RESEARCH ARTICLE

WILEY

Assessing the performance of a physically based hydrological model using a proxy-catchment approach in an agricultural environment

Veronica Escobar-Ruiz[1] 🆔 | Hugh G. Smith[1,2] 🆔 | William H. Blake[3] 🆔 | Neil Macdonald[1] 🆔

[1]School of Environmental Science, University of Liverpool, Liverpool, UK

[2]Landcare Research, Palmerston North, New Zealand

[3]School of Geography, Earth and Environmental Sciences, University of Plymouth, Plymouth, UK

Correspondence
Veronica Escobar-Ruiz, School of Environmental Science, University of Liverpool, Liverpool L697ZT, UK.
Email: v.escobar-ruiz@liverpool.ac.uk

Abstract

Physically based models are useful frameworks for testing intervention strategies designed to reduce elevated sediment loads in agricultural catchments. Evaluating the success of these strategies depends on model accuracy, generally established by a calibration and evaluation process. In this contribution, the physically based SHETRAN model was assessed in two similar U.K. agricultural catchments. The model was calibrated on the Blackwater catchment (18 km^2) and evaluated in the adjacent Kit Brook catchment (22 km^2) using 4 years of 15 min discharge and suspended sediment flux data. Model sensitivity to changes in single and multiple combinations of parameters and sensitivity to changes in digital elevation model resolution were assessed. Model flow performance was reasonably accurate with a Nash–Sutcliffe efficiency coefficient of 0.78 in Blackwater and 0.60 in Kit Brook. In terms of event prediction, the mean of the absolute percentage of difference (μAbs_{diff}) between measured and simulated flow volume (Qv), peak discharge (Qp), sediment yield (Sy), and peak sediment flux (Sp) showed larger values in Kit Brook (48% [Qv], 66% [Qp], 298% [Sy], and 438% [Sp]) compared with the Blackwater catchment (30% [Qv], 41% [Qp], 106% [Sy], and 86% [Sp]). Results indicate that SHETRAN can produce reasonable flow prediction but performs less well in estimation of sediment flux, despite reasonably similar hydrosedimentary behaviour between catchments. The sensitivity index showed flow volume sensitive to saturated hydraulic conductivity and peak discharge to the Strickler coefficient; sediment yield was sensitive to the overland flow erodibility coefficient and peak sediment flux to raindrop/leaf soil erodibility coefficient. The multiparameter sensitivity analysis showed that different combinations of parameters produced similar model responses. Model sensitivity to grid resolution presented similar flow volumes for different digital elevation model resolutions, whereas event peak and duration (for both flow and sediment flux) were highly sensitive to changes in grid size.

KEYWORDS
catchment modelling, model sensitivity, proxy catchment, sediment yield, SHETRAN

1 | INTRODUCTION

Surface flow and soil erosion are natural processes affected by changes in agricultural land use and management in catchments. Increases in surface run-off, event peak flows, and suspended sediment loads have been related to altered agricultural practices and management interventions (Deasy, Brazier, Heathwaite, & Hodgkinson, 2009; Deasy, Titman, & Quinton, 2014; Smith et al., 2018). Hydrological models provide useful frameworks for testing, understanding, and predicting catchment behaviour under different agricultural change scenarios and in response to mitigation measures designed to reduce surface run-off, erosion, and suspended sediment loads.

Physically based spatially distributed (PBSD) hydrological models represent processes using fundamental physical equations (Merritt, Letcher, & Jakeman, 2003) usually simulating the hydrological cycle and sediment generation processes (Daniel et al., 2011). These types of models are generally classified according to their spatial scale, temporal distribution, variables, and process descriptions (Aksoy & Kavvas, 2005). PBSD models have the advantage, over empirical or conceptual models, in their ability to better represent the effect of changes in catchment conditions as their parameters relate to physical properties (Bathurst, 2011). However, a frequent problem with PBSD models is the large number of parameters required; some of which may be unavailable or inaccessible, particularly in data scarce regions (Tarawneh, Bridge, & Macdonald, 2016). This data uncertainty can be reduced by a model calibration approach, followed by an evaluation procedure, providing increased confidence in model outputs for further application.

Klemeš (1986) described several approaches for model calibration and evaluation. For example, the split-sample test uses measured data, divided into two periods of similar length, one for calibration and the other for evaluation, or the proxy-catchment test that comprises model calibration in one catchment and evaluation in a nearby catchment. The proxy-catchment approach is recommended as a framework for testing model performance (Pechlivanidis, Jackson, McIntyre, & Wheater, 2011; Xu & Singh, 2004) but is less commonly applied as it requires more resources than the split-sample test. Comparatively few studies use the proxy-catchment approach for model evaluation (Arsenault, Brissette, & Martel, 2018; Gumindoga, Rwasoka, Nhapi, & Dube, 2017; Yang, Herath, & Musiake, 2002) with limited applications using PBSD models (Refsgaard, Storm, & Refsgaard, 1995). Moreover, studies that adopt this model testing approach in combination with high-resolution measurements of discharge and sediment flux were not found in the available literature.

Another common criticism of PBSD model is grid size applications (Beven, 1991; Brazier et al., 2011), as parameter values applied to a catchment grid are usually an average of a physical property (measured or calibrated). This generalization fails to capture natural variability in a given property and may affect hydrological process representation (Thomas et al., 2016). Thompson, Bell, and Butler (2001) found that decreasing the digital elevation model (DEM) resolution tends to produce a smoother topography, as larger grids average the elevation of the covered area, losing important landscape detail. The loss of the topographic features can also lead to changes in overland flow pathways affecting sediment transport and deposition processes (Thomas et al., 2017). Moreover, the change in spatial resolution may affect model representation of tributaries (Daniel et al., 2011). Errors can be minimized by using the finest possible grid resolution. However, this requires more computing capacity and increases model run times (e.g., De Figueiredo & Bathurst, 2007). Despite the scale resolution problem being addressed in several studies (Lesschen, Schoorl, & Cammeraat, 2009; Valeo & Moin, 2000; Zhang & Montgomery, 1994), there is a lack of information on the effects of changes in DEM resolution on flow and sediment flux at an event scale with comparison against measured data. Furthermore, understanding PBSD model sensitivity to DEM resolution may be important for assessing predictions of the effectiveness of management interventions to reduce agricultural impacts, particularly when accurate representation of an intervention (e.g., minimum width of field buffer strips) is dependent on grid size. The selection of one PBSD model over another often depends on the available information and accessibility. The SHETRAN model was selected in the present study for its capability to simulate flow and sediment flux on an event and continuous basis, at any given resolution (with a maximum limit of 300 × 300 grids [Windows Version v4.4.5x64]). Moreover, event-based sediment prediction on a continuous basis can reduce errors related to initial conditions (Bussi, Francés, Montoya, & Julien, 2014). The SHETRAN model has been designed to simulate a wide range of land uses (e.g., agricultural, forest, or urban) and has been used globally, with a limited number of applications in the United Kingdom (Bathurst, Ewen, Parkin, O'Connell, & Cooper, 2004; Birkinshaw, 2008; Janes, Holman, Birkinshaw, O'Donnell, & Kilsby, 2017).

This study represents the first application of SHETRAN using a proxy-catchment approach with a semicoarse spatial resolution (50 × 50 m) at a high temporal resolution (15 min). Continuous flow and suspended sediment data from two nearby agricultural catchments in the United Kingdom were used for model calibration and evaluation, with the assessment of model efficiency focussing on flow volume, peak discharge, sediment yield, peak sediment flux, and event duration (flow and sediment) for a range of DEM grid sizes.

2 | MATERIALS AND METHODS

2.1 | The SHETRAN model

SHETRAN uses deterministic equations and finite-difference solutions to simulate water flow, sediment transport, and contaminant transport processes (Ewen, Parkin, & O'Connel, 2000). It has a distributed response at a catchment scale limited by 300 × 300 grids, each one containing soil and vegetation information. Soil profiles represented by columns of stacked grids (up to six layers) allow lateral and vertical flow transport for a three-dimensional subsurface flow simulation. The river network is modelled as channel links along the edges of grids (Birkinshaw, 2010).

Water flow comprises the subprocesses of evapotranspiration/-interception, overland/channel, and variably saturated subsurface. The first uses climatological data of rainfall and potential evapotranspiration (PE). The PE can be introduced by the user or calculated by the model using the Penman–Monteith equation (Monteith, 1965; Penman, 1948). The interception process that calculates the amount of rainfall stored by the canopy is based on the canopy surface storage capacity (Rutter, Morton, & Robins, 1975). Evaporation from soil occurs at a rate determined by the proportion of bare ground. This soil moisture control is assessed by the state of the unsaturated flow limits; the losses on the soil are calculated by a simple linear relation between the soil moisture tension (Ψ) and the actual evapotranspiration (AE) and potential evapotranspiration (PE) ratio (AE/PE), specified in the model as a table of AE/PE values against Ψ. SHETRAN simulates soil erosion by raindrop/leaf drip impact and overland flow, overland and channel transport of eroded sediment, and deposition (Lukey, Bathurst, Hiley, & Ewen, 1995).

In SHETRAN, the reported parameters to which the model is most sensitive on the water flow component are AE/PE ratio, saturated hydraulic conductivity (K_{sat}), and overland Strickler coefficient (Stk; Anderton, Latron, & Gallart, 2002; Bathurst, 1986; Op de Hipt et al., 2017). The AE/PE ratio influences water balance; the saturated hydraulic conductivity (K_{sat}) affects the surface–subsurface infiltration and storage; and the overland Strickler coefficient (Stk) has an effect on the surface flow velocity (Ewen et al., 2000). It is suggested by SHETRAN developers that AE/PE and soil depth are calibrated in the first instance, afterwards, if necessary, K_{sat} and Stk (Birkinshaw, n.d.). In the sediment transport component, the reported parameters to which model is most sensitive are raindrop soil erodibility coefficient (K_r) and overland flow erodibility coefficient (K_f; Adams & Elliott, 2006; Op de Hipt et al., 2017) with the parameters' values associated with catchment soil conditions and vegetation characteristics. For example, Verhaegen (1987) established a relationship between K_r and soil texture. However, relationships between soil texture and K_f are not commonly reported; hence, calibration of this coefficient is often used.

Performance of the SHETRAN water flow component has been reported in a number of studies using the Nash–Sutcliffe efficiency (NSE) coefficient. The term ranges from 1, for a perfect match, to $-\infty$. The reported NSE values for SHETRAN are between 0.6 and 0.8 (Adams, Western, & Seed, 2012; Janes et al., 2017; Mourato, Moreira, & Corte-Real, 2015). Inaccurate peak discharge prediction has been addressed in previous SHETRAN applications where better estimation of base flow and poorer prediction of discharge peaks have been reported (Zhang, Santos, Moreira, Freire, & Corte-Real, 2013). Underestimation of event peak discharge was also found by comparing simulated flow with data from eight gauging stations within a 2,400 km² catchment in the United Kingdom (Janes et al., 2017). Conversely, successful double-peak discharge predictions were obtained in a U.K. catchment (Birkinshaw, 2008). However, in this study, estimation of phreatic level against two measured sites was not accurate. This phreatic surface level is related to base flow. To assess model uncertainty, the "blind validation" method (Ewen & Parkin, 1996) was developed, which consists of establishing parameter bounds for the prediction of certain hydrological variables (e.g., hydrographs, phreatic surface level, peak discharge, soil water potential, and run-off) and calculating the degree (perceptual value established by the user) to which a measured variable lies within the predicted range. Using this method, a successful prediction of phreatic level (90%) but poorer prediction of peak discharge (81% [outlet] and 55% [channel]) was obtained for a U.K. catchment (Bathurst et al., 2004).

The sediment transport component efficiency has also been assessed. However, as continuous measurements of sediment flux are more difficult to obtain, evaluation has largely focussed on individual events with different coefficients. For example, in a small catchment in New Zealand, the difference between observed and predicted sediment loads (kg) for seven events varied between −3.3 (−97%) and 55 (134%) depending on the event (Adams & Elliott, 2006). Similarly, a single event in four nearby catchments in New Zealand (Elliott, Oehler, Schmidt, & Ekanayake, 2012) showed sediment yield error of 1%, 3%, 13%, and 110% and NSE values of 0.61, 0.70, 0.86, and 0.65, respectively, for each catchment. Additionally, the coefficient of determination (R^2) between simulated and measured daily sediment yield on two subcatchments of the Brazilian Sumé catchment showed values of 0.35 and 0.28 (De Figueiredo & Bathurst, 2007). Furthermore, using the blind validation, errors associated with 5-year sediment yields were reported with values ranging from 1% to 198% in a catchment in France (Lukey, Sheffield, Bathurst, Hiley, & Mathys, 2000).

