Study of overland flow with uncertain infiltration using stochastic tools

Marie Rousseau, Olivier Cerdan, Alexandre Ern, Olivier Le Maitre, Pierre Sochala

Study of overland flow with uncertain infiltration using stochastic tools

M. Rousseau[a,b], O. Cerdan[b], A. Ern[a], O. Le Maître[c], P. Sochala[b]

[a] *Université Paris-Est, Cermics, Ecole des Ponts ParisTech, 77455 Marne la Vallée Cedex 2, France*
[b] *BRGM, RIS, 45060 Orléans Cedex 2, France*
[c] *LIMSI-CNRS, 91403 Orsay, France*

Abstract

The saturated hydraulic conductivity is one of the key parameters in the modelling of overland flow water fluxes. In this study, this parameter is defined as a stochastic parameter, idealized as a piecewise constant random field with uniform distribution. This paper aims at investigating the effects of the spatial and temporal scales in uncertainty propagation within overland flow models, and at identifying the localization of the most influential saturated hydraulic conductivity using sensitivity analysis. The results show that the influence of saturated hydraulic conductivity depends on the soil saturation and its spatial localization. For instance, in case of low saturated soils, the most influent parameter is the one located downslope, whereas in case of high saturated soils, the most influent one is either the most infiltrating or the intermediate one. The results indicate where efforts should be concentrate when collecting input parameters to reduce modelling uncertainties.

1. Introduction

Water fluxes are a fundamental part of natural ecosystems and are essential to support human activities. Many research efforts are therefore devoted to the development and application of physically-based models able to improve our understanding and modelling of these fluxes. One of the main obstacles to the application of such models is the difficulty to describe the spatial and temporal (non-linear) variability of the input parameters [24]. Indeed, the performance of models directly depends on the validity of the input parameters. Even if the technological progress in sensor development regularly improves the resolution at which we are able to measure the different natural and anthropogenic factors [3, 13], we are still far from being able to capture all their spatial and temporal variability. In recent years, many efforts have been undertaken to evaluate the rainfall input through the development and implementation of rainfall radars [30]. Furthermore, several plant growth models, such as the Soil Vegetation Atmosphere Transfer scheme (SVAT) [6], permit to determine operationally input parameters related to vegetation with a reasonable accuracy. It is more difficult to estimate the soil parameters, principally because of their heterogeneity and their high variability in space and time. For rainfall-runoff prediction models, numerous studies show that the saturated hydraulic conductivity, which is deduced from soil properties, is the most influent input parameter [11, 31].

The saturated hydraulic conductivity, herein denoted by K_s, provides a quantitative measure of the soil ability to transmit water. Indeed, K_s is one of the key parameters in the infiltration process and in water transfer through the unsaturated and saturated parts of the soil. The parameter K_s yields the maximum value of the infiltration rate, which is obtained for a saturated state of the soil, and influences predominantly the infiltration capacity [11]. Different methodologies have been elaborated to measure directly saturated hydraulic

Email addresses: m.rousseau@brgm.fr (M. Rousseau), o.cerdan@brgm.fr (O. Cerdan), ern@cermics.enpc.fr (A. Ern), olm@limsi.fr (O. Le Maître), p.sochala@brgm.fr (P. Sochala)

conductivities. However, the obtained values for K_s depend on the chosen methodology and most importantly, the spatial representativity of these measurements remains rather limited [43]. In most model applications, values for K_s are estimated through the application of pedo-transfer functions (PTF) using basic soil properties [8, 9, 37]. By testing different PTF's to predict K_s, it was concluded [35] that predicting K_s using a PTF is not always accurate owing to the inherent variability of K_s. Furthermore, using a set of data to compare different measurement and estimation methods, a high variability of K_s (more than 79%) has been observed [4]. To overcome this lack of accuracy, a possible approach consists in calibrating parameters, but the resulting values are often valid only for the used configuration and moderate variations thereof.

