AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1002/2017MS001120

A Multialgorithm Approach to Land Surface Modeling of Suspended Sediment in the Colorado Front Range

J. R. Stewart[1], B. Livneh[1,2], J. R. Kasprzyk[1], B. Rajagopalan[1,2], J. T. Minear[2], and W. J. Raseman[1]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO, USA, [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

Key Points:
- Multialgorithm sediment modeling is a novel way to incorporate uncertainty into land surface model estimates of sediment
- There are trade-offs when finding an equilibrium between suspended sediment and the partitioning of streamflow into runoff and base flow
- Transferability of calibrated suspended sediment parameters is sensitive to the variability of the observed data

Supporting Information:
- Supporting Information S1
- Data Set S1–S18

Correspondence to:
B. Livneh,
ben.livneh@colorado.edu

Citation:
Stewart, J. R., Livneh, B., Kasprzyk, J. R., Rajagopalan, B., Minear, J. T., & Raseman, W. J. (2017). A multialgorithm approach to land surface modeling of suspended sediment in the Colorado Front Range. *Journal of Advances in Modeling Earth Systems*, *9*, 2526–2544. https://doi.org/10.1002/2017MS001120

Received 10 JUL 2017
Accepted 26 SEP 2017
Accepted article online 29 SEP 2017
Published online 12 NOV 2017

Abstract A new paradigm of simulating suspended sediment load (SSL) with a Land Surface Model (LSM) is presented here. Five erosion and SSL algorithms were applied within a common LSM framework to quantify uncertainties and evaluate predictability in two steep, forested catchments (>1,000 km^2). The algorithms were chosen from among widely used sediment models, including empirically based: monovariate rating curve (MRC) and the Modified Universal Soil Loss Equation (MUSLE); stochastically based: the Load Estimator (LOADEST); conceptually based: the Hydrologic Simulation Program—Fortran (HSPF); and physically based: the Distributed Hydrology Soil Vegetation Model (DHSVM). The algorithms were driven by the hydrologic fluxes and meteorological inputs generated from the Variable Infiltration Capacity (VIC) LSM. A multiobjective calibration was applied to each algorithm and optimized parameter sets were validated over an excluded period, as well as in a transfer experiment to a nearby catchment to explore parameter robustness. Algorithm performance showed consistent decreases when parameter sets were applied to periods with greatly differing SSL variability relative to the calibration period. Of interest was a joint calibration of all sediment algorithm and streamflow parameters simultaneously, from which trade-offs between streamflow performance and partitioning of runoff and base flow to optimize SSL timing were noted, decreasing the flexibility and robustness of the streamflow to adapt to different time periods. Parameter transferability to another catchment was most successful in more process-oriented algorithms, the HSPF and the DHSVM. This first-of-its-kind multialgorithm sediment scheme offers a unique capability to portray acute episodic loading while quantifying trade-offs and uncertainties across a range of algorithm structures.

1. Introduction

1.1. Background

Soil erosion adds constituents to streams that alter water chemistry and streambed geometry, interacting with aquatic life (Rice et al., 2001) and adversely affecting water treatment efforts (Delpla et al., 2009) and reservoir operations (Podolak & Doyle, 2015). Over the past century, climate extremes and land cover changes have been intensifying sediment loading in streams across the globe (Walling, 2006) and are of particular concern in montane systems like the Colorado Front Range which is expected to warm (Pachauri et al., 2014) and are likely to experience higher suspended sediment loads (SSLs) than historically (Smith et al., 2011; Whitehead et al., 2009). Therefore, accurate prediction of SSL will become increasingly important for future water resources management.

Many sediment modeling efforts have estimated erosion and transport over individual hillslopes or within small channel reaches or catchments (<50 km^2) (Morgan et al., 1998; Nearing et al., 1989; Smith et al., 1995; Wicks & Bathurst, 1996), while continental and global efforts have typically focused on single, spatially lumped models (Cohen et al., 2013, 2014; De Vente et al., 2013; Syvitski & Milliman, 2007), or excluded consideration of the surface energy balance (Cohen et al., 2014) which is an integral part of Land Surface Model (LSM) application, and is important under a changing climate. A gap exists for expanding small-scale, spatially distributed calculations to intermediate scales (~1,000 km^2) because of the difficulties in acquiring high-quality input data and characterizing key processes (De Vente et al., 2013; Yang, 2006). This work provides a quantitative intercomparison of lumped and distributed erosion and sediment transport methods under a common LSM framework and a critical evaluation with intermediate-scale catchments, representing an important step in sediment modeling.

1.2. Existing Modeling Efforts

This study compares the structural uncertainty among model estimates because existing erosion and sediment transport models incorporate different spatial and temporal scales of application, parameter requirements, and model structures. For detailed explanations of erosion and sediment transport modeling initiatives, see Merritt et al. (2003), Aksoy and Kavvas (2005), and De Vente et al. (2013). Due to the reliance on hydrologic inputs, erosion and sediment transport models are sometimes coupled with hydrologic or land surface models that incorporate different scales and model structures (Arnold et al., 1998; Doten et al., 2006; Nearing et al., 1989). As highlighted below, four of the common model sediment structures are empirical, conceptual, stochastic, and physically based. To date, many modeling initiatives are hybrids of the four model categories, combined to increase accuracy and efficiency (Kabir et al., 2011; Pechlivanidis et al., 2011; Zuliziana et al., 2015).

Empirical methods, such as the Modified Universal Soil Loss Equation (MUSLE) (Williams & Berndt, 1977; Wischmeier et al., 1960) or the Monovariate Rating Curve (MRC) (Glysson, 1987), 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).

Conceptual models such as the Hydrological Simulation Program–Fortran (HSPF) (Bicknell et al., 1996; Johanson & Davis, 1980) or the Agricultural Non-Point Source Model (AGNPS) (Young et al., 1989) represent a catchment or hillslope through storage systems. Conceptual models are limited by parametric uncertainty due to their lack of in situ observations (Aksoy & Kavvas, 2005). However, the conceptual and parameterized structure lends itself well to calibration.

Stochastic methods such as the Load Estimator (LOADEST) (Runkel et al., 2004) or multiple linear regression (Gartner et al., 2009) can relate multiple predictors, such as climate or catchment characteristics, to a response using regression techniques. These methods are also reliant on historical data (Helsel & Hirsch, 2002). Like empirical methods, stochastic estimation is frequently spatially lumped.

Physically based models such as the Distributed Hydrology Soil Vegetation Model (DHSVM) (Doten et al., 2006; Wigmosta et al., 1994) or the Water Erosion Prediction Project (WEPP) (Nearing et al., 1989) 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 et al., 2012).

There are different techniques for comparing erosion and sediment model outputs with observed data depending on the spatial and temporal scales of application. These methods include using maps of landslides for comparing erosion outputs (Doten et al., 2006), or using reservoir sedimentation surveys or instantaneous suspended samples for SSL outputs (Strand & Pemberton, 1982). Though sediment transport models can be calibrated to sediment volumes from reservoir sedimentation surveys over long time scales (De Vente et al., 2008), comparison with instantaneous loads allows for analysis of SSL over shorter time scales (Cohen et al., 2013). In ungauged catchments, studies have estimated model parameters from catchment characteristics (Morehead et al., 2003; Srinivasan et al., 2010) or transferred parameters from nearby catchments or climatically and physically similar areas through regionalization schemes (Hundecha & Bárdossy, 2004; Samaniego et al., 2010). However, transferring parameters to ungauged catchments often lacks validation.

To date, few studies have quantitatively compared and analyzed the output of multiple erosion and sediment transport models over the same period and catchment scale, though they were not run under a unified framework. For example, Jetten et al. (1999) ran a small-scale soil erosion model comparison using seven field scale (0.01–10 ha) and seven catchment scale (40 ha) models, incorporating empirical, conceptual and physical models in their study. Two of the physically based models, the WEPP (Nearing et al., 1989) and the KINEROS2 (Smith et al., 1995), performed best for the field scale and catchment scale, respectively.

A larger, regional-scale quantitative study was completed by De Vente et al. (2008) using three models with structures ranging from empirical, to factorial scoring, to physically based. The models were applied to 61 catchments in Spain with areas from 30 to 13,000 km^2. Findings from the study indicate that a lumped modeling approach performed better than a physically based distributed approach. This conclusion was

also found through an evaluation of 14 soil erosion and sediment transport models run over catchments in Spain, Italy, Ethiopia, and Belgium (De Vente et al., 2013). By analyzing results from previous studies, De Vente et al. (2013) documented that lumped models were better able to balance out extreme events and details over larger scales than distributed models. Using a distributed approach, however, helped to identify sources and sinks (De Vente et al., 2008, 2013).

Though the model comparisons by Jetten et al. (1999), De Vente et al. (2008), and De Vente et al. (2013) allowed for quantitative analyses, the studies were limited because they were not run under a unified framework. For example, the inputs and parameters varied from model to model, making it difficult to directly compare model structures and performances. Therefore, comparison of models under a unified framework is important to address structural uncertainties under different scenarios and over multiple catchments.

1.3. Objectives

The objectives of this study are: (i) to evaluate five different SSL algorithms (three hillslope and two catchment catchment scale) within a consistent LSM framework for moderately sized catchments ($\sim$1,000 km^2); (ii) to use multiobjective optimization to efficiently calibrate the algorithms; (iii) to validate the algorithms over an ancillary time period and to transfer the calibrated parameters to a neighboring catchment; and (iv) to evaluate the results of the SSL algorithms with respect to the multiobjective calibration and LSM framework. This study advances a single model coupling scheme from Mao et al. (2010) that estimated WEPP hillslope erosion by coupling the erosion computations with the VIC LSM. Given the interalgorithm differences, it is expected that calibration of LSM hydrology simultaneously with parameters from each algorithm will produce unique insights relevant for constraining SSL predictions. Of additional interest is whether the more complex algorithms will out-perform the simpler algorithms, and whether the computationally expensive versus cheap algorithms will be most successful during calibration, validation, and the transfer to an uncalibrated catchment. Inter-algorithm comparisons are presented that aim to quantify parametric and structural uncertainties relevant for large LSM-scale prediction and a discussion follows into the implications and future directions of this research.

