



Quantifying the uncertainty of nonpoint source attribution in distributed water quality models: A Bayesian assessment of SWAT's sediment export predictions



Christopher Wellen^{a,*}, George B. Arhonditsis^a, Tanya Long^b, Duncan Boyd^b

^a Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario M1C 1A4, Canada

^b Great Lakes Unit, Water Monitoring & Reporting Section, Environmental Monitoring and Reporting Branch, Ontario Ministry of the Environment, Toronto, Ontario M9P 3V6, Canada

ARTICLE INFO

Article history:

Received 30 June 2013

Received in revised form 2 July 2014

Accepted 1 October 2014

Available online 12 October 2014

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:

SWAT model

Bayesian inference

Bayesian model averaging

Extreme events

Hamilton Harbour

Water quality

SUMMARY

Spatially distributed nonpoint source watershed models are essential tools to estimate the magnitude and sources of diffuse pollution. However, little work has been undertaken to understand the sources and ramifications of the uncertainty involved in their use. In this study we conduct the first Bayesian uncertainty analysis of the water quality components of the SWAT model, one of the most commonly used distributed nonpoint source models. Working in Southern Ontario, we apply three Bayesian configurations for calibrating SWAT to Redhill Creek, an urban catchment, and Grindstone Creek, an agricultural one. We answer four interrelated questions: can SWAT determine suspended sediment sources with confidence when end of basin data is used for calibration? How does uncertainty propagate from the discharge submodel to the suspended sediment submodels? Do the estimated sediment sources vary when different calibration approaches are used? Can we combine the knowledge gained from different calibration approaches? We show that: (i) despite reasonable fit at the basin outlet, the simulated sediment sources are subject to uncertainty sufficient to undermine the typical approach of reliance on a single, best fit simulation; (ii) more than a third of the uncertainty of sediment load predictions may stem from the discharge submodel; (iii) estimated sediment sources do vary significantly across the three statistical configurations of model calibration despite end-of-basin predictions being virtually identical; and (iv) Bayesian model averaging is an approach that can synthesize predictions when a number of adequate distributed models make divergent source apportionments. We conclude with recommendations for future research to reduce the uncertainty encountered when using distributed nonpoint source models for source apportionment.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Water quality management often relies upon spatially distributed watershed models for an estimation of nonpoint source pollutant fluxes, sources and fates under current conditions and possible future scenarios (Rode et al., 2010). In one particularly high-profile study, the largest algal bloom in Lake Erie's history was attributed in large part to a nonpoint source labile phosphorus pulse event (Michalak et al., 2013). The magnitude of this pulse event was estimated with the popular Soil–Water Assessment Tool (SWAT), a distributed nonpoint source water quality model. SWAT has been applied worldwide, including sites in Europe (Nasr et al., 2007),

Asia (Cheng et al., 2007; Talebizadeh et al., 2010), Africa (Setegn et al., 2010), and North America (Arabi et al., 2007).

There have been a large number of studies validating the end-of-basin predictions of the SWAT model and other nonpoint source models for many water quality constituents (Gassman et al., 2007). A few studies have sought to quantify the uncertainty of the discharge predicted by such models (Yang et al., 2007a,b; Yang et al., 2008). However, models such as SWAT are also used for source attribution and scenario analysis of water quality. There is a dearth of work which validates or quantifies the uncertainty of the source attributions of water quality impairments made by distributed water quality models. Studies with small chains of lumped models suggest that positive and negative errors of upstream stations can compensate for each other (Freni et al., 2009). For distributed, nonpoint source models such as SWAT, the much larger number of potential sources may mean that many source attributions could result in a reasonable fit at the basin outlet (Beven,

* Corresponding author at: Watershed Hydrology Group, School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario L8S 4L8, Canada. Tel.: +1 647 239 5138; fax: +1 416 287 7279.

E-mail address: wellenc@mcmaster.ca (C. Wellen).

2006). If the sources of observed pollutants are highly uncertain, this calls into question any scenario analyses based on possible future land uses.

Little is known about how the inference drawn regarding pollutant sources and fates can vary across different implementations of the SWAT model. In an influential piece, Rykiel (1996) asserts that a model is validated if it has been shown to be adequate for a particular task. Most literature on SWAT and other distributed models approaches validation with this approach (Beven, 2006; Gassman et al., 2007). However, this definition does not preclude a situation where multiple competing models all pass a validation test, yet fail to agree on pertinent aspects of system functioning. For instance, both Easton et al. (2008) and White et al. (2011) achieved an equally satisfactory fit to flow discharge measurements with two distributed watershed models. However, these models gave very different estimates of runoff generating areas. Clearly, what is needed is a methodological strategy that allows achieving predictive statements which integrate information from multiple models.

Recent work in statistics allows us to draw inference from such integrative statements. A number of methods exist to synthesize predictions across groups of models (ensembles), including sequential data assimilation approaches, such as the ensemble Kalman filter and ensemble particle filters (Vrugt and Robinson, 2007; Moradkhani et al., 2006) and post hoc ensemble integration strategies, such as the Bayesian Model Averaging (BMA) commonly used in weather forecasting (Raftery et al., 2005). While the watershed modeling community has advanced a number of ensemble-type approaches, there has not been a study evaluating their ability to reduce the ambiguity of distributed predictions of complex water quality models. Existing applications have typically been focused on simple, heuristic flow discharge models for the purposes of illustrating and/or evaluating methodological frameworks (Vrugt and Robinson, 2007; Parrish et al., 2012).

Little is known about the propagation of uncertainty through the different submodels of SWAT and other spatially distributed watershed models of nonpoint source pollution. However, results from integrated urban water quality models are instructive. Freni and Mannina (2010) decomposed the total uncertainty of an integrated model of an urban water system. They found that the water quality submodels contributed more uncertainty than its water quantity counterpart to the overall model uncertainty, though the contribution of the water quantity submodels was not negligible. It is unknown if this is the case with non-point source models. An understanding of the interplay of water quality and quantity submodels is important to guide the development and application of nonpoint source models of water quality.

We investigate these issues in the context of suspended sediment, one nonpoint source pollutant responsible for many beneficial use impairments. We address four inter-related questions: Can the SWAT model determine sediment sources with confidence when end of basin data is used for calibration? How does uncertainty propagate from the discharge submodel to the water quality submodels? Do the estimated sediment sources vary when different calibration approaches are used? Can we combine the knowledge gained from different calibration approaches? We address the first two questions by conducting the first Bayesian calibration of any of the water quality components of the SWAT model. To answer the third question, we use three SWAT model formulations Wellen et al. (2014a) advanced for accommodating extreme watershed states. We couple these formulations to the submodels which simulate suspended sediment load. To answer the fourth question, we take an ensemble approach. We adopt a post hoc BMA approach to ensemble integration, as demonstrated in the context of lake water quality modeling by Ramin et al. (2012). Synthesis proceeds by weighting the posterior densities of the model predictions inversely to their residual variance.

2. Case study

The study site is a pair of catchments situated in the drainage basin of Hamilton Harbour, a large embayment at the western end of Lake Ontario. The Harbour is designated as one of 17 Canadian Areas of Concern in the Great Lakes Basin, due to its long history of eutrophication problems (Hiriart-Baer et al., 2009; Ramin et al., 2011). Substantial uncertainty exists in regards to the projections of the future water quality conditions, due to the poorly defined nutrient loadings from the drainage basin (Gudimov et al., 2010, 2011).

In Fig. 1, we present a map of the two study catchments, Redhill and Grindstone Creeks. Aside from the land use, the two Creeks are fairly similar. The soils of the Harbour basin are mainly loams (25%), sandy loams (28%), and silty loams (20%), while organic soils, silty clay loams, and clay loams together make up about 10% of the basin soils (Soil Landscapes of Canada dataset v.3.2 from Agriculture and Agri-Food Canada; <http://sis.agr.gc.ca/cansis/nsdb/slc/index.html>). The slopes of the Harbour basin are mild and average 4.4%. Elevation ranges from 74 to 318 m above sea level. The basin has a humid continental climate, with daily temperatures ranging from -10 to -2 °C in January and 15 to 26 °C in July. The Harbour basin receives 910 mm of precipitation annually, 146 mm of which occurs as snowfall.

2.1. Redhill Creek

Redhill Creek drains an area of approximately 63 km², 66% of which is urban residential area and 17% is urban greenspace. The remaining 10% is a mixture of agriculture and forested areas. Of the urban area, 50% is impervious and 40% of the total urban area is directly connected to a storm sewer system.

2.2. Grindstone Creek

Grindstone Creek drains an area of approximately 87 km², 60% of which is agricultural land split evenly between pasture and cropland. Of the remainder, 30% is forested and 9% is urban.

2.3. Data sets used

The meteorological data for this study come from Environment Canada's Hamilton Airport station (WMO ID 71263; http://www.climate.weatheroffice.gc.ca/climateData/canada_e.html). The daily flow information comes from the Water Survey of Canada's gauges at Redhill (02HA014) and Grindstone Creeks (02HB012; <http://www.ec.gc.ca/rhc-wsc/default.asp?lang=En>). The suspended sediment load dataset used for calibration was gathered by the Ontario Ministry of the Environment between July 2010 and May 2012 (Long et al., 2014). Samples were collected as level-weighted composites over a 24-h period roughly once per week. High-flow events were targeted by the sampling, although baseflow periods were included as well. The average time between samples was about 1 week. The discharges during sampling periods spanned the 1st to the 99th percentiles of the overall discharges for both Creeks.

3. Model description

3.1. Overview of SWAT

SWAT is a semi-distributed and semi-process based model typically used to evaluate the effects of alternative management practices on watershed functioning in agricultural landscapes (Arnold et al., 1998; Neitsch et al., 2011). Watersheds are disaggregated into subbasins. Subbasins are disaggregated into hydrological

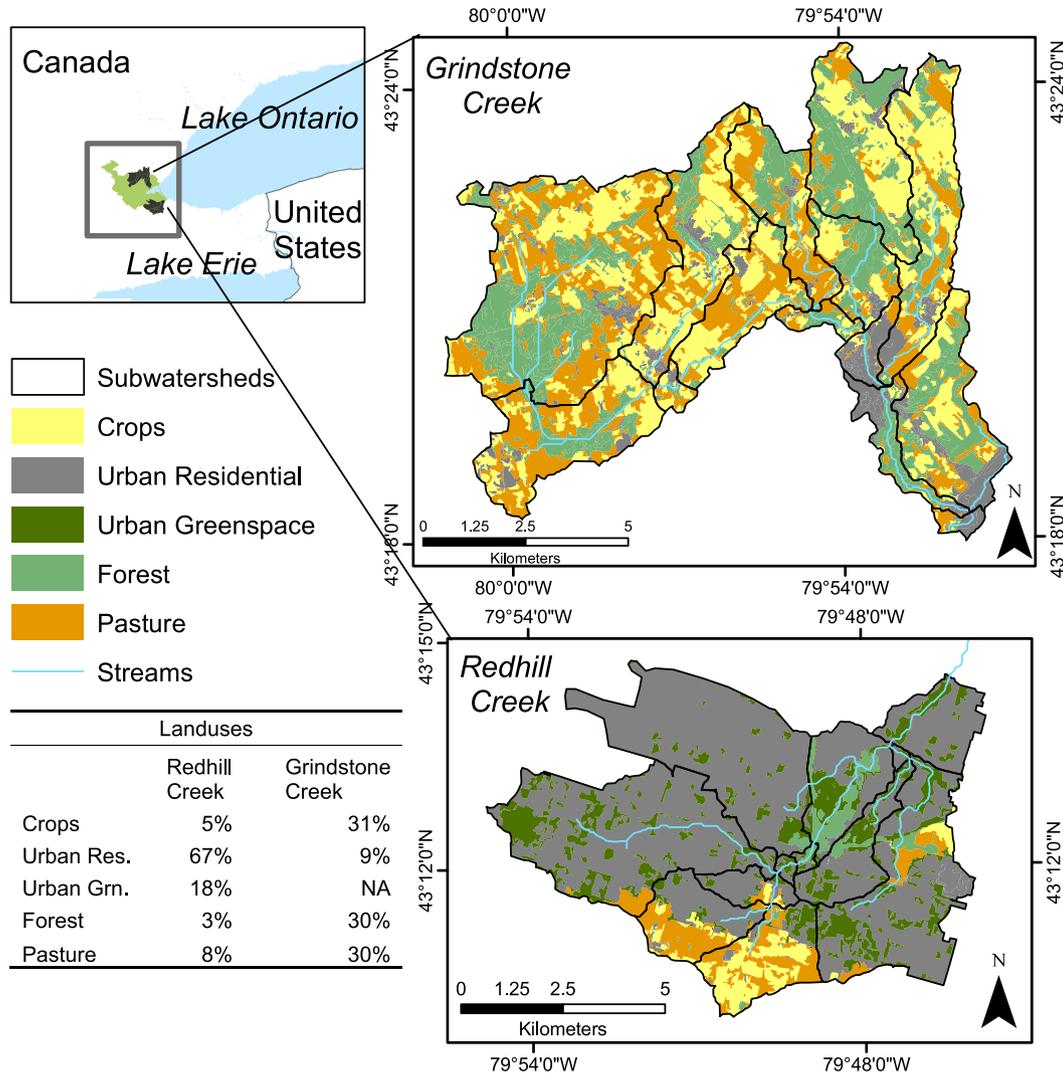


Fig. 1. Map of study site. Reproduced from Wellen et al. (2014a).

response units (HRU) on the basis of land use, soil type, slope, and land management. Surface runoff is computed using a version of the United States' National Resources Conservation Service's Curve Number (CN) methodology. Overland soil erosion for each HRU is estimated using the Modified Universal Soil Loss Equation (MUSLE, Williams, 1995):

$$sed = 11.8(Q_{surf} \times q_{peak} \times area_{HRU})^{0.56} \times K \times C \times P \times LS \times CFRG \quad (1)$$

where sed is the sediment load (metric tons day^{-1}), Q_{surf} is the surface runoff volume ($mm\ day^{-1}$), q_{peak} is the peak surface runoff rate ($m^3\ s^{-1}$), $area_{HRU}$ is the area of the hrU (ha), K is the USLE soil erodibility factor, C is the USLE crop management factor, P is the USLE support practice factor, LS is the USLE topographic factor, and $CFRG$ is the coarse fragment factor. Additional details of our SWAT implementation are contained in the Electronic Supplementary Material (ESM).

3.2. Threshold configuration of SWAT

Theoretical and empirical work provides evidence that watershed systems may be governed by threshold dynamics (Lehmann et al., 2007; Zehe and Sivapalan, 2009; Oswald et al., 2011; Ali et al., 2013). This means that system response can be qualitatively different beyond a threshold level of a relevant variable (e.g., precipitation or catchment storage). This kind of system

response has been investigated in aquatic systems with empirical changepoint models, where model parameters take on different values above and below a critical forcing value (Qian et al., 2003,2004). We accommodate this behavior by incorporating a precipitation threshold into the SWAT model. We assume that above some threshold of precipitation θ_p , an extreme state exists and a subset of the parameter takes on different values than in the normal state. This essentially postulates that watersheds can be characterized by multiple discrete states of response. For this application, we allowed the curve number parameters to vary between states. We averaged the precipitation over 2 days for Redhill Creek and 3 days for Grindstone Creek:

$$CN2\ (Multiplicative\ Effect)_t = CN2_{low}\ for\ Log_{10}(n\text{-day\ averaged\ precipitation} + 1) \leq \theta_p \quad (2a)$$

$$CN2\ (Multiplicative\ Effect)_t = CN2_{high}\ for\ Log_{10}(n\text{-day\ averaged\ precipitation} + 1) > \theta_p \quad (2b)$$

where $CN2\ (Multiplicative\ Effect)_t$ refers to the value of the multiplicative effect for the curve numbers at time t , θ_p refers to the threshold between the two states, $CN2_{low}$ and $CN2_{high}$ refer to the state-specific values of the multiplicative effects applied to

the curve number parameters, and n is equal to 2 for Redhill Creek and 3 for Grindstone Creek. We detail the derivation of the values of n in the ESM.

4. Bayesian inference framework

The Bayesian approach treats statistical inference as a quantitative update of prior beliefs after taking measurements into account. Beliefs are expressed as probability distributions (i.e., random variables), with the central tendency of these distributions corresponding to the degree of certainty that the expected value of the distribution is correct (Gelman et al., 2004). Mathematically, Bayesian inference is founded upon Bayes' Theorem, expressed as:

$$\pi(\theta|Y) = \frac{\pi(\theta)L(Y|\theta)}{\int_0 \pi(\theta)L(Y|\theta)d\theta} \quad (3)$$

where $\pi(\theta)$ represents our prior statements regarding the probability distribution that depicts the existing knowledge of the model parameters (θ), $L(Y|\theta)$ corresponds to the likelihood of observing the data given the different θ values, and $\pi(\theta|Y)$ is the posterior probability that expresses our updated beliefs on the θ values after the existing data from the system are considered. The denominator of Eq. (3) is a constant and acts as a scaling factor.

The typical practice with the SWAT model (and many other distributed watershed models) is to first calibrate the discharge component, followed by the sediment component, and then the nutrient components (Santhi et al., 2001). While this approach has the advantage of accommodating the causal linkages of the various model components, it does not allow any revision of the discharge parameter specification when proceeding to the water quality components. This practice is adopted despite the fact that distributed watershed models are overparameterized, so selecting a single parameter vector of discharge parameters for use in subsequent calibrations may not be optimal. Here, we demonstrate two sequential model updates. The first update involves the hydrological model parameters, while the second one uses the posterior discharge parameters from an earlier calibration exercise as informative priors and calibrates the model to sediment loading data. This sequential update has as an additional advantage that parameter vectors judged acceptable but sub-optimal for modeling discharge are retained and tested against the sediment load data.

Eq. (3) is typically evaluated empirically using a process called Markov chain Monte Carlo (MCMC) sampling. Rather than arriving at analytic expressions of the joint posterior density, samples from this distribution are generated using a (Markov) random walk through the parameter space. In this study, we used the Differential Evolution Adaptive Metropolis Algorithm-ZS (DREAM-ZS) as presented by Laloy and Vrugt (2012). This algorithm is based on the original DREAM algorithm presented by Vrugt et al. (2009). DREAM adapts traditional MCMC approaches to the complex, multi-modal likelihood surfaces typically characterizing deterministic watershed models. DREAM proceeds by running multiple Markov chains and deriving the proposal distribution from the distances between chains in the parameter space. When chains are far apart the proposal distribution is very diffuse, but if the chains converge to a single region of the parameter space the proposal distribution is narrowed accordingly. DREAM-ZS further adapts this approach by sampling from an archive of past states to generate proposal locations in the parameter space. This allows the algorithm to be applied with far fewer Markov chains (Laloy and Vrugt, 2012). We developed a MATLAB interface between DREAM-ZS and SWAT. This interface is available from the corresponding author on request. We note that MCMC software for SWAT also exists for the R computing language (Joseph and Guillaume, 2013). We include additional details about our MCMC sampling in the ESM.

4.1. Discharge calibration

For the model update with the flow data, we used the classic AR(1) residual transformation (Sorooshian and Dracup, 1980; Yang et al., 2007a,b):

$$\begin{aligned} \varepsilon_t &= \rho\varepsilon_{t-1} + \delta_t \\ \delta_t &\sim N(0, \sigma_v^2) \\ \delta_t &= \varepsilon_t - \rho\varepsilon_{t-1}, \end{aligned} \quad (4)$$

where ε_t denotes the residuals, ρ the daily correlation, and δ_t are the daily innovations. We used a Student's t distribution for the likelihood function, as the normal distribution typically does not have thick enough tails (Yang et al., 2007b; Schoups and Vrugt, 2010):

$$\begin{aligned} L(Y|\theta) &= \frac{(1-\rho^2)\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma_v}} \times \left(1 + (1-\rho^2) \times \frac{\varepsilon_1^2}{\nu\sigma_v^2}\right)^{-\frac{\nu+1}{2}} \\ &\times \prod_{t=2}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma_v}} \times \left(1 + \frac{\delta_t^2}{\nu\sigma_v^2}\right)^{-\frac{\nu+1}{2}} \end{aligned} \quad (5)$$

where ν refers to the degrees of freedom and Γ refers to the gamma function. Our preliminary investigations found that 7 was an acceptable number of degrees of freedom. Finally, we opted for a natural logarithm transformation, $Y' = \ln(Y + 1)$, for both measurements and model predictions.

We considered three statistical configurations for discharge parameter estimation and inference. The first statistical formulation (Formulation 1) sets $CN2_{low}$ equal to $CN2_{high}$. This can be thought of as the 'standard' version of SWAT. All prior parameter distributions were uniform over their range. The second statistical formulation (Formulation 2) allows different values for $CN2_{low}$ and $CN2_{high}$. The third statistical formulation was identical to the second one, except that informative prior distributions were used for θ_p , $CN2_{low}$ and $CN2_{high}$ and other parameters where information was available. The calibration vector and all prior parameter distributions are presented in Table 1. We arrived at these subsets after a review of the literature, including the SWAT manual and sensitivity analysis studies (van Griensven et al., 2006; Arabi et al., 2007; Rouhani et al., 2007; Yang et al., 2007a,b; Ekstrand et al., 2010; Neitsch et al., 2011). We employed a one-year spin-up period, while model calibration and validation were based on daily flows from the year ranges 1992 – 1994 and 1995 – 1998, respectively.

4.2. Updating model parameters to simulate sediment loading

Our prior distribution for the second update was the joint posterior distribution obtained from the first update. This joint posterior consisted of the updated discharge parameters and the priors of the sediment parameters, as no information existed during the first calibration to update the prior knowledge of the sediment parameters. In Table 2 we provide the sediment calibration vector employed. In our Bayesian formulation, we selected four parameters to take on common values between the two Creeks: the USLE P factor for agricultural land (USLE_P), the urban washoff coefficient (URBCOEF), the average slope length (SLSUBBSN), and Manning's n for overland flow (OV_N). This reflected our assumption that the dominant aspects of urban and agricultural land use do not vary significantly between these two Creeks. In doing so, we have the advantage of reducing the overall number of free parameters, and thus controlling the parametric uncertainty. Allowing a site-specific calibration of the other parameters reflects our assumption that channel processes and soil properties may vary between the Creeks.