An alternative approach already suggested in [36] consists in considering K_s as a stochastic parameter instead of being estimated by deterministic approaches. It is today well admitted that probabilistic modelling provides efficient means to quantify parameter uncertainty. Uncertainty Analysis (UA) considers the uncertain parameters of a model as random objects, and the objective is to compute or characterize the induced variability in the model solution or in quantities of interest. For highly uncertain data within non-linear models, as in hydrology applications, so-called global UA methods, which study the effects of all the input parameters simultaneously, are needed. One essential step in UA is the definition of a random model for the uncertain parameters. Random models with different levels of complexity can be considered. For instance, a relatively simple approach is the Generalized Likelihood Uncertainty Estimation (GLUE) procedure [5] which is a Monte Carlo (MC) method generating a high number of parameter sets to compare the predicted model responses with observed responses and to accept or not some simulations through some chosen likelihood measure. Being a Bayesian approach, this likelihood measure can be updated for each new set of observed responses. Numerous studies are based on a Bayesian framework [23, 25, 26, 34]. Bayesian statistics mean that input parameters are considered as probabilistic variables having a joint posterior probability density function (pdf). Different methods exist for sampling posterior pdf's. The Monte Carlo Markov Chain (MCMC) sampler is often used in hydrology models, the earliest general (and most popular) method being the Metropolis–Hastings algorithm [21, 29]. In hydrology, various recent studies have aimed at improving MCMC samplers: the Shuffled Complex Evolution (SCE) Metropolis algorithm (SCEM) [39], which is a modified version of the SCE global optimization algorithm [14], or the Differential Evolution Adaptive Metropolis algorithm (DREAM) [40]. In contrast with Bayesian statistics, an alternative approach providing a complete probabilistic description considers the unknown parameters as random variables described by a fixed pdf. This approach, which is more adapted to problems where little data is available, is undertaken in the present work. Once the probabilistic framework is settled, it remains to characterize the model output variability in terms of input variability. To this purpose, we rely on MC methods since such methods provide an effective and robust methodology to generate a sample set of model solutions by sampling input parameters. We mention another recent methodology based on stochastic spectral methods [18]. The advantage is that a more complete probabilistic description of model output is achieved, but the methodology becomes computationally demanding when the input parameters are described by a large number of random variables.

In the present paper, we focus on the impact of the variability in the saturated hydraulic conductivity K_s on overland flows produced by various rainfall events in various geometries. A general probabilistic description of the saturated hydraulic conductivity is to model it as a random field. Although very rich, this type of model needs a substantial amount of information for its description, and is, therefore, not well adapted to the present setting. Moreover, extracting simple information in view of practical hydrology purposes from complex probabilistic models is an intricate issue. For these reasons, we rely on simpler probabilistic models where the saturated hydraulic conductivity is idealized as a piecewise constant random field over distinct portions of the simulation domain which are specified a priori. Each spatial subdomain is characterized by a unique Random Variable (RV) yielding the corresponding saturated hydraulic conductivity in the subdomain. Specifically, the random conductivity

field $K_s(x,\theta)$, where θ is a random event, is written as

$$K_s(x,\theta) = \sum_i \mathbb{1}_{\Omega_i}(x) K_{s,i}(\theta), \tag{1}$$

where $\mathbb{1}_{\Omega_i}$ is the indicator function of the i-th spatial subdomain and $\{K_{s,i}(\theta)\}$ is a set of (positive) real-valued random variables. In addition, the saturated hydraulic conductivity within a subdomain is assumed to be statistically independent from the others. As a result, the uncertain hydraulic conductivity field is modelled using a finite set of independent random variables, whose cardinality is equal to the number of subdomains considered in the simulation, and the joint probability density function (pdf) of the random variables $K_{s,i}$ factorizes in the form

$$p(\{K_{s,i}\}) = \prod_i p_i(K_{s,i}). \tag{2}$$

This idealization of the actual conductivity field is motivated by the physical reality. If one thinks of subdomains as homogeneous agricultural plots (or parcels of vegetation), the variability of K_s inside the parcel is usually negligible compared to the variability from a parcel to another because of the homogenization created by agricultural practices. Moreover, the present model can be subsequently refined by introducing inner variability within the parcels if additional information on soil properties within parcels is available.

The objective of this work is twofold. Firstly we consider test cases involving different spatial and temporal scales so as to investigate the effect of these scales in uncertainty propagation within overland flow models. Indeed, one salient result discussed hereafter is that the most influential input parameter on model output variability depends on the spatial and temporal scales of the processes of interest. Secondly, with an eye toward practical hydrology issues, we use sensitivity analysis tools as a means to identify the most influential input parameters. This information can be, for instance, valuable to decide on where to concentrate additional measurement efforts to improve field knowledge. Moreover, within a given test case, we consider various possibilities for the spatial organization of the parcels so as to study the effect of this organization on model output variability. One interesting result in this direction, relevant to landscape management issues, is the comparison presented hereafter between three grass strips distributed evenly within a field and a single grass strip located near the field outlet. Finally, we observe that the present work is not dedicated to a specific field study, but the purpose is instead rather general and aims at assessing the possibilities offered by the present methodology to improve our understanding of variability in overland flows due to infiltration. For this reason, we focus on a simple two-dimensional setting where the flow is described by a one-dimensional shallow-water flow model including friction and infiltration, the latter being described by the Green–Ampt model. More elaborate flow models can be considered. We also mention a different approach [41, 42] to compute pdf's of infiltration rates and infiltration depths.

