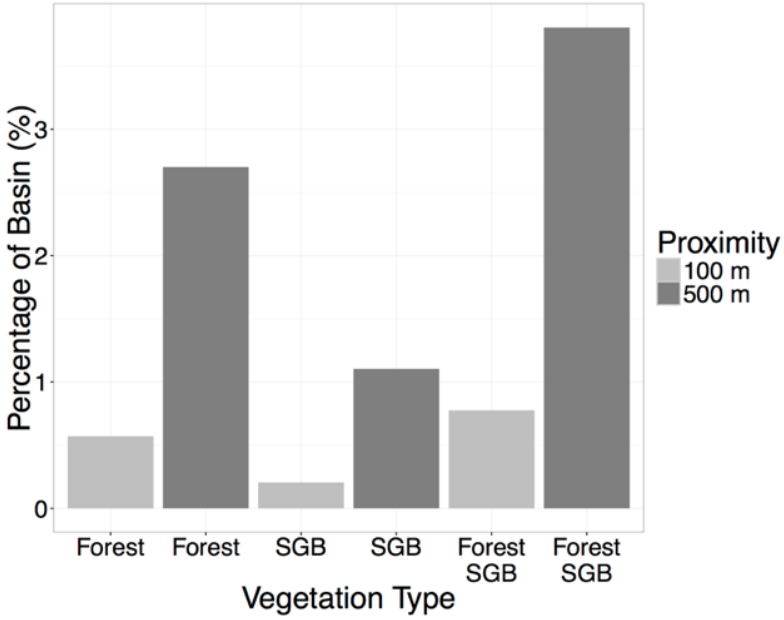

Figure 3. Range in critical area percentages of CC catchment based on varying vegetation types of shrubs, grassland and bare ground (SGB) and forest, and stream proximities.

We calculated slope from a 10 m DEM (Available for the US by USGS National Elevation Dataset[4]) for all pixels situated within each VIC grid cell (Figure 4a). Because our critical areas are dependent on slopes greater than 15˚, we generated slope distributions above that same threshold for each VIC grid cell (Figure 4b, Figure 4c). To incorporate a range of slopes into our erosion estimates, we applied the 25th, 50th and 75th percentiles and developed a mean SSL estimate from the three slopes for each VIC grid cell.

3.5 Algorithm Optimizations

3.5.1 Sensitivity Analysis

The performance of hydrologic and SSL models are dependent on numerous parameters. To minimize computation and maximize efficiency, studies have isolated parameters that hold the largest influence on model performance. These efforts are deemed as sensitivity analyses and include both local and global methods. Local sensitivity analyses change a single parameter at a time and evaluate the model performance. Global sensitivity

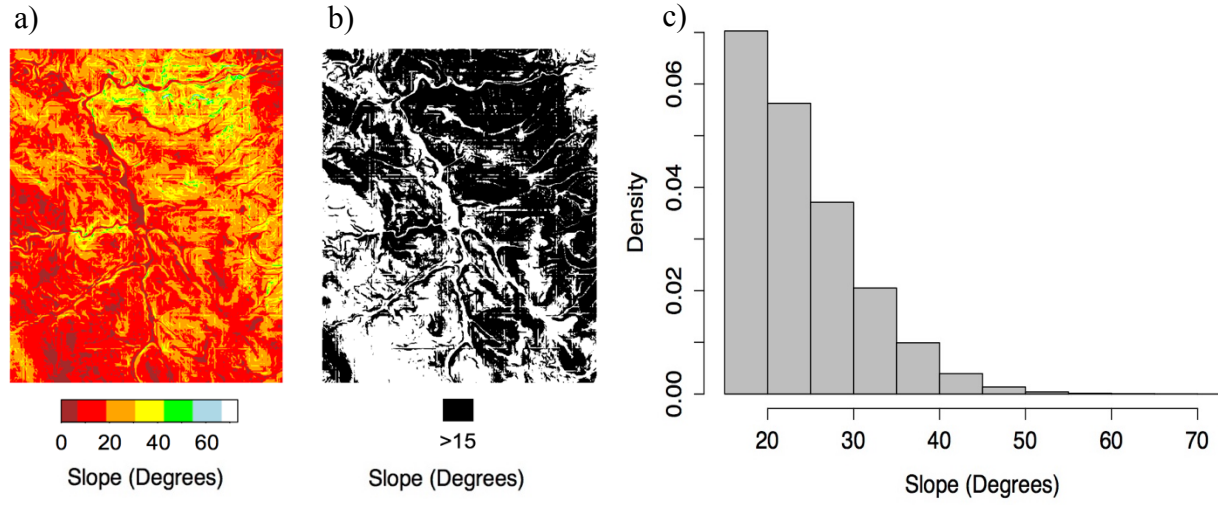

Figure 4. (a) Slope calculations for an example CC VIC grid cell, (b) slopes sorted to greater than 15˚ (c), distribution of slopes greater than 15˚.

[4] https://viewer.nationalmap.gov/basic/

analyses compare performance changing multiple parameters simultaneously. While local methods identify individual parameter effects, global methods are more robust and characterize complex parameter interactions (van Werkhoven, Wagener, Reed, & Tang, 2009). For this study, we ran a Latin Hypercube Sampling (LHS) test, which uses random and stratified sampling to span the entire parameter space and has been effective in identifying global sensitivities (Tang, Reed, van Werkhoven, & Wagener, 2007). To run LHS, we used 500 samples and 8 soil parameters to identify streamflow parameter sensitivities.

3.5.2 Calibration, Validation and Transferability

We used multi-objective optimization to calibrate model outputs in the CLP-L and CC catchments with daily observed streamflow and observed instantaneous SSL. We ran one experiment in CLP-L, and two experiments in CC using different time periods in the calibrations. We elected to run the two CC experiments to explore the effects of large SSL sample sizes in 1981 (CC-Early) versus small SSL samples sizes in the 1990s (CC-Late). We validated CLP-L, CC-Early and CC-Late using the same catchments but over a different time period (Table 2). We then assessed transferability of the parameters by applying the top performing parameter sets in CLP-L, CC-Early and CC-Late to CLP-F. To capture meteorological and streamflow fluctuations in our calibrations, we selected five and six-year periods surrounding the dates of our suspended sediment observations. The experiments were designed to explore the impact of calibrating over different sample sizes, time periods, ranges in data magnitudes, and catchments.

To explore differences in multi-objective optimization techniques, we ran individual calibrations for streamflow and each SSL algorithm, and joint calibrations of streamflow and each SSL algorithm. Though the two optimization methods are fundamentally different

and the results cannot be directly compared, the methods provide insights into how an individual and a joint ensemble can provide different results. The parameters and parameter bounds are listed in Table 3 for each method. Parameter bounds were obtained from the literature for each algorithm. For VIC streamflow, we incorporated eight soil parameters using parameter bounds from Demaria, Nijssen, and Wagener (2007), Troy, Wood and Sheffield (2008), and Yanto, Livneh, Rajagopalan, and Kasprzyk (2017). For DHSVM we incorporated three SSL parameters using parameter bounds from Doten et al. (2006) and Maidment (1993). For MUSLE we incorporated three SSL parameters using parameter bounds from Maidment (1993). For HSPF we incorporated seven SSL parameters and used parameter bounds from Donigian and Love (2003).

Table 2. Description of the calibration, validation and transfer experiments for the Cache La Poudre Livermore (CLP-L) and Fort Collins (CLP-F), and Clear Creek (CC) gauges, including an early period "Early" and late period "Late", used alternatively for calibration, validation and inter-basin transfer or parameters "Transfer". We used two calibration methods: calibration of all algorithms individually "Individual Modules", and calibration of all algorithms simultaneously "Joint Modules."

Experiment	Calibration Type	Calibration Period	Validation Period	Transfer to CLP-F Period
CLP-L	Joint Modules	1987 – 1993	1993 – 1999	1992 – 1997
CC-Early	Joint Modules	1980 – 1985	1992 – 1997	1992 – 1997
CC-Late	Joint Modules	1992 – 1997	1980 – 1985	1992 – 1997
CC-Late	Individual Modules	1992 – 1997	1980 – 1985	--

3.5.3 Borg-VIC Optimization

Single objective calibrations tend to be biased towards fitting certain parts of the hydrograph, for example, using a transformed root mean square error to calibrate low streamflow periods can cause misrepresentations of peak streamflows, whereas using a percent bias objective largely ignores the goodness of fit of variability and timing. To compensate for these errors, multi-objective evolutionary algorithms (MOEAs) have been developed to identify tradeoffs in performance measures and to locate Pareto optimal

Table 3. Parameters used in the multi-objective optimization routine for each calibration. Because sediment modules are dependent on streamflow (VIC Flow), we incorporated sensitive VIC soil parameters into the individual module calibrations as well.

			Individual Calibrations				Joint Calibration
Parameter	Definition	Bounds	VIC Flow	DHSVM	MUSLE	HSPF	All
Binf	Infiltration Capacity	0.0001 – 0.4					
Ds	Fraction of DsMax where non-linear baseflow occurs	0.0001 – 1.0					
DsMax	Maximum baseflow velocity	0.1 – 30.0					
Ws	Fraction of maximum soil moisture where non-linear baseflow occurs	0.1 – 1.0					
C	Baseflow curve exponent	1.0 – 2.0					
Layer 1	Soil layer depth 1	0.1 – 0.3					
Layer 2	Soil layer depth 2	0.3 – 3.0					
Layer 3	Soil layer depth 3	0.3 – 3.0					
K Factor	Erodibility factor	0.02 – 0.6					
C Factor	Cropping management factor	0.0001 – 0.5					
P Factor	Conservation practice factor	0.0 – 1.0					
K Index	Erodibility index	19.0 – 32.0					
D50	Median grain size	0.5 – 2.0					
Soil Cohesion	Soil cohesion	0.00075 – 15.0					
JR	Detachment exponent	1.0 – 3.0					
AFFIX	Attachment fraction	0.01 – 0.50					
KS	Transport coefficient	0.1 – 10.0					
JS	Transport exponent	1.0 – 3.0					
KG	Scour coefficient	0.0 – 10.0					
JG	Scour exponent	1.0 – 5.0					
Critical Area	Catchment critical area	CC: 0.002 – 0.04 CLP-L: 0.003 – 0.02					

solutions, which are non-dominated solutions (Gupta, Sorooshian, & Yapo, 1998). MOEAs are optimization methods developed for using multiple objectives, parameters, and constraints to accurately and efficiently calibrate complex problems. Particular MOEA frameworks incorporate machine learning to actively adapt to finding solutions, such as in the Borg MOEA (Hadka & Reed, 2012). For this study, we used the Borg MOEA to optimize five objective functions described below using multiple streamflow and sediment parameters across all modules. For each calibration initiative, we ran 20,000 – 30,000 function evaluations depending on the complexity of the problem, and incorporated 5 random seeds to ensure the solutions were not an artifact of the Borg parameter search. We used epsilons of 0.1 for all objective functions in this study, as this provides a balance between coarseness and ability to span the range of the data. Epsilon-dominance is a method to bound complex problems by setting precision values for Pareto optimal solutions (Kollat, Reed, & Wagener, 2012). To increase efficiency, we developed a parallelized program using Message Passing Interface to couple Borg and VIC to efficiently calibrate our problems.

We selected five objective functions for our calibration to capture various components of the hydrograph and SSL time series (Table 4). The first objective used in the calibration was Nash Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970):

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{n}\left(x_{s,t} - x_{o,t}\right)^2}{\sum_{t=1}^{n}\left(x_{o,t} - \mu_o\right)^2} \tag{29}$$

where $x_{s,t}$ and $x_{o,t}$ are the simulated and observed values for each time step (t), n is the number of data points, and μ_o is the mean of the observed data. An NSE value of 1 is considered a perfect model simulation, whereas an NSE value of less than 0 is considered to perform worse than using the mean of the observed data as the predictor. An NSE value ≥ 0.50 is considered satisfactory for both streamflow and SSL model performance (Moriasi et

al., 2007).

The second objective used in the calibration was Pearson's Correlation Coefficient (R), which describes the linear correlation between simulated and observed values and is often used to characterize goodness of fit in timing between simulated and observed time series:

$$\mathrm{R} = \frac{n\sum_{t=1}^{n} x_{o,t}x_{s,t} - \sum_{t=1}^{n} x_{o,t}\sum_{t=1}^{n} x_{s,t}}{\sqrt{\left(n\sum_{t=1}^{n} x_{o,t}^2 - \left(\sum_{t=1}^{n} x_{o,t}\right)^2\right)\left(n\sum_{t=1}^{n} x_{s,t}^2 - \left(\sum_{t=1}^{n} x_{s,t}\right)^2\right)}} \quad (30)$$

an R value of 1 implies perfect positive correlation, -1 implies perfect negative correlation, and 0 implies no correlation. R^2 values of > 0.5 are considered satisfactory for hydrologic model performance (Santhi et al., 2001).

The third objective used was percent bias (PBIAS) to estimate the overall simulated magnitude compared with observed (Gupta, Sorooshian, & Yapo, 1999):

$$\mathrm{PBIAS} = \frac{\sum_{t=1}^{n} x_{o,t} - x_{s,t}}{\sum_{t=1}^{n} x_{o,t}}(100) \quad (31)$$

A PBIAS value of 0 is considered to have no magnitude bias in the model. A PBIAS value $\pm$ 25% is considered satisfactory for streamflow, and a PBIAS value $\pm$ 55% is considered satisfactory for SSL (Moriasi et al., 2007).

The fourth objective used was relative variability (VAR), to reflect the relative range in values between the model and the observed data using standard deviations of the data (σ) (Gupta, Kling, Yilmaz, & Martinez, 2009):

$$\mathrm{VAR} = \frac{\sigma_s}{\sigma_o} \quad (32)$$

A VAR value of 1 is considered a perfection representation of model variability.

The final objective used in this calibration was the transformed root mean squared error (tRMSE) selected to target low streamflow and low SSL periods by using a Box-Cox

transformation (Wagener, van Werkhoven, Reed, & Tang, 2009):

$$\mathrm{Z} = \frac{(1+x)^{\lambda} - 1}{\lambda} \tag{33}$$

$$\mathrm{tRMSE} = \frac{1}{n}\sum_{t=1}^{n}\left(Z_{s,t} - Z_{o,t}\right)^2 \tag{34}$$

where λ is 0.3. A tRMSE value of 0 is indicative of good model performance, however, the units and magnitudes of the metric are dependent on the data.

Table 4. Objective functions and their intervals used in this study with optimal values obtained from Moriasi et al. (2007) for NSE and PBIAS.

Objective Function	Reference	Interval	Optimal Value	Moriasi et al. (2007) Satisfactory Value
NSE	Nash and Sutcliffe, 1970	-∞ to 1	1	> 0.50
PBIAS	Gupta et al., 1999	-∞ to ∞	0	< ±25% streamflow < ±55% SSL
VAR	Gupta et al., 2009	0 to ∞	1	--
tRMSE	Wagener et al., 2009	0 to ∞	0	--
R	Pearson, 1920	-1 to 1	1	--

CHAPTER 4

RESULTS

A series of experiments were conducted to explore sediment and streamflow prediction performance in three modes: calibration, validation, and transferability of parameters to a neighboring basin. These experiments were carried out over three catchments, Cache La Poudre Livermore (CLP-L), Clear Creek (CC), and Cache La Poudre Fort Collins (CLP-F) with long-term streamflow observations and suspended sediment load (SSL) observations during multiple time periods. Five sediment transport algorithms were evaluated based on six objective functions in an individual algorithm calibration, and two objective functions in a joint algorithm calibration. The observational data are presented below, followed by a description of the results from the three types of experiments, exploring tradeoffs in performance for each. We present results from the MRC and the LOADEST calibrations, individual algorithm Borg calibrations, and joint algorithm Borg calibrations, and the validation and transferability from the joint calibrations. The different categories of experiments and algorithms tested are in Table 2.

4.1 Observational Data Analysis

Observed SSL exhibited a wide range in magnitude across the four experiments for calibration, validation, and transfer periods, with CC-Late having the highest range in SSL (Figure 5). This disparity in the relative SSL between periods could have been caused by

SSL collection timing, as diurnal streamflow fluxes can impact SSL (Glysson, 1987; Runkel et al., 2004) (Appendix A: Figure 16). Streamflow stayed more consistent across the sites and time periods, with CLP-F having the highest maximum and mean streamflow likely due to the larger basin area.

4.1.1 Suspended Sediment Load Sample Sizes

We tested the statistical representativeness of the SSL sample sizes in each experiment for the calibration, validation, and transfer periods using bootstrap sampling from the empirical cumulative density functions (CDFs). We generated 150 random samples from each empirical CDF and calculated the mean, median, standard deviation (SD), skew, interquartile range (IQR), and probability density function (PDF). For each experiment and period, the empirical measures of accuracy fell within the first and third quartiles of the estimated measures of accuracy (Appendix A: Figure 17 - Figure 21), indicating that the sample sizes were representative of the populations.

4.1.2 Climatic Differences Between Calibration and Validation Periods

We conducted a paired Wilcoxon Rank Sum Test on observed annual streamflow between the calibration and validation periods for CLP-L, and CC-Early and CC-Late, as the data were not normally distributed and we therefore could not use a t-test. For CLP-L, CC-Early and CC-Late we found that the ranks for the calibration period and the validation period were not statistically different (Table 5) though the ranges were qualitatively more variable in CLP-L (Figure 5). These findings indicate that streamflow has similar distributions between calibration and validation periods in both CLP-L and CC-Early. However, the Wilcoxon Rank Sum Test does not assess the relative magnitudes between the ranked data, and therefore might not be representative of the differences in variability

for each period. To assess differences in climate between the calibration and validation periods, we performed similar tests. We analyzed precipitation, maximum temperature and minimum temperature, and found no significant differences between the periods within both catchments other than for minimum temperature in CLP-L (Figure 6, Table 5).

