Journal of Great Lakes Research 40 Supplement 3 (2014) 102–116

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

Accommodating environmental thresholds and extreme events in hydrological models: A Bayesian approach

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 14 June 2013
Accepted 12 February 2014
Available online 6 June 2014

Communicated by Ram Yerubandi

Index words:
SWAT model
Bayesian inference
Uncertainty analysis
Extreme events
Runoff estimation
Hamilton Harbour

ABSTRACT

Extreme events appear to play an important role in pollutant export and the overall functioning of watershed systems. Because they are expected to increase in frequency as urbanization and recent climate change trends continue, the development of techniques that can effectively accommodate the behavior of watersheds during extreme events is one of the challenges of the contemporary modeling practice. In this regard, we present a Bayesian framework which postulates that the watershed response to precipitation occurs in distinct states. Precipitation depth above a certain threshold triggers an extreme state, which is characterized by a qualitatively different response of the watershed to precipitation. Our calibration framework allows us to identify these extreme states and to characterize the different watershed behavior by allowing parameter values to vary between states. We applied this framework to SWAT model implementations in two creeks in the Hamilton Harbour watershed of Redhill Creek, an urban catchment, and Grindstone Creek, an agricultural one. We found that our framework is able to coherently identify watershed states and state-specific parameters, with extreme states being characterized by a higher propensity for runoff generation. Our framework resulted in better model fit above the precipitation threshold, although there were not consistent improvements of model fit overall. We demonstrate that accommodating threshold-type of behavior may improve the use of models in locating critical source areas of non-point source pollution.

Introduction

Hydrology has long been concerned with extreme events in the form of floods (Gumbel, 1954). Recent developments in the field have suggested that extreme events may play an important role in the overall functioning of watershed systems, despite their relatively low frequencies of occurrence (Macrae et al., 2007; Shields et al., 2008). In particular, extreme hydrological events have been found to significantly contribute to the overall export of nitrogen and phosphorus in agricultural (Macrae et al., 2007) as well as urban systems (Duan et al., 2012; Shields et al., 2008). There is evidence that both urbanization (Duan et al., 2012; Shields et al., 2008) and climate change (Kunkel et al., 2013) will make extreme hydrological and nutrient export events more common in the future.

Watershed modeling can play a key role in advancing our understanding of the likely effects of an increased frequency of extreme events on water quality (Rode et al., 2010). For instance, Michalak et al. (2013) used the Soil-Water Assessment Tool (*SWAT*) to estimate the unusually high nonpoint source soluble phosphorus inputs associated with the largest algal bloom in Lake Erie's history. However, continuous watershed models typically focus on the processes or variables responsible for the “average” system behavior. Due to their infrequency, extreme events and any processes (or dynamics) associated with them will usually not be considered in the model development process and are often relegated to the role of “outliers”. This is perhaps why hydrological model parameter studies often find that the ideal parameter set for modeling high flow conditions is different from that used when modeling the entire range of flows (e.g., Cibin et al., 2010; Zhang et al., 2011). There is empirical evidence that this is not an artifact of mathematical models, but a genuine reflection of the thresholds which do in fact operate in hydrological systems at the hillslope and watershed scales. As mentioned previously, extreme events can represent a significant proportion of annual fluxes of water or materials out of a watershed and should be better considered by continuous models.

Empirical work in hydrology has found that extreme events can result from different flow mechanisms than more common events. McDonnell (1990), for instance, found that flow through the soil matrix was responsible for small runoff events, whereas macropore flow tended to be responsible for larger events. Subsequent empirical work has found that the initiation of macropore flow tended to occur when the soil was close to saturation (Zehe et al., 2001). While an explicit

[*] Corresponding author at: Watershed Hydrology Group, School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario, L8S 4L8, Canada. Tel.: +1 647 239 5138.
E-mail address: wellenc@mcmaster.ca (C. Wellen).

http://dx.doi.org/10.1016/j.jglr.2014.04.002
0380-1330/© 2014 International Association for Great Lakes Research. Published by Elsevier B.V. All rights reserved.

introduction of this two-domain conceptualization of flow into numerical watershed models did prove to be feasible, doing so required prohibitively extensive field data (Zehe et al., 2001). Other environments, such as the Canadian Shield, are characterized by “fill-and-spill” mechanisms where certain cascading storages in bedrock depressions must be filled before the catchment as a whole is able to export significant water volumes (Ali et al., 2013; Oswald et al., 2011). In many watersheds of management interest, threshold behavior may be at work in differentiating the response to extreme precipitation events; yet, obtaining a detailed process understanding and explicitly representing this behavior in our mathematical models is typically not feasible. We are focused on such cases in this paper.

In a review of threshold behavior of hydrological systems, Zehe and Sivapalan (2009) identified two strategies for accommodating threshold behavior in watershed models. The first strategy is explicit through the model equations, as done, for instance, by Zehe et al. (2001) with their two-domain model of soil water movement. The second, less commonly pursued, strategy is to assume that, as we do here, the system operates in multiple states or modes of behavior, the identification of which is a component of the model calibration process. Our novel contribution to the study of extreme events in watershed modeling is an example of this second strategy. We posit that extreme events may be modeled as a different response of a system to precipitation inputs above a threshold. That is, the system may be thought of as having distinct states of response to precipitation. This approach to extreme events is in agreement with empirical and theoretical developments in the field (Ali et al., 2013; Zehe and Sivapalan, 2009; Zehe et al., 2001).

Bayesian inference provides an approach to model calibration, which is uniquely suited to the problem of identification of latent states (Gelman et al., 2004; Prado and West, 2010). Applications of Bayesian inference techniques to accommodate different states of a watershed system have typically focused on one of two techniques: mixture likelihoods and time varying parameters. Employing mixture likelihood explicitly accounts for multiple watershed states by assuming the model residuals represent different populations with different statistical properties. Yang et al. (2007a), for instance, showed that the residuals from (pre-specified) dry and wet seasons have distinct variances and temporal correlation patterns, while Schaefli et al. (2007) showed that the class membership of particular residuals could be identified as part of the model training exercise. The main advantage of mixture likelihoods is to essentially weight different residuals more or less strongly, which serves to quantitatively express an expectation that the model would not perform consistently well throughout its temporal domain, e.g., we expect that a model would not reproduce the high flow periods as closely as the baseflow conditions (Yang et al., 2007a). In doing so, we avoid both biasing the model calibration in favor of the more uncertain states of the system and overestimating the residual variance of the states which are characterized by lower uncertainty.

While useful and statistically coherent, mixture likelihoods do not allow us to accommodate the different processes, which may be operating in the different members of the mixture of the residuals. In fact, Schaefli et al. (2007) found that the use of a mixture likelihood led to a higher residual variance for large events. This would serve to decrease the impact of these events on the overall likelihood function value, leading to a model calibration which would tolerate very large residuals during the extreme events instead of reducing them, precisely the opposite of what we here aim to do. A second strategy to accommodate extreme events is to allow the model parameters to vary in time. This type of approach has mainly opted for a continuous evolution of parameter values through time, either in a manner analogous to the Kalman filter (Kalman, 1960), where parameter values are adjusted at each time step to allow a better correspondence of the model and the data, or with a type of random walk, where parameter values may change significantly at each time step. Such approaches may be stationary (Reichert and Mieleitner, 2009) or non-stationary (Lin and Beck, 2007). While suitable for tracking gradual changes in watershed functioning, such approaches may not sensibly identify a number of distinct states of watershed functioning; especially, if we consider that using different parameters for every time step likely results in an over-parameterization of the model.

In this study, we take a state-specific approach, founded upon a multivariate Bayesian approach, which effectively balances the need to relax rigid model structures but does not introduce the complexity typically entailed by continuous parameter evolution in time (Arhonditsis et al., 2008a,b; Wellen et al., 2012). We discuss the mathematical details of our approach in the methodology section, but we essentially posit that a threshold of precipitation exists above which the watershed is characterized by different parameter values relative to those used to parameterize the model below the threshold. However, the values of a particular parameter are not independent between the two watershed states or modes of behavior but are characterized by a covariance structure to be identified during model calibration. Finally, we critically discuss the prospect of the present methodological framework to offer an effective means for reproducing watershed dynamics during extreme events.

Methodology

Incorporating threshold behavior in model parameter estimation

We may think of a deterministic hydrological model as a function, which connects a time series of environmental inputs with a time series of streamflow outputs:

$$Y = f(\gamma, \theta) \tag{1}$$

where Y indicates the time series of model predictions (e.g., streamflow, chloride concentration), f indicates the model, γ indicates the various time series of environmental inputs (e.g., precipitation, temperature, wind speed, crop rotations), and θ indicates the vector of model parameters. Because both the watershed model and the measurements from the system are subject to substantial uncertainty, we typically introduce a term to describe their mismatch, and so we re-write Eq. (1) as:

$$Y = f(\gamma, \theta) + \varepsilon, \tag{2}$$

where ε indicates the time series of model residuals, defined as the difference between measurements and model predictions. The statistical treatment of the ε time series has been the subject of considerable work in hydrology — the emerging consensus is that it is typically autocorrelated and non-Gaussian (Schoups and Vrugt, 2010; Sorooshian and Dracup, 1980; Yang et al., 2007b). The error characterization can guide us in drawing statistically sound inference, and the mismatch of model predictions and system measurements is mainly due to the simplifications employed when constructing models of highly complex natural systems, such as watersheds. Thus, reducing the model structural error, a significant component of the ε time series, requires creating better models. While ultimately better models must be arrived at by using systems of equations which more accurately represent the environmental system in question, we contend that a significant step in this direction can be made by relaxing the assumptions made by the inference procedure. Namely, we relax the assumption that the vector of model parameters (θ) is constant for all time steps.

Frameworks for relaxing this assumption have been proposed before, but all of them tend to favor either the replacement of a subset of the parameter vector θ with a stochastic process in time (e.g., Reichert and Mieleitner, 2009; Wellen et al., 2012) or the relaxation of a subset of the parameter vector θ to evolve in time (e.g., Lin and Beck, 2007). We here postulate that rather than a gradual evolution or a continuum of system responses to climate forcing, watersheds can be thought of as characterized by multiple discrete states of response. We take an approach philosophically similar to a class of models called Markov-switching models, which posit that the model parameters depend on the identity of discrete, unobserved, time-dependent states (Prado and West, 2010). We further postulate that values of precipitation exceeding a threshold can trigger a shift to an alternative state of watershed response. In the future, our framework can easily be generalized to other external forcing factors (e.g., temperature) or even internal state variables (e.g., soil water storage). We postulate that above some threshold of precipitation θ_p, a subset of the parameter vector is characterized by different values than when the system forcing is below the threshold:

$$\begin{aligned}&\theta_t = \theta_{\text{low}} \text{ for } \gamma_{p,t} \leq \theta_p\\&\theta_t = \theta_{\text{high}} \text{ for } \gamma_{p,t} > \theta_p\end{aligned} \tag{3}$$

where $\gamma_{p,t}$ refers to the value of precipitation averaged for some period before time t, θ_t refers to the value of the parameter vector at time t, θ_p refers to the threshold between the two states, and θ_{low} and θ_{high} refer to the state-specific values of the parameter vector θ. Extending this framework to an arbitrary number of states is straightforward although the rest of the manuscript will assume the presence of only two states separated by one threshold. This assumption was made because our case study has evidence of only two states (see Case study & model selection section for more detail). Referring back to our original notation, we may now write the deterministic model as:

$$Y_t = f(\gamma, \theta_{\text{low}}) + \varepsilon_t \text{ for } \gamma_{p,t} \leq \theta_t \tag{4a}$$

$$Y_t = f\left(\gamma, \theta_{\text{high}}\right) + \varepsilon_t \text{ for } \gamma_{p,t} > \theta_t \tag{4b}$$

Note that our framework adds a discrete number of parameters equal to $z + 1$, where z is the number of parameters assumed to vary between states. The example which follows presents the simplest possible implementation of our framework, where only two states exist, and where only one parameter varies by state. This adds an additional two parameters to the model, likely a much smaller increase of complexity than allowing the parameters to vary with each time step.

In a final note regarding our framework, the inclusion of the threshold θ_p as a random variable instead of a single, pre-specified quantity implies that a) membership of each time step in one state or another is to be inferred rather than assumed; and b) the state membership of each time step is described probabilistically rather than deterministically. That is, each model iteration is characterized by a particular value of θ_p and thus each day is classified into the normal or the extreme state. When we have a number of model iterations, the probability of a day being classified as extreme is simply the number of iterations in which the day is classified as extreme divided by the total number of iterations. The same is true for the normal state. Thus, the proposed framework essentially draws inference at a given time step on model parameters and states simultaneously.

Bayesian inference framework

Inference is founded upon Bayes' Theorem, expressed as:

$$\pi(\theta|Y) = \frac{\pi(\theta)L(Y|\theta)}{\int_\theta \pi(\theta)L(Y|\theta)d\theta} \tag{5}$$

where $\pi(\theta)$ represents our prior beliefs regarding 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. We simulated samples from the posterior using Markov chain Monte Carlo (MCMC) sampling. In this study, we used the DiffeRential Evolution Adaptive Metropolis Algorithm-ZS (DREAM-ZS) as presented by Laloy and Vrugt (2012), the details of which we include in the Electronic Supplementary Material (ESM). Our likelihood function accounted for the correlation of the residuals through time using an AR1 approach (Sorooshian and Dracup, 1980):

$$\begin{aligned}&\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t\\&\delta_t \sim N\left(0, \sigma_v^2\right)\end{aligned} \tag{6}$$

Conceptually, this approach assumes that the residuals can be decomposed into two components — an independent one called an innovation (δ_t) and a temporally correlated component ($\rho\varepsilon_{t\ -\ 1}$), which describes how long the effects of past innovations linger in the time series. Our preliminary tests indicated that the distributions of the innovations could be accounted for using Student's t-distribution as the basis for the likelihood (Yang et al., 2007b):

$$\begin{aligned}L(Y|\theta) = &\frac{\left(1-\rho^2\right)\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}\sigma_v} \times \left(1 + \left(1-\rho^2\right) \times \frac{\varepsilon_1^2}{v\sigma_v^2}\right)^{-\frac{v+1}{2}}\\&\times \prod_{t=2}^{T} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}\sigma_v} \times \left(1 + \frac{\delta_t^2}{v\sigma_v^2}\right)^{-\frac{v+1}{2}}\end{aligned} \tag{7}$$

where v refers to the degrees of freedom and Γ refers to the gamma function. During our preliminary model runs we calculated the innovations as $\delta_t = \varepsilon_t - \rho\varepsilon_{t\ -\ 1}$ and fit them to a Student's t distribution. We found that the innovations could be described with a Student's t distribution with 7 degrees of freedom. We accommodated the heteroscedasticity of the residuals using a log transformation $Y' = \ln(Y + 1)$.

Detailed statistical formulations

We tested three statistical formulations for parameter estimation and inference. The first statistical formulation (Formulation 1) is a Bayesian update of the hydrological model which does not allow any parameters to vary when thresholds are crossed. This formulation is intended to serve as a benchmark to compare to our other formulations. All prior parameter distributions were uniform over their range. The second statistical formulation (Formulation 2) implements the deterministic model using the framework presented in Eq. (4a,4b). In keeping with our assumption that the two states of watershed response are distinct but related, we treat θ_{low} and θ_{high} as draws from a bivariate normal distribution. Doing so has the advantage of removing the component of the covariance from the marginal variances of the two parameters, allowing better defined marginal posterior parameter distributions and posterior predictive distributions (Bates et al., 2003; Wellen et al., 2014).

Specifically, we postulate that the parameters θ_{low} and θ_{high} are characterized by a bivariate normal distribution:

$$\Theta \sim MVN(\mu, \textstyle\sum) \tag{8}$$

where $\Theta = \begin{bmatrix}\theta_{\text{low}}\\\theta_{\text{high}}\end{bmatrix}$ represents the vector of state-specific parameters; $\mu = \begin{bmatrix}0.0\\0.0\end{bmatrix}$ is the vector of prior means; and Σ denotes the covariance matrix. We use a Wishart distribution to describe our prior knowledge on the inverse of the covariance matrix:

$$\textstyle\sum^{-1} \sim Wish\left(R^{-1}, v\right) \tag{9}$$

where v denotes the degrees of freedom (set equal to 3) and R denotes the prior covariance matrix. We assumed fairly flat priors and near-independence for the covariance matrix:

$$R = \begin{bmatrix} 1000 & 0.001 \\ 0.001 & 1000 \end{bmatrix} \tag{10}$$

Note that we still posit the independence of the threshold θ_p from the values of the multiplicative effects. All prior parameter distributions other than Θ and Σ were uniform over their range, implying no prior relationship between the threshold θ_p and the change point presented in Fig. 2.

Our third statistical formulation (Formulation 3) uses the same framework as our second formulation, but introduces informative prior parameter distribution. Informative priors, one of the unique advantages of the Bayesian approach to model calibration, allow information about plausible model parameter values to be specified before the model is calibrated to observed data. The informative prior distributions are presented along with the model description in the subsequent section. Note that Formulation 3 uses the posterior of the change point presented in Fig. 2 as the informative prior for the threshold θ_p, in effect assuming that the change point in Fig. 2 represents our best estimate of the threshold θ_p. Formulations 2 and 3 both involve some inference of the state of each time step. To estimate the probability that each time step belonged to each state, we calculated the binary state membership of each time step for each MCMC sample according to Eq. (4a,4b), and then calculated the mean and 5th and 95th percentiles of these binary state memberships for all the MCMC samples.