We pooled the calibration data by constructing a joint likelihood between Redhill and Grindstone Creeks. Of the 149 samples

Table 1
SWAT model parameters of the flow discharge submodel included in the calibration vector. Reproduced from Wellen et al. (2014a).

Parameter	Description	Range	Informative prior ^b	Source
CN2	Curve numbers for antecedent moisture condition two (multiplicative effect)	0.5, 1.5	N(1, 0.41)	Schwab et al. (2002), p.74
ALPHA_BF	Baseflow recession constant (1/days)	0.1, 0.99	B(3, 1.15) (Redhill) N(0.64, 0.18) (Grindstone)	Streamflow measurements
SOL_AWC	Fraction of soil water available for plant uptake (multiplicative effect)	0.25, 2.5 (Redhill)	N(1, 0.455)	Assumed minimum and maximum values of 0.01 and 0.85
GW_REVAP	Revap coefficient	0.5, 1.5 (Grindstone)	N(1, 0.455)	–
ESCO	Soil evaporation compensation factor	0.02, 0.2	U(0.02, 0.2)	–
EPCO	Plant transpiration compensation factor	0.1, 0.99	B(3, 1.22)	Expected value of 0.9, signifying a weak ability of lower soil layers to supply evaporative demand of the top layer
GW_DELAY ^a	Ground water delay time (days; multiplicative effect)	0.1, 0.99	U(0.5, 5)	Expected value of 0.9, signifying a strong ability of lower soil layers to supply evaporative demand of the plants
SOL_KSAT	Soil saturated hydraulic conductivity (mm/hr) (multiplicative effect)	0.5, 5	LN(0.1, 15)	–
SNOWCOVMX	Minimum snow water content corresponding to 100% aerial snow coverage (mm)	0.1, 10 (Redhill)	LN(0, 1.15)	Corresponds to a range of one order of magnitude
SMFMX	Snow melt factor on June 31st (mm water/°C above 0.5 °C)	0.5, 1.5 (Grindstone) 1, 40	LN(2.48, 0.35)	Donald et al. (1995)
SMFMN	Snow melt factor on December 31st (mm water/°C above 0.5 °C)	1, 9	N(5.5, 3.1)	Donald (1992), Conetta (2004), Yang et al. (2007a,b), Hu et al. (2007)
SURLAG	Lag time for surface runoff (days)	1, 5	N(3.1, 1.8)	Ibid
ρ	First order residual correlation coefficient for all days	0.5, 10	LN(0, 1.0)	Assumed 1 day was the most likely value and upper end of 95% credible interval was 1 week
σ	Innovation standard deviation for all days	0.1, 0.99	U(0.1, 0.99)	–
CN2 _{Low}	Curve number for moisture condition 2 on low precipitation days (multiplicative effect)	0.002, 2000	G(0.001, 0.001)	–
CN2 _{High}	Curve number for moisture condition 2 on high precipitation days (multiplicative effect)	0.5, 1.5	N(1, 0.41)	Schwab et al. (2002), p.74
CN2 ρ	Correlation of CN2 _{Low} and CN2 _{High}	0.5, 1.5	N(1, 0.41)	Schwab et al. (2002), p.74
θ_p	Threshold of time averaged precipitation switching between curve numbers	–0.99, 0.99	U(–0.99, 0.99)	–
		0.9, 1.4 (Redhill)	N(0.94, 0.025)	Streamflow and precipitation measurements
		0.6, 1.1 (Grindstone)	N(0.78, 0.047)	

^a The base value of ground water delay time was 1.25 days for urban areas, 10 for forested areas, and 5.25 for other areas.

^b N(μ, σ) refers to the normal distribution with mean μ and standard deviation σ ; B(a, b) refers to the beta distribution with shape parameters a and b ; U(l, u) refers to the uniform distribution with lower bound l and upper bound u ; LN(μ, σ) refers to the lognormal distribution with location parameter μ and scale parameter σ ; G(α, β) refers to the gamma distribution with shape parameter α and rate parameter β .

used for the sediment calibration, 132 were taken from Redhill and Grindstone Creek on the same day. We accounted for the correlation of residuals of loads occurring on the same day by using a joint normal distribution for these days. We computed the density of the residuals of the remaining samples assuming normality and independence, so the entire likelihood function is:

$$\begin{aligned}
 L(\mathbf{x}|\theta) &= \prod p(x_{R,G}|\theta) \times \prod p(x_R|\theta) \times \prod p(x_G|\theta) \\
 p(x_R|\theta) &= \frac{1}{\sigma_R \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{x_R^2}{\sigma_R^2}\right) \\
 p(x_G|\theta) &= \frac{1}{\sigma_G \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{x_G^2}{\sigma_G^2}\right) \\
 p(x_{R,G}|\theta) &= \frac{1}{2\pi \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \varepsilon'_{R,G} \Sigma^{-1} \varepsilon_{R,G}\right) \quad (6) \\
 \varepsilon_{R,G} &= [\varepsilon_R \quad \varepsilon_G] \\
 \Sigma &= \begin{bmatrix} \sigma_R^2 & \rho_s \sigma_R \sigma_G \\ \rho_s \sigma_R \sigma_G & \sigma_G^2 \end{bmatrix}
 \end{aligned}$$

where x_R and ε_R denotes an unpaired observation and residual from Redhill Creek, x_G and ε_G denotes an unpaired observation and residual from Grindstone Creek, $x_{R,G}$ and $\varepsilon_{R,G}$ denote a pair of

observations and corresponding residuals from Redhill and Grindstone Creeks, $\sigma_{R,G}^2$ refers to the residual variance at either Redhill or Grindstone Creek, and ρ_s refers to the between-site correlation coefficient of the residuals. The likelihood function in Eq. (6) assumes independence of the residuals in time, an assumption which is reasonable given the sampling interval of our data approximately 1 week. At a lag of 7 days, the correlation coefficients of daily flows for Redhill and Grindstone Creeks were, respectively, $r = 0.06$ and $r = 0.23$. We used informative priors for the sediment submodel parameters, the details of which are presented in Table 2. We used a natural log transformation of the measured and modeled loads prior to calculating the likelihood. We updated the discharge and sediment submodels with data from July 2010 to June 2012, using a six-year spin-up period.

4.3. Ensemble integration of the three formulations

We adopted an ensemble approach to integrate across the different predictions made by the three Formulations. The basic premise of the integration scheme is that the posterior densities

Table 2
SWAT model parameters of the sediment load submodel included in the calibration vector.

Parameter	Description	Range	Informative prior ^a	Source
<i>Independent parameters</i>				
USLE_K	USLE equation soil erodibility (K) factor (0.013 (ton m ² h) * (m ³ ton cm) ⁻¹ ; multiplicative effect)	0.54, 1.5	N(1, 0.48)	OMAFRA 2012
PRF	Peak rate adjustment factor for sediment routing in the main channel	1, 5	Ln(PRF) ~ LN(-1.5, 1.4) (Redhill)	Streamflow measurements
SPCON	Linear parameter for calculating the maximum amount of sediment that can be reentrained during channel sediment routing	0.0001, 0.01	Ln(PRF) ~ LN(-2.07, 1.01) (Grindstone) U(0.0001, 0.01)	–
CH_N	Manning's n value for the main channel (multiplicative effect)	0.7, 3	LN(0, 0.56)	Schwab et al. (2002), p.74
CH_COV1	Channel erodibility factor	0.1, 1.0	B(3.97, 2.54)	Shugar et al., (2007), Kahn and Kostaschuk (2011)
<i>Shared parameters</i>				
USLE_P	USLE equation support practice factor	0.25, 0.75	U(0.25, 0.75)	OMAFRA 2012
URBCOEF	Washoff coefficient for removal of constituents from impervious area (mm ⁻¹)	0.002, 0.39	U(0.002, 0.39)	–
SLSUBBSN	Average slope length (m; multiplicative effect)	0.5, 5	LN(0, 1.15)	Schwab et al. (2002), p.74
OV_N	Manning's n value for overland flow (multiplicative effect)	0.25, 3	LN(0, 0.35)	Schwab et al. (2002), p.74
<i>Likelihood parameters</i>				
ρ_s	First order residual correlation coefficient between Redhill and Grindstone Creek	0.01, 0.90	U(0.01, 0.90)	–
σ_R, σ_G	Residual standard deviation in Redhill and Grindstone Creeks	0.02, 10	G(0.001, 0.001)	–

^a N(μ, σ) refers to the normal distribution with mean μ and standard deviation σ ; B(α, β) refers to the beta distribution with shape parameters α and β ; U(l, u) refers to the uniform distribution with lower bound l and upper bound u ; LN(μ, σ) refers to the lognormal distribution with location parameter μ and scale parameter σ ; G(α, β) refers to the gamma distribution with shape parameter α and rate parameter β .

representing the model predictions can be treated as members in a mixture of densities. Weights were assigned according to the individual performances during the flow update:

$$\begin{aligned}
 \tau_{ij} &= \sqrt{\frac{\sigma_{ij}^2}{1-\rho_{ij}^2}} \\
 w_i &= \frac{\sum_{j=1}^{MC} \tau_{ij}}{MC} \\
 W_i &= \frac{\frac{1}{w_i}}{\sum_{i=1}^l \frac{1}{w_i}} \\
 \overline{Flow} &= \sum_{i=1}^l W_i Flow_i \\
 \overline{Sed} &= \sum_{i=1}^l W_i Sed_i
 \end{aligned} \quad (7)$$

where σ_{ij} and ρ_{ij} refer to the innovation standard deviation and residual correlation coefficient sampled from formulation i and MCMC run j , and τ_{ij} is the residual variance given the innovation variance and correlation coefficient (Prado and West, 2010); MC refers to the total number of MCMC runs sampled from the posteriors; \bar{Y} refers to the average measured flow; l refers to the number of formulations considered in this analysis ($l=3$); and $Flow_i$ and Sed_i are the predictions from the individual formulations weighted by weights W_i to obtain the averaged predictions \overline{Flow} and \overline{Sed} .

4.4. Model evaluation

We assessed the performance of all models using four metrics: the coefficient of determination (r^2), Nash and Sutcliffe's (1970) index of model efficiency (NSE), the relative error as calculated by Arhonditsis and Brett (RE; 2004) and the logarithm of the likelihood function. Following Hong et al. (2005), we assessed the degree of updating of all informative prior distributions using an index developed by Endres and Schindelin (2003), the so-called

delta index, which quantifies the difference in shape of two parameter distributions:

$$\delta_{\theta_i} = \sqrt{\int \left(\pi(\theta_i) \log \frac{2\pi(\theta_i)}{\pi(\theta_i) + \pi(\theta_i|Y)} + \pi(\theta_i|Y) \log \frac{2\pi(\theta_i|Y)}{\pi(\theta_i) + \pi(\theta_i|Y)} \right) d\theta} \quad (8)$$

where $\pi(\theta_i)$ and $\pi(\theta_i|Y)$ represent the marginal prior and posterior distributions of parameter θ_i , respectively. This metric is equal to zero if there is no difference between the two distributions, and equal to $\sqrt{2 \log 2}$, if there is no overlap between the two distributions. All delta index values are presented as percentages of this maximum value. We also assessed the degree of updating by computing (i) the percent difference between prior and posterior median values, and (ii) the percent change in the width of the 95% credible interval.

5. Results

5.1. Model assessment at the basin outlet

5.1.1. Flow discharge submodel

The metrics of fit of the various statistical characterizations of the flow discharge submodel during calibration and validation are presented in Tables ESM1 and ESM2. For Redhill Creek, the NSE ranged from 0.6 to 0.66 and 0.52 to 0.56 during calibration and validation, respectively. For Grindstone Creek, the corresponding NSE ranges were 0.71–0.74 and 0.44–0.56. Fig. ESM-15 presents time series predictions of the ensemble of the flow discharge predictions for the two Creeks.

5.1.2. Sediment submodel

For Redhill Creek, the NSE ranged from 0.13 to 0.17, with Formulation 2 being characterized with the highest NSE. For Grindstone Creek, the NSE ranged from 0.63 to 0.69, with Formulation

3 being characterized with the highest NSE. The NSE of the updated flow submodel during the period 2010–2012 was fairly high (0.68–0.70 for Redhill Creek, 0.62–0.70 for Grindstone Creek), showing that the quality of our flow discharge estimates was maintained during the update of the sediment submodel. Fig. 2 presents the time series predictions of the sediment load submodel at the basin outlets. Fig. 2 indicates no significant differences between the sediment loading predictions made by the different Formulations at the basin outlet.

For the events in which we have measured loads, the mean sediment loading was 33.7 tons day⁻¹ for Redhill Creek and 17.1 tons day⁻¹ for Grindstone Creek, while the modeled sediment loading for those same days was 30 ± 8.8 tons day⁻¹ for Redhill Creek and 18.1 ± 5.2 tons day⁻¹. The average predictions of the models are in agreement with the data measured from the watersheds under study, so we have reason to believe the long term predictions of the model are likely reasonable estimates. These NSE values are within the range presented by Gassman et al. (2007),

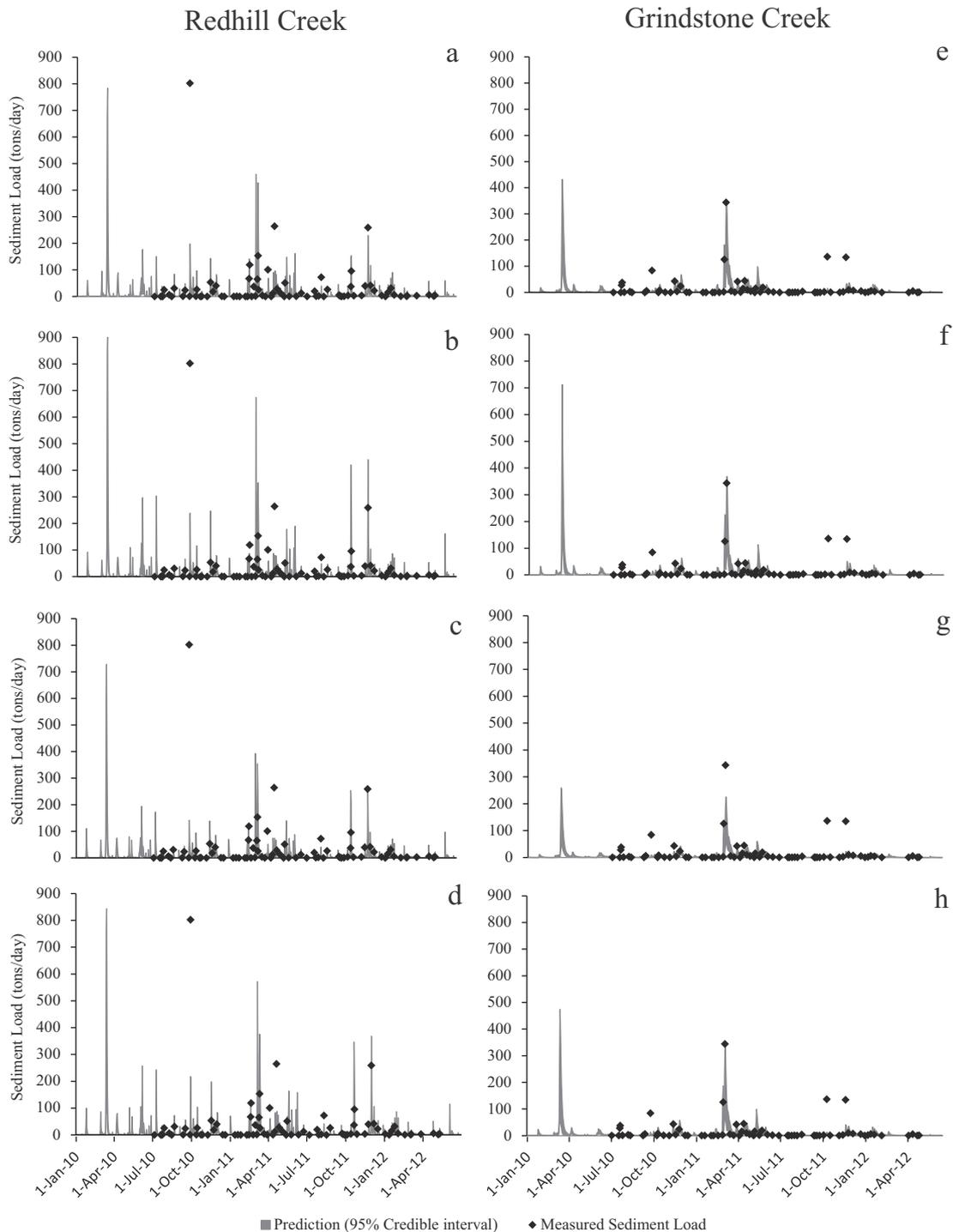


Fig. 2. SWAT calibration to sediment load data for Redhill (a–d) and Grindstone Creeks (e–h). Top to bottom panels represent Formulations 1 (a, e), 2 (b, f), 3 (c, g) and the ensemble prediction (d, h), respectively.

indicating that their fit is comparable to other applications which have informed policy. To evaluate the ability of the model to reproduce seasonal dynamics, we averaged the measured and corresponding ensemble simulated daily loads by month (Fig. ESM-16). The uncertainty of the measured data was represented as the 95% predictive interval of the mean estimated by a single omission jackknife without replacement. Months when a single event dominated the measured average have a high degree of negative skewness (e.g., September in Redhill Creek). The overall seasonal pattern was represented reasonably well.

5.2. The uncertainty of source attributions

In Fig. 3, we present estimates of sediment export to streams for the different landuses in Redhill and Grindstone Creeks across the three formulations examined. We calculate these figures for the entire annual cycle and for the Harbour's growing season

(May–September), during which pulse nutrient input events are most likely to result in phytoplankton blooms. The posterior uncertainty of the sediment yield estimates for cropland was fairly high (see full 95% credible intervals in Figs. ESM 13 and 14). The uncertainty bounds for agricultural land presented in Fig. 3 were calculated by holding the parameters USLE_P and USLE_K constant at their prior modes. This allowed us to draw meaningful inference about sediment sources.

It is clear from Fig. 3 that there is significant uncertainty of the identified sediment yield estimates. There was significant overlap of the 95% credible intervals of the simulated sediment yield of the different landuses in the study (a, b, e, f). With Redhill Creek (c), it is difficult to identify the main source of sediment export to streams even when we take into account the areas of the landuses in the basin. This is despite the fact that the model predictions for sediment export at the basin outlet were reasonably constrained.

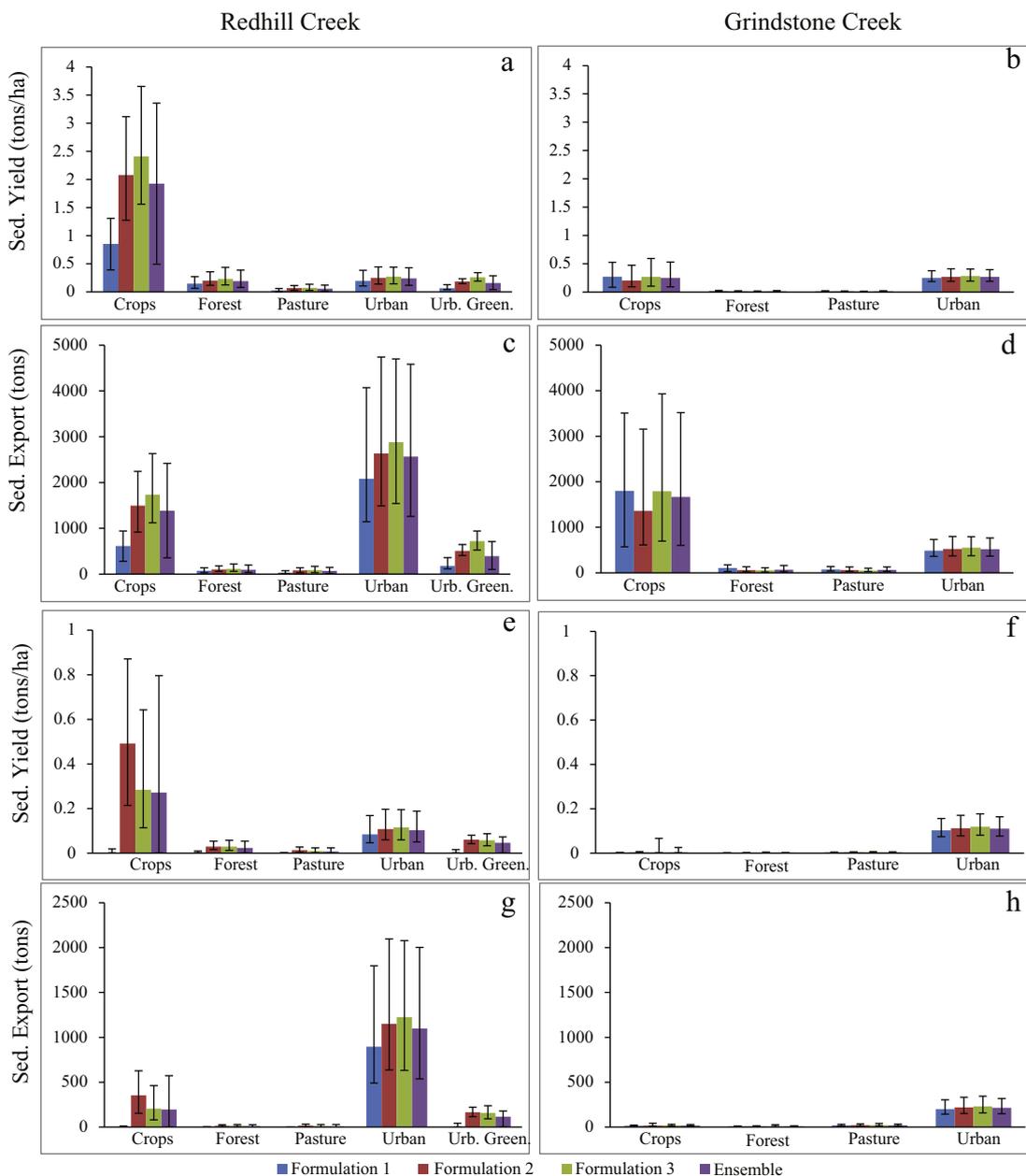


Fig. 3. Estimated sediment source areas across the three formulations by land use for the entire annual cycle (2010–2012; a–d) and for the growing season (May–September, 2010–2012; e–h). Urb. Green. refers to urban greenspace.