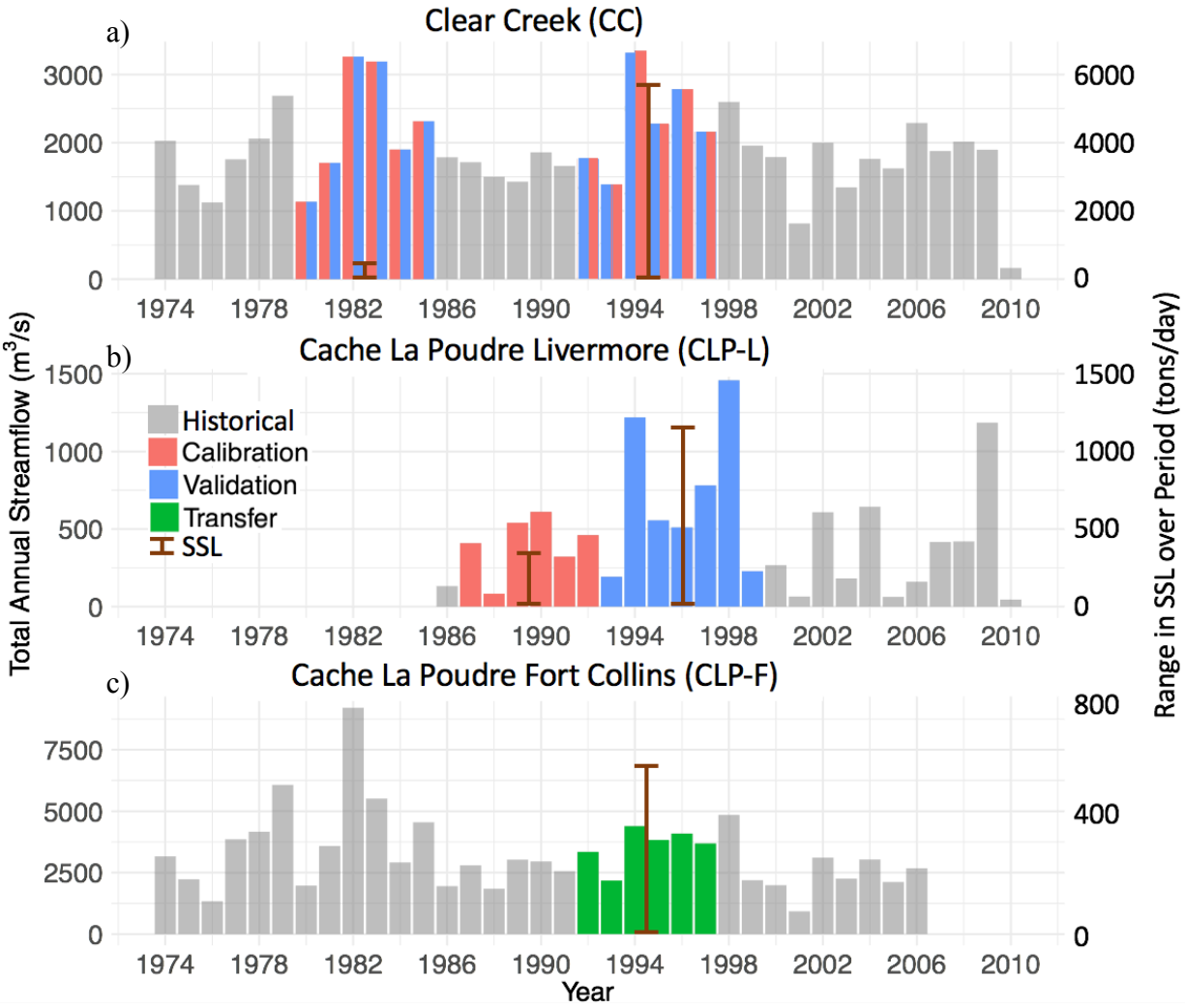

Figure 5. Bar plot of total annual streamflow separated by site into calibration, validation and transfer periods for a) Clear Creek (CC), b) Cache La Poudre Livermore (CLP-L), and c) Cache La Poudre Fort Collins (CLP-F). SSL is depicted by error bars showing the range over the respective period. The lowest annual streamflows occurred in CLP-L, whereas the highest occurred in CLP-L. The largest SSL range occurred in CC over 1992-1997, whereas the lowest occurred in CC over 1980-1985. There is a disparity in the relative SSL in CC between the two periods, which could cause poor performance in the SSL simulations as the streamflows have comparable volumes.

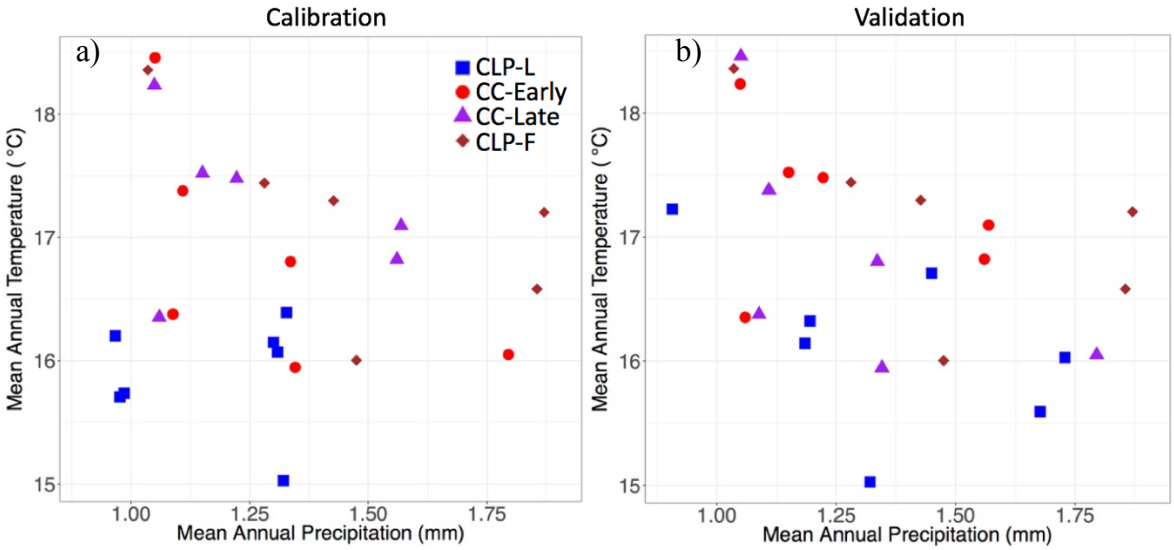

Figure 6. Average daily temperature and precipitation for each experiment during the a) calibration period, and b) validation period. Data was obtained from Livneh et al. (2015). CLP-L has a bias towards cold and dry years during the calibration period and towards cold years during validations, whereas CC-Early, CC-Late and CLP-F are more normally distributed with a central tendency of warmer years than CLP-L.

Table 5. Wilcoxon Rank Sum test for differences in climate and annual streamflow between calibration and validation periods. Streamflow and precipitation were annual sums, whereas temperatures (T) were annual daily means. W is the Wilcoxon Rank Sum score, and the p-value denotes significance. P-values < 0.05 are considered significant and denoted with *. For streamflow, precipitation and Max T, the distributions and spread of the data were similar for both periods, whereas for Min T in CLP-L, the calibration period was colder.

		Calibration Mean	Validation Mean	Wilcoxon W	p-value
CLP-L	Streamflow	404 m^3/s	706 m^3/s	5	0.31
	Precipitation	370 mm/yr	430 mm/yr	9	0.47
	Min T	0.50 ˚C	1.8 ˚C	0	0.016*
	Max T	15 ˚C	15 ˚C	7	0.29
CC-Early	Streamflow	2,275 m^3/s	2,250 m^3/s	10	1
	Precipitation	393 mm/yr	383 mm/yr	12	0.84
	Min T	0.69 ˚C	0.07 ˚C	11	1
	Max T	17 ˚C	17 ˚C	11	1

4.2 Calibration

4.2.1 Monovariate Rating Curve Calibration

We calibrated the MRC for each experiment using historical streamflow and SSL observations to minimize non-linear least squares estimation (Appendix B: Table 7). The MRC performance ranged from R^2 of 0.64 in CC-Early to R^2 of 0.98 in CLP-L over the calibration periods. The MRC in CC-Late required a much greater b exponent than the other experiments to compensate for the non-linearity stemming from large SSL magnitudes (Figure 7). Extrapolations of the MRC for both CLP-L and CC-Early show comparatively smaller estimates than the MRC in CC-Late.

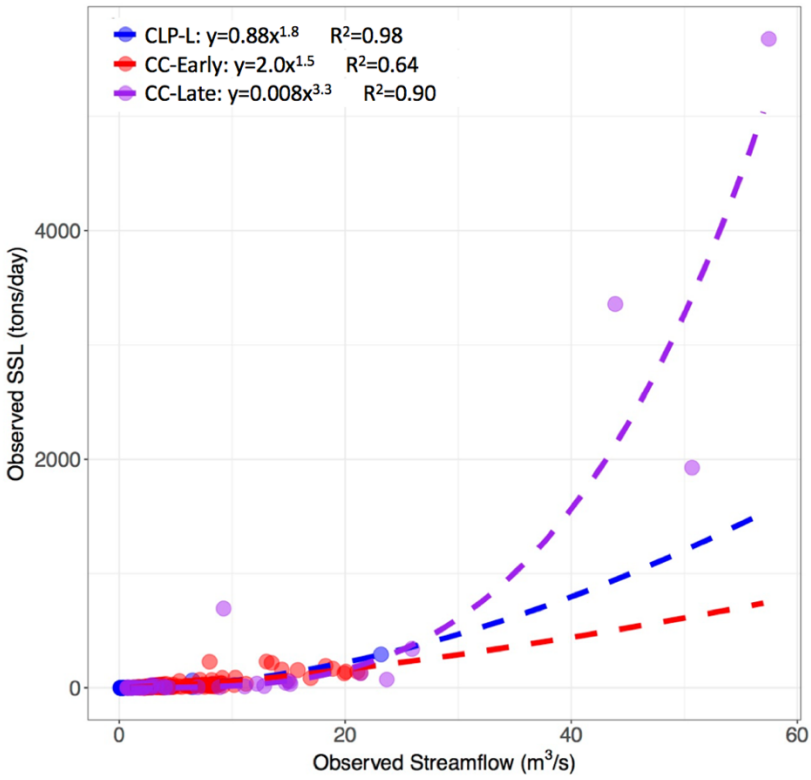

Figure 7. The MRC for each experiment developed over the calibration period. The MRC b exponent was the greatest in CC-Late, causing a larger estimation of SSL than CLP-L and CC-Early at high streamflows.

4.2.2 Load Estimator Calibration

To develop multi-variate regression estimates, we used LOADEST and formulated our experiments with Adjusted Maximum Likelihood Estimates (AMLE). The model selection process used Akaike Information Criterion (AIC). CLP-L, CC-Early and CC-Late selected different models. CLP-L incorporated streamflow and date as

$$\ln \hat{SSL} = a_0 + a_1 lnQ + a_2 lnQ^2 + a_3 \sin(2\pi t) \quad (35)$$

whereas CC-Early required additional permutations of both predictors:

$$\ln \hat{SSL} = a_0 + a_1 lnQ + a_2 lnQ^2 + a_3 \sin(2\pi t) + a_4 \cos(2\pi t) + a_5 t \quad (36)$$

CC-Late required an additional non-linear predictor for date as:

$$\ln \hat{SSL} = a_0 + a_1 lnQ + a_2 lnQ^2 + a_3 \sin(2\pi t) + a_4 \cos(2\pi t) + a_5 t + a_6 t^2 \quad (37)$$

For each model above, a_{0-6} are coefficients, lnQ is the log of streamflow, and t is decimal time, where lnQ and t are defined per Equations 9 and 10. To test normality of residuals, we used Probability Plot Correlation Coefficient (PPCC) (Appendix B: Table 8). The LOADEST performance ranged from R^2 of 0.80 in CLP-L to R^2 of 0.89 in CC-Late over the calibration periods.

4.2.3 Individual Algorithm Borg Calibrations

Using Latin Hypercube Sampling (LHS) with 500 function evaluations, we conducted a sensitivity analysis for VIC streamflow parameters (Appendix C: Figure 22) in which we simultaneously modified all parameters within specified ranges. Through visual assessment of the LHS, we saw general trends in streamflow response to the parameters b_{inf}, Ds, DsMax, and Soil Layer 2 and incorporated these parameters into the optimization of each SSL algorithm, excluding insensitive parameters.

To explore the initial range of values that can be obtained from individual algorithm

calibrations, we used the multi-objective evolutionary algorithm, Borg. We calibrated the streamflow, and each SSL algorithm separately using CC-Late in the experiment using 20,000 function evaluations and 5 random seeds. We used 5 objective functions for streamflow: Nash Sutcliffe Efficiency (NSE), percent bias (Bias), correlation (Cor), variability (Var) and transformed root mean squared error (tRMSE); and 6 objective functions for each SSL algorithm: NSE streamflow, NSE SSL, Bias, Cor, Var, and tRMSE. The suite of objective functions showed tradeoffs for each algorithm (Appendix C: Figure 23 - Figure 25), with the DHSVM algorithm highlighted as an example below (Figure 8). Parameters varied from the lower to the upper bounds across each calibration. To produce the highest SSL NSE, the DHSVM algorithm favored VIC streamflow parameters being small b_{inf}, large DsMax, and small Soil Depth 2. The DHSVM SSL parameters were less sensitive, as the greatest SSL NSE showed a spread across the parameter ranges for K Index, d50, Soil Cohesion (Soil Cosn), and Critical Area. For a description of each parameter, see Table 3. The SSL algorithms showed different patterns in the VIC soil parameters for greatest SSL NSE, indicating that the SSL mechanisms within each algorithm preferred different streamflow characteristics. Within each SSL calibration, there was a tradeoff between SSL NSE and Streamflow NSE.

4.2.4 Joint Algorithm Borg Calibrations

To develop SSL estimates from the algorithms simultaneously, we ran calibration experiments using joint algorithm parameter optimizations together with streamflow optimization. For each joint experiment, we simultaneously calibrated streamflow, the DHSVM, the MUSLE, and the HSPF, using VIC soil parameters and the algorithm specific SSL parameters. We used a subset of the original objective functions in the joint calibration to minimize problem complexity, using NSE and Bias for performance measures across

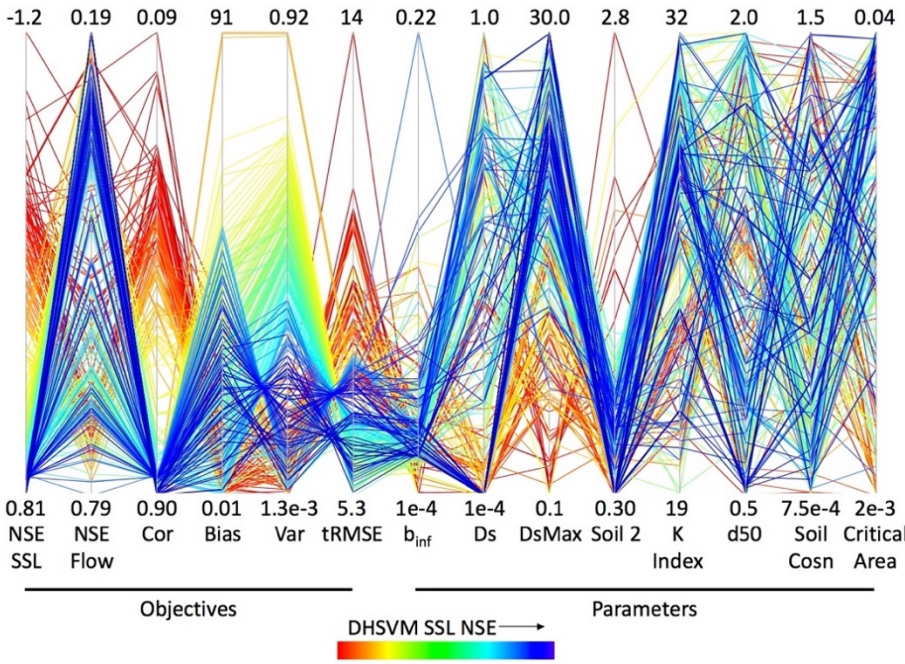

Figure 8. Individual DHSVM SSL algorithm calibration for CC-Late during 1992-1997 with a spin-up period in 1991. There were tradeoffs between the objectives, as well as the VIC soil parameters. DHSVM SSL parameters were less sensitive than VIC soil parameters, as the best performing solutions spanned the ranges of the SSL parameters. Solutions are sorted by NSE SSL, and blue solutions are better performing. Similar plots exist for other algorithms in Appendix C: Figure 23 - Figure 25.

each SSL algorithm and streamflow. We chose these two objective functions to target timing with NSE and magnitude with Bias. We ran the calibrations for 30,000 function evaluations using 5 random seeds and saw that performance was minimally increased after roughly 20,000 evaluations through visual inspection (Appendix D: Figure 26). To generate Pareto optimal solutions, which are non-dominated solutions, we used a combined process of epsilon-dominance and filtering.

Across all algorithms and joint experiments, there were tradeoffs between NSE and Bias (Appendix D: Figure 27 - Figure 29). Our selection process for the best parameter set was the same for all joint experiments, and we describe the procedure below using CC-Late as an example. Each joint experiment had more than 1,000 Pareto optimal solutions, which are too numerous for straightforward selection. To reduce the solutions to a more tractable

number, we used filtration measures for the relative performance of each objective function. For the first filter, we selected the top 50% NSE values across all algorithms (Figure 9). For the second filter, we selected the top 50% Bias values across all algorithms using the previously generated filtered set. In CC-Early and CC-Late, this filtration technique yielded three final solutions. For CLP-L, we identified the three final solutions from the top performing NSE values in the filtered set. The coarse filtration measure of 50% was necessary to generate top solutions that balanced performance measures across algorithms, as the tradeoffs between and within the algorithms were so great.

For the results shown below, we followed the guidelines of Moriasi et al. (2007) by depicting satisfactory performance in daily streamflow estimates as NSE > 0.50 and Bias < ± 25%, and satisfactory performance in daily SSL estimates as NSE > 0.50 and Bias < ± 55%. Overall, the algorithms in the joint calibrations performed worse in NSE and Bias

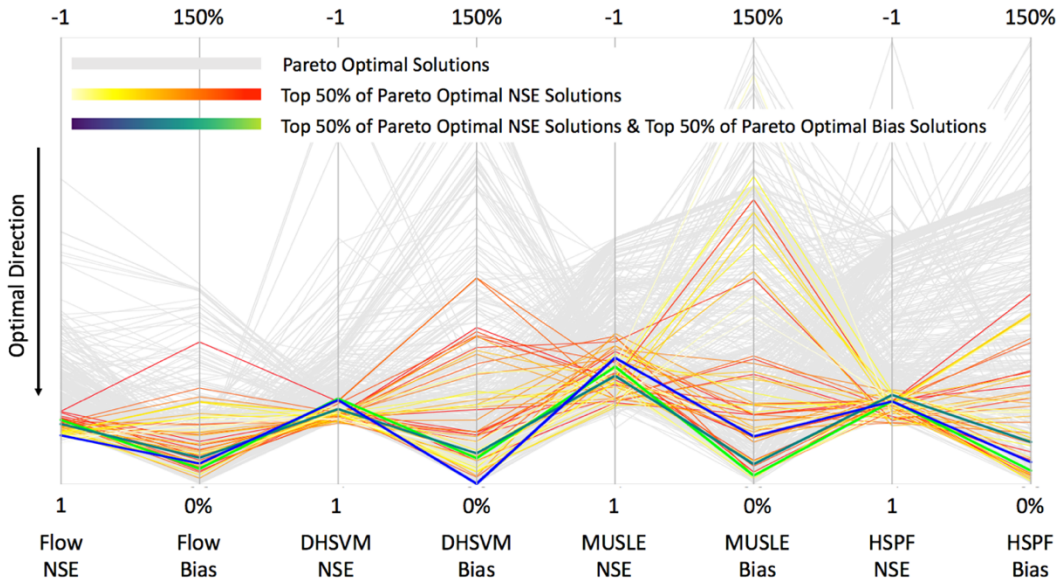

Figure 9. Parallel coordinate plot of Pareto optimal solutions from Borg for CC-Late computed over the calibration period 1992 – 1997 for the joint algorithm calibration with a spin-up year in 1991. Colored solutions indicate the highest performances across the ensemble, with blues and greens representing the final selected solutions. Colored solutions are sorted by streamflow NSE (Flow NSE). The parameter sets from selected solutions were used to run validation and transferability analyses.

than the individual calibrations. This could be attributed to the dynamic dependence of SSL on streamflow, as different equilibrium states with streamflow were optimal for different algorithms. Further, the quantity and type of objective functions varied between methods, and caution should be taken when directly comparing the two calibrations.

We calculated the mean daily streamflow from the top three parameter sets and visually compared the hydrographs with the observed data from the calibration periods for each experiment showing CC-Late as an example below (Figure 10, Appendix D: Figure 31 - Figure 34). We repeated this procedure for SSL (Figure 11, Appendix D: Figure 32 - Figure 35), and also estimated kernel densities for each SSL algorithm to assess distributions (Appendix D: Figure 36, Figure 33, Figure 30).

For CC-Late, the VIC streamflow was flashier than observed and did not capture the peak magnitudes, but the timing was largely captured (Figure 10). The large SSL magnitudes in 1995 biased the algorithms towards higher peak SSL, making them overestimate smaller peaks earlier in the period (Figure 11). From visually assessing the kernel density estimations, LOADEST was best able to capture the distribution of the observed SSL, whereas the HSPF and the DHSVM favored larger magnitudes, the MRC had a flatter distribution, and the MUSLE favored mid-range magnitudes (Appendix D: Figure 30).

