Impact of model simplifications on soil erosion predictions: application of the GLUE methodology to a distributed event-based model at the hillslope scale

L. Cea[1,*], C. Legout[2], T. Grangeon[3], G. Nord[2]

Abstract

In this paper we analyse how the performance and calibration of a distributed event-based soil erosion model at the hillslope scale is affected by different simplifications on the parameterisations used to compute the production of suspended sediment by rainfall and runoff. Six modelling scenarios of different complexity are used to evaluate the temporal variability of the sedimentograph at the outlet of a 60 m long cultivated hillslope. The six scenarios are calibrated within the GLUE framework in order to account for parameter uncertainty, and their performance is evaluated against experimental data registered during five storm events. The NSE, PBIAS and coverage performance ratios show that the sedimentary response of the hillslope in terms of mass flux of eroded soil can be efficiently captured by a model structure including only two soil erodibility parameters which control the rainfall and runoff production of suspended sediment. Increasing the number of parameters makes the calibration process more complex without increasing in a noticeable manner the predictive capability of the model.

Keywords: soil erosion, rainfall runoff, physically-based model, model calibration, model validation, GLUE

[*]E.T.S. Ingenieros de Caminos Canales y Puertos, Universidade da Coruña, 15071, A Coruña, Spain, luis.cea@udc.es, Tlf. +34 981167000 Ext.1492

[1]Environmental and Water Engineering Group, Departamento de Métodos Matematicos y de Representación, Universidade da Coruña, Spain

[2]UJF-Grenoble 1 /CNRS / G-INP / IRD, LTHE UMR 5564, Grenoble, F-38041, France

[3]EA6293 GéoHydrosystèmes Continentaux, Université François Rabelais de Tours, Faculté des Sciences et Techniques, Parc de Grandmont, 37200 Tours, France / Bureau de Recherches Géologiques et Minières, 3 Avenue Claude Guillemin, 45060 Orléans, France

1. Introduction

Soil erosion and subsequent sediment delivery to river systems are increasingly studied as a consequence of both on-site and off-site impacts such as net soil and nutrient losses (Pimentel, 2006), turbidity increase in rivers and reservoir filling (Owens et al., 2005). Understanding sediment production and conveyance at the watershed scale implies a detailed analysis of sediment production on hillslopes. This involves complex processes such as the detachment of soil particles due to rainfall impact and runoff shear, the transport of these particles by overland flow and eventually, their deposition in other regions different from where they were originally eroded. Several physically-based formulations to represent these processes at small scales have been proposed in the last years (Beuselinck et al., 2002, 1999; Foster et al., 1995; Govers, 1992; Hairsine and Rose, 1992b,a; Jomaa et al., 2010; Kinnell, 1990, 2005; Nord and Esteves, 2007; Shaw et al., 2009, 2006), and implemented in distributed event-based (Favis-Mortlock et al., 2000; Laloy and Bielders, 2009; Morgan et al., 1998; Nord and Esteves, 2005; Smith et al., 1995) and continuous (Ascough et al., 1997) soil erosion models. All these formulations require a detailed definition of the soil erodibility properties, as well as an accurate representation of the flow field, including water depth, velocity and bed friction. Moreover, the formulations for the calculation of the production of suspended sediments by rainfall and runoff require the definition of parameters which are difficult to measure, and for which there are no available empirical estimations that can be used in a robust way, since the scarce values reported in the literature vary over a wide range (Rousseau, 2012). At the same time, the calibration of distributed soil erosion models with field data is complex for several reasons as: the large number of parameters which need to be estimated, the high non-linearity of the equations, the interaction between input parameters, the scarcity of comprehensive field data available for calibration, the uncertainty in the experimental measurements and input data, and the spatial and temporal variability of the physical processes involved in soil erosion. In order to make affordable the use of distributed soil erosion models in field applications it is necessary to circumvent the previous difficulties in the determination of model parameters. One possible way is to identify the parameters with a highest impact on model output by means of a global sensitivity analysis (Hantush and Kalin, 2005; Laloy and Bielders, 2009; Rousseau, 2012; Veihe and Quinton, 2000) and then focus the efforts of model calibration on accurately identifying these parameters. An alternative way is to simplify the representation of the most relevant sediment production mechanisms without a significant reduction on model performance, in such a way that the number of input parameters and calibration efforts are reduced.

The aim of this paper is to study the impact of different model structure simplifications on the performance of a distributed event-based soil erosion model at the hillslope scale. Model structure is understood in this context as the selection of processes, formulations and parameterisations used to model the production of suspended sediment on hillslopes. For this purpose six modelling scenarios of different complexity are calibrated and validated within the Generalized Likelihood Uncertainty Estimation (GLUE) framework (Beven and Binley, 1992), which assumes that in field applications different parameter sets can produce acceptable results due to our imperfect knowledge of the system and to the uncertainty in input data and parameters. The use of the GLUE framework is of particular interest in soil erosion studies due to the scarcity of accurate field data, which increases the uncertainty on model calibration and validation. This methodology was applied for the first time to soil erosion models by Brazier et al. (2000), who used it to assess explicitly the uncertainties associated to the predictions of annual soil losses at the plot scale. Other recent applications of the GLUE methodology to soil erosion models are described in Quinton et al. (2011) and Krueger et al. (2012). In this study we use the standard GLUE framework to assess to which extent decreasing the complexity of a soil erosion model impacts its performance at the storm event scale. Model performance is evaluated against field measurements of water discharge and sediment mass flux during five storm events. The increase in model performance relative to the increase in model complexity is evaluated and discussed.

2. Numerical Model

2.1. Hydrodynamic Equations

The overland flow water depth and velocity fields are computed from the two-dimensional shallow water equations, including rainfall and infiltration terms and using Manning formulation to compute the bed friction, which can be written as:

$$\begin{aligned}
\frac{\partial h}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} &= R - i \\
\frac{\partial q_x}{\partial t} + \frac{\partial}{\partial x}\left(\frac{q_x^2}{h}\right) + \frac{\partial}{\partial y}\left(\frac{q_x q_y}{h}\right) &= -gh\frac{\partial z_s}{\partial x} - g\frac{n^2}{h^{7/3}}|\mathbf{q}|q_x \\
\frac{\partial q_y}{\partial t} + \frac{\partial}{\partial x}\left(\frac{q_x q_y}{h}\right) + \frac{\partial}{\partial y}\left(\frac{q_y^2}{h}\right) &= -gh\frac{\partial z_s}{\partial y} - g\frac{n^2}{h^{7/3}}|\mathbf{q}|q_y
\end{aligned} \tag{1}$$

where h is the water depth, z_s is the free surface elevation, (q_x, q_y) are the two components of the unit discharge, n is the Manning coefficient, R is the rainfall intensity, i is the infiltration rate and g is the gravity acceleration. Previous works have shown that the depth-averaged shallow water equations are able to represent properly the spatial distribution of water depth and velocity in overland flows, as long as an accurate characterization of the bed roughness coefficient is used (Cea et al., 2014; Mugler et al., 2011; Tatard et al., 2008).

2.2. Soil Erosion Model

The soil erosion model used in this work considers a vertical structure of the soil composed of a non-cohesive layer of eroded sediment which lays over a cohesive matrix of non-eroded soil. Both layers have different erodibility properties. This kind of soil structure has been used in previous works such as Hairsine and Rose (1992a); Heng et al. (2011); Nord and Esteves (2005); Rose et al. (2007); Sander et al. (2007). In this section we describe all the processes implemented in the soil erosion model. As mentioned in the introduction, the aim of this study is to analyse the influence on model performance of different simplifications on the formulations used to compute the production of suspended sediment. These simplifications assume that some of the processes represented in the full model have a minor effect on model output and can therefore be neglected without a significant performance degradation. The model structure simplifications are described in following sections.

The median size of the particles collected at the outlet of the hillslope during the events selected for this study is of the order of 25 $\mu\mathrm{m}$. Bed load is therefore considered to be negligible relative to suspended load. Thus, this latter will be the only sediment transport mechanism considered in the model. The time and spatial evolution of suspended sediment concentration is computed from the following depth-averaged scalar transport equation, which includes several source terms to account for the production and deposition of suspended sediment:

$$\frac{\partial hC}{\partial t} + \frac{\partial q_x C}{\partial x} + \frac{\partial q_y C}{\partial y} = D_{rdd} + D_{rdrd} + D_{fde} + D_{fdre} + D_{dep} \qquad (2)$$

where C $(\mathrm{kg/m^3})$ is the depth-averaged concentration of sediment in the water column, D_{rdd} is the rainfall driven detachment rate from the cohesive layer, D_{rdrd} is the rainfall driven redetachment rate from the eroded layer, D_{fde} is the flow driven entrainment rate from the cohesive layer, D_{fdre} is the flow driven reentrainment rate from the eroded layer and D_{dep} is the deposition rate of suspended sediment in the eroded layer. All the terms in the right hand side of Equation (2) are expressed in $\mathrm{kg/m^2/s}$.