This paper is organized as follows. Section 2 briefly describes the rainfall-runoff model and the numerical method used in the deterministic overland flow simulations once values for the random input parameters are specified. Section 3 introduces the stochastic approach and the statistic tools used to propagate and analyze the uncertainties in model output. Section 4 presents the two test cases designed to evaluate the impact of uncertainties in K_s and of the spatial localization of these uncertainties on overland flow. Results are discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. The setting

In this section, we present the physical model and its numerical resolution.

2.1. Physical model

We are concerned with overland flows where the water depth is much smaller than the characteristic horizontal size of the field of study (see Figure 1). Such flows can be de-

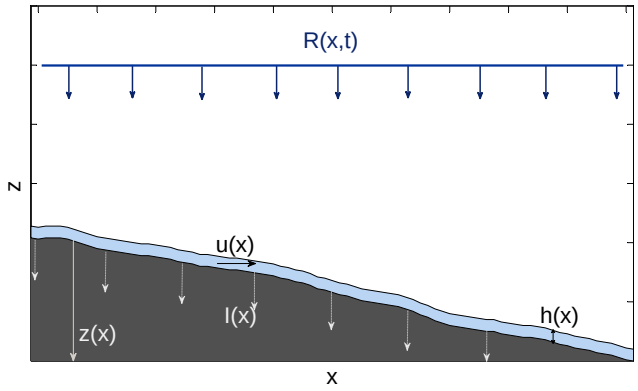


Figure 1: Geometric configuration and basic notation.

scribed by the 2D shallow water (SW) equations which are obtained from the 3D incompressible free-surface Navier–Stokes equations by averaging on the vertical direction under some simplifying assumptions, in particular hydrostatic pressure and negligible vertical velocity [17, 33, 38]. Neglecting also the flow transverse to the main slope direction, we obtain the 1D SW equations which express mass and momentum conservation as follows:

$$\partial_t h + \partial_x (hu) = R - I, \tag{3}$$

$$\partial_t (hu) + \partial_x \left(hu^2 + \tfrac{1}{2} g h^2\right) = -gh \left(\partial_x z + S_\mathrm{f}\right), \tag{4}$$

where h [L] is the water depth, u [L/T] the depth-averaged velocity, z [L] the ground surface elevation, and g [L/T^2] the gravitational constant (where L and T denote length and time units, respectively). The source term $R - I$ [L/T] corresponds to the difference between the rainfall rate R and the infiltration rate I. The quantity S_f [L/L] accounts for friction effects. The value of S_f depends on the properties of the soil surface and can be estimated from calibration or published values. Darcy–Weisbach's formula is often used [12, 15, 16, 38]:

$$S_\mathrm{f} = f \frac{|u|}{8gh} u, \tag{5}$$

where f is the possibly time and space-dependent Darcy–Weisbach's roughness coefficient and $|u|$ the module of the velocity u.

We use the Green–Ampt model [19] to predict cumulative infiltration through dry or wet soils. Herein, we consider the formulation developed in [28] for rainfall-runoff predictive models. This formulation postulates, at any point x in space, a well-defined wetting front propagating vertically and separating a fully saturated zone from a zone at the initial soil moisture. At any point on the surface, the infiltration capacity $I_\mathrm{c}(t)$ [L/T] at time t is calculated as follows (the dependency on the space variable is omitted for simplicity):

$$I_\mathrm{c}(t) = K_\mathrm{s} \Big(1 + (\theta_\mathrm{s} - \theta_\mathrm{i}) \frac{h_\mathrm{f} + h(t)}{I^*(t)}\Big), \tag{6}$$

where $I^*(t)$ [L] is the cumulative infiltration up to time t, K_s [L/T] the saturated hydraulic conductivity, h_f [L] the wetting front capillary pressure head, and θ_i and θ_s the initial and saturated water content. Over the time interval $[t, t + \delta t]$, the model assumes that if the water depth $h(t)$ is smaller than $I_\mathrm{c}(t) \times \delta t$, all the water volume is infiltrated; otherwise, the infiltrated volume is equal to the infiltration capacity and the remaining water streams. Hence, the infiltration rate I over the time interval $[t, t + \delta t]$ is equal to $\min(I_\mathrm{c}(t), h(t)/\delta t)$.