The 95% credible intervals of the cropland sediment yield estimates in Fig. 3(a) are non-overlapping for Redhill Creek, Formulations 1 and 2. The disagreement between formulations about the functioning of Redhill Creek becomes more evident when we examine the growing season (Fig. 3(e)). Formulation 1 predicts a near-cessation of surface runoff and consequent cessation of sediment supply to streams during the growing season. Formulations 2 and 3 predict a continued input of sediment to streams from all the land uses. This is somewhat surprising, given the close agreement of the three Formulations when judged against the basin outlet (Fig. 2).

When we examine the predictions obtained from integrating across the ensemble members, we are able to draw some inference about the functioning of the Creeks. Despite the small aerial coverage of the agricultural areas in Redhill Creek (5%) and the urban areas in Grindstone Creek (9%), these areas were responsible for a disproportionate amount of overland sediment export to streams. Cropland was estimated to contribute between 20% and 30% of Redhill Creek's total sediment export to streams (720–3299 tons), while urban areas were estimated to contribute between 17%

and 36% of Grindstone Creek's total sediment export (410–1830 tons). During the growing season, urban residential areas are the main sources of sediment export to both streams, comprising 70–99% of all sediment exported to streams in Redhill Creek (217–1143 tons) and 60–81% of all estimated sediment exported to Grindstone Creek (74–214 tons).

Integrating across the ensemble allows us to use the SWAT model to pose testable hypotheses about how the simulated catchments function. In Fig. 4, we present estimates of sediment yield and streambed sediment storage status for Redhill and Grindstone Creeks at the subbasin scale. While the sediment routing submodel was calibrated, reliable data were simply not available on stream bankfull width and depth. In order to make meaningful predictive statements while acknowledging this substantial uncertainty, we took the entire posterior distribution of the sediment storage for each subbasin (bed storage = upstream sediment in + erosional sediment in – downstream sediment out). We assumed that if the 95% credible interval of the bed storage distribution was non-overlapping with zero, we could make credible statements about whether the reach was gaining or losing sediment during

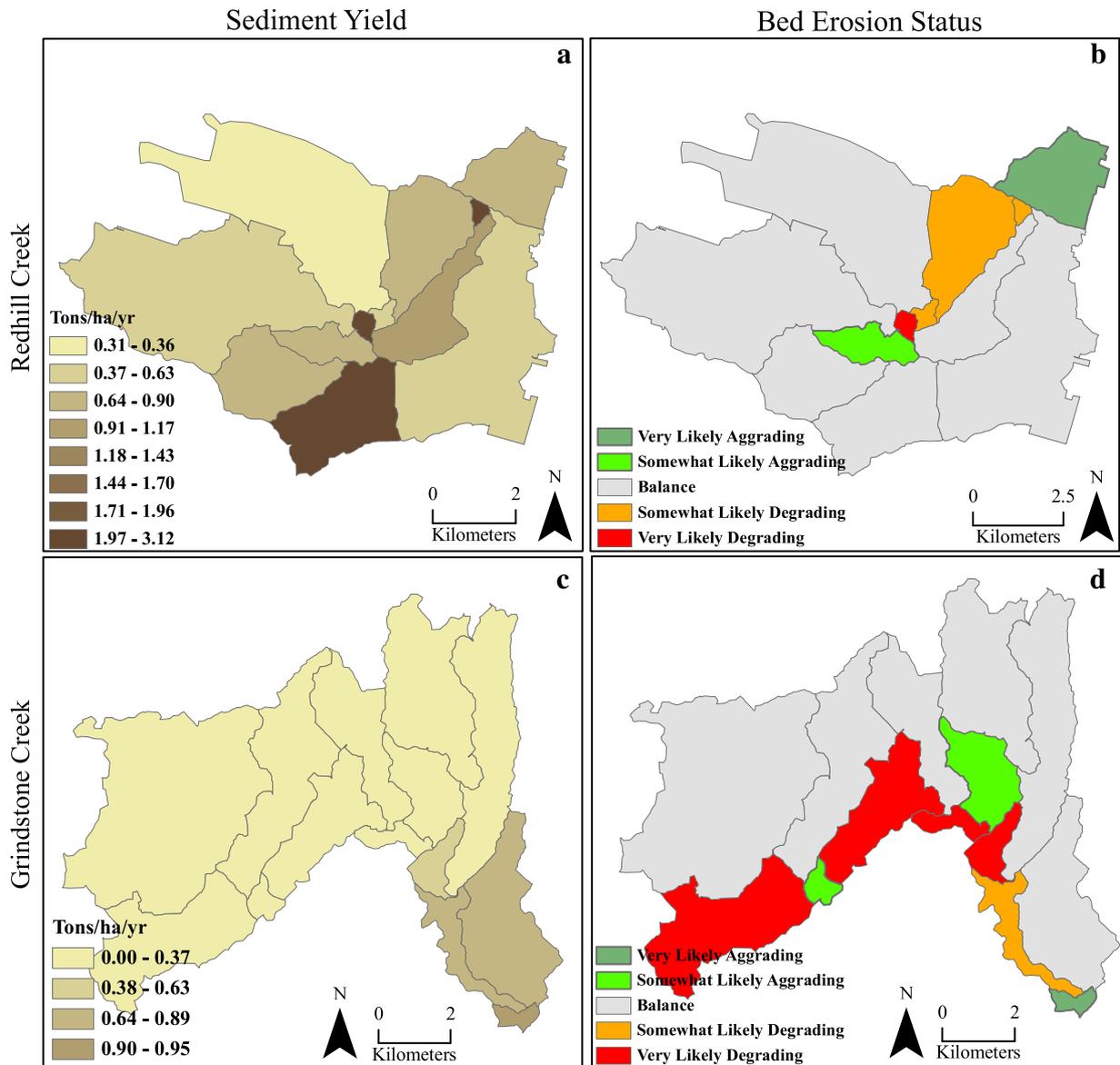


Fig. 4. Estimated sediment yield and bed erosion status for Redhill and Grindstone Creeks.

the period 2010–2012. If the bed storage was positive, we categorized the reach as very likely aggrading, while if the bed storage was negative, we categorized the reach as very likely degrading. If there was overlap with zero, we categorized the reach as likely aggrading or degrading, depending on which side of zero the median of the distribution laid. Some reaches categorized as balanced, as their credible intervals of absolute bed storage were less than 1 ton year^{-1} .

The headwater areas of both Creeks were classified as balanced, while all the reaches losing sediment from their bed are located along the main channel. The final downstream reach was characterized as gaining sediment in both Creeks, reflecting the wider streams and gentler slopes. Note that the subbasin characterized as having the highest class of sediment yield in Redhill Creek's southern end was in balance, indicating that the substantial agricultural sediment mass estimated to be added to the streams in that reach was largely propagated downstream. In Grindstone Creek, there are few reaches which are storing sediment. In particular, the reaches containing most of the urban area towards the mouth of the basin are either at balance or likely degrading, implying that much of the urban sediment added to Grindstone Creek is exported downstream. Despite the substantial uncertainties in many of the model processes, we are able to make some meaningful predictive statements when taking all of these uncertainties into account. The ensemble approach allows us to do so without arbitrarily endorsing one single model formulation.

5.3. Model parameter posteriors

5.3.1. Flow discharge submodel

An examination of the model parameter posteriors shows that they are reasonably well constrained. Figs. 5 and 6 present the 95% credible intervals of the posterior parameter distributions obtained after the update of the flow discharge model. The parameter posteriors were well identified for both Redhill and Grindstone Creeks for all formulations. Using as a reference the first statistical configuration (Formulation 1), Formulation 2 was characterized by lower values of the innovation standard deviation σ and residual correlation ρ and higher values of the likelihood function for Redhill Creek and similar values for Grindstone Creek, showing that (at least) for Redhill Creek there was a better overall fit when accommodating a threshold of runoff generation. We note that there are significant differences of the parameter posteriors between the three formulations.

5.3.2. Sediment submodel

The sediment parameter posteriors were significantly more consistent across the three formulations than the flow parameters were (Fig. 7). This suggests that the differences we present in Fig. 3 regarding the estimated sediment source areas by each Formulation stem from differences in the flow submodel, not the sediment submodel. While the model parameters were not as well identified as those resulting from the update of the flow discharge submodel

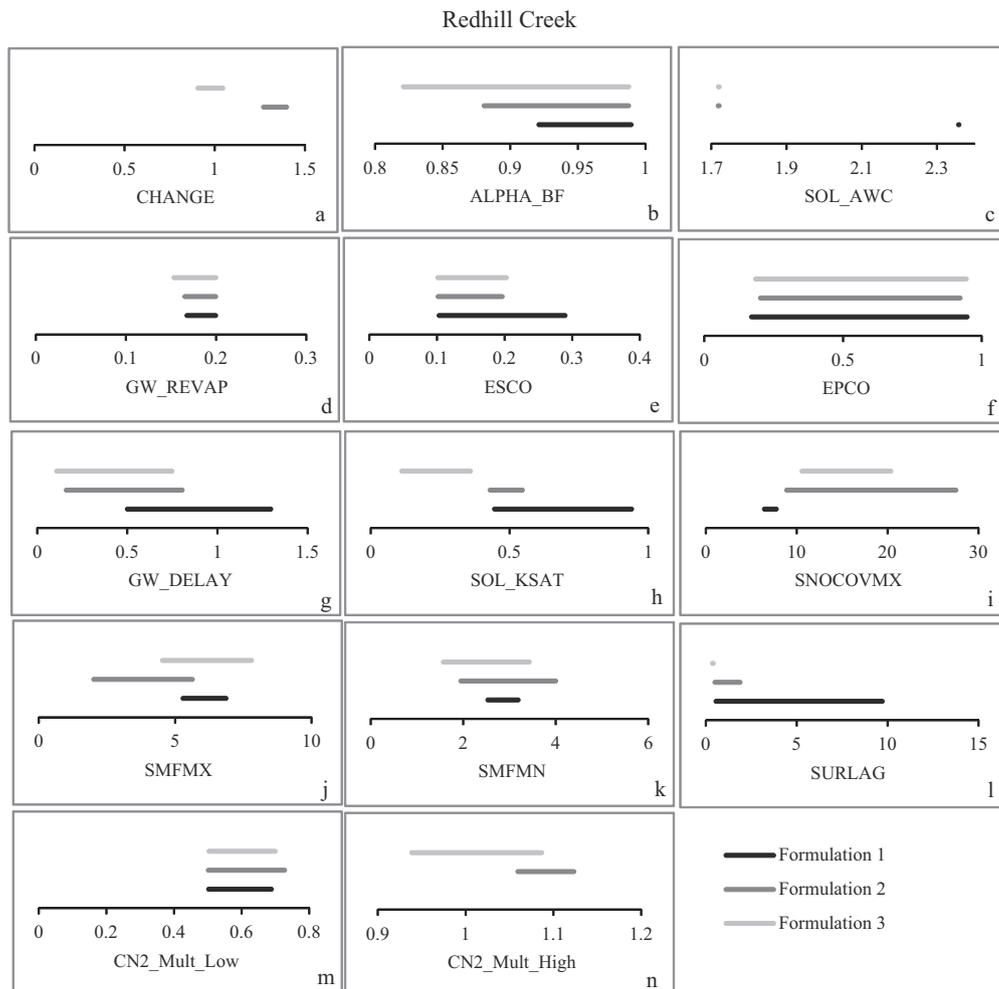


Fig. 5. Posterior 95% credible intervals of the discharge parameters for Redhill Creek after the sediment update.

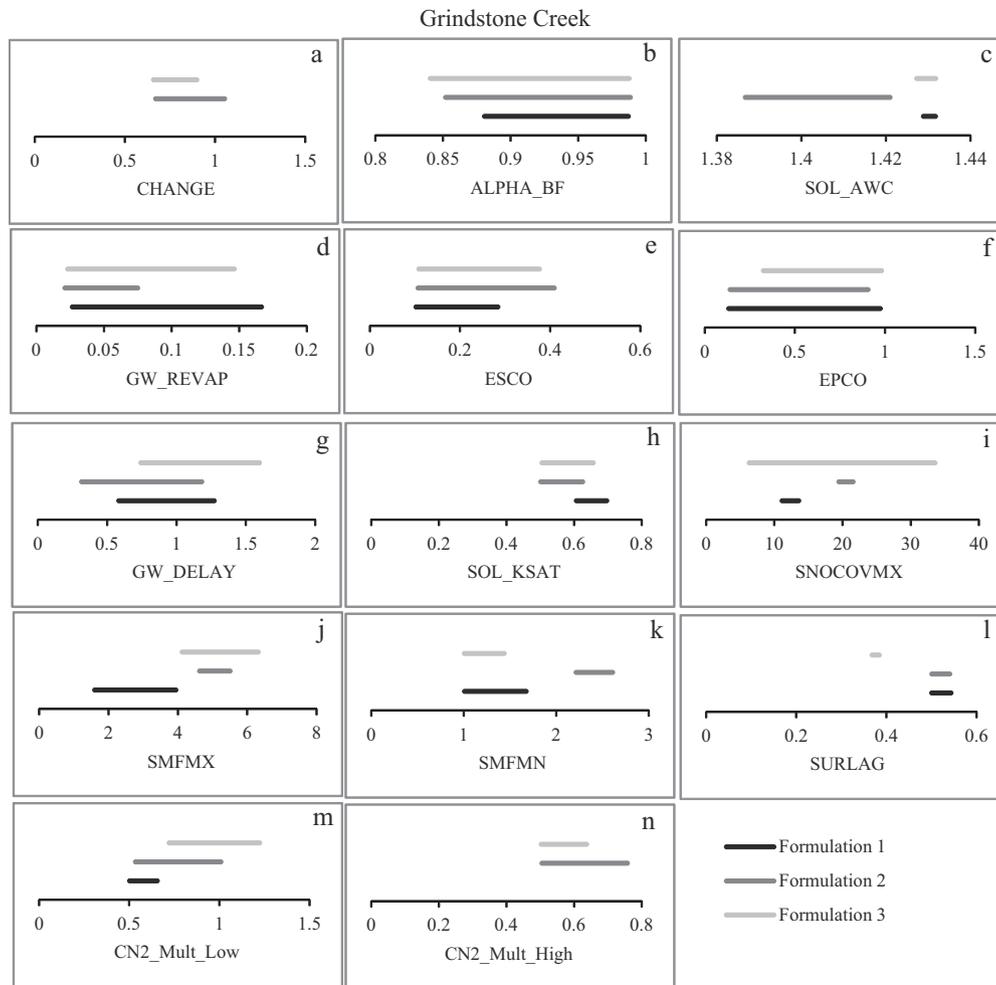


Fig. 6. Posterior 95% credible intervals of the discharge parameters for Grindstone Creek after the sediment update.

alone, we note that the ratios of the posterior mean values relative to the standard deviations were still significantly lower than one (Table ESM 5; Fig. 7). The latter result represents a proxy to assess to what extent our knowledge about the central tendency of the corresponding posteriors compares with the underlying uncertainty.

The likelihood parameters σ_G and σ_R were fairly consistent across the three Formulations, indicating that none of the formulations were strongly favored in both Creeks. However, the lowest values of the residual standard deviation for both Creeks were achieved by Formulation 2 or 3, both of which postulate threshold behavior regarding the runoff generation.

5.4. Propagation of uncertainty

We quantified the change in the flow and discharge parameters after the update to the sediment data, and the results for Formulation 1 are presented in Fig. 8. The results for Formulations 2 and 3 are presented in Tables ESM-3 and ESM-4. The average delta indices were 29% for Formulations 1 and 2, and 18% for Formulation 3, indicating that the information gained about the flow parameters during the second update was approximately similar to that obtained about many sediment parameters (Table ESM-5; average delta index was 31%). However, the credible intervals of the discharge parameters were generally wider after the update with the sediment data – on average by 12%, 5%, and 14% for Formulations 1, 2, and 3, respectively, with maximum increases of 54%, 62%, and 40%. The expected values also shifted by an average of

17%, 25%, and 8%, suggesting that optimal regions of the parameter space for flow discharge may not necessarily be optimal for faithfully reproducing the water quality. Finally, we note that the delta indices for the flow parameters were comparable (and often exceeded) the delta indices for the sediment parameters. This further corroborates our assertions that conditioning flow parameters to discharge may not be sufficient to ensure optimal water quality simulations.

We conducted a post hoc analysis to assess the extent that the uncertainty of the flow submodel controls the uncertainty of the sediment model predictions. We took all the posterior MCMC samples for Formulation 2 but set all the sediment parameters constant at their prior mode. We quantify the uncertainty of the predicted yields as the width of the 95% credible interval divided by the median estimate. In doing so, we were able to estimate the amount of the uncertainty of the sediment yields that stems from the flow submodel (Fig. 9). The percentage of uncertainty in sediment yield predictions in pervious areas stemming from the discharge parameters ranged from 29% to 33% in Redhill Creek and 40–68% in Grindstone Creek. Taken together, these findings call into question the sequential approach to updating integrated watershed models generally practiced by the watershed modeling community (e.g., Santhi et al., 2001).

We also computed the parametric uncertainty of the annual discharge and annual overland flow simulations from the pervious areas of Redhill and Grindstone Creek. The uncertainty was quantified as the width of the 95% credible interval divided by the median estimate. While the discharge simulations showed relatively

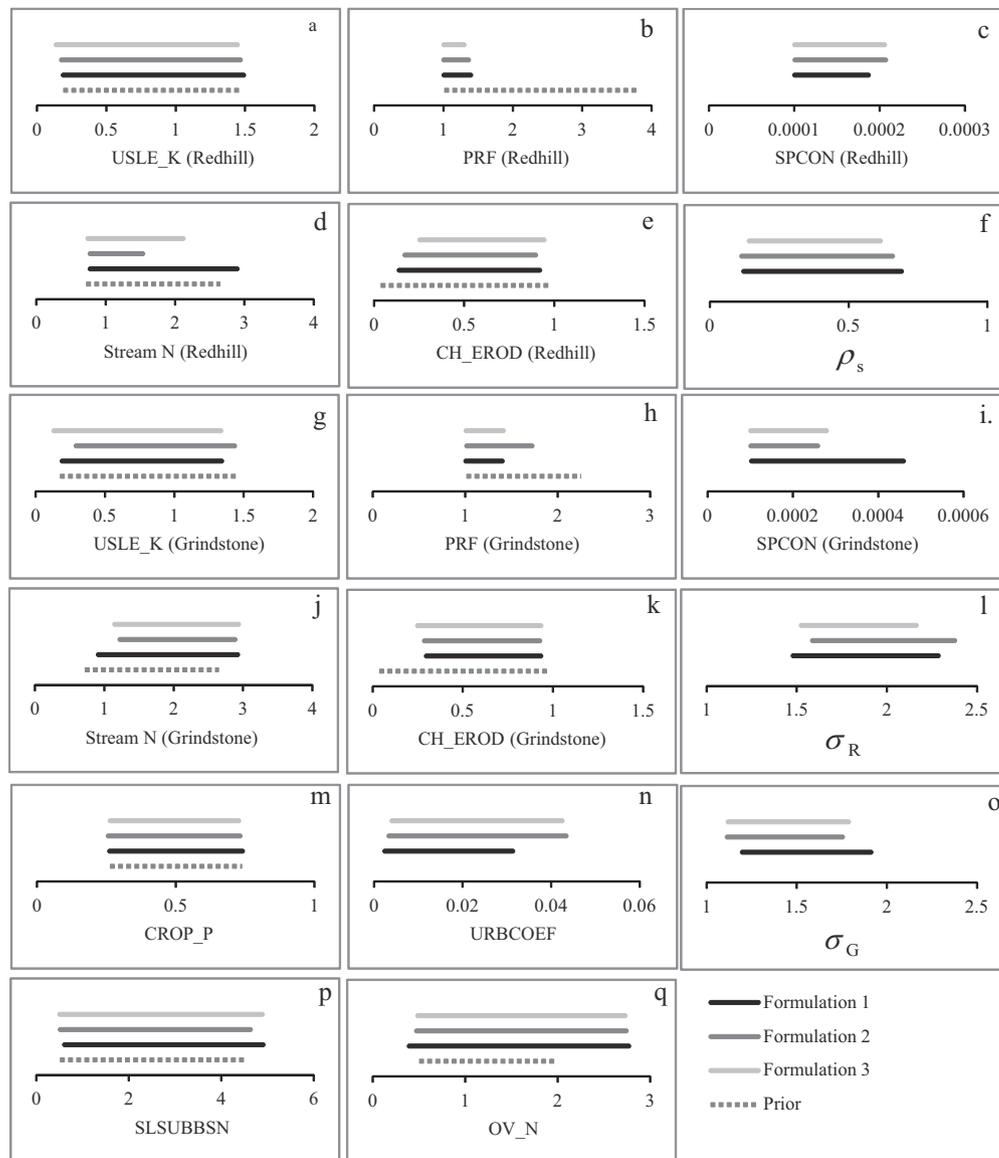


Fig. 7. Prior and posterior 95% credible intervals of the sediment parameters.

low levels of this estimate of parametric uncertainty (0.04 for Redhill Creek and 0.08 for Grindstone Creek), the overland flow estimates showed levels more than ten times as high (1.24 for Redhill Creek and 1.32 for Grindstone Creek). Thus, well constrained discharge estimates can be achieved with fairly uncertain runoff estimates, suggesting that the sole reliance on discharge data does not necessarily constrain the overland flow estimates and thus the derived estimates of sediment flux to streams.

6. Discussion

A handful of studies have conducted a Bayesian update of the discharge submodel of SWAT (Yang et al., 2007a,b), but this is the first study to apply Bayesian updating to one of the water quality components of a process-based, distributed model of the complexity typically used in watershed management. Approaching model calibration as a series of Bayesian updates performed as information becomes available has a number of distinct advantages: (i) the full range of acceptable parameter vectors from previous updates are used at each stage, instead of one well-fitting vector; (ii) the model parameter specification from earlier stages is subject to further refinement; and (iii) the uncertainty associated with the

model predictions and estimated runoff and pollutant source areas can be addressed when the model informs policy.

We also applied the framework for accommodating extreme events first advanced by Wellen et al. (2014a) to the prediction of sediment loading, a water quality variable associated with peak flows. We conducted an ensemble integration of three representations of the threshold which defines an extreme state: one which ignores the extreme state entirely, one which allows the threshold to take on any value, and one which uses measured streamflows to constrain the value of the threshold. Rather than base model predictions on only one of these representations of the watersheds under study, we synthesize the predictions of the entire ensemble using a Bayesian Model Averaging scheme (Ramin et al., 2012).

