Sensitivity Analysis of the Parameter-Efficient Distributed (PED) Model for Discharge and Sediment Concentration Estimation in Degraded Humid Landscapes

Authors:

Boris F. Ochoa-Tocachi[a,b,c], boris.ochoa13@imperial.ac.uk*

Tilashwork Alemie[a,d], t.alemie14@imperial.ac.uk

Christian D. Guzman[c], christian.d.guzman@wsu.edu

Seifu A. Tilahun[d], satadm86@gmail.com

Fasikaw A. Zimale[d,f], fasikaw@gmail.com

Wouter Buytaert[a,b,c], w.buytaert@imperial.ac.uk

Tammo S. Steenhuis[d,f], tss1@cornell.edu

Affiliations:

[a] Department of Civil and Environmental Engineering, Imperial College London, South Kensington Campus, London, UK

[b] Grantham Institute – Climate Change and the Environment, Imperial College London, South Kensington Campus, London, UK

[c] Regional Initiative for Hydrological Monitoring of Andean Ecosystems, iMHEA, Quito, EC

[d] Faculty of Civil and Water Resources Engineering, Bahir Dar University, Bahir Dar, ET

[e] Department of Civil and Environmental Engineering, Washington State University, WA, US

[f] Department of Biological and Environmental Engineering, Cornell University, NY, US

* *Corresponding Author: Boris F. Ochoa-Tocachi, Department of Civil and Environmental Engineering & Grantham Institute – Climate Change and the Environment, Imperial College London, Skempton Building – Room 411, South Kensington Campus, London SW7 2AZ, UK, Tel.: +44 (0) 7463263232, e-mail: boris.ochoa13@imperial.ac.uk.*

Running head: Ochoa-Tocachi et al.: Sensitivity Analysis of the Rainfall-Runoff-Erosion PED Model

Keywords: global sensitivity analysis, erosion model, rainfall-runoff model, sediment modelling, multi-method GSA, SAFE Toolbox, PAWN, Ethiopia

ABSTRACT

Sustainable development in degraded landscapes in the humid tropics require effective soil and water management practices. Coupled hydrological-erosion models have been used to understand and predict the underlying processes at watershed scale and the effect of human interventions. One prominent tool is the parameter-efficient distributed (PED) model, which improves on other models by considering a saturation-excess runoff generation driving erosion and sediment transport in humid climates. This model has been widely applied at different scales for the humid monsoonal climate of the Ethiopian Highlands, with good success in estimating discharge and sediment concentrations. However, previous studies performed manual calibration of the involved parameters without reporting sensitivity analyses or assessing equifinality. The aim of this article is to provide a multi-objective global sensitivity analysis of the PED model using automatic random sampling implemented in the SAFE Toolbox. We find that relative parameter sensitivity depends greatly on the purpose of model application and the outcomes used for its evaluation. Five of the 13 PED model parameters are insensitive for improving model performance. Additionally, associating behavioural parameter values with a clear physical meaning provides slightly better results and helps interpretation. Finally, good performance in one module does not translate directly into good performance in the other module. We interpret these results in terms of the represented hydrological and erosion processes and recommend field data to inform model calibration and validation, potentially improving land degradation understanding and prediction and supporting decision-making for soil and water conservation strategies in degraded humid landscapes.

1. INTRODUCTION

In the Ethiopian Highlands, increasing population and associated rural development have intensified agriculture resulting in land degradation and hardpan formation at shallow depths (Tebebu et al., 2015; Tebebu et al., 2017). As a result, sediment concentrations have increased in rivers that feed important reservoirs (Zimale et al., 2017). To inform decision makers of the best management practices to counteract these effects, several studies have used coupled hydrological-erosion models, with the prominent case of the Parameter-Efficient Distributed (PED) model (e.g., Tesemma et al (2010) Engda et al., 2011; Tilahun et al., 2013a; Tilahun et al., 2013b; Tilahun et al., 2015; Guzman et al., 2017a; Guzman et al., 2017b; Zimale et al., 2017; Zimale et al., 2018). The PED model was conceptualised and developed by Collick et al. (2009), Steenhuis et al. (2009), and Tesemma et al. (2010) to simulate rainfall-runoff processes in the sub-humid, semi-monsoonal climate of the Ethiopian Highlands, and later complemented with an erosion module by Tilahun et al. (2013a) and Tilahun et al. (2013b). The PED model improves over more complex models implemented in the area (e.g., Setegn et al., 2010; Easton et al., 2010) because it represents more accurately the process of saturation excess runoff characteristic of humid climates. In the Ethiopian Highlands, saturation excess runoff occurs in degraded hillslope areas and valley bottoms with a water table close to the land surface (Steenhuis et al., 2013; Tilahun et al., 2013a; Guzman et al., 2017a).

Despite the attractive name, the PED model has up to nine parameters for the rainfall-runoff module and four for the erosion module. This is similar to other conceptual models [e.g. GR4J (Perrin et al., 2003), TOPMODEL (Beven & Kirby, 1979), WEPP (Flanagan &

Nearing, 1995)], but many less than other physically based models [e.g. SWAT (Arnold et al., 2012), EUROSEM (Morgan et al., 1998), DYRIM (Wang et al., 2015; Shi et al., 2016)]. The HBV conceptual model, for instance, may use 15 or more parameters depending upon the application (Bergström & Forsman, 1973). One current limitation, however, is that studies implementing the PED model so far have relied on manual calibration and are very limited in reporting sensitivity or uncertainty analyses. The exception is the manual sensitivity analysis of the PED model in the Supplementary Material of Tilahun et al. (2013a). Based on this material, Guzman et al. (2017b) recognised that four of the nine parameters in the hydrological module are insensitive and highlighted that the relatively few adjustable parameters are one advantage of the model. The remaining parameters can be linked directly to observable processes at catchment scale and are thus less affected by equifinality issues (Beven, 2006). At the same time, Guzman et al. (2017b) recognised that it is difficult to determine input parameters a priori because they are defined at the catchment scale and not at the point scale at which measurements are made. Moreover, a sensitivity analysis of the erosion module parameters has been largely dismissed.

To fill this methodological gap, we present a systematic and comprehensive sensitivity analysis of the PED model using the SAFE Toolbox implemented in MATLAB (Pianosi et al., 2015). This manuscript aims to answer three main questions with respect to model implementation in degraded humid landscapes:

(**Q1**) does good performance in the hydrological module translate directly into good performance in the erosion module and vice versa?

(**Q2**) what are the most influential parameters in providing good performing model simulations?

(**Q3**) what are the most influential parameters in the simulation of low, mean, and peak flows, and of mean and peak sediment concentrations?

The results of this analysis may improve the application of the PED model in the evaluation of the effects of landscape interventions in the hydrology and erosion processes of degraded watersheds, as well as provide a guide to evaluate the implementation of other coupled hydrological-erosion models. Ultimately, this can help inform decision makers more accurately about the implementation of soil and water management practices in degraded humid landscapes.

2. METHODS

2.1 Parameter-Efficient Distributed (PED) Model

The PED model divides a watershed in three distinct lumped landscape units for the simulation of the hydrological response (**Figure 1**): (A_1) lower areas that can saturate during the rainy season, (A_2) degraded hillslopes that saturate promptly after a storm, and (A_3) permeable hillslopes that allow subsurface infiltration. By applying a water balance between precipitation (P), potential evapotranspiration (E_p), and soil moisture (S_{ri}), any excess over maximum soil storage capacity ($S_{max\,i}$) of areas A_1 and A_2 contributes surface runoff (q_{r1} and q_{r2}) directly to streamflow. In the case of the permeable hillsides, any excess is transformed to recharge (q_{r3}) to a baseflow linear reservoir (BS_t) with maximum storage BS_{max} and mean residence time $t_{1/2}$. Excess over BS_{max} percolates (q^*_{r3}) to an interflow zero-order reservoir (IS_t)

with mean duration τ^*. Baseflow (q_b) and interflow (q_i) are generated from their respective subsurface reservoirs and are added to total streamflow.

The sum of the area fractions does not necessarily adds up to the entire catchment, as a fraction of the discharge may leave the watershed in the form of deep regional flow without being accounted for (e.g., Tilahun et al., 2013a; Tilahun et al., 2015; Zimale et al., 2018). In this implementation, when the automatic sampling of the input space (subsection 2.2) yields combinations of area fractions whose sum exceeds 1, values are then normalised. This is equivalent to reducing one parameter in the calibration procedure but will introduce parameter interaction that needs to be accounted for. In addition, in the manual calibration of the model, the initial three root soil moistures and baseflow storage (at *t=0*) need to be defined a priori, which would imply adding 4 extra parameters to the calibration. Generally, daily time step continuous models are calibrated for at least 8 years (Yapo et al., 1996), and the first year is discarded as warmup period. Here, due to the long 8-month dry phase, the moisture in the root profile is always depleted to the same low moisture content at the beginning of the rain phase and independent of the initial values chosen if the first 90 time steps (days) are discarded as a warmup period. Although the baseflow storage may need a longer warmup period, it will affect only the first year of the simulations by the start of the first rainy season.

An erosion module is coupled to the outputs of the hydrological module to calculate sediment concentrations in the stream at daily scale (**Figure 1**). In this module, erosion is generated from the two runoff source areas –the periodically saturated bottom lands (A_1) and the degraded soils on the hillsides (A_2)– by rills that form in the newly plowed lands. Initially,

when the rills are just formed, sediment transport is the highest and the concentration is determined by the transport limit of the water carrying the sediment. Hairsine & Rose (1992a), Hairsine & Rose (1992b), Ciesiolka et al. (1995), Tilahun et al. (2013a), and Tilahun et al. (2013b) have shown that the sediment load at the transport limit is proportional to the unit flux, q, to the *1.4* power (e.g., $a_t q^{1.4}$), where a_t is a coefficient of transport limiting conditions. Once the rills are formed, the sediment concentration is dependent on the delivery of soil from the interrill area and the sediment load can be expressed as $a_s q^{1.4}$, where a_s is a coefficient of source limiting conditions. A function H, which ranges from 0 to 1 and is determined a priori, is used to express the proportion of land in the watershed with active rill formation. Originally, H was developed as a step function by Tilahun et al. (2013b), and then simplified to a curve by Guzman et al. (2017b).