2. Methods

In this section, the study areas, data sources, SSL algorithms, and coupling procedures are described including a description of a *critical area* approach used to spatially discretize the landscape. Lastly, the procedures for calibration, validation, transfer to an uncalibrated catchment, as well as an overview of the optimization package used here are presented.

A set of five diverse SSL algorithms were coupled with a coarse resolution LSM framework. This study uses two empirical methods (the MRC and the MUSLE), one stochastic method (the LOADEST), one conceptual model (the HSPF), and one physically based model (the DHSVM) inserted into a common hydrologic framework, the Variable Infiltration Capacity (VIC) LSM (Liang et al., 1994). Although hillslope sedimentation is the primary focus of this work, it is acknowledged that other sources of sediment exist, including bed load, gullying and mass movement. However, gullying and mass movement are relatively rare and difficult to predict, while with few exceptions, bed load contributes to $<$10% of a river's total solid transport and often $<$1% (Meade et al., 1990; Syvitski et al., 2000).

2.1. Catchment Descriptions and Data Sources

We used the Front Range of the Colorado Rocky Mountains as a test bed to investigate the usefulness of the modeling framework. The Front Range is an important area because it is susceptible to increased wildfire and drought frequency, which can lead to higher SSL. To study sedimentation on the Front Range (Figure 1) it was necessary to identify catchments with minimal reservoir operations as well as available SSL data. The ratio of reservoir storage to mean annual streamflow was calculated to qualify potential catchments that are less affected by reservoirs. Reservoir storage was obtained from the GAGES-II database (Falcone, 2011) and United States Geological Survey (USGS) National Water Information System provided streamflow, suspended sediment concentration (SSC), and SSL data (U.S. Geological Survey, 2017). Storage ratios of 35%, 25%, and 8% for the three catchments selected here were among the lowest for stream gages along the Colorado Front Range that had coincident sediment observations (Table 1).

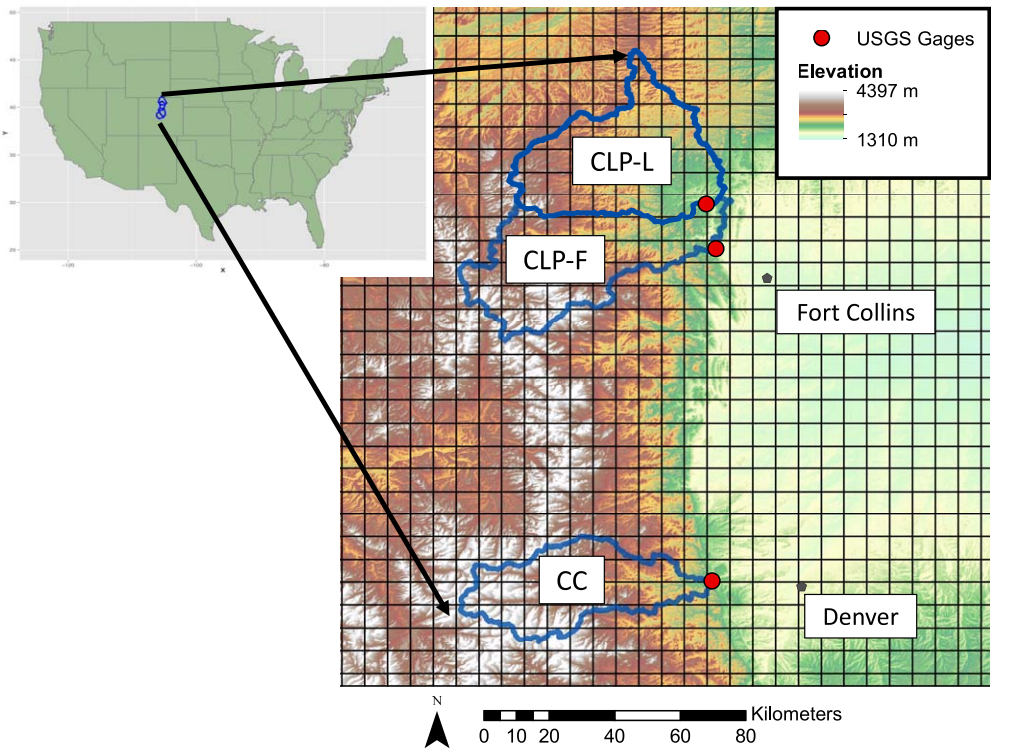

Figure 1. North Fork of Cache La Poudre at Livermore, CO (CLP-L), Cache La Poudre at Mouth of Canyon near Fort Collins, CO (CLP-F), and Clear Creek at Golden, CO (CC) catchments located within the Colorado Front Range overlaid by 1/16° resolution LSM grid cells. Blue outlined regions indicate catchment areas, red dots indicate USGS streamflow and SSL gauges, and gray dots indicate nearby cities.

The study was conducted over three catchments: two for calibration and one to test the transferability of parameters. The two catchments used in the calibration were: a subbasin of the Cache La Poudre catchment delineated to USGS gage 06751490 North Fork Cache La Poudre River at Livermore, CO (CLP-L, 1,393 km^2, SSL data sampled irregularly from 1987 to 1999), and a subbasin of the Clear Creek catchment delineated to USGS gage 06719505 Clear Creek at Golden, CO (CC, 1,024 km^2, SSL data sampled daily in 1981 and irregularly from 1981 to 1995). The calibrated parameter sets were then transferred to a larger subbasin of the Cache La Poudre catchment delineated to USGS gage 06752000 Cache La Poudre River at Mouth of Canyon

Table 1
Description of the Calibration, Validation, and Transfer Experiments for the North Fork of Cache La Poudre at Livermore, CO (CLP-L), Cache La Poudre at Mouth of Canyon Near Fort Collins, CO (CLP-F), and Clear Creek at Golden, CO (CC) Gauges, Including an Early Period "Early" and Late Period "Late," Used Alternatively for Calibration, Validation, and Transfer of Parameters to an Uncalibrated Catchment "Transfer"

Experiment	USGS gage number	Upstream storage (%)	Area (km^2)	Calibration type	Calibration period	Validation period	Transfer to CLP-F period
CLP-L	06751490	35	1,393	Joint	1987–1993	1993–1999	1992–1997
CLP-L	06751490	35	1,393	Individual	1987–1993	–	–
CC-Early	06719505	8	1,024	Joint	1980–1985	1993–1997	1992–1997
CC-Early	06719505	8	1,024	Individual	1980–1985	–	–
CC-Late	06719505	8	1,024	Joint	1993–1997	1980–1985	1992–1997
CC-Late	06719505	8	1,024	Individual	1993–1997	–	–
CLP-F	06752000	25	2,017	–	–	–	1992–1997

Note. Two calibration methods were used: calibration of all algorithms individually "Individual" and calibration of all algorithms simultaneously "Joint."

near Fort Collins, CO (CLP-F, 2,017 km^2, SSL data sampled irregularly from 1963 to 2002) (Figure 1 and Table 1). The CLP-L catchment is nested within the CLP-F catchment. The three selected catchments have similar climate, land use and soil regimes. The steep, high-elevation, catchments exhibit hydrology that is snowmelt dominated under a transition between continental-arid and alpine climate conditions. The land cover is dominated by alpine forests, shrubs, and grasslands, while the geology comprises intrusive igneous and sedimentary rocks that are largely granitic.

Compared to streamflow records, relatively few SSL observations were made within the three catchments. SSL data were sampled at irregular intervals for the three catchments, except for 1981 in Clear Creek (CC) when daily SSC and SSL were inferred from a turbidimeter (A. Duran, USGS, personal communication, 2017). Turbidimeter estimates for SSC and SSL were calculated through regression from turbidity measurements, and therefore the data come with inherent uncertainty.

2.2. Description of LSM and SSL Algorithms

The VIC LSM was selected as the overarching modeling framework, as it is a physically based model of comparable complexity to other state-of-the art LSMs and simulates hydrologic processes including base flow, runoff, evapotranspiration, soil moisture, and snow water equivalent. The VIC LSM is spatially distributed, solves both the water and energy balance equations and includes subgrid parameterization of land cover following a mosaic approach, elevation bands to distribute precipitation and lapse temperature, and variability in the soil infiltration capacity.

The VIC LSM was run in offline-mode with model settings from Livneh et al. (2015), including station-based meteorological forcings gridded at 1/16° (∼6 km) spatial resolution at a daily time step. Lateral transfer between grid cells is not incorporated into the VIC LSM, therefore a routing model was used to transport the streamflow and SSL between the VIC LSM grid cells. The routing model used in this study was RVIC, a model that solves the linearized Saint-Venant equations and uses a unit-hydrograph approach at each grid cell to route the VIC LSM outputs to a user-defined location (Lohmann et al., 1996). For each SSL algorithm, it was assumed that constituent loads were advected with the streamflow. A description of the SSL algorithms follows in order of increasing complexity.

2.2.1. Monovariate Rating Curve (MRC)

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

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

where Q is streamflow, a is a coefficient, and b is an exponent (Gray & Simões, 2008). This method is commonly used by the USGS as a representation of sediment loading at a catchment outlet. In this study, the coefficients and exponents in the MRC were fit using minimized nonlinear least squares in the R software package. To integrate the MRC with the VIC LSM, the VIC LSM and RVIC were run, and then the MRC was applied to the routed streamflow values at the catchment outlet.