Model evaluation

We assessed the performance of all models using four metrics: the coefficient of determination (r^2), Nash and Sutcliff's (1970) index of model efficiency (NSE), the relative error as calculated by Arhonditsis and Brett (2004; RE), and the logarithm of the likelihood function (Eq. (7); LogLike). Following Hong et al. (2005), we assessed the degree of updating of the informative prior distributions of Formulation 3, using the delta index by Endres and Schindelin (2003), which quantifies the difference in shape of two parameter distributions:

$$\delta_{\theta i} = \sqrt{\int \left(\pi(\theta_i) \log \frac{2\pi(\theta_i)}{\pi(\theta_i) + \pi(\theta_i|Y)} + \pi(\theta_i|Y) \log \frac{2\pi(\theta_i|Y)}{\pi(\theta_i) + \pi(\theta_i|Y)} \right) d\theta} \tag{11}$$

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 $(2\log2)^{1/2}$ 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 updating by computing the percent difference between the prior and posterior medians.

Case study & model selection

Case study

The study site is a pair of catchments, Redhill and Grindstone Creeks, situated in the drainage basin of Hamilton Harbour, a large embayment at the western end of Lake Ontario (Fig. 1). Aside from the land use patterns, the two Creeks are quite 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, with most of the remainder composed of rocky outcroppings and ravines. Soils are evenly spread between the four Natural Resources Conservation Service's soil hydrologic runoff groups — groups A and B, those least likely to generate runoff, have 23% coverage, respectively, group C has 29% coverage, and group D, the group most likely to generate runoff, has 24% coverage. The slopes of the Harbour basin are mild, with the exception of the Niagara Escarpment. The average slope of the entire basin is 4.4%, and ignoring all slopes greater than 30% the average is 3.8%. The mean elevation of the basin is 130 m above sea level, with elevation ranging from 318 m above sea level to 74 m above sea level.

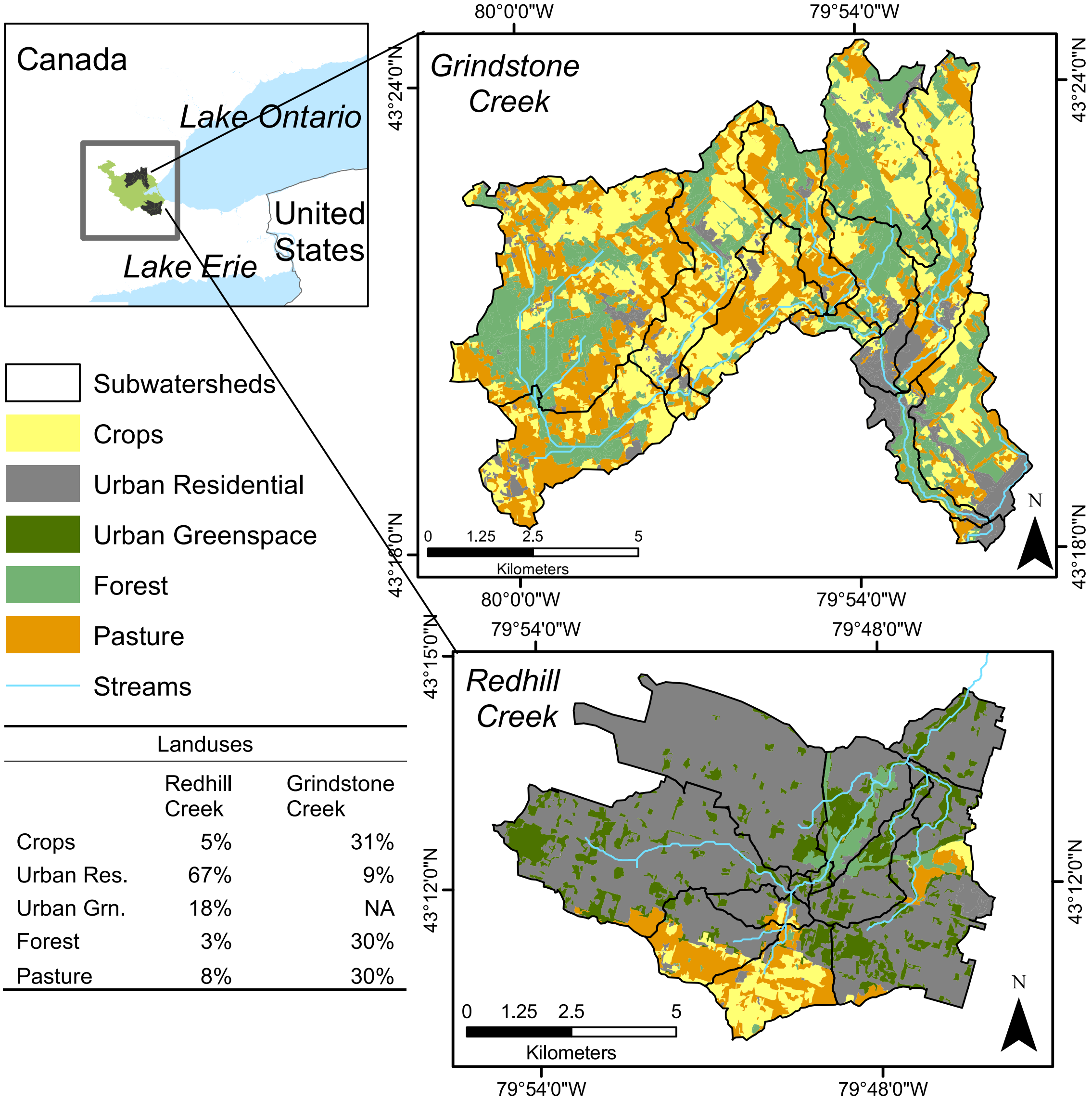
The meteorological data for this study come from Environment Canada's Hamilton Airport station (WMO ID 71263; http://www.climate.weatheroffice.gc.ca/climateData/canada_e.html), while the daily flow information comes from the Water Survey of Canada's gauges at Redhill (02HA014) and Grindstone Creeks (02HB012; http://www.ec.gc.ca/rhc-wsc/default.asp?lang=En). The basin has a humid continental climate, with daily temperatures ranging from −10 °C to −2 °C in January and 15 °C to 26 °C in July. The Harbour basin receives 910 mm of precipitation annually, 146 mm of which occurs as snowfall. To characterize the soils, we used the Soil Landscapes of Canada dataset v.3.2 from Agriculture and Agri-Food Canada (http://sis.agr.gc.ca/cansis/nsdb/slc/index.html).

Redhill Creek drains an area of approximately 63 km^2, of which 66% is urban residential area, 17% is urban greenspace, 12% is agricultural area and 4% is forested. Of the urban area, 50% is impervious and 40% of the total urban area is directly connected to a storm sewer system. Towards the mouth in the City of Hamilton there are some connections with the city's combined sewer system. Grindstone Creek drains an area of approximately 87 km^2, 60% of which is agricultural land evenly split between pasture and cropland. Of the remainder, 30% is forested and 9% is urban. This urban land encompassing the town of Waterdown is serviced by storm sewers. The flows of both Creeks are unregulated.

An examination of the daily flows of Redhill and Grindstone Creeks supports the idea of a single threshold separating two states of response of the two Creeks to precipitation. Fig. 2 shows scatterplots of Log$_{10}$ transformed daily flows and averages of the previous 2 or 3 days of precipitation along with the fitted piecewise regressions. These periods were chosen to implicitly include the effect of antecedent moisture. The data used are from the period 1988–2009, representing the months from May through November. Redhill Creek's threshold was estimated at 0.94 ± 0.05 transformed previous 2-day average precipitation and represents a clear change in response above that threshold. This threshold corresponds to a 2-day average of 7.7 ± 1.1 mm, and would be reached by one day with 15.2 mm of precipitation or two days of 7.7 mm. Grindstone Creek's threshold, estimated at 0.78 ± 0.09 transformed previous 3-day average precipitation. This threshold corresponds to a 3-day average of 5.0 ± 1.2 mm. In light of these findings, the threshold θ_t of watershed response to precipitation was specified in units of $Log_{10}\left(\left(\frac{1}{n}\sum_{t-n}^{t} P_t\right)+1\right)$, where t indicates a particular day in the time series, P refers to the precipitation on day t (mm), and n was set equal to 2 for Redhill Creek and 3 for Grindstone Creek. Note that the actual value of θ_t was subject to calibration.

Model selection

While our inference framework is applicable to any watershed model, we used the Soil-Water Assessment Tool for this application (SWAT; Arnold et al., 1998). SWAT is a semi-distributed model and is typically used to evaluate the effects of alternative management schemes on agricultural landscapes. Watersheds are disaggregated by topography into subbasins, and then further disaggregated into hydrological response units (HRU) on the basis of land use, soil type, slope, and land management. The rationale is that all the area in a given HRU behaves as a homogenous unit, and all aspects of the land phase of the hydrological cycle are executed at the HRU scale and then aggregated to the subbasin scale, where routing is executed. SWAT models the response of a watershed to precipitation or snowmelt using a water



Landuses		
	Redhill Creek	Grindstone Creek
Crops	5%	31%
Urban Res.	67%	9%
Urban Grn.	18%	NA
Forest	3%	30%
Pasture	8%	30%

Fig. 1. Map of study catchments and their land uses.

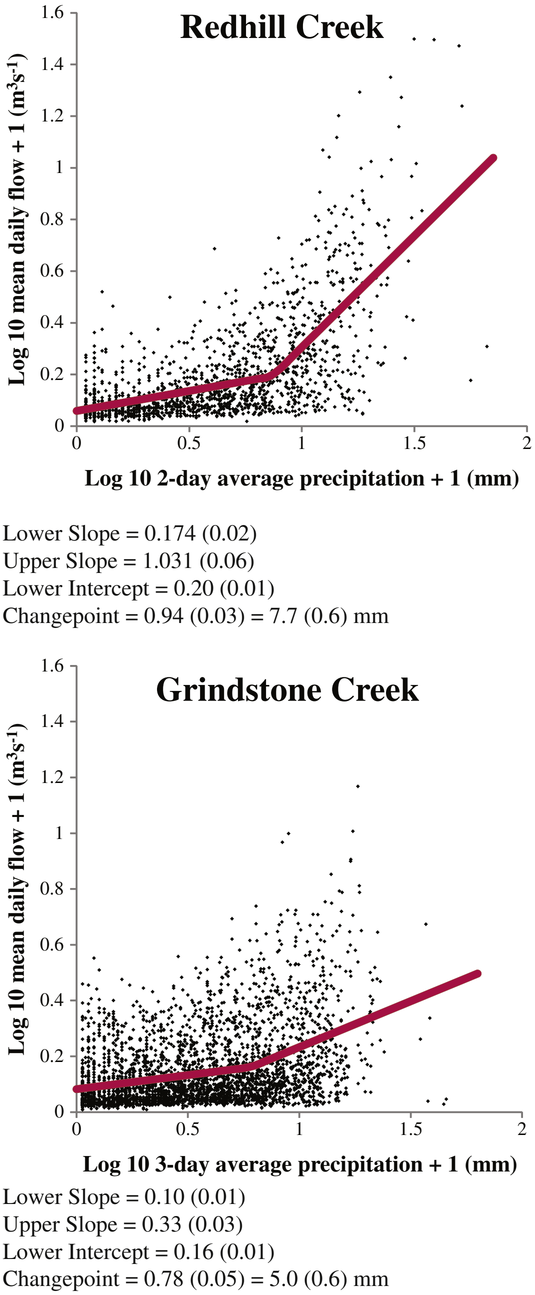
balance approach. Runoff is computed using a version of the United States' National Resources Conservation Service's Curve Number (CN) methodology, an empirical approach where a daily CN varies nonlinearly with soil moisture. The CN for a standardized level of moisture (referred to as moisture condition 2) is treated as a calibration parameter. Readers seeking additional information about the SWAT model may consult Neitsch et al. (2011).

Our calibration approach proceeded by setting realistic starting values for each HRU or subbasin and calibrating by adjusting all the parameters at once with global multiplicative effects (e.g., all the curve numbers were decreased by 5%). These multiplicative effects were the calibration parameters. While this does not allow us to calibrate land use or soil type specific values of each parameter, it has been shown to lead to reasonable predictions at the watershed outlet (Cerucci and Conrad, 2003; Yang et al., 2007a,b). In Table 1, we provide the calibration vector employed. We arrived at this vector after a review of the literature, including the SWAT manual (Arabi et al., 2007; Ekstrand et al., 2010; Neitsch et al., 2011; Rouhani et al., 2007; van Griensven et al., 2006; Yang et al., 2007a,b). Also note that informative priors were employed only in Formulation 3; Formulations 1 and 2 use uniform priors in the range given in Table 1, with the exception of the innovation variance, the inverse of which was given an uninformative gamma prior for all formulations. We used a total of 12 subbasins for Redhill Creek (median area = 4 km^2, interquartile range = 5.36 km^2) and 14 subbasins for Grindstone Creek (median area = 5.58 km^2, interquartile range = 7.03 km^2). We used a total of 41 HRUs for Redhill Creek (median area = 0.57 km^2, interquartile range = 1.3 km^2) and a total of 64 HRUs for Grindstone Creek (median area = 1 km^2, interquartile range = 1.63 km^2).

There appeared to be a threshold of time averaged precipitation above which runoff is generated at a greater rate than below (Fig. 2). Therefore, we allowed the curve number for moisture condition 2 to vary between states of the watershed. Updating our notation in Eq. (3) for this application, we get:

$$\begin{aligned} &CN2\ \ (Multiplicative\ \ Effect)_t \\ &\quad = CN2_{low}\ \text{for}\ \mathrm{Log}_{10}(\mathrm{n{-}day\ average\ precipitation}+1) \le \theta_p \end{aligned} \tag{12a}$$

$$\begin{aligned} &CN2\ \ (Multiplicative\ \ Effect)_t \\ &\quad = CN2_{high}\ \text{for}\ \mathrm{Log}_{10}(\mathrm{n{-}day\ average\ precipitation}+1) > \theta_p \end{aligned} \tag{12b}$$



Lower Slope = 0.174 (0.02)
Upper Slope = 1.031 (0.06)
Lower Intercept = 0.20 (0.01)
Changepoint = 0.94 (0.03) = 7.7 (0.6) mm

Lower Slope = 0.10 (0.01)
Upper Slope = 0.33 (0.03)
Lower Intercept = 0.16 (0.01)
Changepoint = 0.78 (0.05) = 5.0 (0.6) mm

Fig. 2. Piecewise regression graphs relating the 2- or 3-day average precipitation to the daily streamflow. Scatterplots show daily flows from 1988 to 2009. Only data from the months May–November are plotted. Statistics below graphs show the means and, in parentheses, standard deviations of the parameters of the regressions.

where *CN2 (Multiplicative Effect)$_t$* indicates the multiplicative effect employed on the curve numbers at time *t*, and *n* is equal to 2 for Redhill Creek and 3 for Grindstone Creek.

We used SWAT 2009 with the Penman–Monteith submodel for potential evapotranspiration and the variable travel time method of stream routing, and employed a daily time step for model calculations and fitting (likelihood calculations). For the agricultural areas, we used SWAT's generic agricultural crop. This generic crop assumes parameters fairly similar to corn, the second-most common crop grown in the Hamilton Census Division in 2011 (35% of total cropped area). Our preliminary tests indicated corn was an adequate representation of the agricultural area, and we did not have any information about how crops grown varied spatially. Recommended practices for growing corn in this region entail an addition of 200 kg of N fertilizer and 20 kg of P fertilizer per hectare in two equal inputs in the spring and soil tillage once in the spring and once in the fall (www.omafra.gov.on.ca/english/crops/pub811.htm); we assumed these practices were applied here. We also assumed all pervious urban land cover was bluegrass, and all forest was mixed. We obtained estimates of urban soil fertilizer additions from Law et al. (2004). After a 1-year spin-up period, the model was calibrated to daily flows in the years 1992–1994, and subsequently validated with daily flows from 1995 to 1998.

Results

Assessment of MCMC update

The MCMC algorithm showed signs of convergence after 20,000 to 45,000 iterations per chain for Redhill Creek and after 25,000 to 95,000 iterations for Grindstone Creek. The values of the Gelman–Rubin statistic were lower than 1.1 for all parameters, with most of them below 1.05, indicating acceptable convergence. Inspections of trace plots showed stationarity, and marginal density plots of the parameters showed reasonably well-shaped distributions, though as can be seen in the Electronic Supplementary Material (ESM; Figs. S1–S6) some multi-modality is present. This pattern may partly stem from the use of Student's t as the basis of the likelihood function, given that this distribution is not log-concave (Vanhatalo et al., 2009). Further, the DREAM algorithm has been shown to avoid convergence to a single mode in the case of a multi-modal posterior space (Laloy and Vrugt, 2012; Vrugt et al., 2009), so we have cause to think about this multi-modality neither as an artifact of our analysis nor evidence of poor convergence, but as a genuine reflection of the posterior parameter space. We also assessed the assumptions made in the likelihood function, namely that i) the innovations are independent with respect to time, and ii) these innovations come from a Student's t distribution with seven (7) degrees of freedom. We present these assessments in the ESM, as Figs. S7 (Redhill Creek) and S8 (Grindstone Creek), and they show acceptable adherence to the assumptions of the likelihood function.