The concentration of sediment, C [g l^{-1}], in the river is obtained by dividing the sediment yield by the total predicted discharge (including baseflow and interflow) from the hydrological module. The final form is the one shown in **Equation 1**:

$$C = \frac{A_1 q_{r1}^{1.4}\left[a_{s1} + H\left(a_{t1} - a_{s1}\right)\right] + A_2 q_{r2}^{1.4}\left[a_{s2} + H\left(a_{t2} - a_{s2}\right)\right]}{A_1 q_{r1} + A_2 q_{r2} + A_3\left(q_b + q_i\right)} \qquad \textbf{(1)}$$

where the four soil-related parameters (a) are coefficients where the subscripts indicate the saturated (*1*) and degraded (*2*) areas, for transport limiting (*t*) and source limiting (*s*) conditions. Feasible ranges for these parameters have been determined in previous modelling exercises performed by Tilahun et al. (2013a), Tilahun et al. (2013b), and Guzman et al. (2017b), where an extended description of the module can be also found.

2.2. Model evaluation and sensitivity analysis

The SAFE Toolbox implements several established Global Sensitivity Analysis (GSA) methods to support the development and assessment of environmental models (Pianosi et al., 2015). The 3 basic steps of GSA in SAFE, with the criteria implemented in this article, are:

1. **Sampling input space:** We created a matrix *X* of *N* randomly sampled input combinations, each made up of *M* parameter components. Here, *M=9* for the hydrological module (**Table I**) and *M=4* for the erosion module (**Table II**). We sampled *N=10000* parameter combinations from uniform distributions within feasible ranges using Latin Hypercube Sampling (McKay et al., 1979). The automatic sampling covers the entire parameter space with different degrees of density depending on the number of samples. The advantage of using a Latin Hypercube Sampling algorithm is to avoid repeating parameter combinations by using a memory of previously sampled parameters when sampling the next set. Since this is a random exercise, although each run of the automatic sampling will provide different specific results, the overall statistics and global performance will remain consistent. Increasing the number of samples beyond 10000 would likely increase the probability of finding a better performing set compared to a smaller number, but this comes at a computing cost that does not correlate optimally to the overall improvement in model performance.

For the erosion module, we tested the premise that a model not only needs to perform well, but it needs to do so for the right reasons. Therefore, we first identified N_{beh} behavioural parameter sets from the hydrological module as explained below, approximately accounting

for 1% of the original N random sets. Then, we sampled $N_e \approx N/N_{beh}$ independent random sets of erosion parameters and combined them with the behavioural hydrological parameter sets, resulting again in approximately 10000 different combinations effectively assimilated in the erosion module. We compared this to a completely random selection of N combinations of the 13 parameters involved in the entire PED model. This comparison aims to answer the question of whether optimising the erosion module independently will provide behavioural results for the hydrological module as well (**Q1**).

2. **Model evaluation:** We computed a matrix Y of model output samples given the inputs from X. The outputs of the PED model include simulated discharge, flow components, soil moisture status, and sediment concentration. Model outputs are also the objective functions used to evaluate model simulations. We used a multi-objective evaluation (Gupta et al., 1998) consisting of the Nash-Sutcliffe Efficiency (*NSE*) (Nash & Sutcliffe, 1970), Percent Bias (*PBIAS*) (Gupta et al., 1999) and the Kling–Gupta Efficiency (*KGE*) (Gupta et al., 2009). They are defined, respectively, as:

$$NSE = 1 - \left[\frac{\sum_{t=1}^{n} \left(Y_t^{obs} - Y_t^{sim} \right)^2}{\sum_{t=1}^{n} \left(Y_t^{obs} - \mu_{Yobs} \right)^2} \right] \quad \textbf{(2)}$$

$$PBIAS = 100 \left[\frac{\sum_{t=1}^{n} \left(Y_t^{obs} - Y_t^{sim} \right)}{\sum_{t=1}^{n} \left(Y_t^{obs} \right)} \right] \quad \textbf{(3)}$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{Ysim}}{\sigma_{Yobs}} - 1\right)^2 + \left(\frac{\mu_{Ysim}}{\mu_{Yobs}} - 1\right)^2} \quad \textbf{(4)}$$

where Y_t are the observed (*obs*) or simulated (*sim*) variables (e.g., flow or sediment concentration) at time *t*; μ and σ are their correspondent mean and standard deviation, respectively; and, *r* is the correlation coefficient between simulated and observed values. In order to frame the multi-objective optimisation as a minimisation exercise, *NSE* and *KGE* were subtracted from 1 whereas PBIAS was converted to absolute values to place the theoretical optimum at zero for all metrics.

Model calibrations that optimise the *NSE* alone have two problems: (1) the simulated flow variability will be systematically underestimated; and, (2) the bias component has a low weight when the observed flow variability is high. Therefore, the addition of *PBIAS* is to address the flow variability and give more representation to low flow errors. In addition, Gupta et al. (2009) presented a decomposition of the *NSE* criterion expressed in terms of three components: linear correlation, bias, and variability of flow. *KGE* aims to optimise these three components simultaneously, although it may yield lower *NSE* values.

Moriasi et al. (2007) recommend judging parameter sets as satisfactory if *NSE>0.50*, and if *PBIAS<±25%* for streamflow or *PBIAS<±55%* for sediment modelling. In our implementation, behavioural thresholds for the three criteria were defined from the top 5% performing parameter sets. This approach results in better, more exigent values than the recommendation keeping approximately only the top 1% parameter combinations (**Figure 2**). The 'best set' overall was selected as the one which performs simultaneously well for all

metrics. This is achieved by identifying the minimum Euclidian distance from the theoretical optimum (zero) in a 3-dimensional space, which is expressed as:

$$MultiObj_i = \sqrt{\left(NSE_{inorm}\right)^2 + \left(PBIAS_{inorm}\right)^2 + \left(KGE_{inorm}\right)^2} \quad \textbf{(5)}$$

where $MultiObj_i$ is the multi-objective function for parameter combination i. Instead of assigning distinct weights to each individual objective function (NSE_i, $PBIAS_i$, and KGE_i), these were ranked and normalised by the worst and best values obtained from the N parameter combinations. As shown for example for *PBIAS* in **Equation 6**, this means that effectively the best performing set is assigned a normalised objective function value of 0, while the worst performing set is assigned a normalised objective function value of 1.

$$PBIAS_{inorm} = \frac{\left|PBIAS_i\right| - \min_i \left|PBIAS\right|}{\max_i \left|PBIAS\right| - \min_i \left|PBIAS\right|} \quad \textbf{(6)}$$

Lastly, we also compared the results obtained from the automatic random sampling with those of manually calibrated sets from the literature (**Tables I** and **II**, Engda et al., 2011; Tilahun et al., 2013a; Tilahun et al., 2013b; Guzman et al., 2017b).

3. **Sensitivity analysis:** We obtained a matrix S of sensitivity analysis outputs from the model inputs X and outputs Y. The sensitivity of each parameter was assessed applying multio-bjective regional sensitivity analysis (RSA) (Spear & Hornberger, 1980; Sieber & Uhlenbrook, 2005) and the moment-independent PAWN method (Pianosi & Wagener, 2015;

Pianosi & Wagener, 2018). For both methods, we assessed the robustness of the sensitivity indices by bootstrapping (Efron, 1979; Efron & Tibshirani, 1993) using 100 resamples, which approximates the uncertainty in the calculation of *S* when subgroups of parameter combinations are used iteratively in the calculations.

The RSA approach splits the parameter population into two datasets (behavioural and non-behavioural) depending on whether the associated sample *i* in matrix *Y* is better than the threshold defined for the evaluated objective function *j*:

$$Y_{i,j} < threshold\left(j\right) \tag{7}$$

The method then assesses the difference between the cumulative density functions (CDFs) of behavioural ($y_{beh} \in Y$) and non-behavioural ($y_{non\text{-}beh} \in Y$) sets by a suitable measure [e.g., maximum vertical distance (*mvd*) or area between curves (*spread*), (Pianosi et al., 2015)]. The *mvd* for input parameter $x_i \in X$ is calculated using the Kolmogorov-Smirnov statistic (Kolmogorov, 1933; Smirnov, 1939) as:

$$mvd(x_i) = \max_{x_i} \left| CDF_{x_i|y_{beh}}\left(x_i\right) - CDF_{x_i|y_{non-beh}}\left(x_i\right) \right| \tag{8}$$

The larger the sensitivity index (*mvd* or *spread*) of an input parameter, the higher the sensitivity to that parameter. However, in contrast to *spread*, *mvd* is an absolute measure, i.e. it has meaningful value per se, regardless of the units of measures of *X* and *Y* (Pianosi et al., 2015). By definition, *mvd* ranges from 0 to 1; if equal to 0, then the two CDFs are exactly the

same; if equal to 1, then the two CDFs are 'mutually exclusive' (the same value is given probability 0 by one CDF and 1 by the other); the higher the *mvd is*, the higher the sensitivity to that parameter is (Pianosi et al., 2015).

The PAWN method also uses CDFs for sensitivity analysis but differently than RSA. RSA focuses on how input distributions vary when conditioning the output, for instance, using thresholds for objective functions (**Equation 7**). In contrast, PAWN focuses on how the output distribution varies when conditioning an input, i.e., when removing the uncertainty around the input (Pianosi & Wagener, 2015). Whereas RSA quantifies differences in the CDFs of the inputs (model parameters) using **Equation 8**, PAWN uses differences in the CDFs of the output (flow and sediment simulations, or performance metrics) using **Equation 9**:

$$KS(x_i) = \max_y \left| CDF_y\left(y\right) - CDF_{y|x_i}\left(y\right) \right| \quad \textbf{(9)}$$

where KS is the Kolmogorov-Smirnov statistic, $CDF_y(y)$ is the unconditional CDF of output $y \in Y$ when all inputs vary simultaneously, and $CDF_{y|x_i}(y)$ is the conditional CDF obtained when all inputs but x_i vary. As *KS* would depend on the fixed nominal value of x_i, Pianosi & Wagener (2015) and Pianosi & Wagener (2018) recommend using a statistic (e.g., median or maximum) of *KS* over all possible values of x_i to derive the *pawn* sensitivity index:

$$pawn\left(x_i\right) = \underset{x_i}{stat}\left[KS\left(x_i\right) \right] \quad \textbf{(10)}$$

In this numerical implementation of PAWN, the maximum was used as the statistic to define the *pawn* sensitivity index. **Equations 9 and 10** were approximated using the generic sampling approach developed by Pianosi & Wagener (2018), splitting the range of variation of each input parameter x_i into $n=10$ equally spaced intervals, and thus using parameter samples of size N/n for building the unconditional and conditional CDFs. The impact of these approximation errors is inferred by using a *dummy parameter* that, in principle, should not affect the output variability (Zadeh et al., 2017). The value of the dummy *pawn* sensitivity index is used to put all other *pawn* sensitivity results into context (Pianosi & Wagener, 2018). By definition, *pawn* ranges from 0 to 1; the lower *pawn* is, the less influential the evaluated parameter is; if equal to 0, then the parameter has no influence on the output (Pianosi & Wagener, 2015). In addition, if *pawn* is significantly larger than the dummy sensitivity, then the parameter is indeed influential; if equal or even smaller than the dummy sensitivity, then the parameter is potentially uninfluential (Pianosi & Wagener, 2018).