For CC-Early, the VIC streamflow overestimated the falling limbs of the hydrograph for 1982 – 1985, but captured the overall shape during 1981 (Appendix D: Figure 31), which is when the observed SSL was also being calibrated. Calibrating the SSL for only one year of the streamflow calibration period therefore could have impacted the streamflow estimates in subsequent years. The algorithms varied in the SSL calibration, as the MRC and the MUSLE maintained a flat behavior and underestimated peaks, whereas the HSPF and the DHSVM were much flashier and captured the peak magnitude (Appendix D: Figure

32). The LOADEST algorithm was unable to capture the observed data during this period, and greatly underestimated the SSL by orders of magnitude. The kernel density estimations showed that the MUSLE and the MRC favored mid-range magnitudes, whereas the HSPF and the DHSVM favored the peaks and the low values and underestimated the mid-range magnitudes (Appendix D: Figure 33).

For CLP-L, the observed streamflow had a large range in magnitudes, and had a prolonged peak in 1993 and multiple peaks in 1990. This could be caused by the 35% reservoir storage upstream. The VIC streamflow, therefore, was unable to capture these fluctuations, and tended to create a flashier system with underestimated peak flows (Appendix D: Figure 34). The large range in observed SSL values impacted the algorithms, as the algorithms favored the three high peaks and overestimated the lower SSL (Appendix D: Figure 35). The kernel density estimations showed that the algorithms favored large magnitudes (Appendix D: Figure 36).

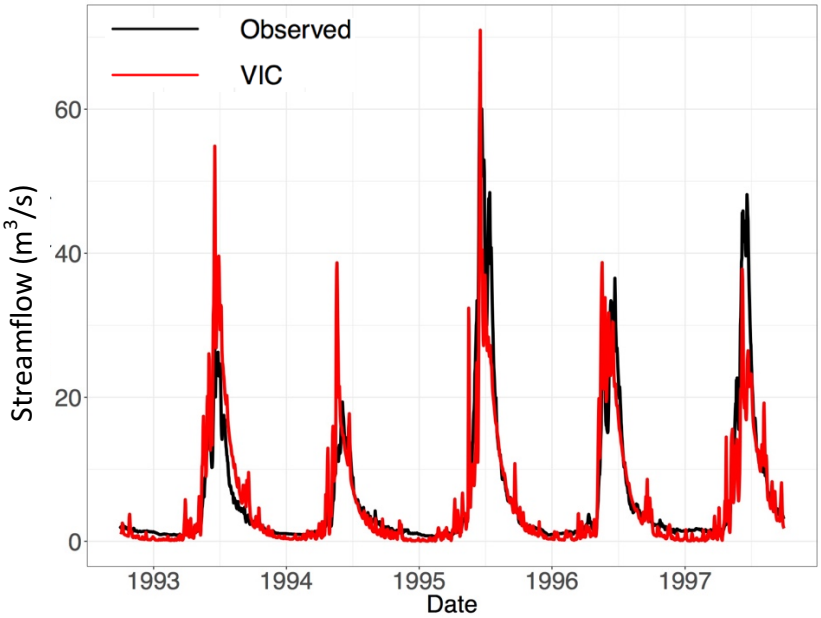

Figure 10. Hydrograph for CC-Late with VIC streamflow computed over the calibration period 1992 – 1997 for the joint algorithm calibration with a spin-up year in 1991. The plotted estimates were calculated from the mean of the VIC streamflow values from the top three performing parameter sets. VIC streamflow was flashier than observed but the timing was largely captured.

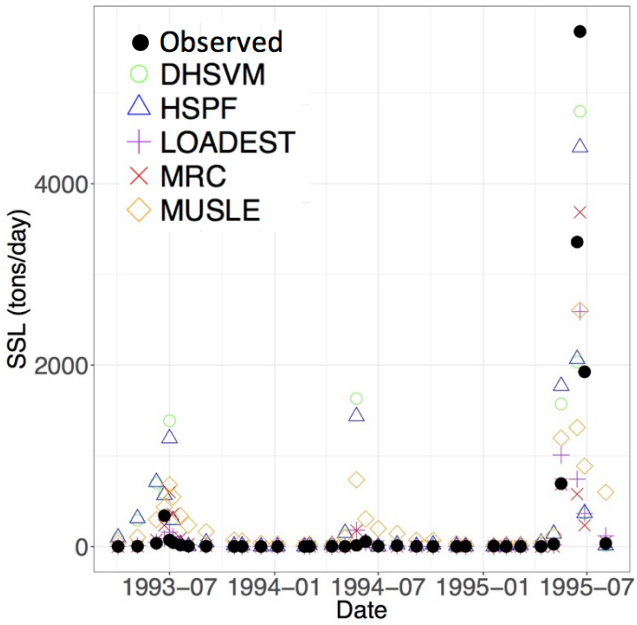

Figure 11. Time series of suspended sediment load (SSL) for CC-Late with each SSL algorithm computed over the calibration period 1992 – 1997 for the joint algorithm calibration with a spin-up year in 1991. The plotted SSL estimates were calculated from the mean of the SSL values from the top three performing parameter sets.

4.3 Validation of the Joint Calibration

Application of the three best performing calibration parameter sets to the validation periods yielded varying results across the experiments and algorithms (Figure 12, Appendix E: Figure 37 - Figure 38). All algorithms generally performed better during the calibration period than the validation period, possibly because the joint calibration did not find the optimal solutions for each catchment system. There were a few exceptions in NSE for the DHSVM and the MUSLE algorithms in CLP-L and streamflow in CC-Early (Figure 12, Appendix E: Table 9 - Table 14). NSE and Bias had similar patterns across algorithms and experiments during the calibration and validation periods. Overall, CC-Late performed best in both NSE and Bias during the calibration. The validation performance, however, was generally poorest in CC-Late for both NSE and Bias. This could be due to the relative variability in streamflow and SSL between the calibration and validation periods. The

performance between the top three parameter sets for each experiment showed a range, though the range was minimal and the three sets generally performed similarly.

4.4 Transferability of the Joint Calibration to Cache La Poudre Fort Collins

We applied the top three calibration solutions from each of CLP-L, CC-Early and CC-Late to the CLP-F catchment to test parameter transferability. Because our filtration technique identified three solutions in two of the basins, we elected to use three solutions for all experiments to represent uncertainty. Streamflow NSE and Bias performance was best in CC-Late, with NSE greater than 0.70 and Bias less than 25%, which we attribute to the fact that the calibration spanned the same period, and hence a similar climate to the CLP-F analysis (Figure 12 - Figure 13). CLP-L had the second-best performance in streamflow NSE and Bias, with NSE greater than 0.50 and Bias less than 50%, which could indicate a connection between streamflow parameters and physical mechanisms in the sub-basin and the larger CLP-F catchment.

Performance of SSL algorithm transferability varied across the sites, though there was a clear trend of poor results coming from the CC-Late transfer across all SSL algorithms and both objectives (Figure 12 - Figure 13). Overall, the DHSVM and the HSPF algorithms had the best NSE performance from the CLP-L and CC-Early transfers (Appendix E: Table 9 - Table 14), with top NSE values exceeding 0.50. In CLP-L, the MUSLE, the LOADEST, and the HSPF algorithms had the best Bias performance, with top values below 50%. In CC-Early, the MUSLE, the HSPF, and the DHSVM algorithms had the best Bias performance, with top values below 50%.

We counted the number of runs per algorithm that met the satisfactory performance criteria of NSE > 0.50 and absolute value of Bias < $\pm$25% for streamflow and < $\pm$55% for

SSL to assess patterns across the experiments and modes (Figure 14). To meet the satisfactory performance criteria, we selected algorithms where one or more runs exceeded the thresholds for both NSE and Bias simultaneously. The clearest signals from the visualization show the criteria were met during calibration for the HSPF in CLP-L, and for calibration of all algorithms in CC-Late. HSPF could have performed best in CLP-L due to its conceptual nature, which lends itself to calibration. For validation, performance criteria were not met in any of the experiments other than the MRC for CC-Late. For transfer, performance criteria were met for the HSPF in CLP-L, the HSPF and the DHSVM in CC-Early, and streamflow in CC-Late.

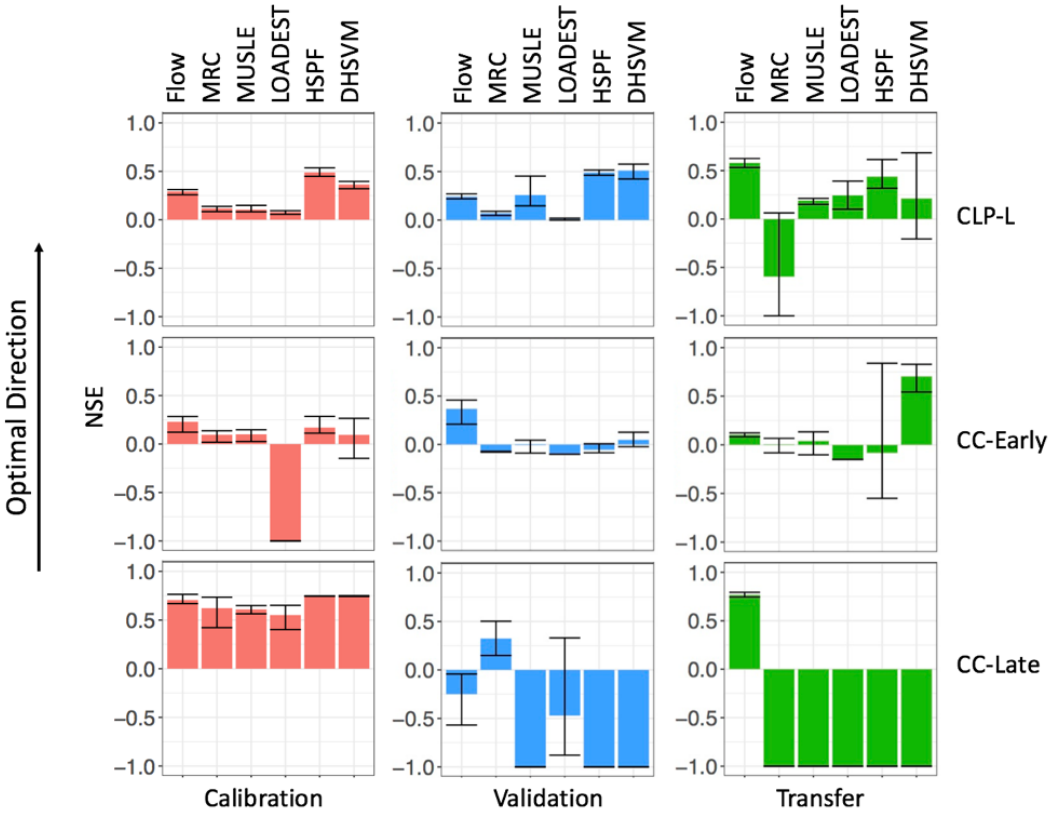

Figure 12. NSE performance for calibration, validation and transfer periods for all algorithms in CLP-L, CC-Early, and CC-Late. Calibration runs used the top three solutions from the joint calibration from each experiment, and validation runs used the same solutions over a different period within the same catchment. Transfer runs applied the top three solutions from each experiment to the CLP-F catchment. Error bars represent the range in performance of the three solutions. Values of NSE < -1.0 are plotted as -1.0 for visualization purposes.

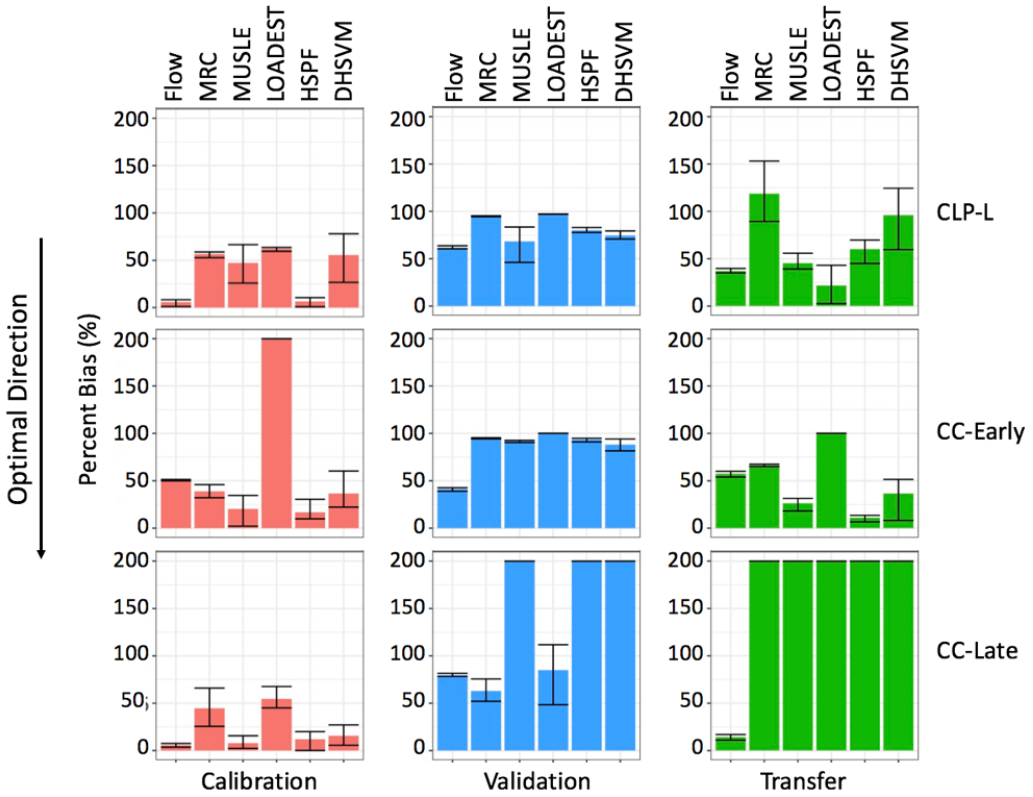

Figure 13. Absolute value of Bias performance for calibration, validation and transfer periods for all algorithms in CLP-L, CC-Early, and CC-Late. Calibration runs used the top three solutions from the joint calibration from each experiment, and validation runs used the same solutions over a different time within the same catchment. Transfer runs applied the top three solutions from each experiment to the CLP-F catchment. Error bars represent the range in performance of the three solutions. Values of |Bias| > 200% are plotted as 200% for visualization purposes.

To determine the role of high SSL magnitudes on objective function performance between the CLP-L calibration and validation periods, we removed the year 1999 from the validation period, as the observed SSL magnitudes exceeded 1,000 tons/day, and was high compared with other years. To compare the ranges in SSL magnitudes in CC-Early and CC-Late, we removed the year 1995 from the CC-Early validation period, as the observed SSL magnitudes exceeded 5,000 tons/day and were high compared with other years. For both removals, the bias of all algorithms decreased, except for LOADEST in CC-Early. We ran a paired t-test for CLP-L before and after the removal of 1999, and followed the same

methodology for 1995 in CC-Early. Both t-tests showed significant differences between the samples, with p-values of 0.03 and 0.02, respectively (Table 6).

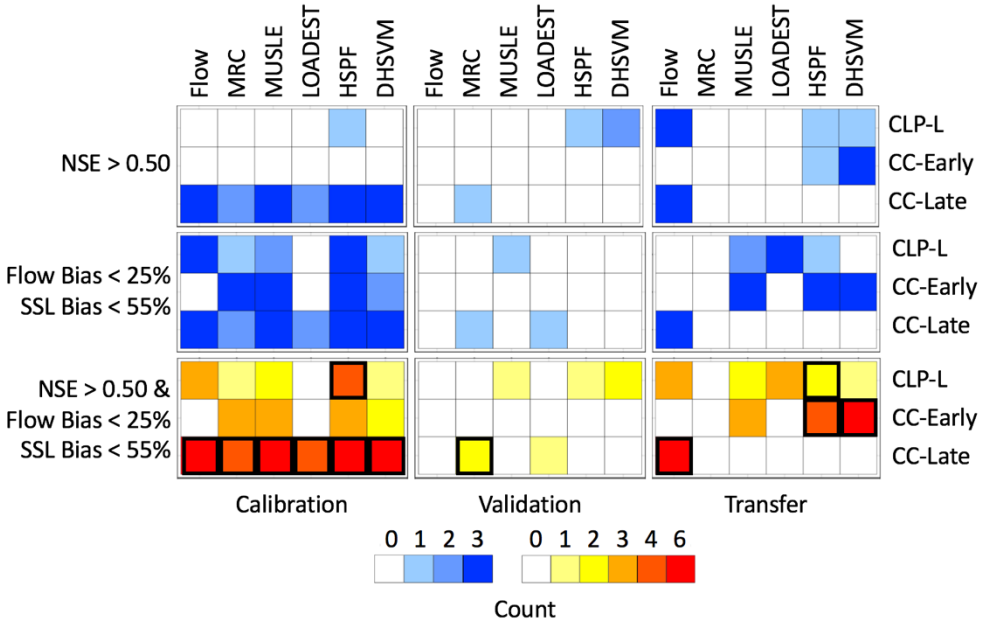

Figure 14. Counts of the top three solutions for each experiment and mode exceeding the performance criteria for NSE, Bias, and NSE and Bias combined. For NSE and Bias individually, a count of 3 indicates that all three solutions were satisfactory, whereas for NSE and Bias combined, a count of 6 indicates that all three solutions were satisfactory in both NSE and Bias. Black boxes indicate an algorithm was satisfactory in both NSE and Bias for at least one solution.

Table 6. Bias estimates for the validation periods of CLP-L and CC-Early with individual years removed to assess the objective performance with and without large SSL. Bias performance increased significantly after the removal of high SSL, with p-values of 0.03 in CLP-L and 0.02 in CC-Early.

	Year	Data Pts Removed / Total	Flow (%)	MRC (%)	MUSLE (%)	LOADEST (%)	HSPF (%)	DHSVM (%)
CLP-L, Validation	1999 Kept	--	60	94	46	97	78	70
	1999 Removed	9 / 74	55	88	15	93	45	29
CC-Early, Validation	1995 Kept	--	39	94	90	100	91	82
	1995 Removed	10 / 39	29	31	0.97	100	29	9.7

4.5 Future Applications

We ran the algorithms over the period of 1950 to 2013 in CC-Late to assess the ability of the ensemble to capture uncertainty within each year and to show long-term variability. We normalized the SSL by the catchment area to represent annual sediment yield. Comparing the sediment yield estimates with historical trends in sediment yield (Jansson, 1988), we saw that the estimates fell within the range of 0 – 100 tons/km^2/year for the Colorado Front Range other than in years with large storm events such as the early 1980s. We demonstrated that the multi-algorithm technique can be used as a forecasting tool by extrapolating the algorithms to periods outside of the calibration and validation periods.