The parameters to which SHETRAN is most sensitive varied according to catchment characteristics (location, topography, vegetation, etc.). Nonetheless, SHETRAN studies showed similar Stk values for land cover with the highest values for urban areas, followed by cropland, grassland, and woodland as the lowest (Bathurst, Moretti, El-Hames, Beguería, & García-Ruiz, 2007; Bathurst, Moretti, El-Hames, Moaven-Hashemi, & Burton, 2005; Birkinshaw, 2008; Elliott et al., 2012; Wicks & Bathurst, 1996). Moreover, on-site studies also indicated a similar Stk trend for land covers (Engman & American Society of Civil Engineers, 1986; Kværnø & Stolte, 2012; Rahimy, 2011; Schob, Schmidt, & Tenholtern, 2006). In SHETRAN, hillslope surface run-off is simulated as a sheet flow (fine, extensive, and surficial) rather than confined flow (e.g., rills or gullies). The flow depth and velocity are an average value in each grid; therefore, a coarse spatial resolution will generate a wider sheet flow (equal to the size of the grid). The dependence of overland flow erosion by surface flow makes the K_f coefficient an important parameter to be considered in the calibration process, with the empirical value representing a combined sheet and rill flow effect (Wicks & Bathurst, 1996). Nonetheless, different erodibility coefficient values were calculated in plot scale studies with different agricultural land uses (Wicks, Bathurst, & Johnson, 1992). Wicks et al. (1992) showed that erodibility coefficient variation depends on the land use (tilled, clipped grass, grazed, and ungrazed). However, there is high variation of soil erodibility coefficients in SHETRAN applications, with K_r ranges

from 0.05 to 70 (J^{-1}) and K_f between 5×10^{-7} and 2×10^{-5} (kg m^{-2} s^{-1}; Bathurst, 2011). Most of the reported SHETRAN studies employ single value for the erodibility coefficients (K_f and K_r). Moreover, the values varied with the spatial resolution of model applications (Bathurst et al., 2005; Bathurst, Burton, Clarke, & Gallart, 2006; Janes et al., 2017; Lukey et al., 2000).

2.2 | Study area

The Blackwater (18 km^2) and Kit Brook (22 km^2) catchments are part of the River Axe hydrological network in south-west England (Figure 1). Catchment DEMs derived from Ordnance Survey (OS) data were obtained through EDINA Digimap with a 5 m resolution. The elevation of the Blackwater catchment ranges from 49 m at the outlet up to 250 m on the crest (south-east), and slopes range from 0° to 36° across the catchment. The Kit Brook catchment elevation rises from 42 m at the outlet to 251 m in the upper northeast reaches, with slope angles ranging between 0° and 36°.

The nearest rainfall station with 15 min data available for the period overlapping hydrological measurements was at Raymond's Hill (Met Office, 2018 [50°46′0.84″N 2°57′46.08″W, 85 m]), approximately 5 km south of the catchments. Mean daily temperature data were obtained from Seavington station (ID 9092 [50°56′26.9″N 2°51′35.6″W, 85 m]) approximately 16 km north of the catchments, which was used to estimate daily PE (mm) using a PET formula (Oudin, Michel, & Anctil, 2005). Climatological data range from 1 October 2009 to 30 September 2014.

Measurements of pressure and turbidity were obtained with a 15 min time step. Troll 9500 probes were installed at the outlet of each catchment in September 2010, and measurements continued until December 2014. Pressure data were converted to flow (m^3 s^{-1}) based on the stage-discharge rating curve using available measurements collected for each catchment. An extrapolation method was used based on three types of regression equation: linear, quadratic, and power. Mean and base flow values were also obtained from similar size catchments in the region (Centre for Ecology and Hydrology, 2018) and compared with estimated flow. The quadratic equation gave values closest to other gauged flow and the highest regression coefficients (R^2 = 0.9 [Blackwater], 0.9 [Kit Brook]). Flow duration curves were calculated and show similar flow responses between the catchments (Figure 2). In Kit Brook, the fraction of daily discharge over annual mean flow was slightly higher than in Blackwater 90% of the time, whereas Blackwater exceeded Kit Brook only during periods of high flow, suggesting a higher permeability in Kit Brook than Blackwater. Turbidity data were converted to suspended sediment concentration (mg L^{-1}) based on the regression equation, $y = 1.1049 \times -15.005$ (R^2 = 0.9; Little, 2012), and suspended sediment flux (kg s^{-1}) was computed from sediment concentration and flow. Issues with turbidity data quality arose during some periods; therefore, it was necessary to exclude some events from further analysis.

The most recent land cover for each catchment was characterized by a 2010 ground-based field survey of the Blackwater catchment from the Westcountry Rivers Trust and by digitizing 2010 imagery from Google Earth for Kit Brook. The 2010 survey map was classified into four land covers: (a) urban, (b) deciduous woodland, (c) arable

FIGURE 1 Location of catchments, meteorological stations, and stream gauging sites instrumented with level and turbidity probes (Troll 9500)
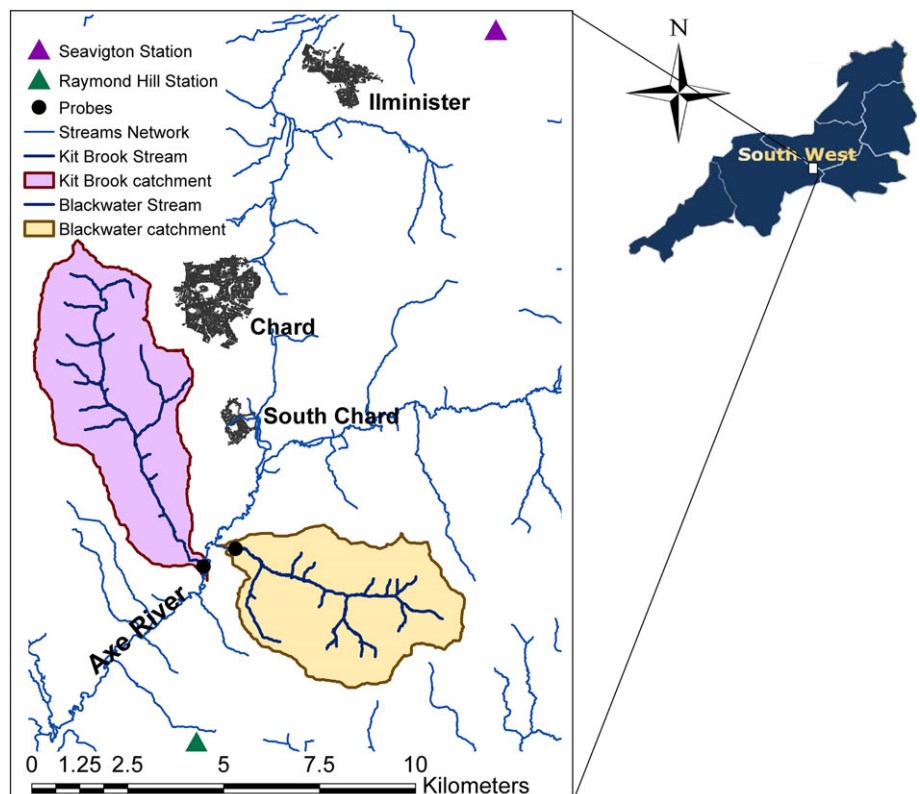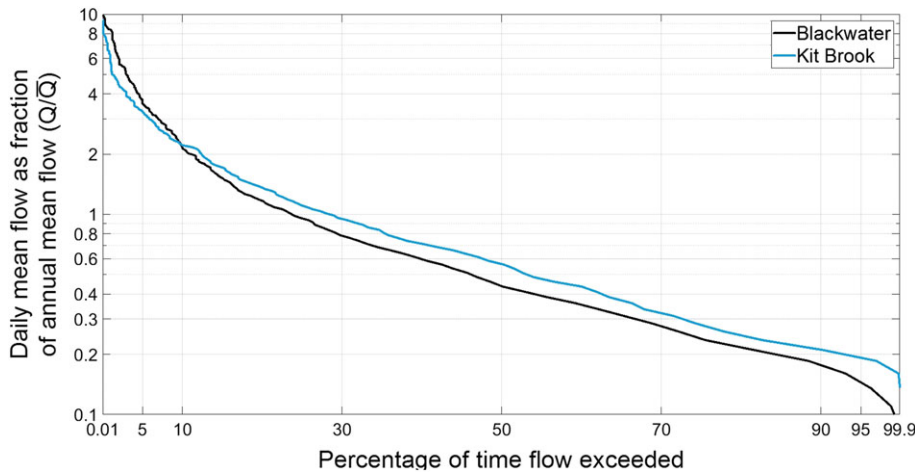
FIGURE 2 Flow duration curve for Blackwater and Kit Brook streams, period 2010–2014

crops, and (d) grass. In Blackwater, the land cover was grass (60%), arable crops (27%), deciduous woodland (12%), and urban (1%) with comparable land cover proportions observed in the Kit Brook catchment: 57%, 29%, 13%, and 1%, respectively. In the south-west region of England, winter cereals are the most common crop type (Department for Environment, Food and Rural Affairs, 2015); hence, winter barley was used as the simulated arable crop for both catchments. Soil data were acquired from the National Soil Resources Institute (NSRI) of Cranfield University (Cranfield University, 2018). Five soil types were identified in the Blackwater catchment; the one with the largest extent (40%), namely, WICKHAM (Eutric Luvic Planosols), had a soil texture of 21% sand, 41% silt, and 30% clay. Each soil is characterized by five depth layers between 0 and 1.5 m; and the properties for each soil type and layer vary depending on the land use (Hollis et al., 2015). Six soil classifications with five depth layers (0–1.5 m) were identified in the Kit Brook catchment; BATCOMBE (18% sand, 58% silt, and 24% clay) and CHARITY (16% sand, 58% silt, and 26% clay) soils cover the 37% and 31% of the catchment, respectively. The NSRI soils database provides the international standard soil classification (International Union of Soil Sciences, 2007) related to the soil classification of England and Wales, in which WICKHAM corresponds to Eutric Luvic Planosols, BATCOMBE to Profundic Chromic Endostagnic Luvisols, and CHARITY to Chromic Luvisols. The QUORND soil that covers 20% of the Kit Brook catchment was substituted for HENCE soil (60% sand, 25% silt, and 15% clay) due to the lack of information available in the NSRI.

2.3 | Event selection

Individual flow events were selected in both catchments to enable event-based analysis of suspended sediment flux for model calibration and testing. Issues with turbidity data quality arose during some periods; for example, sensor fouling or burial, leading to periods of persistent high turbidity values until cleaning, or on occasion, large rapid variations in measured turbidity occurred that were unrelated to any change in flow. Therefore, it was necessary to exclude some events from further analysis due to these data quality issues.

Discharge event analysis focussed on selecting a subset of events in each catchment that were determined to have "good quality" flow and turbidity data. The start of a flow event was defined by when flow exceeded the base flow and where sediment flux increased above the prior baseline data. The end of an event was determined when flow fell to the pre-event level or to a new temporary base level, which exceeded the flow prior to the event (Robson & Reed, 2008). Total flow volume (m^3; Equation (1)), maximum flow peak ($m^3 s^{-1}$), sediment yield (t ha^{-1}; Equation (2)), and maximum sediment flux peak (kg s^{-1}) of each event were obtained. The reference date for each event was stipulated as the day in which the maximum flow peak was found. On the basis of this event selection, 53 events were identified in Blackwater and 46 in Kit Brook. For most selected events, a clockwise suspended sediment concentration–discharge (C-Q) hysteresis behaviour (Williams, 1989) was observed in both catchments (i.e., sediment peak arrives before the discharge at the outlet and/or presenting sediment exhaustion).

$$Qv = \sum [(Q_i)(900)], \tag{1}$$

$$Sy = \left\{ \sum (Sed_i)(900)(0.0011) \right\} / \{(ha)\}, \tag{2}$$

where Q = flow ($m^3 s^{-1}$), Sed = sediment flux (kg s^{-1}), i = number of time steps (15 min) in an event, and ha = hectares.

2.4 | Model set-up

The use of a 25 × 25 m grid size for Blackwater produced a grid number of 239 × 170, fitting with the model grid limit (300 × 300), but this limit was exceeded for Kit Brook (180 × 355). Furthermore, a run time of 15 days was observed for this resolution over the complete simulation period. Therefore, the 50 × 50 m grid size was selected to produce a reasonable running time (24 hr for 5 years); and this grid was converted from the 5 × 5 m DEM in both catchments.