To identify which parameters are the most influential in generating behavioural (good performing) model simulations (**Q2**), we used RSA with the three performance metrics (*NSE*, *PBIAS*, and *KGE*) and PAWN with the multi-objective function (*MultiObj*) as the model output *Y*. More interestingly, we used PAWN to identify which parameters are the most influential in the simulation of different flow magnitudes (minimum, mean, and maximum) and sediment concentrations (mean and maximum) (**Q3**).

2.3. Application study site and available data

We used data from the Anjeni watershed in the Ethiopian Highlands covering the period between 01/01/1990 to 31/12/1993. These data were originally collected by the Soil Conservation Research Programme (SCRP) and made available through the Amhara Regional Agricultural Research Institute (ARARI). The Anjeni watershed has more than 70% of the area under cultivation (Legesse, 2009). The catchment has an area of 1.13 km^2, with elevations between 2405 and 2507 m above sea level and receives an average of 1690 mm of rainfall (Tilahun et al., 2013b). The watershed is characterised by a mixture of deep soils and shallow soils with a hard pan. Soil and water conservation structures were installed in 1985. They consisted of a trench and the removed soil was placed uphill to form an embankment (Bosshart, 1995). Terraces were formed by deposition of soil uphill of the embankment.

For more details on the Anjeni watershed and the available data, the reader is referred to Liu et al. (2008), Engda et al. (2011), Tilahun et al. (2013a), Tilahun et al. (2013b). Guzman et al. (2013), and Guzman et al. (2017b). We focus this article in the evaluation of the sensitivity of the different PED model parameters and the relation between model structure and observable and measurable processes in the field, a procedure that can be applied similarly to other models, catchments, and datasets.

2.4. Code availability

The code for the Parameter-Efficient Distributed (PED) model and for the sensitivity analysis workflow using the SAFE toolbox are in the form of freely available MATLAB scripts in a GitHub repository (Ochoa-Tocachi, 2018).

3. RESULTS

3.1. (Q1) does good performance in the hydrological module translate directly into good performance in the erosion module and vice versa?

When applying automatic sampling of parameter sets for the hydrological module, the top 5% performing sets for each criteria resulted in thresholds of 0.76, ±3.23%, and 0.84 for *NSE*, *PBIAS*, and *KGE*, respectively (**Table III**). Behavioural sets are those whose performance is better than these thresholds simultaneously, which resulted in approximately 1% (105) of the 10000 sampled parameter sets in a single run (**Figure 2a**). The best *NSE* obtained was 0.83 (with corresponding *PBIAS*=10.42% and *KGE*=0.82), the best *PBIAS* was -0.0089% (with *NSE*=0.72 and *KGE*=0.79), and the best *KGE* was 0.91 (with *NSE*=0.82 and *PBIAS*=-2.29%). Although all these values are considerably better than the minima recommended in the literature (Moriasi et al., 2007), this also evidences that different combinations of parameters can yield acceptable performances. This issue, known as equifinality (Beven, 2006), hinders the identification of a single global optimum set. The 'best set', as defined using **Equation 5**, resulted in *NSE*=0.81, *PBIAS*=-0.04%, and *KGE*=0.90 (**Table I**).

Results are more modest for the erosion module. When using behavioural sets from the hydrological module combined with random sampled erosion parameters, only 0.77% (78) of the 10080 combinations were behavioural. This resulted in thresholds of 0.58, ±4.50%, and 0.72 for *NSE*, *PBIAS*, and *KGE*, respectively (**Table III**). The best *NSE* obtained was 0.69 (with *PBIAS*=-2.30% and *KGE*=0.80), the best *PBIAS* was 0.0027% (with *NSE*=0.55 and *KGE*=0.67), and the best *KGE* was 0.80 (with *NSE*=0.66 and *PBIAS*=-3.87%).

Again, although any of these values can still be judged as excellent, they do not necessarily belong to a single set and, therefore, they might not be better than the defined thresholds simultaneously (**Figure 2b**). For this case, and particularly for this exercise run, the 'best set' in the erosion module coincided with the one that provided the best *NSE* (**Table II**). Behavioural sets in the hydrological module are not directly associated to good performance in the erosion module, as evidenced by the high dispersion in **Figure 2b**. Similarly, as seen in **Table II**, values of the four erosion parameters are the most different, which suggests that erosion module performance is not as dependent on the input parameters from the hydrological module as it is on its own erosion parameter values.

Interestingly, the evaluation of the erosion module using completely random combinations of the 13 PED model parameters performs similarly well (**Figure 2c**). Thresholds resulted in 0.57, ±5.58%, and 0.71 for *NSE*, *PBIAS*, and *KGE*, respectively, with 0.91% of the 10000 combinations being behavioural (**Table III**). The best *NSE* obtained was 0.67, the best *PBIAS* was -0.0130%, and the best *KGE* was 0.81. The 'best set' yielded *NSE*=0.64, *PBIAS*=-0.86%, and *KGE*=0.81 (**Table II**). These numbers are similar to those obtained using behavioural hydrological sets, which indicates that the erosion module is insensitive to the hydrological parameters, i.e., it is more sensitive to the rainfall amount than to the exact amount of overland flow. In contrast, the behavioural sets from the erosion module do not perform necessarily well when tested in the hydrological module (**Figure 2d**). Although some parameter sets fell within the performance thresholds recommended by Moriasi et al. (2007), none was behavioural according to the criteria derived from the global population. Behavioural erosion sets can yield hydrological module performances as poor as

NSE=-0.20, *PBIAS*=-65.14%, and *KGE*=-0.01 (**Table III**), indicating that good performance in one module does not necessarily translate to good performance in the other. This is especially true if one module uses only a subset (i.e., surface runoff from areas A_1 and A_2 in the erosion module) of the parameters of the other modules (i.e., interflow, baseflow, evapotranspiration, and surface runoff from areas A_1 and A_2 in the hydrological model). However, under the premise that a model not only needs to perform well but to do so for the right reasons, using behavioural sets from the hydrological module do provide slightly better performances in the erosion module, although this result is not statistically significant.

3.2. (Q2) what are the most influential parameters in providing good performing model simulations?

Equifinality is indeed present. Although the overall statistics (e.g., percentage of behavioural sets, best *NSE*, *PBIAS*, and *KGE*) remain similar when repeating the exercise with new random combinations, the actual absolute parameter values in the identified behavioural and best sets will differ each time. As expected, it is not possible from the discharge signal at the outlet of the watershed to identify a single combination of parameters that optimises all performance metrics at the same time (**Figure 2**), which is even more problematic having the two independent modules in the PED model. Seemingly similar parameter sets (e.g. Tilahun et al., 2013a; Tilahun et al., 2013b; Guzman et al., 2017b) can yield similarly good *NSE* values but very different *PBIAS*, whereas other much different parameter combinations ('best sets' from random sampling) can offer consistently good statistics (**Tables I** and **II**, and **Figure 3**).

Despite this lack of identifiability, the multi-objective RSA (evaluated using the combined thresholds of *NSE*, *PBIAS*, and *KGE*) reveals that behavioural sets might occur in specific regions of the parameter input space. Good performing hydrological simulations were sensitive to the portion of degraded and non-degraded hillside areas in the watershed (A_2 and A_3), the parameter that determines the relative amount of baseflow and interflow (BS_{max}), and the interflow mean duration (τ^*) (**Figure 4a**). The four erosion parameters ($a_{t,1}$, a_{t2}, $a_{s,1}$, and a_{s2}) are the main control of the sediment concentration response, using either behavioural or completely random hydrological sets as input for the erosion module (**Figures 4a** and **4b**). In the latter case, the only hydrological parameter that has a strong influence in the erosion module performance is the maximum soil storage capacity of the degraded hillsides ($S_{max\,2}$) and, in fact, the sediment transport limiting capacity ($a_{t,2}$) and the sediment source limiting capacity ($a_{s,2}$) of these degraded areas were the most influential (**Figure 4b**). Despite the high RSA uncertainty, it is thus possible from the nine parameters that present the highest sensitivity in the performance of the PED model to determine the degradation state of a watershed based on the discharge at the outlet, and to provide good estimations of discharge and sediment concentrations (e.g., **Figure 3**). The other four parameters of the PED model (A_1, $S_{max\,1}$, $S_{max\,3}$, and $t_{1/2}$) were insensitive and, therefore, unidentifiable.

The PAWN sensitivity analysis suggests that, in fact, variations in particular model parameters are less influential in the overall PED model performance (evaluated using MultiObj) (**Figures 4c** and **4d**). Again, the hydrological module outputs were mostly controlled by the portion of degraded and non-degraded hillside areas in the watershed (A_2 and A_3), and to a lesser extent by the interflow mean duration (τ^*). In contrast to RSA, PAWN

determined a much smaller influence of the baseflow reservoir capacity (BS_{max}) (**Figure 4c**). The sediment concentration outputs were much more sensitive to the erosion parameters, remarkably to the sediment transport and source limiting capacity of the degraded hillside areas ($a_{t,2}$ and $as_{,2}$) and, to a lesser extent, to the fractional hillside areas (A_2 and A_3) that control both sediment yield and sediment concentration dilution (**Figure 4d**).

3.3. (Q3) what are the most influential parameters in the simulation of low, mean, and peak flows, and of mean and peak sediment concentrations?

When model outputs are evaluated not in terms of model performance (an objective function metric) but in terms of actual model simulations (flow magnitudes and sediment concentrations), results are more remarkable. The PAWN sensitivity analysis shows that model parameters have different influence in the simulation of outcomes with a clear physical meaning. Peak flow simulation was mostly determined by the distribution of the three fractional saturated, degraded, and permeable areas (A_1, A_2, and A_3), and the interflow mean duration (τ^*), which controls how quickly water moves through the fast subsurface pathway (**Figure 5a**). Mean flow simulation, in contrast, was also controlled by the fractional areas, especially those of the degraded and permeable hillsides (A_2 and A_3), and their correspondent maximum soil moisture storage capacities ($S_{max\,2}$, $S_{max\,3}$) (**Figure 5b**). These parameters are the main control of the catchments' water yield, whereas those involved in daily flow variability (BS_{max}, $t_{1/2}$, and τ^*) had no influence on mean flow simulation. Low flow simulation was insensitive to all but two parameters, namely, the parameter that controls the relative amount of baseflow (BS_{max}) and the residence time in the baseflow reservoir ($t_{1/2}$) (**Figure 5c**). This is

consistent with the choice of the output (the minimum flow), as these two parameters represent the mechanism by which water leaving the catchment is delayed, while it is much less sensitive to the parameterisation of the soil moisture account and of the fast pathway (τ^c).