2.2.2. Modified Universal Soil Loss Equation (MUSLE)

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 five 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). The MUSLE was applied with the VIC LSM by following a similar methodology to that in the Soil Water Assessment Tool (SWAT) (Arnold et al., 1998), whereby the calculation of R for a given streamflow event is based on the peak streamflow rate of an event (q_p) and the total volume of streamflow in the event (Q) as

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

The MUSLE equations were applied within each VIC LSM grid cell providing both Q and q_p. For complete details of the MUSLE see Arnold et al. (1998).

2.2.3. Load Estimator (LOADEST)

Developed by the USGS, the LOADEST is a stochastic method that uses both single and multivariate regression to predict constituent loads such as SSL and nutrients in streams from historical data (Runkel et al., 2004). The LOADEST incorporates methods from Helsel and Hirsch (2002) to predict SSL using eleven predefined algorithms, incorporating variations of linear and nonlinear predictors including streamflow, collection time, and periodicity. In the most generalized form, the linear regression is computed as

$$\ln \hat{SSL} = a_0 + \sum_{j=1}^{n} a_j X_j \tag{4}$$

where $\hat{SSL}$ is a vector of instantaneous loads, a_0 and a_j are algorithm coefficients, X_j is an explanatory variable, and n is the number of explanatory variables (Runkel et al., 2004).

The LOADEST algorithms are validated by three error estimate methods over a calibration period: Maximum Likelihood Estimation (MLE), Adjusted Maximum Likelihood Estimation (AMLE), and Least Absolute Deviation (LAD). The algorithms were run until they reached convergence, and the best algorithm was automatically selected based on Akaike Information Criterion (AIC) (Akaike, 1974). Like the MRC, to integrate the LOADEST with the VIC LSM, the VIC LSM and RVIC were run, and then the LOADEST was applied to the routed streamflow values at the catchment outlet.

2.2.4. Hydrological Simulation Program—Fortran (HSPF)

For the purposes of this study, the HSPF hillslope erosion algorithms were taken from the full HSPF model and embedded into the VIC LSM framework. The HSPF is a conceptual model developed by the Environmental Protection Agency (EPA) to simulate hydrologic processes and the transport of contaminants in catchments for water quality applications (Bicknell et al., 1996; Johanson & Davis, 1980). The HSPF is based on the Stanford Watershed Model, and incorporates conceptualized hydrologic storage within 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 algorithm. The HSPF contains both a hillslope erosion algorithm and a bed load transport algorithm. However, for the purposes of this study only the hillslope component was applied.

Hillslope erosion is divided into two components: (i) detachment by rainfall, and (ii) wash off and scour by overland flow. The algorithm is applied over a pervious land segment, as impervious segments are assumed to not generate sediment. Rainfall detachment (DET, t/ac/interval) is estimated by using the kinetic energy of raindrops as

$$DET = \left(dt \ (1 - CR)(PK)\left(\frac{I}{dt}\right)\right)^{JR} \tag{5}$$

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 on the day following a day without rainfall when DET is zero. Once the soil is detached by rainfall, it can either be redeposited or transported by overland flow. In this context, HSPF uses a conceptualized method for estimating transport capacity (TC, t/ac/interval) as

$$TC = dt(KS)\left(\frac{SU + SO}{dt}\right)^{JS} \tag{6}$$

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 transport capacity is then related to the detached sediment to estimate how much sediment is entrained in the overland flow.

Scour from the soil matrix (SCR, t/ac/interval) uses the overland flow as a metric for detachment:

$$SCR = \frac{SU}{SU + SO} dt(KG)\left(\frac{SU + SO}{dt}\right)^{JG} \tag{7}$$

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

To integrate HSPF within the VIC LSM, erosion was computed within each VIC LSM grid cell for the daily time step, similar to the MUSLE. To obtain SO estimates, subdaily VIC LSM runoff calculations were used for each grid cell, as this represents overland flow.

2.2.5. Distributed Hydrology Soil Vegetation Model (DHSVM)

As with HSPF, the DHSVM hillslope erosion algorithms were taken from the full DHSVM model and embedded into the VIC LSM framework. The DHSVM is a physically based, spatially distributed model that resolves the energy and water balances (Wigmosta et al., 1994)), developed to simulate hydrology over forested montane catchments. The physics routines in the DHSVM are largely similar to the VIC LSM, though unique to DHSVM are topographical shading, dynamic inter-grid cell routing, as well as typical application on a much finer grid cell resolution between 10 and 250 m.

Doten et al. (2006) upgraded the model to include soil loss and sediment transport from four major processes: hillslope erosion, forest road erosion, mass wasting, and channel routing using a discrete approximation to the kinematic wave equation. Hillslope erosion is 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 and excluded kinematic routing.

Hillslope erosion in DHSVM considers detachment from three 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 sediment/m^3 water) as

$$D_{of} = \beta_{de}(dy)(v_s)(TC) \tag{8}$$

where β_{de} is estimated from soil cohesion, and settling velocity is dependent on median particle grain size. TC is based on a unit stream power approach from the KINEROS model (Woolhiser et al., 1990) as

$$TC = \frac{0.05}{d50\left(\frac{PDensity}{WDensity}-1\right)^2}\sqrt{\frac{(slope)(h)}{G}}(SP - critical\ \ threshold) \tag{9}$$

where d50 is median grain size (m), PDensity is particle density (kg/m^3), WDensity is water density (kg/m^3), slope is the slope of the bed (m/m), h is water depth (m), G is gravitational acceleration (m/s^2), (SP) is stream power (m/s), and critical threshold is a critical stream power threshold (m/s). A critical threshold of 0.004 m/s was used following 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). The momentum squared of direct throughfall and leaf drip is combined to estimate overall raindrop detachment (D_R, $kg/m^2/s$) through

$$D_R = Kindex(f)(1 - C_G)((1 - C_c)M_R + M_D) \tag{10}$$

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 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 partial differential conservation of mass equation:

$$\frac{\delta(QC)}{\delta x} + \frac{\delta(AC)}{\delta t} = e(x,t) \tag{11}$$

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 the VIC LSM requires an external routing model, the equation was adapted in this study to solely account for sediment concentration (SSC, m^3/m^3) at a single grid cell during the current time step as:

$$SSC = \frac{D_r + D_{of}}{\frac{\alpha}{2dt}Q^{\beta} + \frac{\theta}{dx}Q + \beta_{de}(dy)(v_s)} \tag{12}$$

where Q is runoff (m^3/s), dx is hillslope dimension (m), dt is time step (s), α is a conceptual channel area determined by Manning's equation using the hillslope dimensions, β is set to 2/3, and θ is a time-weighting factor set to 0.55 per Doten et al. (2006).

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. Like the HSPF, DHSVM was integrated into the VIC LSM within each VIC LSM grid cell for the daily time step.

2.2.6. Critical Area Approach

Spatial discretization can have major bearing on erosion and sediment transport given nonlinear dependencies of processes on scale (Wu et al., 2005). Erosion processes typically occur on the scale of meters; therefore, implementing erosion equations on the scale of kilometers is computationally cumbersome and inefficient. To this end, a *critical area* approach was developed and applied here to isolate algorithm computations over areas where soil erosion is expected to be most productive (Figure 2). This approach was applied for each VIC LSM grid cell for the MUSLE, HSPF and DHSVM algorithms using the characteristics of an individual hillslope (e.g., calculated from a 10 m resolution DEM) incorporating hydrologic inputs from the VIC LSM grid cell scaled down to the hillslope scale by converting the runoff depth to a volume using the hillslope dimensions. The computed hillslope erosion estimates were scaled up based on the total fraction of critical area within the catchment.

If a location met all three of the following criteria, it was classified as a critical area:

1. Steep slopes—A slope greater than 15° in mountainous regions, as steeper slopes increase runoff rates (Dodds, 1997) and the potential for sediment transport: a threshold developed by Larsen et al. (2014), representing ~10% of global topography, estimated to contribute >50% of global sediment (19 Gt/yr) (Figure 2a);
2. Land cover susceptible to erosion—Erosion occurs most frequently when shrubs, grasslands or bare ground (SGB) are present given that hillslope runoff is dependent on land cover type and density (Bosch & Hewlett, 1982) yet to ensure adequate coverage in the study catchments, forest cover was also included as the catchments are heavily forested (Figure 2b);
3. Proximity to channel—The influence of the aforementioned features is limited by their proximity to a channel and the associated travel times and momentum needed to transport sediment. To develop an uncertainty range for the estimates, two stream proximities were applied in the filtering: 100 and 500 m (Figure 2c) and compared them with both individual and combined land cover types. Channel networks were developed using a flow accumulation threshold of 10,000 grid cells.