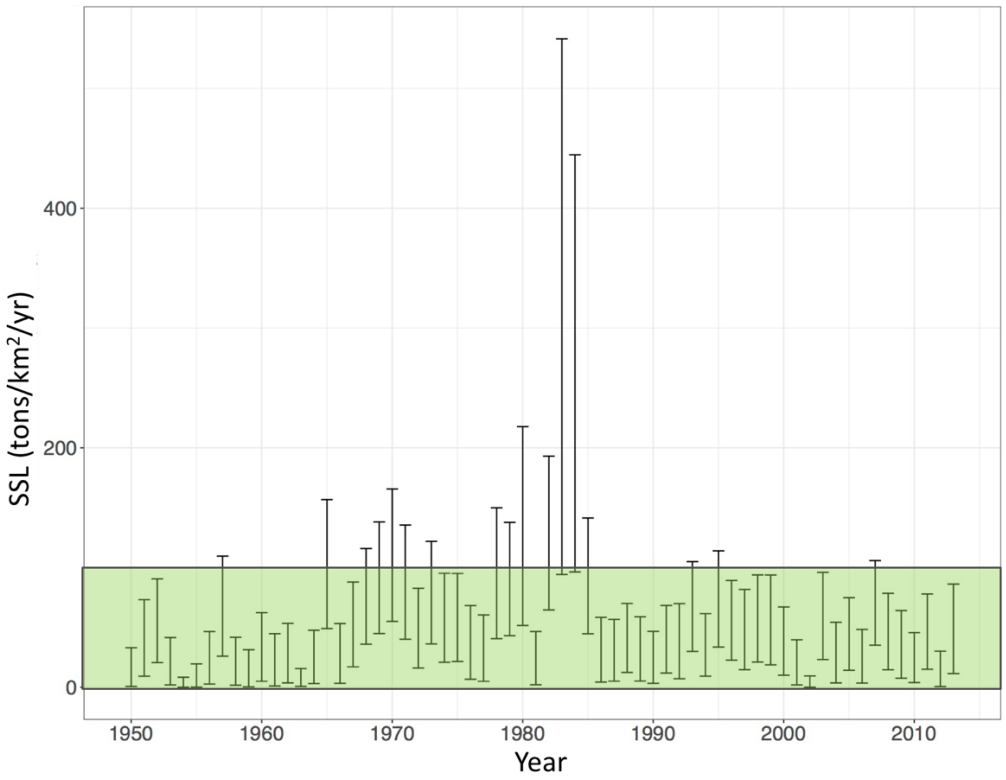

Figure 15. Extension of the model ensemble in CC from 1950 to 2013 using one of the top performing parameter sets from CC-Late. The error bars represent the differences between algorithm estimates, providing a range of uncertainty for each year. The highlighted green box indicates a historical range in sediment yields for the Colorado Front Range visually estimated from maps developed by Jansson (1988).

CHAPTER 5

DISCUSSION, CONCLUSIONS, AND FUTURE WORK

5.1 Discussion

We developed a method for estimating SSL using five algorithms coupled with VIC in three experiments: CLP-L, CC-Early, and CC-Late. The algorithms produced a range in uncertainty across the three modes: calibration, validation, and transfer to CLP-F. Differing performance of each experiment across different modes can be attributed to ranges in SSL observations, static parameterizations, and compromises between preferences for streamflow characteristics among other factors as discussed below. Within each experiment and mode, streamflow and SSL algorithms varied greatly, largely producing no consistent patterns in performance. HSPF and DHSVM reached the satisfactory criteria during the transferability, though, which could indicate a higher level of algorithm flexibility and robustness. We found that the variability in observed SSL magnitudes across calibration, validation, and transfer periods had a large effect on the algorithm performances. In this section, we present a discussion into the performances of the experiments, implications of the study, theoretical and practical limitations. Further, we include future work and conclusions.

5.1.1 Streamflow Performance

Streamflow performance was below the satisfactory criteria from Moriasi et al.

(2007) throughout many of the experiments. This indicates that our joint algorithm calibration did not find optimal solutions, which can be attributed to tradeoffs between streamflow and SSL algorithm partitioning of the hydrograph. Streamflow met the performance criteria during the CC-Late calibration and transfer periods. For CC-Late, the climate and observed annual streamflow between the calibration and validation periods were similar, with ranges in annual streamflow from 1,360 m^3/s – 3,320 m^3/s in calibration and 1,130 m^3/s – 3,260 m^3/s in validation. However, there was a negative bias in the CC-Late validation period, which indicates that the simulated streamflow was greater than the observed streamflow. From Table 5 there were no statistical differences in the observed streamflow and climate between CC-Early and CC-Late. Therefore, we infer that the joint calibration with SSL algorithms could have affected the partitioning of runoff and baseflow, impacting the flexibility and robustness of the streamflow to adapt to different time periods.

CLP-L and CC-Early did not meet the satisfactory performance criteria for streamflow calibration, validation or transfer. This could be because observed streamflow had a high storage ratio upstream of the gauge in CLP-L, which could have impacted the results. Furthermore, in CC-Early we only had SSL data for one of the years of streamflow calibration, and therefore the calibration could have been biased towards that year. However, there were apparent tradeoffs between NSE and Bias for CLP-L in the calibration and transfer periods, where Bias performance was satisfactory in calibration, and NSE performance was satisfactory in transfer. As discussed in more depth below, possible reasons for poor streamflow performance could be due to timing discrepancies between streamflow and SSL peaks. Snowmelt dominated systems are susceptible to hysteresis during spring melts, as noted by Syvitski et al. (2000), having high suspended sediment concentrations on the rising limb of the hydrograph. If the SSL algorithms were improved

by a premature streamflow peak to simulate hysteresis effects, the streamflow performance would therefore be compromised.

5.1.2 Suspended Sediment Load Performances

The magnitude and timing of observed SSL was captured by few algorithms across the experiments. Since streamflow performance was poor under most modes in the joint calibration, the errors could have propagated into SSL estimates as well, thereby causing poor SSL performance.

The MRC was directly impacted by the performance of VIC streamflow, as was the LOADEST, though the LOADEST is also affected by the date of the model simulation. Furthermore, though the MUSLE is linearly dependent on erosion parameters, it also has a nonlinear dependency on total streamflow. In contrast, the HSPF and the DHSVM produce SSL from both precipitation and runoff, as the algorithms incorporate detachment from rainfall and detachment from overland flow.

Of the algorithms, the MRC, the LOADEST, and the MUSLE have a greater reliance on total streamflow performance (combining both runoff and baseflow), making them more susceptible to errors from total streamflow simulations than the HSPF and the DHSVM. Conversely, precipitation rates and VIC partitioning between runoff and baseflow more greatly affect the HSPF and the DHSVM than the MUSLE, the MRC, or the LOADEST.

The HSPF and the DHSVM algorithms had the best performance in the transfer to CLP-F from CLP-L and CC-Early. This could be due to the algorithms' incorporation of both precipitation and runoff (rather than total streamflow). Further, the two algorithms were less correlated with site-specific streamflow and were therefore not as hindered by poor streamflow performance in the transfer process. Additionally, the conceptual and physically-based structure of the HSPF and the DHSVM allowed for diverse

parameterizations and processes, indicating less of a reliance on a single equation generated from empirical data.

Due to the complexity of the HSPF and the DHSVM mechanisms, these algorithms are more robust in their ability to handle various climate and streamflow alterations when compared with the static nature of the MRC and the LOADEST. For example, Glysson et al. (1987) documented how the MRC parameters varied annually within a catchment due to changes in extreme events, implying that one formulation of the MRC is not flexible enough to capture annual differences. Furthermore, applications of the LOADEST are numerically dependent on the dates of the calibration period (Bicknell et al., 1996), thereby affecting transfers to other time periods and limiting flexibility. In contrast, using the DHSVM erosion module, Doten et al. (2006) applied a wildfire scenario to the Rainy Creek catchment, and saw an increase in erosion due to changes in root cohesion and increased surface runoff.

Though the calibration of each SSL algorithm was dependent on streamflow to varying degrees, the calibrations were also affected by the observed SSL magnitudes. CC-Late had the highest SSL magnitudes and the widest range in SSL of the experiments, ranging from ~ 0 tons/day to > 5,000 tons/day. Because we used NSE and Bias as our primary objective functions, the algorithms could have favored the high peak in CC-Late, as under simulation of that peak would cause poor performance. For example, the MRC exponent was ~ 2 times greater in CC-Late than in CLP-L or CC-Early to capture the peak SSL. In the multi-objective optimization for CC-Late, the VIC b_{inf} parameter was optimized towards the highest value, indicating that the system required greater runoff. As higher runoff increases SSL estimates from the MUSLE, the HSPF, and the DHSVM, the optimization could have been generating more runoff to reach the peak SSL in the observed data.

The calibration performance for CC-Late was satisfactory for all algorithms for both NSE and Bias despite the large range in SSL magnitudes during the calibration period. However, we saw a reduction in performance during the validation and transfer periods, which is to be expected. SSL in both the validation and transfer periods was smaller in magnitude than in the calibration periods, and we saw consistent negative Bias in the MUSLE, the HSPF, and the DHSVM for the validation period, and for all SSL modules in the transfer period. A negative Bias implies that the simulated SSL was greater than the observed SSL, as our Bias calculation subtracted simulated from observed. The parameter sets were overestimating runoff and SSL in the validation and transfer periods, as the CC-Late calibration was biased towards high runoff and therefore high SSL peaks.

5.1.3 Implications of this Study

By comparing multiple sites across different time periods, we explored the flexibility of SSL algorithms. A key outcome from this study is the comparison of multiple sediment algorithms within a consistent modeling framework. We added well-known hillslope erosion algorithms into VIC similar to the VIC-WEPP model developed by Mao et al. (2010), and also applied a quantitative multi-model comparison akin to Jetten et al. (1999). We improved SSL predictions at large-scales by combining the incorporation of erosion estimates into VIC with the use of a multi-algorithm routine and multi-objective optimization.

Transferability of parameter sets from a neighboring catchment was explored, in addition to transfer from a sub-catchment into a larger catchment, with the finding that the HSPF and the DHSVM produced satisfactory results in CC-Early and HSPF in CLP-L during the transfer mode. This highlights the potential for transferring calibrated parameter sets to ungauged catchments with similar climatic and land use characteristics.

Future work could also test transferability across climate and land cover gradients.

We found little correlation between calibration, validation and transfer performance. However, by analyzing ranges in the observed SSL data, we suspect the cause being differences in observed magnitudes across time periods and catchments. In experiments with similar ranges of SSL across calibration, validation and transfer, the algorithms performed better. The multi-algorithm routine, therefore, did not validate well over a high SSL period if the calibration was over a low SSL period, and we found the converse to be true as well. This finding has implications for predictions of future climate scenarios (e.g. flooding, droughts, and wildfires) as SSL is known to increase non-linearly with extreme events. Therefore, the most flexible and robust algorithms, the HSPF and the DHSVM, would likely fair better due to their physical inclusion of precipitation rates and vegetation coverage, rather than reliance purely on streamflow as in the MRC.

5.1.4 Theoretical and Practical Limitations

Due to the complexity of erosion and sediment transport processes, we identify the following limitations. First, physical erosion occurs on the order of meters, whereas our implementation of erosion equations is on the order of kilometers. Therefore, our grid cells assume a high degree of landscape homogeneity, inherently underestimating fine scale erosion controls such as channel geometry and heterogeneous soil characteristics.

Second, we assumed SSL was supplied from rainfall and overland flow detachment on hillslopes in the MUSLE, the HSPF and the DHSVM, whereas we ignored the contribution from bedload or mass wasting events. Though the contribution of bedload to suspended load has been found to be smaller than hillslope erosion in mountainous regions, we acknowledge that our algorithms are inherently underestimating total SSL loads.

Third, a tradeoff was evident in streamflow performance between the MUSLE, the

HSPF and the DHSVM for both the individual and joint calibrations. Though it was expected that the performance of streamflow and SSL should align, our model makes a key assumption of an infinite sediment supply, which affects peak timing and performance. In catchments where hysteresis occurs (meaning the peak in SSL comes before the peak in streamflow), the system is often supply limited. However, our infinite supply assumption causes our algorithms to rely on streamflow timing for peak SSL. To compensate for this timing issue, an accounting of sediment supply limitation should be imposed, thereby allowing the peak SSL timing to occur asynchronously with streamflow. However, in the current setting, the streamflow and SSL performances exhibited a tradeoff.

Importantly, the HSPF and the DHSVM algorithms were affected by the partitioning of streamflow into runoff and baseflow. To compensate for a lack of sediment supply from processes such as mass wasting or bedload transport, the HSPF and the DHSVM could compensate by increasing the portion of streamflow that comes from overland flow. However, from our multi-objective optimization and filtering method, we selected solutions that had a balance between the algorithms and the objectives, thereby minimizing the biases from an algorithm's preferences for streamflow partitioning and peak timing. Further, it would be difficult to get reasonable streamflow performance if the HSPF and the DHSVM had extensively altered the hydrograph separation into runoff and baseflow.

Fourth, our joint algorithm calibration method was lacking diversity in objective functions, though we adequately calibrated the NSE and Bias performance of streamflow, the MUSLE, the HSPF and the DHSVM using Borg. The method was more of a multi-model, double objective calibration, though we used eight objective functions in the joint calibration. Our selection of the number and type of objective functions therefore could have impacted our results. We also acknowledge a discrepancy in the manner for which we

calibrated the algorithms, though the MRC and the LOADEST are commonly formulated through statistical regression. However, the MRC and the LOADEST estimate SSL directly, whereas the MUSLE, the HSPF, and the DHSVM estimate grid cell erosion and require routing.

Fifth, the individual calibrations had better overall performance than the joint calibrations. Though this could be an artifact of our parameter and objective function selection, it also implies that there are tradeoffs between the algorithms. Therefore, to get maximum performance from an individual module, one could argue that an individual calibration with more objective functions is preferred. However, this would limit a key novelty of the multi-model technique, allowing for exploration of algorithm structure with consistent settings. We can also directly compare algorithm outputs to see how intertwined they are with streamflow performance, and to identify patterns among the algorithms through application to validation periods and a transfer catchment. Importantly, the multi-algorithm setup allows us to develop ranges of uncertainty from a single VIC simulation.

5.2 Conclusions

We developed a multi-algorithm routine for estimating SSL in mountainous regions from empirical, stochastic, conceptual and physically-based methods. We coupled five erosion and sediment transport algorithms with the VIC model at a resolution of 1/16˚. Using Borg, we performed multi-objective optimizations with individual and joint algorithm calibrations to identify optimal parameter sets in three experiments using two catchments over three time periods. Using the top solutions from each calibration, we applied the parameter sets to validation periods within the same catchments, and to transfer periods within a neighboring catchment.

Algorithm performance varied across the experiments and periods. We found that the more complex algorithms, the HSPF and the DHSVM, were best able to perform well in the transfer, though further testing of the algorithms in a greater number of catchments is required. We conclude that while we generated satisfactory results among the various experiments and modes, our multi-algorithm method was unable to calibrate over periods of low SSL and validate over periods of high SSL and vice versa. Future work will incorporate a greater number of catchments and more diverse SSL data to more broadly evaluate performance of the method.

5.3 Future Work

This study analyzed SSL and streamflow performance in three catchments across varying time periods in the Colorado Front Range. However, to more greatly assess trends in algorithm performance across spatial and temporal scales, future work will include expanding the experiments to the Western United States across varying climate, topography and land use regimes. The SSL algorithms will be evaluated against long-term sediment records from reservoirs, providing larger sample sizes with greater variability. Future work will address limitations in the current study by incorporating a greater number of objective functions, VIC soil parameters, and catchments.

References

Aalto, R., Dunne, T., & Guyot, J. L. (2006). Geomorphic Controls on Andean Denudation Rates. *The Journal of Geology*, *114*(1), 85–99. https://doi.org/10.1086/498101

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Aksoy, H., & Kavvas, M. L. (2005). A review of hillslope and watershed scale erosion and sediment transport models. *Catena*, *64*(2), 247–271.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large Area Hydrologic Modeling and Assessment Part I: Model Development1. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Arnold, J. G., Williams, J. R., & Maidment, D. R. (1995). Continuous-Time Water and Sediment-Routing Model for Large Basins. *Journal of Hydraulic Engineering*, *121*(2), 171–183. https://doi.org/10.1061/(ASCE)0733-9429(1995)121:2(171)

Benda, L., Hassan, M. A., Church, M., & May, C. L. (2005). Geomorphology of Steepland Headwaters: The Transition from Hillslopes to Channels1. *Journal of the American Water Resources Association; Middleburg*, *41*(4), 835–851.

Bicknell, B. R., Imhoff, J. C., Kittle Jr, J. L., Donigian Jr, A. S., & Johanson, R. C. (1996). Hydrological simulation program-FORTRAN. user's manual for release 11. *US EPA*. Retrieved from http://sdi.odu.edu/mbin/hspf/dos/hspf_v11_entirety.pdf

Bracken, L. J., Turnbull, L., Wainwright, J., & Bogaart, P. (2015). Sediment connectivity: a framework for understanding sediment transfer at multiple scales. *Earth Surface Processes and Landforms*, *40*(2), 177–188.

Bryan, R. B. (2000). Soil erodibility and processes of water erosion on hillslope. *Geomorphology*, *32*(3–4), 385–415. https://doi.org/10.1016/S0169-555X(99)00105-1

Burton, A., & Bathurst, J. C. (1998). Physically based modelling of shallow landslide sediment yield at a catchment scale. *Environmental Geology*, *35*(2–3), 89–99. https://doi.org/10.1007/s002540050296

De Vente, J., & Poesen, J. (2005). Predicting soil erosion and sediment yield at the basin scale: Scale issues and semi-quantitative models. *Earth-Science Reviews*, *71*(1–2), 95–125. https://doi.org/10.1016/j.earscirev.2005.02.002

Delpla, I., Jung, A.-V., Baures, E., Clement, M., & Thomas, O. (2009). Impacts of climate change on surface water quality in relation to drinking water production. *Environment International*, *35*(8), 1225–1233. https://doi.org/10.1016/j.envint.2009.07.001

Demaria, E. M., Nijssen, B., & Wagener, T. (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research: Atmospheres*, *112*(D11), D11113. https://doi.org/10.1029/2006JD007534

Donigian, A. S., & Love, J. T. (2003). Sediment calibration procedures and guidelines for watershed modeling. *Proceedings of the Water Environment Federation*, *2003*(4), 728–747.

Doten, C. O., Bowling, L. C., Lanini, J. S., Maurer, E. P., & Lettenmaier, D. P. (2006). A spatially distributed model for the dynamic prediction of sediment erosion and transport in mountainous forested watersheds. *Water Resources Research*, *42*(4), W04417. https://doi.org/10.1029/2004WR003829

Epema, G. F., & Riezebos, H. T. (1987). Fall velocity of waterdrops at different heights as a factor influencing erosivity of simulated rain. Retrieved from http://library.wur.nl/WebQuery/wurpubs/4251

Falcone, J. A. (2011). *GAGES-II: Geospatial attributes of gages for evaluating streamflow*. US Geological Survey. Retrieved from https://pubs.er.usgs.gov/publication/70046617

Ferro, V., & Porto, P. (2000). Sediment Delivery Distributed (SEDD) Model. *Journal of Hydrologic Engineering*, *5*(4), 411–422. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:4(411)

Fryirs, K. A., Brierley, G. J., Preston, N. J., & Kasai, M. (2007). Buffers, barriers and blankets: The (dis)connectivity of catchment-scale sediment cascades. *CATENA*, *70*(1), 49–67. https://doi.org/10.1016/j.catena.2006.07.007

Garcia, M., & Parker, G. (1991). Entrainment of Bed Sediment into Suspension. *Journal of Hydraulic Engineering*, *117*(4), 414–435. https://doi.org/10.1061/(ASCE)0733-9429(1991)117:4(414)

Gartner, J. E., Cannon, S. H., Helsel, D. R., & Bandurraga, M. (2009). *Multivariate statistical models for predicting sediment yields from Southern California watersheds*. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.593.192&rep=rep1&type=pdf

Gibbs, R. J., Matthews, M. D., & Link, D. A. (1971). The relationship between sphere size and settling velocity. *Journal of Sedimentary Research*, *41*(1), 7–18.

Glysson, G. D. (1987). *Sediment-transport curves* (USGS Numbered Series No. 87–218). U.S. Geological Survey,. Retrieved from http://pubs.er.usgs.gov/publication/ofr87218

Goode, J. R., Luce, C. H., & Buffington, J. M. (2012). Enhanced sediment delivery in a changing climate in semi-arid mountain basins: Implications for water resource management and aquatic habitat in the northern Rocky Mountains. *Geomorphology*, *139–140*, 1–15. https://doi.org/10.1016/j.geomorph.2011.06.021

Gordon, E., and Ojima, D. (2015). "Colorado Climate Change Vulnerability Study. A report submitted to the Colorado Energy Office." Western Water Assessment and Colorado State University.