The rainfall driven detachment and redetachment source terms (D_{rdd} and D_{rdrd}) model respectively the sediment transfer from the cohesive and eroded layers to the water column. Both terms are evaluated assuming a linear relationship between the detachment/redetachment rates and the rainfall rate (Sharma et al., 1993, 1995; Gao et al., 2003), as:

$$D_{rdd} = \alpha_d R \left(1 - \varepsilon\right) \qquad\qquad D_{rdrd} = \alpha_{rd} R \varepsilon \tag{3}$$

with:

$$\varepsilon = \min\left[\frac{M_s}{M_{s,cr}}, 1\right] \tag{4}$$

where α_d and α_{rd} ($\mathrm{kg/m^3}$) are the rainfall erodibility coefficients for the cohesive and eroded layers, which represent the flux of sediment mass per unit surface detached by a rainfall intensity of $1\ \mathrm{m/s}$ and ε is a shield factor which represents the protection effect that the eroded layer has over the cohesive layer. The shield factor is assumed to vary linearly with the mass of sediment per unit surface in the eroded layer (M_s). When M_s achieves a critical value ($M_{s,cr}$) the protection effect is maximum, implying that no sediment is eroded from the cohesive layer. This kind of model for rainfall driven erosion has been used in previous works as those presented by Gao et al. (2003), Nord and Esteves (2005), Sharma et al. (1993), Sharma et al. (1995) and Shaw et al. (2006).

The flow driven entrainment and reentrainment rates (D_{fde} and D_{fdre}) model respectively the transfer of sediment particles from the cohesive matrix and from the eroded layer to the water column due to the effect of bed friction. Both terms are computed from the formulation proposed by Hairsine and Rose (1992a) as:

$$D_{fde} = \begin{cases} \dfrac{F}{J}\left(\Omega - \Omega_{cr}\right)\left(1 - \varepsilon\right) & \text{if} \quad \Omega > \Omega_{cr} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$D_{fdre} = \begin{cases} \dfrac{\rho_s r_f F}{\left(\rho_s - \rho\right) g}\left(\dfrac{\Omega - \Omega_{cr}}{h}\right)\varepsilon & \text{if} \quad \Omega > \Omega_{cr} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $\rho_s = 2600\ \mathrm{kg/m^3}$ and $\rho = 1000\ \mathrm{kg/m^3}$ are the densities of sediment particles and water, Ω ($\mathrm{W/m^2}$) is the runoff stream power per unit surface, Ω_{cr} is a critical stream power threshold below which the sediment entrainment rate is zero, F (dimensionless) is the fraction of the stream power excess over Ω_{cr} which contributes to the entrainment and reentrainment of sediment, and J (J/kg) is the specific energy of entrainment, which characterises the resistance offered by the soil matrix to entrainment. In Equation (6) r_f is the Rouse factor, defined as the ratio between the concentration of suspended sediment near the bed and the depth-averaged concentration of sediment. Given the small water depths in overland flow applications, the Rouse factor is often taken to be one, assuming that the sediment is homogeneously distributed over the water column. Therefore, a value of $r_f = 1$ was used in all the simulations presented in this paper. The formulation given by Equations (5-6) assumes that from the total stream power dissipation (Ω), only a fraction given by $F(\Omega - \Omega_{cr})$ contributes to soil erosion, the rest being spent in other head losses. This stream power available for soil erosion is distributed between entrainment and reentrainment according to the thickness of the eroded layer via the parameter ε. In all the simulations done in this study the critical stream power Ω_{cr} in the entrainment and reentrainment equations was set to zero. This simplification is supported by typical values of this parameter reported by other authors, of the order of 0.01 $\mathrm{W/m^2}$ (Heng et al., 2011; Misra and Rose, 1995; Proffitt et al., 1993; Rose et al., 2007; Sander et al., 2007), two orders of magnitude lower than the average stream power in the study hillslope during the storm events analysed here (of the order of 1 $\mathrm{W/m^2}$ according to the numerical simulations).

The deposition of suspended sediment from the water column to the eroded layer is modelled as:

$$D_{dep} = -\rho_s r_f w_s C \tag{7}$$

where w_s is the settling velocity of sediment particles. The settling velocity is computed from the density and diameter of sediment particles using the formulation of van Rijn (1984), which for a particle size of 25 $\mu\mathrm{m}$ gives a value of $w_s = 0.34$ mm/s.

Once the suspended sediment concentration has been evaluated from Equation (2), the following mass balance equation is solved to compute the time evolution of the mass of sediment per unit surface in the eroded layer:

$$\frac{\partial M_s}{\partial t} = -\left(D_{rdrd} + D_{fdre} + D_{dep}\right) \tag{8}$$

The evolution of the bed elevation (z_b) is computed from the following mass conservation equation, which includes all the terms implying movement of sediment particles from either the eroded or the cohesive layer to the water column:

$$\frac{\partial z_b}{\partial t} = -\frac{D_{rdd} + D_{rdrd} + D_{fde} + D_{fdre} + D_{dep}}{\rho_s \left(1 - \phi\right)} \tag{9}$$

where ϕ is the soil porosity. At each time step the new topography computed from Equation (9) is updated in the hydrodynamic equations to ensure an appropriate coupling between the movement of sediment and water.

2.3. Numerical Solver

The overland flow equations are solved with an explicit finite volume solver for the two-dimensional shallow water equations presented and validated in previous works. The reader is referred to Cea et al. (2010) and the references therein for a detailed description of the numerical schemes, including experimental validation under rainfall-runoff conditions.

The suspended sediment transport equation is solved with the explicit finite volume scheme described in Cea and Vázquez-Cendón (2012) for scalar transport equations, which guarantees a mass conservative discretisation of the advection terms. The main singularity of Equation (2) with respect to the standard depth-averaged transport equation is the presence of the source terms which account for the production and deposition of suspended sediment ($D_{rdd}, D_{rdrd}, D_{fde}, D_{fdre}, D_{dep}$). In order to avoid negative values of the suspended sediment concentration during the computation, special care must be placed in the discretisation of the term D_{dep}, which is the only sink in Equation (2). A bad numerical practice is to reset to zero the concentration at any computational cell in which its value becomes negative. This procedure is not used in the present solver because it generates a gain of sediment mass, which can be very relevant if successive wetting and drying cycles occur, as it is often the case in rainfall runoff applications. Alternatively an implicit discretisation of the suspended sediment concentration in the deposition term (D_{dep}) is used, which guarantees the positivity of the suspended sediment concentration and the conservation of sediment mass. At the same time, the source terms D_{rdrd} and D_{fdre} are limited to the availability of sediment in the eroded layer in order to avoid negative values of the sediment mass M_s in Equation (8). Since the all the equations are solved with an explicit scheme, the time step in the calculations was restricted by a CFL condition (Courant et al., 1967) automatically implemented in the solver.

3. Methodology

3.1. Study Site and Observations

The numerical model described in the previous section was applied to compute soil erosion at the outlet ($4^o29\prime43.2\prime\prime E; 44^o34\prime47.3\prime\prime N$) of a hillslope located in the south eastern part of France (Cevennes-Vivarais, Figure 1) during five storm runoff events (Table 1). This instrumented hillslope is part of a wider network of nested catchments, itself being part of the Mediterranean Hydrometeorological Observatory (Boudevillain et al., 2011). Vineyard spreads over the whole hillslope. The brown calcareous soils underlain by marly-limestones are composed with 34% of swelling clays, 41% of silt and 25% of sand particles. The vegetation cover between the vine rows varied between years but remained very sparse. The instrumented hillslope is 60 m long and 2.2 m width, which corresponds to the distance between two vines rows. The topography of the hillslope (Figure 1) was measured twice (2012 and 2014) with a theodolite with a spatial resolution of 1 m^2 and uncertainties of 1 cm in the three dimensions. No significant evolution of the topography occurred between 2012 and 2014. The average slope in the longitudinal direction is about 15% and, as shown in Figure 1, there is a clear rill which collects and conveys the overland flow to the hillslope outlet, avoiding runoff losses through the lateral sides of the hillslope. Rainfall was measured with a raingauge (Précis Mécanique) having a 0.2 mm resolution, whose location is represented in Figure 1. Runoff was collected in the bottom part of the hillslope. The water heights were measured every minute with a 1 mm resolution using a limnimeter (OTT Thalimede) within a H-flume designed following the US Soil Conservation Service recommendations. The discharge rating curve was built experimentally and allowed to calculate discharges with a median relative uncertainty of 10%. A sequential sampler (ISCO 3700 Teledyne) containing 24 bottles of 1 l capacity sampled water and soil aggregates within the H-flume, the intake of the pipe being placed horizontally at the bottom of the flume. When critical thresholds of water heights or water heights variations were exceeded, the data logger (Campbell CR 800) triggered the sampling of water and soil aggregates. Thus, the time intervals between each two samples were irregular, depending on the shape of the hydrograph. The suspended sediment concentrations were estimated by weighting the water samples after drying them during 24 h at 105^oC with a median relative uncertainty of 15%. While the discharges were available continuously, the sediment fluxes were only calculated for the times where suspended sediment concentrations were available.