2.2. Numerical resolution

A well-balanced finite volume method is used to discretize the SW equations, which we rewrite in the general form $\partial_t U + \partial_x F(U) = S(U)$, where U is the vector of conservative variables, F the flux vector, and S the source term. Specifically,

$$U = \begin{pmatrix} h \\ hu \end{pmatrix}, \qquad F(U) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{pmatrix}, \qquad S(U) = \begin{pmatrix} R - I \\ -gh(\partial_x z + S_{\mathrm{f}}) \end{pmatrix}.$$

The domain is divided into cells (indexed by i) of the form $C_i = [x_{i-1/2}, x_{i+1/2}]$ and of length $\Delta x > 0$ taken constant for simplicity. The Green–Ampt model is applied locally in each mesh cell. To obtain a second-order scheme, the variables need to be reconstructed at cell interfaces. We denote by $U_{i+1/2\,\pm}$ the conservative variables computed at either side of the interface $x_{i+1/2}$ using an ENO-type (Essentially Non Oscillatory) reconstruction [20]. Moreover, the ground surface elevation z is described as a piecewise constant function, and cell-interface values $z_{i+1/2\,\pm}$ are also reconstructed. This yields the following scheme written here in space semi-discrete form:

$$\Delta x \frac{d}{dt} U_i(t) + F_{i+1/2} - F_{i-1/2} = \Delta x \begin{pmatrix} R_i - I_i \\ -g h_i S_{\mathrm{f,i}} \end{pmatrix} + \Delta x \begin{pmatrix} 0 \\ S_{\mathrm{s,i}} \end{pmatrix},$$

where the subscript i refers to the value in the cell C_i and the fluxes $F_{i\pm1/2}$ are computed using the HLL flux (see e.g., [7]). The source term $S_{\mathrm{s,i}}$ accounts for the term $-gh\partial_x z$ in the source term S. To evaluate the fluxes $F_{i\pm1/2}$ and the source term $S_{s,i}$, an hydrostatic reconstruction scheme is applied, as described in [1, 2, 7, 27]. Specifically, we set

$$\begin{cases} h_{i+1/2\,L} = \max\big(0, h_{i+1/2\,-} + z_{i+1/2\,-} - \max(z_{i+1/2\,-}, z_{i+1/2\,+})\big), \\ h_{i-1/2\,R} = \max\big(0, h_{i-1/2\,+} + z_{i-1/2\,+} - \max(z_{i-1/2\,-}, z_{i-1/2\,+})\big), \\ U_{i+1/2\,L} = \big(h_{i+1/2\,L}, h_{i+1/2\,L} u_{i+1/2\,-}\big)^t, \\ U_{i-1/2\,R} = \big(h_{i-1/2\,R}, h_{i-1/2\,R} u_{i-1/2\,+}\big)^t. \end{cases}$$

Then, the HLL flux is evaluated using $(U_{i+1/2\,L}, U_{i+1/2\,R})$, and the source term $S_{\mathrm{s,i}}$ is evaluated as

$$S_{\mathrm{s,i}} = \frac{1}{\Delta x}\frac{g}{2}\Big(\big(h^2_{i+1/2\,L} - h^2_{i+1/2\,-}\big) + \big(h^2_{i-1/2\,+} - h^2_{i-1/2\,R}\big) + \big(h_{i-1/2\,+} - h_{i+1/2\,-}\big)\big(z_{i+1/2\,-} - z_{i-1/2\,+}\big)\Big).$$

Finally, for time discretization, we use a second-order explicit Runge–Kutta method based on the Heun scheme, except for the friction term S_{f} which is treated semi-implicitly at each stage of the Heun scheme [10]. This leads to a second-order accurate overland flow model with infiltration that we now use to study uncertainty propagation.

3. Stochastic model and statistic tools

In this section, we describe the stochastic model and the statistic tools used to analyze the results.