The simulation of peak sediment concentration was largely influenced by the four erosion parameters ($a_{t,1}$, $a_{t,2}$, $a_{s,1}$, and $a_{s,2}$) and insensitive to the hydrological parameters (**Figure 5d**). This is sensible since model performance in the erosion module can be assessed only when sediment transport occurs, as including periods without sediment would overestimate model performance. However, in contrast to the performance analysis, the simulation of peak sediment concentration is insensitive to the maximum soil storage capacity of the degraded hillsides ($S_{max\,2}$) and less sensitive to the hillside area fractions (A_1 and A_2). Nevertheless, the sediment transport limiting capacity ($a_{t,2}$) and the sediment source limiting capacity ($a_{s,2}$) of the degraded areas were still more influential than the erosion capacities of the saturated areas ($a_{t,1}$ and $a_{s,1}$) (**Figure 5d**). Lastly, the simulation of mean sediment concentration shows a sensitivity pattern quite similar to that of erosion module performance and, in contrast to peak sediment concentration sensitivity, area fractions controlling sediment yield (A_1 and A_2) and sediment concentration dilution (A_3) gain more influence (**Figure 5e**). The inconclusion of some parameter sensitivity results in the context of the dummy sensitivity (**Figures 4** and **5**) indicates that uncertainty in individual parameters may easily contribute to total output uncertainty (Pianosi & Wagener, 2015; Pianosi & Wagener, 2018).

4. DISCUSSION

4.1. Interpretation of physical processes behind parameter sensitivity

Tilahun et al. (2013a) performed a manual sensitivity analysis of the PED model based on single parameter modifications of ±10%, ±20%, and ±30% from their optima and looking at changes in *NSE*. They also found that the distribution of the three fractional areas –saturated, degraded, and permeable (A_1, A_2, and A_3)– are more sensitive than their correspondent water storage capacities ($S_{max\,i}$). They attributed such sensitivity to errors in closing the water balance when area fractions are modified. In addition, their higher sensitivities may be determined by the different structure of the model for each landscape unit. A_1 represents the fraction of the watershed that are in the valley bottom, A_2 the hillsides with limited infiltration capacity through the hardpan (degraded areas), and A_3 the remaining hillsides with percolation rates in excess of the daily rainfall intensities.

We observed here that behavioural parameter combinations occur when fractional areas A_1 and A_2 are the smallest and A_3 the largest, which could be attributed to the higher complexity of the model structure for A_3 that provides more degrees of freedom for the daily flow simulation than for the annual water yield. Runoff generation in the PED model occurs immediately after rainfall exceeds soil storage in the portions of the watershed in the valley bottom (A_1), and on the degraded hillside (A_2), whereas percolation of the remaining hillside are routed through a linear reservoir (baseflow) and a zero-order reservoir (interflow). Therefore, when most of the simulated flows are contributed by A_3, the more complex time series is more able to match the daily flow variability in the observed discharge, in contrast to the more rapid runoff response offered by A_1 and A_2 (Tesemma et al., 2010). Although this might suggest that a simpler model would perform similarly well, overland flow from A_1 and A_2 are an accurate and a preferred representation of the rapid runoff response that occurs in

the semi-humid Ethiopian Highlands (Collick et al. 2009; Steenhuis et al., 2009; Tesemma et al., 2010).

The relative contribution of each fractional area in **Table I** should not be a surprise and could be a direct consequence of the monsoonal climate and specific watershed being simulated (i.e., Anjeni in this case). Field observations have shown that the saturated areas in the valley bottom are small in the Anjeni watershed (Guzman et al., 2017b), because of the deep incisions of the streams in the deep soil profile and the lack of a flat area near the stream, which produces little contributions to the outflow signal at the outlet. In other catchments where saturated areas are hydrologically more important, the sensitivity of parameter A_i might become greater (Zimale et al., 2018). Even here, the uncertainty associated to the sensitivity calculation of A_i was quite large (**Figure 4a**). As the fractional areas are the initial control of water yield and flow variability, these are highly influential parameters in the simulation of mean and peak flows (**Figures 5a** and **5b**).

A plausible explanation for the insensitivity of the three soil moisture capacity parameters ($S_{max\,i}$)**,** is related to the fact that soils remain near full capacity during the frequent storms in the rainy season of the simulated watershed (Tilahun et al., 2013a). Indeed, $S_{max\,1}$ and $S_{max\,2}$ were always insensitive (**Figures 4** and **5**). The best hydrological performance was obtained when the permeable soil moisture capacity ($S_{max\,3}$) was smaller. The maximum water storage parameters of the root zone are inconsequential during most of the year because the recharge and the overland flow are equal to the precipitation minus the evapotranspiration on the period prior to the last rain that filled up the soil. The actual evapotranspiration during

the rain phase is always close to the potential rate because the soil is always near the maximum storage value ($S_{max\,1}$).

Only shortly after the dry phase the first few rainfalls are dependent on the soil storage variation. Therefore, the higher sensitivity of the degraded soil moisture capacity ($S_{max\,2}$) when evaluating the erosion module performance (**Figure 4b**) can be linked to the early transportation of sediments during the first rains that form eroding rills in the newly plowed lands (Guzman et al., 2013). In fact, one would expect sediment transport to be determined essentially by how small or large the degraded soil moisture storage capacity is, i.e., how fast or slow the degraded soil saturates and runoff generation in the degraded hillsides starts. This signal is later lost in the overall runoff signal and the soil moisture capacity parameters become mostly insensitive.

In contrast to Tilahun et al. (2013a), who found that the PED model was not greatly dependent on the subsurface flow parameters, we observe that maximum baseflow storage (BS_{max}) and interflow mean duration (τ) were quite sensitive (**Figure 4a**). This was due to the nature of the watershed with flow stopping within 24 hour of the rain event. Behavioural sets tended to present rather small values of these parameters with respect to their entire sampling range. This implies that the relatively large amount of water infiltrated in the permeable hillsides fills the baseflow reservoir quickly, and any excess above its storage capacity generates percolation to the interflow reservoir leaving the catchment more rapidly. Lower volumes in the linear reservoir also represent smaller baseflow amounts, which will then be less important than interflow in the contribution to total flows. We coincide with Tilahun et

al. (2013a), nonetheless, in that the residence time of the baseflow reservoir ($t_{1/2}$) is less sensitive when evaluating model performance and thus unidentifiable.

The subsurface reservoirs are more influential in the simulation of different flow magnitudes. Peak flow simulation was highly sensitive to variations in τ, which controls water movement through the fast subsurface pathway (**Figure 5a**), whereas minimum flow simulation was highly sensitive to BS_{max} and $t_{1/2}$, which represent the mechanism by which water is delayed through the slow subsurface pathway (**Figure 5c**). The higher sensitivity of the baseflow reservoir capacity (BS_{max}) compared to that of the soil moisture storage capacities ($S_{max\,i}$) may be also related to the higher importance of the entire permeable hillside landscape unit (A_3) in the PED model compared to the other fractional areas.

Lastly, we observe that the four erosion parameters were highly sensitive, especially when using behavioural sets from the hydrological module as input to the erosion module (**Figures 4a** and **4c**). The good performance of the erosion module was mostly influenced by the sediment transport limiting capacity ($a_{t,2}$) and the sediment source limiting capacity ($a_{s,2}$) in the degraded areas (**Figure 4**). The latter was also the case when analysing the sensitivity in the simulation of peak and mean sediment concentration (**Figures 5d** and **5e**). In the watershed, when the first rains are forming rills, sediment transport is the greatest and the concentration is determined by the transport limit of the water carrying the sediment. Once the rills are established, the sediment concentration depends on the delivery of soil from the interrill area and is controlled by the source limiting conditions (Tilahun et al., 2013a; Tilahun et al., 2013b; Guzman et al., 2017b). Sensitivity in the erosion module is also linked to the fractional areas to which the erosion limiting conditions are connected to (**Table II**,

and **Figures 4d** and **5e**), and to the area of permeable hillsides (A_s) that control baseflow and interflow that ultimately dilute total sediment concentration in the stream.

Tilahun et al. (2013a) did not interpret the sensitivity of the erosion parameters, partly because of the lower *NSE* efficiencies obtained for the erosion module compared to those of the hydrological module. However, so few parameters involved in the erosion module of the PED model do provide good estimations of sediment concentrations. Other models have a much poorer fit, even though they use many more parameters [e.g., SWAT (Arnold et al., 2012), EUROSEM (Morgan et al., 1998), DYRIM (Wang et al., 2015; Shi et al., 2016)]. Other studies implementing the PED model for sediment concentration estimation need to consider the characteristics of their catchments. For instance, the depth of the soil to bedrock is quite large in the Anjeni watershed, whereas it might be shallower in other watersheds for which the PED model has been tested (Tilahun et al., 2015; Guzman et al., 2017a; Zimale et al., 2018). Future research might consider the application of a multi-basin sensitivity analysis of the PED model or even a multi-model evaluation.

4.2. Potential ways to improve parameter identifiability and model performance

There are several reasons for observing different sets of similar performing parameters. Besides those discussed before, two other possible explanations are: (1) there are strong interaction effects between the PED model parameters that the sensitivity analysis does not capture [see, e.g. Zadeh et al. (2017)] and (2) the objective functions are not diverse enough to identify all parameters. In fact, some parameters can be insensitive in providing good model performance when this is only informed by *NSE*. As shown in **Figure 2**,

parameter sets that yield good *NSE* values may be highly biased, thus adding *PBIAS* as a complementary performance metric provides more insights on the sensitivity and potential identifiability of the parameters. In the task of finding a global optimum set, Gupta et al. (2009) recommends the modeller to move from aggregated metrics towards the use of multiple measures of model performance, for instance, to reduce the simulation bias even at the expense of impairing the *NSE* (e.g., **Figure 2**). As shown by Hoang et al. (2017), a nearly unique set of parameters can be obtained in this way. It is worth highlighting that the manually calibrated optima from previous studies provide excellent performance, comparable to that of the best set from automatic random sampling (**Figure 3**).