With mean runoff rates of 110 mm/yr for 890 mm of rainfall and mean erosion rates of the order of 5 t/yr/ha, the site is representative of this type of land in a mediterranean context (Cerdan et al., 2010). It is important to note that these erosion rates are smaller than those recorded in mediterranean areas for bare soils.

3.2. Numerical Model Setup

In order to solve the hydrodynamic and soil erosion equations the study hillslope was discretised with a finite volume mesh formed by 3300 quadrilateral elements of 20 cm length (Figure 1). Given the small size of the hillslope all the parameters included in the hydrodynamic and soil erosion equations were assumed to be spatially homogeneous. Rainfall intensity was also considered to be uniform in the whole hillslope, and it was defined in the model with the same temporal resolution as it was measured by the raingauge. As initial condition it was assumed that the soil surface was completely dry at the beginning of the events. The only boundary condition imposed was a critical flow condition at the hillslope outlet, which is coherent with the experimental conditions.

3.3. Calibration of the Overland Flow Equations

In order to focus our analysis in the soil erosion model, an independent calibration of the Manning coefficient and infiltration parameters was done for each of the five rainfall events to reproduce as best as possible the hydrograph measured at the hillslope outlet. This means that the hydrological performance of the model is prioritized over its sedimentological performance, since the water depth, velocity and bed stress are essential variables in soil erosion modelling. Numerical-experimental agreement was evaluated in terms of the Nash-Sutcliffe Efficiency (NSE), defined as the ratio of the error variance to the variance of the observed time series (Nash and Sutcliffe, 1970).

A Dunne type infiltration model (Dunne and Black, 1970) was used in all the computations. It considers that the soil has a very large infiltration capacity (larger than the rainfall rate) at the beginning of the event, until it gets completely saturated. For modelling purposes, the rainfall depth which is infiltrated in the soil during the first stage is considered as an initial abstraction (I_a). After the soil is fully saturated the infiltration capacity is reduced to a constant value (k_s) of a few $\mathrm{mm/h}$. Since the number of parameters of the overland flow equations is limited (Manning coefficient, initial abstraction, and constant infiltration rate) and the effect of each one in the outlet hydrograph is quite distinctive, their calibration was performed manually to their optimal value. The beginning of the rising limb of the hydrograph is determined mainly by the initial abstraction (I_a) and can therefore be identified quite easily. The Manning coefficient (n) has its major effect on the sharpness of the outlet hydrograph, while the total runoff volume is largely influenced by the constant infiltration capacity (k_s). There is obviously some interaction between these three parameters, mostly between the Manning coefficient and the infiltration capacity, which could be captured by performing a GLUE-based calibration instead of an *optimal value* calibration. However, calibrating independently the hydraulic parameters to their optimal value allows us to detach the hydraulic calibration from the estimation of the soil erosion parameters, and is justified by the fact that the experimental hydrograph is very well captured by the numerical model, with NSE values higher than 0.95 in all cases (Table 2 and Figure 2).

The calibrated parameters are shown in Table 2. The order of magnitude of the Manning coefficient, which oscillates between 0.2 and 0.8 $\mathrm{sm}^{-1/3}$, is consistent with values reported in the literature for overland flows (Engman, 1986; Fraga et al., 2013; Muñoz-Carpena et al., 1999; Wilson and Horritt, 2002), which depend on the vegetative cover, micro-topography, rainfall intensity and water depth. The differences in the calibrated Manning coefficient from one event to another might be explained by differences in the characteristics of the micro-topography during the four years in which the five events took place, as well as by the fact that its numerical calibration might account for all sorts of model deficiencies (Lane, 2014). The constant infiltration rate varies from 0 to 1.8 mm/h in four of the five events, which is consistent with the clayey nature of the soil and with the values measured by Braud et al. (2014) and Braud and Vandervaere (2012), but raises to 18 mm/h in the event R4. It should be noted that R4 is the only event which occurs in early september, corresponding to the beginning of the rainy season in this south eastern part of France. The presence of many desiccation cracks in these dry clayey soils at this period would increase the infiltration capacity of soils due to a dual permeability structure. The variability of I_a from one event to another is also consistent with already published litterature in this catchment, highlighting the high influence of the antecedent soil moisture conditions on the generation of runoff (Huza et al., 2014). In all the events model output agrees very well with the experimental measures (Figure 2), the NSE being larger than 0.97 (Table 2) and the mean absolute error lower than 10% of the maximum discharge in all cases.

3.4. Modelling Scenarios

In the following, model structure is understood as the formulations and parameterisations used to model the production of suspended sediment by rainfall and runoff, which include all the terms on the right hand side of Equation (2).

As justified in section 2.2, in all the scenarios considered in this study the critical stream power and the Rouse factor were fixed respectively to zero and one ($\Omega_{cr} = 0$ and $r_f = 1$). With this assumption, the complete erosion model given by Equations (3), (4), (5) and (6) has 5 parameters, namely: α_d, α_{rd}, F, J and $M_{s,cr}$. Rough estimates for some of the previous parameters can be found in the scientific literature (Heng et al., 2011; Proffitt et al., 1993; Sander et al., 2007). Such estimates are generally obtained after calibration of simplified analytical solutions of soil erosion models with data from laboratory experiments which were undertaken under rather different conditions from those of field applications (constant rainfall intensity, no vegetation, unstructured soils, uniform bed slope, no macro-roughness features). The extrapolation of these estimates to field studies is not evident and thus, parameter calibration is mandatory.

The simplified Model Scenarios (MS) considered in this paper are shown in Table 3. The simplest scenarios are MS1 and MS2, which include a single parameter to model respectively the production of suspended sediment by rainfall and runoff. The modelling scenario MS1 can be easily obtained from the full soil erosion model described in section 2 by forcing the parameters $F = 0$ (neglect runoff production) and $\alpha_d = \alpha_{rd}$, which implies to assume a single soil layer with uniform erodibility characteristics. In this case the total sediment production in the right hand side of Equation (2) is equal to $\alpha_{rd} \times R$ and therefore, the model is insensitive to J and $M_{s,cr}$. With this simplification the only model parameter is α_{rd}. Notice that there are other ways of simplifying the full erosion model to obtain the scenario MS1 which can lead to different physical interpretations. For instance, setting $F = \alpha_d = 0$ and $\varepsilon = 1$ leads also to MS1. Even though the physical interpretation of these choices of parameters is different, they are exactly equivalent from a mathematical point of view. The only parameter of the second modelling scenario (MS2) is F, the fraction of stream power which is spent on soil erosion. In this scenario the rainfall erodibility parameters are set to zero ($\alpha_d = \alpha_{rd} = 0$), and the shield factor to 1 ($\varepsilon = 1$). The scenario MS3 is a combination of the first two scenarios (2 parameters: α_{rd} and F).

The first three scenarios assume that only one soil layer with uniform erodibility properties is active during a single storm event. The other three modelling scenarios (MS4, MS5 and MS6) are respectively the two soil layer extensions of MS1, MS2 and MS3. The fact of considering two layers with different resistance to erosion doubles the number of erodibility parameters in the model and in addition, it is necessary to introduce the shielding parameter $M_{s,cr}$. MS4 is obtained from the full erosion model by just setting $F = 0$, while in MS5 the parameters which are set to zero are α_d and α_{rd}. The most complex scenario is MS6, which corresponds to the full erosion model.

3.5. Calibration of the Soil Erosion Parameters

The parameters of the five modelling scenarios defined in Table 3 were calibrated using the standard Generalised Likelihood Uncertainty Estimation (GLUE) methodology of Beven and Binley (1992), which allows for many acceptable (or behavioural) parameter sets in the calibration process. The parameter sets identified as behavioural are then used to assess the uncertainty on parameter identification and model predictions. The original GLUE methodology was extended in the last years in the so-called limits of acceptability approach, which was first proposed in Beven (2006) to account for observational errors in the field data used to evaluate model performance and as model input. Different implementations of the limits of acceptability approach have been applied to hydrological (Liu et al., 2009; Blazkova and Beven, 2009) and sediment transport (Quinton et al., 2011; Krueger et al., 2012) studies. The advantages of using a limits of acceptability approach within GLUE are discussed in detail in Beven (2006).