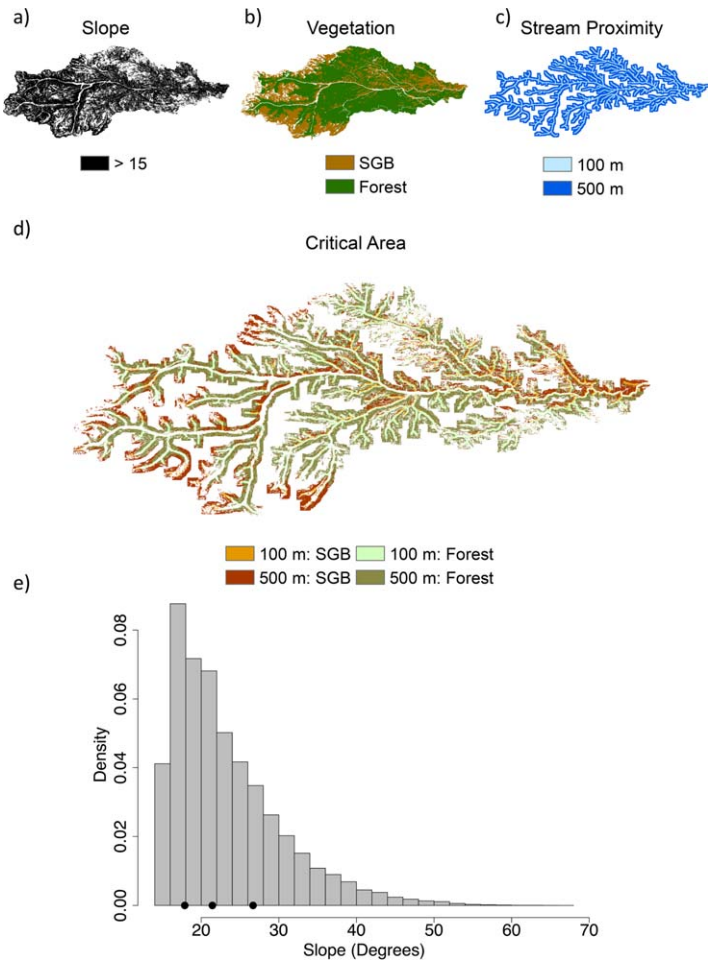

Figure 2. Critical areas computed for CC using (a) slope threshold of 15° and above (41% of the catchment had slopes below 15°), (b) vegetation types including shrub, grassland, bare ground (SGB), and forest, and (c) stream proximities of 100 and 500 m. (d) Critical area included four estimated regions. (e) For each grid cell, slope estimates were applied from the 25th, 50th, and 75th percentiles of slopes above 15° indicated by the black points.

We acknowledge that additional factors may further contribute to sedimentation (aspect, soil texture, etc.), however, the above three were the most straightforward to quantify with the least uncertainty and were applied to all catchments. For demonstration, the intersection of the above criteria depicts the critical area for CC in Figure 2d.

Distributed soil loss and sediment transport algorithms account for topography to varying degrees depending on the grid cell resolution. It has been shown that a coarse grid cell resolution produces reduced slope steepness estimates and ultimately causes substantial underestimations in soil erosion (Aalto et al., 2006; Larsen et al., 2014; Wu et al., 2005). In this context, a 10 m Digital Elevation Model DEM was used in this study for hillslope erosion estimates, as this resolution was found to be adequate in balancing topographical representation with computational expense (Wu et al., 2005; Zhang & Montgomery, 1994).

Slope was calculated from a 10 m DEM (USGS National Elevation Dataset) for all pixels within each grid cell. Because the critical areas were chosen for slopes greater than 15°, slope distributions were generated above this threshold for each grid cell. Importantly, the

methodology was implemented to account for the variability in slopes and ensuing sediment fluxes computing SSL for slopes at the 25th, 50th, and 75th percentiles (Figure 2e). SSL was computed from each slope percentile, and then aggregated by weighting the SSL estimates by 25%, 50%, and 25%, respectively to more heavily weight the SSL from the mean slope. For each grid cell, the aggregated SSL estimates from the hillslope (*HE SSL*) were then scaled up from the hillslope resolution to the grid cell resolution as

$$SSL = (HE\ \ SSL)\left(\frac{Grid\ \ Area}{HE\ \ Area}\right)(Critical\ \ Area) \tag{13}$$

where SSL is in (kg/m^2/s), *Grid Area* is the area of an individual VIC LSM 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.

2.3. Method Approach and Rationale

The MRC and LOADEST algorithms were applied after routing the VIC LSM streamflow, whereas the others used the hydrometeorology, soil, and vegetation fields from each VIC LSM grid cell, similar to VIC-WEPP (Mao et al., 2010). For these latter algorithms, the *critical area* approach was applied to calculate SSL using characteristics of hillslopes in productive regions of the catchments, rather than with aggregated properties from the large homogenous LSM grid cells. Parameters were then identified using a multiobjective calibration routine and performance was validated using a different time period, and parameter transferability was tested at a nearby catchment.

2.4. Calibration, Validation, and Transferability

To assess the robustness and uncertainties in algorithm portrayals of streamflow and suspended sediment load, calibration, validation, and a transfer to an uncalibrated catchment were performed as detailed in Table 1. Five and six-year periods surrounding the dates of the SSL observations were selected to ensure the simulation periods captured a broad range of climate variability (see supporting information Figure S1) representative of both wet and dry regimes. Additionally, the algorithm was run for a "spin-up" period of 1 year in order to minimize impacts of the initial algorithm settings. The experiments were designed to explore the impact of calibrating over different sample sizes, time periods, and ranges in data magnitudes, as shown in Figure 3.

Streamflows were not found to be statistically different (Wilcoxon Rank Sum Test) across the experimental periods at each gage ($p = 0.31$ for CLP-L, $p = 1.0$ for CC, Table 1 and Figure 3), yet observed SSL exhibited a wide range in magnitude, possibly attributed to variations in sediment supply. This disparity between streamflow and SSL magnitudes is fairly common, yet it represents a challenging issue for modeling and predictability. The SSL observations for the early period in Clear Creek (CC-Early, 1980–1985) (Table 1) were estimated through regression from turbidity measurements, thereby introducing more uncertainty into the data, which could have exacerbated the differences in magnitude between the Clear Creek time periods. Furthermore, the CC-Early SSL covered a year with very low streamflow whereas CC-Late (1993–1997) included a more representative range of streamflow. Precipitation, maximum temperature and minimum temperature showed no significant differences across periods, other than for minimum temperature in Cache La Poudre at Livermore.

The overarching goal of calibration was to identify parameter settings for the Land Surface Model and SSL algorithms simultaneously, or "jointly," that produced the best overall performance of streamflow and SSL, so that the resulting five diverse SSL estimates would provide a useful range of SSL prediction. It was expected that joint parameter calibration across the diverse algorithms would produce trade-offs in performance, as the SSL algorithms were reliant on different partitions of streamflow. Because of this, additional calibrations were performed on the individual algorithms to realize their best possible performance, from which trade-offs could be assessed relative to the joint calibration.

2.5. Multiobjective Optimization

Multiobjective evolutionary algorithms (MOEAs) have been utilized for multibjective calibration of hydrological models, using the concept of nondomination (a nondominated solution is one that is not equaled or exceeded in all objective function values by any other feasible solution) (Gupta et al., 1998). The Borg MOEA, an autoadaptive algorithm, was selected for this study because it is among the top performing MOEAs (Hadka & Reed, 2012a). A single Borg function evaluation represents an algorithm simulation from a unique set of parameters. Each calibration experiment involved 20,000–30,000 function evaluations depending on the complexity of the experiment, and incorporated five random seeds to ensure the

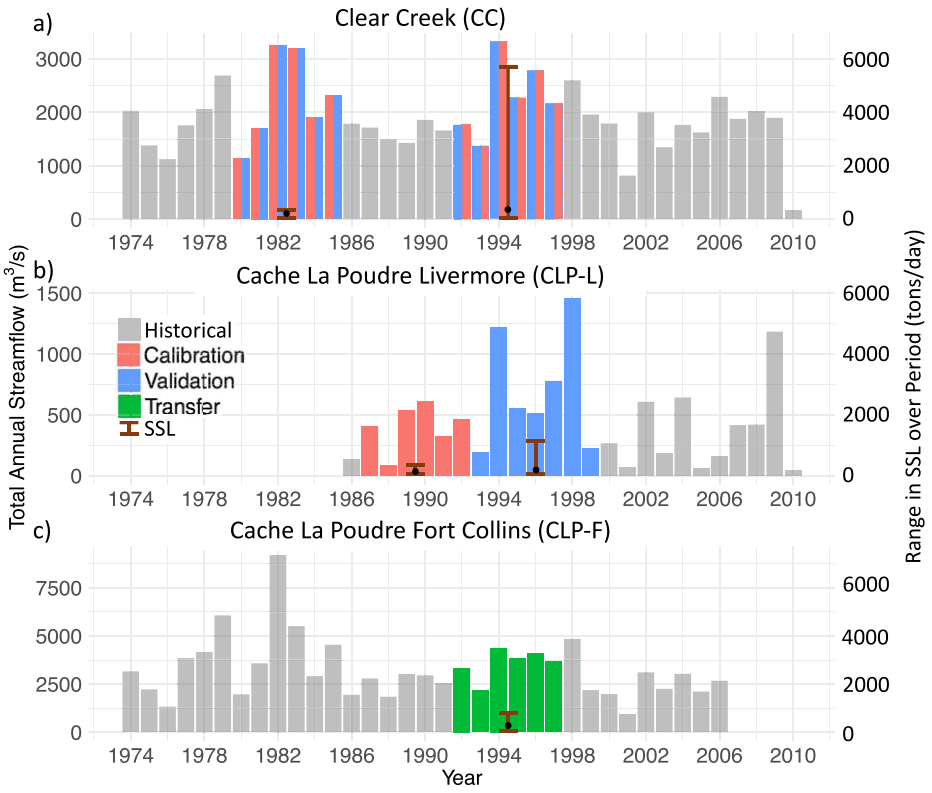

Figure 3. Bar plots of total annual streamflow separated by site into calibration, validation, and transfer periods for (a) Clear Creek (CC), (b) Cache La Poudre Livermore (CLP-L), and (c) Cache La Poudre Fort Collins (CLP-F). SSL is depicted by error bars showing the range over the respective period with the mean SSL shown as black points. The lowest annual streamflows occurred in CLP-L. The largest SSL range occurred in CC over 1993–1997, whereas the lowest occurred in CC over 1980–1985, resulting in a disparity in the relative SSL between the two periods. Note: different vertical axes are used; for consistency with USGS data, a 2000-lb ton was used = 0.907 metric tons.

solutions were not an artifact of the Borg parameter search, for a total of 100,000–150,000 simulations per calibration experiment. An epsilon value of 0.1 was used 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 nondominated solutions (Kollat et al., 2012). Maximum and minimum calibration bounds for the SSL and LSM parameters were obtained from the literature for each algorithm (Demaria et al., 2007; Donigian & Love, 2003; Doten et al., 2006; Maidment, 1993; Troy et al., 2008; Yanto et al., 2017) and are listed along with parameter descriptions in (supporting information Table S1).