Vegetation and sediment parameter values (Appendix A) were selected from a literature review (Birkinshaw, 2008; Lukey et al.,

2000; Wicks et al., 1992). NSRI data were used in the model with a soil depth up to 1.5 m; parameter values (e.g., K_{sat}, θ_{res}, and θ_{sat}) varied according to each soil layer and land cover (Appendix B). Rainfall data with a 15 min time step and calculated daily potential evapotranspiration were used for an initial simulation from October 2010 to September 2014 with a "spin-up" hydrological year (October 2009–September 2010) to obtain phreatic surface level equilibrium.

2.5 | Calibration and evaluation processes

Calibration aims to improve model performance by changing values of selected parameters, through either "manual" or automated methods. A manual process is commonly applied, in which parameters are changed "one at a time" (OAT; Rabitz, 1989). This necessitates determination of which parameters require calibration and consideration of the relationship between parameters. It also requires a choice of the model output to be calibrated (e.g., run-off, phreatic level, sediment concentration, discharge peaks, and sediment flux peaks). Furthermore, multiple different combinations of parameter values might give a good fit (Beven & Binley, 1992; Jetten & Maneta, 2011). The decision on parameters to be calibrated depends on information about previous applications (i.e., sensitivity analysis) and/or user experience.

The calibration process was performed in the Blackwater catchment. The first simulation used five soil layers (0–1.5 m), although for an accurate base flow, subsequent simulations required the addition of a layer (sixth [1.5–20 m]) to represent soil from subsoil to bedrock (Birkinshaw, n.d.). The water flow component was calibrated by comparing the measured 4-year discharge record to the model simulation and by quantitative comparison of the selected discharge events. Afterwards, the sediment transport component was calibrated using the selected events. Using the proxy-catchment approach, model evaluation was undertaken in the Kit Brook catchment for the same spin up and run period using parameter values from the final Blackwater calibration.

The NSE (Equation (3)) was used to assess model fit with the continuous discharge record. For assessing event-based performance, the coefficient of determination (R^2) between measured and simulated data was calculated for flow (R^2_Q) and sediment flux (R^2_{Sed}), respectively. Coefficient values (NSE and R^2) higher than 0.5 are considered as good fit (Moriasi, Arnold, Van Liew, Harmel, & Veith, 2007).

$$NSE = 1 - \left\{ \left[\sum \left((S_i - O_i)^2 \right) \right] / \left[\sum \left((O_i - \mu O)^2 \right) \right] \right\}, \qquad (3)$$

where O_i = observe measurement at the time i (i = 1 … m), S_i = model simulation at the time i (i = 1 … m), and μO = mean of observe measurements (1 … m).

Measured and simulated flow volume (Qv [m³]), maximum flow discharge (Qp [m³ s⁻¹]), sediment yield (Sy [t ha⁻¹]), and maximum sediment flux (Sp [kg s⁻¹]) were obtained for each selected event. The absolute percentage difference between measured and simulated data (Qv, Qp, Sy, and Sp) for each event (Abs_{diff}) was calculated. The mean

of the Abs_{diff} was used to quantify event model performance (Equation (4)).

$$Abs_{diff} = (1/n) \left\{ \sum \left((O_j - S_j)/O_j \right) (100) \right\} \qquad (4)$$

where n = number of events, O = measured: Qv, Qp, Sy, and Sp, S = simulated: Qv, Qp, Sy, and Sp, and j = event (1 … n).

Simulations were run by changing parameters values using the OAT method with the objective of reaching the highest possible NSE and the minimum μAbs_{diff}. Model accuracy in flow and sediment flux prediction was assessed for the final calibration (Blackwater) and for the evaluated simulation (Kit Brook) using the previous coefficients and additionally by comparing measured against simulated data using (a) monthly mean run-off; (b) a monthly window of the coefficient of determination (R^2) for discharge and sediment flux (i.e., R^2_Q and R^2_{Sed}) for the selected events; and (c) R^2 for Qv, Qp, Sy, and Sp.

2.6 | Sensitivity analysis

The parameters reported as the most sensitive in other studies of SHETRAN sensitivity (Bathurst, 1986; Bathurst et al., 2004; Op de Hipt et al., 2017; Parkin et al., 1996; Wicks et al., 1992; Wicks & Bathurst, 1996; Wicks, Bathurst, Johnson, & Ward, 1988) include AE/PE ratio, Strickler coefficient (Stk), saturated hydraulic conductivity (K_{sat}), raindrop soil erodibility coefficient (K_r), and overland flow erodibility coefficient (K_f). In the present study, the sensitivity analysis consisted of varying values of selected parameters upward and downward, relative to the base run (final calibration), by factors of 10 (F^{10}), 0.1 ($F^{0.1}$), and 0.01 ($F^{0.01}$) with the OAT method. The factorial changes were applied to the parameters' value for each of the four land cover types, and the maximum possible value was used when a parameter value exceeded the physical limit.

An adaptation of the sensitivity index (SI; Sheikh, van Loon, Hessel, & Jetten, 2010) was implemented using model outputs of the sensitivity simulations (F^{10} and $F^{0.01}$) and the base run (F^1). The SI was calculated for each assessed parameter according to Equation (5). Model outputs analysed included 4-year cumulative flow volume (Qv) and the mean of flow peaks (μQp) for the parameters in the water flow component (AE/PE, Stk, and K_{sat}) and 4-year sediment yield (Sy) and mean of sediment flux peaks (μSp) for the parameters in the sediment transport component (K_r and K_f).

$$SI = \left(F^{10}_{out} - F^{0.01}_{out} \right) / F^1_{out} \qquad (5)$$

where F^{10}_{out} = model output by factor change of 10 on the corresponding parameter, $F^{0.01}_{out}$ = model output by the factor change of 0.01 on the corresponding parameter, and F^1_{out} = model output for the base run on the corresponding parameter.

The multiparameter approach to sensitivity analysis consisted of running simulations with a combination of parameter changes for the water flow (AE/PE, Stk, and K_{sat}) and sediment transport

components (K_r and K_f). Factorial change of parameters from $F^{0.1}$ to F^{10} on the water flow component produced 27 simulations and nine simulations on the sediment transport component (Figure 6). The percentage of difference (Pct_{diff}) for each model output relative to the base run was calculated by Equation (6). The base run simulation ($S_{out}{}^{br}$) for the parameters on the water flow and sediment transport components corresponds to simulation numbers 14 and 5, respectively (Figure 6).

Model sensitivity to grid resolution was examined by running simulations of the Blackwater catchment with grid sizes of (a) 25 × 25 m, (b) 50 × 50 m, (c) 100 × 100 m, and (d) 200 × 200 m. Considering the longer model run time associated with the finest resolution (15 days), the climate records were restricted to 3 years (October 2009–September 2010 [spin up] and October 2010–September 2012). Model sensitivity to changes in grid resolution was assessed by comparing outputs of the base run (50 × 50 m, final calibration) against each resolution (Equation (6)). The model output considered was 4-year flow volume (Qv), mean of discharge peaks (μQp), mean of event flow duration (μQw), 4-year sediment yield (Sy), mean of sediment flux peaks (μSp), and mean of event sediment duration (μSw). The width of the fluxes (flow [Qw] and sediment [Sw]) was measured as the distance between the intercept points to the left and right of the one-half peak height (Robson & Reed, 2008).

$$Pct_{diff} = \left(\left(S_{out}^{br} - S_{out}^{i} \right) / S_{out}^{br} \right) 100, \qquad (6)$$

where $S_{out}{}^{br}$ = model output of base run simulation and $S_{out}{}^{i}$ = model output of simulation i (i = 1 … m), m = 27 for the multiparameter analysis on the water flow component, 9 on the sediment transport component, and 4 for the simulation on model sensitivity to grid resolution.

The catchment representation of grid size resolution (25 × 25, 50 × 50, 100 × 100, and 200 × 200 m) was assessed using probability density plots of the distribution of topographic slope and percentage of land cover represented. Model comparison with measured flow and sediment flux was undertaken for each of the grid size simulations using the same coefficients as in the model calibration and evaluation processes (NSE, μAbs_{diff} [Qv, Qp, Sy, and Sp]).

3 | RESULTS

3.1 | Model calibration in Blackwater catchment

The water flow component was calibrated by changing four parameters. The parameter value and NSE for continuous discharge and the μAbs_{diff} (Qv and Qp) of the best fit simulation are presented in Table 1. The sequence in the table follows the order in which parameters were calibrated. After the best-fit discharge is obtained, four parameters were calibrated for the sediment transport component in which the μAbs_{diff} (Sy and Sp) was calculated for each simulation. The second baseline simulation appearing in Table 1 corresponds to the last calibration for the water flow component

and set-up values in the sediment transport component. The K_f value required an escalation of the plot measurements reported for tillage and clipped grass (Wicks et al., 1992). An over prediction for the majority of events was observed with the K_f value of 59 × 10^{-10} (kg m^{-2} s^{-1}) and an under prediction with 6.9 × 10^{-10} (kg m^{-2} s^{-1}). The highest K_f was chosen for cropland and the lowest for grass and woodland. A similar procedure was performed for K_r parameter using plot values reported in Wicks et al. (1992).

Monthly mean run-off (Figure 3a) showed the lowest differences between simulated and measured data during autumn (2 mm) and winter period (−6 mm), whereas simulated run-off during spring and summer exceeded measured by 23 and 17 mm, respectively. The simulated annual run-off showed an over estimation of 5%.

Model performance on event scale showed that 83% of the discharge events and 64% of the sediment flux events have R^2 values higher than 0.5 (Figure 4a,b). This indicates an accurate representation of event timing by the model. Moreover, model presented better prediction of flow volume (Qv [R^2 = 0.8]) than flow peak (Qp [R^2 = 0.6; Figure 4c,d). Correspondingly, the absolute percentage of difference between the measured and simulated data (μAbs_{diff}) showed lower values for Qv (30%) than Qp (41%). A high variability was observed in the representation of sediment export compared with measured data; better prediction of sediment yield (Sy [R^2 = 0.4]) than sediment flux peaks (Sp [R^2 = 0.2]) was observed using the coefficient of determination (Figure 4e-f). In contrast, the μAbs_{diff} was higher for Sy (106%) than Sp (86%; Table 2.

3.2 | Model evaluation in Kit Brook catchment

Model simulation of flow and sediment flux in Kit Brook showed a lower NSE, larger μAbs_{diff} (Qv, Qp, Sy, and Sp) and similar R^2 compared with the calibrated results for the Blackwater catchment (Table 2). Model runs in Kit Brook used the final calibrated parameter values obtained from the Blackwater simulation. The measured mean monthly run-off (mm) was higher than simulated, especially for winter months. In contrast to Blackwater, simulated mean annual run-off in Kit Brook was 10% under predicted (Figure 3b).

The model showed poorer performance when comparing Kit Brook with Blackwater on an event basis. Events with R^2 higher than 0.5 represent the 80% of the total discharge events (Figure 5a) and 41% of the sediment flux events (Figure 5b). The 46 measured events in Kit Brook presented a higher model predictive performance for Qv (R^2 = 0.8) than Qp (R^2 = 0.6; Figure 5c,d), which correspond to μAbs_{diff} values for Qv and Qp of 48% and 66%, respectively. Lower performance was observed for sediment prediction (R^2 = 0.6 [Sy] and R^2 = 0.3 [Sp]; Figure 5e,f), corresponding to μAbs_{diff} values for Sy and Sp of 298% and 438%, respectively.