Gray, J. R., & Simões, F. J. (2008). Appendix D Estimating Sediment Discharge. Retrieved from http://ascelibrary.org/doi/abs/10.1061/9780784408148.apd

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97WR03495

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *Journal of Hydrologic Engineering*, *4*(2), 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)

Hadka, D., & Reed, P. (2012). Borg: An Auto-Adaptive Many-Objective Evolutionary Computing Framework. *Evolutionary Computation*, *21*(2), 231–259. https://doi.org/10.1162/EVCO_a_00075

Hamman, J.J., Nijssen, B., Roberts, A., Osinski, R., Lettenmaier D.P., and Maslowski., W. (in press). *Land - Ocean Coupling in the Regional Arctic System*.

Helsel, D. R., & Hirsch, R. M. (2002). *Statistical methods in water resources: US Geological Survey Techniques of Water Resources Investigations, book 4, chap*. A3.

Jansson, M. B. (1988). A Global Survey of Sediment Yield. *Geografiska Annaler. Series A, Physical Geography*, *70*(1/2), 81–98. https://doi.org/10.2307/521127

Jetten, V., de Roo, A., & Favis-Mortlock, D. (1999). Evaluation of field-scale and catchment-scale soil erosion models. *CATENA*, *37*(3–4), 521–541. https://doi.org/10.1016/S0341-8162(99)00037-5

Johanson, R. C., & Davis, H. H. (1980). *Users manual for hydrological simulation program-Fortran (HSPF)* (Vol. 80). Environmental Research Laboratory, Office of Research and Development, US Environmental Protection Agency. Retrieved from https://books.google.com/books?hl=en&lr=&id=cdXuAAAAMAAJ&oi=fnd&pg=PR3&dq=johannson+hspf&ots=o1yFR-olOG&sig=uxbWu5W6BokqktI_sB-siB3sKiU

Kabir, M. A., Dutta, D., & Hironaka, S. (2011). Process-based distributed modeling approach for analysis of sediment dynamics in a river basin. *Hydrology and Earth System Sciences*, *15*(4), 1307–1321.

Kemper, W. D., & Chepil, W. S. (1965). Size distribution of aggregates. *Methods of Soil Analysis. Part 1. Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling*, (methodsofsoilana), 499–510.

Khan, S. J., Deere, D., Leusch, F. D. L., Humpage, A., Jenkins, M., & Cunliffe, D. (2015). Extreme weather events: Should drinking water quality management systems adapt to changing risk profiles? *Water Research*, *85*, 124–136. https://doi.org/10.1016/j.watres.2015.08.018

Knighton, D. (2014). *Fluvial forms and processes: a new perspective*. Routledge.

Kollat, J. B., Reed, P. M., & Wagener, T. (2012). When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resources Research*, *48*(3). Retrieved from http://onlinelibrary.wiley.com/doi/10.1029/2011WR011534/full

Lane, L. J., Hernandez, M., & Nichols, M. (1997). Processes controlling sediment yield from watersheds as functions of spatial scale. *Environmental Modelling & Software*, *12*(4), 355–369. https://doi.org/10.1016/S1364-8152(97)00027-3

Larsen, I. J., Montgomery, D. R., & Greenberg, H. M. (2014). The contribution of mountains to global denudation. *Geology*, *42*(6), 527–530. https://doi.org/10.1130/G35136.1

Laws, J. O., & Parsons, D. A. (1943). The relation of raindrop-size to intensity. *Eos, Transactions American Geophysical Union*, *24*(2), 452–460.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415–14428. https://doi.org/10.1029/94JD00483

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., … Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data*, *2*. https://doi.org/10.1038/sdata.2015.42

Lohmann, D., NOLTE-HOLUBE, R., & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus A*, *48*(5), 708–721.

Maidment, D. R. (1993). *Handbook of hydrology*. New York, McGraw-Hill, Inc.

Mao, D., Cherkauer, K. A., & Flanagan, D. C. (2010). Development of a coupled soil erosion and large-scale hydrology modeling system. *Water Resources Research*, *46*(8). Retrieved from http://onlinelibrary.wiley.com/doi/10.1029/2009WR008268/full

Martin, D. A., & Moody, J. A. (2001). Comparison of soil infiltration rates in burned and unburned mountainous watersheds. *Hydrological Processes*, *15*(15), 2893–2903. https://doi.org/10.1002/hyp.380

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software*, *18*(8), 761–799.

Meybeck, M., Laroche, L., Dürr, H. H., & Syvitski, J. P. M. (2003). Global variability of daily total suspended solids and their fluxes in rivers. *Global and Planetary Change*, *39*(1–2), 65–93. https://doi.org/10.1016/S0921-8181(03)00018-3

Middleton, H. E. (1930). *Properties of soils which influence soil erosion*. US Dept. of Agriculture.

Moore, R. D., & Demuth, M. N. (2001). Mass balance and streamflow variability at Place Glacier, Canada, in relation to recent climate fluctuations. *Hydrological Processes*, *15*(18), 3473–3486. https://doi.org/10.1002/hyp.1030

Morgan, R. P. C. (2009). *Soil Erosion and Conservation*. John Wiley & Sons.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Murdoch, P. S., Baron, J. S., & Miller, T. L. (2000). Potential Effects of Climate Change on Surface-Water Quality in North America1. *JAWRA Journal of the American Water Resources Association*, *36*(2), 347–366. https://doi.org/10.1111/j.1752-1688.2000.tb04273.x

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290.

Nearing, M. A., Foster, G. R., Lane, L. J., & Finkner, S. C. (1989). A process-based soil erosion model for USDA-Water Erosion Prediction Project technology. *Transactions of the ASAE*, *32*(5), 1587–1593.

Nearing, M. A., Pruski, F. F., & O'neal, M. R. (2004). Expected climate change impacts on soil erosion rates: a review. *Journal of Soil and Water Conservation*, *59*(1), 43–50.

Park, S. W., Mitchell, J. K., & Scarborough, J. N. (1982). Soil erosion simulation on small watersheds: a modified ANSWERS model. *Transactions of American Society of Agricultural Engineers*, *25*(6), 1581–1588.

Pechlivanidis, I., Jackson, B., McIntyre, N., & Wheater, H. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *ResearchGate*, *13*(3), 193–214.

Podolak, C. J., & Doyle, M. W. (2015). Reservoir sedimentation and storage capacity in the United States: Management needs for the 21st century. *Journal of Hydraulic Engineering*, *141*(4), 2515001.

Ranzi, R., Le, T. H., & Rulli, M. C. (2012). A RUSLE approach to model suspended sediment load in the Lo river (Vietnam): Effects of reservoirs and land use changes. *Journal of Hydrology*, *422–423*, 17–29. https://doi.org/10.1016/j.jhydrol.2011.12.009

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., Yoder, D. C., & others. (1997). *Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE)* (Vol. 703). US Government Printing Office Washington, DC. Retrieved from https://naldc.nal.usda.gov/naldc/download.xhtml?id=CAT10827029&content=PDF

Rice, S. P., Greenwood, M. T., & Joyce, C. B. (2001). Macroinvertebrate community changes at coarse sediment recruitment points along two gravel bed rivers. *Water Resources Research*, *37*(11), 2793–2803.

Rieke-Zapp, D. H., & Nearing, M. A. (2005). Slope shape effects on erosion. *Soil Science Society of America Journal*, *69*(5), 1463–1471.

Runkel, R. L., Crawford, C. G., & Cohn, T. A. (2004). *Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers*. Retrieved from https://pubs.er.usgs.gov/publication/tm4A5

Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R., & Hauck, L. M. (2001). Validation of the Swat Model on a Large Rwer Basin with Point and Nonpoint Sources1. *JAWRA Journal of the American Water Resources Association*, *37*(5), 1169–1188. https://doi.org/10.1111/j.1752-1688.2001.tb03630.x

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical Modeling of Watershed Hydrology. *Journal of Hydrologic Engineering*, *7*(4), 270–292. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270)

Smith, H. G., Sheridan, G. J., Lane, P. N. J., Nyman, P., & Haydon, S. (2011). Wildfire effects on water quality in forest catchments: A review with implications for water supply. *Journal of Hydrology*, *396*(1–2), 170–192. https://doi.org/10.1016/j.jhydrol.2010.10.043

Smith, R. E., Goodrich, D. C., Woolhiser, D. A., & Unkrich, C. L. (1995). KINEROS–a kinematic runoff and erosion model. *Computer Models of Watershed Hydrology*, *20*, 627–668.

Syvitski, J. P. M., Peckham, S. D., Hilberman, R., & Mulder, T. (2003). Predicting the terrestrial flux of sediment to the global ocean: a planetary perspective. *Sedimentary Geology*, *162*(1–2), 5–24. https://doi.org/10.1016/S0037-0738(03)00232-X

Syvitski, J. P. M., Vörösmarty, C. J., Kettner, A. J., & Green, P. (2005). Impact of Humans on the Flux of Terrestrial Sediment to the Global Coastal Ocean. *Science*, *308*(5720), 376–380. https://doi.org/10.1126/science.1109454

Syvitski, J. P., Morehead, M. D., Bahr, D. B., & Mulder, T. (2000). Estimating fluvial sediment transport: The rating parameters. *Water Resources Research*, *36*(9), 2747–2760. https://doi.org/10.1029/2000WR900133

Tang, Y., Reed, P., van Werkhoven, K., & Wagener, T. (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research*, *43*(6), W06415. https://doi.org/10.1029/2006WR005813

Toy, T. J., Foster, G. R., & Renard, K. G. (2002). *Soil Erosion: Processes, Prediction, Measurement, and Control*. John Wiley & Sons.

Troy, T. J., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. *Water Resources Research*, *44*(9). Retrieved from http://onlinelibrary.wiley.com/doi/10.1029/2007WR006513/full

Turowski, J. M., Rickenmann, D., & Dadson, S. J. (2010). The partitioning of the total sediment load of a river into suspended load and bedload: a review of empirical data. *Sedimentology*, *57*(4), 1126–1146. https://doi.org/10.1111/j.1365-3091.2009.01140.x

van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2009). Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Advances in Water Resources*, *32*(8), 1154–1169. https://doi.org/10.1016/j.advwatres.2009.03.002

Wagener, T., van Werkhoven, K., Reed, P., & Tang, Y. (2009). Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling. *Water Resources Research*, *45*(2), W02501. https://doi.org/10.1029/2008WR007347

Walling, D. E. (1983). The sediment delivery problem. *Journal of Hydrology*, *65*(1), 209–237. https://doi.org/10.1016/0022-1694(83)90217-2

Walling, D. E. (2006). Human impact on land–ocean sediment transfer by the world's rivers. *Geomorphology*, *79*(3–4), 192–216. https://doi.org/10.1016/j.geomorph.2006.06.019

Whitehead, P. G., Wilby, R. L., Battabee, R. W., Kernan, M., & Wade, A. J. (2009). A review of the potential impacts of climate change on surface water quality. *Hydrological Sciences Journal*, *54*(1), 101–123. https://doi.org/10.1623/hysj.54.1.101

Wicks, J. M. (1988). Physically-based mathematical modelling of catchment sediment yield. Retrieved from https://theses.ncl.ac.uk/dspace/handle/10443/152

Wicks, J. M., & Bathurst, J. C. (1996). SHESED: a physically based, distributed erosion and sediment yield component for the SHE hydrological modelling system. *Journal of Hydrology*, *175*(1), 213–238. https://doi.org/10.1016/S0022-1694(96)80012-6

Wigmosta, M. S., Vail, L. W., & Lettenmaier, D. P. (1994). A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, *30*(6), 1665–1679. https://doi.org/10.1029/94WR00436

Wilcock, P. R., & McArdell, B. W. (1993). Surface-based fractional transport rates: Mobilization thresholds and partial transport of a sand-gravel sediment. *Water Resources Research*, *29*(4), 1297–1312. https://doi.org/10.1029/92WR02748

Williams, J. R., & Berndt, H. D. (1977). Sediment yield prediction based on watershed hydrology. *Transactions of the ASAE*, *20*(6), 1100–1104.

Wischmeier, W. H., Smith, D. D., & others. (1960). A universal soil-loss equation to guide conservation farm planning. *Transactions 7th Int. Congr. Soil Sci.*, *1*, 418–425.

Wolman, M. G., & Miller, J. P. (1960). Magnitude and Frequency of Forces in Geomorphic Processes. *The Journal of Geology*, *68*(1), 54–74.

Woolhiser, D. A., Smith, R. E., & Goodrich, D. C. (1990). *KINEROS: a kinematic runoff and erosion model: documentation and user manual*. US Department of Agriculture, Agricultural Research Service Washington^ eD. C DC. Retrieved from https://naldc.nal.usda.gov/naldc/catalog.xhtml?id=47330

Wu, S., Li, J., & Huang, G. (2005). An evaluation of grid size uncertainty in empirical soil loss modeling with digital elevation models. *Environmental Modeling & Assessment*, *10*(1), 33–42. https://doi.org/10.1007/s10666-004-6595-4

Yang, C. T. (2006). Erosion and sedimentation manual. *US Dep. of the Interior, Bureau of Reclamation, Denver, CO*.

Yanto, Livneh, B., Rajagopalan, B., & Kasprzyk, J. (2017). Hydrological model application under data scarcity for multiple watersheds, Java Island, Indonesia. *Journal of Hydrology: Regional Studies*, *9*, 127–139. https://doi.org/10.1016/j.ejrh.2016.09.007

Zhang, W., & Montgomery, D. R. (1994). Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research*, *30*(4), 1019–1028. https://doi.org/10.1029/93WR03553

Zuliziana, S., Tanuma, K., Yoshimura, C., & Saavedra, O. C. (2015). Distributed model of hydrological and sediment transport processes in large river basins in Southeast Asia. *Hydrology and Earth System Sciences Discussions*, *12*, 6755–6797.

Appendix

Appendix A: Analysis of Observed SSL

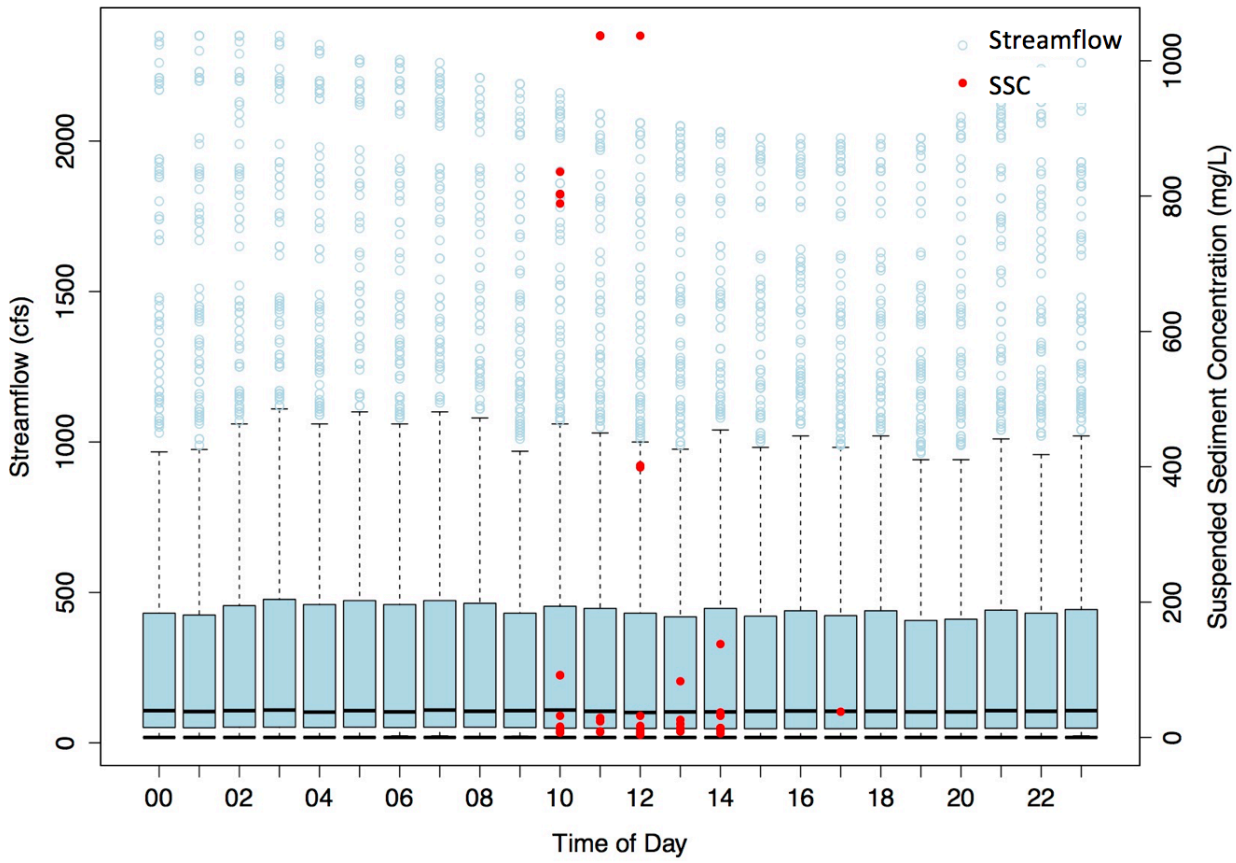

Figure 16. Boxplots of observed streamflow in CC-Late during 1992 – 1997 separated by time of day showing suspended sediment concentration (SSC) sampling times. Streamflow shows a diurnal flux. Therefore, the SSC samples might not be representative of the wave celerity in the catchment, as the samples were not taken during peak streamflow hours.

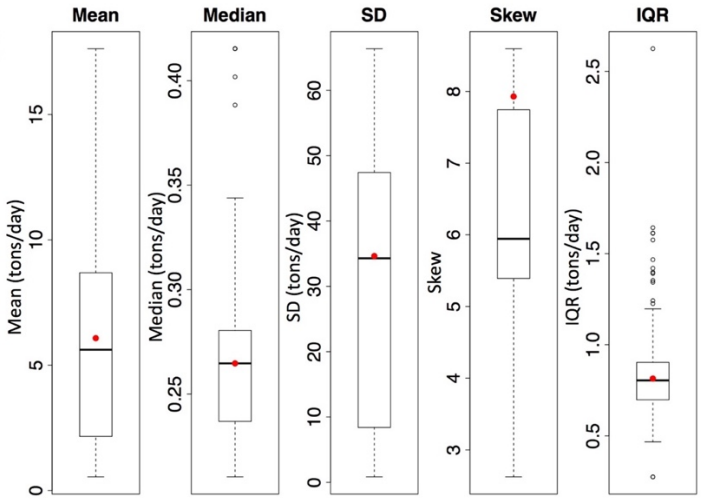

Figure 17. Bootstrap sampling with 150 evaluations of randomly selected samples from the CLP-L calibration period in 1987 – 1993. Results suggest the sample size adequately represents the population, with a slight deviation in skew.

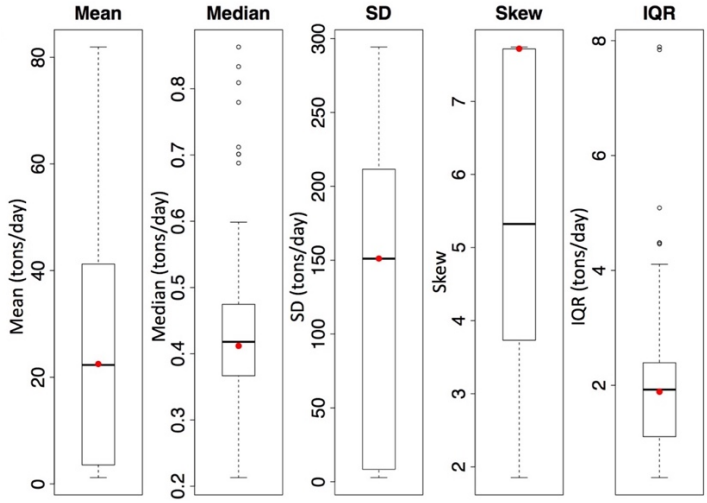

Figure 18. Bootstrap sampling with 150 evaluations of randomly selected samples from the CLP-L validation period in 1993 – 1999. Results suggest the sample size adequately represents the population, with a slight deviation in skew.