Two objective functions were selected for each SSL algorithm in the joint calibration to capture important components of the hydrograph and SSL time series for a total of eight objective functions. Using more than ten objective functions has been found to reduce the efficiency of an MOEA to find solutions (Hadka & Reed, 2012b), therefore we elected to use eight. The first objective was Nash Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970), where an NSE value of 1 is considered a perfect 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 $\geq$ 0.50 is considered satisfactory for both streamflow and SSL algorithm performance (Moriasi et al., 2007). The second objective was percent bias (bias) to estimate the overall simulated magnitude compared with observed (Gupta et al., 1999), where a bias value of 0 is considered to have no magnitude bias in the algorithm. A bias value $\pm$25% is considered satisfactory for streamflow, and a bias value $\pm$55% is considered satisfactory for SSL (Moriasi et al., 2007).

For the individual calibration, three additional objectives were used for each algorithm, including Pearson's Correlation Coefficient (R) (Pearson, 1920), variability (Gupta et al., 2009), and transformed root mean squared error (tRMSE) (Wagener et al., 2009). There was a total of five objective functions for streamflow,

and six objective functions for SSL because streamflow NSE was also included in each individual SSL calibration.

To reduce the number of MOEA-produced solutions to a tractable number, filters were applied to solutions to eliminate poor performance in individual objective functions (Herman et al., 2014; Kasprzyk et al., 2013). For the first filter, a threshold approach was applied by selecting solutions with NSE >0 and bias $<100\%$. If there were still more than 15 solutions in the filtered set, a second filter was applied. For the second filter, the range of each objective function was calculated, and solutions were filtered to be within a percentage threshold (e.g., the top 70% of the objective function range). The threshold was decreased until less than 15 solutions were selected for each experiment. This technique was compared with the method of minimizing Euclidean Distance and similar selections were found.

3. Results

We first present the calibration results using parallel coordinate plots to identify trade-offs in objective functions, followed by analyses of SSL data distributions and time series for streamflow and SSL. Analysis of the validation and transfer experiments are then presented using objective function performances. Finally, we explore applications of the multialgorithm routine across longer temporal scales, and within a catchment to identify SSL sources.

3.1. Calibration Results

To explore the relative sensitivity of SSL and streamflow parameters, we first present a representative parallel coordinate plot for CC-Late for an individual algorithm calibration in Figure 4 followed by a joint calibration in Figure 5. We used parallel coordinate plots to show performance of objective functions simultaneously, which helps to identify trade-offs. The individual calibration showed a trade-off between SSL NSE and Streamflow NSE, indicating different equilibrium states needed for successful SSL and streamflow simulations. Across all SSL algorithms, the VIC LSM streamflow parameters were more sensitive than

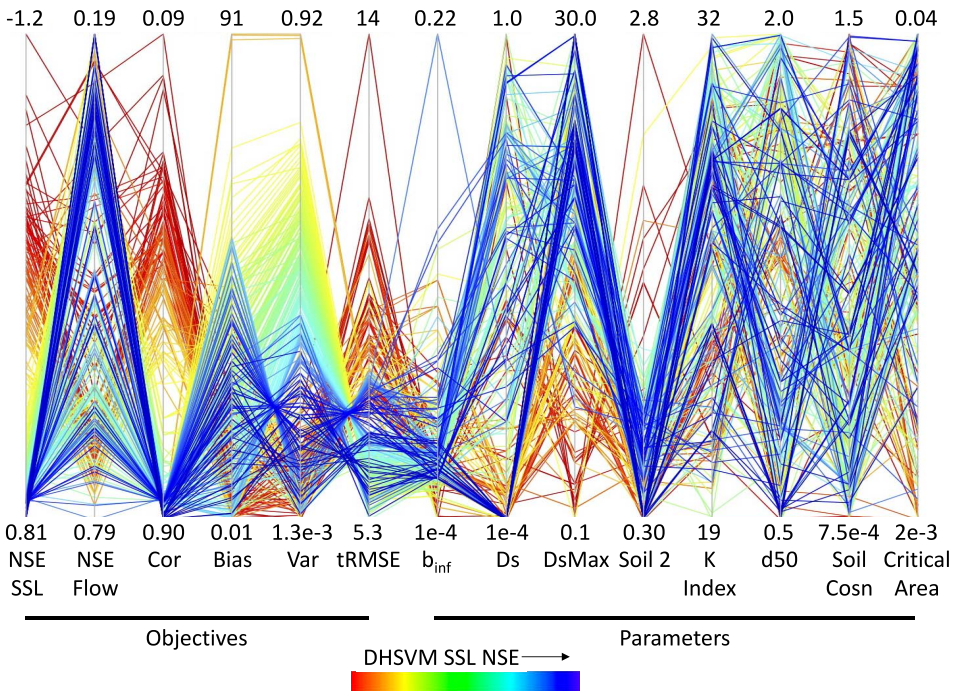

Figure 4. Parallel axis plot for the DHSVM individual SSL algorithm calibration for CC-Late during 1993–1997 with a spin-up period in 1992, solutions sorted by SSL NSE with blue solutions performing best. There were trade-offs between the objectives, as well as the VIC LSM soil parameters. DHSVM SSL parameters were less sensitive than the VIC LSM soil parameters, as the best performing solutions spanned the ranges of the SSL parameters (see supporting information Table S1 for parameter descriptions).

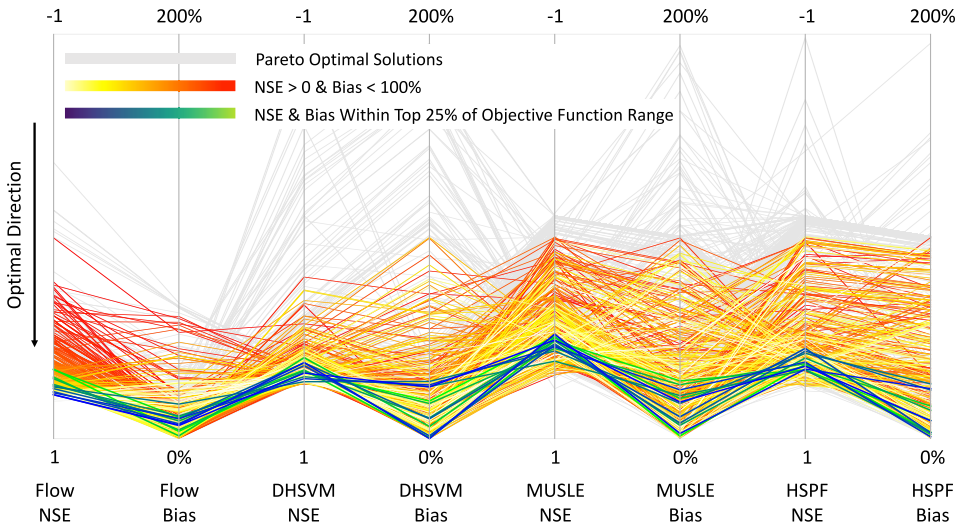

Figure 5. Parallel coordinate plot of Pareto optimal solutions from Borg for CC-Late computed for the joint algorithm calibration, 1993–1997, with a spin-up year in 1992. The Pareto optimal solutions were initially filtered to NSE $>$ 0 and bias $<$ 100%, and then filtered again to identify solutions within the top 25% of each filtered objective function range. 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 SSL parameters. The SSL algorithms showed different patterns in the VIC LSM soil parameters for greatest SSL NSE, indicating that the SSL mechanisms within each algorithm preferred different streamflow characteristics.

Across all algorithms and joint experiments, there were trade-offs between NSE and bias (Figure 5). Each joint experiment had more than 1,000 Pareto optimal solutions identified from the combined solutions of the random seeds. Applying the filtering technique yielded 12–14 top performing parameter sets for each experiment.

Overall, the algorithms in the joint calibrations performed worse in NSE and bias than the individual calibrations. This could be attributed to the dynamic dependence of SSL on streamflow, as different equilibrium states of streamflow were optimal for different algorithms. Importantly, the quantity and type of objective functions varied between individual and joint calibrations, and caution should be taken when directly comparing the two calibrations.

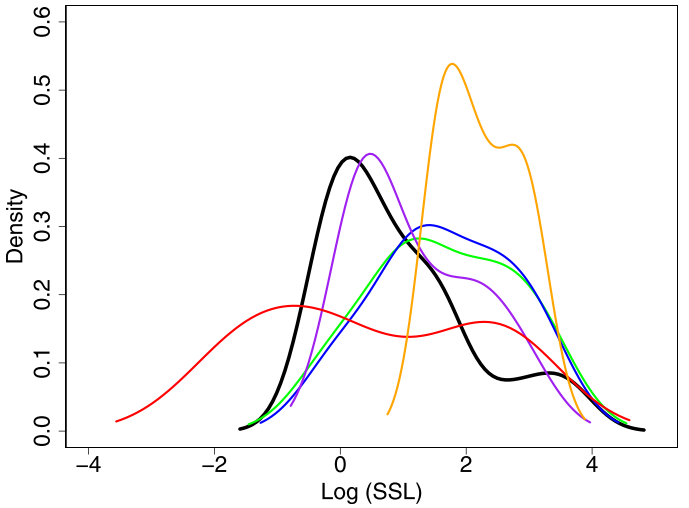

Figure 6. Kernel density estimation for the CC-Late observed data and each SSL algorithm computed over the calibration period 1993–1997 for the joint algorithm calibration with a spin-up year in 1992. The kernel density distributions were calculated from the mean of the SSL values from the top performing parameter sets.