In order to focus the analysis on the soil erosion model the overland flow parameters were kept equal to the values detailed in Table 2 in all the simulations, which guarantee an optimal representation of the experimental hydrograph at the plot outlet. Other authors as Quinton et al. (2011) perform an ensemble hydrological and sedimentological calibration within the GLUE framework, varying at the same time the hydraulic and soil erosion parameters of the model. The procedure followed in Quinton et al. (2011) allows to account for the uncertainty on the hydraulic parameters estimation on the calibration of the soil erosion model. However, it increases the number of parameters to calibrate within GLUE and therefore the number of Monte Carlo simulations, which is computationally very expensive in a fully distributed model as the one used in this study, and was for that reason not applied in this study.

Since no prior estimation of the model parameters in our study site was available, a uniform prior distribution over the ranges of variation defined in Table 4 was assumed for all the parameters. These ranges of variation were chosen after some preliminary simulations in which the parameters were varied over wider intervals. To verify that the search of behavioural parameter sets was not restricted by these interval limits, after the calibration process it was verified that the latter probability density functions were not limited by the chosen variation ranges and that the most probable parameter values were located well inside the search interval. The random parameter sets were generated using a Sobol quasi-Monte

Carlo low-discrepancy sequence (Sobol, 1998; Saltelli et al., 2008). This sampling method is very adequate for computationally demanding models, because it allows for the extraction of a large amount of information with a smaller number of parameter sets than traditional Monte Carlo random sampling (Saltelli et al., 2008). The number of random sets was different for each modelling scenario, since the number of input parameters increases with the complexity of the model structure. A total number of 100 sets were generated for scenarios MS1 and MS2, which have a single parameter, 250 for MS3 (2 parameters), 1000 for MS4 and MS5 (3 parameters) and 5000 for MS6 (5 parameters). Each storm event was run with all the previous parameter sets, and behavioural simulations were defined as those with a positive NSE. This implies a quite loose rejection level, which is justified in our case by the large uncertainties involved in measuring and modelling suspended sediment fluxes, including model structural errors. As it will be shown in the results section, the NSE values computed from the suspended sediment fluxes are significantly lower than those computed from the water discharge. Nevertheless, the GLUE methodology is flexible in the definition of the threshold of model rejection, which should be fixed considering data availability and modellers criterion (Beven, 2006).

Each behavioural simulation was assigned a weight w_i computed as:

$$w_i(\theta_i) = \frac{L(\theta_i \mid C^*)}{\sum_{j=1}^{m} L(\theta_j \mid C^*)} \qquad i = 1, m \tag{10}$$

where m is the number of behavioural simulations, θ_i is a behavioural parameter set and C^* is the measured sediment flux. The generalized likelihood measure for each parameter set θ_i was computed as the inverse error variance $L(\theta_i \mid C^*) = \sigma_e^{-2}$, where σ_e is the root mean square error computed from the numerical-experimental agreement of sediment fluxes at the hillslope outlet.

All behavioural parameter sets are run to compute the cumulative density function (cdf) of model predictions at any time step as:

$$P[\hat{C}_t < C] = \sum_{i=1}^{m} w_i(\theta_i \mid \hat{C}_{t,i} < C) \tag{11}$$

where $\hat{C}_{t,i}$ is the model prediction at time t obtained with the parameter set θ_i. The deterministic model prediction and its associated uncertainty are characterised respectively by the median of the cdf and the 95% confidence interval, as it is usually done when applying the GLUE methodology.

3.6. Evaluation of Model Performance

The performance of each modelling scenario was evaluated by comparing the computed and measured sedimentographs for each of the storm events shown in Table 2. Model performance was evaluated in both, calibration and validation phases. In the calibration phase the behavioural parameter sets of each scenario were computed independently for each storm event, and model performance was evaluated in terms of the NSE, the percent bias (PBIAS) and the coverage of the 95% confidence interval, all of them being computed for the suspended sediment fluxes at the hillslope outlet. The PBIAS (Gupta et al., 1999) measures whether there is a tendency in the numerical predictions to be larger or smaller than the experimental observation, while the coverage ratio, defined as the percentage of experimental measurements included in the 95% prediction confidence interval, is an indicator of the model performance considering output uncertainty (Vrugt et al., 2009). In order to compute the NSE and PBIAS performance indices, the median of the output sediment flux cdf was used as the deterministic model prediction.

It should be stressed that the calibration phase allows us to evaluate what is the best performance expected for a given model structure during each rainfall event (within the calibration framework used in this paper), but it does not account for the predictive capability of the model. However, in the validation phase the parameters obtained from the calibration of the storm event R1 were used to model the other four events, and the predictive performance of each modelling scenario was evaluated in terms of the NSE, PBIAS (Gupta et al., 1999) and coverage ratios.

4. Results and Discussion

4.1. Evaluation of the Modelling Scenarios

4.1.1. Calibration

The NSE, PBIAS and coverage performance ratios in calibration phase for the six modelling scenarios and for each rainfall event are shown in Table 5. By considering the rainfall and runoff production parameters (α_{rd} and F) the scenarios MS3 and MS6 are able to reproduce the experimental observations with NSE values larger than 0.85 in all the events except in R1, with no significant bias in the prediction, even if the sedimentary response of the hillslope in terms of sediment flux variability is very different from one event to another, as illustrated in Figure 3. Regarding the percentage of field measurements captured by the estimated

95% confidence interval, there is a general departure of the coverage from 95%, which is usual in the application of the GLUE methodology, and may be explained in our case by the uncertainty on the experimental measurements and imperfections on model structure. This is illustrated by the first experimental measure in the event R2 (Figure 3), which is not captured by any of the modelling scenarios, and reveals either a limitation of the model structure or an erroneous measurement. Notice also that the number of field data points from which the coverage is computed varies within 19 for the event R3 and 6 for the event R4 and therefore, missing or hitting an additional data point implies a variation between 5% and 15% in the coverage. With these limitations in mind, the coverage ratios obtained with MS3 and MS6 (in all cases except one larger than 65%) can be considered very satisfactory in the context of soil erosion modelling field applications. It is remarkable that these two scenarios give a very similar level of agreement with the experimental data, since MS6 assumes a two layer soil structure with 5 parameters calibrated from 5000 Monte Carlo runs, while MS3 includes only 2 parameters calibrated from 250 Monte Carlo runs. The calibration process is therefore simpler and less computationally demanding in MS3.

The two scenarios which consider only the production of suspended sediment due to rainfall impact (MS1 and MS4) achieve also high NSE values in all the rainfall events (in general larger than 0.70), although they give in general lower coverage values (Table 5). Both scenarios show a systematic trend to underpredict the experimental observations (with PBIAS of the order of -20% in average) because they do not consider the production of sediment by runoff. At the same time the parameter α_{rd} is higher in MS1 than in MS3 (Table 6). This could indicate that the calibration of MS1 tries to compensate the fact of neglecting the runoff production by increasing the rainfall production. A noticeable point is that considering a two layer soil structure in MS4 does not improve its performance relative to MS1 in terms of NSE, PBIAS and coverage.

The two scenarios which give the worst performance levels are those which only consider runoff as the soil erosion mechanism (MS2 and MS5). Again, both scenarios have a very similar performance, despite considering one and two soil layers respectively. The fact of neglecting the rainfall driven production of suspended sediment prevents these model structures to reproduce multiple peaks in the sediment flux time series as the one which appears in R2 (Figure 3), which is clearly related to the peak in rainfall intensity and is well captured by all the other scenarios. The calibrated value of the parameter F is of the order of $2 \cdot 10^{-6}$. This is a very low value compared to results reported by other authors in laboratory studies (Hairsine and Rose, 1992b; Heng et al., 2011; Rose et al., 2007;

Sander et al., 2007), and it means that a very small fraction of the total stream power dissipated by bed friction is used for the entrainment and reentrainment of sediment particles in the water column. This is because most of the bed friction is due to the head losses induced by the micro-topography of the terrain, or even the vegetation, and only a very small fraction is caused by skin roughness. Since the total streampower is used in Equations (5-6), the parameter F has to account for the difference between total and skin roughness. In laboratory experiments this difference is in general small, but in field applications of overland flow over rough terrains it can be very relevant. As mentioned in previous sections, this is also related to the large values of the Manning coefficient reported in Table 2.