Model parameter posteriors

Tables 2 and 3 present the model parameter posteriors. The model parameter posteriors were fairly well identified for both Redhill and Grindstone Creeks for all formulations. In fact, many of the posteriors represent a very small segment of the parameter space (e.g., SOL_AWC, a parameter often identified as particularly sensitive; e.g., Arabi et al., 2007; van Griensven et al., 2006), although others (e.g., EPCO) tended to be less well constrained. Two of the parameters of our calibration vector were multiplicative effects on measured data — SOL_AWC, the proportion of plant available soil water, and SOL_KSAT, the saturated hydraulic conductivity of the soil layers. It is also interesting to note how far from the value of 1.0 these multiplicative effects are in all formulations, suggesting that the effective values of model parameters at the catchment scale can deviate significantly from those measured at the scale of soil surveys.

Our framework for capturing extreme events as implemented in Formulation 2 led to coherent parameter posteriors in both Creeks. 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 of these quantities for Grindstone Creek in relation to Formulation 1. In both cases, the threshold θ_p was well identified. In Tables 2 and 3 as well as Fig. 2, we present the estimated 2-day or 3-day averaged precipitation required to cause a shift to the extreme state in untransformed units (mm). The threshold θ_p tended to be on the upper edge of its range and well outside the credible interval of the change point of the piecewise regressions (Fig. 2). Note that the credible intervals of this change point were 0.89–0.99 transformed units for Redhill Creek and 0.68–0.87

Table 1
SWAT model parameters included in the calibration vector.

Parameter	Description	Range	Informative prior[**]	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) 0.5,1.5 (Grindstone)	N(1,0.455) N(1,0.455)	Assumed minimum and maximum values of 0.01 and 0.85.
GW_REVAP	Revap coefficient.	0.02, 0.2	U(0.02, 0.2)	–
ESCO	Soil evaporation 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.
EPCO	Plant transpiration compensation factor.	0.1, 0.99	B(3,1.22)	Expected value of 0.9, signifying a strong ability of lower soil layers to supply evaporative demand of the plants.
GW_DELAY	Ground water delay time (days; Multiplicative Effect).	0.5, 5	U(0.5, 5)	–
SOL_KSAT	Soil saturated hydraulic conductivity (mm/h). (Multiplicative Effect).	0.1, 10 (Redhill) 0.5,1.5 (Grindstone)	LN(0,1.15) LN(0,1.15)	Corresponds to a range of one order of magnitude.
SNOWCOVMX	Minimum snow water content corresponding to 100% aerial snow coverage (mm).	1, 40	LN(2.48,0.35)	Donald et al. (1995)
SMFMX	Snow melt factor on June 31st (mm water/°C above 0.5 °C).	1, 9	N(5.5,3.1)	Conetta (2004)[1]; Donald (1992); Yang et al. (2007b)
SMFMN	Snow melt factor on December 31st (mm water/°C above 0.5 °C).	1, 5	N(3.1,1.8)	Conetta (2004); Donald (1992); Yang et al. (2007b)
SURLAG	Lag time for surface runoff (days).	0.5, 10	LN(0,1.0)	Assumed one day was the most likely value and upper end of 95% credible interval was one week.
ρ	First order residual correlation coefficient for all days.	0.1, 0.99	U(0.1,0.99)	–
σ	Innovation standard deviation for all days.	0.002, 2000	G(0.001,0.001)	–
CN2 Low	Curve number for moisture condition 2 on low precipitation days. (Multiplicative Effect).	0.5, 1.5	N(1,0.41)	Schwab et al. (2002, p. 74)
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	−0.99,0.99	U(−0.99,0.99)	–

θ_p	Threshold of time averaged precipitation switching between curve numbers.	0.9, 1.4 (Redhill)
		0.6, 1.1 (Grindstone)
			N(0.94,0.025)
			N(0.78,0.047)
				Streamflow and Precipitation Measurements

[*] 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.

[**] 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.

[1] Conetta, M., Unpublished. Snow Disposal Sites, Conceptual Designs Part A — Snow Meltwater Characteristics and Treatment Technologies. Report Submitted to the City of Toronto, October 29, 2004.

transformed units for Grindstone Creek. Even though the change points were well identified when using 22 years of data, it is clear that the three years of data in the calibration period were not sufficient to locate the change points. The multiplicative effects of the curve numbers were generally higher above the threshold θ_p than below (e.g., CN2 High > CN2 Low), indicating that above the threshold of precipitation input, a greater amount of rainfall is converted into surface runoff than below. For both Creeks, the values of the multiplicative effects on

Table 2
Parameter posterior means and standard deviations, Redhill Creek study site.

Parameter	Formulation 1		Formulation 2		Formulation 3		Delta index (3)	Expected value shift (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	67%	2%
SOL_AWC (Mult. Eff.)	2.358	0.001	1.719	0.001	1.720	0.001	97%	72%
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	98%	−86%
EPCO	0.705	0.215	0.662	0.150	0.663	0.167	25%	−26%
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	86%	−79%
SNOWCOVMX	6.812	0.234	16.745	2.279	16.174	1.371	75%	1%
SMFMX	3.387	0.213	3.566	0.237	3.646	0.258	85%	−34%
SMFMN	2.961	0.112	3.082	0.261	2.807	0.243	76%	−9%
SURLAG	3.748	3.194	0.530	0.034	0.380	0.011	95%	−62%
θ_p	–	–	1.350 (21 mm)	0.027 (1 mm)	0.971 (8.3 mm)	0.025 (1 mm)	51%	3%
CN2 σ_{Low}	–	–	22.462	5.104	17.347	8.789	–	–
CN2 ρ	–	–	0.036	0.395	−0.250	0.626	–	–
CN2 σ_{High}	–	–	24.994	4.275	16.551	8.276	–	–
CN2 Low (Mult. Eff.)	–	–	0.572	0.051	0.538	0.039	89%	−46%
CN2 High (Mult. Eff.)	–	–	1.100	0.011	1.013	0.024	87%	1%
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	–	–
Logged Likelihood	224.377	3.900	297.162	4.724	273.937	4.151	–	–

Table 3
Parameter posterior means and standard deviations, Grindstone Creek study site.

Parameter	Formulation 1		Formulation 2		Formulation 3		Delta index (3)	Expected value shift (3)
	Mean	SD	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	87%	47%
SOL_AWC (Mult. Eff.)	1.430	0.001	1.407	0.008	1.430	0.001	99%	41%
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	92%	−80%
EPCO	0.627	0.224	0.488	0.261	0.701	0.196	12%	−46%
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	77%	−45%
SNOWCOVMX	10.926	0.400	21.064	0.369	15.872	5.057	65%	76%
SMFMX	2.871	0.226	2.794	0.111	4.085	0.535	70%	−49%
SMFMN	1.286	0.151	2.512	0.083	1.254	0.330	81%	−20%
SURLAG	0.509	0.007	0.511	0.012	0.373	0.005	97%	−49%
θ_p	–	–	1.050 (10.2 mm)	0.092 (1 mm)	0.746 (4.6 mm)	0.042 (1 mm)	34%	35%
CN2 σ_{Low}	–	–	25.016	4.850	17.134	7.783	–	
CN2 ρ	–	–	−0.026	0.393	0.061	0.641	–	
CN2 σ_{High}	–	–	25.102	4.740	15.703	9.135	–	
CN2 Low (Mult. Eff.)	–	–	0.572	0.120	1.093	0.108	61%	−43%
CN2 High (Mult. Eff.)	–	–	0.770	0.066	0.540	0.041	87%	−23%
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	–	–
Logged Likelihood	1054.631	3.436	1053.227	4.854	1074.392	5.689	–	–

the soil parameters SOL_AWC and SOL_KSAT were closer to 1.0 for Formulation 2 than for Formulation 1, providing evidence that when a threshold of catchment response is explicitly considered we are able to use more physically realistic values of soil parameters in order to reach an acceptable model fit at the basin outlet. Taken together, these results show that Formulation 2 resulted in coherent results in both Creeks, and a significantly better fit and more realistic parameterization in Redhill Creek.

With Formulation 3, we sought to incorporate informative priors into the model calibration process in order to arrive at a more realistic and better constrained parameterization of Formulation 2. Doing so allowed us to provide information about the location of the change point which the calibration data alone may not have been able to provide. The most obvious difference between Formulations 2 and 3 lies in the values of the parameters associated with the threshold of runoff generation. Formulation 3 used as informative priors the values of the change points from the regressions presented in Fig. 2, and the posteriors of θ_p lie within the 95% credible interval of the change points. With Redhill Creek, the latter result was also associated with a decrease of both curve number parameters relative to Formulation 2, and thus more days were characterized as extreme when using a lower threshold. The rest of the parameters were characterized by generally high delta index values and significant shifts of the most likely values, suggesting a significant update of the priors. Interestingly, the posteriors of Formulation 3 were generally fairly close to those of Formulation 2 for both Creeks, suggesting that Formulation 2 did converge to a realistic parameterization. In Grindstone Creek, four parameters did appear to be better constrained or in better agreement with empirical data as a result of the informative priors. The parameter EPCO controls the extent to which lower soil layers may supply water in demand by plants which cannot be supplied by the upper soil layers. Values closer to 1 indicate more water may be supplied by lower layers. For vascular plants, we would expect a value close to 1 except in the case of very deep soils. Formulation 3 had a significantly higher and better constrained value of EPCO than the other formulations. The parameter SNOWCOVMX has been measured empirically elsewhere in Southern Ontario. The posterior estimate of SNOWCOVMX for Grindstone Creek obtained with Formulation 2 is high for this area and land cover, while the value obtained with Formulation 3 is more realistic (Donald et al., 1995).

Watershed scale model predictions

The models were characterized by respectable performances, as depicted by the NSE application on their mean predictions. [The metrics of fit of the models during calibration and validation are presented in the ESM as Tables S1 and S2.] The optimal model formulation varied by case study. For Redhill Creek, the NSE ranged from 0.6 to 0.66 during calibration and from 0.52 to 0.56 during validation. During both calibration and validation, Formulation 2 resulted in the best overall fit.

The results differed somewhat for Grindstone Creek, where the NSE ranged from 0.71 to 0.74 during calibration and from 0.44 to 0.56 during validation. Formulation 3 had the highest NSE during the calibration phase but the lowest during the validation; a result that is typically perceived as “model overfitting”. During the validation, Formulation 1 resulted in the best overall fit. Figs. 3 and 4 present the time series predictions of the various statistical formulations.

The mathematical framework introduced for accommodating extreme events allows us to estimate the number of events which can be classified as extreme. Figs. 3 and 4 present in dark black bars the precipitation on days having a 5% or greater probability of belonging to the extreme state. For Redhill Creek's validation period (1461 days), when using Formulation 2, 1436 days had less than 5% probability of belonging to the extreme state, 7 days had greater than 99% probability of belonging to the extreme state, and 10 days had between 5% and 99% chance of belonging to the extreme state. With Formulation 3, 1347 days had less than 5% probability of belonging to the extreme state, 35 days had greater than 99% probability of belonging to the extreme state, and 79 days had between 5% and 99% chance of belonging to the extreme state. We quantified the Root Mean Squared Error (RMSE) of the mean model predictions on the days with at least a 5% chance of being classified as extreme. With Formulation 2, the RMSE on these days was 8.25, compared to 9.99 for the corresponding days simulated with Formulation 1. With Formulation 3, the RMSE on the days with at least a 5% chance of being classified as extreme was 4.09, and on the corresponding days for Formulation 1 it was 4.92. From this finding, we can conclude that Formulations 2 and 3 were both improvements over the standard calibration technique regarding the estimation of daily flows on extreme days for Redhill Creek.

For Grindstone Creek's validation period (1461 days), when using Formulation 2, 1305 days had less than 5% probability of belonging to

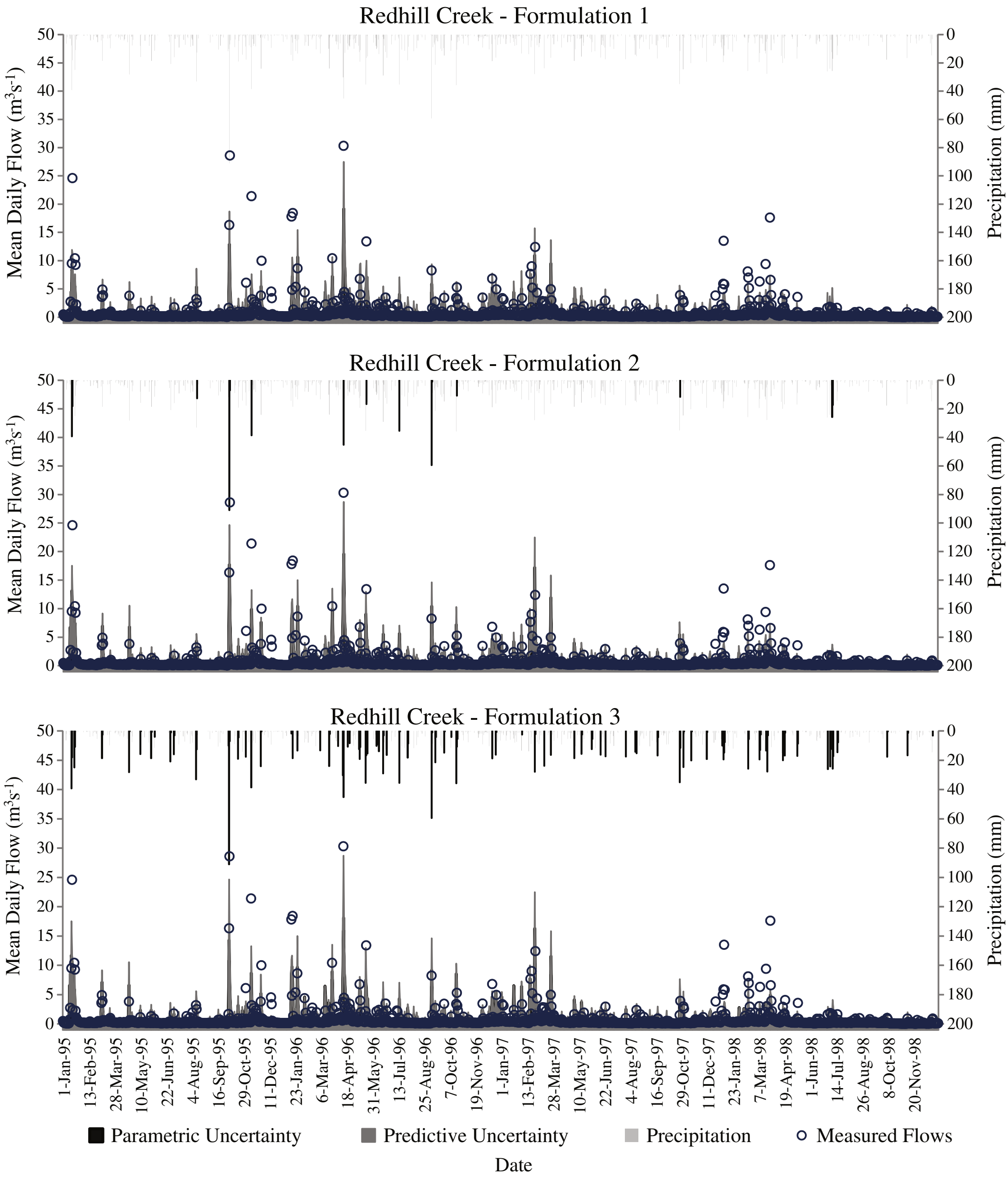


Fig. 3. Flow validation, Redhill Creek study site. Black precipitation bars indicate days with at least a 5% chance of exceeding the threshold for extreme events. Note that the remainder of the validation period (1995–1996) is included in the Electronic Supplementary Material (Fig. S9).

the extreme state, 63 days had greater than 99% probability of belonging to the extreme state, and 80 days had between 5% and 99% chance of belonging to the extreme state. With Formulation 3, 1065 days had less than 5% probability of belonging to the extreme state, 133 days had greater than 99% probability of belonging to the extreme state, and 182 days had between 5% and 99% chance of belonging to the extreme state. With Formulation 2, the RMSE on the days with at least a 5% chance of being classified as extreme was 2.14, compared to 2.39 for the corresponding days simulated with Formulation 1. With Formulation 3, the RMSE on the days with at least a 5% chance of being classified as extreme was 1.87, and on the corresponding days simulated with Formulation 1 it was 1.95. Thus, Formulations 2 and 3 were both improvements over the standard calibration technique regarding the estimation of daily flows on extreme days for Grindstone and Redhill Creeks, despite the fact that these formulations were not always characterized as having a better overall fit.