3.1. Stochastic model

We are interested in uncertainty propagation stemming from the uncertain parameter K_{s} in the SW equations. As discussed in the introduction, we consider a simple approach where we subdivide the physical domain into p (with typically $p = 2$ or 3 in our numerical results) subdomains and assign to each subdomain a single uncertain parameter K_{s} which is a RV with known pdf. The uncertain input parameters of the model are then regrouped into a single vector-valued RV such that $\boldsymbol{X} = (X_1, \ldots, X_p) = (K_{\mathrm{s},1}, \ldots, K_{\mathrm{s},p})$. We assume that the RV's X_i are independent, but can have different pdf's denoted by p_{X_i}. We consider a uniform

distribution for all the subdomains because of the relatively low range of values taken by K_s within each subdomain (however, high contrasts are considered between subdomains). In the present setting, the pdf p_{X_i} depends on its corresponding subdomain i only through the minimal and maximal bounds on $K_{s,i}$. Moreover, since the RV's X_i are assumed to be independent, the pdf of $\boldsymbol{X}$ factorizes into the form

$$p_{\boldsymbol{X}}(x_1, \cdots, x_p) = \prod_{i=1}^{p} p_{X_i}(x_i). \tag{7}$$

The uncertain output quantities of the model are the peak runoff rate and the runoff coefficient for a specific rainfall event. Let Y denote any of these output quantities. Once a realization of $\boldsymbol{X}$, say $\boldsymbol{x}$, is known, a realization of Y, say $y(\boldsymbol{x})$, is obtained by solving numerically the corresponding deterministic problem described in Section 2.

3.2. Moments and pdf

Assuming that Y has finite second-moment, the expectation and the variance of Y are defined as

$$E[Y] = \int y(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \quad \text{and} \quad V(Y) = \int \Big(y(\boldsymbol{x}) - E[Y]\Big)^2 p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x},$$

so that $V(Y) = E[Y^2] - E[Y]^2$. We are interested in evaluating various statistical quantities related to the model output Y. To this purpose, we use Monte Carlo (MC) simulations. Let $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\}$ be a sample set of the input stochastic parameters, where M is the sample set dimension and $\boldsymbol{x}^{(m)}$, $1 \leq m \leq M$, are realizations of $\boldsymbol{X}$. Let $\mathcal{Y} = \{y^{(1)}, \ldots, y^{(M)}\}$ be the corresponding sample set of the model output such that, for each $1 \leq m \leq M$, $y^{(m)} = y(\boldsymbol{x}^{(m)})$ is the model response to the vector of input parameters $\boldsymbol{x}^{(m)}$. The empirical estimators for the expectation and the variance are

$$\hat{E}[Y] = \frac{1}{M} \sum_{m=1}^{M} y^{(m)} \quad \text{and} \quad \hat{V}(Y) = \frac{1}{M} \sum_{m=1}^{M} \Big(y^{(m)} - \hat{E}[Y]\Big)^2. \tag{8}$$

To estimate the pdf of a random variable, we use the kernel density estimator, also called Parzen–Rozenblatt method, which is a generalization of the histogram method. The pdf of Y is estimated as

$$\hat{p}_\eta(y) = \frac{1}{M\eta} \sum_{m=1}^{M} G\left(\frac{y - y^{(m)}}{\eta}\right),$$

where G is a specific pdf used as kernel and η is a smoothing parameter. The most commonly used kernel is the Gaussian function $G(x) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right)$. Thus, the pdf at a point y is estimated by the number of observations close to y and counterbalanced by the distance of these observations to y. The kernel distribution function allocates more important weights to observations near the point y and weaker weights to distant observations. The parameter η fixes the kernel function width and, therefore, controls the smoothness of the estimated pdf $\hat{p}_\eta$. The smaller the parameter, the more accurate the estimation of the pdf; however, too small values for η can generate spurious data artifacts if the sample set is not sufficiently large. An illustration is presented in Section 4.2.