To assess the ability of the algorithms to capture the distributions of SSL, kernel densities were estimated for each SSL algorithm and performance was computed across quantiles compared with observed (Figure 6). When analyzing the 50th, 75th, and 95th percentiles of SSL outputs from each algorithm, the LOADEST and the MRC had the highest performance for the 50th percentile, and the LOADEST had the highest performance for the 75th percentile. However, the DHSVM had the highest performance at the 95th percentile followed closely by the HSPF.

Simulations of daily streamflow from the top parameter sets for each joint calibration experiment are shown for CC-Late as an example (Figure 7). The VIC LSM streamflow was generally flashier than observed and did not capture several peaks, though the seasonal cycle was well portrayed (Figure 7a). Capturing the large SSL magnitudes in 1995 biased the algorithms toward higher peak SSL, resulting in overestimation of smaller peaks earlier in the period (Figure 7b).

3.2. Validation and Transferability of the Joint Calibration

Algorithms generally performed better during the calibration period than the validation period (Figure 8). This could indicate that the joint

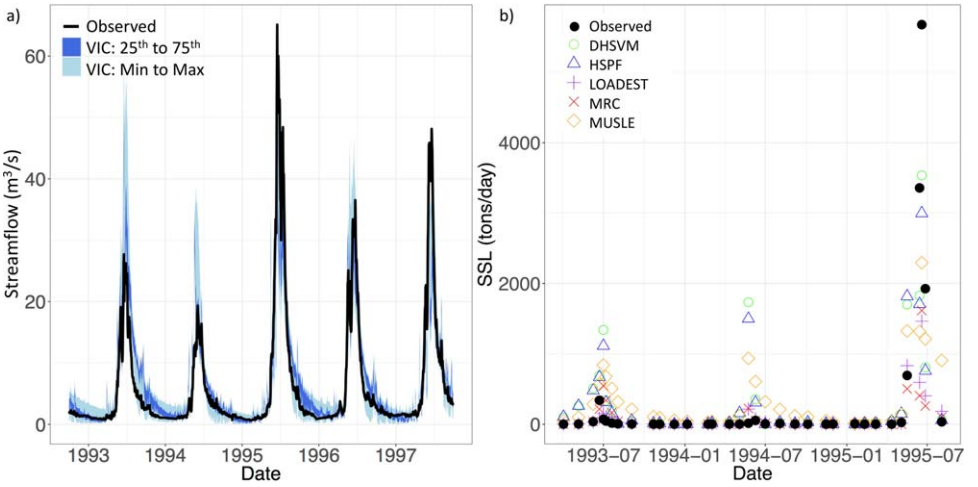

Figure 7. (a) Hydrograph for CC-Late with the VIC LSM streamflow computed over the calibration period 1993–1997 for the joint algorithm calibration with a spin-up year in 1992. The plotted minimum to maximum ranges and 25th to 75th percentiles were calculated from the VIC LSM streamflow values from the top 12 performing parameter sets. (b) Time series of suspended sediment load (SSL) for CC-Late with each SSL algorithm computed over the calibration period 1993–1997 for the joint algorithm calibration with a spin-up year in 1992; only the years with SSL observation shown, 1993–1995. The plotted SSL estimates were calculated from the mean of the SSL values from the top 12 performing parameter sets.

calibration did not find the optimal solutions for each catchment system, but is more likely the result of drastically different SSL magnitudes between the periods in both CC and CLP-L. Transferred streamflow performance (NSE in Figure 8, bias in supporting information Figure S2) was best in CC-Late, which we attribute to the calibration and transfer spanning overlapping periods and hence having similar climate. CLP-L had the second-best performance in transferred streamflow NSE and bias, which we attribute to the nesting of CLP-L within CLP-F, as the two catchments have overlapping catchment characteristics. Consistently poor results were observed from the CC-Late transfer across all SSL algorithms and both objectives. We attribute the poor results to the large magnitude of SSL observations in the CC-Late calibration period.

Overall, the two process-based algorithms—the DHSVM and the HSPF—had the best performance in transfer (CLP-L and CC-Early), meeting the satisfactory performance criteria. Both algorithms' incorporation of both precipitation and runoff (rather than total streamflow) drivers of sediment mobilization makes their SSL estimates less correlated with site-specific streamflow. The two algorithms were therefore not as affected by poor streamflow performance and were less reliant on a single empirical equation like the other algorithms. It should be noted that the HSPF and the MUSLE had two overlapping SSL parameters, which could have affected the joint calibration results.

Individual calibrations (shown as circles in Figure 8) consistently outperformed the joint calibrations, highlighting an important artifact of the streamflow-dependence of SSL algorithm performance. Compensatory behavior could arise during optimization of individual SSL algorithms with streamflow, whereby parameters that produce unique and sometimes erroneous streamflow performance yield improved SSL performance. In this context, the joint algorithm calibration met a significant challenge to simultaneously align both streamflow as well as five—sometimes divergent—SSL algorithms, often resulting in suboptimal results overall, despite numerous iterations (>20,000).

Streamflow met the performance criteria from Moriasi et al. (2007) during the CC-Late calibration and transfer periods, but performed poorly in CLP-L and CC-Early for streamflow calibration, validation or transfer. CLP-L had a high storage ratio (35%), while CC-Early only had overlapping SSL data for one of streamflow calibration years therefore biasing performance toward that year. Streamflow performance in the joint calibration was clearly impacted by the SSL algorithms' different preferences for streamflow timing and equilibrium states, as the algorithm performance for the individual calibrations exceeded the performance of the joint calibration, except in CLP-L due to the high storage ratio (Figure 8).

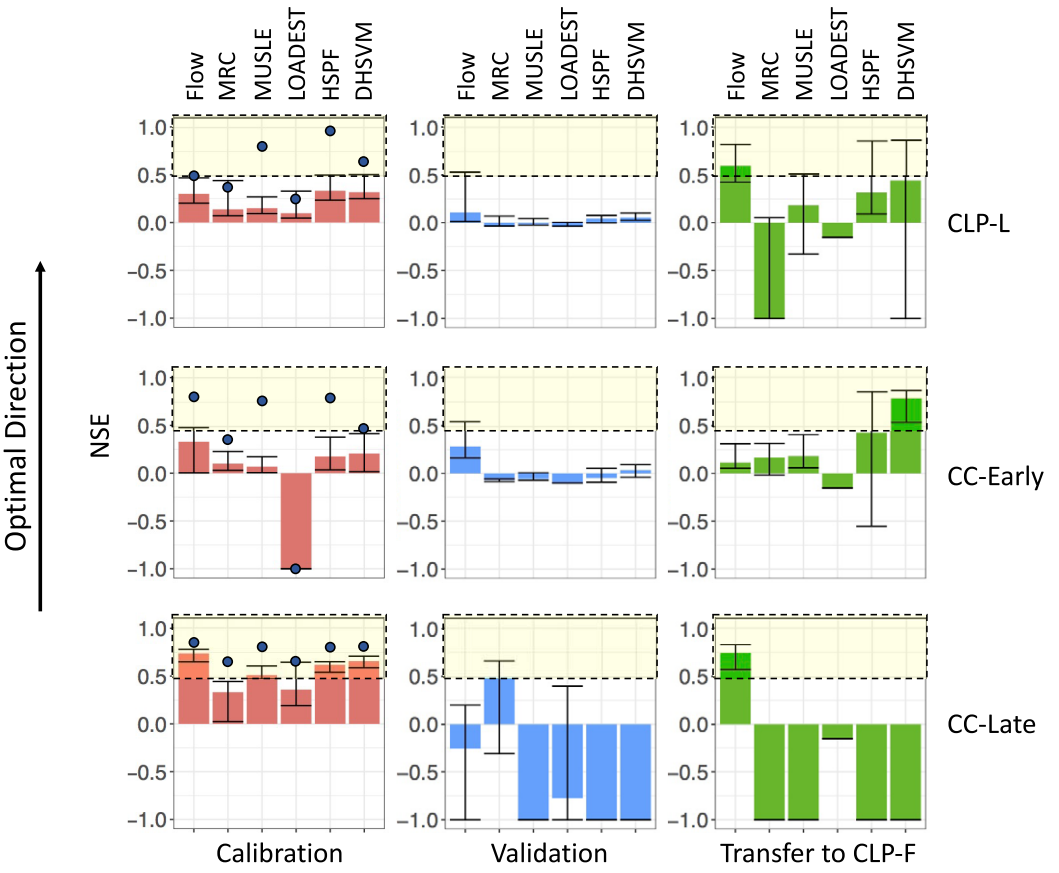

Figure 8. NSE performance for calibration, validation, and transfer periods for all algorithms in CLP-L, CC-Early, and CC-Late. Calibration runs used the top solutions from the joint calibration from each experiment, and validation runs used the same solutions over a different period within the same catchment. Transfer runs applied the top solutions from each experiment to the CLP-F catchment. Error bars represent the range in performance surrounding the median of the solutions. Values of NSE < -1.0 are plotted as −1.0 for visualization purposes. The blue points in the calibrations represent the maximum NSE obtained from the individual algorithm calibrations. The highlighted boxes indicate satisfactory criteria from Moriasi et al. (2007).