The issue of identifiability in rainfall-runoff modelling theory is also related to the correct understanding and selection of acceptable conceptual representations of the hydrological processes in the system (Beven, 2006). The PED model offers a process-based model structure whose conceptualisation is consistent with field observations and has proved useful in providing better understanding and predictions of the degree of landscape degradation and its effects on hydrological and erosion processes at different scales (Easton et al., 2010; Tilahun et al., 2013a; Tilahun et al., 2013b; Tilahun et al., 2015; Guzman et al., 2017a; Guzman et al., 2017b; Zimale et al., 2018). In contrast to structural identifiability, the issue of parameter identifiability can be addressed by informing calibration with the use of complementary data.

Guzman et al. (2017b) suggest that the problem of parameter identifiability can be addressed by fixing insensitive parameters in advance, which can be done, for example, by using auxiliary soil or groundwater data (Shin et al., 2015). These data, together with

geographical information and local knowledge, can be also used to constrain the extent of the different landscape units. In practical terms, it may be possible to fix the fractional areas to approximate their real spatial distribution in the watershed, and to avoid obtaining unrealistic areal values. This was done here by restricting the ranges from which the different area fractions were sampled. However, fixing sensitive parameters in advance compromises model calibration and may result in poor performances, irrespectively of the combination of the other parameters. In addition, the fractional area parameters represent the average of these areas throughout the watershed. While saturated and degraded areas might seem apparent, observations made above ground may not capture the transient dynamics over the season or over the years (Guzman et al., 2017b). For instance, reports from the Palouse (Eastern Washington) indicate that the distribution of the argillic layer/fragipan is erratic and consequently the perched water table dynamics are highly variable as well (O'Geen et al., 2005). Research to map these areas using remote sensing and other methods is still ongoing work.

In addition, advances in ecohydrological monitoring technologies, such as isotope tracing (e.g., McGuire & McDonnell, 2006; Klaus & McDonnell, 2013; Tekleab et al., 2014), can be used to quantify soil and groundwater storage volumes and to estimate transit times in different catchment components. This information may indeed be used to inform the quantification of soil moisture capacities and subsurface flow parameters, by fixing insensitive parameters or narrowing their feasible ranges substantially. However, in areas with extensive data scarcity, more fundamental data such as precipitation, evapotranspiration, or discharge, can account for a large proportion of the uncertainty in the

simulations than the uncertainty added by model structural error and parameter identifiability (Buytaert et al., 2014). Recent research by the authors is directed to improving the interflow component of the PED model calibrating the parameters using groundwater levels.

4.3. Usefulness of modelling and sensitivity analyses for decision-making

Understanding runoff and erosion processes will aid the placement of soil and water conservation practices. The model helps to determine what these processes are and thus has the potential to optimise the effectiveness of interventions. The sensitivity analysis shows that the fractional area of degraded soils determines the amount of overland flow and the amount of sediment leaving the watershed. Increasing the permeable hillside area is important, as this will decrease the sediment yield of a watershed. Therefore, using soil and water conservation practices that convert degraded areas into more permeable ones will decrease runoff generation and soil leaving the watershed.

The (in)sensitivity of the soil moisture capacities shows that difference between saturated and degraded areas (A_1 and A_2) is less important than the sum of both areas. As the water storage in both soils will saturate during the rainy season, the model will likely provide good performances if they were combined. However, a clear distinction between the two areas can be useful to plan and intervene physically in the watershed. For example, installing infiltration furrows in saturated areas will enhance gully formation and erosive processes, while the same practice is beneficial on degraded lands. Practices that prevent rills from forming, initially, and sediment contributions from the interrill areas, later, will decrease sediment concentrations. In addition, replanting the most severely degraded areas and using

natural fertilizers to maintain more vegetation will greatly decrease sediment concentrations. The sensitivity analysis also shows indirectly that the permeable hillsides are not in need of soil and water conservation practices because these areas contribute zero sediment as demonstrated by the good fit of the erosion module. Although the sensitivity of the subsurface parameters is interesting, it does not have a practical significance as these parameters are determined by the geology of the landscape and are not influenced by interventions at the surface level.

In addition to the work from Zegeye et al. (2010), Tebebu et al. (2010), as well as the Bahir Dar University program research, Nigussie et al. (2018) found a mismatch between planned design of soil and water conservation activities and actual processes occurring in the field. Since decision-making is often based on models –rather than on more personal and time-intensive participatory programs–, understanding how models represent the dominant patterns in the watersheds is key to providing design advice that is more likely to make improvements in the landscape; thereby being favourable for the community.

5. CONCLUSIONS

The PED model is a useful coupled hydrological-erosion tool that has proved convenient in modelling discharge and sediment concentrations in humid monsoonal climates. Studies that have implemented the PED model for discharge and sediment concentration estimation had dismissed the issue of equifinality, parameter identifiability, and uncertainty and sensitivity analyses. The contribution of this article is to present the application of a systematic and comprehensive global sensitivity analysis of the PED model

to answer several questions about model implementation and interpretation in degraded humid landscapes.

We find that good performance in the hydrological module does not translate directly to good performance in the erosion module and vice versa. Nevertheless, good performing sets from the hydrological module, which are associated with a good physical representation of the watershed processes, yield slightly better results when input to the erosion module. Furthermore, parameters in the PED model have clear physical meaning, and therefore, can be associated to real physical processes and variables. We have identified which parameters are more important in the optimisation of the PED model performance and, more importantly, which of them are associated to different flow and sediment concentration outcomes. We have interpreted these results in terms of the underlying physical processes in the application watershed.

The identification of model parameters can be complemented with new data sources such as spatially distributed soil moisture and groundwater levels or leveraging recent advances in tracer hydrology to estimate currently insensitive parameters. Nevertheless, even without these additional data and with a relatively small number of parameters, the PED model is able to estimate with outstanding accuracy the discharge and, more importantly, the sediment concentration, which is generally extremely difficult to predict. Sediment concentration simulation is highly sensitive to the extent and characteristics of degraded hillside areas and, therefore, can be useful to support decision-making on soil and water conservation practices.

For water and soil management purposes in data scarce regions, such as many tropical and humid mountainous environments, providing a group of parameter sets that reasonably agree could be more effective than finding a unique set of parameters. Improving the parameter set representativeness of the real system would provide relevant outputs that help understand and predict land degradation processes and watershed management impacts more effectively. Moreover, generating and improving the quality of more fundamental data may have a higher impact in reducing uncertainties for discharge and sediment concentration predictions than the selection of optimum sets in model calibration.

6. ACKNOWLEDGEMENTS

The authors would like to thank Amhara Regional Agricultural Research Institute for providing data and the Soil Conservation Research Program for establishing the study site and source of data. BOT was funded by an Imperial College President's PhD Scholarship and the UK Natural Environment Research Council (grant NE/L002515/1) "Science and Solutions for a Changing Planet DTP". We also acknowledge funding from the UK Natural Environment Research Council, "Mountain-EVO" project (grant NE-K010239-1). The authors declare no conflict of interest.

7. REFERENCES

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N., & Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the*

American Society of Agricultural and Biological Engineers, 55, 1491–1508. doi:10.13031/2013.42675

Bergström, S., & Forsman, A. (1973). Development of a conceptual deterministic rainfall-runoff model. *Hydrology Research*, 4(3), 147–170. doi:10.2166/nh.1973.0012

Beven, K. J. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1–2), 18–36. doi:10.1016/j.jhydrol.2005.07.007

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 241(1), 43–69. doi:10.1080/02626667909491834

Bosshart, U. (1995). *Catchment discharge and suspended sediment transport as indicators of physical soil and water conservation in the Minchet catchment, Gojam Research Unit: A case study in the north-western Highlands of Ethiopia*. *Soil Conservation Research Programme Research Report*. Bern, Switzerland: University of Bern.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T., Bastiaensen, J., De Bièvre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., VanHecken, G., & Zhumanova, M. (2014). Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, 2(26), 1–21. doi:10.3389/feart.2014.00026

Ciesiolka, C. A. A., Coughlan, K. J., Rose, C. W., Escalante, M. C., Hashim, G. M., Paningbatan, Jr. E. P., & Sombatpanit, S. (1995). Methodology for a multi-country study of soil erosion management. *Soil Technology*, 8, 179–192. doi:10.1016/0933-3630(95)00018-6

Collick, A. S., Easton, Z. M., Ashagrie, T., Biruk, B., Tilahun, S. A., Adgo, E., Awulachew, S. B., Zeleke, G., & Steenhuis, T. S. (2009). A simple semidistributed water balance model for the Ethiopian highlands. *Hydrological Processes*, 23, 3718–3727. doi:10.1002/hyp.7517

Easton, Z. M., Fuka, D. R., White, E. D., Collick, A. S., Ashagre, B., McCartney, M., Awulachew, S. B., Ahmed, A. A., & Steenhuis, T. S. (2010). A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia. *Hydrology and Earth System Sciences*, 14, 1827–1841. doi:10.5194/hess-14-1827-2010.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26. doi:10.1214/aos/1176344552

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL, USA: Chapman & Hall/CRC.

Engda, T. A., Bayabil, H. K., Legesse, E. S., Ayana, E. K., Tilahun, S. A., Collick, A. S., Easton, Z. M., Rimmer, A., Awulachew, S. B., & Steenhuis, T. S. (2011). *Watershed hydrology of the (semi) humid Ethiopian Highlands*. In Melesse, A. M. (ed.) *Nile River Basin: Hydrology, Climate and Land Use*, pp. 145–162. New York, NY, USA: Springer Science Publisher.

Flanagan, D. C., & Nearing, M. A. (1995). *USDA Water Erosion Prediction Project hillslope and watershed model documentation. NSERL Report No. 10*. West Lafayette, IN, USA: USDA-ARS National Soil Erosion Research Laboratory.