Runoff areas and water balance

A distributed model, such as SWAT, allows insights into the magnitudes and uncertainties of various water flow pathways as well as the sources of runoff within the catchment. We present here an overall water balance of the two catchments and then estimate the sources of surface runoff by land use, soil type, and subbasin. We also note that although these estimates are based on only one rain gauge, averaging predictions across space and time, as we do here, has been shown to considerably reduce the error associated with a sparse rain gauge network (Chaplot et al., 2005). Table 4 presents an annual water balance

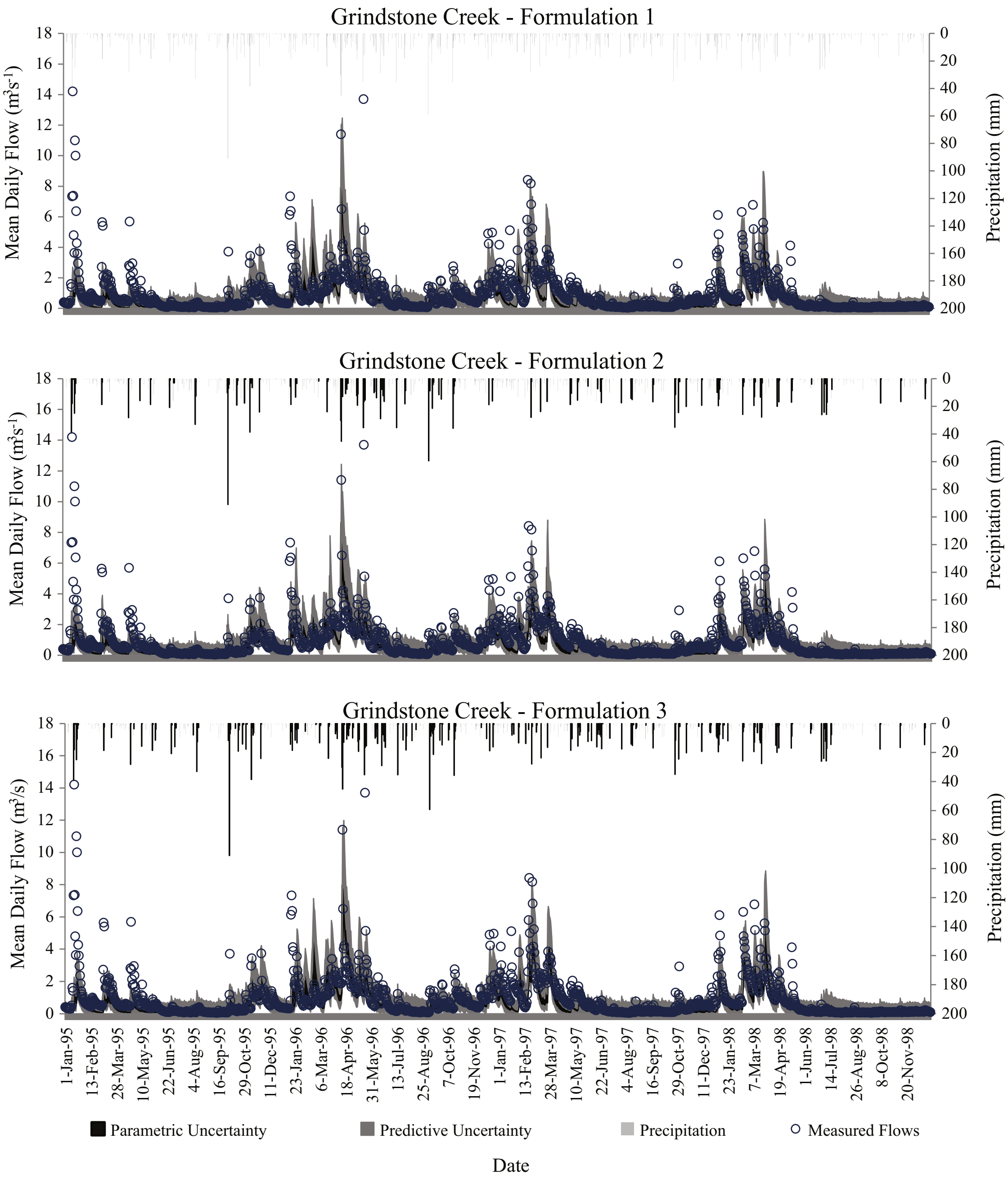


Fig. 4. Flow validation, Grindstone Creek study site. Black precipitation bars indicate days with at least a 5% chance of exceeding the threshold for extreme events. Note that the remainder of the validation period (1995–1996) is included in the Electronic Supplementary Material (Fig. S10).

for Redhill and Grindstone Creeks for the validation period, 1995–1998. The various fluxes in Table 4 represent surface runoff due to overland flow; shallow groundwater flow to streams; evapotranspiration to the atmosphere; and deep drainage out of the catchment to aquifers or lake beds. [Note that streamflow is generated by a combination of surface runoff and groundwater flow, but in this paper we refer to overland flow as surface runoff or simply runoff.] Both Creeks received about 850 mm of precipitation with substantial inter-annual variability. Redhill Creek overall shows a much greater rate of surface runoff generation (21% to 25% of precipitation) than Grindstone Creek (3% to 5% of precipitation), largely due to the predominantly urban land cover of the former site. This higher rate of surface runoff came at the cost of both evapotranspiration and groundwater discharge to streams. Evapotranspiration ranged from 48% to 52% of precipitation in Redhill Creek and from 61% to 62% in Grindstone Creek. Ground water discharge to streams was estimated at between 21% and 23% of precipitation for Redhill Creek, and between 27% and 28% for Grindstone Creek. The estimates of the various pathways through which water exits the catchment are also fairly similar across the three formulations.

Surface runoff is the primary pathway through which many pollutants (including phosphorus) enter waterways, and so identifying sources of surface runoff can aid in locating possible pollutant source areas (McDowell and Srinivasan, 2009). We locate surface runoff source areas by land use and soil type. In Fig. 5, we present estimates of surface runoff generation for the different land uses in Redhill and Grindstone Creeks across the three formulations. We distinguish between runoff generated during the entire year and runoff generated during the growing season of Hamilton Harbour, the receiving waterbody (May–September), as this is the period when the receiving waterbody is most sensitive to eutrophication. In both Creeks, urban land use generated

Table 4
Annual water balance from the three statistical formulations for Redhill and Grindstone Creeks, during the validation period.

Validation Period (1995–1998)	Formulation 1 Depth (mm)	Formulation 2 Depth (mm)	Formulation 3 Depth (mm)
Redhill Creek			
Precipitation	853.0 ±188.0	853.0 ±188.0	853.0 ±188.0
Surface Runoff	176.1 ±6.0	203.9 ±5.0	216.9 ±6.2
Ground Water	193.4 ±2.3	196.0 ±4.6	182.2 ±5.5
Evapotranspiration	446.0 ±1.3	409.4 ±1.2	410.7 ±1.2
Deep Drainage	9.7 ±0.1	9.8 ±0.2	9.1 ±0.3
Grindstone Creek			
Precipitation	853.0 ±188.0	853.0 ±188.0	853.0 ±188.0
Surface Runoff	35.4 ±1.9	44.7 ±7.0	25.0 ±3.3
Ground Water	226.8 ±5.8	224.1 ±7.3	237.3 ±8.7
Evapotranspiration	532.0 ±4.3	527.1 ±6.9	533.6 ±6.7
Deep Drainage	11.3 ±0.3	11.2 ±0.4	11.9 ±0.4

the greatest depth of runoff; 245–262 mm for Redhill Creek and 202–240 mm for Grindstone Creek. For Redhill Creek, this compares to 51–183 mm for crops, 26–76 mm for forest, 34–149 mm for pasture, and 34–106 mm for urban greenspace. For Grindstone Creek, our estimate for the urban runoff compares to 11–45 mm for crops, 3–16 mm for forest, and 3–21 mm for pasture. During the growing season, this disparity became more acute, particularly in Grindstone Creek. Between May and September, runoff generation in Redhill Creek ranged from 8–51 mm for crops, 4–16 mm for forest, 6–37 mm for pasture, and 6–29 mm for urban greenspace. For Grindstone Creek, this compares to 1 mm for crops, <1 mm for forest, and <1 mm for pasture. Urban areas effectively bypass catchment storage, as nearly all the precipitation falling on them becomes surface runoff and reaches the stream in less than one day, leaving little time for evapotranspiration. While the importance of urban areas as a surface runoff source increased slightly during the growing season in Redhill Creek, it is somewhat surprising that the model predicts that almost no surface runoff reaches the stream from any of the pervious surfaces in Grindstone Creek from May to September. While it is likely that the contribution of runoff for Grindstone Creek is somewhat underestimated, the posterior multiplicative effects for the curve number parameters were not close to the maximum of their allowable range, indicating that model solutions with higher rates of runoff generation did not result in a better fit at the basin outlet. This suggests that there may be important differences in soil type and/or vegetation cover between the two catchments which may be responsible for generating the markedly different amounts of runoff during the growing season. There were also noteworthy differences between statistical formulations. Formulations 2 and 3 both accommodated differences in response of the catchments to extreme events, and both generated more runoff than Formulation 1.

Accounting for the different proportions of the various land uses in each watershed, we can estimate how much water volume originated from various land uses. This piece of information allows us to assess the overall importance of each type of land use with respect to the generation of runoff reaching Hamilton Harbour (Fig. 6). The results for Redhill Creek are not surprising and underscore the large proportion of urban area in that watershed, i.e., 1040–1115 m^3 of 1115–1333 m^3 total runoff originated from urban residential areas. However, the results from Grindstone Creek indicate that i) the volume of runoff generated therein is a small fraction of that generated in Redhill Creek (21%–29%), and ii) despite their small areal coverage (~10%), urban areas in Grindstone generate a disproportionate amount of total runoff (157–187 m^3 of 233–388 m^3). When examining the growing season virtually all (>98%) of the surface runoff reaching Grindstone Creek is estimated to originate from urban areas. During the entire annual cycle, urban areas in Grindstone Creek generated between 37 and 74% of surface runoff as estimated with Formulation 2. We also quantified runoff depths generated by soil type (Fig. 7). Soil types refer to the SLC Polygon Identifier of the Soil Landscapes of Canada dataset version 3.2. [Note that we exclude any urban areas from the summaries presented in Fig. 7.] It is interesting to note that the most common soils in Redhill Creek appear to be more prone to runoff generation than the most

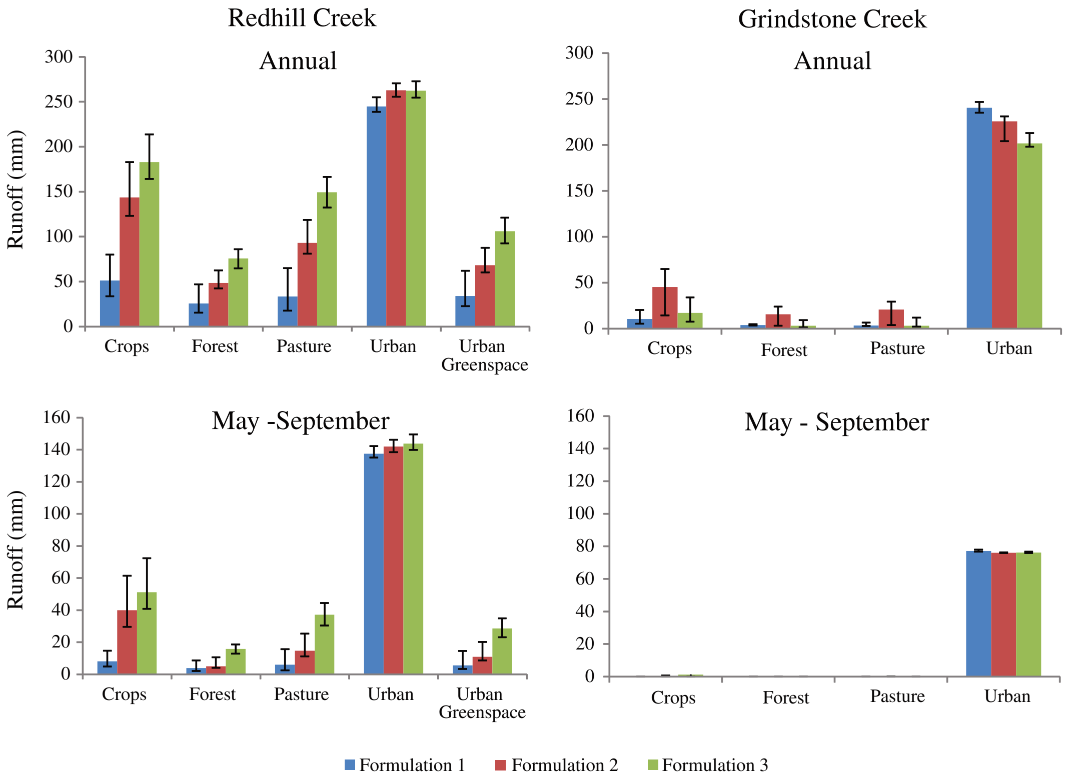


Fig. 5. Surface runoff depths generated in Redhill and Grindstone Creeks during the validation period (1995–1998) by different land uses.

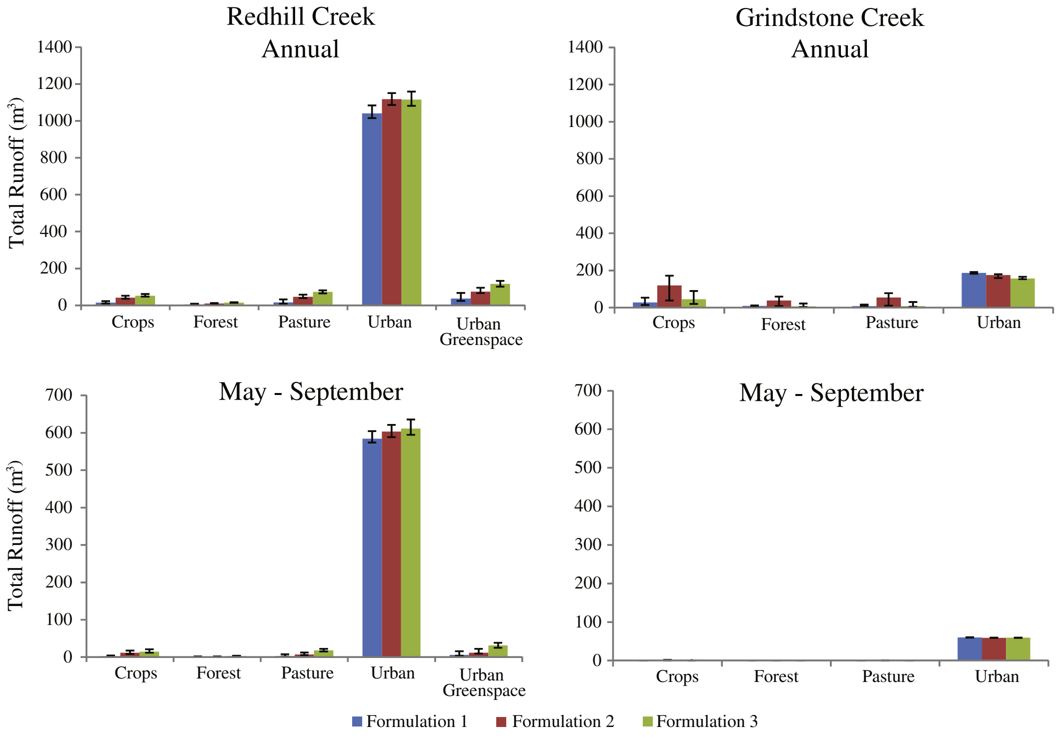


Fig. 6. Surface runoff volumes generated in Redhill and Grindstone Creeks during the validation period (1995–1998) by different land uses.

common ones in Grindstone Creek. Depending on which of these soil types undergoes urban development, it is reasonable to surmise that a greater increase in runoff is likely to occur in Grindstone Creek than in Redhill Creek, per unit addition of urban area.

Discussion

We have presented a framework to accommodate the effect of extreme precipitation events on watershed response in hydrological models by identifying distinct states of a system and allowing some parameters to vary between states. Relative to strategies that opt for time-varying parameters (e.g., Reichert and Mieleitner, 2009), our intent was to offer a parsimonious means to augment model capacity to reproduce system response in ways which mixture-likelihood type approaches cannot (Schaefli et al., 2007). Our framework allowed us to better reproduce the flows of the days identified as extreme, though the overall model fit was not consistently improved when compared to a model calibration consisting of only one state. The remainder of the discussion is structured to assess our framework relative to other modeling strategies, to examine the role of parametric uncertainty in SWAT model, and to delve into what can actually be learned about the system studied.

An assessment of our framework for accommodating extreme events

There is evidence that extreme events will become more common and more dominant in watershed systems as a result of climate change and urbanization (Duan et al., 2012; Kunkel et al., 2013; Shields et al., 2008). Accommodating the role of extreme events will become critical to assessing future land use and climate change scenarios (Rode et al., 2010). A logical first step is to analyze catchment-scale data to arrive at an empirical understanding of how the watershed of interest responds to extreme events and then to incorporate this behavior in the model. The approach presented here has two specific advantages over calibration schemes, which ignore any state-specific parameter variability.