To assess the importance of anomalous SSL magnitudes on objective function performance, several large outlier years were removed from the validation: 1999 in CLP-L had observed SSL magnitudes exceeding 1,000 t/d; and 1995 in CC-Late had observed SSL magnitudes exceeding 5,000 t/d. For both removals, performance of the algorithms generally improved. A paired *t* test before and after removal showed significant differences between the samples, with p-values of 0.03 (CLP-L) and 0.02 (CC-Late). The CLP-L and CC-Late SSL calibrations were therefore at least partially biased toward the anomalously high SSL during the calibration periods thereby overestimating smaller SSL values in the validation and transfer periods.

3.3. Applications of the Multialgorithm Routine

To explore the capabilities of the multialgorithm technique, we present a long-term application followed by a spatial application (Figure 9). The algorithms were applied over the period of 1950–2013 in CC-Late to assess the inter-algorithm uncertainty and to show long-term variability. Comparing the annual sediment yield estimates with a traditional estimate of long-term SSL (Jansson, 1988), the estimates fell within the range of 0–100 t/km^2/yr for most years. However, in years with large storm events such as the early 1980s (Figure 9a), the algorithms provide and important expression of the acute episodic loading and associated uncertainties associated with extreme climate conditions, relevant for water management and planning. To identify grid cells that were more productive for SSL, we spatially analyzed the variability in SSL outputs. Using the DHSVM as an example, the calibrated SSL in CC-Late follows elevation within the catchment, with high elevations yielding high SSL (Figures 9b and 9c). This could be due to the vegetation types and climate differences at high elevation where mean precipitation increases and forest cover is absent. Higher

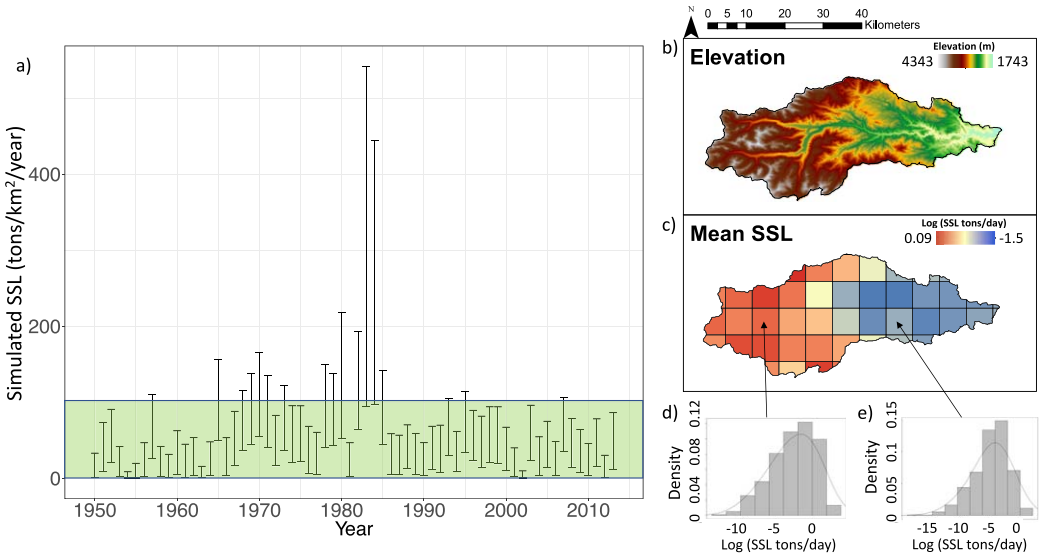

Figure 9. (a) Extension of the algorithm ensemble in CC from 1950 to 2013. The error bars represent the differences between algorithm estimates, providing a range of uncertainty for each year. The highlighted green box indicates a historical range in sediment yields for the Colorado Front Range estimated from maps developed by Jansson (1988). (b) Elevation for the CC catchment, (c) mean simulated DHSVM SSL for each VIC LSM grid cell computed for CC-Late over the calibration period 1993–1997 for an individual parameter set from the joint algorithm calibration, (d) distribution of SSL $>$ 0 for a high elevation grid cell, and (e) distribution of SSL $>$ 0 for a low elevation grid cell. Simulated SSL was zero for the majority of time steps.

precipitation increases sediment mobilization through raindrop impact and overland flow, while the lack of forest cover exposes the land surface to precipitation, further increasing the potential for sediment erosion and transport. The temporal variability of SSL during the calibration period showed negative skew for both high and low SSL grid cells (Figures 9d and 9e).

4. Discussion and Conclusions

A key outcome from this study is the comparison of multiple sediment algorithms within a consistent modeling framework. This work added multiple diverse, well-known hillslope erosion algorithms into the VIC LSM, improving upon previous work that used a single erosion algorithm within the VIC LSM (Mao et al., 2010). The work also contributed a quantitative multialgorithm comparison akin to Jetten et al. (1999) and De Vente et al. (2008). Variable performance of each algorithm was attributed to ranges in SSL observations, static parameterizations, and compromises between preferences for streamflow characteristics, yet some consistent patterns in performance emerged. Most notably, the more process-based algorithms appeared to perform best when transferring parameters to new catchments. Streamflow performance was below the satisfactory criteria from Moriasi et al. (2007) throughout many of the experiments. This indicates that either the joint algorithm calibration did not find optimal solutions, or else there may be no optimal streamflow partitioning (e.g., between surface runoff and base flow) that satisfies all SSL algorithms. Given the rigor of the calibration procedure, we are inclined to conclude that the joint calibration with SSL algorithms impacted the flexibility and robustness of the streamflow to adapt to different periods. This conclusion is supported by the consistent, satisfactory streamflow performance obtained from single algorithm calibrations (Figure 8). Future efforts could incorporate a more robust partitioning-focused calibration of streamflow into the joint calibration to better confront the issue (Shafii et al., 2017).

A confounding factor in the joint calibration is the timing discrepancy between streamflow and SSL peaks. Snowmelt dominated systems are susceptible to hysteresis during spring melts, as noted by Syvitski et al. (2000), having high SSC on the rising limb of the hydrograph. In this context, the SSL algorithms may have seen improved performance by an erroneous early streamflow peak in order to simulate the early SSL peak and associated hysteresis effects, resulting in a deterioration of streamflow performance.

The process-oriented algorithms, HSPF and the DHSVM were slightly more robust in their SSL portrayals under different climate and streamflow conditions as compared with the more static MRC and LOADEST. Glysson (1987) suggested that MRC parameters should vary annually due to changes in extreme events. Similarly, LOADEST performance is dependent on the dates of the calibration period (Runkel et al., 2004), limiting successful transferability to other time periods and locations. In contrast, the comparatively robust DHSVM erosion algorithm has been applied to simulate dynamic changes in land cover (from wildfire) and climate (Doten et al., 2006) realistically simulating increased erosion due to changes in root cohesion and increased surface runoff. This finding has implications for predictions of future climate scenarios (e.g., flooding, droughts, and wildfires) as SSL is known to increase nonlinearly with extreme events (Moody & Martin, 2009; Smith et al., 2011; Whitehead et al., 2009). Therefore, the most flexible and process-oriented algorithms, the HSPF and the DHSVM, are expected to perform better due to their explicit treatment of precipitation rates and vegetation coverage.

4.1. Theoretical and Practical Limitations

Attempting to capture the complexity of sediment transport processes has yielded the following limitations. First, physical erosion occurs on the order of meters, and despite the use of critical areas, this implementation assumes varying degrees of landscape homogeneity on the order of kilometers, inherently underestimating fine-scale erosion controls such as channel geometry and heterogeneous soil characteristics.

Second, SSL was assumed to result from rainfall and overland flow detachment on hillslopes in the MUSLE, the HSPF and the DHSVM, whereas the contribution from bed load and mass wasting events were ignored. Though the contribution of bed load to SSL has been found to be smaller than hillslope erosion in mountainous regions (Turowski et al., 2010), it is acknowledged that the present framework inherently underestimates total sediment load.

Third, the implementation of MUSLE, the HSPF and the DHSVM were based on the assumption of an infinite sediment supply, which resulted in SSL and streamflow varying together and which greatly limited the portrayal of hysteresis. In catchments where hysteresis occurs (meaning the peak in SSL comes before the peak in streamflow), the system is often supply limited. To compensate for this issue, an accounting of sediment supply limitation could be imposed, thereby allowing the peak SSL timing to occur asynchronously with streamflow. This limitation was at least partially responsible for the trade-offs between streamflow and SSL performance in the joint calibrations.

Fourth, the selected parameter sets exhibited equifinality, whereby multiple combinations of parameters yielded similar performance (Beven & Binley, 1992). To address this issue, future work could apply weighted likelihood estimates into the calibration procedure, identify parameter sets that are the most spatially physically representative, or employ more rigorous parameter constraints based on physical measurements or empirical assessments. Furthermore, the individual calibrations had better overall performance than the joint calibrations, possibly because the joint algorithm calibration method was lacking diversity in objective functions (for more explanation, see supporting information).

4.2. Conclusions

Algorithm performance varied across the experiments and periods. The more complex, process-oriented algorithms, the HSPF and the DHSVM, were best able to perform well in the transfer to a different catchment, though further testing of the algorithms in a greater number of catchments is required. Performance trade-offs between streamflow and SSL greatly limited the joint calibration performance. The trade-offs were attributed to different equilibrium states most likely resulting from (i) a lack of hysteresis due to an infinite supply assumption, (ii) differing algorithm complexity stemming from compensatory reliance on either total flow or the combination of precipitation and overland flow, and (iii) numerical issues resulting from the number and type of objective functions. Despite the trade-offs between algorithms, top performing parameters produced multidecadal SSL estimates that generally agreed with historical estimates, while providing an important estimate of episodic response to extreme years. Future work will incorporate a greater number of catchments and more diverse SSL data to more broadly evaluate performance of the method.