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

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2), 135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135)

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

Guzman, C. D., Tilahun, S. A., Zegeye, A. D., & Steenhuis, T. S. (2013). Suspended sediment concentration–discharge relationships in the (sub-) humid Ethiopian highlands. *Hydrology and Earth System Sciences*, 17, 1067–1077. doi:10.5194/hess-17-1067-2013

Guzman, C. D., Tilahun, S. A., Dagnew, D. C., Zegeye, A. D., Tebebu, T. Y., Yitaferu, B., & Steenhuis, T. S. (2017a). Modeling sediment concentration and discharge variations in a small Ethiopian watershed with contributions from an unpaved road. *Journal of Hydrology and Hydromechanics*, 65(1), 1–17. doi:10.1515/johh-2016-0051

Guzman, C. D., Zimale, F. A., Tebebu, T. Y., Bayabil, H. K., Tilahun, S. A., Yitaferu, B., Rientjes, T. H. M., & Steenhuis, T. S. (2017b). Modeling discharge and sediment concentrations after landscape interventions in a humid monsoon climate: The Anjeni watershed in the highlands of Ethiopia. *Hydrological Processes*, 31, 1239–1257. doi:10.1002/hyp.11092

Hairsine, P. B., & Rose, C. W. (1992a). Modeling water erosion due to overland flow using physical principles 1. Sheet flow. *Water Resources Research*, 28, 237–243. doi:10.1029/91WR02380

Hairsine, P. B., & Rose, C. W. (1992b). Modeling water erosion due to overland flow using physical principles 2. Rill flow. *Water Resources Research*, 28, 245–250. doi:10.1029/91WR02381

Hoang, L., Schneiderman, E. M., Moore, K. E. B., Mukundan, R., Owens, E. M., & Steenhuis, T. S. (2017). Predicting saturation-excess runoff distribution with a lumped hillslope model: SWAT-HS. *Hydrological Processes*, 31, 2226–2243. doi:10.1002/hyp.11179

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 83–91. doi:n/a

Legesse, E. S. (2009). *Modeling rainfall-runoff relationships for the Anjeni watershed in the Blue Nile basin. M.P.S. Thesis*. Ithaca, NY, USA: Cornell University.

Liu, B. M., Collick, A. S., Zeleke, G., Adgo, E., Easton, Z. M., & Steenhuis, T. S. (2008). Rain-fall discharge relationships for a monsoonal climate in the Ethiopian highlands. *Hydrological Processes*, 22, 1059–1067. doi:10.1002/hyp.7022

Klaus, J., & McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, 505, 47–64. doi:10.1016/j.jhydrol.2013.09.006

McGuire, K. J., & McDonnell, J. J. (2006). A review and evaluation of catchment transit time modelling. *Journal of Hydrology*, 330(3–4), 543–563. doi:10.1016/j.jhydrol.2006.04.020

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2), 239–245. doi:10.2307/1268522

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

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the American Society of Agricultural and Biological Engineers*, 50, 885–900. doi:10.13031/2013.23153

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. doi:10.1016/0022-1694(70)90255-6

Nigussie, Z., Tsunekawa, A., Haregeweyn, N., Adgo, E., Cochrane, L., Floquet, A., & Abele, S. (2018). Applying Ostrom's institutional analysis and development framework to soil and water conservation activities in north-western Ethiopia. *Land Use Policy*, 71, 1–10. doi:10.1016/j.landusepol.2017.11.039

Ochoa-Tocachi, B. F. (2018, October 2). MATLAB code for the Parameter-Efficient Distributed (PED) model and sensitivity analysis routine using the SAFE toolbox [GitHub repository]. Retrieved from https://github.com/topicster/PED_model

O'Geen, A. T., McDaniel, P. A., Boll, J., & Keller, C. K. (2005). Paleosols as deep regolith: Implications for ground-water recharge across a loessial climosequence. *Geoderma*, 126, 85–99. doi:10.1016/j.geoderma.2004.11.008

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1–4), 275–289. doi:10.1016/S0022-1694(03)00225-7

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling & Software*, 70, 80–85. doi:10.1016/j.envsoft.2015.04.009

Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67, 1–11. doi:10.1016/j.envsoft.2015.01.004

Pianosi, F., & Wagener, T. (2018). Distribution-based sensitivity analysis from a generic input-output sample. *Environmental Modelling & Software*, 108, 197–207. doi:10.1016/j.envsoft.2018.07.019

Setegn, S. G., Dargahi, B., Srinivasan, R., Melesse, A. M. (2010). Modeling of sediment yield from Anjeni-gauged watershed, Ethiopia using SWAT model. *Journal of the American Water Resources Association*, 46, 514–526. doi:10.1111/j.1752-1688.2010.00431.x

Shi, H. Y., Li, T. J., Wang, K., Zhang, A., Wang, G. Q., & Fu, X. D. (2016). Physically based simulation of the streamflow decrease caused by sediment-trapping dams in the middle Yellow River. *Hydrological Processes*, 30(5), 783–794. doi:10.1002/hyp.10649

Shin, M. J., Guillaume, J. H. A., Croke, B. F. W., Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, 520, 1–16. doi:10.1016/j.jhydrol.2014.11.040

Sieber, A., & Uhlenbrook, S. (2005). Sensitivity analysis of a distributed catchment model to verify the model structure. *Journal of Hydrology*, 310(1–4), 216–235. doi:10.1016/j.jhydrol.2005.01.004

Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin*, 2(2), 3–14. doi:n/a

Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peel inlet–II, identification of critical uncertianties via generalized sensitivity analysis. *Water Research*, 14, 43–49. doi:10.1016/0043-1354(80)90040-8

Steenhuis, T. S., Collick, A. S., Easton, Z. M., Leggesse, E. S., Bayabil, H. K., White, E. D., Awulachew, S. B., Adgo, E., & Ahmed, A. A. (2009). Predicting discharge and sediment for the Abay (Blue Nile) with a simple model. *Hydrological Processes*, 23(26), 3728–3737. doi:10.1002/hyp.7513

Steenhuis, T. S., Hrncir, M., Poteau, D., Romero Luna, E. J., Tilahun, S. A., Caballero, L. A., Guzman, C. D., Stoof, C. R., Sanda, M., Yitaferu, B., & Cislerova, M.

(2013). A saturated excess runoff pedotransfer function for vegetated watersheds. *Vadose Zone Journal*, 12(4), 1–10. doi:10.2136/vzj2013.03.0060

Tebebu, T. Y., Steenhuis, T. S., Dagnew, D. C., Guzman, C. D., Bayabil, H. K., Zegeye, A. D., Collick, A. S., Langan, S., MacAlister, C., Langendoen, E. J., Yitaferu, B., & Tilahun, S. A. (2015). Improving efficacy of landscape interventions in the (sub) humid Ethiopian highlands by improved understanding of runoff processes. *Frontiers in Earth Science*, 3(49), 1–13. doi:10.3389/feart.2015.00049

Tebebu, T. Y., Bayabil, H. K., Stoof, C. R., Giri, S. K., Gessess, A. A., Tilahun, S. A., & Steenhuis, T. S. (2017). Characterization of degraded soils in the humid Ethiopian highlands. *Land Degradation and Development*, 28, 1891–1901. doi:10.1002/ldr.2687

Tekleab, S., Wenninger, J., & Uhlenbrook, S. (2014). Characterisation of stable isotopes to identify residence times and runoff components in two meso-scale catchments in the Abay/Upper Blue Nile basin, Ethiopia. *Hydrology and Earth System Sciences*, 18, 2415–2431. doi:10.5194/hess-18-2415-2014

Tesemma, Z. K., Mohamed, Y. A., & Steenhuis, T. S. (2010). Trends in rainfall and runoff in the Blue Nile basin: 1964–2003. *Hydrological Processes*, 24(25), 3747–3758. doi:10.1002/hyp.7893

Tilahun, S. A., Guzman, C. D., Zegeye, A. D., Engda, T. A., Collick, A.S., Rimmer, A., & Steenhuis, T. S. (2013a). An efficient semi-distributed hillslope erosion model for the sub-humid Ethiopian highlands. *Hydrology and Earth System Sciences*, 9(2), 2121–2155. doi:10.5194/hess-17-1051-2013

Tilahun, S. A., Mukundan, R., Demisse, B. A., Engda, T. A., Guzman, C. D., Tarakegn, B. C., Easton, Z. M., Collick, A. S., Zegeye, A. D., Schneiderman, E. M., Parlange, J. Y., & Steenhuis, T. S. (2013b). A saturation excess erosion model. *Transactions of the American Society of Agricultural and Biological Engineers*, 56, 681–695. doi:10.13031/2013.42675

Tilahun, S. A., Guzman, C. D., Zegeye, A. D., Dagnew, D. C., Collick, A. S., Yitaferu, B., & Steenhuis, T. S. (2015). Distributed discharge and sediment concentration predictions in the sub-humid Ethiopian highlands: the Debre Mawi watershed. *Hydrological Processes*, 29, 1817–1828. doi:10.1002/hyp.10298

Wang, G. Q., Fu, X. D., Shi, H. Y., & Li, T. J. (2015). In Yang, C. T., & Wang, L. K. (eds.) *Advances in Water Resources Engineering, Handbook of Environmental Engineering*, 14, pp. 1–40. London, UK: Springer International Publishing.

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1996). Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology*, 181(1-4), 23–48. doi:10.1016/0022-1694(95)02918-4

Zadeh, F. K., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., Bauwens, W. (2017). Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. *Environmental Modelling and Software*, 91, 210–222. doi:10.1016/j.envsoft.2017.02.001

Zegeye, A. D., Steenhuis, T. S., Blake, R. W., Kidnau, S., Collick, A. S., & Dadgari, F. (2010). Assessment of soil erosion processes and farmer perception of land conservation

in Debre Mewi watershed near Lake Tana, Ethiopia. *Ecohydrology & Hydrobiology*, 10(2–4), 297–306. doi:10.2478/v10104-011-0013-8

Zimale, F. A., Moges, M. A., Alemu, M. L., Ayana, E. K., Demissie, S. S., Tilahun, S. A., & Steenhuis, T. S. (2018). Budgeting suspended sediment fluxes in tropical monsoonal watersheds with limited data: the Lake Tana basin. *Journal of Hydrology and Hydromechanics*, 66(1), 65–78. doi:10.1515/johh-2017-0039

Zimale, F. A., Tilahun, S. A., Tebebu, T. Y., Guzman, C. D., Hoang, L., Schneiderman, E. M., Langendoen, E. J., & Steenhuis, T. S. (2017). Improving watershed management practices in humid regions. *Hydrological Processes*, 31, 3294–3301. doi:10.1002/hyp.11241

Tables:

Table I. Parameter sets in the hydrological module, feasible ranges, manually calibrated 'optima', and 'best set' from automatic sampling. See main text and Figure 1 for the parameter and objective function names.

Parameter	Units	Min	Max	Engda et al. (2012)	Tilahun et al. (2013a, 2013b)	Guzman et al. (2017b)	Best behavioural hydrological set	Best behavioural erosion set
A_1	%	0	20	0	2	2	14.09	0.49
A_2	%	5	40	20	14	10	7.29	11.25
A_3	%	20	80	60	50	47	40.08	68.53
$S_{max\,1}$	mm	50	250	-	200	200	219.45	192.78
$S_{max\,2}$	mm	5	150	150	10	10	132.53	20.41
$S_{max\,3}$	mm	50	300	250	100	65	52.69	68.18
BS_{max}	mm	50	200	70	100	100	75.30	89.93
$t_{1/2}$	day	20	100	70	70	75	84.63	33.20
τ*	day	1	50	20	10	10	11.16	25.20
NSE	-			0.68	0.80	0.82	0.81	0.64
PBIAS	±%			-19.43	-8.64	1.13	-0.04	-34.58
KGE	-			0.70	0.87	0.87	0.90	0.60

Table II. Parameter sets in the erosion module, feasible ranges, manually calibrated 'optima', and 'best set' from automatic random sampling. See main text for the parameter and objective function names.