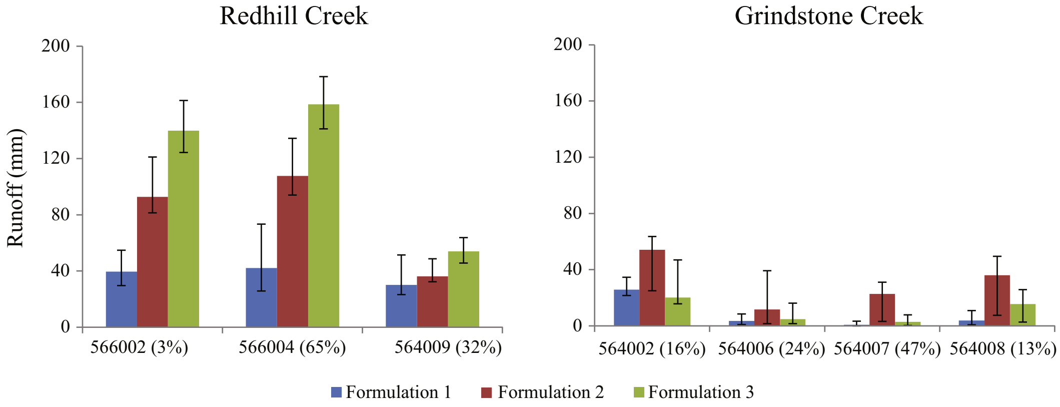


Fig. 7. Surface runoff depths generated in Redhill and Grindstone Creeks during the validation period (1995–1998) by different soil types. Soil types refer to the SLC Polygon Identifier of the Soil Landscapes of Canada dataset version 3.2. Percentages indicate the percent coverage of the non-urban part of the watershed.

First, using state-specific parameters allowed us to reduce the RMSE of days classified as extreme by roughly 17%, although there were not improvements in overall fit noted across all case studies. Applications of hydrological models primarily concerned with peak flows, either due to their importance as hazards, their importance to habitat quality, or their importance in the downstream transport of a number of waterborne constituents may find our approach useful. Second, our approach also classifies days into normal or extreme on the basis of their response to precipitation. This could also be of interest as a metric of the "flashiness" of the system.

Note that for both piecewise regressions presented in Fig. 2, there is no overlap of the 95% credible intervals of the two slopes, indicating that there was a significantly different response of the watershed to precipitation above the change point. However, there was also no overlap of the 95% credible intervals of the change points from the piecewise regressions with the posteriors of the thresholds θ_p. It is possible that the calibration period (1992–1994) had a different effective change point than the period from 1988 to 2009. We calibrated piecewise regressions for Redhill and Grindstone Creeks for the period from 1992 to 1994 and obtained change points with 95% credible intervals of 0.99–1.11 for Redhill Creek and 1.04–1.21 for Grindstone Creek. We found that the discrepancy between the change point and θ_p was reduced for Redhill Creek and was practically eliminated for Grindstone Creek. This finding may be evidence of substantial year-to-year variability, or simply highlights the lack of reliability in determining thresholds of watershed behavior with only three years of data.

While it would have been simpler to postulate prior independence between the state-specific model parameters, our statistical framework explicitly considered the parameter covariance between the two watershed states. We evaluated the benefits gained through our bivariate normal approach by calibrating a version of the Grindstone Creek model which was identical to Formulation 2 except that the curve number parameters below and above the threshold had independent priors (Table S3). While the inference drawn when postulating prior independence was fairly similar for most parameters, the assumption of prior independence significantly increased the posterior uncertainty of the threshold θ_p (0.17 relative to 0.09 in Table 3) and *CN2 Low (Mult. Eff.)* (0.20 relative to 0.12 in Table 3). The mean value of the threshold θ_p was also reduced to 0.87 (1.05 in Table 3), which in turn increased the number of days identified as having at least a 5% chance of being extreme during the validation phase from 123 to 547. Of these potentially extreme days, only 1 had a 100% chance of being extreme, with the rest having an average of 35% chance of being extreme. Clearly, our approach resulted in a reduced posterior parametric uncertainty (diagonal elements of Σ), thereby providing a much better separation of the estimated states of the watershed, but we also note that the covariance estimates (off-diagonal elements of Σ) were poorly identified; a result frequently reported in the literature (Bates et al., 2003; Wellen et al., 2014).

Zehe and Sivapalan (2009) highlighted the need to understand the reason for thresholds of catchment response. The existence of thresholds at the process scale is a necessary but not sufficient condition for the existence of thresholds at the catchment scale. Process thresholds include the classical mechanisms of overland flow generation — infiltration excess (Horton, 1933) and saturation excess (Dunne and Black, 1970). However, in order for excess water to contribute to streamflow, some degree of hydrological connection to the stream must be established. This often involves the (over)filling of various catchment storages (Ali et al., 2013; Lehmann et al., 2007; Oswald et al., 2011). Fig. 2 suggests that a storage threshold controls the generation of discharge in Redhill Creek, as the 2-day average precipitation controls the catchment response. The threshold response presented in Fig. 2 for Redhill Creek was also observed at an upstream monitoring station on Redhill Creek (Fig. S11), which drains about 40% of the catchment (Albion Falls station, Water Survey of Canada station 02HA023, drainage area 23.5 km^2) and at a monitoring station on Indian Creek, a nearby urban catchment monitored by the Ontario Ministry of Environment (drainage area 23 km^2, 72% urban residential and urban greenspace). The Albion Falls drainage is dominated by urban area (~80% including greenspace), but its later construction throughout the 1980s means that the sanitary and combined sewers are separated. Likewise, Indian Creek's urban area was constructed during roughly the same time frame, and it is also served by separated sewer systems. Thus, it is reasonable to hypothesize that the common denominator of the threshold behavior is a set of processes associated with urbanization. It is possible that this threshold corresponds to the storage capacity of the soils in the pervious areas within the urban matrix. When this capacity is exceeded, additional precipitation runs off directly into streets and other areas drained by sewers. In the absence of these interruptions of the natural drainage network, this runoff would be forced to flow overland to a channel and risk infiltration or evaporation en route. Despite significant attention given to anthropogenic effects on catchment thresholds in agricultural areas, Zehe and Sivapalan (2009) reported very little work in urban systems. More research is needed to understand the cause of the threshold behavior identified herein, as the hypothesis we advanced describes a clear pathway by which contaminants applied to urban pervious areas (e.g., phosphorus fertilizer, pesticides) may be entering waterways. Whatever the cause of this threshold behavior is, our framework for accommodating extreme events in model calibration was able to capture it better than the mathematics of the SWAT model alone.

This framework should not be considered solely as a manner of accommodating the functioning of urban areas in the SWAT model, as non-urban watersheds may also be characterized by threshold behavior in their response to extreme events (Oswald et al., 2011; Zehe and Sivapalan, 2009). For example, there is some increase of the variability of streamflow above a threshold of 3-day average precipitation for Grindstone Creek, and Formulations 2 and 3 both led to improvements in the prediction of extreme events for the same creek.

While our work has focused on the variability of runoff generation between response states, it is possible to allow any parameters to vary between states. In a parameter sensitivity and identifiability analysis, Cibin et al. (2010) found that the curve numbers dominated the model predictions in the two watersheds considered. Cibin et al. (2010) divided the period of record into days with low, medium, and high flow and found that the optimum region of the curve number parameters varied with flow regime. Low and high flow periods were better modeled with low and high curve numbers, respectively. Their result strongly bolsters our decision to include the curve number parameters in the threshold configuration.

On the value of informative priors

Informative priors are an advantage of the Bayesian approach to model calibration, as they allow the results of previous investigations to be included in the model calibration and inference process (Gelman et al., 2004). This can ensure that model calibration is not simply a data fitting exercise, but an update of prior knowledge. Informative priors decrease parametric uncertainty and can avoid misleading model calibrations (Arhonditsis et al., 2007, 2008a,b, 2011; Wellen et al., 2014). Despite these advantages, very seldom are informative priors used in watershed modeling. We found that the calibration data generally played a greater role in deciding the posteriors than the priors did. This conclusion is based on the consistently large values of the delta index (Tables 3 and 4) as well as the relative similarity of the inference drawn between Formulations 2 and 3 in both Creeks. There was enough information in only three years of continuous streamflow records to overcome any differences due to informative priors. However, future research may conclude that informative priors are valuable when streamflow records are of lower quality, e.g. consisting of a handful of spot flow records.

What can we learn about the watersheds under study?

To substantiate our estimates of water balance partitioning, we compared our estimates of surface runoff to those estimated by Parkin et al.'s (1999) water balance model. Parkin et al. (1999) estimated the water balance for corn crop grown in Guelph, Ontario, roughly 40 km from the study sites presented in this paper. They estimated 81 ± 64 mm of annual runoff, 25% of which occurred in the spring (April and May). Our estimates of annual agricultural runoff are consistent with this figure (33–183 mm in Redhill Creek and 5–64 mm in Grindstone Creek), although our estimates of growing season agricultural runoff for Grindstone Creek (<4 mm) are lower than the Parkin et al.'s (1999) findings (20.25 ± 16 for April and May). It is possible that we underestimate surface runoff during the growing season for Grindstone Creek. To examine the robustness of the latter results, we considered an alternative approach that postulates the ratio of growing season runoff depth to annual runoff depth estimated in Redhill Creek applies in Grindstone Creek as well. Adjusting the growing season runoff estimates, we estimate agricultural areas in Grindstone Creek generate between 4 and 18 mm of runoff, with a median estimate of 12 mm. These estimates are within the range obtained by Parkin et al. (1999) and are still considerably less than those estimated for Redhill Creek.

The most obvious conclusion of the differences in watershed functioning between the urban and agricultural creek pertains to the disproportionate role of the urban impervious surfaces in generating surface runoff. For Redhill Creek, between 81% and 93% of all surface runoff volume was estimated to be generated in urban residential areas over the entire annual cycle, whereas this proportion varied between 90% and 98% during the growing season (May–September). For Grindstone Creek, an approximate fraction between 45% and 80% of all surface runoff was estimated to be generated in urban residential areas over the entire annual cycle, and 95% and 99% during the growing season (May–September). The latter result is surprising given the low coverage of urban area in Grindstone Creek (~9%), and arises largely as a consequence of a near cessation of surface runoff generation in non-urban areas in Grindstone Creek during the growing season. If we use our alternative estimates of growing season runoff, urban residential areas still generate between 47% and 82% of the total runoff volume.

The impervious surfaces were not the sole factor responsible for the differences in functioning between the two Creeks. When we compare the pervious land covers of both Creeks, we see dramatic seasonal differences. While Redhill Creek generates about one half of its surface runoff volume during the growing season, Grindstone Creek generates a significantly lower fraction (15–29%). Despite the larger size, Grindstone Creek generates about 10% of the surface runoff volume Redhill Creek generates during the growing season. This difference cannot entirely be explained with reference to the greater urban area of Redhill Creek, as similar land uses appear to behave differently between the two Creeks. Agricultural land uses, for instance, generated between 33 and 183 mm of runoff annually in Redhill Creek and between 5 and 64 mm in Grindstone Creek. During the growing season, agricultural areas in Redhill Creek generated between 4 and 72 mm of runoff, whereas during this same period agricultural land uses in Grindstone Creek were estimated to generate between 0 and 3 mm. Even using our upwardly revised estimates of Grindstone Creek's growing season runoff (4–18 mm), this is significantly lower than the agricultural area in Redhill Creek during the growing season. The relative disconnect between the pervious areas of Grindstone Creek and the receiving waterbody during the growing season can be partially explained with reference to soil properties. The most common soil types in Redhill Creek tend to be more runoff prone than those in Grindstone Creek. The soils in Redhill Creek had base (prior) values of hydraulic conductivity nearly an order of magnitude lower than those in Grindstone Creek (1.7–3.4 mm/h for Redhill Creek; 16–27 mm/h for Grindstone); the lower posterior estimates of *SOL_KSAT (Mult. Eff.)* for Redhill and Grindstone Creeks exacerbated this difference. Clearly, the soils in Redhill Creek are not able to drain to field capacity after a significant wetting event as fast as those in Grindstone Creek. This built-up of moisture would lead to higher daily curve numbers being estimated by SWAT. Further, the Natural Resources Conservation Service Hydrologic Runoff Group values for the soils in Grindstone Creek were typically B, while those in Redhill Creek were typically C. This led to the selection of higher base value curve numbers in Redhill Creek, while the posterior values of *CN2 (Mult. Eff.)* were typically higher in Redhill Creek. These factors have the combined effect of making similar land uses more runoff prone in Redhill Creek than in Grindstone Creek. The effect of future development on runoff production thus depends partially on which soil type is developed and in which creek.

In conclusion, we presented a framework to accommodate the effect of extreme events on watershed response to precipitation by identifying distinct states of a system and allowing some parameters to vary between states. As climate and land use changes together are likely to make extreme events more common, it will become necessary to evaluate their impact on watershed functioning and downstream water quality (Zehe and Sivapalan, 2009). Our framework improved the reproduction of flow conditions in days identified as extreme, though the overall model fit was not improved. Accommodating the extreme event behavior resulted in significantly different runoff source apportionments, as the runoff generated by pervious areas tended to be greater. The Bayesian nature of our approach allows a probabilistic estimation of critical runoff generating areas that may be responsible for greater pollutant export to receiving waterbodies (McDowell and Srinivasan, 2009) while accounting for the different sources of uncertainty (model structure imperfection, model input uncertainty) as well as natural system variability. Our future work will couple the framework presented here with a water quality model to evaluate its efficiency in predicting the impact of non-point source pollution to the eutrophication patterns of the receiving water body.

Acknowledgements

This project has received funding support from the Ontario Ministry of the Environment (Canada Ontario Grant Agreement, Contract 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jglr.2014.04.002.

References

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.

Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: application to a mesotrophic lake. Ecol. Model. 208, 215–229.

Arhonditsis, G.B., Papantou, D., Zhang, W., Perhar, G., Massos, E., Shi, M., 2008a. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. J. Mar. Syst. 73, 8–30. http://dx.doi.org/10.1016/j.jmarsys.2007.07.004.

Arhonditsis, G.B., Perhar, G., Zhang, W., Massos, E., Shi, M., Das, A., 2008b. Addressing equifinality and uncertainty in eutrophication models. Water Resour. Res. 44, W01420. http://dx.doi.org/10.1029/2007WR005862.

Arhonditsis, G., Stremilov, S., Gudimov, A., Ramin, M., Zhang, W., 2011. Integration of Bayesian inference techniques with mathematical modelling. In: Wolanski, E., McLusky, D.S. (Eds.), Treatise on Estuarine and Coastal Science, 9. Academic Press, Waltham, pp. 173–192.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modelling and assessment. Part I: Model development. J. Am. Water Resour. Assoc. 34 (1), 73–89.

Bates, S.C., Cullen, A., Raftery, A.E., 2003. Bayesian uncertainty assessment in multi-compartment deterministic simulation models for environmental risk assessment. Environmetrics 14, 355–371.

Cerucci, M., Conrad, J.M., 2003. The use of binary optimization and hydrologic models to form riparian buffers. J. Am. Water Resour. Assoc. 39 (5), 1167–1180.

Chaplot, V., Saleh, A., Jaynes, D.B., 2005. Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO3–N loads at the watershed level. J. Hydrol. 312, 223–234.

Cibin, R., Sudheer, K.P., Chaubey, I., 2010. Sensitivity and identifiability of stream flow generation parameters of the SWAT model. Hydrol. Process. 24, 1133–1148. http://dx.doi.org/10.1002/hyp.7568.

Donald, J.R., 1992. Snowcover depletion curves and satellite snowcover estimates for snowmelt runoff modelling. (Ph.D. Thesis) University of Waterloo, ON, Canada (232 pp.).

Donald, J.R., Soulis, 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, S., Kaushal, S.S., Groffman, P.M., Band, L.E., Belt, K.T., 2012. Phosphorus export across an urban to rural gradient in the Chesapeake Bay watershed. J. Geophys. Res. 117, G01025. http://dx.doi.org/10.1029/2011JG001782.

Dunne, T., Black, R.D., 1970. Partial area contributions to storm runoff in a small New England watershed. Water Resour. Res. 6 (5), 1296–1311.

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 Trans. Inf. Theory 49, 1858–1860. http://dx.doi.org/10.1109/tit.2003.813506.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman &Hall/CRC Press, Boca Raton, FL.

Gumbel, E.J., 1954. Statistical theory of extreme values and some practical applications. Applied Mathematics Series, 33. U.S. Department of Commerce, National Bureau of Standards.

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.

Horton, R.E., 1933. The role of infiltration in the hydrologic cycle. Trans. AGU 14, 446–460.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 35–45.

Kunkel, K.E., Karl, T.R., Easterling, D.R., Redmond, K., Young, J., Yin, X., Hennon, P., 2013. Probable maximum precipitation and climate change. Geophys. Res. Lett. 40. http://dx.doi.org/10.1002/grl.50334.

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.

Law, N., Band, L., Grove, M., 2004. Nitrogen input from residential lawn care practices in suburban watersheds in Baltimore county, MD. J. Environ. Plan. Manag. 47 (5), 737–755.