The higher observed permeability in Kit Brook, based on comparison of the flow duration curves (Figure 2), supported adjustment of the K_{sat} for the sixth soil layer in this catchment. Increasing the K_{sat} value in this layer showed a difference of 7% between simulated and measured annual run-off. Furthermore, better NSE and μAbs_{diff} coefficients (Table 2) were observed with this

TABLE 1 Calibrated simulations with the corresponding best fit parameter values (NSE and μAbs_{diff} [Qv, Qp, Sy, and Sp])

Parameter	Value				NSE	μAbs_{diff} (%) Qv	Qp
	Baseline[a]				0.60	38	47
K_{sat}[b] (m day^{-1})	0.05				0.76	30	41
	Urban	Deciduous	Arable	Grass			
Stk (m$^{1/3}$ s^{-1})	1.2	0.6	0.8	1.0	0.77	30	41
CLAI[c]	0	1[d]	1[d]	1	0.79[e]	29	40
PLAI[f]	0.1	0.9[d]	0.8[d]	0.5	0.78[g]	30	41
						Sy	Sp
	Baseline[a]					556	340
K_f (×10^{-10} kg m^{-2} s^{-1})	6.5					91	82
K_f (×10^{-10} kg m^{-2} s^{-1})	59					245	162
	Urban	Deciduous	Arable	Grass			
K_f (×10^{-10} kg m^{-2} s^{-1})	0.0	6.5	59	6.5			
K_r (J^{-1})	0.0	1.3	11.8	1.3			
F_{CG}[h]	0.0	0.9	0.3	0.8		106	86
F_{CROCK}[i]	1.0	0.0	0.0	0.0			

[a]See Appendix A.
[b]With addition of a sixth homogenous layer with 18.5 m depth (1.5–20 m).
[c]Ratio of total leaf area to area of ground covered by vegetation.
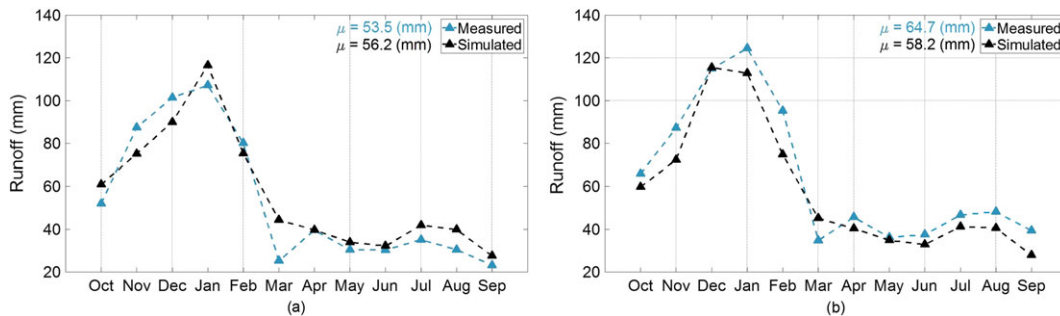[d]Values varied over time.
[e]Twenty-four events with $NSE_Q > 0.5$.
[f]Proportion of ground covered by vegetation.
[g]Twenty-seven events with $NSE_Q > 0.5$.
[h]Proportion of ground shielded by near-ground cover.
[i]Proportion shielded by ground-level cover.

FIGURE 3 Measured and simulated monthly mean runoff in (a) Blackwater catchment and (b) Kit Brook catchment

K_{sat} adjustment. However, the coefficient of determination for the event analysis (R^2_Q and R^2_{Sed}) did not vary significantly compared with the nonadjusted Kit Brook simulation (Appendix C).

3.3 | Model sensitivity

The SI (Table 3) showed that variations of K_{sat} affected flow volume (Qv) the most and Stk affected the mean of discharge peaks (μQp). Variations in AE/PE had little effect on both flow volume and mean of flow peaks. For the sediment transport component, sediment yield

(Sy) was most sensitive to the K_f parameter, whereas the mean of sediment flux peaks (μSp) was more sensitive to K_r. Nonetheless, SI differences between the erodibility coefficients (K_r and K_f) were relatively small for both model responses (Sy and μSp).

The multiparameter approach to sensitivity analysis on the water flow component showed how different combinations of the parameters, especially Stk and K_{sat}, can produce similar percentage of differences (Pct_{diff}) with respect to the base run (Figure 6a). For example, simulations S8 and S16 reduced flow volume (Qv) by 15%. Likewise, similar Pct_{diff} in mean of flow peaks (μQp) was observed in simulations
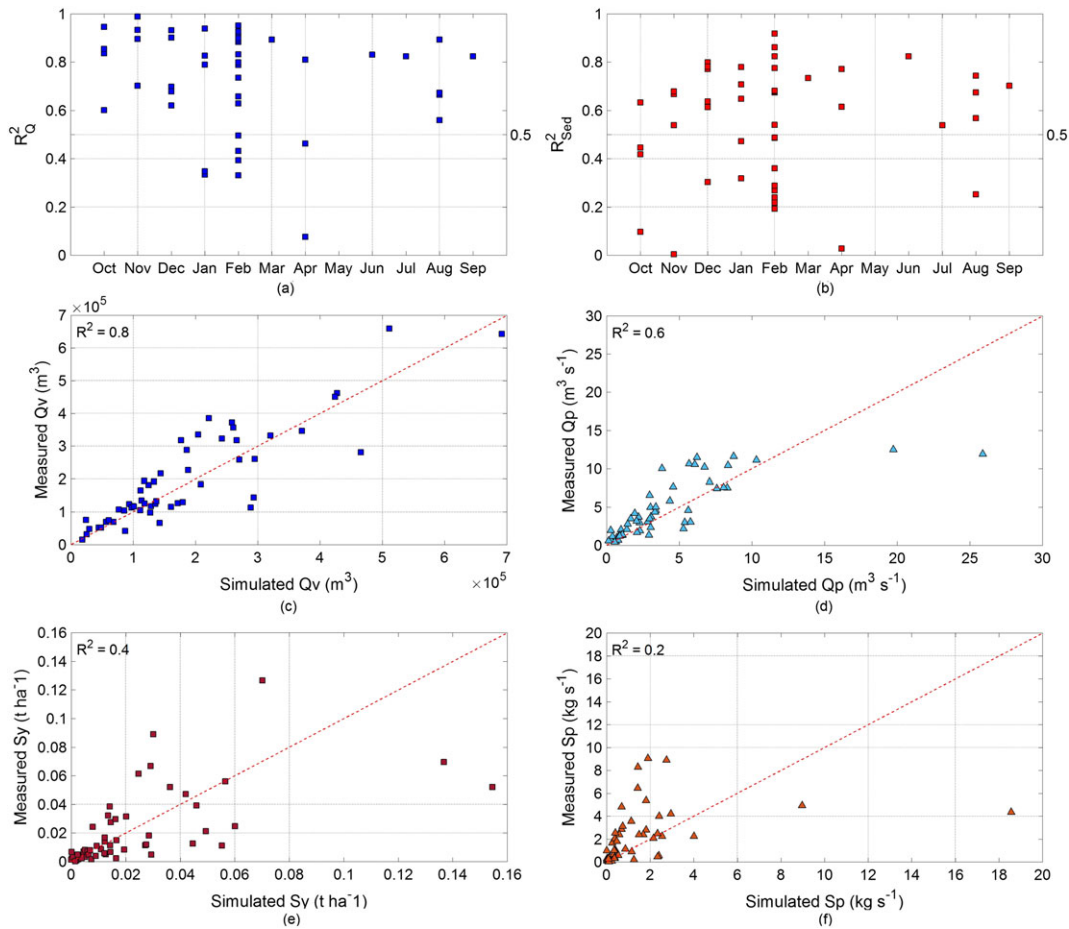
FIGURE 4 (a) Coefficient of determination (R^2_Q) between the measured and simulated discharge for each event in a monthly window, (b) coefficient of determination (R^2_{Sed}) between the measured and simulated sediment flux for each event in a monthly window and plots of simulated versus measured events for (c) flow volume (Qv), (d) peak flow (Qp), (e) sediment yield (Sy), and (f) peak sediment flux (Sp) for the Blackwater catchment

TABLE 2 Comparison of model performance coefficients for the Blackwater and Kit Brook catchment simulations

Catchment	NSE	μAbs_{diff} (%) (R^2)			
		Qv	Qp	Sy	Sp
Blackwater[a]	0.78	30 (0.8)	41 (0.6)	106 (0.4)	86 (0.2)
Kit Brook[a]	0.60	48 (0.8)	66 (0.6)	298 (0.6)	438 (0.3)
Kit Brook[b,c]	0.69	44 (0.8)	58 (0.6)	202 (0.6)	319 (0.3)

[a]K_{sat} = 0.05 m day^{-1} for the lowermost soil layer (sixth).
[b]K_{sat} = 0.10 m day^{-1} for the lowermost soil layer (sixth).
[c]See Appendix C for the event-based performance.

with different combination of K_{sat} and Stk values. For example, the simulations S8 and S15 presented μQp changes of −66%; and S2 and S21 produced changes of approximately 53%. An equifinality problem, in terms of flow volume and peaks, appears when increases in K_{sat} occur in combination with decreases in Stk, and vice versa. Although the shape of hydrographs varies between simulations, specifically, S8 presented more prolonged events with lower flow peaks compared with S16, resulting in similar Qv changes, whereas S15 presented

higher base flow during the 4-year period than S8; as a consequence, the mean of flow peaks showed similar μQp reductions when compared with the base simulation (S14).

The multiparameter analysis of parameters on the sediment transport component (Figure 6b) showed that minor changes in K_r combined with high K_f values produced similar sediment yield with respect to the base run (Pct_{diff} [Sy = 100%]), for example, simulations S3 (K_r [$F^{0.1}$] and K_f [F^{10}]) and S6 (K_r [F^1] and K_f [F^{10}]). Likewise, on the mean of sediment flux peaks (μSp), large changes in K_r combined with small changes in K_f (e.g., S7 and S8) presented similar differences compared with the base run (Pct_{diff} [μSp = 100%]).

In terms of sensitivity to DEM resolution, the probability density distribution of slopes for each DEM resolution shows that decreasing grid size leads to increasing variance in the distribution of slope angles (Figure 7a). The change in grid size produced only a very minor effect on land cover proportions represented in simulations (Figure 7b). The largest difference between grid resolutions was in channel network lengths (Figure 7c–f).

The DEM resolution simulations show that increasing grid size from 25 to 200 m leads to increases in μQw and decreases in μQp and
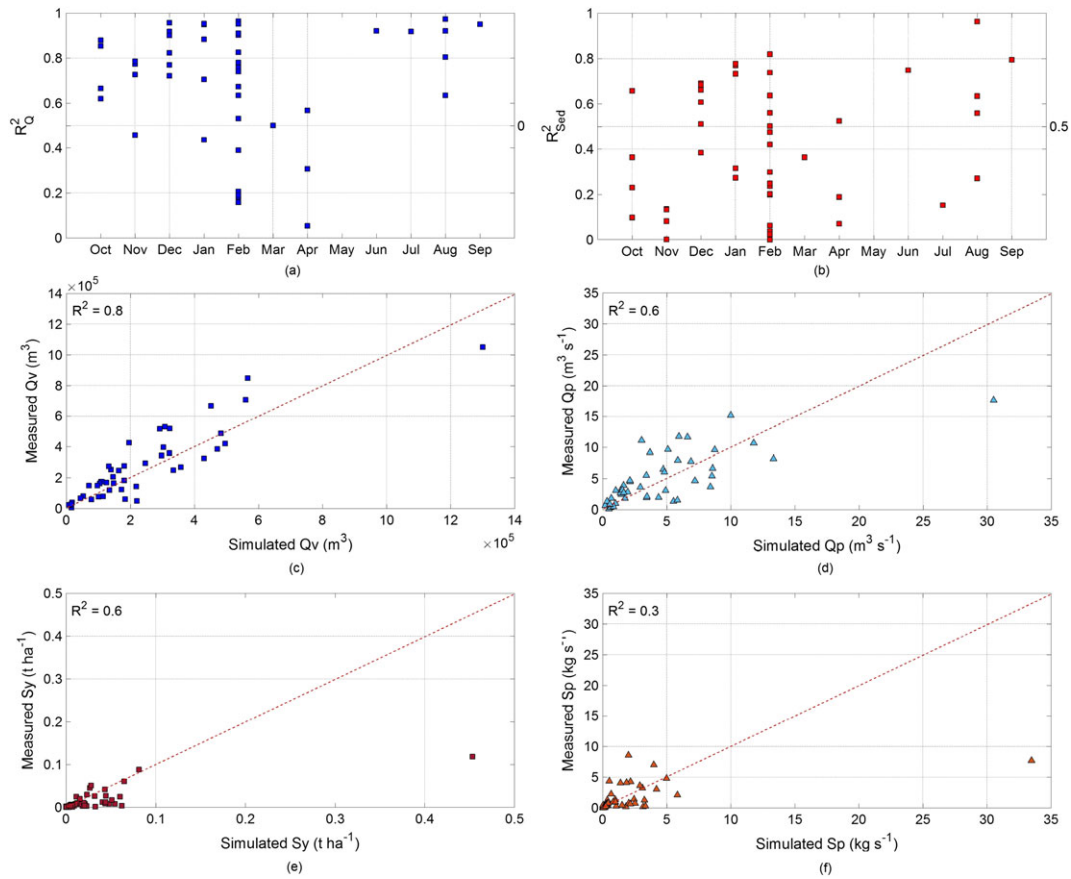
FIGURE 5 (a) Coefficient of determination (R^2_Q) between the measured and simulated discharge for each event in a monthly window, (b) coefficient of determination (R^2_{Sed}) between the measured and simulated sediment flux for each event in a monthly window and plots of simulated versus measured events for (c) flow volume (Qv), (d) peak flow (Qp), (e) sediment yield (Sy), and (f) peak sediment flux (Sp) for Kit Brook catchment

TABLE 3 Sensitivity index (SI) of cumulative flow (Qv) and sediment yield (Sy), and mean of event peak discharge (μQp) and peak sediment flux (μSp) of the complete simulation period (October 2010–September 2014)

Parameter (unit)	SI	
	Qv	μQp
AE/PE (dimensionless)	0.28	0.33
K_{sat} (m day^{-1})	0.75	1.09
Stk (kg m^{-2} s^{-1})	0.15	1.33
	Sy	μSp
K_r (1 J^{-1})	1.17	1.50
K_f (kg m^{-2} s^{-1})	1.27	1.30

Abbreviations: AE, actual evapotranspiration; PE, potential evapotranspiration.