Parameter	Units	Min	Max	Tilahun et al. (2013a)	Tilahun et al. (2013b)	Guzman et al. (2017b)	Best from behavioural hydrological sets	Best from random hydrological sets
a_{t1}	g L^{-1} (mm d^{-1})$^{0.4}$	a_{s1}	15	0.2	4	6	6.25	11.53
a_{t2}	g L^{-1} (mm d^{-1})$^{0.4}$	a_{s2}	15	3.4	4	5	5.87	8.74
a_{s1}	g L^{-1} (mm d^{-1})$^{0.4}$	0	10	-	3	5.5	1.41	8.77
a_{s2}	g L^{-1} (mm d^{-1})$^{0.4}$	0	10	-	3	4.5	4.55	5.96
NSE	-			0.57	0.60	0.66	0.69	0.64
PBIAS	±%			37.51	31.61	14.49	-2.30	-0.86
KGE	-			0.38	0.45	0.66	0.80	0.81

Table III. Range of objective function results for behavioural parameter sets in the hydrological and erosion modules of the PED model (**Figure 2**). Behavioural simulations are those that perform simultaneously better than thresholds defined as the 5% top results for the Nash-Sutcliffe Efficiency (*NSE*), Percent Bias (*PBIAS*), and the Kling–Gupta Efficiency (*KGE*). M2007 are the recommendations from Moriasi et al. (2007).

Objective function	Units	Hydrological module with behavioural hydrological sets	Erosion module with behavioural hydrological sets	Erosion module with random hydrological sets	Hydrological module with behavioural erosion sets
M2007 threshold	%	37.29	18.94	15.36	0.23
5% threshold	%	1.05	0.77	0.91	0.00
NSE	-	0.76 – 0.83	0.58 – 0.69	0.57 – 0.67	-0.20 – 0.81
PBIAS	±%	0.0089 – 3.23	0.0027 – 4.50	0.0130 – 5.58	0.37 – 65.14
KGE	-	0.84 – 0.91	0.72 – 0.80	0.71 – 0.81	-0.01 – 0.88
Figure 2 location	-	(a)	(b)	(c)	(d)

Figure captions:

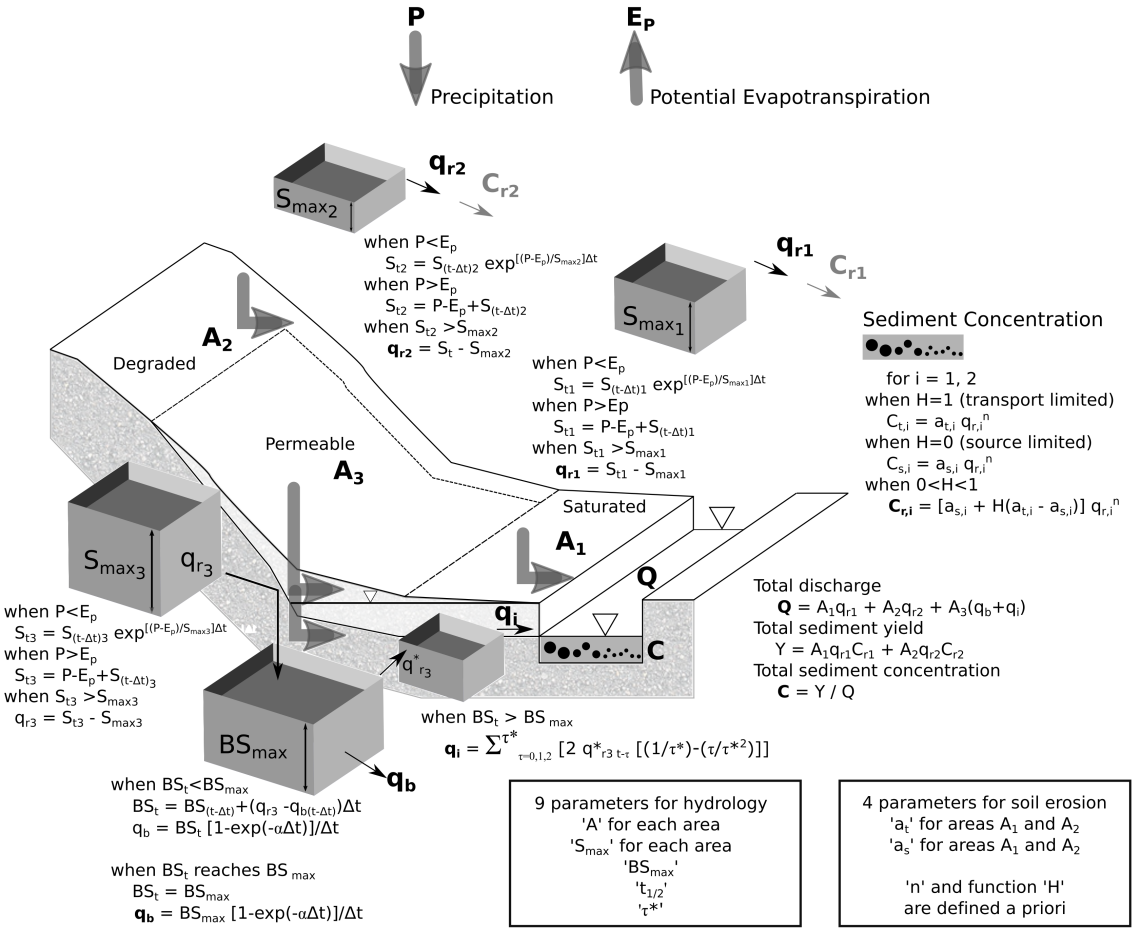

Figure 1. Complete diagram of the Parameter-Efficient Distributed (PED) model. P is precipitation; E_P is potential evapotranspiration; Q is discharge; and, C is sediment concentration. The hydrological module uses 9 parameters: A_i are area fractions of zone i, where i is *1* (saturated), *2* (degraded), and *3* (permeable); $S_{max,i}$ are maximum water storage capacities of area i; BS_{max} is the maximum baseflow storage capacity; $t_{1/2}$ ($=ln(2)/\alpha$) is the time in days required to reduce the volume of the baseflow linear reservoir by a factor of 2 under no-recharge conditions; and, τ is the duration of the period for interflow to cease after percolation, controlled by a zero-order reservoir. The erosion module uses 4 parameters: $a_{t,i}$ are coefficients for transport limiting conditions, and $a_{s,i}$ are coefficients for source limiting conditions, where i is for areas A_1 and A_2. Parameter n and function H, which represents rill formation, are defined a priori.

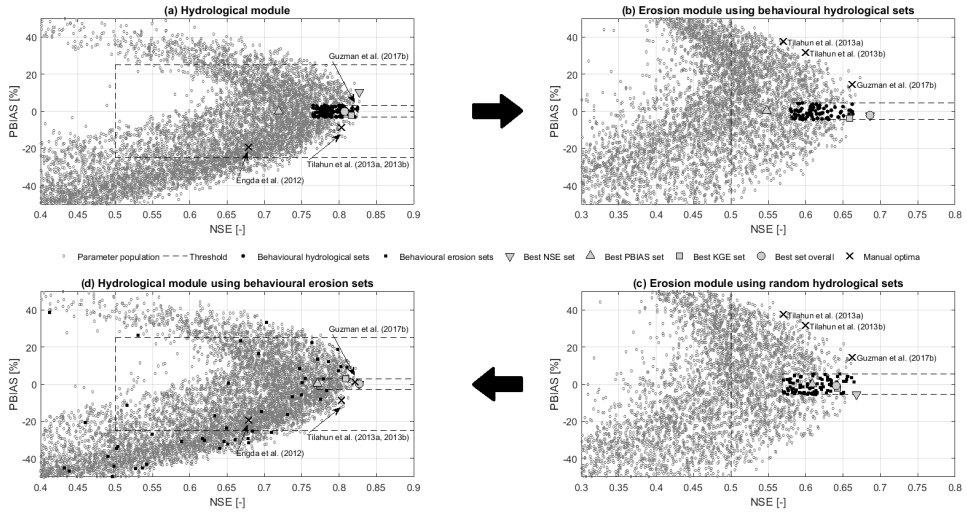

Figure 2. Scatter plots of Nash-Sutcliffe Efficiency (*NSE*) and Percent Bias (*PBIAS*), for the entire population of parameter sets (grey), behavioural parameter sets (black), manual optima from the literature (black cross), and best performing sets for different criteria, including the Kling–Gupta Efficiency (*KGE*). Thresholds (dashed lines) are defined from the top 5% performance of the total population (inner limits) and from Moriasi et al. (2007) recommendations (outer limits). Behavioural simulations are those that perform simultaneously better than thresholds defined as the 5% top results of each objective function. Black circles in plots (a) and (b) represent behavioural hydrological sets which were subsequently input to the erosion module. Black squares in plot (c) represent behavioural erosion sets that were then mapped in the hydrological parameter set shown in plot (d). The larger occurrence of negative *PBIAS* indicates systematic model overestimation with respect to the observed values.