6.1. Research questions revisited

6.1.1. Can the SWAT model determine sediment sources with confidence when end of basin data is used for calibration?

As we showed in Section 5.2, the annual sediment sources generally could not be determined with confidence in either Redhill Creek or Grindstone Creek. This is despite the fact that the end-of-basin fluxes were reasonably well constrained. SWAT and other

distributed models are often used to estimate the effect of land use change on pollutant sources (including sediment) and end of basin fluxes. If our estimates of source apportionment and predictive uncertainty are typical, our results imply that the magnitude of the effect of land use change would need to be fairly large to be reliably reflected in the model outputs. While more research is needed to determine if the uncertainty amounts we present in Section 5 are typical, we do stress that the fit of our mean estimates of sediment load and discharge are within acceptable and typical limits (Gassman et al., 2007).

6.1.2. How does uncertainty propagate from the discharge submodel to the water quality submodels?

The typical calibration approach advocated for SWAT and other distributed models is sequential. The discharge components are calibrated first, followed by the water quality components (Santhi et al., 2001). However, we found that when the model was updated with the sediment data, the parameters pertaining to flow exhibited significant changes of shape. Some flow parameters had wider 95% credible intervals after the update to the sediment data. We must therefore conclude that the regions of the parameter space optimal for discharge simulation may not be optimal for water quality simulation. More research is needed to determine whether a simultaneous calibration of water quantity and water quality would yield better results than a sequential calibration. We note that a simultaneous update of sediment and flow with a joint likelihood is technically straightforward, though care must be taken to ensure that the correlation structure of the residuals is properly addressed (Ramin and Arhonditsis, 2013).

We also found that the uncertainty of the discharge submodel was responsible for roughly a third of the uncertainty of the sediment load predictions. It is also likely that the disagreement between the three Formulations regarding sediment source areas derives from the disagreement regarding overland flow source areas. This is because the sediment parameters were quite similar across the three Formulations, while the runoff parameters exhibited some disagreement. Gains in the confidence of our ability to predict water quality will probably come by improving our hydro-metric simulations.

6.1.3. Do the estimated sediment sources vary when different calibration approaches are used?

Even with slight variations of the model structure and calibration approach, the magnitudes of the estimated sediment sources varied significantly. Fig. 3 shows that the seasonal cycle of sediment export to streams also varied significantly across the different statistical Formulations. However, Fig. 2 shows that the end-of-basin estimates of sediment loading were nearly identical across the different Formulations. Clearly, the details of model calibration can have a significant effect on the estimated pollutant source attribution without having such an effect on the predictions themselves.

Small variations in end-of-basin model fit can correspond to large variations in estimated source apportionment. This undermines the deterministic paradigm in which SWAT and much distributed modeling is typically conducted within. In this paradigm, an optimal solution is sought and used in scenario analysis. However, solutions which are trivially worse than the optimal could result in significantly different inference regarding pollutant source apportionment. This highlights the need for information on pollutant sources to help constrain model predictions.

6.1.4. Can we combine the knowledge gained from different calibration approaches?

Fig. 3 shows that there is significant disagreement regarding the estimated sediment sources to streams between the three

Formulations. Our ensemble integration approach was able to resolve some of this disagreement without requiring us to endorse any one Formulation. The subsequent subsection discusses the use of ensemble methods in the hydrological sciences in depth.

6.2. Ensemble methods and distributed modeling

The importance of including multiple representations of hydrological systems under study has long been acknowledged in the literature (Beven, 2006; Vrugt and Robinson, 2007). Numerous frameworks have been advocated for integrating predictions from multiple models, including Ensemble Kalman Filters (Vrugt and Robinson, 2007), Particle Filters (Parrish et al., 2012), and BMA (Hoeting et al., 1999; Raftery et al., 2005; Vrugt and Robinson, 2007; Parrish et al., 2012; Ramin et al., 2012). This body of literature has developed promising techniques. However, the authors tend to use relatively simple, lumped conceptual models that mainly offer proofs of the concept rather than solutions to real management problems. A very small number of studies have applied ensemble methods to the SWAT model (Zhang et al., 2011; Strauch et al., 2012). The present study offers a first ensemble approach to the water quality predictions of a distributed, non-point source watershed model.

In this paper, we have shown that an ensemble approach can be an effective way to address critical concerns of equifinality raised in the distributed modeling literature. Beven (2006) argues that model structural ambiguity can lead to a situation where contradictory mechanistic foundations can predict different estimates of internal fluxes and source areas, but similar end-of-basin fluxes. The subsequent experience of watershed modelers demonstrates this assertion to be true. For instance, Easton et al. (2008) introduced the soil topographic wetness index as an alternative way of generating hydrological response units in SWAT. This strategy is in contrast to the standard SWAT hydrological response unit delineation, where land use and soil type are assumed to control the generation of runoff. Paradoxically, while the predicted runoff areas were quite different than those obtained with the standard SWAT hydrological response units, the basin discharge predictions were largely unchanged. The authors were able to conclude their study in favor of their topographic wetness index approach by comparing measurements against the predictions of water table depth. However, this variable will not be available in all watersheds and in all scales. Barring a judgment based solely on theoretical soundness, predictions from both watershed delineations would need to be taken into account when using the models to arrive at an understanding of watershed functioning. White et al. (2011) present a case where the SWAT model code was altered so that runoff was calculated with a physical water balance, instead of the empirical curve number approach standard to SWAT modeling. While the water balance has presumably a better theoretical foundation relative to the original curve numbers, it did not quantitatively improve discharge estimates.

Our experience is similar in that we found that different SWAT parameter specifications led to different estimated runoff and sediment generation areas, but very similar end-of-basin time series of discharge and sediment. An encouraging result from the present study is that when integrating across the three parameterizations, we are able to make some meaningful predictive statements. For instance, our work suggests that urban areas are the main source of sediment to Redhill and Grindstone Creeks during the growing season, despite the substantial differences among the sediment yields estimated by the three formulations.

A critical decision when synthesizing the predictions of multiple distributed models is the selection of an appropriate averaging scheme. We chose to use a Bayesian model averaging scheme akin to that presented by Ramin et al. (2012), where the weights of the

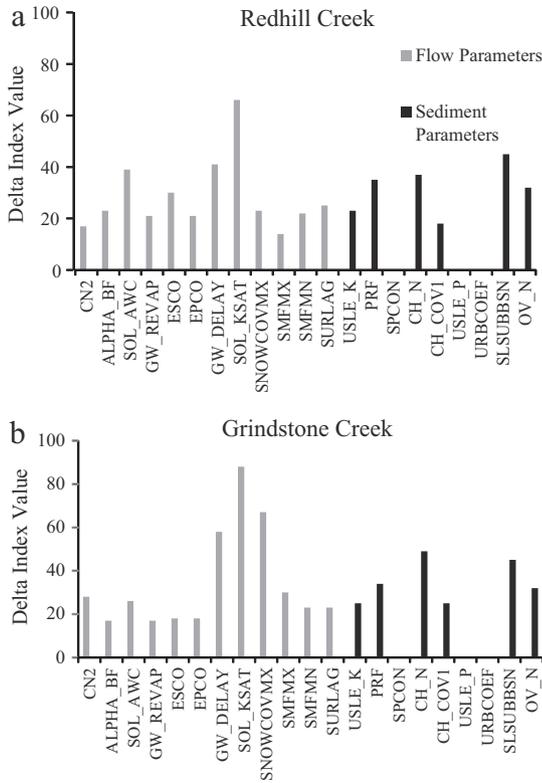


Fig. 8. Delta index of flow and sediment parameters after the sediment update for (a) Redhill and (b) Grindstone Creeks.

individual ensemble members are calculated post hoc as functions of the residual variance of the models. This approach allowed to credibly integrating the predictions of the ensemble members, while accounting for parametric and structural uncertainty. This is philosophically similar to the approach of Fortin et al. (2006), who advocated that the ensemble weights be estimated separately for each ensemble member. This is in contrast to the best member approach of Roulston and Smith (2003), where the ensemble weights were all the same and derived from the best ensemble member. Future work seeking to apply ensemble methods to distributed watershed models will benefit from the innovative methodological work being conducted with lumped models. Of particular interest for our approach are methods which allow the ensemble weights to vary with time. Raftery et al. (2005) used a sliding window, which based the weights on a criterion of

performance evaluated during a historical period. Likewise, Duan et al. (2007) divided the training period into different flow intervals and used separate weights for each interval. Coupling either approach to the framework presented here could emphasize the predictions of Formulations 2 and 3 during episodic events, while relying on Formulation 1 for smaller events or baseflow, should that formulation prove superior in such situations.

6.3. On the value of complex, distributed models in the hydrological sciences

Distributed models are a key tool in watershed management and hydrological science (Rode et al., 2010). They will likely retain this place of importance, as no other tool is able to estimate the impact of land use changes on water quality at daily or monthly timescales. While this paper had presented evidence which calls their use into question, we do not believe that models such as SWAT are unreliable or ‘useless arithmetic.’ We simply advocate an improvement in the common practices of distributed model use, beginning with model validation.

The typical criterion for judging whether a model is adequate is the degree of fit (Rykiel, 1996), despite reservations that such a test is philosophically inconclusive (Oreskes et al., 1994). In the case of distributed models, the degree of fit is typically judged only at the basin outlet (Gassman et al., 2007). Skeptical voices in the watershed modeling literature have cautioned that this may not be sufficient to decide whether a model satisfactorily reproduces the relationship between diffuse upland fluxes and basin outlet fluxes (Beven, 2006). We show in this paper that even with a nearly identical model structure, small changes in model parameterization may lead to divergent source apportionments but nearly identical estimated end-of-basin fluxes. Despite our use of informative priors and nearly 150 measurements, we were not able to adequately constrain the sediment source apportionment. A reasonable hypothesis as to why this is the case is that the measured loading or concentration at the basin outlet contains too many ‘mixed signals’ to constrain the model predictions. In light of this, we recommend that future studies consider integrating three types of additional information: information in space, information of additional model fluxes, and tracer information to help constrain model sources.

Constraining simulated in-stream fluxes of water and waterborne constituents may be improved by incorporating information from a greater variety of locations in space. This has been the strategy of the USGS’ SPARROW regression model (Alexander et al., 2004). Unlike the estimated pollutant export rates we present here, the export rates estimated by SPARROW tend to be

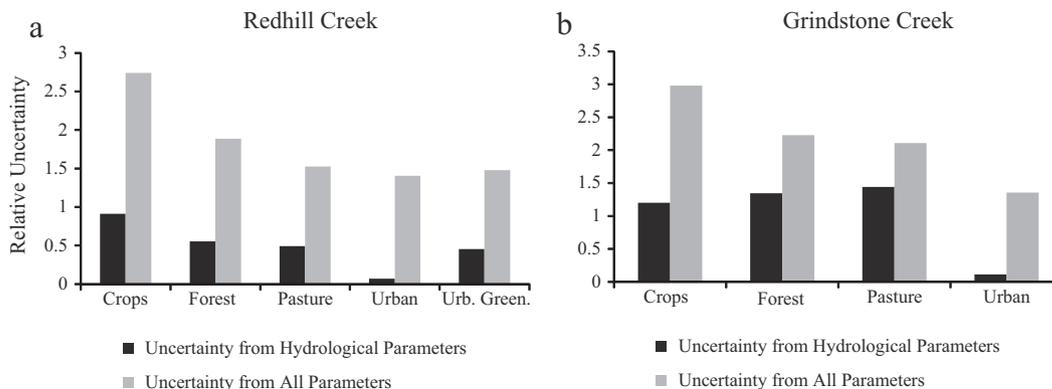


Fig. 9. Proportion of uncertainty of sediment yield due to parametric uncertainty of the hydrological submodel. Relative uncertainty was computed as the width of the 95% credible interval of the sediment yield of each landuse divided by the median estimate.

reasonably well constrained (Wellen et al., 2012, 2014b). This is most likely owing to a fairly large amount of information in space (24 water quality monitoring stations spanning orders of magnitude of size) as well as a simpler model structure with far fewer parameters (less than 10). This allows export rates to be estimated from smaller watersheds where export is dominated by one or two sources and applied to larger, 'mixed signal' watersheds. More research is needed to determine whether more information in space can improve the source attribution of complex models such as SWAT. We recommend future work focus on the use of nested basin sampling structures coupled with Bayesian hierarchical inference frameworks to accomplish this (Zhang and Arhonditsis, 2009).

Complex models such as SWAT simulate all the major fluxes of the hydrological cycle (e.g., evapotranspiration, groundwater flow, overland flow, return flow), yet the common practice is to calibrate only to their sum (streamflow). By incorporating additional hydrological fluxes such as evapotranspiration into the model calibration, it might be possible to arrive at more constrained estimates of the other hydrological fluxes, such as overland flow. This may constrain the estimates of pollutant export. Model calibration may be aided by incorporating empirical information about the sources of water, sediment, and nutrients. There are a variety of techniques developed for making such inference, including end member mixing analysis (EMMA, Burns et al., 2001), sediment fingerprinting (Davis and Fox, 2009), and isotope analysis of some dissolved nutrients such as phosphate (McLaughlin et al., 2006). Empirical source attributions could be used to constrain model predictions by calibrating the model's summary statistics to statistics of source estimates (Csillery et al., 2010).

In conclusion, we conducted the first Bayesian calibration of the water quality components of SWAT. We show that end-of-basin information is not sufficient to constrain the source attribution of suspended sediment. We also presented evidence that more than a third of the uncertainty of SWAT model sediment load predictions may stem from the discharge submodel. We presented a case where three different conceptualizations of the watershed under study produce fairly similar end of basin predictions but divergent source apportionments. In such conditions, an ensemble approach can offer meaningful statements about the watershed functioning while accounting for the different sources of uncertainty. We conclude by suggesting a number of improvements to the current practice of calibrating distributed models to fluxes at the basin outlet. These improvements include incorporating additional information in space, information of additional model fluxes, and tracer information to help constrain model sources.

Acknowledgements

This project has received funding support from the Ontario Ministry of the Environment (Canada Ontario Grant Agreement 091211). Such support does not indicate endorsement by the Ministry of the contents of this material. Christopher Wellen has also received support from Ontario Graduate Scholarships. The authors wish to thank Jasper Vrugt for generously sharing the code for the DREAM(ZS) sampler. The authors also wish to thank Tom Arsenault and the Water Survey of Canada for providing the hourly streamflow data used to calculate event loads.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jhydrol.2014.10.007>.

References

- Alexander, R.B., Smith, R.A., Schwarz, G.E., 2004. Estimates of diffuse total phosphorus sources in surface waters of the United States using a spatially referenced watershed model. *Water Sci. Tech.* 49, 1–10.
- Ali, G., Oswald, C.J., Spence, C., Cammeraat, E.L.H., McGuire, K.J., Meixner, T., Reaney, S.M., 2013. Towards a unified threshold-based hydrological theory: necessary components and recurring challenges. *Hydrol. Process.* 27, 313–318. <http://dx.doi.org/10.1002/hyp.9560>.
- Arabi, M., Govindaraju, R.S., Hantush, M.M., 2007. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. *J. Hydrol.* 333, 459–471. <http://dx.doi.org/10.1016/j.jhydrol.2006.09.012>.
- Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modelling. *Mar. Ecol.-Prog. Ser.* 271, 13–26.
- Arnold, J.G., Srinivasan, R., Mutiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment. Part I: model development. *J. Am. Water Resour. As.* 34 (1), 73–89.
- Beven, K., 2006. A manifesto for the equifinality thesis. *J. Hydrol.* 320 (1–2), 18–36. <http://dx.doi.org/10.1016/j.jhydrol.2005.07.007>.
- Burns, D., McDonnell, J., Hooper, R., Peters, N., Freer, J., Kendall, C., Beven, K., 2001. Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA). *Hydrol. Process.* 15, 1903–1924. <http://dx.doi.org/10.1002/hyp.246>.
- Cheng, H., Ouyang, W., Hao, F., Ren, X., Yang, S., 2007. The non-point source pollution in livestock-breeding areas of the Heihe River basin in Yellow River. *Stoch. Environ. Res. Risk Assess.* 21, 213–221. <http://dx.doi.org/10.1007/s00477-006-0057-2>.
- Conetta, M., 2004. Snow disposal sites, conceptual designs Part A – Snow meltwater characteristics and treatment technologies. Report Submitted to the City of Toronto, October 29, 2004, unpublished.
- Csillery, K., Blum, M., Gaggiotti, O., Francois, O., 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* 25 (7), 410–418. <http://dx.doi.org/10.1016/j.tree.2010.04.001>.
- Davis, C., Fox, J., 2009. Sediment fingerprinting: review of the method and future improvements for allocating nonpoint source pollution. *J. Environ. Eng.* 135 (7), 490–504. [http://dx.doi.org/10.1061/\(ASCE\)0733-9372\(2009\)135:7\(490\)](http://dx.doi.org/10.1061/(ASCE)0733-9372(2009)135:7(490)).
- Donald, J.R., 1992. Snow Cover Depletion Curves and Satellite Snow Cover Estimates for Snowmelt Runoff Modelling. Ph.D. Thesis, University of Waterloo, ON, Canada, pp. 232.
- Donald, J.R., Souliis, E.D., Kouwen, N., Pietroniro, A., 1995. A land cover-based snow cover representation for distributed hydrologic models. *Water Resour. Res.* 31 (4), 995–1009.
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* 30, 1371–1386.
- Easton, Z.M., Fuka, D.R., Walter, M.T., Cowan, D.M., Schneiderman, E.M., Steenhuis, T.S., 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *J. Hydrol.* 348 (3), 279–291.
- Ekstrand, S., Wallenberg, P., Djodjic, F., 2010. Process based modelling of phosphorus losses from arable land. *Ambio* 39, 100–115. <http://dx.doi.org/10.1007/s13280-010-0016-5>.
- Endres, D.M., Schindelin, J.E., 2003. A new metric for probability distributions. *IEEE T. Inform. Theory* 49, 1858–1860. <http://dx.doi.org/10.1109/tit.2003.813506>.
- Fortin, V., Favre, A., Said, M., 2006. Probabilistic forecasting from ensemble prediction systems: improving upon the best-member method by using a different weight and dressing kernel for each member. *Quart. J. Roy. Meteorol. Soc.* 132 (617), 1349–1369.
- Freni, G., Mannina, G., 2010. Uncertainty in water quality modelling: The applicability of Variance Decomposition Approach. *J. Hydrol.* 394, 324–333. <http://dx.doi.org/10.1016/j.jhydrol.2010.09.006>.
- Freni, G., Mannina, G., Viviani, G., 2009. Uncertainty assessment of an integrated urban drainage model. *J. Hydrol.* 373, 392–404. <http://dx.doi.org/10.1016/j.jhydrol.2009.04.037>.
- Gassman, P., Reyes, M., Green, C., Arnold, J., 2007. The soil and water assessment tool: historical development, applications, and future research directions. *Trans. ASABE* 50 (4), 1211–1250.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*. Chapman & Hall/CRC Press, Boca Raton, FL.
- Gudimov, A., Stremilov, S., Ramin, M., Arhonditsis, G.B., 2010. Eutrophication risk assessment in Hamilton Harbour: system analysis and evaluation of nutrient loading scenarios. *J. Gt. Lakes Res.* 36 (3), 520–539. <http://dx.doi.org/10.1016/j.jglr.2010.04.001>.
- Gudimov, A., Ramin, M., Labencki, T., Wellen, C., Shelar, M., Shimoda, Y., Boyd, D., Arhonditsis, G.B., 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions: a modeling analysis of the "ecological unknowns". *J. Gt. Lakes Res.* 37 (3), 494–506. <http://dx.doi.org/10.1016/j.jglr.2011.06.006>.
- Hiriart-Baer, V.P., Milne, J., Charlton, M.N., 2009. Water quality trends in Hamilton Harbour: two decades of change in nutrients and chlorophyll a. *J. Gt. Lakes Res.* 35 (2), 293–301. <http://dx.doi.org/10.1016/j.jglr.2008.12.007>.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. *Statist. Sci.*, 382–401.
- Hong, B.G., Strawderman, R.L., Swaney, D.P., Weinstein, D.A., 2005. Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six. *Water Resour. Res.* 41, W03007. <http://dx.doi.org/10.1029/2004wr003551>.