Qv (Figure 8a). Only small differences (Qv, μQp, and μQw) were observed between grid sizes simulations of 25 and 50 m. In terms of sediment exports, increasing grid size (from 25 to 200 m) produced a decrease in μSp and increase in μSw. In contrast, Sy exhibited a much smaller but more varied response (Figure 8b).

In general, a better estimation (NSE, μAbs_{diff}, and R^2) was obtained for the 50 m grid size when comparing simulations of each resolution with measured data (Table 4). The coefficients showed a better fit to measured data with decreasing grid size, except R^2. Moreover, the two finest resolutions simulations (25 and 50 m) presented only slight differences between coefficients, especially NSE and μAbs_{diff}.

4 | DISCUSSION

SHETRAN predicted streamflow well in both catchments (Figure 3). The addition of a deep, low conductivity, and homogenous soil–rock layer (1.5 to 20 m depth) below the upper soil profile (parameterized using NSRI soil hydraulic data) enabled more accurate simulation of base flow (Table 2), which decreased gradually during extended dry periods. Moreover, the hydraulic conductivity of bedrock (2.7 to 20 m) in previous SHETRAN applications (Birkinshaw, 2008; Birkinshaw & Ewen, 2000) was similar to the present study.

Flow measurements derived from the stage-discharge rating curve may contribute to event-scale model uncertainty. Nevertheless, underestimation of discharge peaks and accurate base flow was
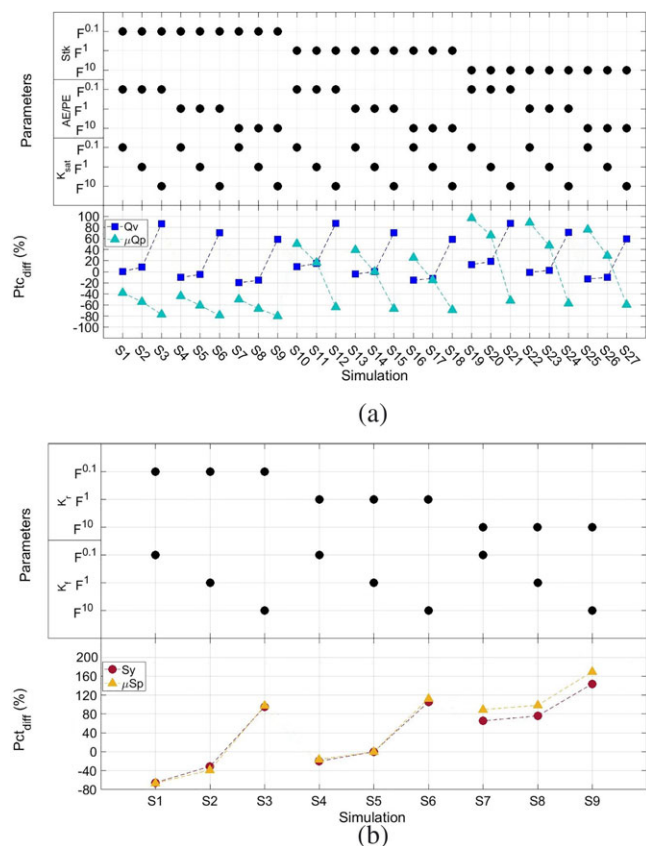
FIGURE 6 (a) Combinations of the changes in parameters (*Stk*, AE-PE, and K_{sat}) and the difference in flow volume (*Qv*) and mean of discharge peaks (*μQp*) for each simulation (S1–S27) with respect to the base run (S14), (b) combinations of the changes in parameters (K_r and K_f) and the difference in sediment yield (*Sy*) and mean of sediment flux peaks (*μSp*) for each simulation (S1–S9) with respect to the base run (S5)

obtained in both catchments. By contrast, accurate peak discharge estimation and under prediction of phreatic level (base flow) in SHETRAN have previously been reported, and vice versa (Adams, Parkin, Rutherford, Ibbitt, & Elliott, 2005; Bathurst et al., 2004; Birkinshaw, 2008; Ewen & Parkin, 1996; Janes et al., 2017; Zhang et al., 2013). Furthermore, similar behaviour has been observed with other hydrological models (Croke, Merritt, & Jakeman, 2004; De Roo & Jetten, 1999; Marsik & Waylen, 2006).

The proxy-catchment test in this study showed better performance in the calibrated catchment (Blackwater) than the evaluated catchment (Kit Brook) probably due to a higher soil permeability in Kit Brook than Blackwater (Figure 2). Nonetheless, SHETRAN showed reasonably accurate flow prediction in both catchments, which are similar in terms of area, relief, and land cover characteristics. The prediction errors (e.g., *Qv* and *Sy*) may be attributed more to the parametrization and DEM resolution rather than lack of model process representation. An important factor to consider for model predictions is that soil (NSRI) under grass represents physical and hydraulic properties of permanent grass land cover, whereas the land cover may include permanent and temporary grass. Soils under temporary grass could have different properties (water retention curve, bulk density,

and soil porosity) than under permanent grass. Moreover, temporary grass may be located in rotation fields, where soil physical properties can vary depending on time since last cultivation, and take between 2 and 9 years to recover (Thorud & Frissell, 1976; Tuzzin de Moraes et al., 2016) and up to 20 years in loamy soils (Froehlich, Miles, & Robbins, 1985).

In general, simulated discharge in both catchments compared well with measured discharge in terms of event timing and reproduced discharge peaks well during wet periods (Figures 4a–5a). However, the underestimation of peak discharge by SHETRAN could also be attributed to local infiltration variability, in terms of subfield scale run-off and run-on patches not represented in the model, as well as the absence of representation of impervious and hydrologically smooth roads and paths. These features could act as efficient flow pathways connecting run-off-generating areas to the stream network (Croke, Mockler, Fogarty, & Takken, 2005; Jordán-López, Martínez-Zavala, & Bellinfante, 2009) and have been implemented in other models (Elliot, 2004; Tiemeyer, Moussa, Lennartz, & Voltz, 2007) but not identified in any SHETRAN application.

The best-fit calibration for the Blackwater catchment resulted in lower mean errors at the event scale (i.e., $μAbs_{diff}$ [*Sy* and *Sp*]) than Kit Brook (Table 2). The lower model performance in Kit Brook (i.e., $μAbs_{diff}$) may be attributed to higher soil permeability in the catchment compared with its neighbour (Figure 2). A K_{sat} adjustment in the sixth soil layer produced an improvement in model performance (Table 2). Nonetheless, coefficients of the adjusted simulation presented lower values when compared with the best-fit calibration in Blackwater. This could be related to the necessary substitution of the HENSE soil type (covering 20% of the catchment) with the QUORND, based on their similar soil description as a result of the absence of soil parameters in the NSRI dataset (Cranfield University, 2018). Moreover, the lack of a land cover map produced by field-scale survey for Kit Brook (in contrast to Blackwater) could limit distinction between fields with permanent versus temporary grass cover. It is also possible that NSRI soil parameters in Kit Brook are less accurate (i.e., data extrapolation for arable soil) as cultivated soils in this catchment have been reported to be more degraded than soils in Blackwater catchment (Palmer, 2007).

Model sediment yield (i.e., $μAbs_{diff}$) showed lower estimation errors in the calibration process (Blackwater) and higher errors in the evaluation process (Kit Brook) compared with other studies (Adams & Elliott, 2006; Elliott et al., 2012; Lukey et al., 2000). In relation to sediment flux peaks, considerable uncertainty has been described in other SHETRAN applications (Elliott et al., 2012; Lukey, Bathurst, et al., 1995; Wicks & Bathurst, 1996) and is a common problem in PBSD models (e.g., Phomcha, Wirojanagud, Vangpaisal, & Thaveevouthti, 2011; Rankinen et al., 2010). Adams and Elliott (2006) explain that this lack of prediction performance in SHETRAN is a consequence of variation in soil cohesive strength, where increases in soil moisture can reduce soil cohesion and increase erodibility. The SHETRAN model does not represent spatial-temporal variability in soil erodibility within overland flow erosion. Seasonal changes in erosion are represented via variability in rainfall intensity in the raindrop/leaf drip

FIGURE 7 (a) Probability density plots of the grid slope for each Blackwater catchment digital elevation model (DEM) grid size; (b) land cover percentage for each Blackwater catchment DEM grid size; and Blackwater stream network (blue line) of each DEM resolution: (c) 25 × 25 m, (d) 50 × 50 m, (e) 100 × 100 m, and (f) 200 × 200 m
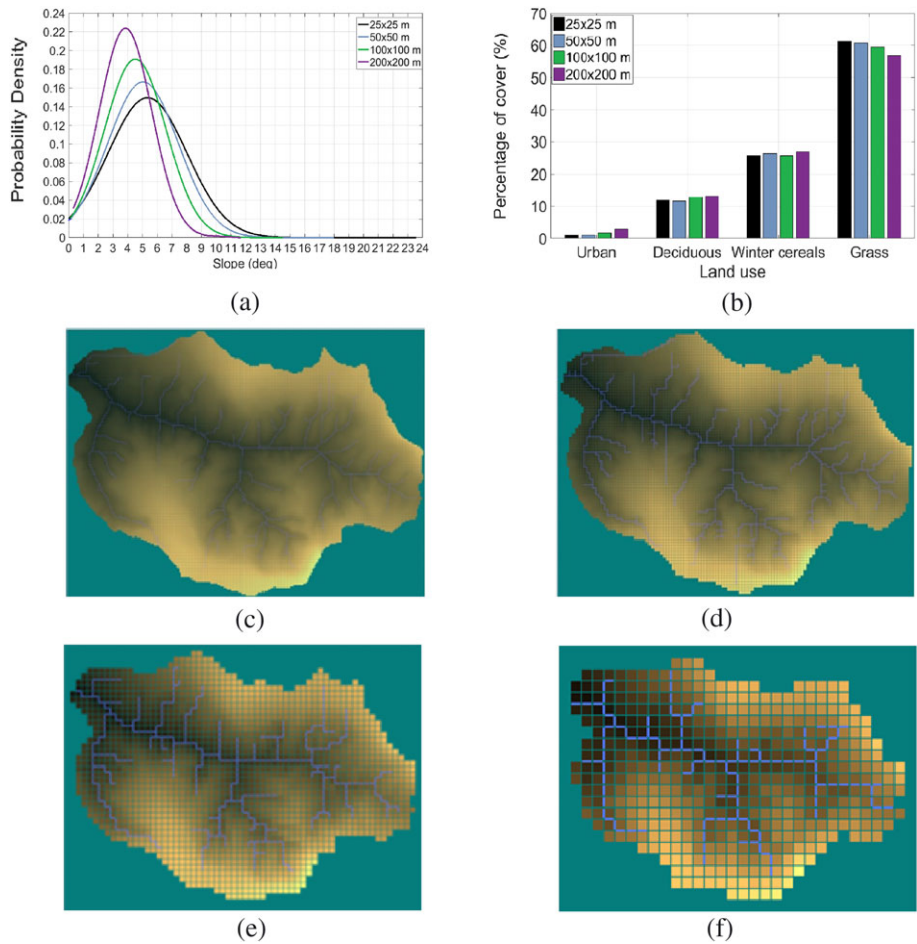
FIGURE 8 Percentage of changes between base run (50 × 50 m) and each DEM grid resolution simulation; (a) flow volume (Qv), mean of flow peaks (μQp) and mean of event flow duration (μQw), and (b) sediment yield (Sy), mean of sediment flux peaks (μSp), and mean of event sediment duration (μSw)
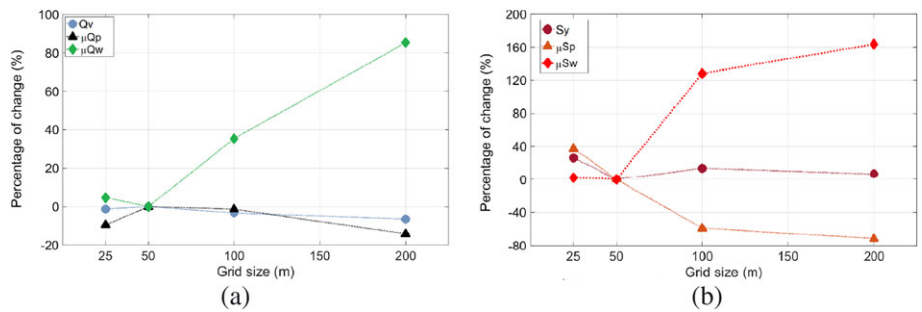
TABLE 4 Evaluation coefficients for each digital elevation model grid size simulation

| Grid size (m) | NSE_Q | μAbs_{diff} (%) (R^2) | | | |
		Qv	Qp	Sy	Sp
25 × 25	0.66	38 (0.5)	49 (0.8)	105 (0.5)	86 (0.6)
50 × 50 (base run)	0.65	37 (0.5)	49 (0.8)	88 (0.4)	81 (0.4)
100 × 100	0.61	46 (0.5)	42 (0.7)	105 (0.4)	64 (0.5)
200 × 200	0.49	58 (0.4)	42 (0.6)	101 (0.6)	61 (0.6)

impact equation and by temporal changes in the proportion of ground shielded by vegetation canopy (Lukey et al., 1995).