Regarding parameter uncertainty, the average value of α_{rd} and F can be clearly identified in the single layer scenarios MS1 and MS3 (Figure 4). On the other hand, in the two soil layer scenarios the spread on the value of these parameters after calibration increases in a significant way. The reason for that is illustrated in Figure 5, which shows the behavioural simulations obtained after calibrarion of MS4 with the event R2, plotted in parameter space. This figure shows that no modal value can be easily identified for α_{rd} in MS4, as it can be in MS1 and MS3. It also shows that when the value of $M_{s,cr}$ is below a certain threshold (which in this specific case is around 0.5 $\mathrm{kg/m^2}$) the model is insensitive to the erodibility of the cohesive layer (α_d), and model performance is determined only by the erodibility of the eroded layer (α_{rd}). This gives rise to a strong equifinality problem in the identification of the model parameters, which is further illustrated with the results shown in Figure 6, which represents the timeseries of the rainfall production terms in three different calibration runs of MS4. The first column of Figure 6 shows the suspended sediment production terms in a behavioural run with a very good performance (NSE=0.94) and a small value of $M_{s,cr}$ (0.048 $\mathrm{kg/m^2}$). In this simulation the mass of eroded soil (M_s) rapidly exceeds the value of $M_{s,cr}$. At that moment the cohesive layer becomes fully protected by the eroded layer and therefore, it does not contribute to the production of suspended sediment, the total production being equal to the production from the eroded layer (D_{rdrd}). The second column of Figure 6 represents the production terms in another behavioural run which also attains a very good performance (NSE=0.97), but in this case with a large value of $M_{s,cr}$ (2.5 $\mathrm{kg/m^2}$). In this case the average mass of sediment in the eroded layer built up during the whole event (M_s=0.17 $\mathrm{kg/m^2}$) is much lower than $M_{s,cr}$ and therefore, it is not enough to protect the cohesive layer. In this situation it is the production from the cohesive layer (parameter α_d) the most relevant process which determines the model performance. Therefore, the relevance of the parameters α_d and α_{rd} in the model is completely conditioned by the value of the parameter $M_{s,cr}$. Moreover, Figure 5 shows that in the scenario MS4 when $M_{s,cr}$ is lower than 0.5 $\mathrm{kg/m^2}$ the average value of α_{rd} is the same as that of α_d when $M_{s,cr}$ is larger than 1, and both are very similar to the average value of α_{rd} in the scenario MS1. This indicates that the scenarios MS1 and MS4 are equivalent for small or large values of $M_{s,cr}$, which explains the similar performance results reported in Table 5. The effect on model calibration of this sort of insensitivity to input parameters conditioned by the value of $M_{s,cr}$ is that the parameter distribution of α_d and α_{rd} extend over the whole parameter range. This is clearly reflected in the median and standard deviation of the calibrated parameters in MS4 (Table 6).

It is interesting to notice in Table 6 that the median value of the parameters F and α_{rd} is quite similar in the scenarios MS3 and MS6. This confirms that the most relevant processes which drive the sediment flux variability at the hillslope outlet during a single storm event are well represented in MS3 with just two parameters and a single soil layer structure. Although slightly higher, the value of α_{rd} in MS1 is also consistent with the two previous scenarios, which stresses the fact that this parameter alone is able to explain properly most of the sedimentary response of the hillslope in terms of mass flux variability.

The previous results are confirmed by the relative contribution in the scenarios MS3 and MS6 of rainfall driven production (via the terms D_{rdd} and D_{rdrd}) to the total suspended sediment production (Table 7). The fact that for each event the relative contribution of rainfall is similar in the scenarios MS3 and MS6 reinforces the conclusions concerning the strong equifinality between these two scenarios. Rainfall production represents at least 60% of the total production in all the events, achieving rates of 90% in the event R1. This is consistent with the fact that the scenarios MS2 and MS5, which do not consider rainfall production, obtain very low performance levels in the event R1 (Table 5). The performance of these scenarios is also low in the events R2 and R5, in which the relative contribution of rainfall is still significant (of the order of 70%).

The results obtained in calibration suggest that in our study site the most relevant sediment production process which determines the good or bad performance of a model scenario is rainfall impact. This is consistent with the results of the sensitivity analysis on synthetic data presented by Rousseau (2012), which suggest that rainfall erodibility is the parameter which explains most of the variability of the eroded mass at the hillslope outlet. Adding the reentrainment of sediment due to overland flow (MS3) improves model performance in the events R3, R4 and R5, which are the ones with highest peaks of runoff discharge. The scenario MS3 is therefore more versatile and would be more adequate than MS1 in other sites dominated by rill and gully erosion. Increasing further the complexity of the model by adding a two layer soil structure (MS4, MS5 and MS6) does not imply a significant improvement on model performance.

4.1.2. Validation

The predictive capability of the five model structure scenarios was analysed by modelling the rainfall events R2, R3, R4 and R5 with the parameters inferred from the calibration of the event R1, which are shown in Table 6. Model performance was evaluated in terms of the NSE, PBIAS and coverage of the 95% confidence interval.

According to the performance ratios reported in Table 8 the model structure with the best predictive capabilities is MS3. This modelling scenario gives NSE values larger than 0.6 and PBIAS values lower than 40%, with the only exception of event R4. But even in this case the coverage ratio (67%) is acceptable. In the events R2 and R3 the NSE and coverage results obtained with MS3 in validation are similar to those obtained in calibration (Table 5), while in R5 they are slightly worse, but still satisfactory. The scenarios MS1 and MS6 also produce high NSE values and low PBIAS ratios (again except in the event R4), but their coverage ratios are lower than those obtained with MS3. Excluding the event R4, the PBIAS results shown in Table 8 indicate that none of these three scenarios (MS1, MS3 and MS6) tend to systematically overpredict or underpredict the experimental observations. On the other hand MS2 and MS5 give very low NSE values and PBIAS ratios larger than 50% and cannot therefore be considered as satisfactory models. In addition the coverage ratios obtained with these scenarios are very low (in general lower than 30%) and the PBIAS is always negative, which indicates a systematic underestimation of the mass flux of sediment at the hillslope outlet.

The sediment flux time series predicted by the three scenarios MS1, MS3 and MS6 in the 4 validation events are shown in Figure 7. The differences between the median predictions of MS3 and MS6 are minimal at all time steps. Since the spread of the uncertainty bounds is slightly larger in the case of MS3, the coverage ratios obtained in this scenario are higher (Table 8). The suspended sediment flux computed with MS1 responds somewhat stronger to peaks in the rainfall intensity, since the calibrated rainfall erodibility coefficient (α_{rd}) is larger in this scenario (Table 6) because it has to account by itself alone for all the sediment production. Some measured points are clearly not captured by any scenario, for instance the first point in the event R2 or the third one in R5. This might be due to some process which is not captured by any of the sediment production formulations, or to exceptionally large errors in the rainfall or sediment flux data. Also the last four experimental points in the recession curve of event R5 are not captured by the 95% confidence interval of any model structure. The fact that the model predictions on the recession curve systematically underestimate the measured sediment flux, suggests that the flow driven production parameter (F) might be slightly underestimated in the calibration of the model. Nevertheless, this numerical-experimental disagreements are quite restrained if we look at the whole validation data shown in Figure 7, with the exception of the event R4, which is overestimated by all the scenarios.

We have not found a clear reason to explain why the model performs worse during validation with the event R4. The model results presented in section 4.1.1 do not show any significant difference in the contribution of rainfall and runoff driven erosion from one event to another. In all the events the performance of the model in calibration is good (as shown in Table 5) and the percentage contribution of the rainfall driven production terms ($D_{rdd} + D_{rdrd}$) is similar in R2, R3 and R4 (Table 7). The problem during validation with the event R4 might be related with a limitation on the mathematical representation of physical processes within the numerical model. As argued in Quinton et al. (2011), the reasons of finding non-overlapping parameter distributions when applying the GLUE methodology to different events in the same catchment might be related to errors in the model structure, errors in the input data and initial conditions, or real variations in the hydraulic and soil characteristics between events. Another possible reason could be added in our study case: the poor experimental representation of the sedimentograph measured during the event R4, which is defined by only 6 field data points, with only two of them during the main part of the hydrograph. The number of experimental measures in the other events is noticeably larger, ranging from 11 to 19. Experimental measurement errors will have a much higher impact on model performance when modelling the event R4.

4.2. *Implications for soil erosion model calibration and application*

The NSE, PBIAS and coverage performance ratios show that considering just two erodibility parameters which account for the production of suspended sediment due to rainfall and runoff, offers a good compromise between model performance and calibration efforts. If only one parameter should be retained, the most meaningful one in our study site is the rainfall erodibility coefficient, although in other sites in which rill or gully erosion dominates, the relative importance of runoff production on model output will probably be higher. The low variability on the value of the rainfall erodibility parameter from one event to another, and from one modelling scenario to another, points out the robustness of the calibration methodology and confirms the fact that this parameter is representing correctly the production of sediment in the hillslope. Nevertheless, considering just the rainfall erodibility coefficient (MS1) in rill erosion dominated sites would fail to give accurate predictions and therefore, the scenario MS3 should be in general preferred unless the modeller is sure about the dominant soil erosion processes in a specific study site.