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.

Lin, Z., Beck, M.B., 2007. On the identification of model structure in hydrological and environmental systems. Water Resour. Res. 43, W02402. http://dx.doi.org/10.1029/2005WR004796.

Macrae, M.L., English, M.C., Schiff, S.L., Stone, M., 2007. Capturing temporal variability for estimates of annual hydrochemical export from a first-order agricultural catchment in southern Ontario, Canada. Hydrol. Process. 21, 1651–1663. http://dx.doi.org/10.1002/hyp.6361.

McDonnell, J.J., 1990. A rationale for old water discharge through macropores in a steep, humid catchment. Water Resour. Res. 26 (11), 2821–2832.

McDowell, R.W., Srinivasan, M.S., 2009. Identifying critical source areas for water quality: 2. Validating the approach for phosphorus and sediment losses in grazed headwater catchments. J. Hydrol. 379 (1–2), 68–80. http://dx.doi.org/10.1016/j.jhydrol.2009.09.045.

Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloğlu, 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. Proc. Natl. Acad. Sci. U. S. A. 110, 6448–6452.

Nash, J.E., Sutcliff, J.V., 1970. River flow forecasting through conceptual models. Part 1 — A discussion of principles. J. Hydrol. 10, 282–290.

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 (Retrieved from http://twri.tamu.edu/reports/2011/tr406.pdf January 17th, 2013).

Oswald, C.J., Richardson, M.C., Branfireun, 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.

Parkin, G.W., Wagner-Riddle, C., Fallow, D.J., Brown, D.M., 1999. Estimated seasonal and annual water surplus in Ontario. Can. Water Resour. J. 24 (4), 277–292.

Prado, R., West, M., 2010. Time Series: Modelling, Computation, and Inference. CRC Press, Boca Raton, FL (353 pp.).

Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resour. Res. 45 (10), W10402. http://dx.doi.org/10.1029/2009wr007814.

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.

Schaefli, B., Balin, D.T., Musy, A., 2007. Quantifying hydrological modelling errors through a mixture of normal distributions. J. Hydrol. 332 (3–4), 303–315. http://dx.doi.org/10.1016/j.jhydrol.2006.07.005.

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 edition. Jon Wiley and Sons, Toronto (507 pp.).

Shields, C.A., Band, L.E., Law, N., Groffman, P.M., Kaushal, S.S., Savvas, K., Fisher, G.T., Belt, K.T., 2008. Streamflow distribution of non-point source nitrogen export from urban–rural catchments in the Chesapeake Bay watershed. Water Resour. Res. 44 (9), 13.

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/WR016i002p00430.

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.

Vanhatalo, J., Jylänki, P., Vehtari, A., 2009. Gaussian process regression with Student-t likelihood. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (Eds.), Proceedings of the twenty-third annual conference on Neural Information Processing Systems, British Columbia, Canada, December 7–10, 2009. Advances in Neural Information Processing Systems, 22. Neural Information Processing Systems Foundation.

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 to accommodate inter-annual nutrient loading variability with the SPARROW model. Water Resour. Res. 48, W10505. http://dx.doi.org/10.1029/2012WR011821.

Wellen, C., Arhonditsis, G.B., Labencki, T., Boyd, D., 2014. 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.

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 modelling: a case study in the Thur River basin (Switzerland). Water Resour. Res. 43, W10401. http://dx.doi.org/10.1029/2006WR005497.

Zehe, E., Sivapalan, M., 2009. Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications. Hydrol. Earth Syst. Sci. 13 (7), 1273–1297.

Zehe, E., Maurer, T., Ihringer, J., Plate, E., 2001. Modeling water flow and mass transport in a loess catchment. Phys. Chem. Earth B 26 (7–8), 487–507.

Zhang, X., Srinivasan, R., Arnold, J., Izaurralde, R.C., Bosch, D., 2011. Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework. Hydrol. Process. 25, 2313–2320. http://dx.doi.org/10.1002/hyp.8058.

ACCOMMODATING ENVIRONMENTAL THRESHOLDS AND EXTREME EVENTS IN HYDROLOGICAL MODELS: A BAYESIAN APPROACH

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 (ESM)

Details of the MCMC sampling algorithm:

We simulated samples from the posterior using Markov chain Monte Carlo (MCMC) sampling. In this study, we used the DiffeRential Evolution Adaptive Metropolis Algorithm-ZS (DREAM-ZS) as presented by Laloy and Vrugt (2011), the details of which we include in the ESM. This algorithm is based on the original DREAM algorithm presented in Vrugt et al. (2008, 2009). DREAM adapts more traditional MCMC approaches to the complex, multi-modal likelihood surfaces characterizing deterministic watershed models by running multiple Markov chains and basing the proposal distribution on the distances between chains in the parameter space. DREAM-ZS further adapts this approach by sampling from an archive of past states to generate proposal locations in the parameter space. We used a total of 5 chains for each model realization and collected between 32,000 and 64,000 samples per chain. Convergence was assessed qualitatively by visually inspecting plots of the posterior Markov chains for mixing and stationarity and by inspecting density plots of the pooled posterior Markov chains for unimodality. We also assessed convergence quantitatively using the modified Gelman–Rubin convergence statistic (Brooks and Gelman, 1998). The first half of our posterior samples was discarded to ensure no significant effect of initial conditions and imposed a thin of 10 to minimize the effect of sample autocorrelation.

References

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative Simulations. J. Comput. Graph. Stat. 7, 434–455.

Table S-1: Fit statistics for the SWAT application in Redhill Creek.

Formulation		**NSE**	**RE**	**r^2**
Formulation 1	Calibration	0.64	0.54	0.66
	Validation	0.52	0.60	0.54
Formulation 2	Calibration	0.66	0.52	0.71
	Validation	0.56	0.58	0.57
Formulation 3	Calibration	0.60	0.72	0.63
	Validation	0.56	0.57	0.57

Table S-2: Fit statistics for the SWAT application in Grindstone Creek.

Formulation		**NSE**	**RE**	**r^2**
Formulation 1	Calibration	0.71	0.44	0.72
	Validation	0.56	0.47	0.56
Formulation 2	Calibration	0.71	0.45	0.71
	Validation	0.49	0.44	0.52
Formulation 3	Calibration	0.74	0.43	0.75
	Validation	0.44	0.47	0.48

Table S-3: Comparison of Formulation 2 posteriors with those obtained when postulating prior independence of CN2 Low (Mult. Eff.) and CN2 High (Mult. Eff.) in Grindstone Creek.

Parameter	**Formulation 2**		**Formulation 2, prior independence**	
	Mean	SD	Mean	SD
ALPHA_BF	0.94	0.05	0.94	0.04
SOL_AWC (Mult. Eff.)	1.41	0.01	1.40	0.01
GW_REVAP	0.05	0.03	0.04	0.02
ESCO	0.18	0.08	0.15	0.04
EPCO	0.49	0.26	0.47	0.23
GW_DELAY (Mult. Eff.)	1.75	0.27	1.55	0.14
SOL_KSAT (Mult. Eff.)	0.55	0.07	0.54	0.05
SNOWCOVMX	21.06	0.37	23.93	0.83
SMFMX	2.79	0.11	5.09	0.83
SMFMN	2.51	0.08	2.68	0.32
SURLAG	0.51	0.01	0.51	0.01
θ_p	1.05	0.09	0.83	0.17
CN2 Low (Mult. Eff.)	0.57	0.12	0.76	0.20
CN2 High (Mult. Eff.)	0.77	0.07	0.64	0.07
σ	0.07	0.00	0.066	0.002
ρ	0.91	0.01	0.918	0.011
Logged Likelihood	1053.23	4.85	1038.51	4.62

LIST OF FIGURES

drainage area 23 km^2). Redhill Creek scatterplots show daily flows between 1989 – 2003. Indian Creek scatterplot shows flows from the period August 2010 – June 2012. For all graphs, only data from the months May – November are plotted.

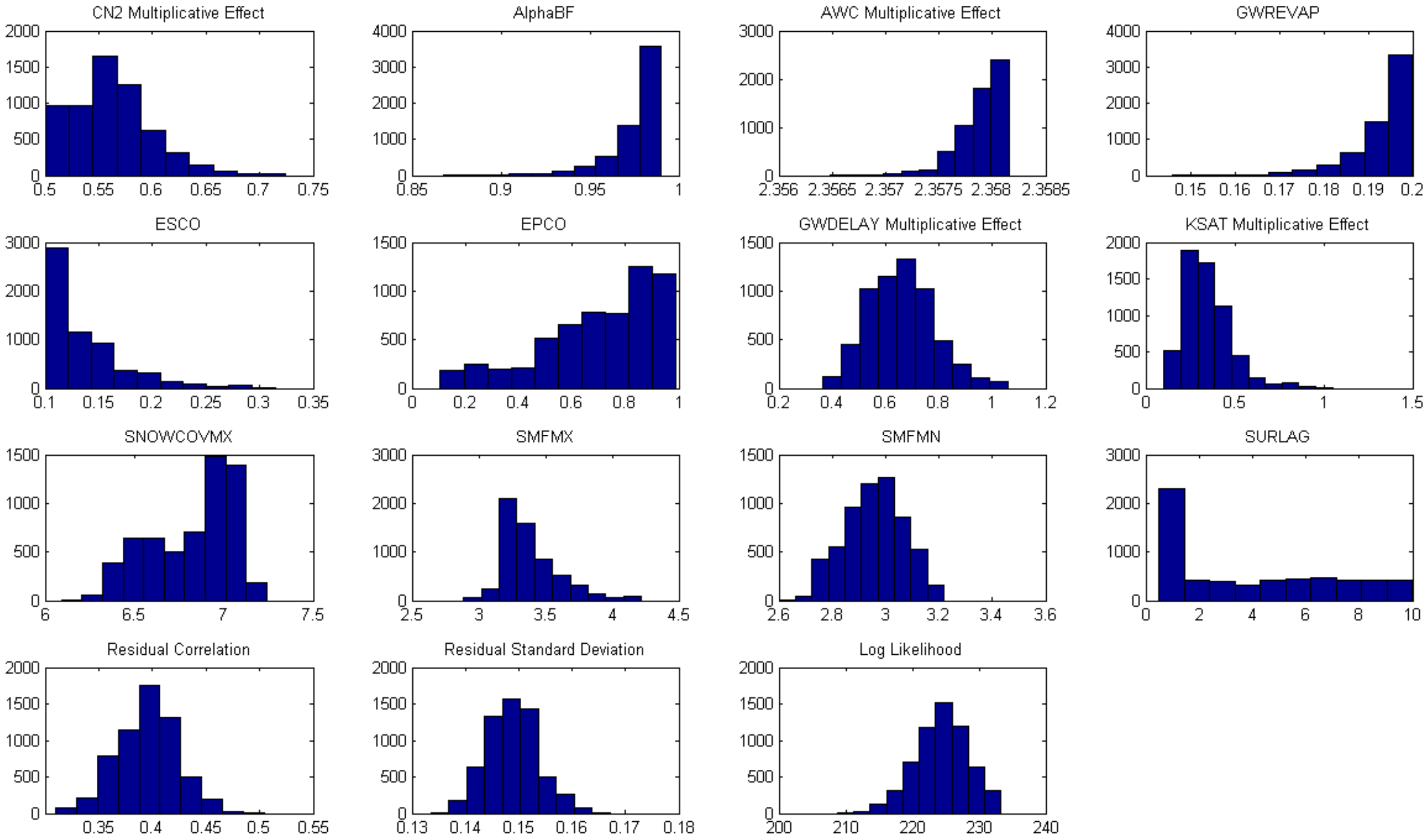


Figure S-1: Posterior marginals for Redhill Creek Formulation 1.

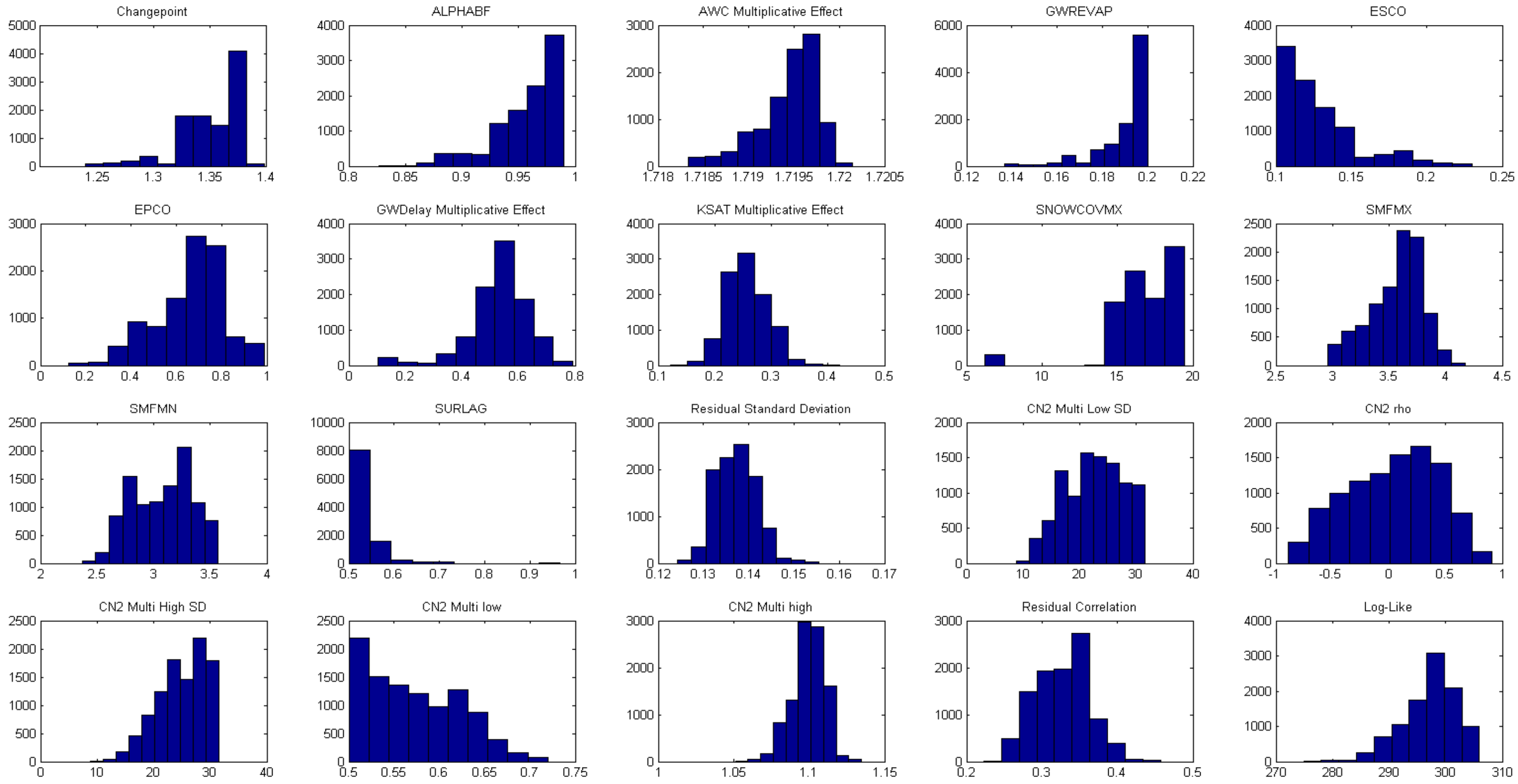


Figure S-2: Posterior marginals for Redhill Creek Formulation 2.