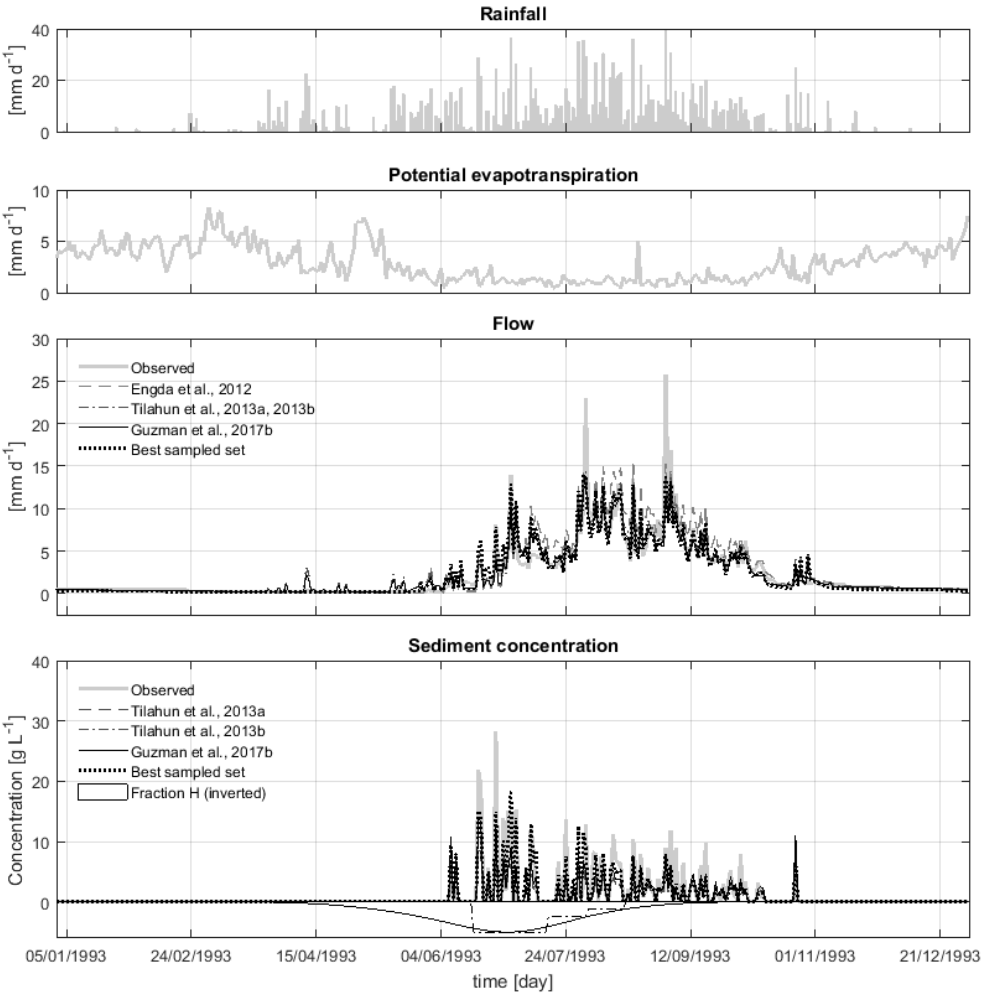

Figure 3. Section of the time series of rainfall, potential evapotranspiration, flow, and sediment concentration. The simulated time series were calculated using the parameter sets from **Tables I** and **II**.

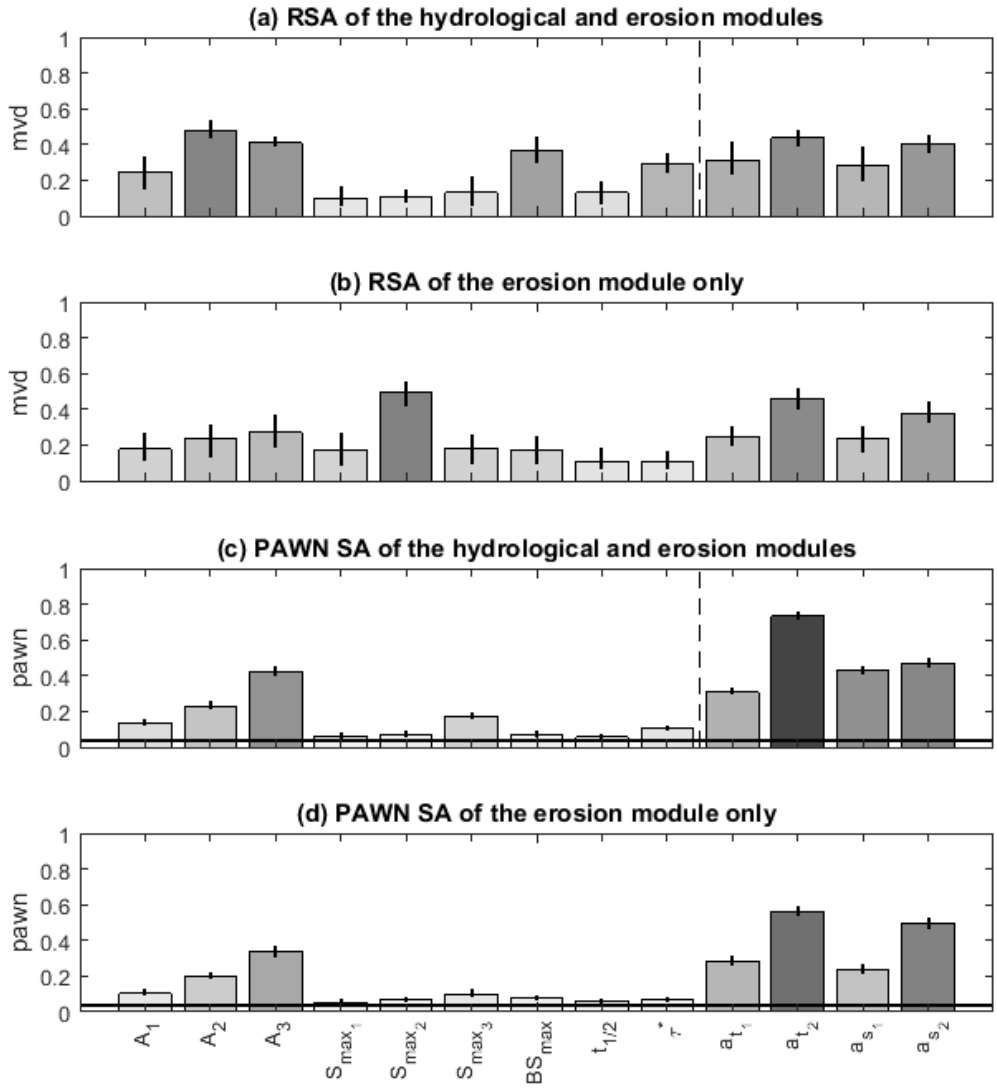

Figure 4. Sensitivity indices of the parameters in the PED model for model performance. Plots (a) and (b) show sensitivity as the maximum vertical distance (*mvd*) index applying regional sensitivity analysis (RSA). Plots (c) and (d) show sensitivity following the PAWN method and using MuiltiObj (**Equation 5**) as the evaluated model output. In (a) and (c), performance for each module is evaluated independently. In (b) and (d), model performance is evaluated on the basis of the erosion module. Larger *mvd* and *pawn*, illustrated by bar height and darker grey shade, represent higher sensitivities. Uncertainty is represented by the vertical line on

top of each bar at the 95% confidence level. The bold black horizontal line depicts the dummy parameter result. See main text and Figure 1 for the parameter names.

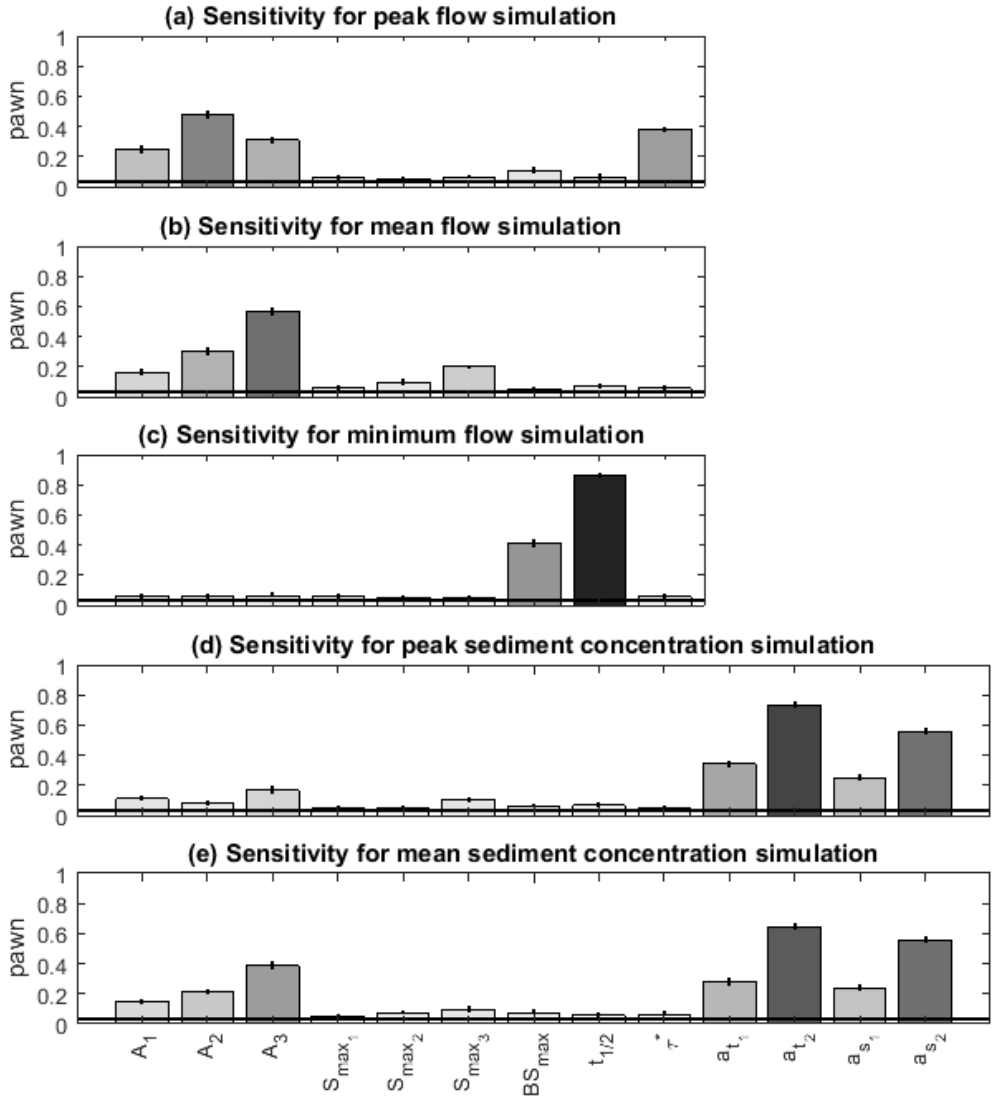

Figure 5. Sensitivity indices of the parameters in the PED model using the PAWN method. Evaluated model outputs are maximum flow (a), mean flow (b), and minimum flow (c) from the hydrological module, and maximum sediment concentration (d), and mean sediment concentration (e) from the erosion module. Larger *pawn*, illustrated by bar height and darker grey shade, represents higher sensitivity. Uncertainty is represented

by the vertical line on top of each bar at the 95% confidence level. The bold black horizontal line depicts the dummy parameter result. See main text and Figure 1 for the parameter names.