While conceptually appealing, a two-layer soil structure is difficult to implement in field applications because detachment and redetachment are very interrelated processes which are difficult to isolate and therefore, to characterise by field or laboratory measurements. The fact that the rainfall erodibility parameters (α_{rd} and α_d) calibrated in the two-layer scenarios (MS4 and MS6) have the same value for the cohesive and eroded layers indicates that in our study case it is not necessary to consider a double layer structure. This is further confirmed by the fact that in all the modelling scenarios and events the calibrated values of the rainfall erodibility parameter are very similar. These results suggest that it might be reasonable in field applications of event-based soil erosion models to consider the erodibility properties homogeneous over the soil depth, without the need of distinguishing two layers of soil with different properties. A two layer soil structure makes the calibration process more complex and might be a constraint in the application of this type of models at larger scales. Even though the relevance of the sediment production processes and model parameters might vary from one study site to another, similar conclusions regarding the compromise between the structural complexity and efficiency of soil erosion models might apply to other hillslopes within the Mediterranean context and for similar land use.

As previously said, the average median sediment diameter measured at the outlet of the studied hillslope during the 5 storm events was 25 $\mu\mathrm{m}$. However, the median diameter measured from the suspended sediment samples varied during the same event (mostly between 10 and 50 $\mu\mathrm{m}$), which could introduce a relevant source of uncertainty when computing its average value during a whole event from a few soil samples. In addition, the size of particles is not always available from field measurements. The soil texture is used in that cases to estimate the characteristic sediment diameter, even if it does not guarantee that the value obtained is the most representative of the eroded particles. Considering these sources of uncertainty, the sediment diameter might be considered as a calibration parameter in an attempt to improve model performance and to define its characteristic value for modelling purposes more precisely. For this purpose we have included the sediment diameter as an additional parameter in the scenarios MS1 and MS3, which are those showing a better performance with just a single soil layer. The calibration procedure of these *augmented* scenarios was the same as the one described in previous sections, using in this case 500 and 1000 Monte Carlo runs respectively for MS1 and MS3. The sediment diameter was varied between 10 and 50 $\mu\mathrm{m}$, which corresponds to settling velocities of 0.054 and 1.4 mm/s respectively.

The results obtained after calibration and validation of these new *augmented* scenarios (Table 9) show that model performance is not improved when using the sediment diameter as an additional calibration parameter. The NSE performance ratios shown in Tables 5 and 9 for the scenarios MS1 and MS3 are almost identical. On the other hand, the sensitivity of the erodibility parameters to the sediment diameter is very high, as shown by the plot of the sediment diameter against the rainfall erodibility parameter (Figure 8). When including the sediment diameter as a model parameter there is a whole set of equifinal behavioural simulations which expand over the whole range of variation of the erodibility parameters. A representative average value of the parameters α_{rd} and F can no longer be identified in Figure 8, as it was possible in MS3 (Figure 4), and the marginal parameter distribution of the erodibility parameters becomes almost flat. This implies that one of the three parameters (α_{rd}, F, D_s) should be known in order to calibrate the others correctly. Given that no easily achievable measurements exist to characterise α_{rd} and F, measurements efforts should focus on the settling velocity (or other size related properties) of eroded particles.

5. Conclusions

The performance of different simplified parameterisations of the production terms in a distributed event-based soil erosion model has been analysed in an agricultural hillslope during five storm events. Model performance was evaluated in terms of the sediment mass flux at the hillslope outlet. Calibration and validation was performed within the GLUE methodology in order to account for the uncertainties inherent to soil erosion modelling. The results show the capabilities in terms of model calibration and validation of the GLUE framework in soil erosion studies, an area in which accounting for modelling uncertainties is of paramount importance given the complexity of the physical processes which are being modelled and the scarcity of accurate field data available for calibration.

The results presented show that in our study site a model structure considering a single soil layer with just two erodibility parameters accounting for the production of suspended sediment due to rainfall impacts and runoff shear offers a good compromise between calibration efforts and model performance. A two layer soil structure makes the calibration process more complex without improving significantly model performance, while it might be a constraint in the application of these type of models at larger scales. In cases in which the modeller is sure about the dominant soil erosion processes a single parameter (in our study site the rainfall erodibility coefficient) can offer appropriate numerical predictions. It should be noticed that the number of simulations needed to account for uncertainty on model output, and consequentely the total amount of computer time required to perform the Monte Carlo runs within GLUE, increases greatly with the number of model parameters. It seems thus reasonable to diminish the number of input parameters as long as the remaining model structure provides an appropriate mathematical representation of the physical processes involved in soil erosion in the study site, especially if one of the next objectives is to apply the model to larger scales (e.g. small catchments). This decision should be made by the modeller based on his expertise, on the expected accuracy on model output and on his knowledge about the hydrological and sedimentological properties of the study site to be modelled.

Even though including the sediment diameter as a calibration parameter does not improve model performance, model calibration is very sensitive to the chosen characteristic particle size, due to the strong interaction between the sediment diameter and the calibrated soil erodibility parameters. However, the precise definition of an effective sediment diameter is still one of the biggest unknowns in soil erosion modelling for a number of reasons as soil aggregation, aggregate stability, change in aggregate size due to the stresses induced by rain drop impact and overland flow transport, among others. This poses a relevant equifinality problem when trying to obtain representative values of the soil erodibility parameters from model calibration.

The methodology and results presented here should incite numerical modellers to incorporate model uncertainty in soil erosion studies, as it has been strongly encouraged by other authors which have already been cited throughout this paper. A further step, not considered in this study, would be to incorporate data uncertainty in the analysis in order to account for observational errors, as it is done for example in the limits of acceptability approach.

Acknowledgments

Field data were obtained within the HyMeX program and sponsored by Rhône-alpes région, Grants MISTRALS/HyMeX, ANR-2011-BS56-027 FLOODSCALE project and OHMCV (http://www.ohmcv.fr). OHMCV is supported by the Institut National des Sciences de l'Univers (INSU/CNRS), the French Ministry for Education and Research, the Observatoire des Sciences de lŠUnivers de Grenoble (OSUG/Grenoble University) and the SOERE Réseau des Bassins Versants (Alliance Allenvi). Data collection and sediment concentration measurements were carried out by Simon Gérard, who had a HyMeX CNRS contract. The assistance of Jean Marie Miscioscia and Henri Mora in building and setting up the flume on the field is acknowledged. The help of Romain Biron and Fred Malinur during field measurements was appreciated.

References

Ascough, J., Baffaut, C., Nearing, M., and Liu, B. (1997). The WEPP watershed model: I. Hydrology and erosion. *Transactions of the ASAE*, 40(4):921–933.

Beuselinck, L., Govers, G., Hairsine, P. B., Sander, G. C., and Breynaert, M. (2002). The influence of rainfall on sediment transport by overland flow over areas of net deposition. *Journal of Hydrology*, 257:145–163.

Beuselinck, L., Govers, G., Steegen, A., Hairsine, P. B., and Poesen, J. (1999). Evaluation of the simple settling theory for predicting sediment deposition by overland flow. *Earth Surface Processes and Landforms*, 24:993–1007.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1):18–36.

Beven, K. J. and Binley, A. M. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6:279–298.

Blazkova, S. and Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, 45(12).

Boudevillain, B., Delrieu, G., Galabertier, B., Bonnifait, L., Bouilloud, L., Kirstetter, P.-E., and Mosini, M.-L. (2011). The Cévennes-Vivarais Mediterranean Hydrometeorological Observatory database. *Water Resources Research*, 47(7).

Braud, I., Ayral, P.-A., Bouvier, C., Branger, F., Delrieu, G., Le Coz, J., Nord, G., Vandervaere, J.-P., Anquetin, S., Adamovic, M., et al. (2014). Multi-scale hydrometeorological observation and modelling for flash flood understanding. *Hydrology and Earth System Sciences*, 18:3733–3761.

Braud, I. and Vandervaere, J. (2012). Analysis of infiltration tests and performed in the claduègne catchment in May-June 2012. http://mistrals.sedoo.fr/?editDatsId=1321&datsId=1321&projectpacename=HyMeX&q=cladu%C3%A

Brazier, R. E., Beven, K. J., Freer, J., and Rowan, J. S. (2000). Equifinality and uncertainty in physically based soil erosion models: Application of the GLUE methodology to WEPP–the Water Erosion Prediction Project–for sites in the UK and USA. *Earth Surface Processes and Landforms*, 25(8):825–845.

Cea, L., Garrido, M., and Puertas, J. (2010). Experimental validation of two-dimensional depth-averaged models for forecasting rainfall-runoff from precipitation data in urban areas. *Journal of Hydrology*, 382:88–102.

Cea, L., Legout, C., Darboux, F., Esteves, M., and Nord, G. (2014). Experimental validation of a 2D overland flow model using high resolution water depth and velocity data. *Journal of Hydrology*, 513:142–153.

Cea, L. and Vázquez-Cendón, M. E. (2012). Unstructured finite volume discretisation of bed friction and convective flux in solute transport models linked to the shallow water equations. *Journal of Computational Physics*, 231:3317–3339.

Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., Saby, N., Gobin, A., Vacca, A., Quinton, J., Auerswald, K., et al. (2010). Rates and spatial variations of soil erosion in europe: a study based on erosion plot data. *Geomorphology*, 122(1):167–177.

Courant, R., Friedrichs, K., and Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal*.

Dunne, T. and Black, R. D. (1970). Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research*, 6:1296–1311.

Engman, E. T. (1986). Roughness coefficients for routing surface runoff. *Journal of Irrigation and Drainage Engineering*, 112(1):39–53.

Favis-Mortlock, D. T., Boardman, J., Parsons, A. J., and Lascelles, B. (2000). Emergence and erosion: A model for rill initiation and development. *Hydrological Processes*, 14(11-12):2173–2205.

Foster, G. R., Flanagan, D. C., Nearing, M. A., Lane, L. J., Risse, L. M., and Finkner, S. C. (1995). *Water Erosion Prediction Project: Hillslope Profile* *and Watershed Model Documentation*, chapter Hillslope erosion component. USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, Indiana.

Fraga, I., Cea, L., and Puertas, J. (2013). Experimental study of the water depth and rainfall intensity effects on the bed roughness coefficient used in distributed urban drainage models. *Journal of Hydrology*, 505:266–275.

Gao, B., Walter, M. T., Steenhuis, T. S., Parlange, J. Y., Nakano, K., Hogarth, W. L., and Rose, C. W. (2003). Investigating ponding depth and soil detachability for a mechanistic erosion model using a simple experiment. *Journal of Hydrology*, 277(1-2):116–124.

Govers, G. (1992). *Overland flow, Hydraulics and Erosion Mechanics*, chapter Evaluation of transporting capacity formulae for overland flow. University College London Press.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2):135–143.

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

Hairsine, P. B. and Rose, C. W. (1992b). Modeling water erosion due to overland-flow using physical principles. 2. Rill flow. *Water Resources Research*, 28(1):245–250.

Hantush, M. M. and Kalin, L. (2005). Uncertainty and sensitivity analysis of runoff and sediment yield in a small agricultural watershed with KINEROS2. *Hydrological sciences journal*, 50(6).

Heng, B., Sander, G. C., Armstrong, A., Quinton, J., Chandler, J. H., and Scott, C. F. (2011). Modeling the dynamics of soil erosion and size-selective sediment transport over nonuniform topography in flume-scale experiments. *Water Resources Research*, 47(2).

Huza, J., Teuling, A. J., Braud, I., Grazioli, J., Melsen, L. A., Nord, G., Raupach, T. H., and Uijlenhoet, R. (2014). Precipitation, soil moisture and runoff variability in a small river catchment (ardèche, france) during hymex special observation period 1. *Journal of Hydrology*, 516:330–342.

Jomaa, S., Barry, D. A., Brovelli, A., Sander, G. C., Parlange, J. Y., Heng, B. C. P., and van Meerveld, H. J. T. (2010). Effect of raindrop splash and transversal width on soil erosion: Laboratory flume experiments and analysis with the Hairsine-Rose model. *Journal of Hydrology*, 395:117–132.

Kinnell, P. I. A. (1990). The mechanics of raindrop induced flow transport. *Australian Journal of Soil Research*, 28:497–516.

Kinnell, P. I. A. (2005). Raindrop-impact-induced erosion processes and prediction: A review. *Hydrological Processes*, 19:2815–2844.

Krueger, T., Quinton, J., Freer, J., Macleod, C., Bilotta, G., Brazier, R., Hawkins, J., and Haygarth, P. (2012). Comparing empirical models for sediment and phosphorus transfer from soils to water at field and catchment scale under data uncertainty. *European Journal of Soil Science*, 63(2):211–223.

Laloy, E. and Bielders, C. (2009). Modelling intercrop management impact on runoff and erosion in a continuous maize cropping system: Part I. Model description, global sensitivity analysis and Bayesian estimation of parameter identifiability. *European journal of soil science*, 60(6):1005–1021.

Lane, S. N. (2014). Acting, predicting and intervening in a socio-hydrological world. *Hydrology and Earth System Sciences*, 18(3):927–952.

Liu, Y., Freer, J., Beven, K., and Matgen, P. (2009). Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. *Journal of Hydrology*, 367(1):93–103.

Misra, R. and Rose, C. (1995). An examination of the relationship between erodibility parameters and soil strength. *Soil Research*, 33(4):715–732.

Morgan, R., Quinton, J., Smith, R., Govers, G., Poesen, J., Auerswald, K., Chisci, G., Torri, D., and Styczen, M. (1998). The European Soil Erosion Model (EUROSEM): A dynamic approach for predicting sediment transport from fields and small catchments. *Earth Surface Processes and Landforms*, 23(6):527–544.

Mugler, C., Planchon, O., Patin, J., Weill, S., Silvera, N., Richard, P., and Mouche, E. (2011). Comparison of roughness models to simulate overland flow and tracer transport experiments under simulated rainfall at plot scale. *Journal of Hydrology*, 402:25–40.

Muñoz-Carpena, R., Parsons, J. E., and Gilliam, J. W. (1999). Modeling hydrology and sediment transport in vegetative filter strips. *Journal of Hydrology*, 214(1):111–129.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I - A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Nord, G. and Esteves, M. (2005). PSEM 2D: A physically based model of erosion processes at the plot scale. *Water Resources Research*, 41(8).

Nord, G. and Esteves, M. (2007). Evaluation of sediment transport formulae and detachment parameters in eroding rills using PSEM2D and the Water Erosion Prediction Project (WEPP) database. *Water Resources Research*, 43(W08420):1–14.

Owens, P., Batalla, R., Collins, A., Gomez, B., Hicks, D., Horowitz, A., Kondolf, G., Marden, M., Page, M., Peacock, D., et al. (2005). Fine-grained sediment in river systems: Environmental significance and management issues. *River Research and Applications*, 21(7):693–717.

Pimentel, D. (2006). Soil erosion: A food and environmental threat. *Environment, development and sustainability*, 8(1):119–137.

Proffitt, A., Hairsine, P., and Rose, C. (1993). Modeling soil erosion by overland flow: Application over a range of hydraulic conditions. *Transactions of the ASAE (USA)*, 36:1743–1753.

Quinton, J., Krueger, T., Freer, J., Bilotta, G., and Brazier, R. (2011). A case study of uncertainty: applying glue to eurosem. *Handbook of erosion modelling. Blackwell Publishing Ltd, Chichester*, pages 80–97.

Rose, C. W., Yu, B., Ghadiri, H., Asadi, H., Parlange, J., Hogarth, W., and Hussein, J. (2007). Dynamic erosion of soil in steady sheet flow. *Journal of Hydrology*, 333(2):449–458.

Rousseau, M. (2012). *Propagation d'incertitudes et analyse de sensibilité pour la modélisation de l'infiltration et de l'érosion*. PhD thesis, Université Paris-Est.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons.

Sander, G., Parlange, J.-Y., Barry, D., Parlange, M., and Hogarth, W. (2007). Limitation of the transport capacity approach in sediment transport modeling. *Water Resources Research*, 43(2).

Sharma, P. P., Gupta, S. C., and Foster, G. R. (1993). Predicting soil detachment by raindrops. *Soil Science Society of America Journal*, 57:674–680.

Sharma, P. P., Gupta, S. C., and Foster, G. R. (1995). Raindrop-induced soil detachment and sediment transport from interrill areas. *Soil Science Society of America Journal*, 59:727–734.

Shaw, S. B., Parlange, J. Y., Lebowitz, M., and Walter, M. T. (2009). Accounting for surface roughness in a physically-based urban wash-off model. *Journal of Hydrology*, 367:79–85.

Shaw, S. B., Walter, M. T., and Steenhuis, T. S. (2006). A physical model of particulate wash-off from rough impervious surfaces. *Journal of Hydrology*, 327:618–626.

Smith, R., Goodrich, D., and Quinton, J. N. (1995). Dynamic, distributed simulation of watershed erosion: The KINEROS2 and EUROSEM models. *Journal of Soil and Water Conservation*, 50(5):517–520.

Sobol, I. M. (1998). On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation*, 47:103–112.

Tatard, L., Planchon, O., Wainwright, J., Nord, G., Favis-Mortlock, D., Silvera, N., Ribolzi, O., Esteves, M., and Huang, C. H. (2008). Measurement and modelling of high-resolution flow-velocity data under simulated rainfall on a low-slope sandy soil. *Journal of Hydrology*, 348(1-2):1 – 12.

van Rijn, L. C. (1984). Sediment transport, Part II: Suspended load transport. *J. Hydraul. Eng.*, 110(11):1613–1641.

Veihe, A. and Quinton, J. (2000). Sensitivity analysis of EUROSEM using Monte Carlo simulation I: Hydrological, soil and vegetation parameters. *Hydrological Processes*, 14(5):915–926.