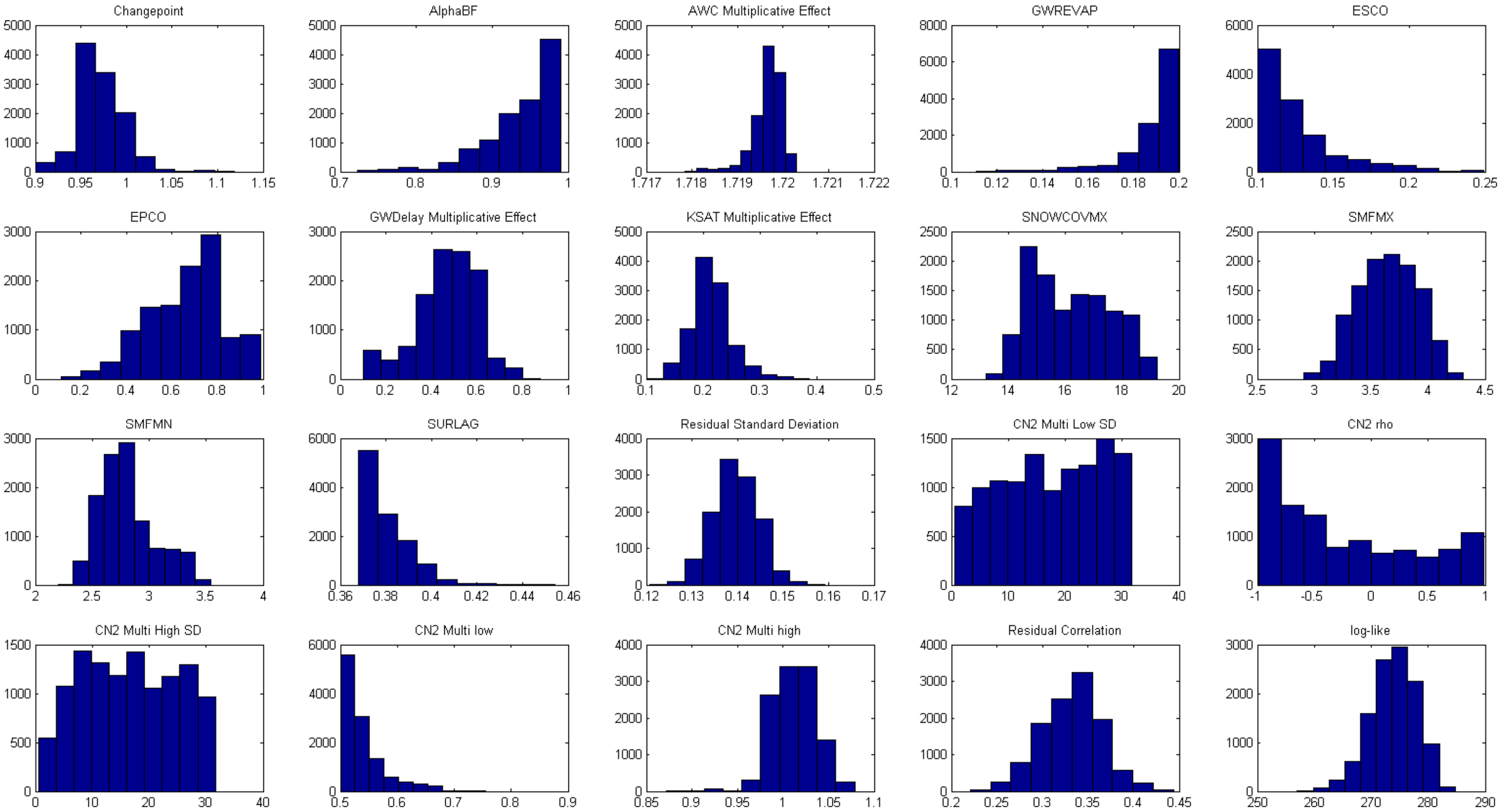


Figure S-3: Posterior marginals for Redhill Creek Formulation 3.

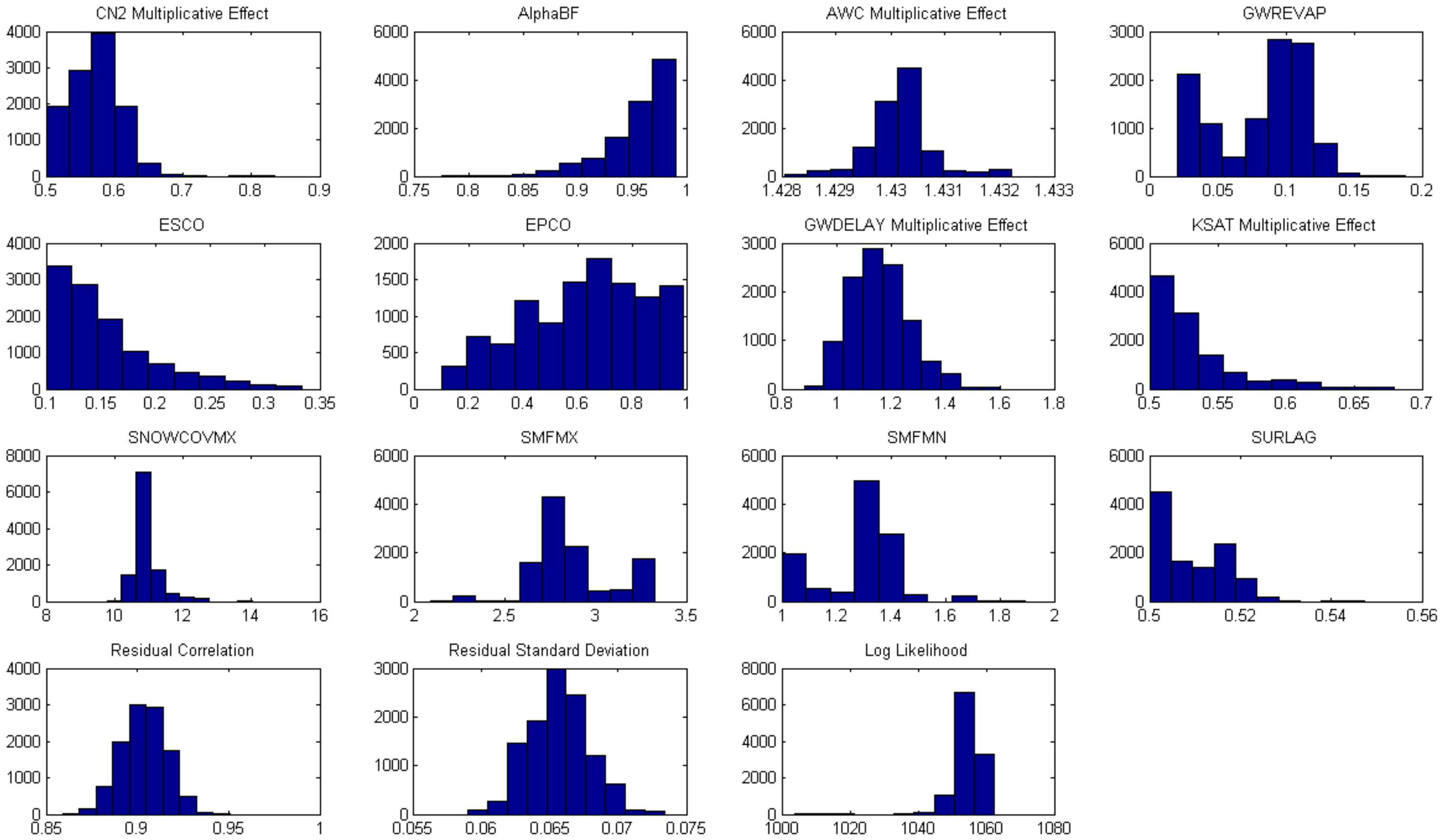


Figure S-4: Posterior marginals for Grindstone Creek Formulation 1.

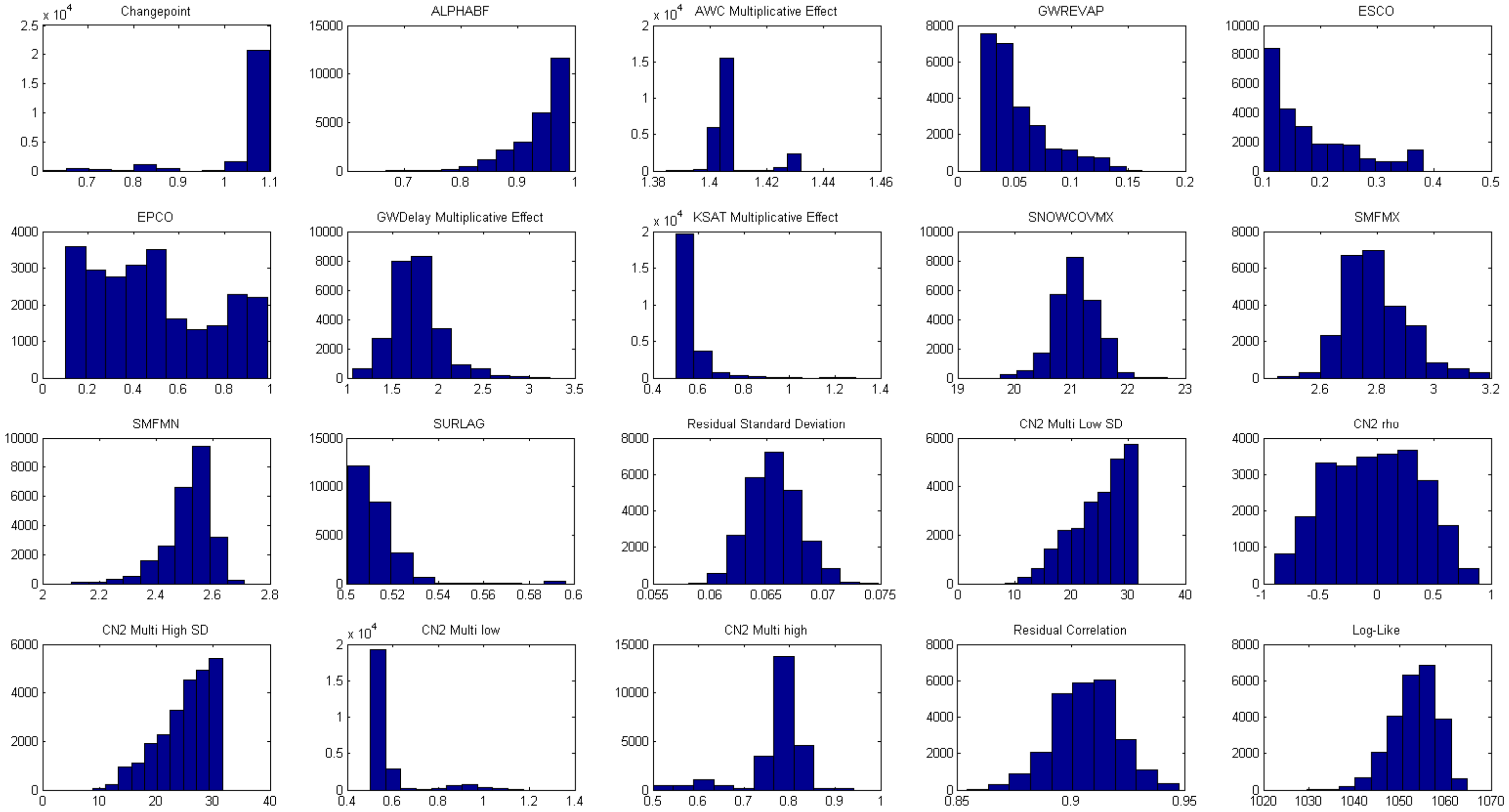


Figure S-5: Posterior marginals for Grindstone Creek Formulation 2.

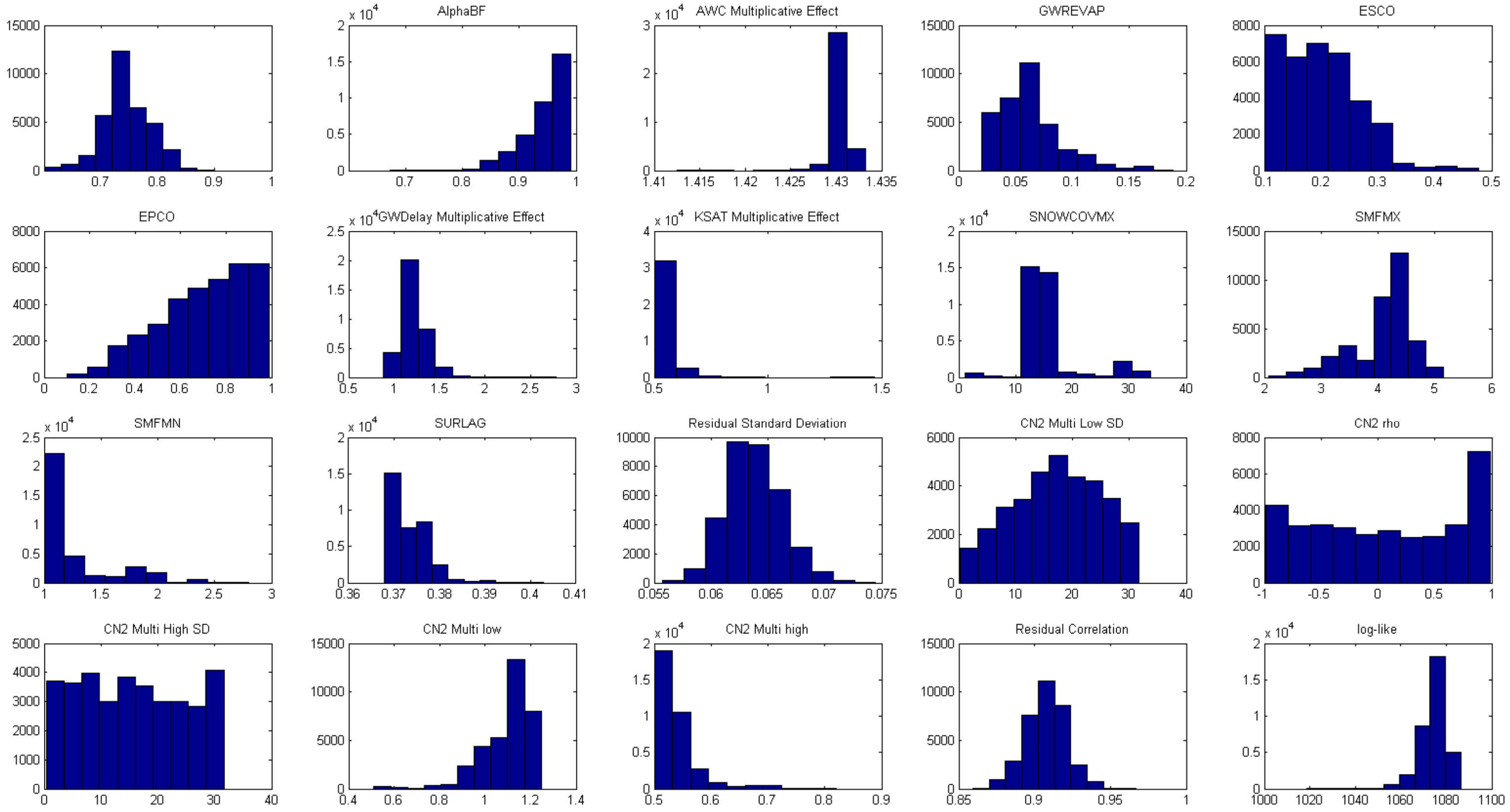


Figure S-6: Posterior marginals for Grindstone Creek Formulation 3.

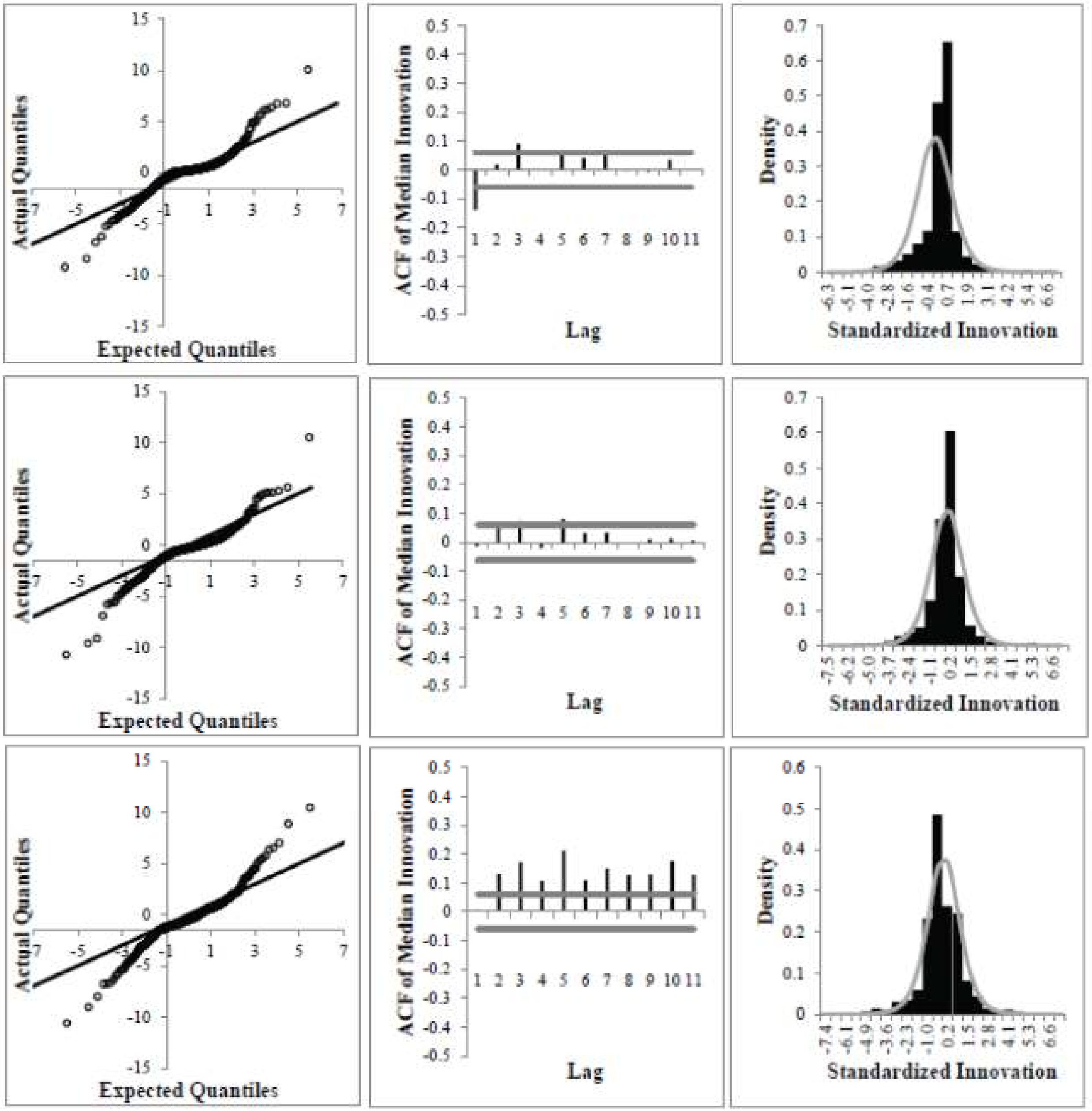


Figure S-7: Likelihood Assessment for Redhill Creek. The top row reports on Formulation 1, the middle row on Formulation 2, and the bottom on Formulation 3. The left column presents quantile-quantile plots for the expected and actual standardized innovations, the middle column presents autocorrelation 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.