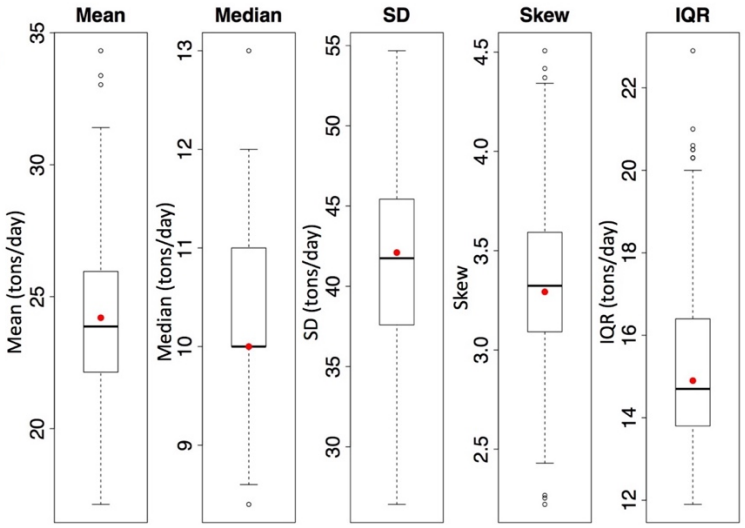

Figure 19. Bootstrap sampling with 150 evaluations of randomly selected samples from the CC-Early calibration period in 1980 – 1985. Results suggest the sample size adequately represents the population.

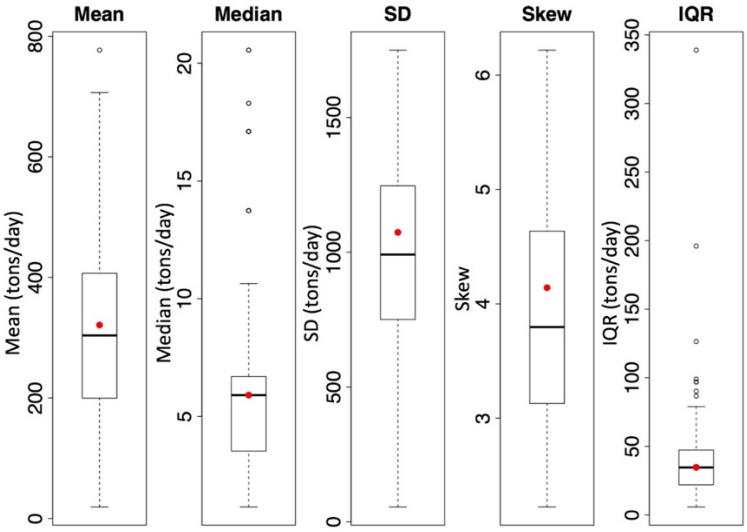

Figure 20. Bootstrap sampling with 150 evaluations of randomly selected samples from the CC-Late calibration period in 1992 – 1997. Results suggest the sample size adequately represents the population.

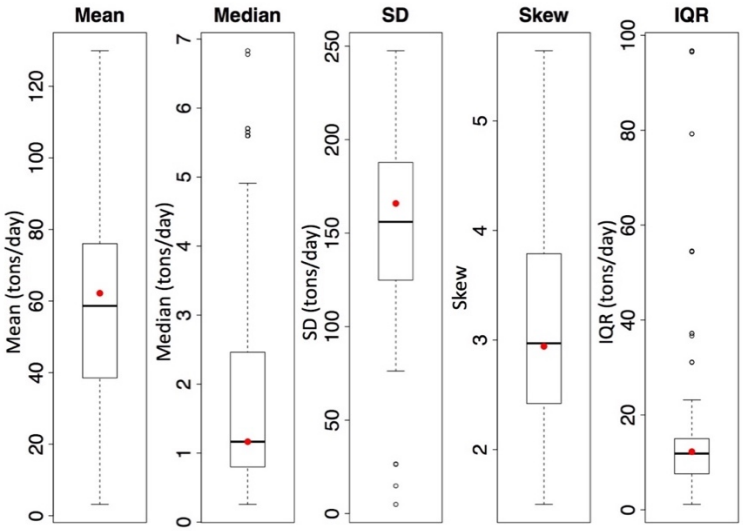

Figure 21. Bootstrap sampling with 150 evaluations of randomly selected samples from the CLP-F transfer period in 1992 – 1997. Results suggest the sample size adequately represents the population.

Appendix B: MRC and LOADEST Calibrations

Table 7. MRC parameter estimation for each experiment using non-linear least squares. Levels of significance were '***' for p-value < 0.001.

Experiment	Parameter	Estimate	Standard Error	t-value	p-value	R^2
CLP-L	a	0.88	0.21	4.3	5.5e-05 ***	0.98
CLP-L	b	1.8	0.075	25.	< 2.0e-16 ***	
CC-Early	a	2.0	0.52	3.9	0.00013 ***	0.64
CC-Early	b	1.5	0.095	15.	< 2.0e-16 ***	
CC-Late	a	0.0080	0.015	0.52	0.60	0.90
CC-Late	b	3.3	0.48	6.9	3.9e-08 ***	

Table 8. LOADEST parameter estimation for each experiment using AMLE with residual variance, PPCC and PPCC significance.

Experiment	a_0	a_1	a_2	a_3	a_4	a_5	a_6	Res. Var	PPCC	Sig	R^2
CLP-L	6.8	1.4	0.09	-0.09	--	--	--	0.65	0.99	6.9e-01	0.80
CC-Early	10.2	1.9	-0.40	2.1	-0.81	-14.5	--	0.26	0.99	3.9e-01	0.84
CC-Late	7.1	1.8	0.12	0.04	0.54	-0.03	0.06	0.82	0.99	3.7e-01	0.89

Appendix C: Individual Algorithm Calibrations

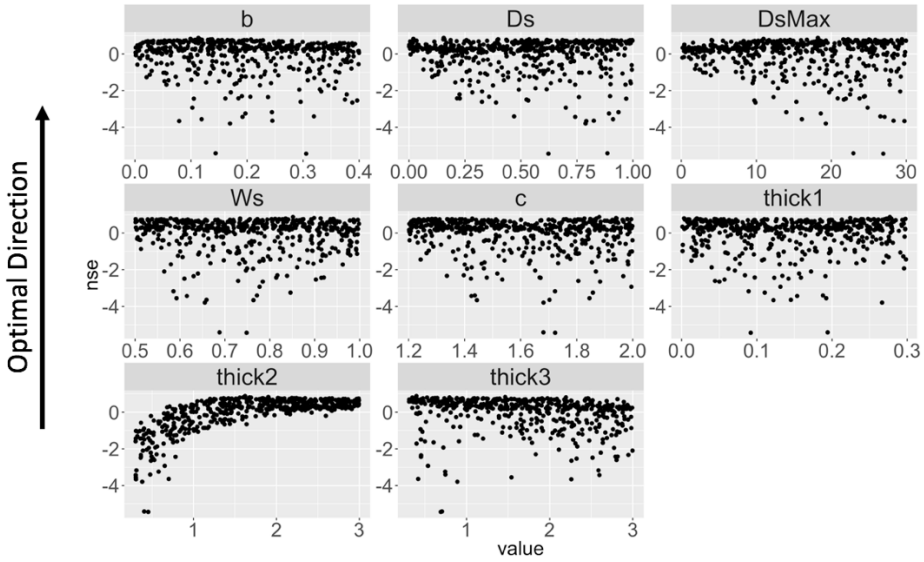

Figure 22. Demonstration of the sensitivity analysis completed in CC using NSE and LHS with 500 function evaluations. Though NSE performance is frequently below the satisfactory criteria, the intention is to show the spread in model performance from the various parameters. We performed similar analyses using correlation, bias and variability.

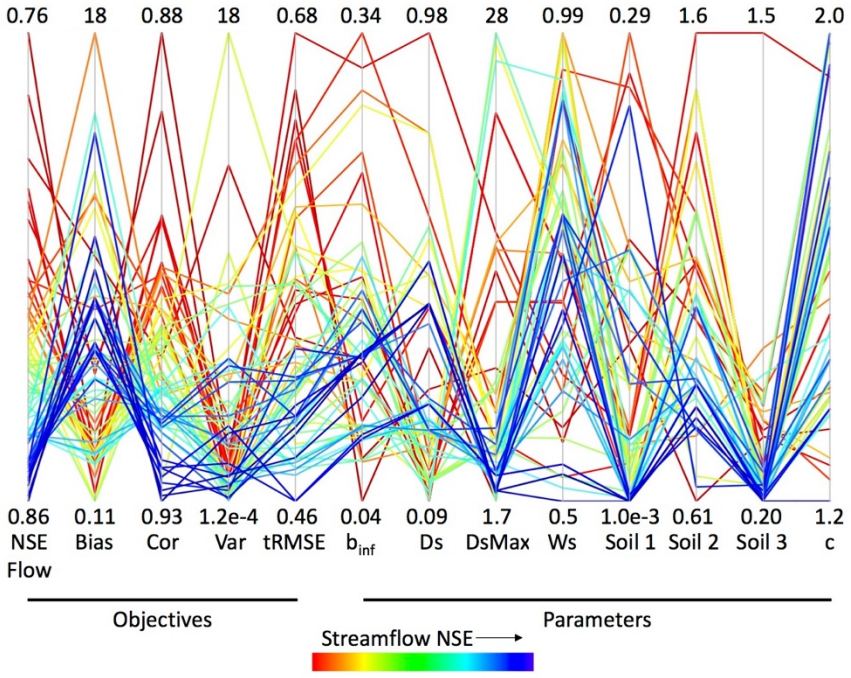

Figure 23. Individual streamflow calibration for CC-Late during 1992 – 1997 with a spin-up period in 1991. There were tradeoffs between the objectives, as well as the values of the VIC soil parameters. Solutions are sorted by streamflow NSE (NSE Flow), and present better performance than in the joint calibration.

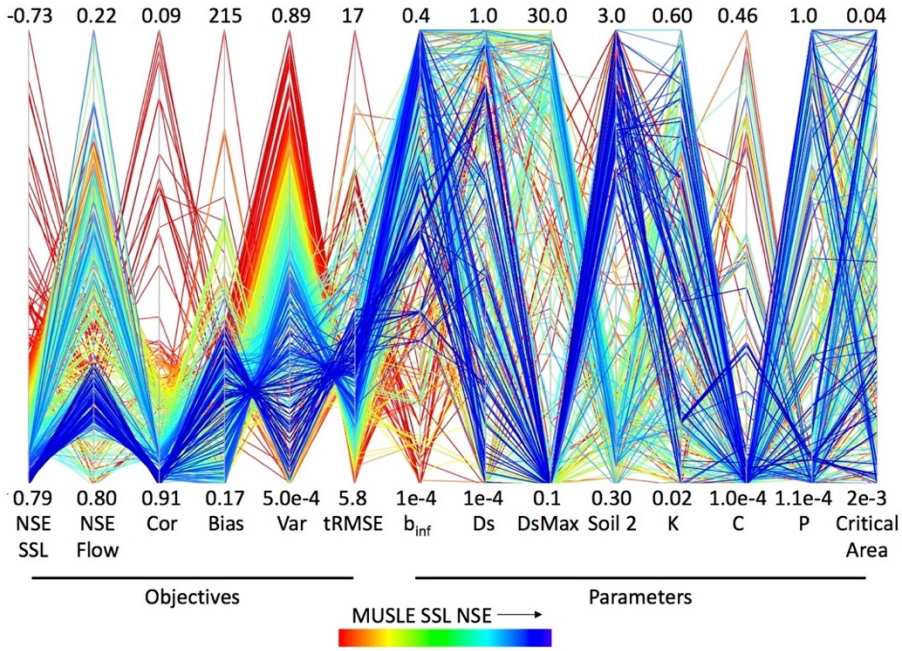

Figure 24. Individual MUSLE SSL algorithm calibration for CC-Late during 1992 – 1997 with a spin-up period in 1991. There were tradeoffs between the objectives, as well as the VIC soil parameters. MUSLE SSL parameters K and P were less sensitive, and covered the parameter bounds, there were tradeoffs with the C parameter. Solutions are sorted by SSL NSE, and present better performance than in the joint calibration.

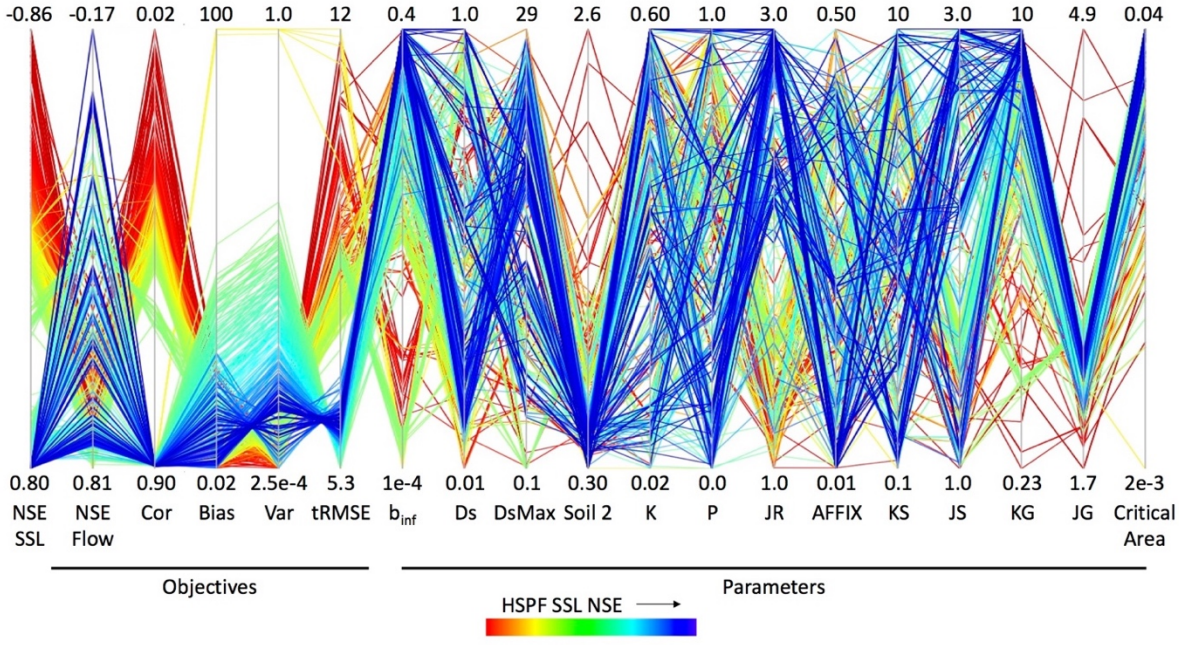

Figure 25. Individual HSPF SSL algorithm calibration for CC-Late during 1992 – 1997 with a spin-up period in 1991. There were tradeoffs between the objectives, as well as the VIC soil parameters. HSPF SSL parameters were generally less sensitive, except for JR, KG and JG. Solutions are sorted by SSL NSE, and present better performance than in the joint calibration.

Appendix D: Joint Algorithm Calibrations

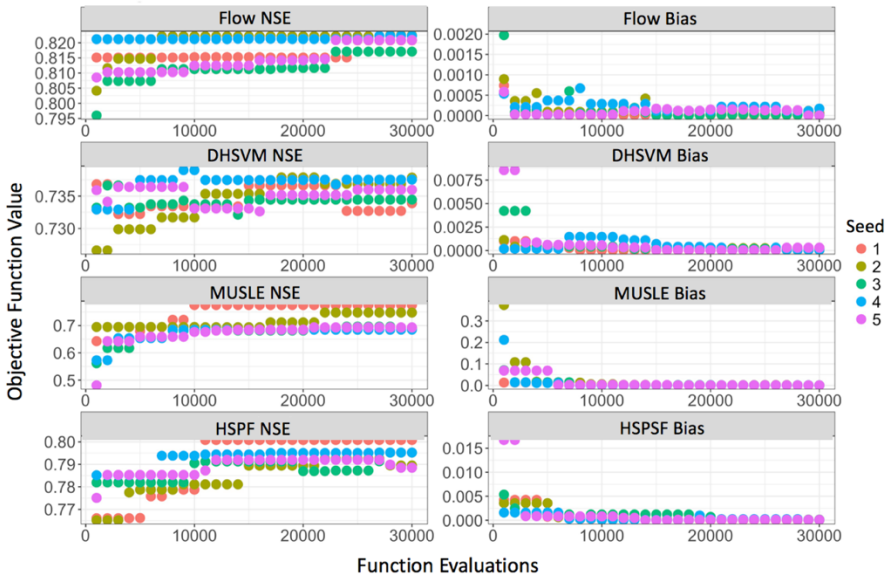

Figure 26. Minimum objective function performance as a measure of the number of Borg function evaluations. The daily NSE and Bias estimates met our satisfactory criteria of NSE > 0.50, and Bias < ±25% for streamflow and Bias < ±55% for SSL.

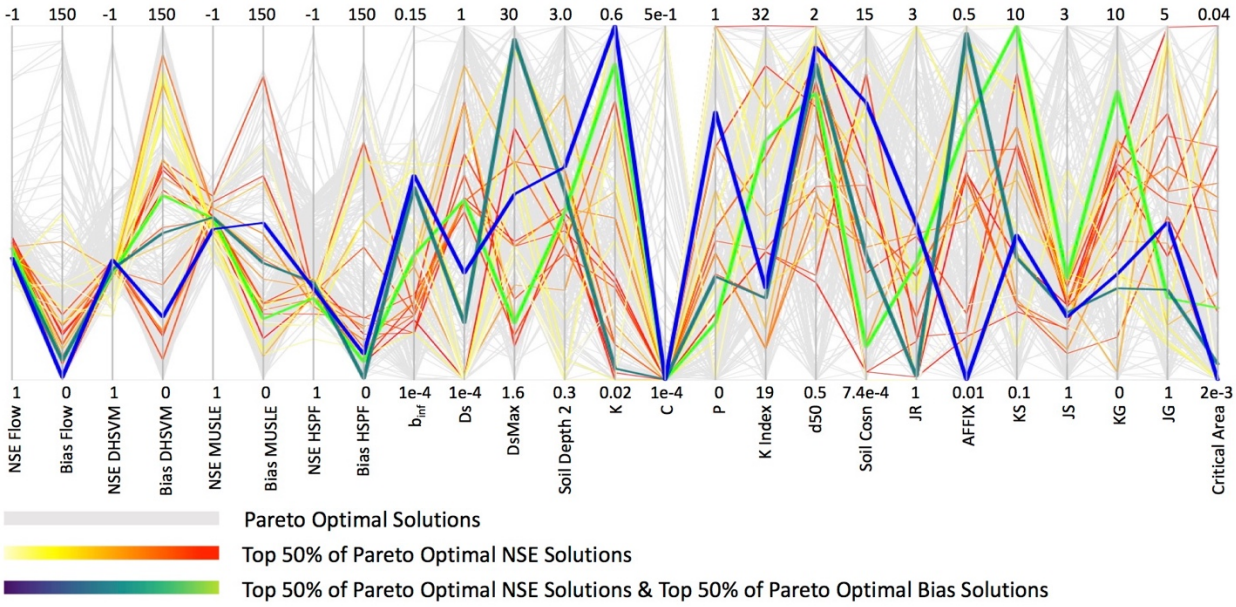

Figure 27. Parallel coordinate plot of Pareto optimal solutions from Borg for CLP-L computed over the calibration period 1987 – 1993 for the joint algorithm calibration with a spin-up year in 1986. Colored solutions indicate the highest performances across the ensemble, with blues and greens representing the final selected solutions. Colored solutions are sorted by Flow NSE. The parameter sets from selected solutions were used to run validation and transferability analyses.