3.3. Sensitivity analysis

Sensitivity Analysis (SA) allows one to assess the relative contribution of each uncertain input parameter to model output variability and, in particular, to identify key parameters by establishing a hierarchy within the input parameters according to their influence on the output variability. Variance-based global SA methods based on Sobol indices [32] determine which part of the response variance results from the variance of each input or group of inputs. The sensitivity of the response Y to the input parameter X_i can be quantified by

the first-order sensitivity index S_i defined as

$$S_i = \frac{V_i}{V(Y)}, \qquad V_i = E\Big[E[Y|X_i]^2\Big] - E[Y]^2,$$

where $E[Y|X_i]$ is the conditional expectation of Y given the value of X_i (see (9) below for its definition). More generally, higher-order sensitivity indices quantify the sensitivity of the model response to interactions among input parameters. Let $\boldsymbol{i}$ denote a non-empty subset of indices such that $\boldsymbol{i} \subseteq \{1,\dots,p\}$ and let $\sim\boldsymbol{i} = \{1,\dots,p\} \setminus \boldsymbol{i}$. The sensitivity index $S_{\boldsymbol{i}}$ is defined as

$$S_{\boldsymbol{i}} = \frac{V_{\boldsymbol{i}}}{V(Y)}, \qquad V_{\boldsymbol{i}} = V\left(E\left[Y|X_{\boldsymbol{i}}\right]\right) - \sum_{\emptyset \neq \boldsymbol{j} \subsetneq \boldsymbol{i}} V_{\boldsymbol{j}},$$

where $V\left(E\left[Y|X_{\boldsymbol{i}}\right]\right)$ is the variance of the conditional expectation of Y given the value of $X_{\boldsymbol{i}}$. This conditional expectation is defined as

$$E\left[Y|X_{\boldsymbol{i}}\right] = \int y(\boldsymbol{x}) p_{\boldsymbol{X}_{\sim \boldsymbol{i}}}(\boldsymbol{x}_{\sim \boldsymbol{i}}) d\boldsymbol{x}_{\sim \boldsymbol{i}}, \tag{9}$$

where $p_{\boldsymbol{X}_{\sim \boldsymbol{i}}}$ and $d\boldsymbol{x}_{\sim \boldsymbol{i}}$ are, respectively, the density and the probability measure of $\boldsymbol{x}_{\sim \boldsymbol{i}}$ (conventionally, $E\left[Y|X_{\boldsymbol{i}}\right] = Y$ if $\boldsymbol{i} = \{1,\dots,p\}$ and $\sim\boldsymbol{i}$ is empty). Observing that $E\left[E[Y|X_{\boldsymbol{i}}]\right] = E[Y]$, we obtain

$$V(E[Y|X_{\boldsymbol{i}}]) = E\Big[E[Y|X_{\boldsymbol{i}}]^2\Big] - E[Y]^2.$$

Furthermore, the law of total variance states that $\sum_{\emptyset \neq \boldsymbol{i} \subseteq \{1,\dots,p\}} V_{\boldsymbol{i}} = V(Y)$, so that

$$\sum_{\emptyset \neq \boldsymbol{i} \subseteq \{1,\dots,p\}} S_{\boldsymbol{i}} = 1.$$

Following Homma and Saltelli [22], it is convenient to consider for a single index $i \in \{1,\dots,p\}$, the total sensitivity index $S_{\mathrm{T},i}$ which evaluates the total sensitivity of the model response Y to the input parameter X_i, including X_i alone and all interactions with the other input parameters X_j, $j \neq i$. Computing this index instead of the high-order sensitivity indices allows one to reduce computational costs by avoiding tedious calculations. The total sensitivity index $S_{\mathrm{T},i}$ is evaluated as follows:

$$S_{\mathrm{T},i} = 1 - \frac{V_{\sim i}}{V(Y)}, \qquad V_{\sim i} = E\Big[E[Y|X_{\sim i}]^2\Big] - E[Y]^2,$$

where $V_{\sim i}$ is the variance of the conditional expectation of Y given all the parameters except X_i. The interpretation of the indices S_i and $S_{\mathrm{T},i}$ is the following: X_i is an influential parameter if S_i is important, whereas X_i is not an influential parameter if $S_{\mathrm{T},i}$ is small. Moreover, S_i close to $S_{\mathrm{T},i}$ means that interactions between X_i and the other parameters are negligible.