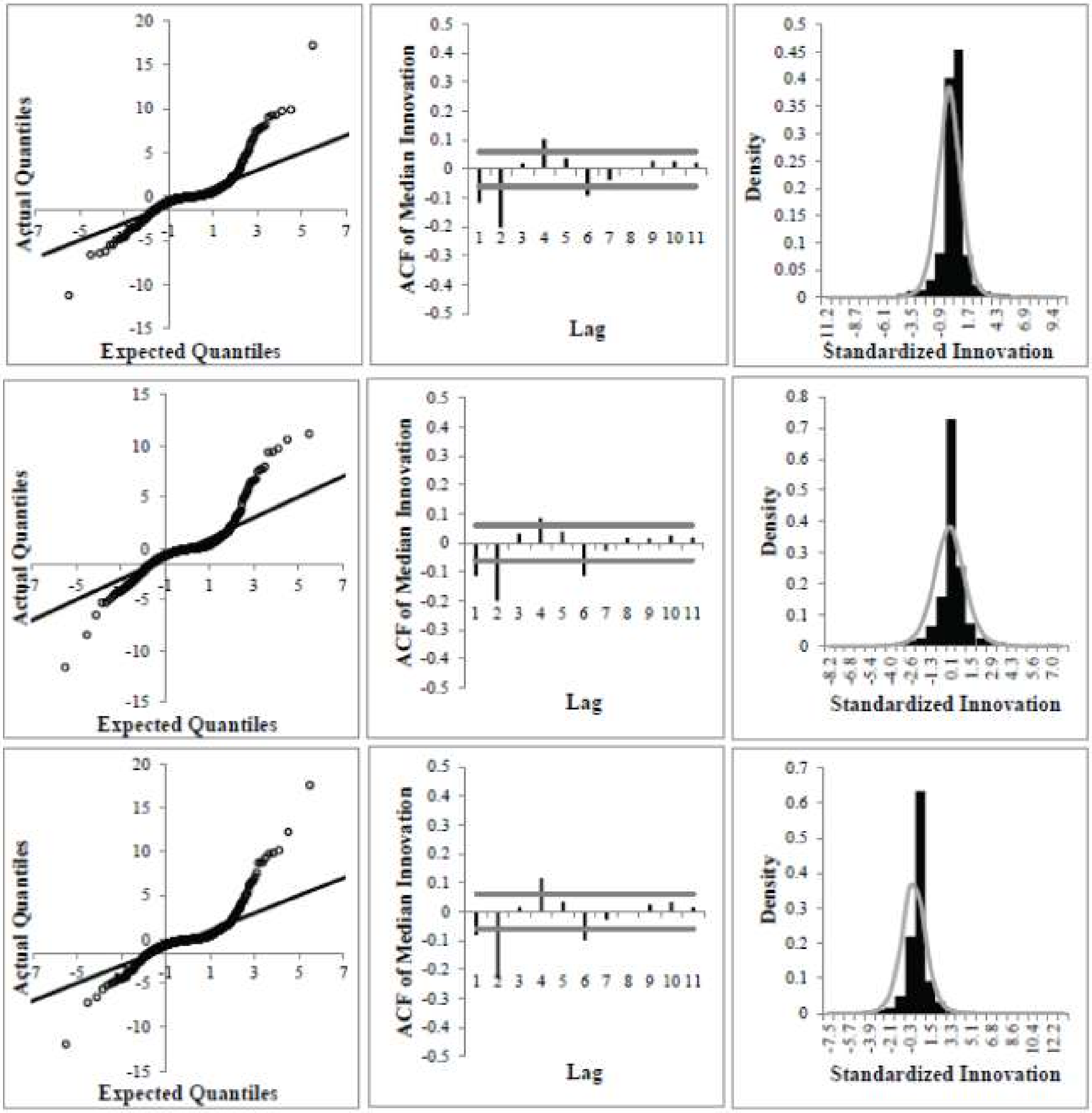


Figure S-8: Likelihood Assessment for Grindstone Creek. The top row reports on Formulation 1, the middle row on Formulation 2, and the bottom on Formulation 3. The left column presents quantile-quantile plots for the expected and actual standardized innovations, the middle column presents autocorrelation 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.

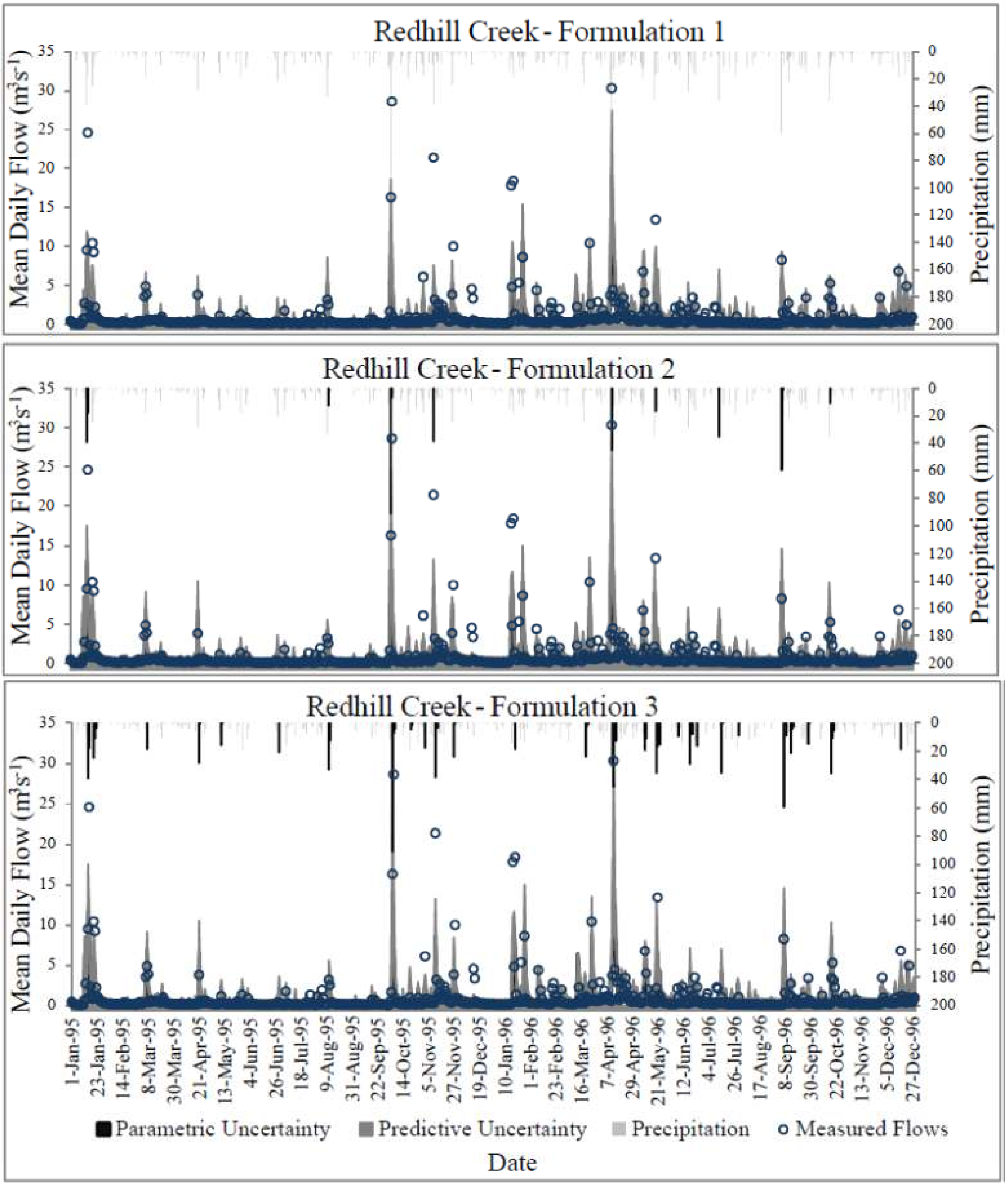


Figure S-9: Flow validation for Redhill Creek for (1995 – 1996). Black precipitation bars indicate days with at least a 5% chance of exceeding the threshold for extreme events.

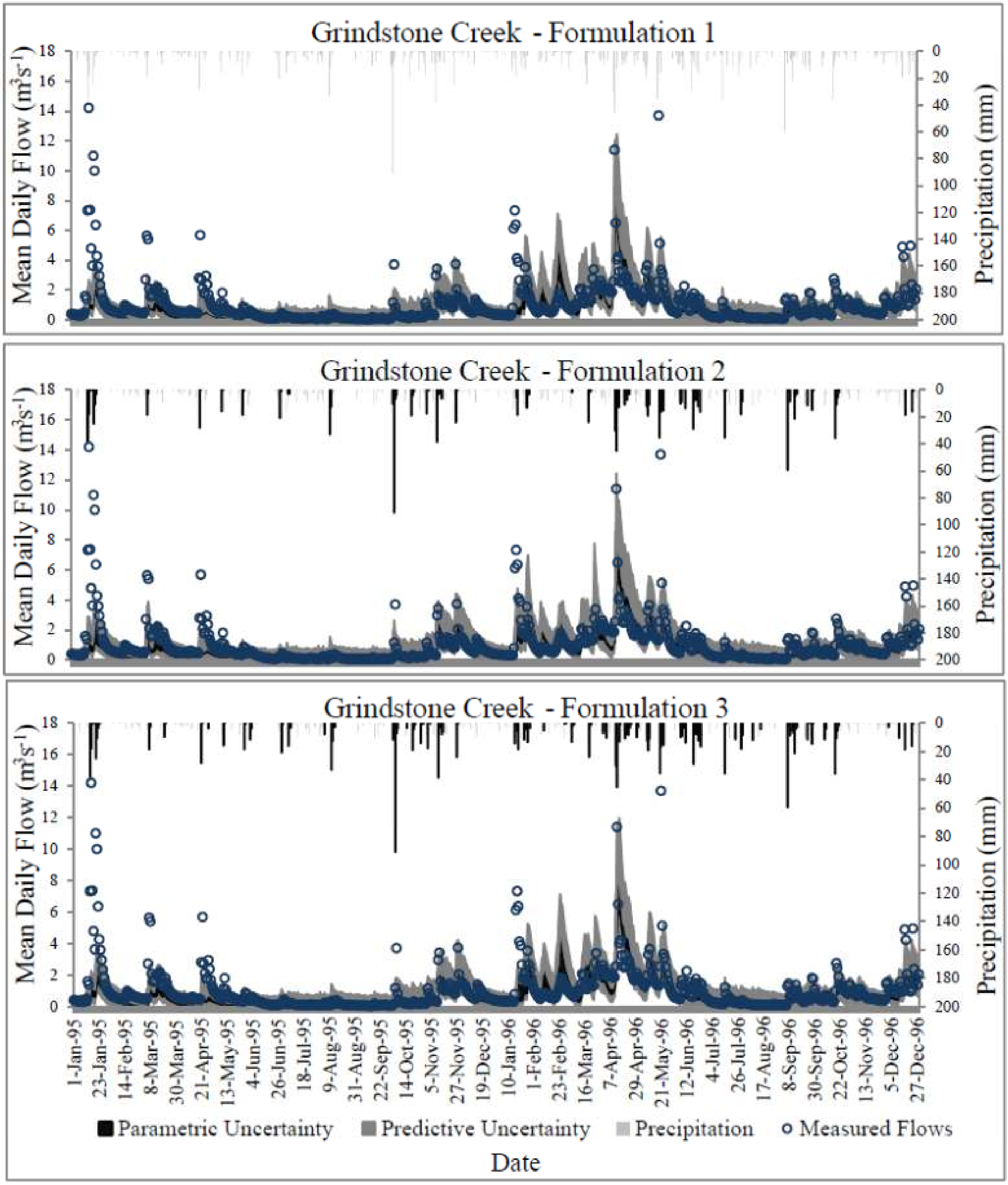


Figure S-10: Flow validation for Grindstone Creek for (1995 – 1996). Black precipitation bars indicate days with at least a 5% chance of exceeding the threshold for extreme events.

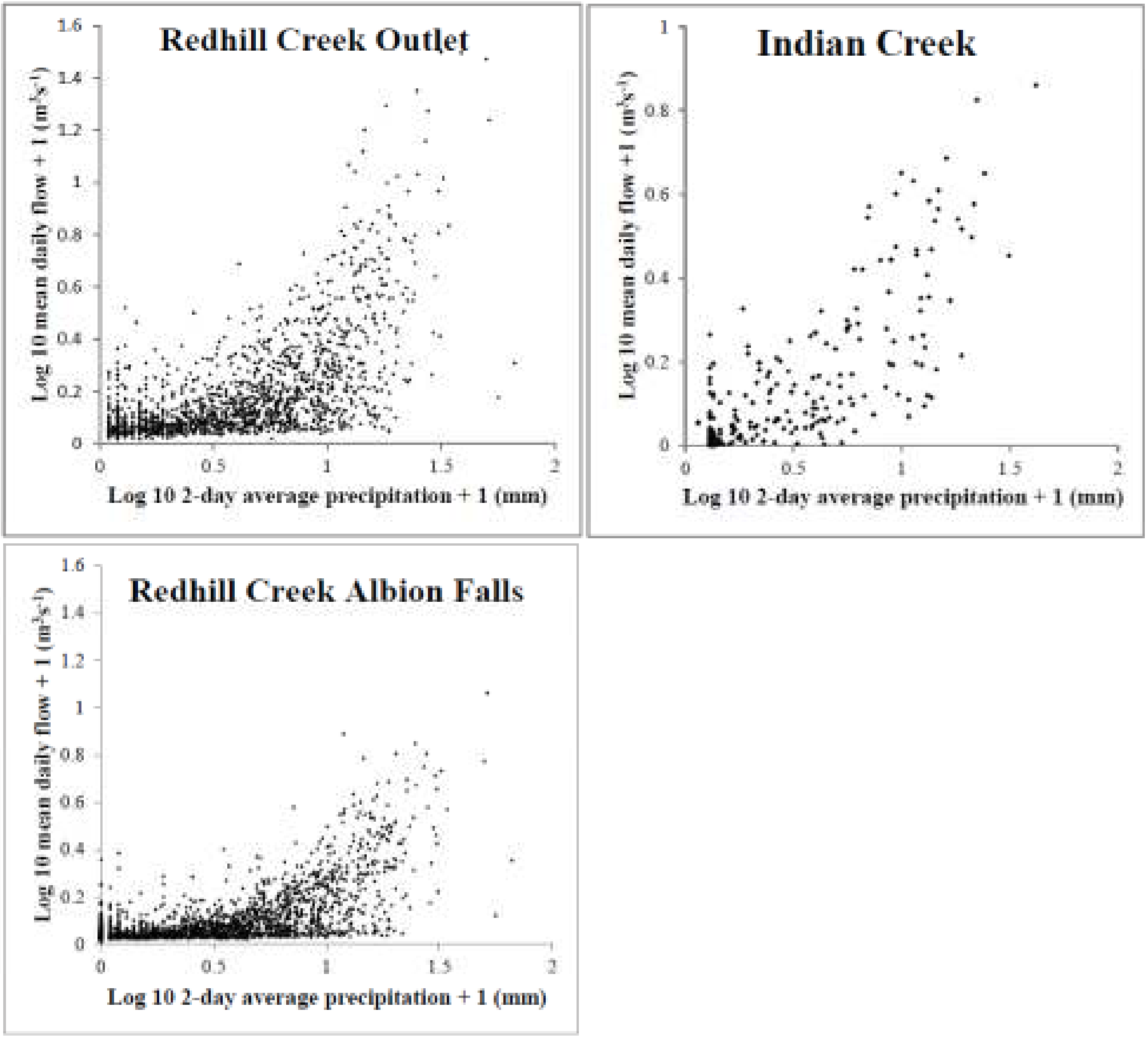


Figure S-11: Scatterplot relating 2 day averaged precipitation with daily streamflow at the Albion Falls station on Redhill Creek (Water Survey of Canada Station 02HA023, drainage area 23.5 km^2) and below the Hagar-Rambo diversion of Indian Creek (Ontario Ministry of Environment, drainage area 23 km^2). Redhill Creek scatterplots show daily flows between 1989 – 2003. Indian Creek scatterplot shows flows from the period August 2010 – June 2012. For all graphs, only data from the months May – November are plotted.