Suspended sediment concentration (C)–discharge (Q) hysteresis behaviour was present in the majority of measured events and could influence SHETRAN sediment load predictions as the model predicts arrival of both fluxes at the same time (Adams et al., 2012; Elliott et al., 2012; Sheikh et al., 2010), which might be expected based on the sediment routing equation. It is possible that if major run-off

sources occur in a more distant location from the channel network than sediment sources, then some hysteresis effects could be observed between simulated flow and sediments. This difference in run-off and sediment source locations could in part be related to crop rotation, which was not represented in the calibration, as the 2010 land cover map was used for the 5 years of simulation (2009–2014). Moreover, urban areas (e.g., village and roads) can act as sources of flow and sediment transport (Jones, Swanson, Wemple, & Snyder, 2000), and parameters representing the hydraulic properties of these urban areas are not captured in the NSRI dataset. Furthermore, it was not possible to represent roads at the catchment scale due to the model-limited minimum grid size (>25 m). The above factors could explain some C-Q hysteresis behaviour that was observed in the measured data but not reproduced by SHETRAN.

The parameters to which the model is most sensitive on the water flow component were saturated hydraulic conductivity (K_{sat}) and the Strickler coefficient (Stk; Table 3). These findings were consistent with other studies (Bathurst, 1986; Bathurst et al., 2006). The sediment transport component showed similar SI values between the erodibility coefficients (K_f and K_r). This similarity in sensitivity means that increases in either erodibility coefficient affect sediment output and sediment peaks to a similar extent. It is important to note that SHETRAN simulated a sheet surface flow as wide as the grid size (i.e., 50 × 50 m). Therefore, calibration of K_f was prioritized over K_r. This is consistent with the reported increase in K_f values (kg m^{-2} s^{-1}) with decreasing grid size (Bathurst et al., 2005; Bathurst et al., 2007; Janes et al., 2017; Lukey et al., 2000). The studies also showed increases in raindrop/leaf erodibility coefficient (K_r) values with decreases in grid size but with a lower ratio of change.

The multiparameter sensitivity analysis of the SHETRAN water flow component (Figure 6a) showed that different parameter combinations, particularly those including Stk and K_{sat}, can produce similar changes in flow volume and discharge peaks. Anderton et al. (2002) followed a comparable approach, although parameters were limited to the infiltration module on the water flow component (i.e., K_{sat} and Van Genuchten α and n). The authors found similar modelled flow response to some combinations of K_{sat} and soil depth (0–0.2 and 0.2–3.3 m). Furthermore, an automatic calibration procedure in Zhang et al. (2013) showed an equifinality problem with the adjustment of Stk, K_{sat}, and soil depth. Likewise, combinations of erodibility coefficients (K_r and K_f) produced similar changes in sediment yield and sediment peaks (Figure 6b). The method did not consider feedback from the water flow component to the sediment transport component as this required an excessively large number of simulations (i.e., 243). However, this interaction may represent an important effect in the prediction of overland flow erosion with surface run-off.

The results of the multiparameter sensitivity analysis demonstrate how different combinations of Stk-K_{sat} can produce similar changes in simulated water yield and event flow peaks. This occurred with some parameters having very high values compared with calibrated values (e.g., F^{10} K_{sat} = 0.5 m day^{-1}). Nonetheless, comparable measured values (i.e., 0.3 m day^{-1}) have been reported at field scale (Chappell &

Franks, 1996) in the nearby Slapton Wood catchment. Similarly, K_r-K_f combinations can provide comparable changes in sediment yield and in the mean of sediment flux peaks without necessarily selecting the appropriate parameter value.

Testing sensitivity to grid resolution showed that decreasing grid size increases the variance in the slope angle distribution (Figure 7a) leading to increases in simulated event peaks (μQp and μSp) and a decrease in event duration (μQw and μSw) with only slight variations in flow volume (Qv; Figure 8a). More complex behaviour was observed for Sy (Figure 8b), which could be explained by differences in the channel network produced by the contrasting DEM grid resolutions (Figure 7c–f). The differences between base run (50 × 50 m) coefficients in Tables 2 and 4 relate to the use of the first 2 years (2010–2012) only for DEM simulations. The model performed better in wet than dry periods, and during the first year, rainfall (694 mm) was 68% of the annual average total during the complete measurement period (1,021 mm). Compared with measured data (2010–2012), the simulation with the best performance was 50 × 50 m (Table 4); this was expected as it was calibrated. Nonetheless, similar model evaluation coefficients were observed between the 25 × 25- and 50 × 50 m simulations. It is possible that a better event prediction could be obtained by using the 25 × 25 m resolution for the calibration process but this represents an important compromise between model run times and grid size.

5 | CONCLUSION

SHETRAN showed reasonable performance in Blackwater and Kit Brook catchments by producing an accurate flow prediction for the 4-year period. Moreover, SHETRAN simulations of discharge and sediment flux performed well in terms of event timing in both catchments. Larger errors in event-scale estimation for both discharge and sediment yield were observed in Kit Brook than Blackwater. Adjustment of K_{sat} for the lowermost subsoil–bedrock layer improved model event-scale accuracy in Kit Brook. Nonetheless, prediction of sediment yield and sediment flux peaks remained overestimated. The proxy-catchment test demonstrated that SHETRAN can predict event-scale flow with reasonable accuracy but requires catchment-specific calibration for sediment flux prediction.

The SI showed that flow volume (Qv) was most sensitive to saturated hydraulic conductivity (K_{sat}) and flow peaks (μQp) to the Strickler coefficient (Stk), whereas sediment yield (Sy) was slightly more sensitive to the overland flow erodibility coefficient (K_f) and sediment flux peaks (Sp) to the raindrop/leaf soil erodibility coefficient (K_r). The multiparameter sensitivity analysis showed that potential equifinality behaviour occurring with K_{sat} increases in combination with Stk decreases and vice versa. Similar effects were observed for combinations of K_r and K_f.

The model showed better performance for the base run (50 × 50 m) than the other grid size simulations (range 25- to 200 m grid size) with small differences between the base run and the

finest resolution (25 × 25 m). The DEM grid size variations showed the largest effects in event peaks and duration for both flow and sediment flux, with least variability in predicted event sediment yields.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Veronica Escobar-Ruiz https://orcid.org/0000-0003-1336-0921
Hugh G. Smith https://orcid.org/0000-0003-3049-9810
William H. Blake https://orcid.org/0000-0001-9447-1361
Neil Macdonald https://orcid.org/0000-0003-0350-7096

REFERENCES

Adams, R., & Elliott, S. (2006). Physically based modelling of sediment generation and transport under a large rainfall simulator. *Hydrological Processes*, *20*(11), 2253–2270. https://doi.org/10.1002/hyp.6050

Adams, R., Parkin, G., Rutherford, J. C., Ibbitt, R. P., & Elliott, A. H. (2005). Using a rainfall simulator and a physically based hydrological model to investigate runoff processes in a hillslope. *Hydrological Processes*, *19*(11), 2209–2223. https://doi.org/10.1002/hyp.5670

Adams, R., Western, A. W., & Seed, A. W. (2012). An analysis of the impact of spatial variability in rainfall on runoff and sediment predictions from a distributed model. *Hydrological Processes*, *26*(21), 3263–3280. https://doi.org/10.1002/hyp.8435

Aksoy, H., & Kavvas, M. L. (2005). A review of hillslope and watershed scale erosion and sediment transport models. *Catena*, *64*(2–3), 247–271. https://doi.org/10.1016/j.catena.2005.08.008

Anderton, S., Latron, J., & Gallart, F. (2002). Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model. *Hydrological Processes*, *16*(2), 333–353. https://doi.org/10.1002/hyp.336

Arsenault, R., Brissette, F., & Martel, J. L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology*, *566*, 346–362. https://doi.org/10.1016/j.jhydrol.2018.09.027

Bathurst, J. C. (1986). Sensitivity analysis of the Systeme Hydrologique Europeen for an upland catchment. *Journal of Hydrology*, *87*, 103–123. https://doi.org/10.1016/0022-1694(86)90117-4

Bathurst, J. C. (2011). Predicting impacts of land use and climate change on erosion and sediment yield in river basins using SHETRAN. In R. P. C. Morgan, & M. A. Nearing (Eds.), *Handbook of erosion modelling* (pp. 263–288). Oxford: Blackwell Publishing Ltd. DOI: 10.1002/9781444328455.ch14

Bathurst, J. C., Burton, A., Clarke, B. G., & Gallart, F. (2006). Application of the SHETRAN basin-scale, landslide sediment yield model to the Llobregat basin, Spanish Pyrenees. *Hydrological Processes*, *20*(14), 3119–3138. https://doi.org/10.1002/hyp.6151

Bathurst, J. C., Ewen, J., Parkin, G., O'Connell, P. E., & Cooper, J. D. (2004). Validation of catchment models for predicting land-use and climate change impacts. 3. Blind validation for internal and outlet responses. *Journal of Hydrology*, *287*, 74–94. https://doi.org/10.1016/j.jhydrol.2003.09.021

Bathurst, J. C., Moretti, G., El-Hames, A., Beguería, S., & García-Ruiz, J. M. (2007). Modelling the impact of forest loss on shallow landslide sediment yield, Ijuez river catchment, Spanish Pyrenees. *Hydrology and Earth System Sciences*, *11*(1), 569–583. https://doi.org/10.5194/hess-11-569-2007

Bathurst, J. C., Moretti, G., El-Hames, A., Moaven-Hashemi, A., & Burton, A. (2005). Scenario modelling of basin-scale, shallow landslide sediment yield, Valsassina, Italian Southern Alps. *Natural Hazards and Earth System Sciences*, *5*(2), 189–202. https://doi.org/10.5194/nhess-5-189-2005

Beven, K. (1991). Scale Considerations. In D. S. Bowles, & P. E. O'Connell (Eds.), *Recent advances in the modelling of hydrologic systems* (pp. 357–371). Netherlands: Springer Netherlands. DOI: 10.1007/978-94-011-3480-4_16

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

Birkinshaw, S. J. (2008). Physically-based modelling of double-peak discharge responses at Slapton Wood catchment. *Hydrological Processes*, *22*(10), 1419–1430. https://doi.org/10.1002/hyp.6694

Birkinshaw, S. J. (2010). Technical note: Automatic river network generation for a physically-based river catchment model. *Hydrology and Earth System Sciences*, *14*(9), 1767–1771. https://doi.org/10.5194/hess-14-1767-2010

Birkinshaw, S. J. (2016). Producing a good hydrological simulation in SHETRAN. Newcastle: University of Newcastle upon Tyme. Retrieved from http://research.ncl.ac.uk/shetran/Producing a good hydrological simulation.pdf

Birkinshaw, S. J., & Ewen, J. (2000). Modelling nitrate transport in the Slapton Wood catchment using SHETRAN. *Journal of Hydrology*, *230*(1–2), 18–33. https://doi.org/10.1016/S0022-1694(00)00173-6

Brazier, R., Anderson, K., Bellamy, P., Ellis, M., Evans, M., Quinton, J., ... Rickson, J. (2011). *A review of soil erosion monitoring and modelling in England and Wales*. Exeter: DEFRA.

Bussi, G., Francés, F., Montoya, J. J., & Julien, P. Y. (2014). Distributed sediment yield modelling: Importance of initial sediment conditions. *Environmental Modelling & Software*, *58*, 58–70. https://doi.org/10.1016/j.envsoft.2014.04.010

Centre for Ecology and Hydrology (2018). National river flow archive. Retrieved March 20, 2016, from https://nrfa.ceh.ac.uk

Chappell, N. A., & Franks, S. W. (1996). Property distributions and flow structure in the Slapton Wood catchment. *Field Studies*, *8*, 559–575.