Acknowledgments
This research was funded by the U.S. Environmental Protection Agency, "National Priorities: Systems-Based Strategies to Improve the Nation's Ability to Plan and Respond to Water Scarcity and Drought Due to Climate Change," grant R835865. We would also like to acknowledge Janus, the supercomputer, which is supported by the NSF (CNS-0821794) and University of Colorado, Boulder. The Janus supercomputer is a joint effort of University of Colorado, Boulder, University of Colorado, Denver, and the National Center for Atmospheric Research. The data used are listed in the references, tables, and supporting information.

References

Aalto, R., Dunne, T., & Guyot, J. L. (2006). Geomorphic Controls on Andean Denudation Rates. *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. https://doi.org/10.1016/j.catena.2005.08.008

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: Model development1. *Journal of the American Water Resources Association*, *34*(1), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Bicknell, B. R., Imhoff, J. C., Kittle, J. L., Jr, Donigian, A. S., Jr., & Johanson, R. C. (1996). *Hydrological simulation program-FORTRAN*. User's manual for *release 11*. Washington, DC: U.S. Environmental Protection Agency.

Bosch, J. M., & Hewlett, J. D. (1982). A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology*, *55*(1–4), 3–23. https://doi.org/10.1016/0022-1694(82)90117-2

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

Cohen, S., Kettner, A. J., & Syvitski, J. P. M. (2014). Global suspended sediment and water discharge dynamics between 1960 and 2010: Continental trends and intra-basin sensitivity. *Global and Planetary Change*, *115*, 44–58. https://doi.org/10.1016/j.gloplacha.2014.01.011

Cohen, S., Kettner, A. J., Syvitski, J. P. M., & Fekete, B. M. (2013). WBMsed, a distributed global-scale riverine sediment flux model: Model description and validation. *Computers & Geosciences*, *53*, 80–93. https://doi.org/10.1016/j.cageo.2011.08.011

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*, *112*, D11113. https://doi.org/10.1029/2006JD007534

De Vente, J., Poesen, J., Verstraeten, G., Govers, G., Vanmaercke, M., Van Rompaey, A., . . . Boix-Fayos, C. (2013). Predicting soil erosion and sediment yield at regional scales: Where do we stand? *Earth-Science Reviews*, *127*, 16–29. https://doi.org/10.1016/j.earscirev.2013.08.014

De Vente, J., Poesen, J., Verstraeten, G., Van Rompaey, A., & Govers, G. (2008). Spatially distributed modelling of soil erosion and sediment yield at regional scales in Spain. *Global and Planetary Change*, *60*(3), 393–415. https://doi.org/10.1016/j.gloplacha.2007.05.002

Dodds, W. K. (1997). Distribution of runoff and rivers related to vegetative characteristics, latitude, and slope: A global perspective. *Journal of the North American Benthological Society*, *16*(1), 162–168. https://doi.org/10.2307/1468248

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. https://doi.org/10.2175/193864703784828345

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*, W04417. https://doi.org/10.1029/2004WR003829

Falcone, J. A. (2011). *GAGES-II: Geospatial attributes of gages for evaluating streamflow*. Washington, DC: U.S. Geological Survey.

Gartner, J. E., Cannon, S. H., Helsel, D. R., & Bandurraga, M. (2009). *Multivariate statistical models for predicting sediment yields from southern California watersheds, USGS numbered series* (Open-File Rep. 2009-1200). Washington, DC: U.S. Geological Survey.

Glysson, G. D. (1987). *Sediment-transport curves*, *USGS numbered series* (Open-File Report). Washington, DC: U.S. Geological Survey.

Gray, J. R., & Simões, F. J. (2008). Estimating sediment discharge. In *Sedimentation engineering: Processes, measurements, modeling, and practice* (pp. 1067–1088). Reston, VA: American Society of Civil Engineers.

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 Hydraulic Engineering*, *4*(2), 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)

Hadka, D., & Reed, P. (2012a). Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation*, *21*(2), 231–259. https://doi.org/10.1162/EVCO_a_00075

Hadka, D., & Reed, P. (2012b). Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. *Evolutionary Computation*, *20*(3), 423–452. https://doi.org/10.1162/EVCO_a_00053

Helsel, D. R., & Hirsch, R. M. (2002). *Statistical methods in water resources: US Geological Survey techniques of water resources investigations* (Book 4, Chap. A3). Washington, DC: U.S. Geological Survey.

Herman, J. D., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2014). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, *50*, 7692–7713. https://doi.org/10.1002/2014WR015338

Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1), 281–295.

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. 15). Washington, DC: Environmental Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

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. https://doi.org/10.5194/hess-15-1307-2011

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71. https://doi.org/10.1016/j.envsoft.2012.12.007

Kollat, J. B., Reed, P. M., & Wagener, T. (2012). When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resources Research*, *48*, W03520. https://doi.org/10.1029/2011WR011534

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

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*, *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 US, and Southern Canada 1950–2013. *Scientific Data*, *2*, 150042. 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, Series A*, *48*(5), 708–721. https://doi.org/10.1034/j.1600-0870.1996.t01-3-00009.x

Maidment, D. R. (1993). *Handbook of hydrology*. New York, NY: 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*, W08543. https://doi.org/10.1029/2009WR008268

Meade, R. H., Yuzyk, T. R., & Day, T. J. (1990). Movement and storage of sediment in rivers of the United States and Canada. In *Surface water hydrology* (pp. 255–280). Boulder, CO: Geological Society of America.

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. https://doi.org/10.1016/S1364-8152(03)00078-1

Moody, J. A., & Martin, D. A. (2009). Synthesis of sediment yields after wildland fire in different rainfall regimes in the western United States. *International Journal of Wildland Fire*, *18*(1), 96–115. https://doi.org/10.1071/WF07162

Morehead, M. D., Syvitski, J. P., Hutton, E. W. H., & Peckham, S. D. (2003). Modeling the temporal variability in the flux of sediment from ungauged river basins. *Global and Planetary Change*, *39*(1–2), 95–110. https://doi.org/10.1016/S0921-8181(03)00019-5

Morgan, R. P. C., Quinton, J. N., Smith, R. E., Govers, G., Poesen, J. W. A., Auerswald, K., . . . Styczen, M. E. (1998). The European Soil Erosion Model (EUROSEM): A dynamic approach for predicting sediment transport from fields and small catchments. *Earth Surface Processes and Landforms*, *23*(6), 527–544. https://doi.org/10.1002/(SICI)1096-9837(199806)23:6<527::AID-ESP868>3.0.CO;2-5

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

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Geology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

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. https://doi.org/10.13031/2013.31195

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., . . . van Ypserle, J. P. (2014). Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, *13*(1), 25–45. https://doi.org/10.2307/2331722

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. https://doi.org/10.5194/hess-19-4559-2015

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). https://doi.org/10.1061/(ASCE)HY.1943-7900.0000999

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

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. https://doi.org/10.1029/2000WR000079

Runkel, R. L., Crawford, C. G., & Cohn, T. A. (2004). *Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers*, *USGS numbered series* (*Techniques Methods 4-A5*). Washington, DC: U.S. Geological Survey.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

Shafii, M., Basu, N., Craig, J. R., Schiff, S. L., & Van Cappellen, P. (2017). A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. *Water Resources Research*, *53*, 3279–3301. https://doi.org/10.1002/2016WR019736

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.

Srinivasan, R., Zhang, X., & Arnold, J. (2010). SWAT ungauged: Hydrological budget and crop yield predictions in the upper Mississippi river basin. *Transactions of the ASABE*, *53*(5), 1533–1546. https://doi.org/10.13031/2013.34903

Strand, R. I., & Pemberton, E. L. (1982). *Reservoir sedimentation technical guidelines for Bureau of Reclamation* (48 pp.). Denver, CO: U.S. Bureau of Reclamation.

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

Syvitski, J. P. M., & Milliman, J. D. (2007). Geology, geography, and humans battle for dominance over the delivery of fluvial sediment to the coastal ocean. *Journal of Geology*, *115*(1), 1–19. https://doi.org/10.1086/509246

Troy, T. J., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. *Water Resources Research*, *44*, W09411. https://doi.org/10.1029/2007WR006513

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

U.S. Geological Survey. (2017). National water information system data available on the world wide web (*USGS Water Data for the Nation*). Washington, DC: U.S. Geological Survey. Retrieved from http://waterdata.usgs.gov/nwis/

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*, W02501. https://doi.org/10.1029/2008WR007347

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., & 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

Williams, J. R., & Berndt, H. D. (1977). Sediment yield prediction based on watershed hydrology. *Transactions of the ASAE*, *20*(6), 1100–1104. https://doi.org/10.13031/2013.35710

Wischmeier, W. H., Smith, D. D., *and others*. (1960). A universal soil-loss equation to guide conservation farm planning. *Transactions of the 7th International Congress of Soil Science*, *1*, 418–425.

Woolhiser, D. A., Smith, R. E., & Goodrich, D. C. (1990). *KINEROS, A kinematic runoff and erosion model: Documentation and user manual* (USDA-ARS, ARS-77). Washington, DC: United States Department of Agriculture, Agricultural Research Service.

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*. Denver, CO: U.S. Department of the Interior's Bureau of Reclamation.

Yanto, B., Livneh, B., Rajagopalan, J. & Kasprzyk, (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

Young, R. A., Onstad, C. A., Bosch, D. D., & Anderson, W. P. (1989). AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. *Journal of Soil and Water Conservation*, *44*(2), 168–173.

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 Discussion*, *12*, 6755–6797. https://doi.org/10.5194/hessd-12-6755-2015