Vrugt, J. A., Ter Braak, C. J., Gupta, H. V., and Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23(7):1011–1026.

Wilson, C. and Horritt, M. (2002). Measuring the flow resistance of submerged grass. *Hydrological Processes*, 16(13):2589–2598.

Table 1: Characteristics of the five storm runoff events.

Event	Start	Rain before runoff starts (mm)	Time before runoff starts (h)	Rain since runoff starts (mm)	Max. 1 min rain intensity (mm/h)	Runoff duration (h)	Runoff depth (mm)	Q_{max} (l/s)
R1	09/11/2012 22:00	43	14.3	22	24	10.0	12	0.30
R2	04/11/2011 12:00	113	45.0	16	79	3.9	17	0.98
R3	18/05/2013 08:00	19	7.0	27	80	5.0	29	1.73
R4	07/09/2010 19:00	91	24.6	61	92	2.7	12	1.12
R5	20/10/2013 06:00	20	5.0	44	92	2.6	29	1.35

Table 2: Value of the hydraulic parameters after calibration of each storm runoff event.

Rainfall event	n ($sm^{-1/3}$)	I_a (mm)	k_s (mm/h)	NSE (-)
R1	0.60	43	1.8	0.97
R2	0.30	113	1.4	0.98
R3	0.20	19	0.0	0.99
R4	0.40	91	18.0	0.99
R5	0.80	20	0.9	0.99

Table 3: Model structure scenarios of the soil erosion model.

Modelling Scenario	Rainfall production	Runoff production	Number of layers	Model parameters	MC runs for calibration
MS1	Yes	No	1	α_{rd}	100
MS2	No	Yes	1	F	100
MS3	Yes	Yes	1	α_{rd},F	250
MS4	Yes	No	2	α_{rd},α_d,$M_{s,cr}$	1000
MS5	No	Yes	2	F,J,$M_{s,cr}$	1000
MS6	Yes	Yes	2	α_{rd},α_d,F,J,$M_{s,cr}$	5000

Table 4: Parameter ranges used in the definition of the prior parameter distribution for calibration purposes.

Parameter	α_{rd}	α_d	F	J	$M_{s,cr}$
Units	kg/m^3	kg/m^3	-	J/kg	kg/m^2
Sampling range	[0-50]	[0-50]	[0-0.001]	[1-10]	[0-2.8]

Table 5: NSE, PBIAS and coverage performance ratios obtained after calibration of the six modelling scenarios. The NSE and PBIAS are computed from the median of the output sediment flux cdf. The coverage is computed as the % of experimental measures lying within the 95% confidence interval computed from the model output.

NSE						
Event	MS1	MS2	MS3	MS4	MS5	MS6
R1	0.72	0.07	0.57	0.78	0.21	0.56
R2	0.98	0.32	0.97	0.98	0.26	0.96
R3	0.67	0.91	0.91	0.66	0.93	0.94
R4	0.75	0.90	0.90	0.86	0.83	0.91
R5	0.78	0.64	0.86	0.77	0.65	0.85
PBIAS						
Event	MS1	MS2	MS3	MS4	MS5	MS6
R1	-4.45	-17.9	1.41	0.04	-12.9	-0.8
R2	-13.68	0.47	9.7	-7.7	-15.2	7.5
R3	-25.8	2.54	11.4	-16.7	5.7	5.7
R4	-33.9	-0.24	-6	-18.8	-4.8	-4.0
R5	-18.9	-13.75	-6.7	-14.9	-1.9	0.4
Coverage 95 (%)						
Event	MS1	MS2	MS3	MS4	MS5	MS6
R1	73	18	73	73	33	82
R2	47	41	65	65	18	60
R3	56	72	72	61	89	89
R4	50	50	67	67	33	67
R5	33	42	75	42	67	75

Table 6: Model parameters (median $\pm$ standard deviation) for each modelling scenario after calibration for the event R1.

	MS1	MS2	MS3	MS4	MS5	MS6
$M_{s,cr}$ x10^{-3}	-	-	-	910 $\pm$ 840	820 $\pm$ 339	710 $\pm$ 382
F x10^{-6}	-	2.5 $\pm$ 0.3	1.0 $\pm$ 0.9	-	51.0 $\pm$ 24.0	4.8 $\pm$ 5.5
J	-	-	-	-	6.5 $\pm$ 2.8	6.9 $\pm$ 2.5
α_d	-	-	-	20.1 $\pm$ 20.3	-	11.5 $\pm$ 9.2
α_{rd}	17.6 $\pm$ 5.0	-	12.9 $\pm$ 6.4	18.5 $\pm$ 25.9	-	12.1 $\pm$ 13.1

Table 7: Contribution of the rainfall driven production terms ($D_{rdd} + D_{rdrd}$ in Equation (2)) to the gross erosion. Only the scenarios MS3 and MS6 are considered since they are the only ones which account simultaneously for rainfall and runoff production.

Scenario	R1	R2	R3	R4	R5
MS3	88%	67%	60%	60%	77%
MS6	92%	70%	61%	71%	69%

Table 8: NSE, PBIAS and coverage performance ratios in validation, using the parameters obtained from the calibration of event R1. The coverage is computed as the % of experimental measures lying within the 95% confidence interval computed from the model output.

			NSE			
Event	MS1	MS2	MS3	MS4	MS5	MS6
R2	0.90	0.14	0.97	0.89	0.04	0.98
R3	0.60	0.92	0.81	0.58	0.91	0.72
R4	-10.80	0.32	-3.60	-11.00	0.05	-4.9
R5	0.78	-0.49	0.69	0.77	-0.21	0.77
			PBIAS			
Event	MS1	MS2	MS3	MS4	MS5	MS6
R2	17.2	-55.9	-0.8	18.9	-68.1	5.6
R3	0.7	2.5	14.8	4.9	-8.5	23.5
R4	213.8	-56.3	131.3	218	-67.1	1.5
R5	-14.4	-85.3	-35.7	-11.5	-74.2	-27.0
			Coverage 95 (%)			
Event	MS1	MS2	MS3	MS4	MS5	MS6
R2	59	0	65	71	6	59
R3	44	6	67	67	39	72
R4	17	17	67	33	17	33
R5	33	0	42	42	0	42

Table 9: NSE in validation and calibration for the scenarios MS1 and MS3 *augmented* with the sediment diameter as an additional calibration parameter.

	NSE calibration		NSE validation	
Event	MS1 + Ds	MS3 + Ds	MS1 + Ds	MS3 + Ds
R1	0.72	0.56	-	-
R2	0.97	0.95	0.88	0.91
R3	0.63	0.92	0.59	0.88
R4	0.79	0.86	-12.80	-2.50
R5	0.72	0.79	0.66	0.51

Figure 1: Location of the study site, finite volume mesh used in the numerical simulations, hillslope topography and typical water depth pattern computed during the storm runoff events.

Figure 2: Measured and computed hydrographs at the hillslope outlet during the five rainfall events. Values in the x-axis refer to the time passed since the beginning of the storm event. Notice that there is a time lag between the beginning of the rainfall and the start of the surface runoff due to the initial abstraction.

Figure 3: Comparison of sediment flux median predictions and 95% confidence intervals obtained in the calibration of the rainfall events R1 (left), R2 (middle) and R4 (right) with model structures (from top to bottom) MS1, MS2, MS3 and MS6. Values in the x-axis refer to the time passed since the beginning of the storm event.

Figure 4: Distribution of the parameter sets used in the behavioural simulations after calibration of the event R2 with the modelling scenarios MS1 (left) and MS3 (middle and right). On the right plot (MS3) each dot represents a behavioural simulation, the size and colour of the dot being proportional to the NSE.

Figure 5: Posteriori distribution of the parameter α_{rd} and behavioural simulations plotted in parameter space, after calibration of the event R2 with the modelling scenario MS4. In the middle and right plots each dot represents a behavioural simulation, the size and colour of the dot being proportional to the NSE.

Figure 6: Time series of the rainfall production terms ($D_{tot} = D_{rdd} + D_{rdrd}$, D_{rdd} and D_{rdrd}) and sediment mass in the eroded layer (M_s) in 3 behavioural simulations computed during the calibration of the modelling scenario MS4 with the storm event R2. Each column corresponds to one simulation. The model parameters and NSE of each simulation are indicated in the first row.

Figure 7: Comparison of sediment flux median predictions and 95% confidence intervals obtained with model structures MS1 (left), MS3 (middle) and MS6 (right). Model parameters calibrated for the event R1. From top to bottom, events R2, R3, R4 and R5. Values in the x-axis refer to the time passed since the beginning of the storm event.

Figure 8: Behavioural simulations after calibration of the event R2 with the modelling scenario MS3 considering the characteristic particle size (D_s) as an additional calibration parameter. Each dot represents a behavioural simulation, the size and colour of the dot being proportional to the NSE.