MC simulations are used to estimate the first-order sensitivity indices S_i and the total sensitivity indices $S_{\mathrm{T},i}$. To this purpose, we estimate the quantities V_i and $V_{\sim i}$. To save computational costs when evaluating these variances [32], the expectations $E\Big[E[Y|X_i]^2\Big]$ and $E\Big[E[Y|X_{\sim i}]^2\Big]$ are computed as a unique integral by making use of two independent M-samples of input variables, $\mathcal{X}$ and $\mathcal{X}^*$, in such a way that

$$\hat{V}_i = \frac{1}{M}\sum_{m=1}^{M} Y\left(\boldsymbol{x}^{(m)}\right) \times Y\left(x_i^{(m)}, \boldsymbol{x}_{\sim i}^{*(m)}\right) - \hat{E}[Y]^2,$$

$$\hat{V}_{\sim i} = \frac{1}{M}\sum_{m=1}^{M} Y\left(\boldsymbol{x}^{(m)}\right) \times Y\left(x_i^{*(m)}, \boldsymbol{x}_{\sim i}^{(m)}\right) - \hat{E}[Y]^2.$$

Finally, the first-order sensitivity index S_i and the total sensitivity index $S_{\mathrm{T},i}$ are estimated as

$$\hat{S}_i = \frac{\hat{V}_i}{\hat{V}(Y)}, \qquad \hat{S}_{\mathrm{T},i} = 1 - \frac{\hat{V}_{\sim i}}{\hat{V}(Y)}.$$

In practice, the computational procedure requires two samples of input parameters, each of dimension M, and $M \times (2p+1)$ deterministic model evaluations to calculate all the first-order and total sensitivity indices.

4. Test cases

The section presents the test cases and a brief performance evaluation of the methodology.

4.1. Presentation

To evaluate how uncertainties on the values of K_s and its spatial localization can impact the surface runoff during various types of rainfall events, we focus on two output quantities: the peak runoff rate at the outlet, $Q_{\max}$, and the runoff coefficient, RC, defined as the total volume of runoff divided by the total volume of rainfall.

A one-dimensional flat slope of length L with an inclination gradient $\partial_x z = 2\%$ is considered. Uniform friction coefficient and infiltration parameters (except K_s) are chosen with values

$$f = 0.25, \qquad \theta_\mathrm{s} - \theta_\mathrm{i} = 0.3, \qquad h_\mathrm{f} = 0.023.$$

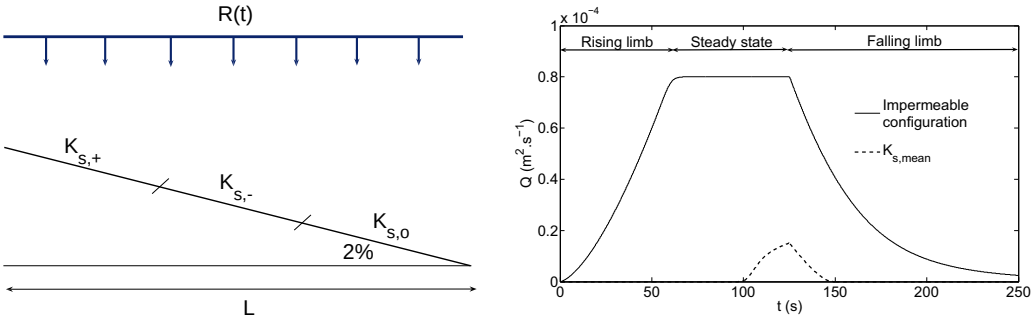
A constant rainfall intensity $R(t)$ [L/T] is imposed during a time T_R [T] and stopped afterwards. The simulation time is denoted by T [T]. Two configurations are studied for the spatial distribution of the uncertainty parameter K_s. For each spatial configuration, two rainfall events are simulated, a short rainfall event (SRE) and a long rainfall event (LRE). The values of the rainfall intensity R, the rainfall duration T_R, and the simulation time T are specified in Tables 1 and 3 for the two spatial configurations described below.

For the first spatial configuration, named "Three-field", we consider a 4.8 m long domain divided into three subdomains, each one with its own saturated hydraulic conductivity $K_{\mathrm{s},i}$, $i \in \{1,2,3\}$, which is a RV independent of $K_{\mathrm{s},j}$, $j \neq i$. Each $K_{\mathrm{s},i}$ has a uniform distribution $K_{\mathrm{s},i} \sim \mathcal{U}[K_{\mathrm{s},i}^{\min}, K_{\mathrm{s},i}^{\max}]$, where $K_{\mathrm{s},i}^{\min}$ and $K_{\mathrm{s},i}^{\max}$ are the minimal and maximal values which can be taken by $K_{\mathrm{s},i}$ in its corresponding subdomain. To assign these minimal and maximal values, we consider three choices, each representing realistic values for a given soil type. We refer to these choices using an index $-$, o, or $+$ indicating respectively low, intermediate or high values for K_s. The corresponding values are listed in Table 2. Then, we consider the six possible spatial localizations of soil types: $+\ o\ -$, $+\ -\ o$, $o\ +\ -$, $o\ -\ +$, $-\ +\ o$, and $-\ o\ +$. For instance, $+\ o\ -$ means that $K_{\mathrm{s},+}$ is located upslope, $K_{\mathrm{s},o}$ is located midslope, and $K_{\mathrm{s},-}$ is located downslope, see Figure 2(a). Figure 2(b) presents the hydrographs for the case where $K_{\mathrm{s},+}$, $K_{\mathrm{s},o}$, and $K_{\mathrm{s},-}$ are all equal to their respective mean values (Table 2), and the impermeable configuration. The significant differences observed emphasize the importance of infiltration processes.