- Hu, X., McIsaac, G.F., David, M.B., Louwers, C.A.L., 2007. Modeling riverine nitrate export from and East-Central Illinois watershed using SWAT. *J. Environ. Qual.* 36, 996–1005.
- Joseph, J., Guillaume, J., 2013. Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environ. Modell. Software* 46, 292–298.
- Kahn, I., Kostaschuk, R., 2011. Erodibility of cohesive glacial till bed sediments in urban stream channel systems. *Can. J. Civ. Eng.* 38, 1363–1372. <http://dx.doi.org/10.1139/L11-099>.
- Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high performance computing. *Water Resour. Res.* 48, W01526, <http://dx.doi.org/10.1029/2011WR010608>.
- Lehmann, P., Hinz, C., Mcgrath, G., Tromp-Van Meerveld, H.J., Mcdonnell, J.J., 2007. Rainfall threshold for hillslope outflow: an emergent property of flow pathway connectivity. *Hydrol. Earth Syst. Sci.* 11, 1047–1063.
- Long, T., Boyd, D., Wellen, C.C., Arhonditsis, G.B., 2014. Characterization of water quality dynamics in the urban and agricultural watersheds of Hamilton Harbour following an intensive two year event-based monitoring program in Hamilton and Burlington, Ontario, Canada. *J. Gt. Lakes Res.* <http://dx.doi.org/10.1016/j.jglr.2014.09.017>.
- McLaughlin, K., Kendall, C., Silva, S., Young, M., Paytan, A., 2006. Phosphate oxygen isotope ratios as a tracer for sources and cycling of phosphate in North San Francisco Bay, California. *J. Geophys. Res.* 111, G03003, <http://dx.doi.org/10.1029/2005JG000079>.
- Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloglu, I., DePinto, J.V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., LaPorte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., Zagorski, M.A., 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *PNAS* 110 (16), 6448–6452. <http://dx.doi.org/10.1073/pnas.121600>.
- Moradkhani, H., Hsu, K., Hong, Y., Sorooshian, S., 2006. Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting. *Geophys. Res. Lett.* 33 (12), L12401.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part 1 – A discussion of principles. *J. Hydrol.* 10, 282–290.
- Nasr, A., Bruen, M., Jordan, P., Moles, R., Kiely, G., Byrne, P., 2007. A comparison of SWAT, HSPF and SHETRAN/GOPC for modelling phosphorus export from three catchments in Ireland. *Water Res.* 41, 1065–1073, <http://dx.doi.org/10.1016/j.watres.2006.11.026>.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool Theoretical Documentation, Version 2009. Texas Water Resources Institute Technical Report No. 406. Texas A&M University System, College Station, Texas. <<http://twri.tamu.edu/reports/2011/tr406.pdf>> (retrieved 17.01.13).
- Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263 (5147), 641–646.
- Oswald, C.J., Richardson, M.C., Branfiren, B.A., 2011. Water storage dynamics and runoff response of a boreal Shield headwater catchment. *Hydrol. Process.* 25, 3042–3060, <http://dx.doi.org/10.1002/hyp.803>.
- Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. *Water Resour. Res.* 48 (3).
- Prado, R., West, M., 2010. Time Series: Modeling, Computation, and Inference. CRC Press, Boca Raton, FL, pp. 353.
- Qian, S., King, R., Richardson, C., 2003. Two statistical methods for the detection of environmental thresholds. *Ecol. Model.* 166 (1), 87–97.
- Qian, S., Pan, Y., King, R., 2004. Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data. *Ecol. Ind.* 4 (1), 29–37.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133 (5), 1155–1174.
- Ramin, M., Arhonditsis, G., 2013. Bayesian calibration of mathematical models: optimization of model structure and examination of the role of process error covariance. *Ecol. Inform.* 18, 107.
- Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of numerical modelling and Bayesian analysis for setting water quality criteria in Hamilton Harbour, Ontario, Canada. *Environ. Modell. Softw.* 26 (4), 337–353, <http://dx.doi.org/10.1016/j.envsoft.2010.08.006>.
- Ramin, M., Labencki, T., Boyd, D., Trolle, D., Arhonditsis, G.B., 2012. A Bayesian synthesis of predictions from different models for setting water quality criteria. *Ecol. Model.* 242, 127–145.
- Rode, M., Arhonditsis, G., Balin, D., Kebede, T., Krysanova, V., van Griensven, A., van der Zee, S.E.A.T.M., 2010. New challenges in integrated water quality modelling. *Hydrol. Process.* 24, 3447–3461, <http://dx.doi.org/10.1002/hyp.7766>.
- Rouhani, H., Willems, P., Wyseure, G., Feyen, J., 2007. Parameter estimation in semi-distributed hydrological catchment modelling using a multi-criteria objective function. *Hydrol. Process.* 21, 2998–3008, <http://dx.doi.org/10.1002/hyp.6527>.
- Roulston, M., Smith, L., 2003. Combining dynamical and statistical ensembles. *Tellus* 55A, 16–30.
- Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecol. Model.* 90 (3), 229–244.
- Santhi, C., Arnold, J.G., Williams, J.R., Dugas, W.A., Hauck, L., 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. Am. Water Resour. Assoc.* 37 (5), 1169–1188.
- Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.* 46, W10531, <http://dx.doi.org/10.1029/2009WR008933>.
- Schwab, G.O., Fangmeier, D.D., Elliott, W.J., Frevert, R.K., 2002. Soil and Water Conservation Engineering, fourth ed. Jon Wiley and Sons, Toronto, 507 pp.
- Setegn, S., Dargahi, B., Srinivasan, R., Melesse, A., 2010. Modeling of sediment yield from Anjeni-gauged watershed, Ethiopia using SWAT model. *J. Am. Water Resour. Assoc.* 1–13, <http://dx.doi.org/10.1111/j.1752-1688.2010.00431.x>.
- Shugar, D., Kostaschuk, R., Ashmore, P., Desloges, J., Burge, L., 2007. In situ jet-testing of the erosional resistance of cohesive streambeds. *Can. J. Civ. Eng.* 34, 1192–1195, <http://dx.doi.org/10.1139/L07-024>.
- Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resour. Res.* 16 (2), 430–442, <http://dx.doi.org/10.1029/WR016i02p00430>.
- Strauch, M., Bernhofer, C., Koide, S., Volk, M., Lorz, C., Makeschin, F., 2012. Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. *J. Hydrol.* 414–415, 413–424, <http://dx.doi.org/10.1016/j.jhydrol.2011.11.014>.
- Talebizadeh, M., Morid, S., Ayyoubzadeh, S., Ghasemzadeh, M., 2010. Uncertainty analysis in sediment load modeling using ANN and SWAT model. *Water Resour. Manage* 24, 1747–1761, <http://dx.doi.org/10.1007/s11269-009-9522-2>.
- van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. *J. Hydrol.* 324 (1–4), 10–23, <http://dx.doi.org/10.1016/j.jhydrol.2005.09.008>.
- Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* 43 (1), W01411.
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Higdon, D., Robinson, B.A., Hyman, J.M., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Int. J. Nonlin. Sci. Num.* 10, 271–288, 4028, 4029, 4033.
- Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2012. A Bayesian methodological framework for accommodating interannual variability of nutrient loading with the SPARROW model. *Water Resour. Res.* 48, W10505, <http://dx.doi.org/10.1029/2012WR011821>.
- Wellen, C.C., Labencki, T., Boyd, D., Arhonditsis, G.B., 2014a. Accommodating environmental thresholds and extreme events in hydrological models: a Bayesian approach. *J. Gt. Lakes Res.* <http://dx.doi.org/10.1016/j.jglr.2014.04.002>.
- Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2014b. Application of the SPARROW model in watersheds with limited information: a Bayesian assessment of the model uncertainty and the value of additional monitoring. *Hydrol. Process.* 28, 1260–1283, <http://dx.doi.org/10.1002/hyp.9614>.
- White, E.D., Easton, Z.M., Fuka, D.R., Collick, A.S., Adgo, E., McCartney, M., Awulachew, S.B., Selassie, Y.G., Steenhuis, T.S., 2011. Development and application of a physically based landscape water balance in the SWAT model. *Hydrol. Process.* 25, 915–925, <http://dx.doi.org/10.1002/hyp.7876>.
- Williams, J., 1995. Chapter 25: The EPIC model. In: V.P. Singh (Ed.), Computer models of watershed hydrology. Water Resources Publications, pp. 909–1000.
- Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007a. Hydrological modelling of the Chaohe basin in China: statistical model formulation and Bayesian inference. *J. Hydrol.* 340, 167–182, <http://dx.doi.org/10.1016/j.jhydrol.2007.03.006>.
- Yang, J., Reichert, P., Abbaspour, K.C., 2007b. Bayesian uncertainty analysis in distributed hydrologic modeling: a case study in the Thur River basin (Switzerland). *Water Resour. Res.* 43, W10401, <http://dx.doi.org/10.1029/2006WR005497>.
- Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *J. Hydrol.* 358, 1–23, <http://dx.doi.org/10.1016/j.jhydrol.2008.05.012>.
- Zehe, E., Sivapalan, M., 2009. Threshold behavior in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications. *Hydrol. Earth Syst. Sc.* 13 (7), 1273–1297.
- Zhang, W., Arhonditsis, G., 2009. A Bayesian hierarchical framework for calibrating aquatic biogeochemical models. *Ecol. Model.* 220, 2142–2161.
- Zhang, X., Srinivasan, R., Arnold, J., Izaurrealde, R.C., Bosch, D., 2011. Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework. *Hydrol. Process.* 25 (14), 2313–2320.

**QUANTIFYING THE UNCERTAINTY OF NONPOINT SOURCE ATTRIBUTION
IN DISTRIBUTED WATER QUALITY MODELS: A BAYESIAN ASSESSMENT OF
SWAT'S SEDIMENT EXPORT PREDICTIONS**

Electronic Supplementary Material (ESM)

Christopher Wellen* and George B. Arhonditsis

Ecological Modelling Laboratory,
Department of Physical & Environmental Sciences, University of Toronto,
Toronto, Ontario, Canada, M1C 1A4

Tanya Long and Duncan Boyd

Great Lakes Unit, Water Monitoring & Reporting Section, Environmental Monitoring and
Reporting Branch, Ontario Ministry of the Environment
Toronto, Ontario, Canada, M9P 3V6

* Corresponding author.

e-mail: wellenc@mcmaster.ca, Tel.: +1 647 239 5138; Fax: +1 416 287 7279.

*Current address: Watershed Hydrology Group, School of Geography and Earth Sciences,
McMaster University, Hamilton, Ontario, Canada, L8S 4L8

ELECTRONIC SUPPLEMENTARY MATERIAL

Table ESM-1: Model fit statistics for all model calibrations in Redhill Creek.

Formulation		NSE*	RE*	r²
Standard (1)	Flow Calibration (1992 – 1994)	0.64	54%	0.66
	Flow Validation (1995 – 1998)	0.52	60%	0.54
	Flow Update (2010 – 2012)	0.68	57%	0.69
	Sediment Calibration	0.14	81%	0.17
Event (2)	Flow Calibration (1992 – 1994)	0.66	52%	0.71
	Flow Validation (1995 – 1998)	0.56	58%	0.57
	Flow Update (2010 – 2012)	0.70	58%	0.70
	Sediment Calibration	0.17	82%	0.18
Informative Priors (3)	Flow Calibration (1992 – 1994)	0.60	72%	0.63
	Flow Validation (1995 – 1998)	0.56	57%	0.57
	Flow Update (2010 – 2012)	0.68	59%	0.69
	Sediment Calibration	0.13	79%	0.15
Ensemble	Flow Validation (1995 – 1998)	0.56	56%	0.62
	Sediment Calibration	0.15	80%	0.17

*Nash Sutcliffe Efficiency **Relative Error

Table ESM-2: Model fit statistics for all model calibrations in Grindstone Creek.

Formulation		NSE	RE	r²
Standard (1)	Flow Calibration (1992 – 1994)	0.71	44%	0.72
	Flow Validation (1995 – 1998)	0.56	47%	0.56
	Flow Update (2010 – 2012)	0.70	50%	0.71
	Sediment Calibration	0.63	71%	0.71
Event (2)	Flow Calibration (1992 – 1994)	0.71	45%	0.71
	Flow Validation (1995 – 1998)	0.49	44%	0.52
	Flow Update (2010 – 2012)	0.70	50%	0.71
	Sediment Calibration	0.64	68%	0.73
Informative Priors (3)	Flow Calibration (1992 – 1994)	0.74	43%	0.75
	Flow Validation (1995 – 1998)	0.44	47%	0.48
	Flow Update (2010 – 2012)	0.62	53%	0.64
	Sediment Calibration	0.69	64%	0.71
Ensemble	Flow Validation (1995 – 1998)	0.47	49%	0.35
	Sediment Calibration	0.68	66%	0.71

Table ESM-3: Posterior results of the flow discharge submodel in Redhill Creek. These results are posterior to the sediment update.

Parameter	Standard (1)			Event (2)			Informative Priors (3)		
	MLV*	CI**	Del***	MLV	CI	Del	MLV	CI	Del
CN2 (Mult. Eff.)	1	7	17	-	-	-	-	-	-
ALPHA_BF	-3	19	23	-4	17	17	-6	20	13
SOL_AWC (Mult. Eff.)	0	-36	39	0	-2	30	0	12	40
GW_REVAP	-5	42	21	-7	-2	14	-8	6	11
ESCO	22	54	30	39	4	16	31	18	12
EPCO	-12	5	21	-13	-2	16	-11	1	15
GW_DELAY (Mult. Eff.)	31	24	41	-5	8	14	-22	15	13
SOL_KSAT (Mult. Eff.)	93	-19	66	89	-50	99	-4	21	19
SNOWCOVMX	4	-2	23	17	23	44	-3	22	22
SMFMX	7	12	14	-80	22	25	0	-3	16
SMFMN	-5	23	22	1	62	27	-6	40	21
SURLAG	17	15	25	23	-22	17	1	0	14
θ_p	-	-	-	0	10	20	-1	6	17
CN2 Low (Mult. Eff.)	-	-	-	4	14	16	8	27	20
CN2 High (Mult. Eff.)	-	-	-	0	1	14	0	10	13

*Percentage change of the most likely value between marginal prior and posterior distributions.

**Percentage change in the width of the 95% credible interval between marginal prior and posterior distributions.

***Delta index as presented by Hong et al. (2005), expressed as percent of maximum value.

Table ESM-4: Posterior results of the flow discharge submodel in Grindstone Creek. These results are posterior to the sediment update.

Parameter	Standard (1)			Event (2)			Informative Priors (3)		
	MLV*	CI**	Del***	MLV	CI	Del	MLV	CI	Del
CN2 (Mult. Eff.)	-3	-11	28	-	-	-	-	-	-
ALPHA_BF	-4	4	17	-5	9	14	-5	22	14
SOL_AWC (Mult. Eff.)	0	6	26	0	-13	57	0	2	14
GW_REVAP	13	-8	17	-96	-15	16	44	24	19
ESCO	5	22	18	97	54	23	10	3	13
EPCO	-10	4	18	3	-3	13	-26	18	18
GW_DELAY (Mult. Eff.)	-22	13	58	-64	-27	81	-6	36	43
SOL_KSAT (Mult. Eff.)	26	-18	88	9	-34	28	5	1	10
SNOWCOVMX	11	16	67	-3	21	40	7	10	19
SMFMX	-48	68	30	1	13	14	-22	4	18
SMFMN	3	0	23	-4	14	24	344	-23	20
SURLAG	1	-16	23	1	-2	16	1	10	11
θ_p	-	-	-	-19	37	54	4	20	16
CN2 Low (Mult. Eff.)	-	-	-	55	52	55	-6	6	17
CN2 High (Mult. Eff.)	-	-	-	-22	12	66	11	-7	20

*Percentage change of the most likely value between marginal prior and posterior distributions.

**Percentage change in the width of the 95% credible interval between marginal prior and posterior distributions.

***Delta index as presented by Hong et al. (2005), expressed as percent of maximum value.

Table ESM-5: Parameter posterior means and standard deviations for sediment parameters in Redhill and Grindstone Creeks.

	Formulation 1					Formulation 2					Formulation 3				
	Mean	SD	MLV*	CI**	Del***	Mean	SD	MLV	CI	Del	Mean	SD	MLV	CI	Del
Redhill Parameters															
USLE_K	0.99	0.39	-1	2	23	0.92	0.36	-8	1	14	0.86	0.36	-14	2	11
PRF	1.11	0.10	-11	-86	35	1.10	0.10	-12	-87	37	1.08	0.09	-14	-89	51
SPCON****	1.24	0.23	-	-	-	1.29	0.28	-	-	-	1.26	0.28	-	-	-
CH_N	1.26	0.58	26	10	37	1.00	0.22	0	-60	45	0.99	0.30	-1	-28	43
CH_COV1	0.58	0.19	-4	16	18	0.53	0.20	-13	8	20	0.67	0.18	9	3	18
Grindstone Parameters															
USLE_K	0.80	0.29	-20	-10	25	0.83	0.29	-17	-11	22	0.70	0.31	-30	-6	29
PRF	1.10	0.10	-3	-69	34	1.20	0.19	6	-45	16	1.11	0.11	-2	-69	29
SPCON****	1.83	0.89	-	-	-	1.49	0.42	-	-	-	1.49	0.50	-	-	-
CH_N	1.98	0.60	98	7	49	2.06	0.45	106	-12	62	2.06	0.50	106	-5	57
CH_COV1	0.68	0.18	11	-5	25	0.66	0.17	9	-5	19	0.63	0.19	3	2	15
Shared Parameters															
USLE_P	0.52	0.15	-	-	-	0.46	0.14	-	-	-	0.47	0.15	-	-	-
URBCOEF	0.011	0.007	-	-	-	0.016	0.01	-	-	-	0.019	0.01	-	-	-
SLSUBBSN	3.05	1.39	205	10	45	1.60	1.15	60	5	17	2.17	1.54	117	12	33
OV_N	1.33	0.60	33	58	32	1.43	0.58	43	51	33	1.33	0.63	33	49	29
Likelihood Parameters															
ρ_s	0.44	0.15	-	-	-	0.42	0.14	-	-	-	0.40	0.14	-	-	-
σ_R	1.84	0.20	-	-	-	1.88	0.20	-	-	-	1.79	0.16	-	-	-
σ_G	1.51	0.18	-	-	-	1.37	0.17	-	-	-	1.45	0.17	-	-	-

* Percentage change of the most likely value between marginal prior and posterior distributions.

** Percentage change in the width of the 95% credible interval between marginal prior and posterior distributions.

*** Delta index as presented by Hong et al. (2005), expressed as percent of maximum value.

**** In units of ten thousandths (10^{-4})

Table ESM-6: Parameter posterior means and standard deviations of the flow discharge submodel for Redhill Creek.

Parameter	Standard (1)		Event (2)		Informative Priors (3)	
	Mean	SD	Mean	SD	Mean	SD
CN2 (Mult. Eff.)	0.564	0.038	-	-	-	-
ALPHA_BF	0.974	0.018	0.957	0.028	0.938	0.047
SOL_AWC (Mult. Eff.)	2.358	0.001	1.719	0.001	1.720	0.001
GW_REVAP	0.193	0.007	0.190	0.012	0.189	0.013
ESCO	0.135	0.034	0.129	0.026	0.126	0.027
EPCO	0.705	0.215	0.662	0.150	0.663	0.167
GW_DELAY (Mult. Eff.)	0.652	0.129	0.532	0.108	0.470	0.138
SOL_KSAT (Mult. Eff.)	0.331	0.123	0.257	0.037	0.214	0.035
SNOWCOVMX	6.812	0.234	16.745	2.279	16.174	1.371
SMFMX	3.387	0.213	3.566	0.237	3.646	0.258
SMFMN	2.961	0.112	3.082	0.261	2.807	0.243
SURLAG	3.748	3.194	0.530	0.034	0.380	0.011
			1.350	0.027	0.971	0.025
θ_p	-	-	(21 mm)	(1 mm)	(8.3 mm)	(1 mm)
CN2 Low (Mult. Eff.)	-	-	0.572	0.051	0.538	0.039
CN2 High (Mult. Eff.)	-	-	1.100	0.011	1.013	0.024
σ	0.149	0.005	0.137	0.005	0.140	0.005
ρ	0.396	0.030	0.328	0.036	0.332	0.033
Log(Model Likelihood)	224.377	3.900	297.162	4.724	273.937	4.151

Table ESM-7: Parameter posterior means and standard deviations of the flow discharge submodel for Grindstone Creek.

Parameter	Standard (1)		Event (2)		Informative Priors (3)	
	Mean	SD	Mean	SD		
CN2 (Mult. Eff.)	0.572	0.036	-	-	-	-
ALPHA_BF	0.957	0.030	0.940	0.046	0.944	0.039
SOL_AWC (Mult. Eff.)	1.430	0.001	1.407	0.008	1.430	0.001
GW_REVAP	0.082	0.034	0.052	0.028	0.062	0.028
ESCO	0.154	0.046	0.181	0.076	0.200	0.066
EPCO	0.627	0.224	0.488	0.261	0.701	0.196
GW_DELAY (Mult. Eff.)	1.161	0.108	1.747	0.274	1.219	0.135
SOL_KSAT (Mult. Eff.)	0.532	0.033	0.553	0.069	0.538	0.041
SNOWCOVMX	10.926	0.400	21.064	0.369	15.872	5.057
SMFMX	2.871	0.226	2.794	0.111	4.085	0.535
SMFMN	1.286	0.151	2.512	0.083	1.254	0.330
SURLAG	0.509	0.007	0.511	0.012	0.373	0.005
			1.050	0.092	0.746	0.042
θ_p	-	-	(10.2 mm)	(1 mm)	(4.6 mm)	(1 mm)
CN2 Low (Mult. Eff.)	-	-	0.572	0.120	1.093	0.108
CN2 High (Mult. Eff.)	-	-	0.770	0.066	0.540	0.041
σ	0.066	0.002	0.066	0.002	0.064	0.003
ρ	0.904	0.012	0.906	0.014	0.908	0.013
Log(Model Likelihood)	1054.631	3.436	1053.227	4.854	1074.392	5.689

FIGURE CAPTIONS

Figure ESM 1: Posterior marginals of the flow discharge submodel for Redhill Creek, Formulation 1.

Figure ESM 2: Posterior marginals of the flow discharge submodel for Redhill Creek, Formulation 2.

Figure ESM 3: Posterior marginals of the flow discharge submodel for Redhill Creek, Formulation 3.

Figure ESM 4: Posterior marginals of the flow discharge submodel for Grindstone Creek, Formulation 1.

Figure ESM 5: Posterior marginals of the flow discharge submodel for Grindstone Creek, Formulation 2.

Figure ESM 6: Posterior marginals of the flow discharge submodel for Grindstone Creek, Formulation 3.

Figure ESM 7: Posterior marginals of the sediment load submodel, Formulation 1.

Figure ESM 8: Posterior marginals of the sediment load submodel, Formulation 2.

Figure ESM 9: Posterior marginals of the sediment load submodel, Formulation 3.

Figure ESM 10: Likelihood Assessment of the flow discharge submodel for Redhill Creek. Top to bottom rows report on Formulations 1 to 3, respectively. The left column presents quantile-quantile plots for the expected and actual standardized innovations, the middle column presents autocorrelation functions for the innovations, and the right column presents density plots of the expected and actual standardized innovations. As described in the methodology section, all likelihoods were based on a first order autocorrelation of the residuals and a Student's t-distribution with 7 degrees of freedom for the innovations. All standardization was performed with the relevant posterior estimates of the first order correlation coefficient (ρ) and the scale parameter (σ) for the innovations.

Figure ESM 11: Likelihood Assessment of the flow discharge submodel for Grindstone Creek. Top to bottom rows report on Formulations 1 to 3, respectively. The left column presents quantile-quantile plots for the expected and actual standardized innovations, the middle column presents autocorrelation functions for the innovations, and the right column presents density plots of the expected and actual standardized innovations. As described in the methodology section, all likelihoods were based on a first order autocorrelation of the residuals and a Student's t-distribution with 7 degrees of freedom for the innovations. All standardization was performed with the relevant posterior estimates of the first order correlation coefficient (ρ) and the scale parameter (σ) for the innovations.