Cranfield University (2018). The soils guide. Retrieved May 20, 2015, from http://www.landis.org.uk/services/soilsguide/mapunit_list.cfm?sorttype_association=map_unit_name

Croke, B. F. W., Merritt, W. S., & Jakeman, A. J. (2004). A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments. *Journal of Hydrology*, *291*(1–2), 115–131. https://doi.org/10.1016/j.jhydrol.2003.12.012

Croke, J., Mockler, S., Fogarty, P., & Takken, I. (2005). Sediment concentration changes in runoff pathways from a forest road network and the resultant spatial pattern of catchment connectivity. *Geomorphology*, *68*(3–4), 257–268. https://doi.org/10.1016/j.geomorph.2004.11.020

Daniel, E. B., Camp, J. V., Leboeuf, E. J., Penrod, J. R., Dobbins, J. P., & Abkowitz, M. D. (2011). Watershed modeling and its applications: A state-of-the-art review. *The Open Hydrology Journal*, *5*, 26–50. https://doi.org/10.2174/1874378101105010026

De Figueiredo, E. E., & Bathurst, J. C. (2007). Runoff and sediment yield predictions in a semiarid region of Brazil using SHETRAN. In *Predictions in ungauged basins: PUB kick-off* (pp. 258–266). Brasilia: IAHS Publ.

De Roo, A. P. J., & Jetten, V. G. (1999). Calibrating and validating the LISEM model for two data sets from The Netherlands and South Africa. *Catena*, *37*(3–4), 477–493. https://doi.org/10.1016/S0341-8162(99)00034-X

Deasy, C., Brazier, R. E., Heathwaite, A. L., & Hodgkinson, R. (2009). Pathways of runoff and sediment transfer in small agricultural catchments. *Hydrological Processes*, *23*(9), 1349–1358. https://doi.org/10.1002/hyp.7257

Deasy, C., Titman, A., & Quinton, J. N. (2014). Measurement of flood peak effects as a result of soil and land management, with focus on experimental issues and scale. *Journal of Environmental Management*, *132*, 304–312. https://doi.org/10.1016/j.jenvman.2013.11.027

Department for Environment, Food and Rural Affairs (2015). *DEFRA statistics: Agricultural facts—Commercial holdings at June 2015. South West*. London: DEFRA. https://doi.org/10.1063/1.4704926.

Elliot, W. J. (2004). WEPP internet interfaces for forest erosion prediction. *Journal of the American Water Resources Association*, *40*(2), 299–309. https://doi.org/10.1111/j.1752-1688.2004.tb01030.x

Elliott, A. H., Oehler, F., Schmidt, J., & Ekanayake, J. C. (2012). Sediment modelling with fine temporal and spatial resolution for a hilly catchment. *Hydrological Processes*, *26*(24), 3645–3660. https://doi.org/10.1002/hyp.8445

Engman, E. T., & American Society of Civil Engineers, M (1986). Roughness coefficients for routing surface runoff. *Journal of Irrigation and Drainage Engineering*, *112*, 39–53.

Ewen, J., & Parkin, G. (1996). Validation of catchment models for predicting land-use and climate change impacts. 1. Method. *Journal of Hydrology*, *175*, 583–594. https://doi.org/10.1016/S0022-1694(96)80026-6

Ewen, J., Parkin, G., & O'Connel, P. E. (2000). SHETRAN: Distributed river basin flow modeling system. *Journal of Hydrologic Engineering*, *5*(3), 250–258. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:3(250)

Froehlich, H. A., Miles, D. W. R., & Robbins, R. W. (1985). Soil bulk density recovery on compacted skid trails in Central Idaho 1. *Soil Science Society of America Journal*, *49*(4), 1015. https://doi.org/10.2136/sssaj1985.03615995004900040045x

Gumindoga, W., Rwasoka, D. T., Nhapi, I., & Dube, T. (2017). Ungauged runoff simulation in Upper Manyame Catchment, Zimbabwe: Application of the HEC-HMS model. *Physics and Chemistry of the Earth*, *100*, 371–382. https://doi.org/10.1016/j.pce.2016.05.002

Hollis, J. M., Lilly, A., Higgins, A., Jones, R. J., Keay, C. A., & Bellamy, P. (2015). Predicting the water retention characteristics of UK mineral soils. *European Journal of Soil Science*, *66*, 239–252. https://doi.org/10.1111/ejss.12186

International Union of Soil Sciences (2007). World Reference Base for Soil Resources 2006, first update 2007. World soil resources reports no. 103. FAO, Rome. *World Soil Resources Reports*. https://doi.org/10.1017/S0014479706394902

Janes, V., Holman, I., Birkinshaw, S., O'Donnell, G., & Kilsby, C. (2017). Improving bank erosion modelling at catchment scale by incorporating temporal and spatial variability. *Earth Surface Processes and Landforms*, *43*(1), 124–133. https://doi.org/10.1002/esp.4149

Jetten, V. G., & Maneta, M. P. (2011). Calibration of erosion models. In *Handbook of erosion modelling* (pp. 33–51). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444328455.ch3

Jones, J. A., Swanson, F. J., Wemple, B. C., & Snyder, K. U. (2000). Effects of roads on hydrology, geomorphology, and disturbance patches in stream networks. *Conservation Biology*, *14*(1), 76–85. https://doi.org/10.1046/j.1523-1739.2000.99083.x

Jordán-López, A., Martínez-Zavala, L., & Bellinfante, N. (2009). Impact of different parts of unpaved forest roads on runoff and sediment yield in a Mediterranean area. *Science of the Total Environment*, *407*(2), 937–944. https://doi.org/10.1016/j.scitotenv.2008.09.047

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13–24. https://doi.org/10.1080/02626668609491024

Kværnø, S. H., & Stolte, J. (2012). Effects of soil physical data sources on discharge and soil loss simulated by the LISEM model. *Catena*, *97*, 137–149. https://doi.org/10.1016/j.catena.2012.05.001

Lesschen, J. P., Schoorl, J. M., & Cammeraat, L. H. (2009). Modelling runoff and erosion for a semi-arid catchment using a multi-scale approach based on hydrological connectivity. *Geomorphology*, *109*(3–4), 174–183. https://doi.org/10.1016/j.geomorph.2009.02.030

Little, R. (2012). *Sediment and water delivery to a second order agricultural stream following catchment restoration and land use management scenarios*. Plymouth University. https://doi.org/http://dx.doi.org/10.1111/j.1469-0691.2011.03557.x

Lukey, B. T., Bathurst, J. C., Hiley, R. A., & Ewen, J. (1995). *SHETRAN sediment transport component: Equations and algorithms*. Newcastle upon Tyne.

Lukey, B. T., Sheffield, J., Bathurst, J. C., Hiley, R. A., & Mathys, N. (2000). Test of the SHETRAN technology for modelling the impact of reforestation on badlands runoff and sediment yield at Draix, France. *Journal of Hydrology*, *235*(1–2), 44–62. https://doi.org/10.1016/S0022-1694(00)00260-2

Lukey, B. T., Sheffield, J., Bathurst, J. C., Lavabre, J., Mathys, N., & Martin, C. (1995). Simulating the effect of vegetation cover on the sediment yield of Mediterranean catchments using SHETRAN. *Physics and Chemistry of the Earth*, *20*(3–4), 427–432. https://doi.org/10.1016/0079-1946(95)00056-9

Marsik, M., & Waylen, P. (2006). An application of the distributed hydrologic model CASC2D to a tropical montane watershed. *Journal of Hydrology*, *330*(3–4), 481–495. https://doi.org/10.1016/j.jhydrol.2006.04.003

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling and Software*, *18*, 761–799. https://doi.org/10.1016/S1364-8152(03)00078-1

Monteith, J. L. (1965). Evaporation and environment. *Symposia of the Society for Experimental Biology*, *19*, 205–234. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5321565

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed siulations. *American Society of Agricultural and Biological Engineers*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Mourato, S., Moreira, M., & Corte-Real, J. (2015). Water resources impact assessment under climate change scenarios in Mediterranean watersheds. *Water Resources Management*, *29*(7), 2377–2391. https://doi.org/10.1007/s11269-015-0947-5

Op de Hipt, F., Diekkrüger, B., Steup, G., Yira, Y., Hoffmann, T., & Rode, M. (2017). Applying SHETRAN in a tropical West African catchment (Dano, Burkina Faso)—Calibration, validation, uncertainty assessment. *Water*, *9*(2), 101. https://doi.org/10.3390/w9020101

Oudin, L., Michel, C., & Anctil, F. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology*, *303*(1–4), 275–289. https://doi.org/10.1016/j.jhydrol.2004.08.025

Palmer, R. C. (2007). A comparison of soil structural conditions in the Axe and Char catchments during late-winter 2004 and 2007. York.

Parkin, G., O'Donnell, G., Ewen, J., Bathurst, J. C., O'Connell, P. E., & Lavabre, J. (1996). Validation of catchment models for predicting land-use and climate change impacts. 2. Case study for a Mediterranean catchment. *Journal of Hydrology*, *175*(1–4), 595–613. https://doi.org/10.1016/S0022-1694(96)80027-8

Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., & Wheater, H. S. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal*, *13*(3), 193–214. https://doi.org/10.1002/hyp

Penman, H. L. (1948). Natural evapotranspiration from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, *193*(1032), 120–145. https://doi.org/10.1007/s13398-014-0173-7.2

Phomcha, P., Wirojanagud, P., Vangpaisal, T., & Thaveevouthti, T. (2011). Predicting sediment discharge in an agricultural watershed: A case

study of the Lam Sonthi watershed, Thailand. *ScienceAsia*, *37*(1), 43. https://doi.org/10.2306/scienceasia1513-1874.2011.37.043

Rabitz, H. (1989). Systems analysis at the molecular scale. *Science*, *246*(4927), 221–226. https://doi.org/10.1126/science.246.4927.221

Rahimy, P. (2011). *Effects of soil depth and saturated hydraulic conductivity spatial variation on runoff simulation by the Limburg soil erosion model (LISEM), a case study in Faucon catchment, France.* University of Twente. Retrieved from https://www.itc.nl/library/papers_2011/msc/aes/rahimy.pdf

Rankinen, K., Thouvenot-Korppoo, M., Lazar, A., Lawrence, D. S. L., Butterfield, D., Veijalainen, N., … Lepistö, A. (2010). Application of catchment scale sediment delivery model INCA-Sed to four small study catchments in Finland. *Catena*, *83*(1), 64–75. https://doi.org/10.1016/j.catena.2010.07.005

Refsgaard, J. C., Storm, B., & Refsgaard, A. (1995). Validation and applicability of distributed hydrological models. In *Modelling and management of sustainable basin-scale water resource systems (Proceedings of a Boulder Symposium, July 1995)* (pp. 387–397). IAHS Publ. https://doi.org/10.4103/0972-5229.160263

Robson, A. J., & Reed, D. W. (2008). Statistical procedures for flood frequency estimation. In *Centre for Ecology & Hydrology, flood estimation handbook: Vol. 3* (pp. 367–380). Wallingford: Centre for Ecology & Hydrology.

Rutter, A. J., Morton, A. J., & Robins, P. C. (1975). A predictive model of rainfall interception in forests. II. Generalization of the model and comparison with observations in some Coniferous and Hardwood stands. *The Journal of Applied Ecology*, *12*(1), 367–380. https://doi.org/10.2307/2401739

Schob, A., Schmidt, J., & Tenholtern, R. (2006). Derivation of site-related measures to minimise soil erosion on the watershed scale in the Saxonian loess belt using the model EROSION 3D. *Catena*, *68*(2–3), 153–160. https://doi.org/10.1016/j.catena.2006.04.009

Sheikh, V., van Loon, E., Hessel, R., & Jetten, V. (2010). Sensitivity of LISEM predicted catchment discharge to initial soil moisture content of soil profile. *Journal of Hydrology*, *393*(3–4), 174–185. https://doi.org/10.1016/j.jhydrol.2010.08.016

Smith, H. G., Peñuela, A., Sangster, H., Sellami, H., Boyle, J., Chiverrell, R., … Riley, M. (2018). Simulating a century of soil erosion for agricultural catchment management. *Earth Surface Processes and Landforms*, *43*(10), 2089–2105. https://doi.org/10.1002/esp.4375

Tarawneh, E., Bridge, J., & Macdonald, N. (2016). A pre-calibration approach to select optimum inputs for hydrological models in data-scarce regions. *Hydrology and Earth System Sciences*, *20*, 4391–4407. https://doi.org/10.5194/hess-20-4391-2016

Thomas, I. A., Jordan, P., Mellander, P. E., Fenton, O., Shine, O., Ó hUallacháin, D., … Murphy, P. N. C. (2016). Improving the identification of hydrologically sensitive areas using LiDAR DEMs for the delineation and mitigation of critical source areas of diffuse pollution. *Science of the Total Environment*, *556*, 276–290. https://doi.org/10.1016/j.scitotenv.2016.02.183

Thomas, I. A., Jordan, P., Shine, O., Fenton, O., Mellander, P. E., Dunlop, P., & Murphy, P. N. C. (2017). Defining optimal DEM resolutions and point densities for modelling hydrologically sensitive areas in agricultural catchments dominated by microtopography. *International Journal of Applied Earth Observation and Geoinformation*, *54*, 38–52. https://doi.org/10.1016/j.jag.2016.08.012

Thompson, J. A., Bell, J. C., & Butler, C. A. (2001). Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma*, *100*(1–2), 67–89. https://doi.org/10.1016/S0016-7061(00)00081-1

Thorud, D. B., & Frissell, S. S. (1976). Time changes in soil density following compaction under an Oak Forest. Retrieved from http://hdl.handle.net/11299/58319.