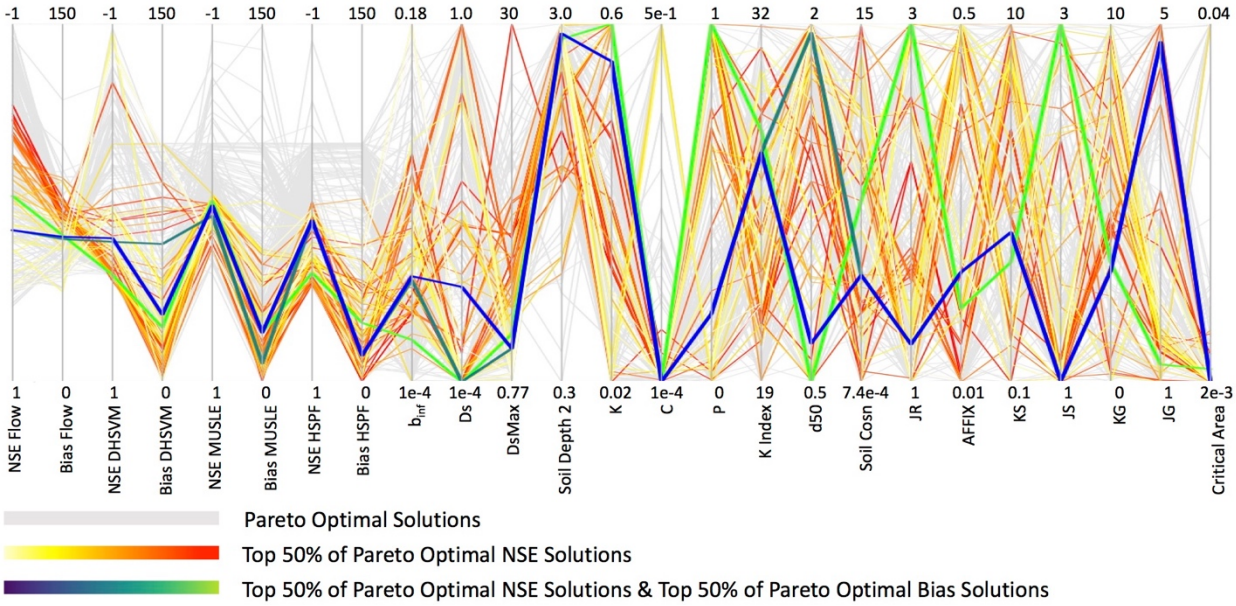

Figure 28. Parallel coordinate plot of Pareto optimal solutions from Borg for CC-Early computed over the calibration period 1980 – 1985 for the joint algorithm calibration with a spin-up year in 1979. Colored solutions indicate the highest performances across the ensemble, with blues and greens representing the final selected solutions. Colored solutions are sorted by Flow NSE. The parameter sets from selected solutions were used to run validation and transferability analyses.

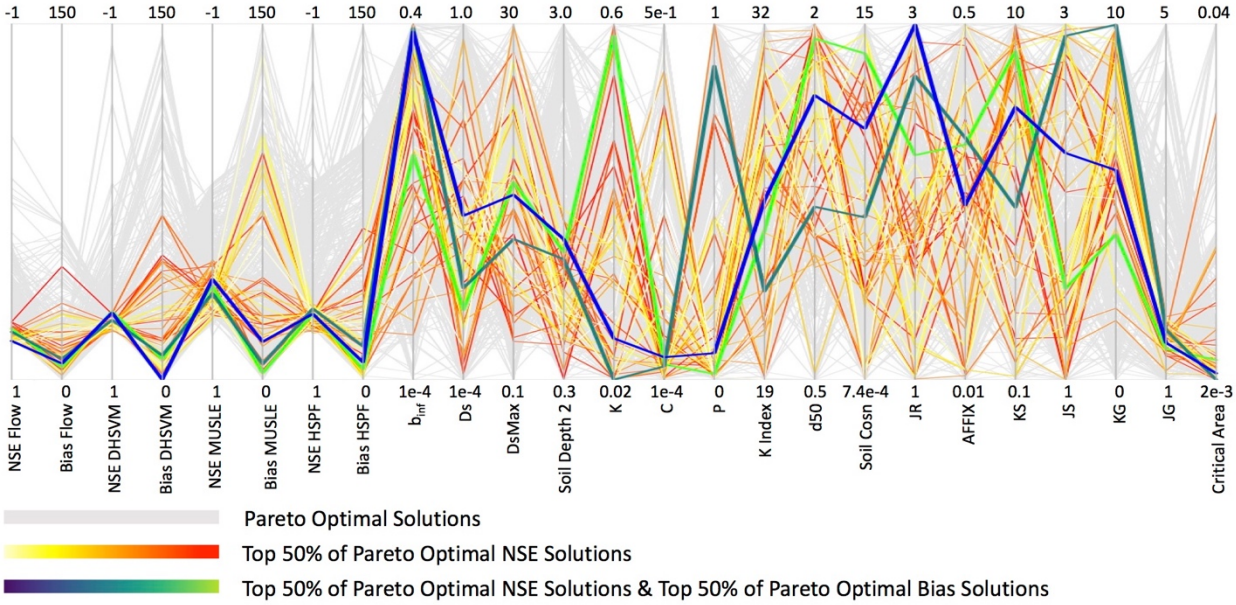

Figure 29. Parallel coordinate plot of Pareto optimal solutions from Borg for CC-Late computed over the calibration period 1992 – 1997 for the joint algorithm calibration with a spin-up year in 1991. Colored solutions indicate the highest performances across the ensemble, with blues and greens representing the final selected solutions. Colored solutions are sorted by Flow NSE. The parameter sets from selected solutions were used to run validation and transferability analyses.

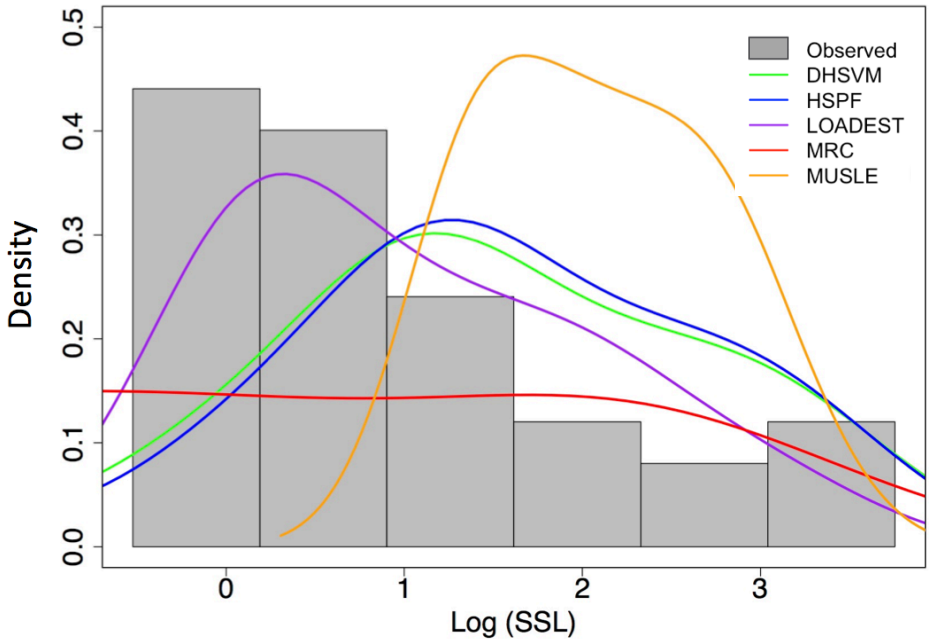

Figure 30. Histogram of observed data for CC-Late with kernel density estimation for each SSL algorithm computed over the calibration period 1992 – 1997 for the joint algorithm calibration with a spin-up year in 1991. The kernel density distributions were calculated from the mean of the SSL values from the top three performing parameter sets.

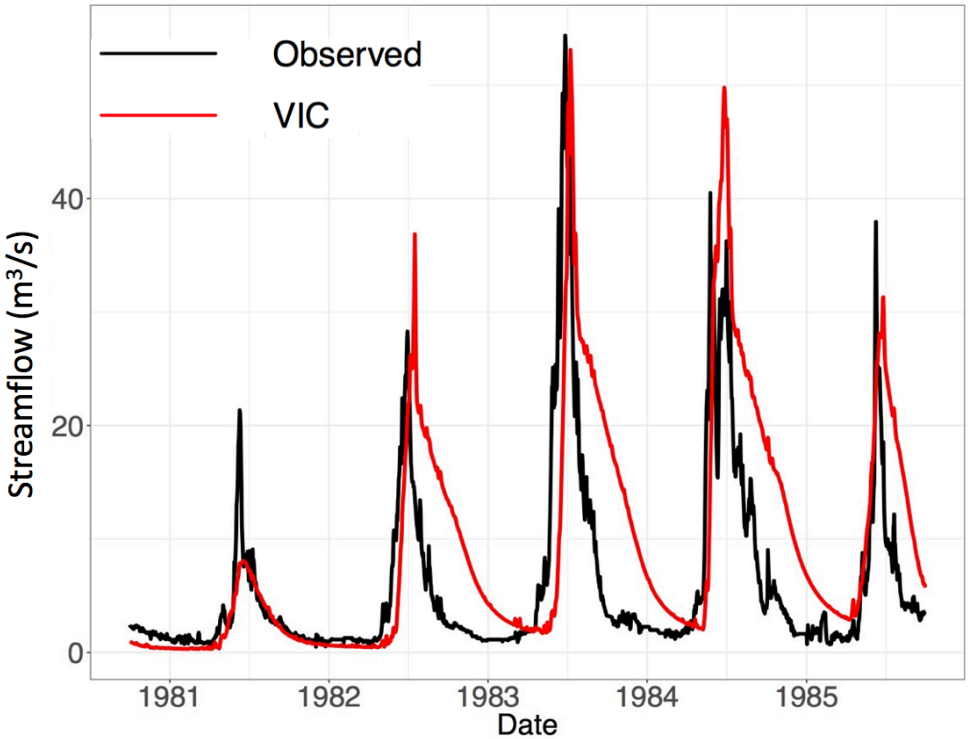

Figure 31. Hydrograph for CC-Early with VIC streamflow computed over the calibration period 1980 – 1985 for the joint algorithm calibration with a spin-up year in 1979. Falling limbs of the hydrograph were overestimated other than in 1981 when SSL was calibrated.

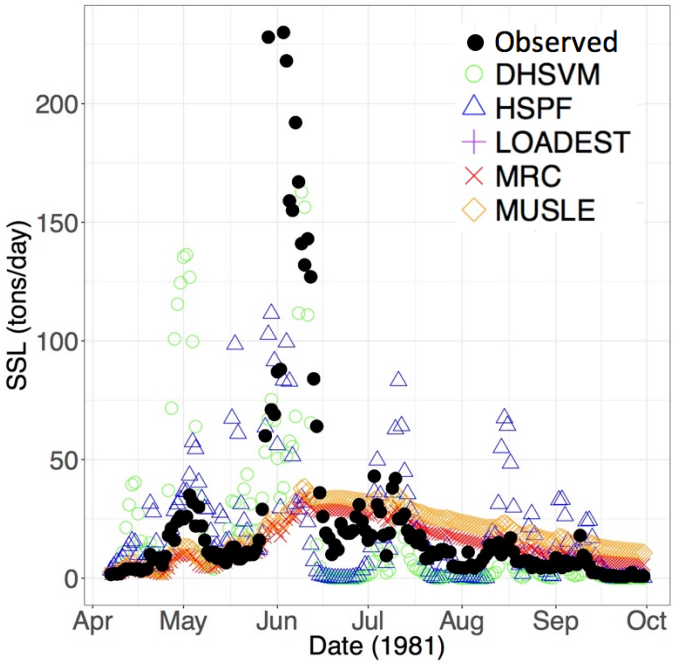

Figure 32. Time series of suspended sediment load (SSL) for CC-Early with each SSL algorithm computed over the calibration period 1980 – 1985 for the joint algorithm calibration with a spin-up year in 1979. The plotted SSL estimates were calculated from the mean of the SSL values from the top three performing parameter sets.

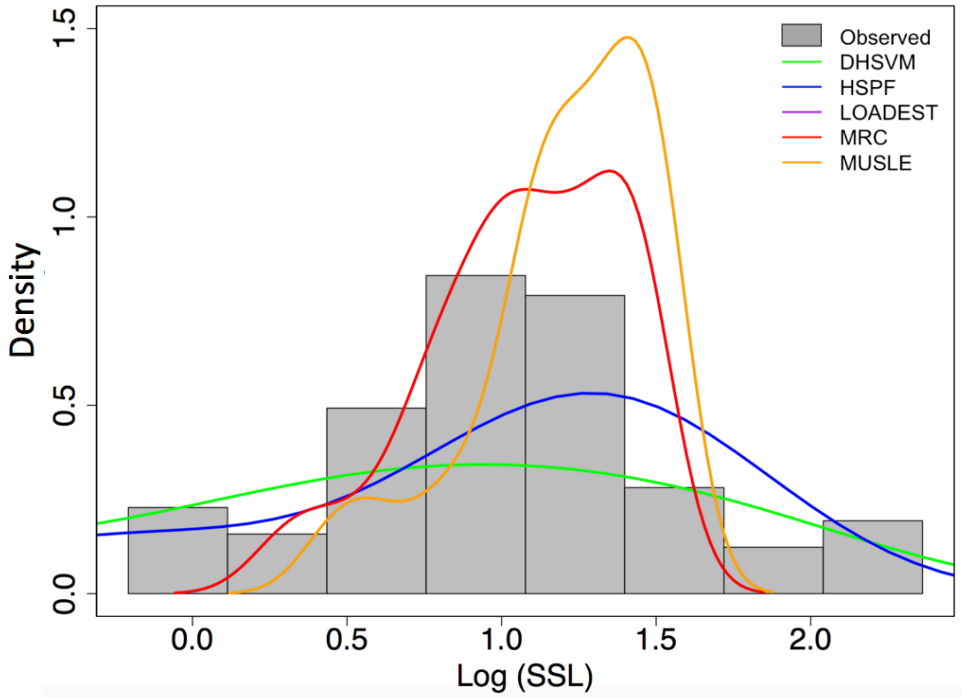

Figure 33. Histogram of observed data for CC-Early with kernel density estimation for each SSL algorithm computed over the calibration period 1980 – 1985 for the joint algorithm calibration with a spin-up year in 1979. The kernel density distributions were calculated from taking the mean of the SSL values from the top three performing parameter sets.

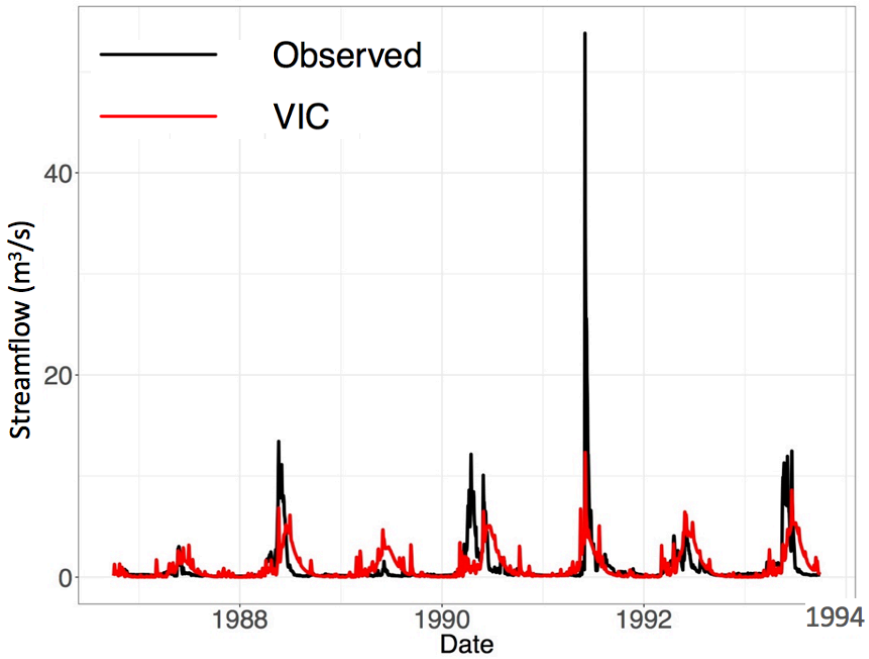

Figure 34. Hydrograph for CLP-L with VIC streamflow computed over the calibration period 1987 – 1993 for the joint algorithm calibration with a spin-up year in 1986. Observed streamflow was impacted by 35% reservoir storage upstream, which could be impacting the results.

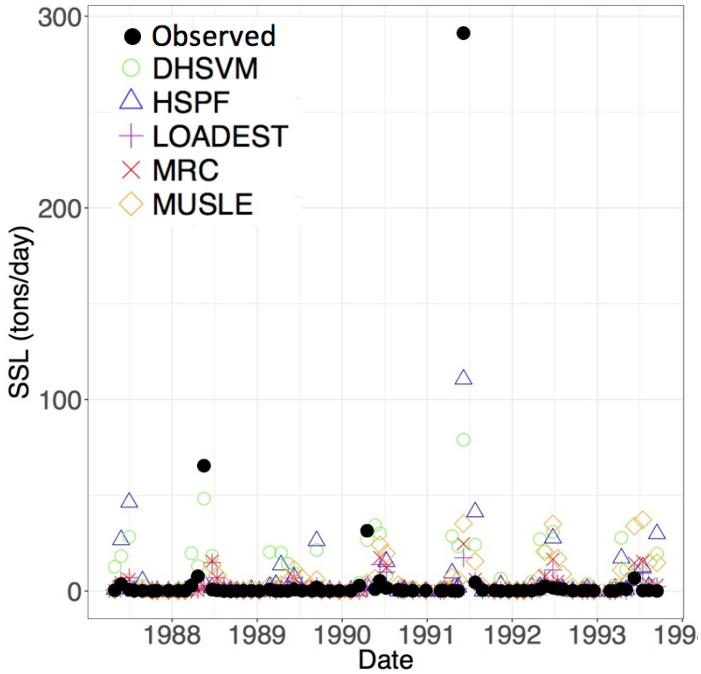

Figure 35. Time series of suspended sediment load (SSL) for CLP-L with each SSL algorithm computed over the calibration period 1987 – 1993 for the joint algorithm calibration with a spin-up year in 1986. The plotted SSL estimates were calculated from the mean of the SSL values from the top three performing parameter sets.