Figure ESM 12: Quantile-quantile plots of the residuals pertaining to the sediment load submodel for both Creeks. All plots compare actual mean residuals to the residuals which would be expected from the normal distribution. The top row presents Redhill Creek, and the bottom row presents Grindstone Creek. The left column presents residuals from Formulation 1, the middle column residuals from Formulation 2, and the right column from Formulation 3. All lines indicate 1:1 relationship. The normal distribution is adequate to describe the residuals.

Figure ESM 13: Estimated runoff and sediment source areas across the three formulations by land use for the entire study period (2010 – 2012). Error bars depict all sources of parametric uncertainty.

Figure ESM 14: Estimated runoff and sediment source areas across the three formulations by land use for the growing season (May – September, 2010 – 2012). Error bars depict all sources of parametric uncertainty.

Figure ESM 15: Ensemble predictions of flow discharge for the validation period.

Figure ESM 16: Ensemble model predictions of events and measured data aggregated by month.

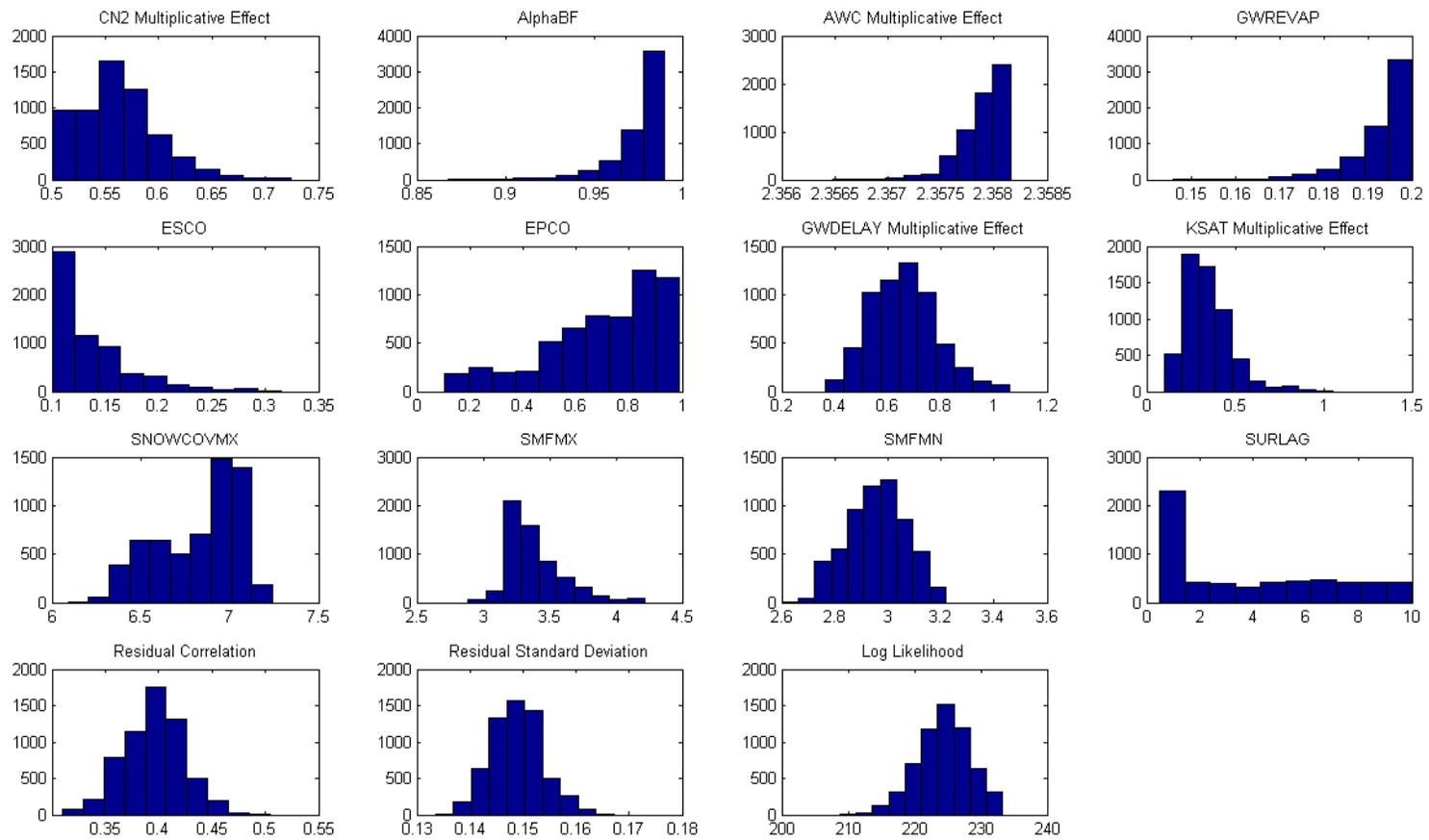


Figure ESM 1: Posterior marginal distributions of the flow discharge submodel for Redhill Creek, Formulation 1.

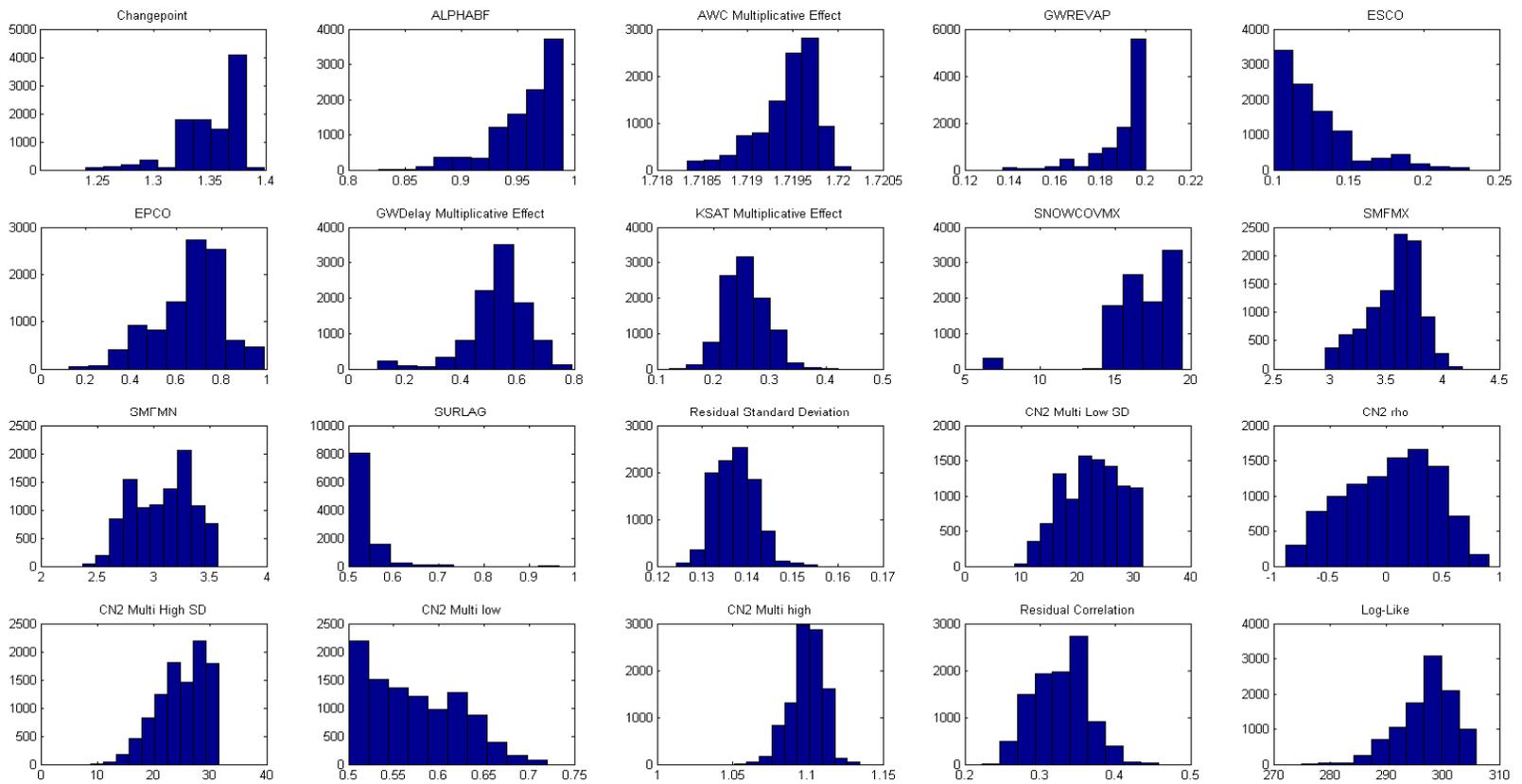


Figure ESM 2: Posterior marginal distributions of the flow discharge submodel for Redhill Creek, Formulation 2.

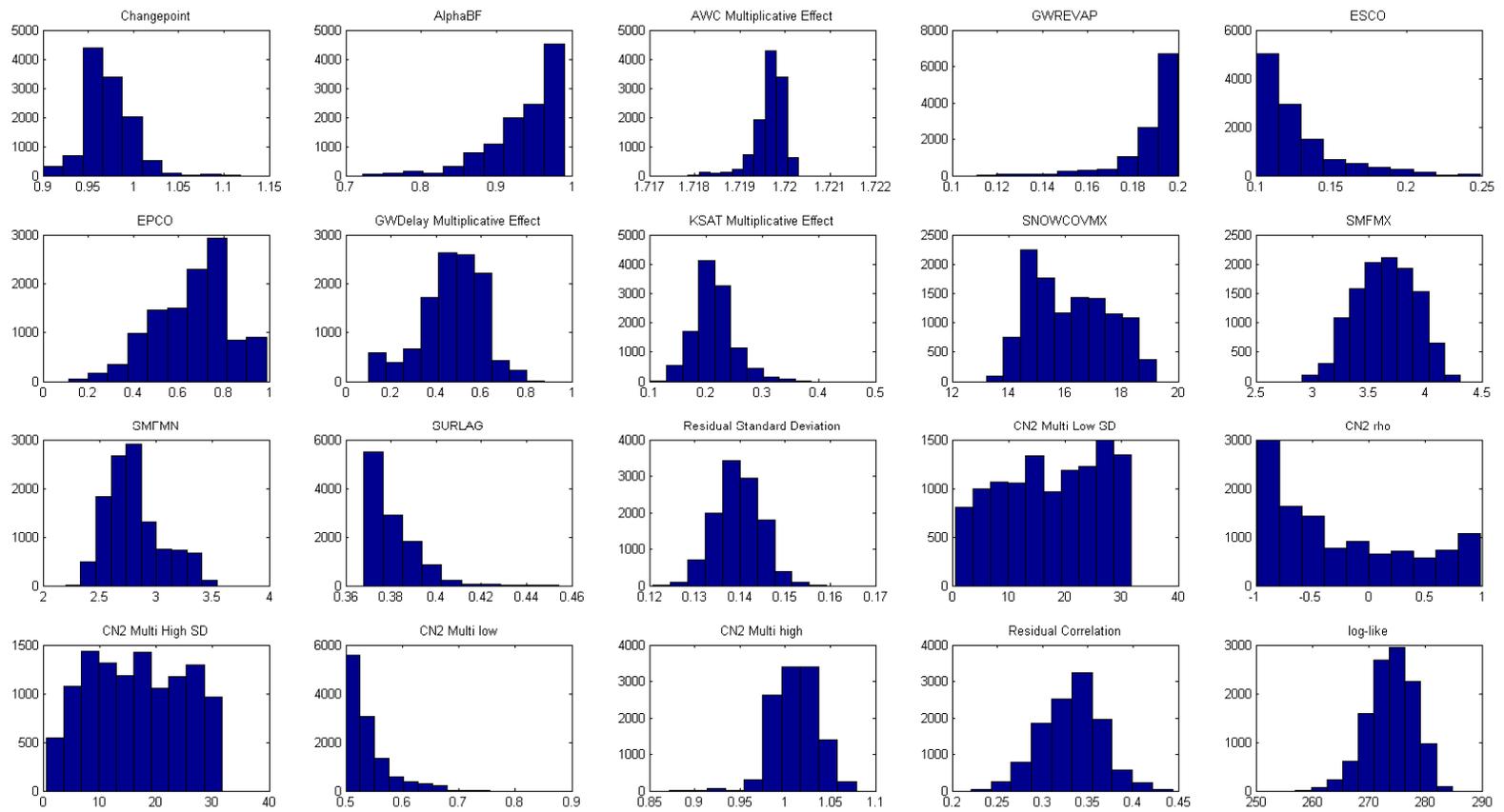


Figure ESM 3: Posterior marginal distributions of the flow discharge submodel for Redhill Creek, Formulation 3.

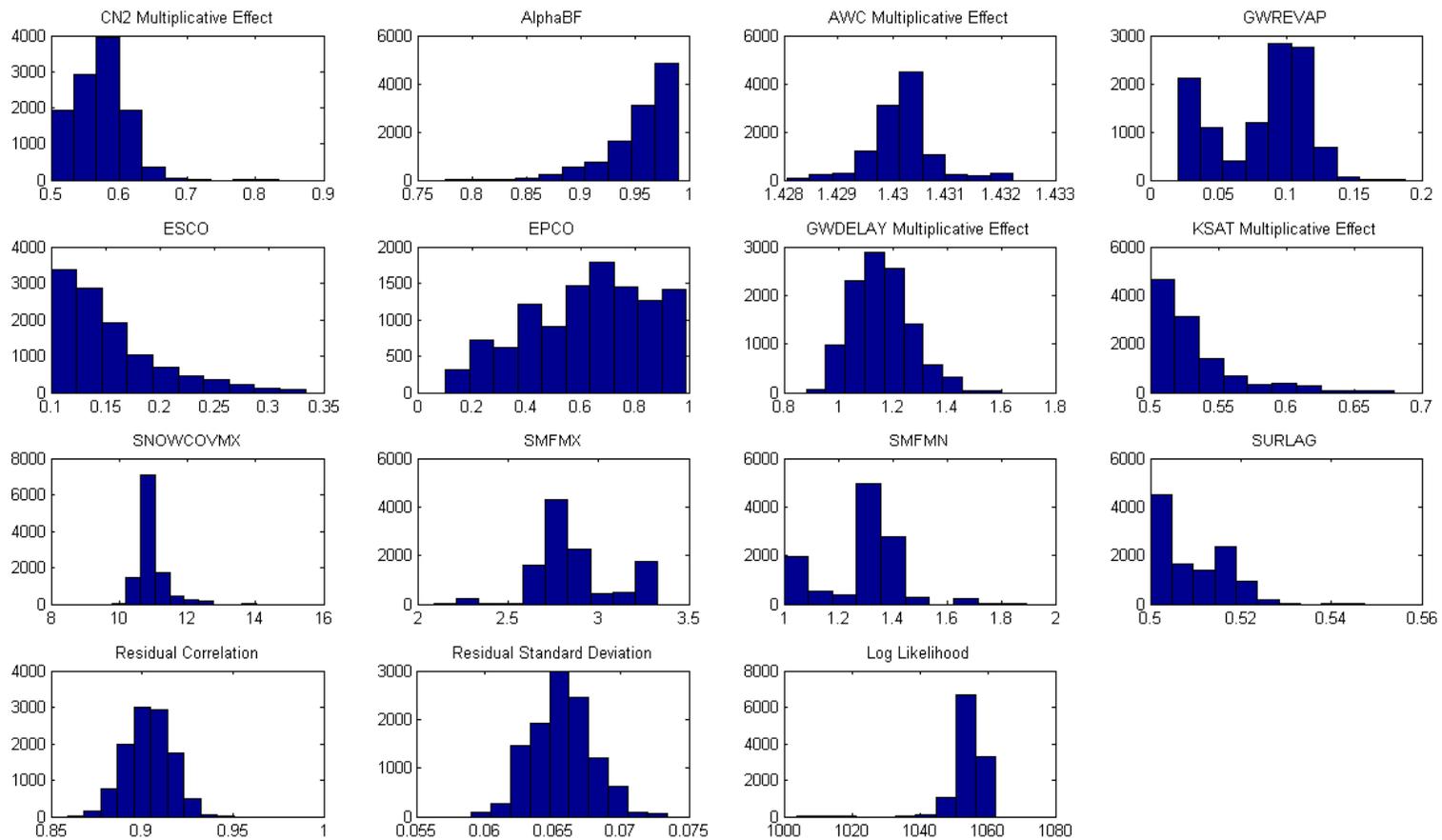


Figure ESM 4: Posterior marginal distributions of the flow discharge submodel for Grindstone Creek, Formulation 1.

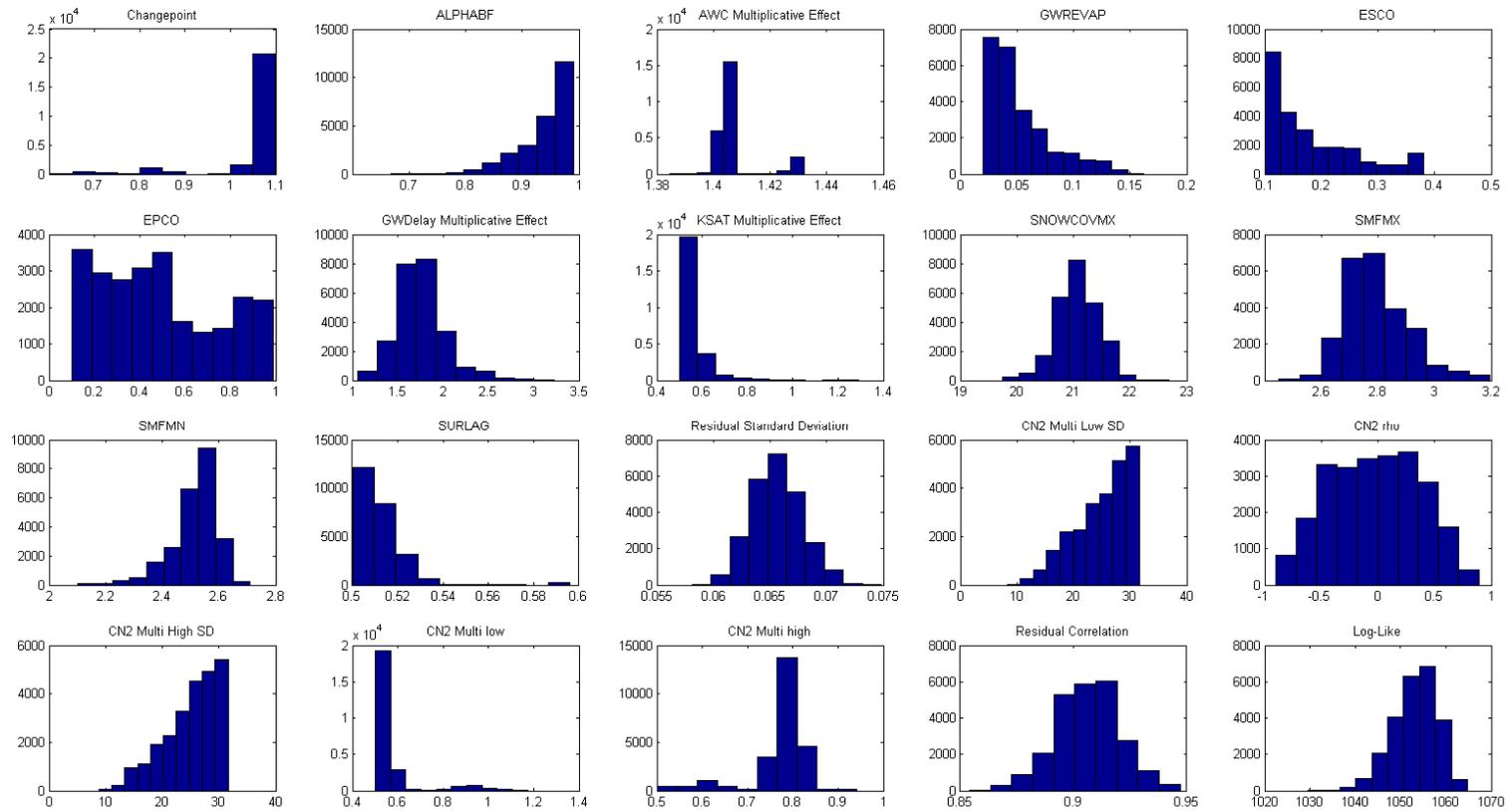


Figure ESM 5: Posterior marginal distributions of the flow discharge submodel for Grindstone Creek, Formulation 2.