Tiemeyer, B., Moussa, R., Lennartz, B., & Voltz, M. (2007). MHYDAS-DRAIN: A spatially distributed model for small, artificially drained lowland catchments. *Ecological Modelling*, *209*(1), 2–20. https://doi.org/10.1016/j.ecolmodel.2007.07.003

Tuzzin de Moraes, M., Debiasi, H., Carlesso, R., Cezar Franchini, J., Rodrigues da Silva, V., & Bonini da Luz, F. (2016). Soil physical quality on tillage and cropping systems after two decades in the subtropical region of Brazil. *Soil and Tillage Research*, *155*, 351–362. https://doi.org/10.1016/j.still.2015.07.015

Valeo, C., & Moin, S. M. A. (2000). Grid-resolution effects on a model for integrating urban and rural areas. *Hydrological Processes*, *14*, 2505–2525. https://doi.org/10.1002/1099-1085(20001015)14:14<2505::AID-HYP111>3.0.CO;2-3

Verhaegen, T. (1987). The use of small flumes for the determination of soil erodibility. *Earth Surface Processes and Landforms*, *12*(2), 185–194. https://doi.org/10.1002/esp.3290120208

Wicks, J. M., & Bathurst, J. C. (1996). SHESED: A physically based, distributed erosion and sediment yield component for the SHE hydrological modelling system. *Journal of Hydrology*, *175*(1–4), 213–238. https://doi.org/10.1016/S0022-1694(96)80012-6

Wicks, J. M., Bathurst, J. C., & Johnson, C. W. (1992). Calibrating SHE soil-erosion model for different land covers. *Journal of Irrigation and Drainage Engineering*, *118*, 708–723. https://doi.org/10.1061/(ASCE)0733-9437(1992)118:5(708), DOI: https://doi.org/10.1061/(ASCE)0733-9437(1992)118:5(708)

Wicks, J. M., Bathurst, J. C., Johnson, C. W., & Ward, T. J. (1988). Application of two physically-based sediment yield models at plot and field scales. *Sediment Budgets*, *174*, 583–591. Retrieved from http://hydrologie.org/redbooks/a174/iahs_174_0583.pdf

Williams, G. P. (1989). Sediment concentration versus water discharge during single hydrologic events in rivers. *Journal of Hydrology*, *111*(1–4), 89–106. https://doi.org/10.1016/0022-1694(89)90254-0

Xu, C.-Y., & Singh, V. P. (2004). Review on regional water resources assessment models under stationary and changing climate. *Water Resources Management*, *18*, 591–612. https://doi.org/10.1007/s11269-004-9130-0

Yang, D., Herath, S., & Musiake, K. (2002). A hillslope-based hydrological model using catchment area and width functions. *Hydrological Sciences Journal*, *47*(1), 49–65. https://doi.org/10.1080/02626660209492907

Zhang, R., Santos, C. A. G., Moreira, M., Freire, P. K. M. M., & Corte-Real, J. (2013). Automatic calibration of the SHETRAN hydrological modelling system using MSCE. *Water Resources Management*, *27*(11), 4053–4068. https://doi.org/10.1007/s11269-013-0395-z

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

APPENDIX A

BASELINE SIMULATION PARAMETER VALUES

Water flow component				
Parameter				**Value**
Str[a] ($m^{1/3}$ s^{-1})				1.0
$CStr$[b] ($m^{1/3}$ s^{-1})				20
	Urban	**Deciduous**	**Arable**	**Grass**
$CLAI$[c]	1.0	4.0	1.0	1.6
$PLAI$[d]	0.3	1.0	1.0	1.0
Sediment transport component				
Parameter	**Value**			
	Urban	**Deciduous**	**Arable**	**Grass**
K_f[e] ($\times 10^{-8}$ kg m^{-2} s^{-1})	65	65	65	65
K_r[f] (J^{-1})	1.3	1.3	1.3	1.3
F_{CG}[g]	0.3	0.3	0.3	0.3
F_{CROCK}[h]	0.0	0.0	0.0	0.0

[a]Overland Strickler coefficient (inverse of Manning number).
[b]Channel Strickler coefficient.
[c]Ratio of total leaf area to area of ground covered by vegetation.
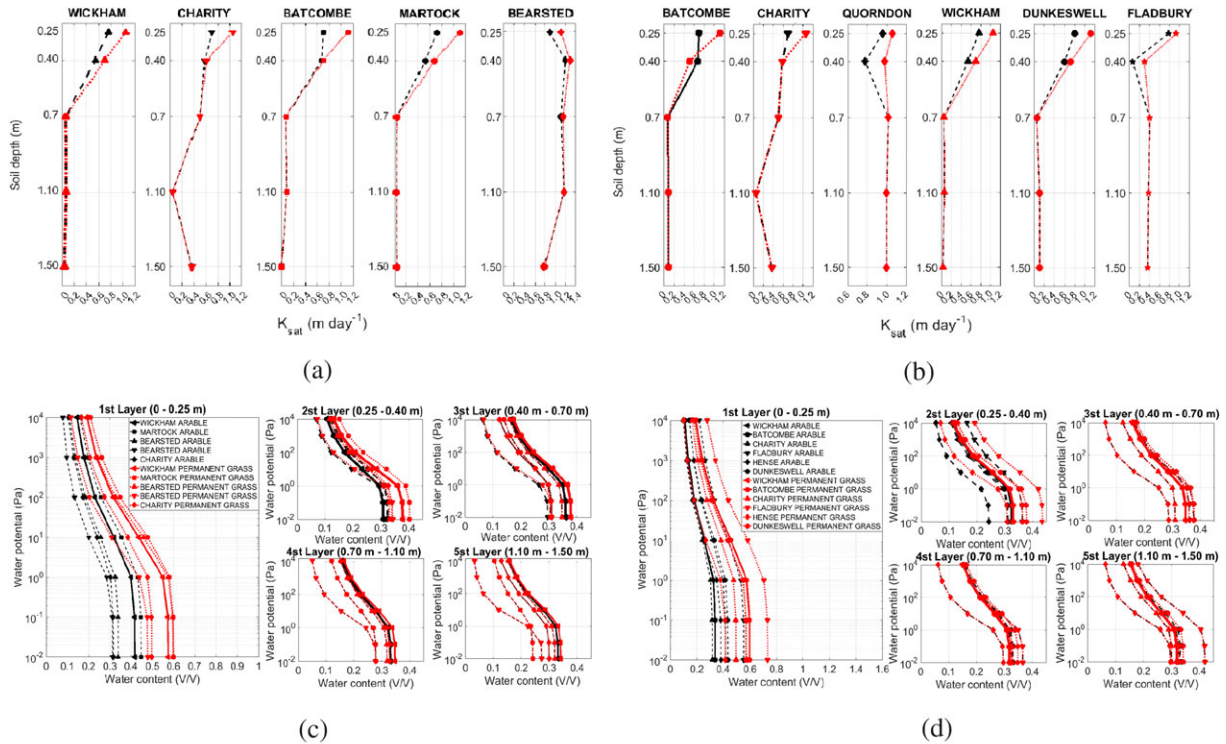[d]Proportion shielded by ground-level cover.
[e]Overland flow erodibility coefficient.
[f]Raindrop/leaf erodibility coefficient.
[g]Proportion of ground shielded by near-ground cover.
[h]Proportion shielded by ground-level cover.

APPENDIX B

Soil data © Cranfield University and for the Controller of HMSO 2019 available at www.landis.org.uk

FIGURE B1 Saturated hydraulic conductivity K_{sat} at different soil depth of each soil type (arable crops on black and permanent grass on red), (a) Blackwater and (b) Kit Brook, and water retention curve at different soil depth of each soil type at National Soil Resources Institute dataset, (c) Blackwater and (d) Kit Brook

APPENDIX C
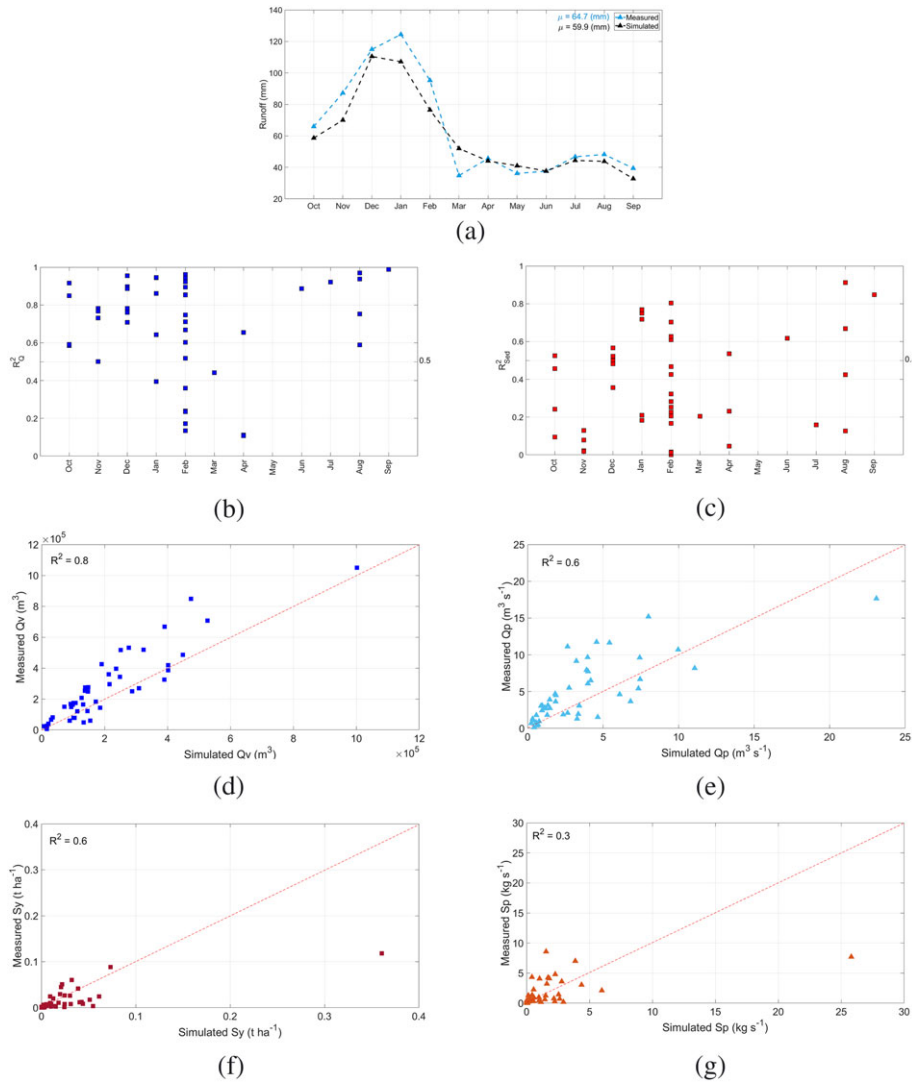
FIGURE C1 Kit Brook results following the K_{sat} adjustment for the sixth soil layer. (a) Measured and simulated monthly mean runoff (b) coefficient of determination (R^2Q) between the measured and simulated discharge for each event in a monthly window, (c) coefficient of determination (R^2Sed) between the measured and simulated sediment flux for each event in a monthly window and plots of simulated versus measured events for (d) flow volume (Qv), (e) peak flow (Qp), (f) sediment yield (Sy), and (g) peak sediment flux (Sp) for Kit Brook catchment