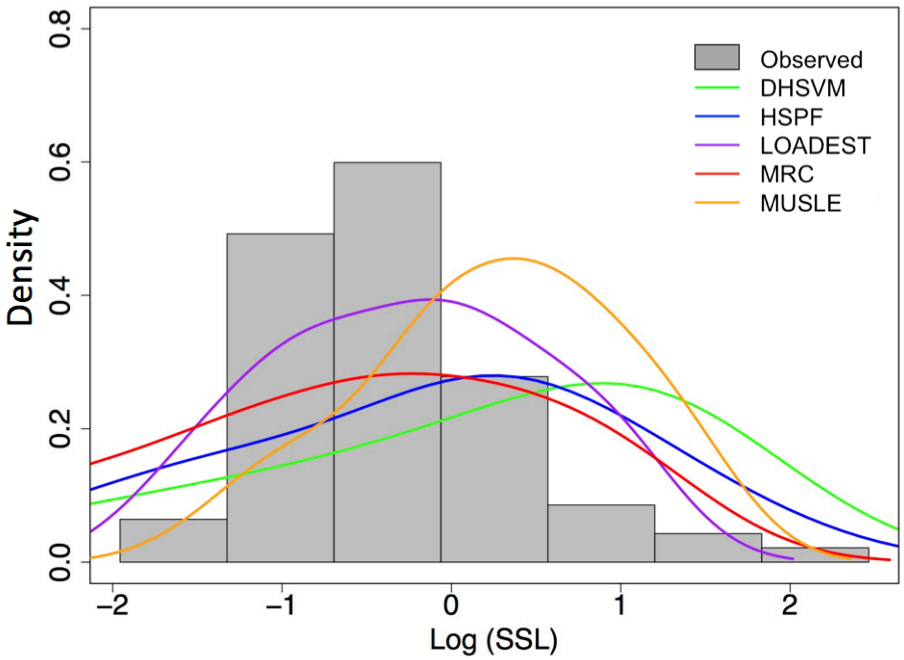

Figure 36. Histogram of observed data for CLP-L with kernel density estimation for each SSL algorithm computed over the calibration period 1987 – 1993 for the joint algorithm calibration with a spin-up year in 1986. The kernel density distributions were calculated from taking the mean of the SSL values from the top three performing parameter sets.

Appendix E: Joint Algorithm Validations and Transfers

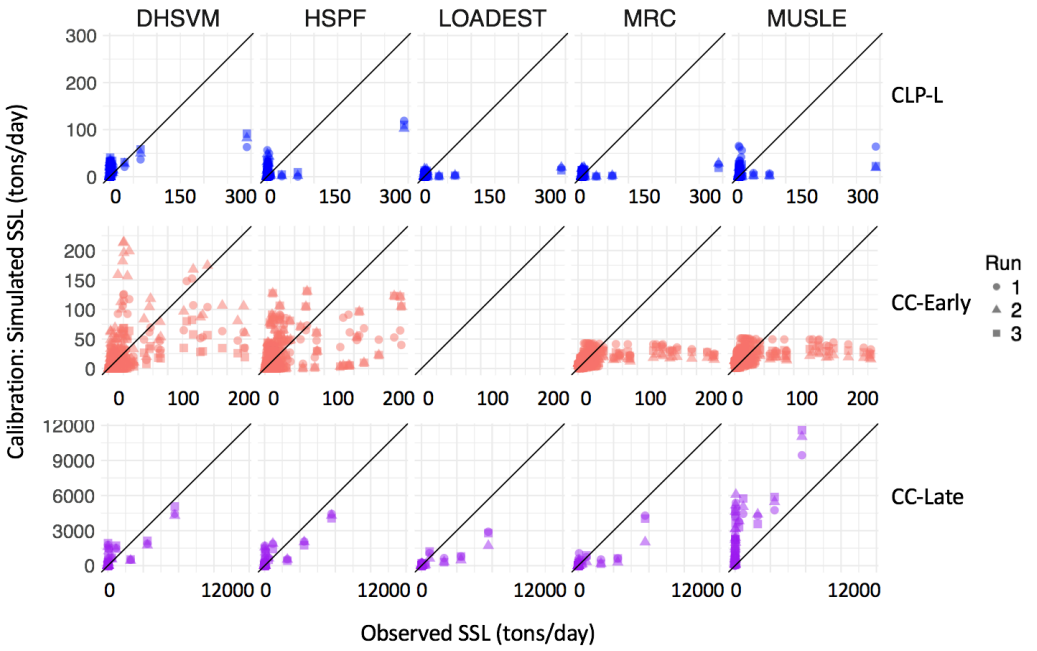

Figure 37. One-to-one plots for calibration periods in CLP-L, CC-Early, and CC-Late for the top three selected parameter sets indicated by shape. CC-Late had the best overall performance in NSE of all modules compared with CLP-L and CC-Early.

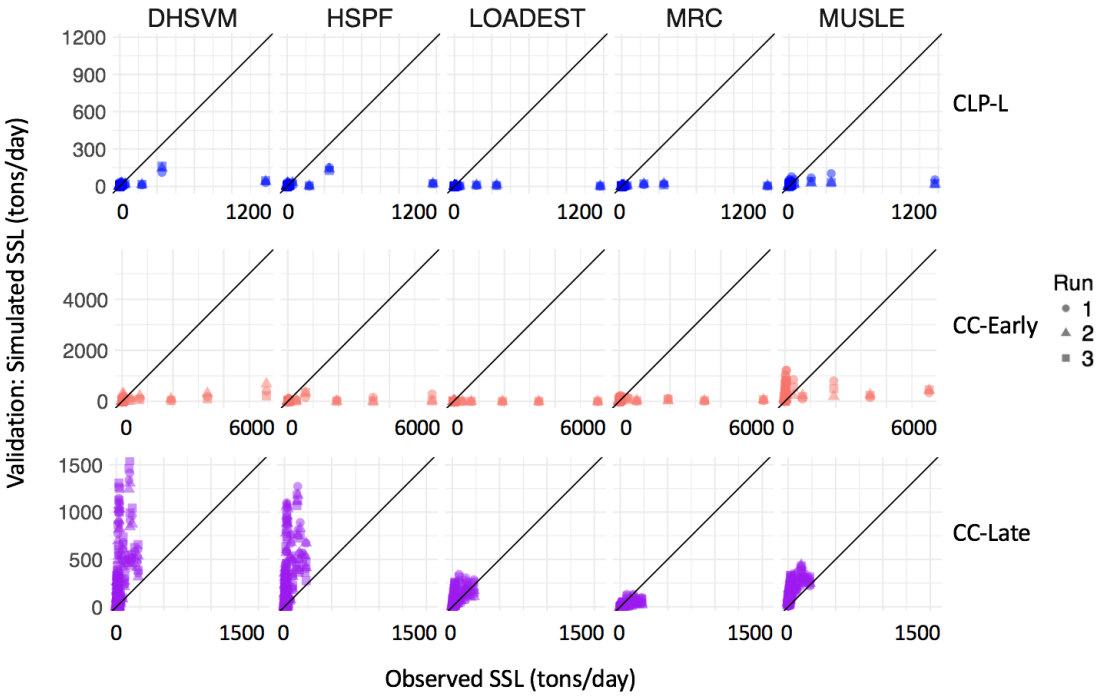

Figure 38. One-to-one plots for validation periods in CLP-L, CC-Early, and CC-Late for the top three selected parameter sets indicated by shape. The best performances in NSE were the DHSVM and HSPF algorithms in CLP-L, and MRC in CC-Late.

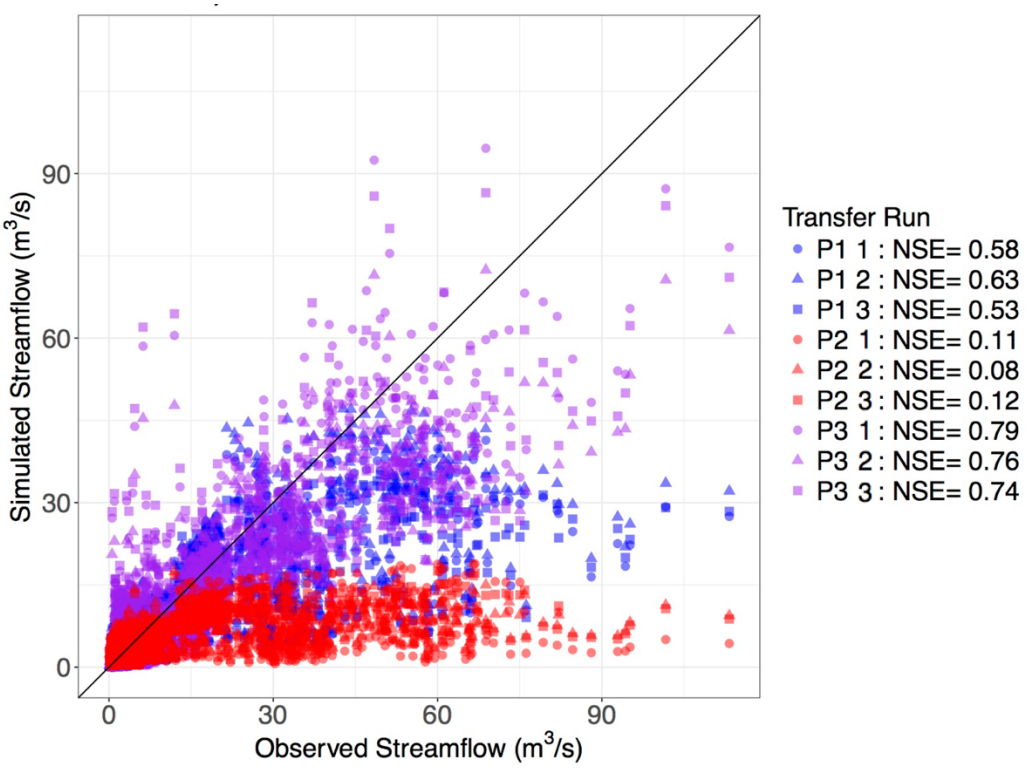

Figure 39. One-to-one plot of simulated versus observed streamflow for the transfer of the top three selected parameter sets indicated by shape from P1: CLP-L, P2: CC-Early, and P3: CC-Late to CLP-F. P1 and P3 met the satisfactory criteria for NSE. P2 performed poorly, potentially due to the joint calibration scheme where SSL and streamflow performances had a tradeoff, and therefore a balance between objectives required a compromise.

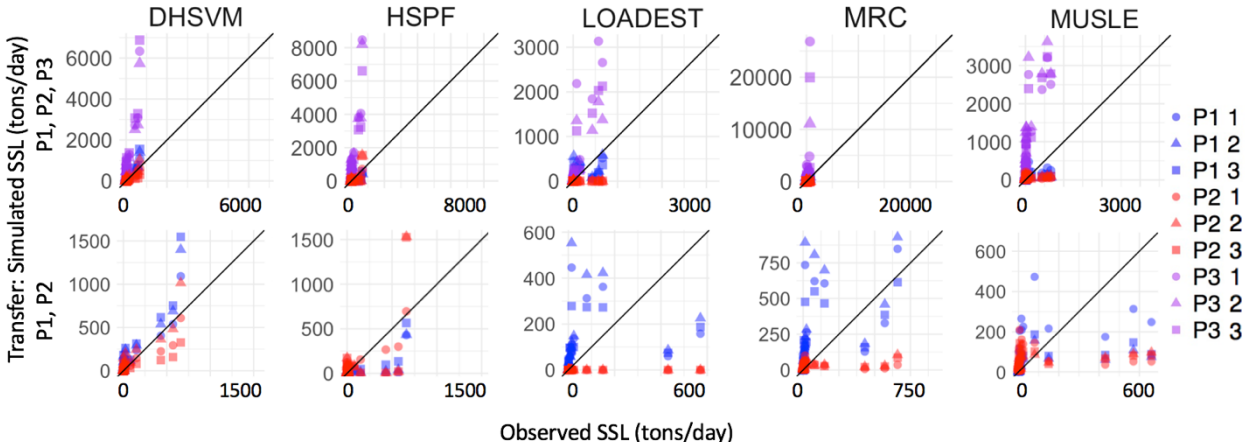

Figure 40. (top) Transferability test for SSL from P1: CLP-L, P2: CC-Early, and P3: CC-Late to CLP-F using the top three calibrated Borg parameter sets for P1, P2 and P3, (bottom) with P3 removed to emphasize performance of P1 and P2.

Table 9. NSE scores for CLP-L calibration, validation, and transfer to CLP-F for all algorithms and the top three selected parameter sets (Runs).

Experiment	Run	Module	Calibration	Validation	Transfer
CLP-L	1	DHSVM	0.32	0.42	0.68
CLP-L	2	DHSVM	0.37	0.53	0.16
CLP-L	3	DHSVM	0.40	0.58	-0.21
CLP-L	1	Flow	0.31	0.23	0.58
CLP-L	2	Flow	0.31	0.27	0.63
CLP-L	3	Flow	0.26	0.22	0.53
CLP-L	1	HSPF	0.48	0.52	0.32
CLP-L	2	HSPF	0.45	0.49	0.38
CLP-L	3	HSPF	0.54	0.46	0.62
CLP-L	1	LOADEST	0.09	0.01	0.24
CLP-L	2	LOADEST	0.09	0.02	0.10
CLP-L	3	LOADEST	0.06	0.00	0.39
CLP-L	1	MRC	0.13	0.08	-0.59
CLP-L	2	MRC	0.14	0.09	-1.26
CLP-L	3	MRC	0.08	0.05	0.06
CLP-L	1	MUSLE	0.15	0.45	0.21
CLP-L	2	MUSLE	0.08	0.15	0.15
CLP-L	3	MUSLE	0.09	0.17	0.21

Table 10. Percent bias scores for CLP-L calibration, validation, and transfer to CLP-F for all algorithms and the top three selected parameter sets (Runs).

Experiment	Run	Module	Calibration	Validation	Transfer
CLP-L	1	DHSVM	-26.6	79.3	-59.6
CLP-L	2	DHSVM	-62.2	73.4	-104.0
CLP-L	3	DHSVM	-78.1	70.8	-124.4
CLP-L	1	Flow	-1.1	63.7	39.7
CLP-L	2	Flow	-8.3	60.7	36.4
CLP-L	3	Flow	-8.2	60.4	35.0
CLP-L	1	HSPF	-10.5	77.8	69.7
CLP-L	2	HSPF	0.8	79.7	66.0
CLP-L	3	HSPF	8.1	82.9	45.0
CLP-L	1	LOADEST	63.7	97.3	-19.7
CLP-L	2	LOADEST	59.6	96.9	-43.1
CLP-L	3	LOADEST	63.5	97.0	-2.6
CLP-L	1	MRC	58.3	95.2	-113.7
CLP-L	2	MRC	52.8	94.4	-153.1
CLP-L	3	MRC	58.9	94.9	-89.4
CLP-L	1	MUSLE	-66.5	46.2	-41.1
CLP-L	2	MUSLE	49.5	83.5	55.9
CLP-L	3	MUSLE	25.8	75.3	39.1

Table 11. NSE scores for CC-Early calibration, validation, and transfer to CLP-F for all algorithms and the top three selected parameter sets (Runs).

Experiment	Run	Module	Calibration	Validation	Transfer
CC-Early	1	DHSVM	0.27	0.04	0.83
CC-Early	2	DHSVM	-0.15	0.13	0.74
CC-Early	3	DHSVM	0.17	-0.02	0.54
CC-Early	1	Flow	0.12	0.40	0.11
CC-Early	2	Flow	0.28	0.16	0.08
CC-Early	3	Flow	0.28	0.37	0.12
CC-Early	1	HSPF	0.28	0.01	0.84
CC-Early	2	HSPF	0.11	-0.09	-0.55
CC-Early	3	HSPF	0.11	-0.09	-0.54
CC-Early	1	LOADEST	-1.1E+86	-0.10	-0.15
CC-Early	2	LOADEST	-4.7E+85	-0.10	-0.15
CC-Early	3	LOADEST	-7.4E+85	-0.10	-0.15
CC-Early	1	MRC	0.14	-0.08	-0.08
CC-Early	2	MRC	0.02	-0.08	0.07
CC-Early	3	MRC	0.13	-0.07	0.03
CC-Early	1	MUSLE	0.15	-0.06	-0.10
CC-Early	2	MUSLE	0.03	-0.06	0.13
CC-Early	3	MUSLE	0.13	-0.05	0.09

Table 12. Percent bias scores for CC-Early calibration, validation, and transfer to CLP-F for all algorithms and the top three selected parameter sets (Runs).

Experiment	Run	Module	Calibration	Validation	Transfer
CC-Early	1	DHSVM	27.3	88.9	8.1
CC-Early	2	DHSVM	-22.2	81.6	-51.5
CC-Early	3	DHSVM	60.2	94.0	50.4
CC-Early	1	Flow	-51.4	42.5	60.0
CC-Early	2	Flow	-50.9	38.8	53.9
CC-Early	3	Flow	-50.0	42.7	58.6
CC-Early	1	HSPF	30.5	91.0	-6.6
CC-Early	2	HSPF	9.8	94.8	-13.5
CC-Early	3	HSPF	9.9	94.8	-13.1
CC-Early	1	LOADEST	-1.1E+45	100.0	100.0
CC-Early	2	LOADEST	-7.9E+44	100.0	100.0
CC-Early	3	LOADEST	-9.1E+44	100.0	100.0
CC-Early	1	MRC	32.1	94.2	67.1
CC-Early	2	MRC	45.9	95.4	64.9
CC-Early	3	MRC	38.7	95.0	67.5
CC-Early	1	MUSLE	2.1	90.4	18.1
CC-Early	2	MUSLE	34.5	92.6	29.5
CC-Early	3	MUSLE	24.1	92.4	31.3

Table 13. NSE scores for CC-Late calibration, validation, and transfer to CLP-F for all algorithms and the top three selected parameter sets (Runs).

Experiment	Run	Module	Calibration	Validation	Transfer
CC-Late	1	DHSVM	0.74	-46.02	-55.56
CC-Late	2	DHSVM	0.75	-37.96	-44.02
CC-Late	3	DHSVM	0.75	-58.20	-67.37
CC-Late	1	Flow	0.76	-0.57	0.79
CC-Late	2	Flow	0.69	-0.14	0.76
CC-Late	3	Flow	0.67	-0.04	0.74
CC-Late	1	HSPF	0.75	-45.84	-107.82
CC-Late	2	HSPF	0.75	-43.50	-100.39
CC-Late	3	HSPF	0.75	-28.44	-62.36
CC-Late	1	LOADEST	0.65	-0.86	-19.35
CC-Late	2	LOADEST	0.40	0.33	-3.97
CC-Late	3	LOADEST	0.60	-0.88	-6.85
CC-Late	1	MRC	0.73	0.50	-852.72
CC-Late	2	MRC	0.42	0.15	-135.74
CC-Late	3	MRC	0.71	0.32	-447.81
CC-Late	1	MUSLE	0.56	-3.96	-30.81
CC-Late	2	MUSLE	0.62	-7.83	-44.70
CC-Late	3	MUSLE	0.65	-6.20	-31.82

Table 14. Percent bias scores for CC-Late calibration, validation, and transfer to CLP-F for all algorithms and the top three selected parameter sets (Runs).

Experiment	Run	Module	Calibration	Validation	Transfer
CC-Late	1	DHSVM	-14.4	-552.9	-845.7
CC-Late	2	DHSVM	-5.6	-490.1	-751.4
CC-Late	3	DHSVM	-27.0	-618.7	-930.0
CC-Late	1	Flow	6.2	-79.1	15.2
CC-Late	2	Flow	7.5	-78.3	17.0
CC-Late	3	Flow	3.5	-81.6	11.0
CC-Late	1	HSPF	-20.0	-597.8	-1180.3
CC-Late	2	HSPF	-15.8	-596.1	-1149.0
CC-Late	3	HSPF	-0.1	-422.4	-882.6
CC-Late	1	LOADEST	45.0	-111.7	-474.9
CC-Late	2	LOADEST	67.7	-48.4	-243.9
CC-Late	3	LOADEST	51.3	-94.7	-299.5
CC-Late	1	MRC	25.5	52.1	-2018.8
CC-Late	2	MRC	66.0	75.6	-837.6
CC-Late	3	MRC	42.4	61.2	-1290.3
CC-Late	1	MUSLE	15.7	-307.6	-833.7
CC-Late	2	MUSLE	-2.3	-451.8	-1073.0
CC-Late	3	MUSLE	5.6	-388.6	-912.4