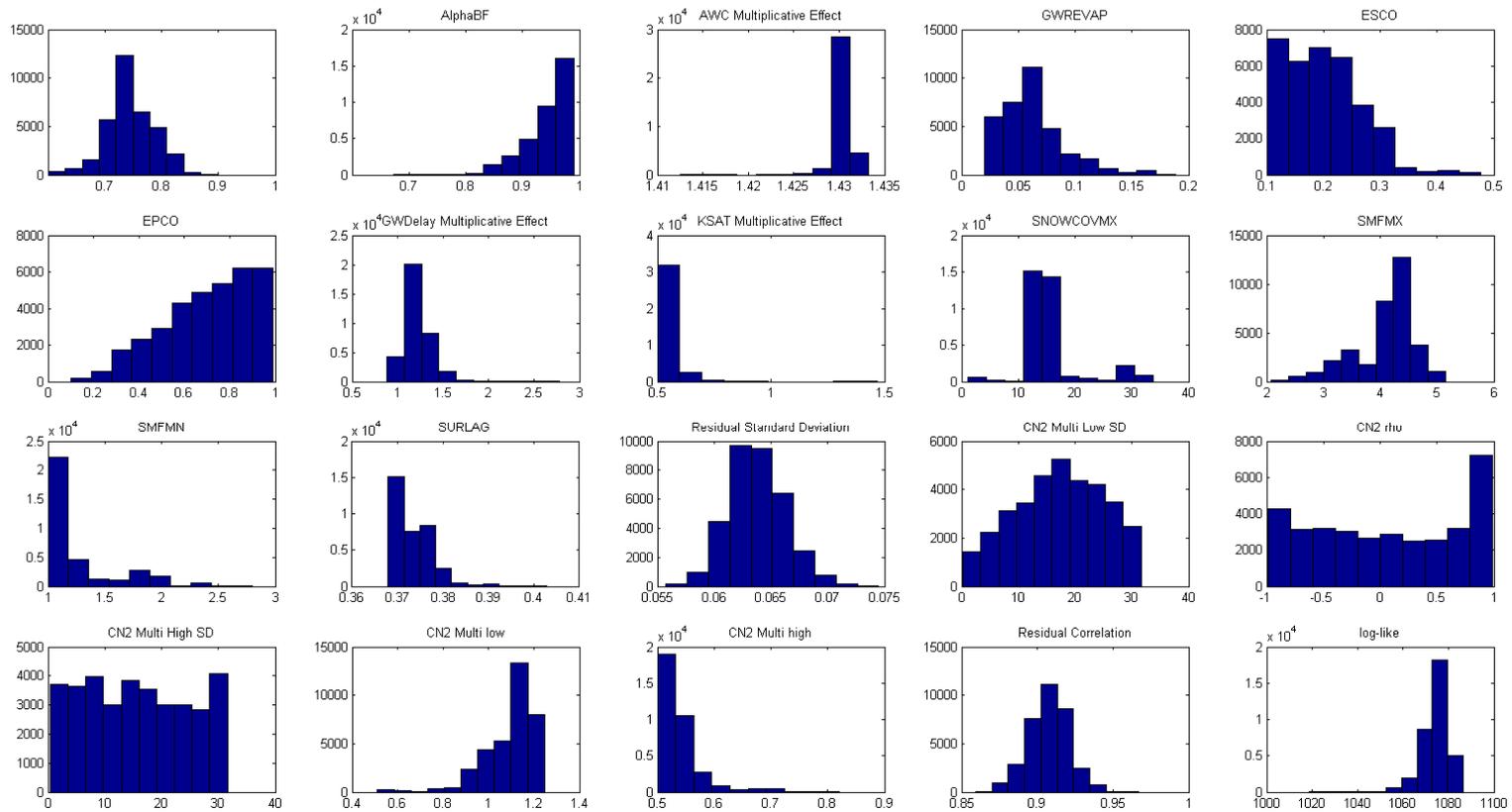


Figure ESM 6: Posterior marginal distributions of the flow discharge submodel for Grindstone Creek, Formulation 3.

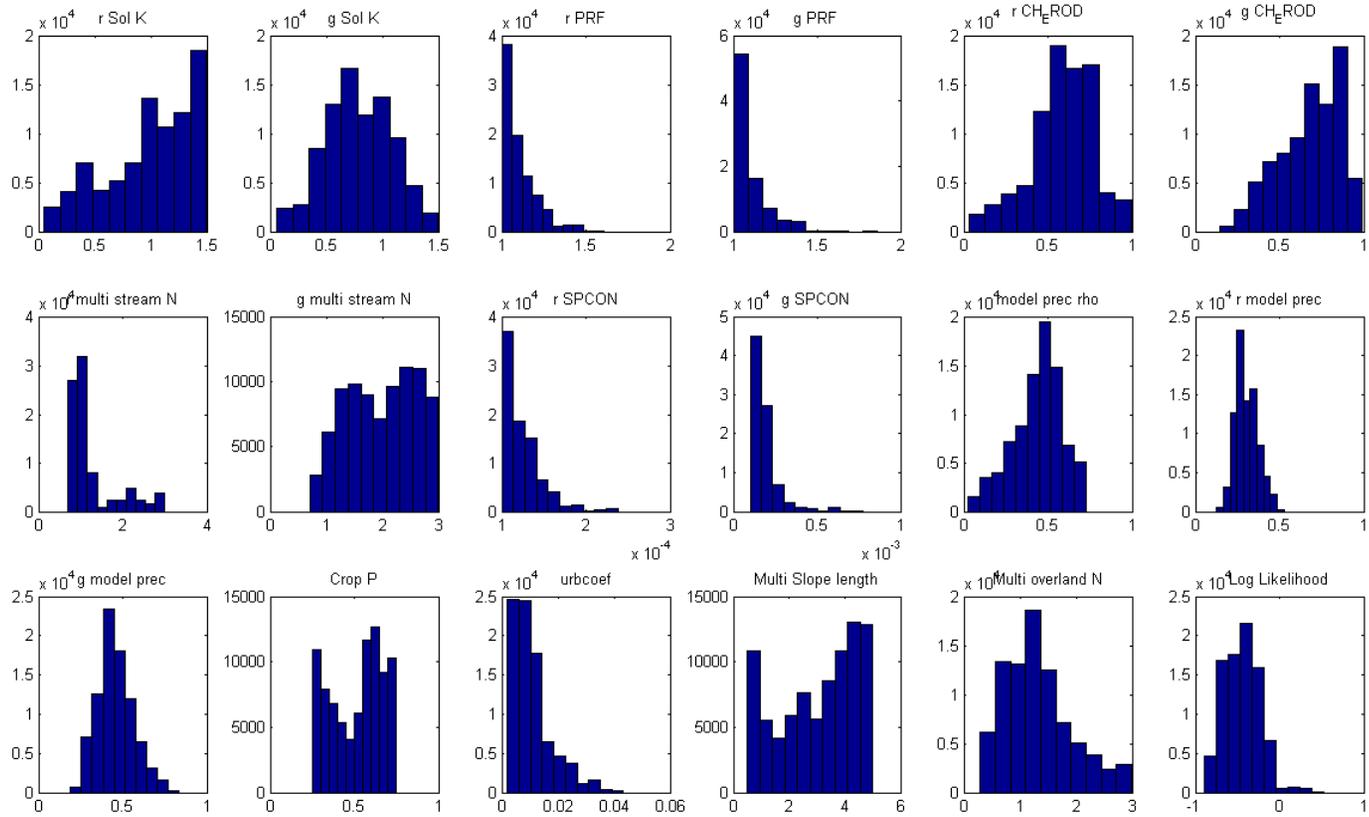


Figure ESM 7: Posterior marginal distributions of the sediment load submodel, Formulation 1.

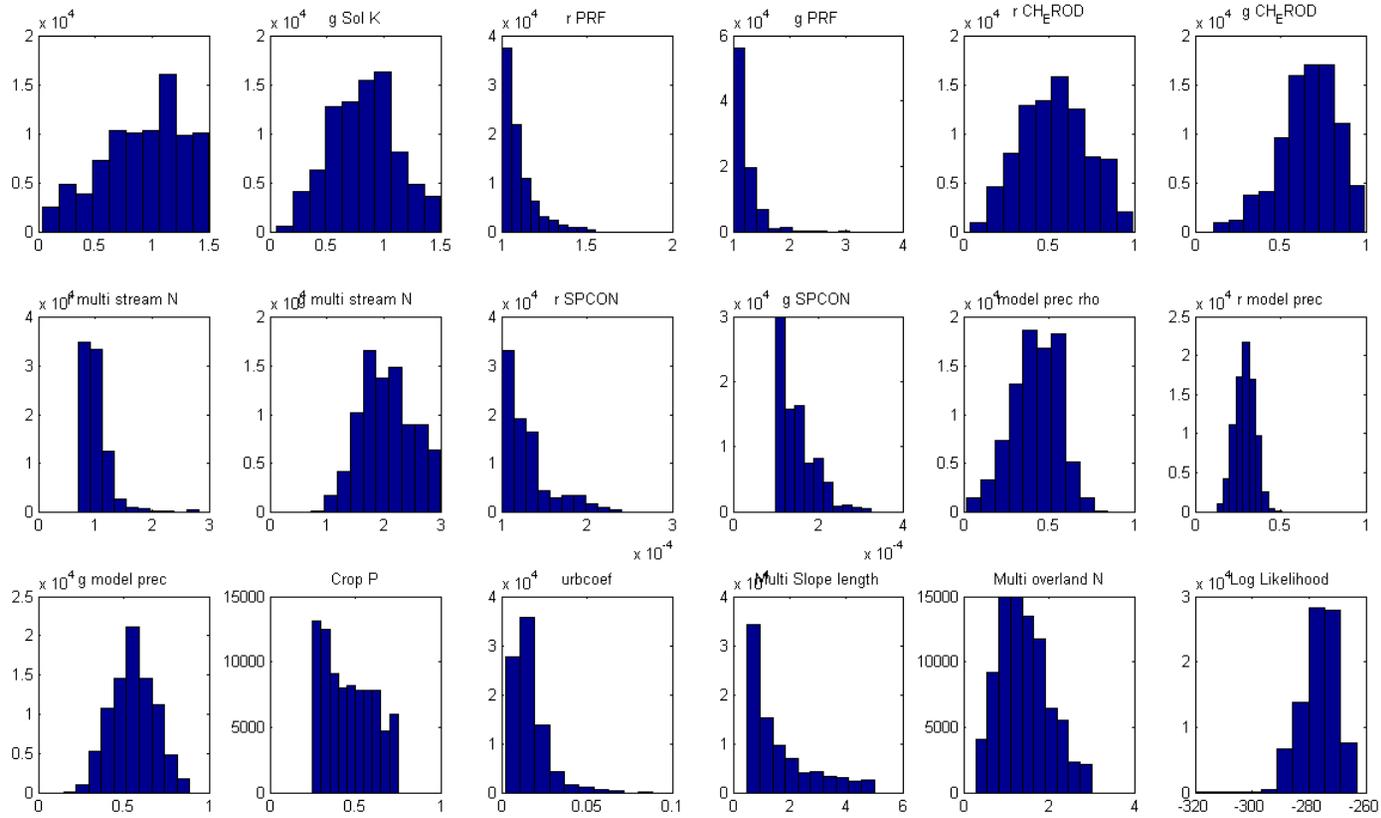


Figure ESM 8: Posterior marginal distributions of the sediment load submodel, Formulation 2.

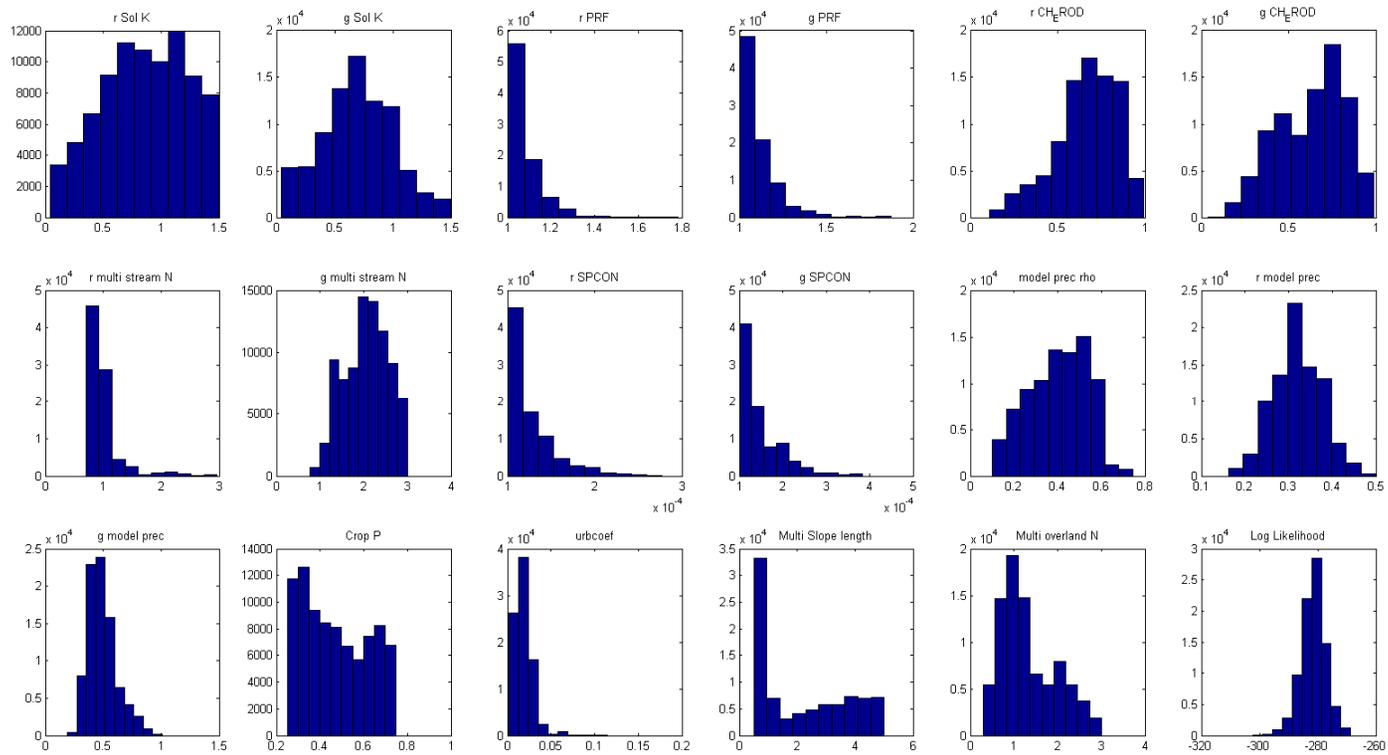


Figure ESM 9: Posterior marginal distributions of the sediment load submodel, Formulation 3.

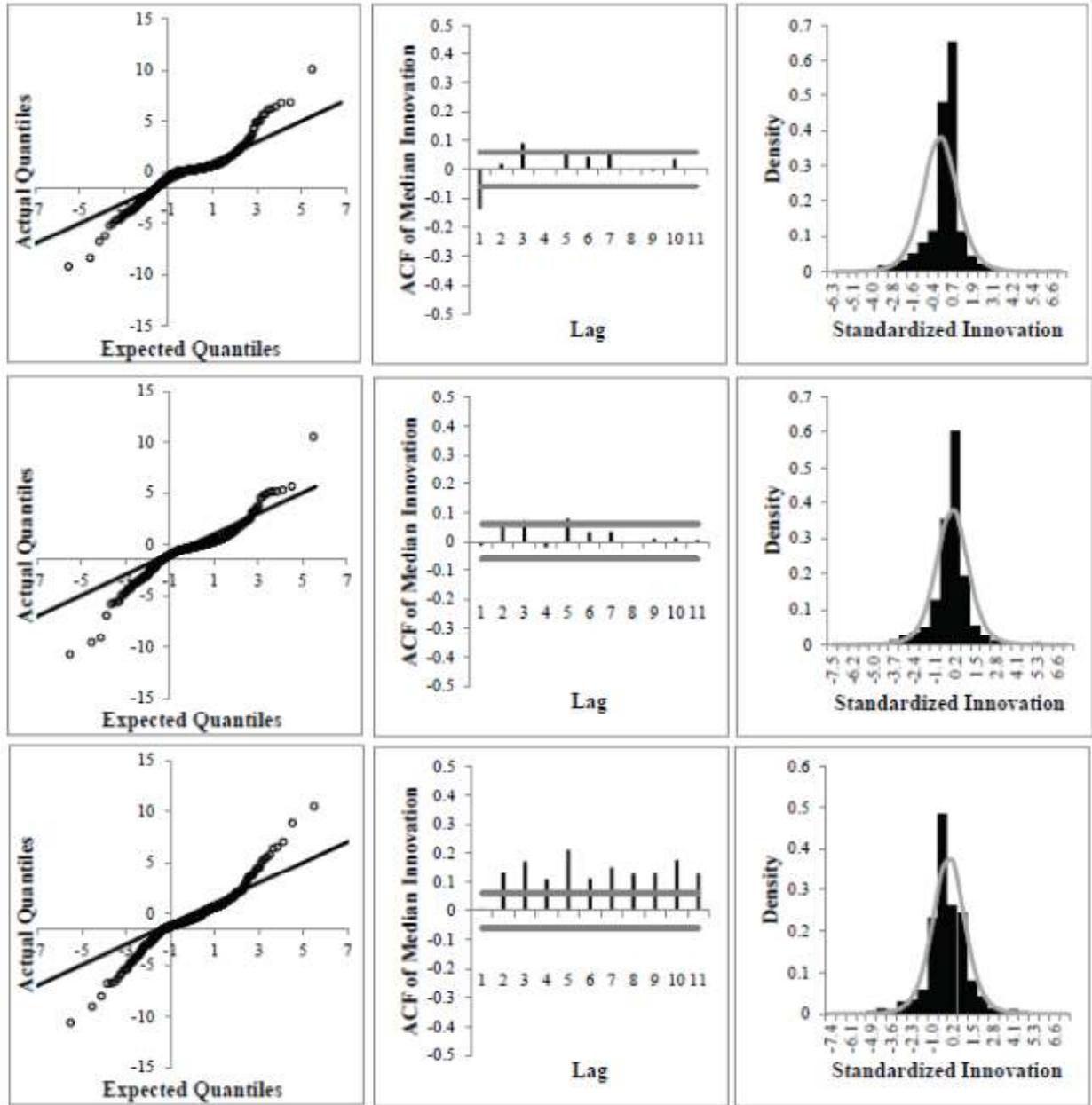


Figure ESM 10: Likelihood Assessment of the flow discharge submodel for Redhill Creek. Top to bottom rows report on Formulations 1 to 3, respectively. The left column presents quantile-quantile plots for the expected and actual standardized innovations, the middle column presents autocorrelation functions for the innovations, and the right column presents density plots of the expected and actual standardized innovations. As described in the methodology section, all likelihoods were based on a first order autocorrelation of the residuals and a Student's t -distribution with 7 degrees of freedom for the innovations. All standardization was performed with the relevant posterior estimates of the first order correlation coefficient (ρ) and the scale parameter (σ) for the innovations.

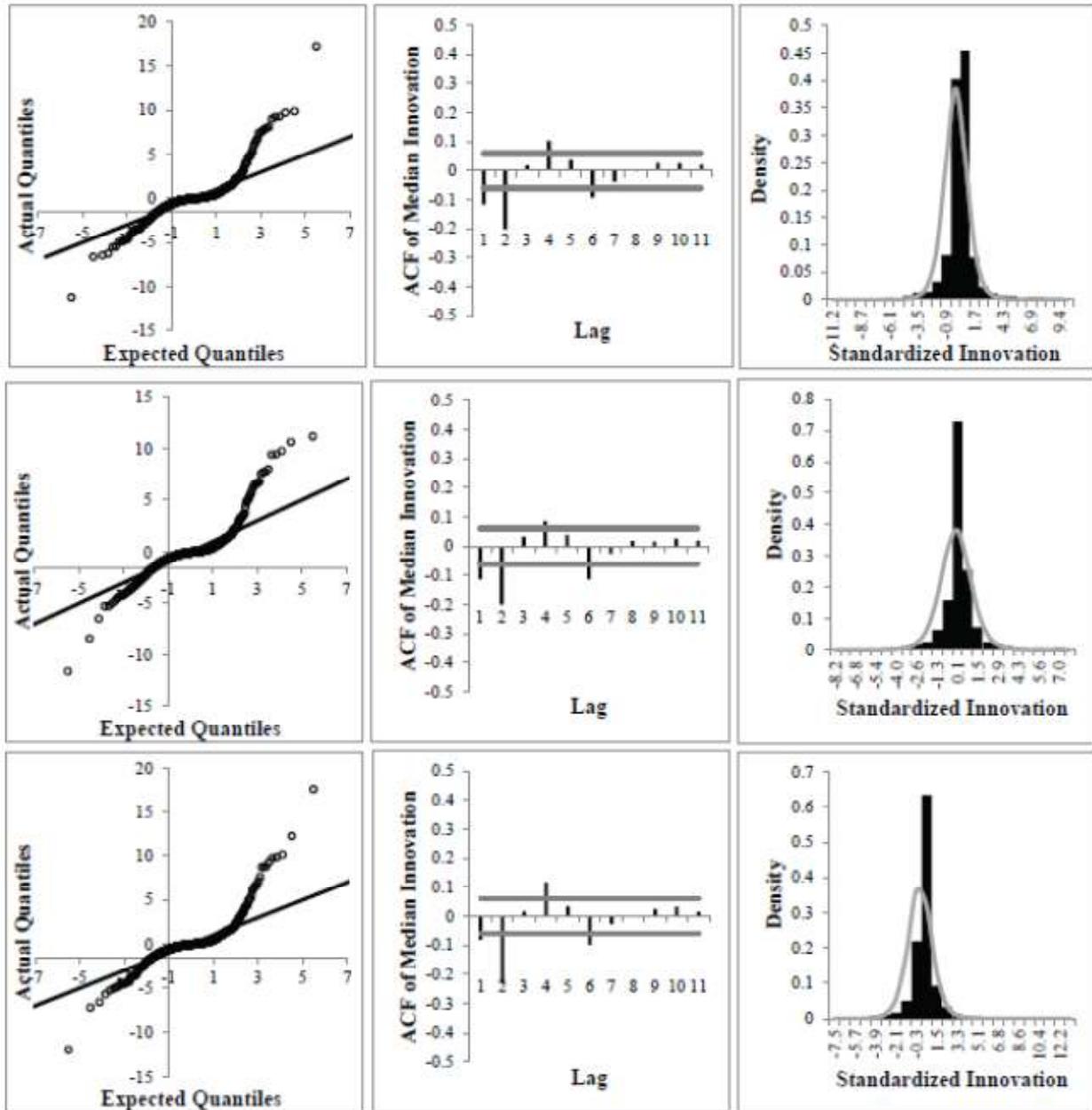


Figure ESM 11: Likelihood Assessment of the flow discharge submodel for Grindstone Creek. Top to bottom rows report on Formulations 1 to 3, respectively. The left column presents quantile-quantile plots for the expected and actual standardized innovations, the middle column presents autocorrelation functions for the innovations, and the right column presents density plots of the expected and actual standardized innovations. As described in the methodology section, all likelihoods were based on a first order autocorrelation of the residuals and a Student's t-distribution with 7 degrees of freedom for the innovations. All standardization was performed with the relevant posterior estimates of the first order correlation coefficient (ρ) and the scale parameter (σ) for the innovations.

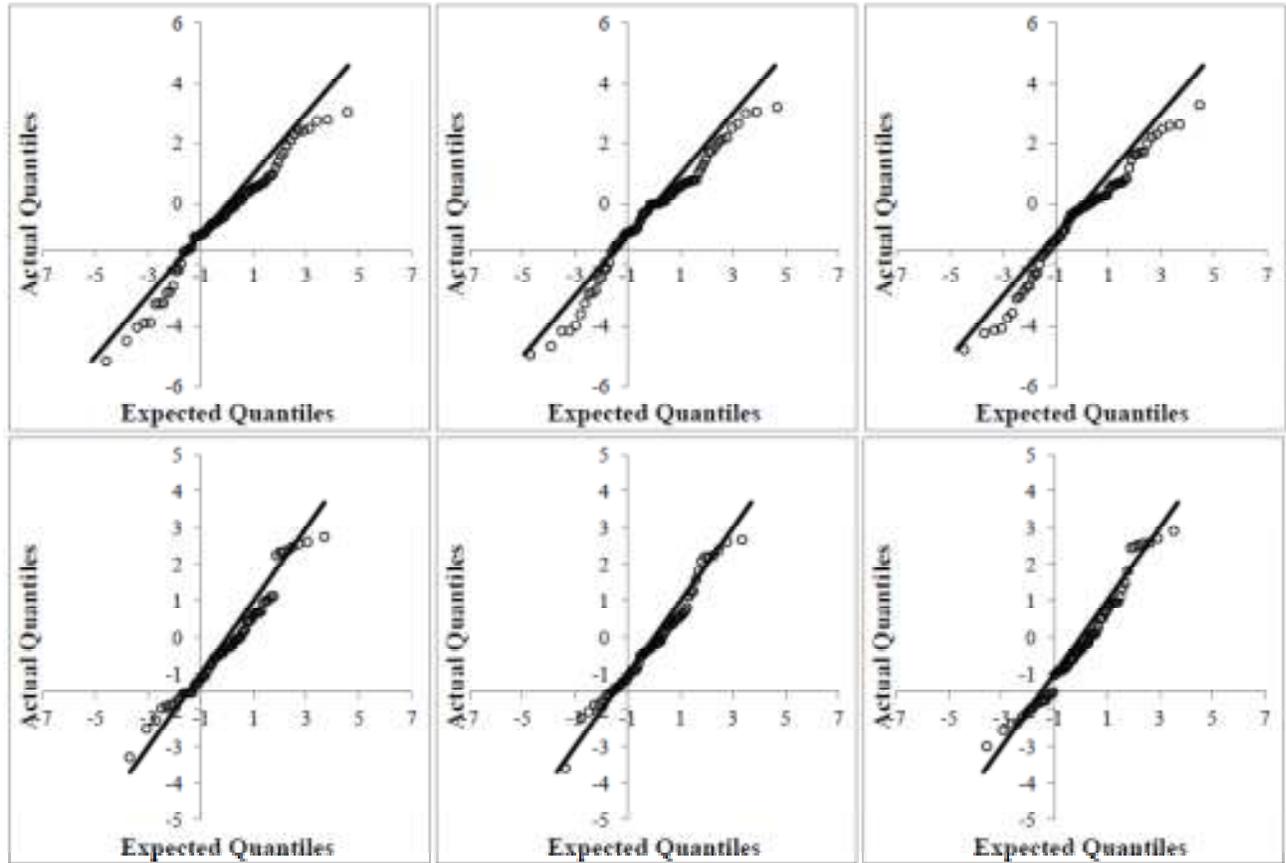


Figure ESM 12: Quantile-quantile plots of the residuals pertaining to the sediment load submodel for both Creeks. All plots compare actual mean residuals to the residuals which would be expected from the normal distribution. The top row presents Redhill Creek, and the bottom row presents Grindstone Creek. The left column presents residuals from Formulation 1, the middle column residuals from Formulation 2, and the right column from Formulation 3. All lines indicate 1:1 relationship. The normal distribution is adequate to describe the residuals.

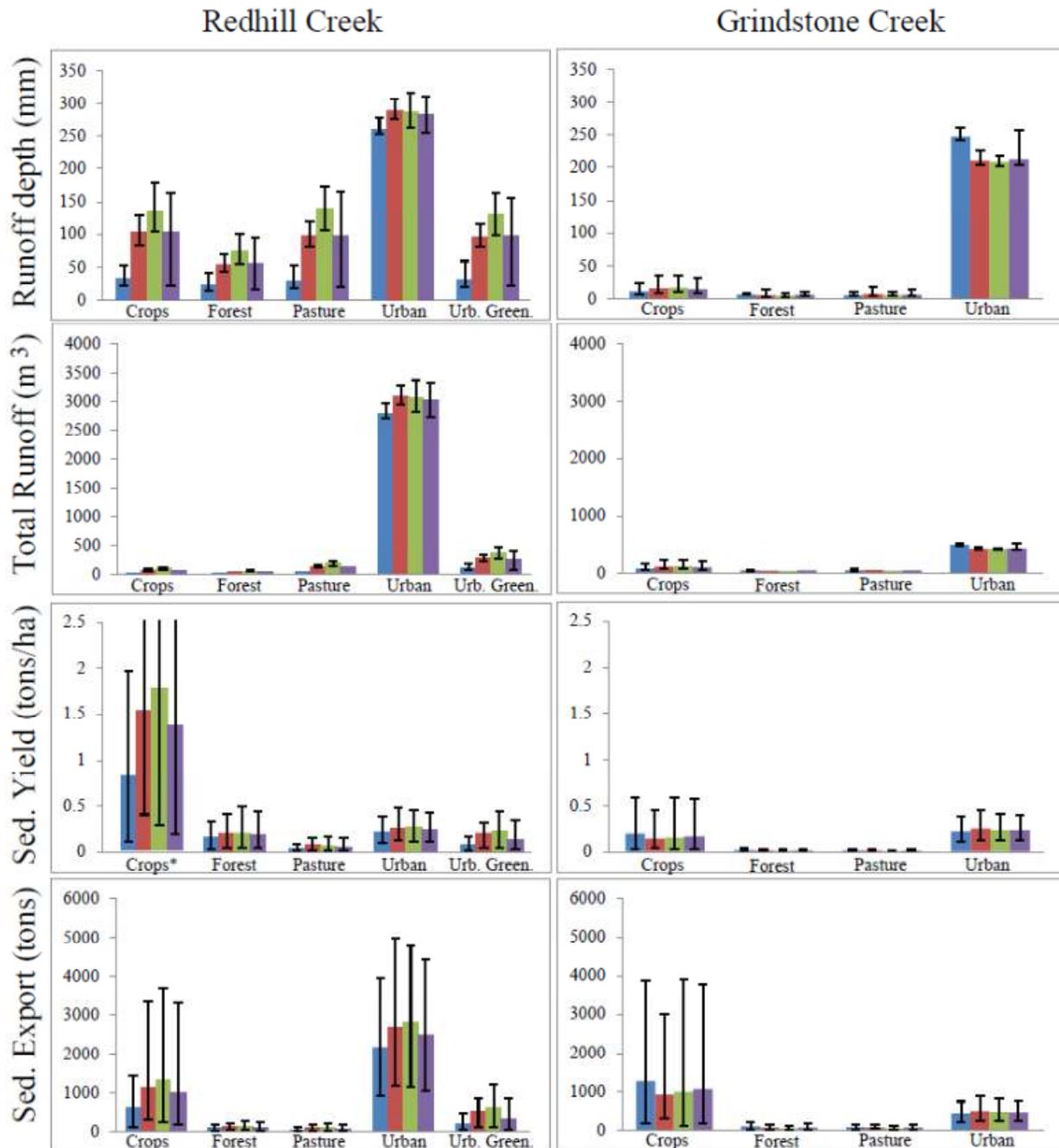


Figure ESM 13: Estimated runoff and sediment source areas across the three formulations by land use for the entire study period (2010 – 2012). Error bars depict all sources of parametric uncertainty.

*95% credible intervals for Crop sediment yield in Redhill Creek for Formulations 2, 3, and the ensemble are 4.6, 5.0, and 4.5 tons ha⁻¹ yr⁻¹.

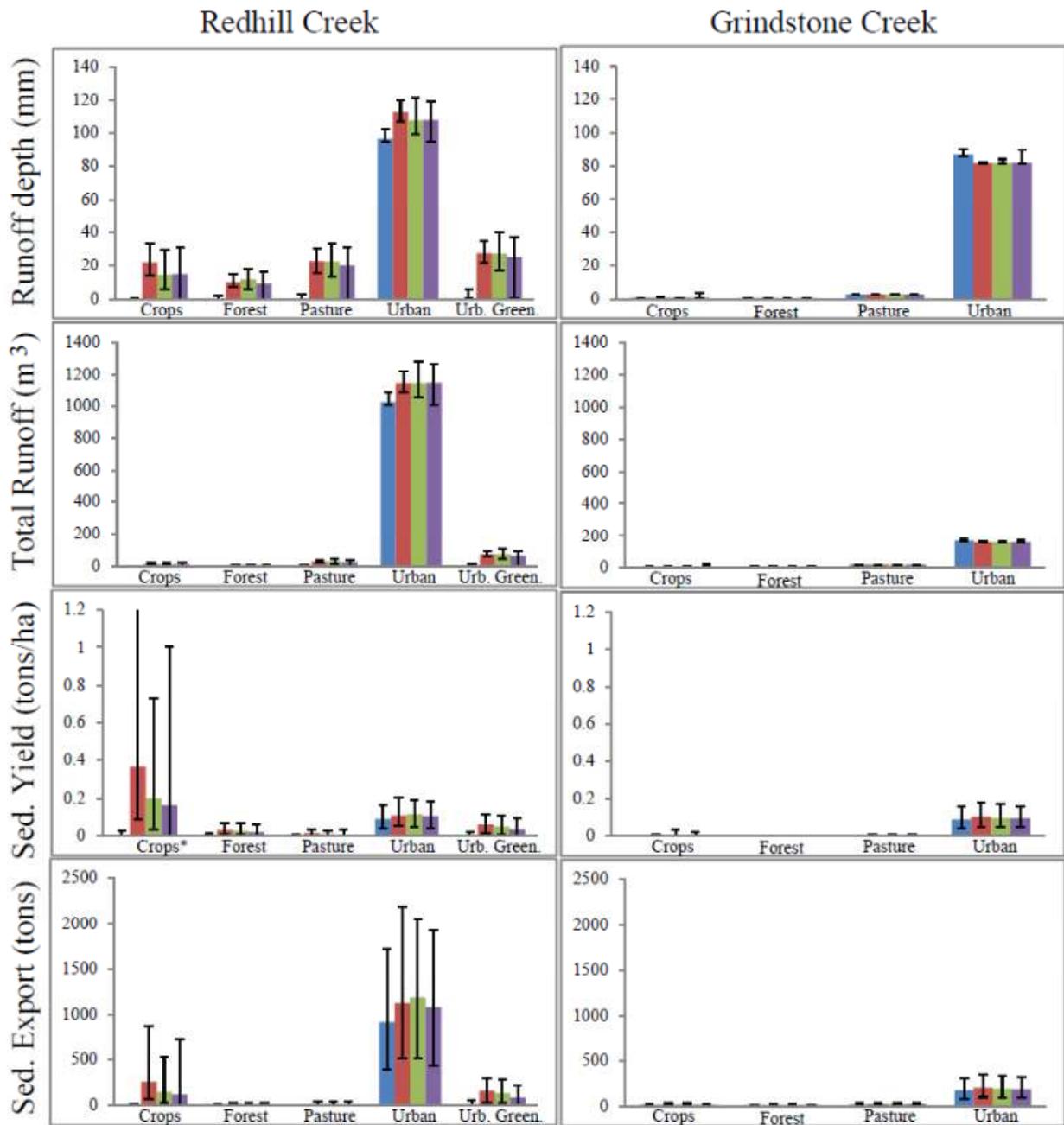


Figure ESM 14: Estimated runoff and sediment source areas across the three formulations by land use for the growing season (May – September, 2010 – 2012). Error bars depict all sources of parametric uncertainty.

*95% credible intervals for Crop sediment yield in Redhill Creek for Formulation 2 is 1.2 tons ha⁻¹ yr⁻¹.

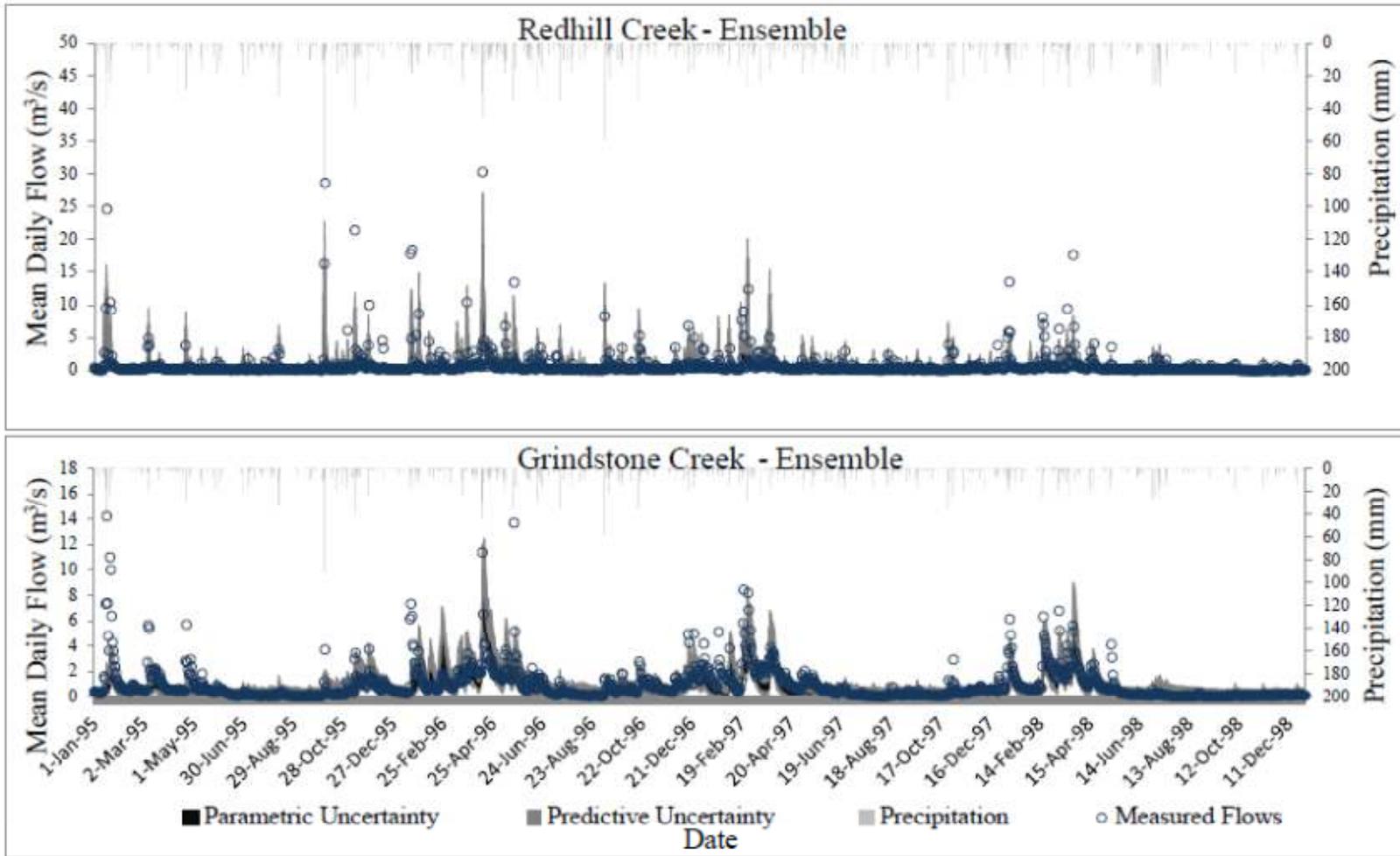


Figure ESM 15: Ensemble predictions of flow discharge for the validation period.

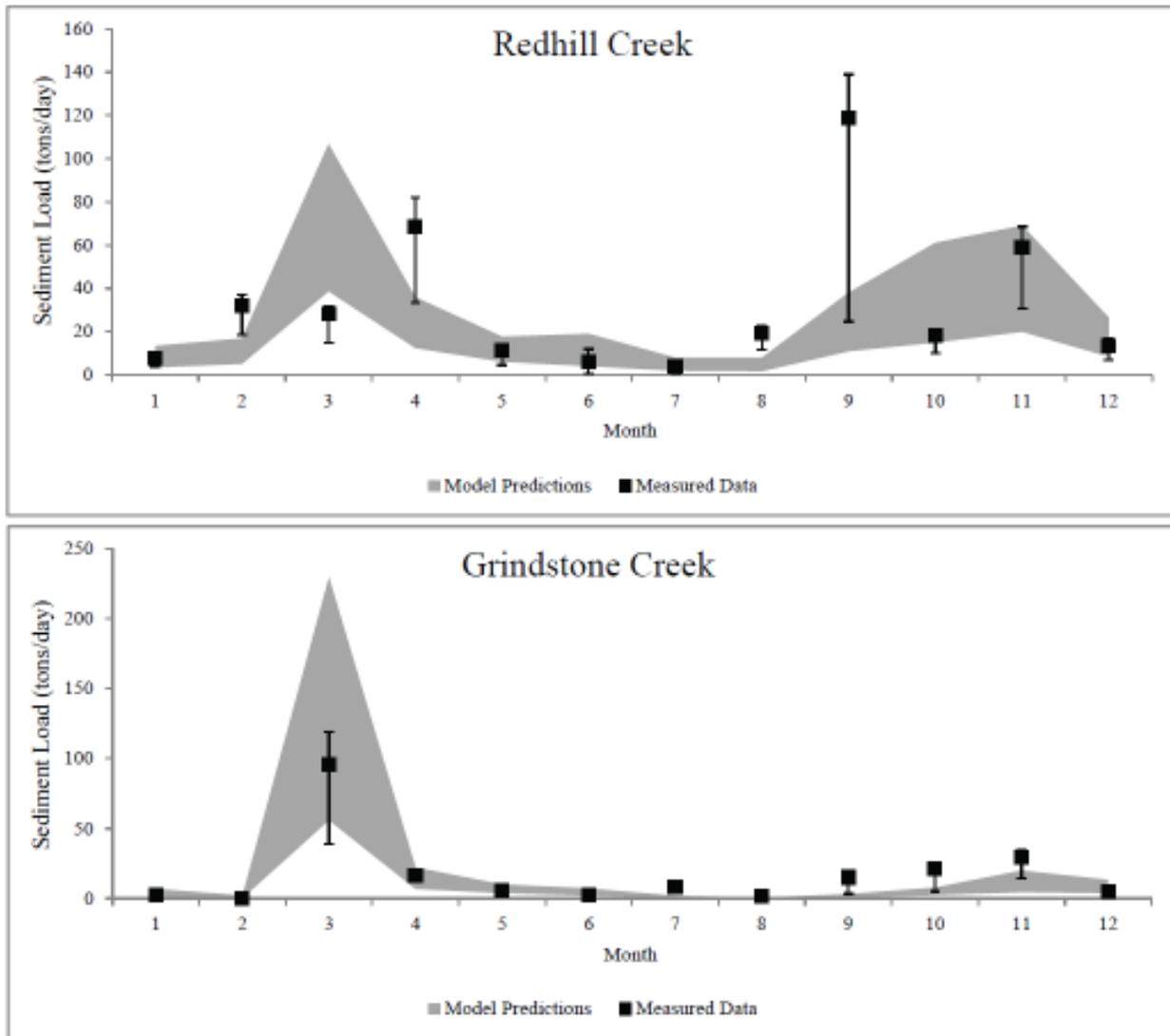


Figure ESM 16: Ensemble model predictions of events and measured data aggregated by month.

Threshold configuration of SWAT

We accommodate threshold behavior by incorporating a precipitation threshold into the *SWAT* model. We assume that above some threshold of precipitation θ_p , an extreme state exists and a subset of the parameter takes on different values than in the normal state. This essentially postulates that watersheds can be characterized by multiple discrete states of response. For this application, we allowed the curve number parameters to vary between states, as they represent the *SWAT* parameters used to calculate surface runoff. We averaged the precipitation over 2 days for Redhill Creek and 3 days for Grindstone Creek:

$$CN2 \text{ (Multiplicative Effect)}_t = CN2_{low} \text{ for } \text{Log}_{10}(\text{n-day averaged precipitation} + 1) \leq \theta_p \text{ (1a)}$$

$$CN2 \text{ (Multiplicative Effect)}_t = CN2_{high} \text{ for } \text{Log}_{10}(\text{n-day averaged precipitation} + 1) > \theta_p \text{ (1b)}$$

where $CN2 \text{ (Multiplicative Effect)}_t$ refers to the value of the multiplicative effect for the curve numbers at time t , θ_p refers to the threshold between the two states, $CN2_{low}$ and $CN2_{high}$ refer to the state-specific values of the multiplicative effects applied to the curve number parameters, and n is equal to 2 for Redhill Creek and 3 for Grindstone Creek.

We quantified this hydrologic response time n using measured streamflows. Using the daily flows measured at Redhill and Grindstone Creeks between the years 1988 and 2009, we compute a 1-day correlation coefficient of $\rho = 0.43$ for Redhill Creek and $\rho = 0.83$ for Grindstone Creek. Following Yang et al. (2007a,b), we may transform these estimates of daily correlation to a characteristic correlation time using the equation $\rho = \exp\left(-\frac{\Delta t}{\tau}\right)$, where Δt is the time step (1 day) and τ is the characteristic correlation time in days. This yields $\tau = 1.15$ days for Redhill Creek, which we rounded up to two days to incorporate some degree of memory in the system. For Grindstone Creek, this method gives us $\tau = 5.37$ days. We assessed the differences obtained when using $n = 3$ and $n = 5$ days for Grindstone Creek using a linear changepoint regression with

Log_{10} (n-day averaged precipitation + 1) as the independent variable and Log_{10} (daily streamflow + 1) as the dependent variable. The values of the model precisions were within the 95% credible intervals of each other, meaning neither model fit better than the other. We used the mean values of the changepoints together with the respective average precipitation time series to classify each day into the normal or extreme state. Of all the days between 1988 and 2009 used to calibrate the changepoint models, only 479 of 3305 days (less than 15% of the total) were not classified into the same state by both the 5 day and 3 day averaging period. We admit that this is not a trivial difference. It is however small enough to assume that using a 5 day averaging period would not have resulted in a significantly different outcome for the study overall.

References

- Yang, J., Reichert, P., Abbaspour, K.C. and Yang, H., 2007a. Hydrological modelling of the Chaohe basin in China: Statistical model formulation and Bayesian inference. *J. Hydrol*, 340:167-182, doi:10.1016/j.jhydrol.2007.03.006.
- Yang, J., Reichert, P., Abbaspour, K.C., 2007b. Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland), *Water Resour. Res.*, 43, W10401, doi:[10.1029/2006WR005497](https://doi.org/10.1029/2006WR005497).