For the second spatial configuration, named "Grass strip(s)", we consider a 318 m long domain in which grass strips (GS) of total width equal to 6 m are added. Two spatial localizations of the GS are considered, as represented in Figure 3: either three narrow, 2 m wide GS are equally spaced or one large, 6 m wide GS is located at the outlet. We assign a saturated hydraulic conductivity to the GS and another one to the remaining part of the soil surface called the field. Both K_s are independent RV's with uniform distribution. The values taken by K_s on the GS are higher than those on the field (see Table 4).

4.2. Performance evaluation

Before discussing our results in the next session, we verify the numerical procedure on the "Three-field" test case with SRE and the spatial localization $+\ -\ o$ for K_s. Figure 4(a) presents the convergence of the first-order and total sensitivity indices as a function of the sample set dimension M. For both order indices (first and total), the convergence is obtained



(a) (b)

Figure 2: "Three-field" test case with the spatial localization $+ \; - \; o$: (a) initial configuration ; (b) rainfall hydrograph for the impermeable configuration and the case where $K_{s,+}$, $K_{s,o}$, and $K_{s,-}$ all take their respective mean value.

Event	R (m.s^{-1})	T_R (s)	T (s)
SRE	$1.66{\cdot}10^{-5}$	125	250
LRE	$1.66{\cdot}10^{-5}$	1,250	2,500

Table 1: "Three-field" test case, data for the two rainfall events: rainfall intensity R, rainfall duration T_R, and total simulation time T.

	$-$	o	$+$
K_s^{min}	$2.78{\cdot}10^{-7}$	$2.78{\cdot}10^{-6}$	$1.10{\cdot}10^{-5}$
K_s^{max}	$1.10{\cdot}10^{-6}$	$5.50{\cdot}10^{-6}$	$1.66{\cdot}10^{-5}$

Table 2: "Three-field" test case: minimal and maximal values of K_s (m.s^{-1}) for the soil types.

quickly around $M = 1000$. Figure 4(b) presents the MC estimate of the expectation and standard deviation of Q_{max} with ± 3 bootstrap standard error bounds plotted against the sample set dimension M. A sample set dimension equal to 100,000 appears to be sufficient to achieve convergence, and in agreement with the previous convergence of the sensitivity indices. This value for M is used in what follows. Figure 4(c) illustrates the influence of the bandwidth η on the pdf estimation. An under-smoothed pdf is obtained with a small value ($\eta = 0.01$) whereas an over-smoothed pdf is obtained with a large value ($\eta = 0.5$). The value $\eta = 0.05$ yields a sufficiently smoothed pdf without spurious oscillations. This value for η is used in what follows.

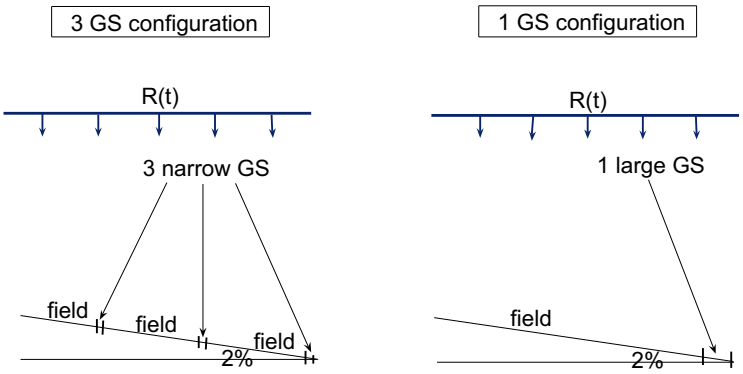


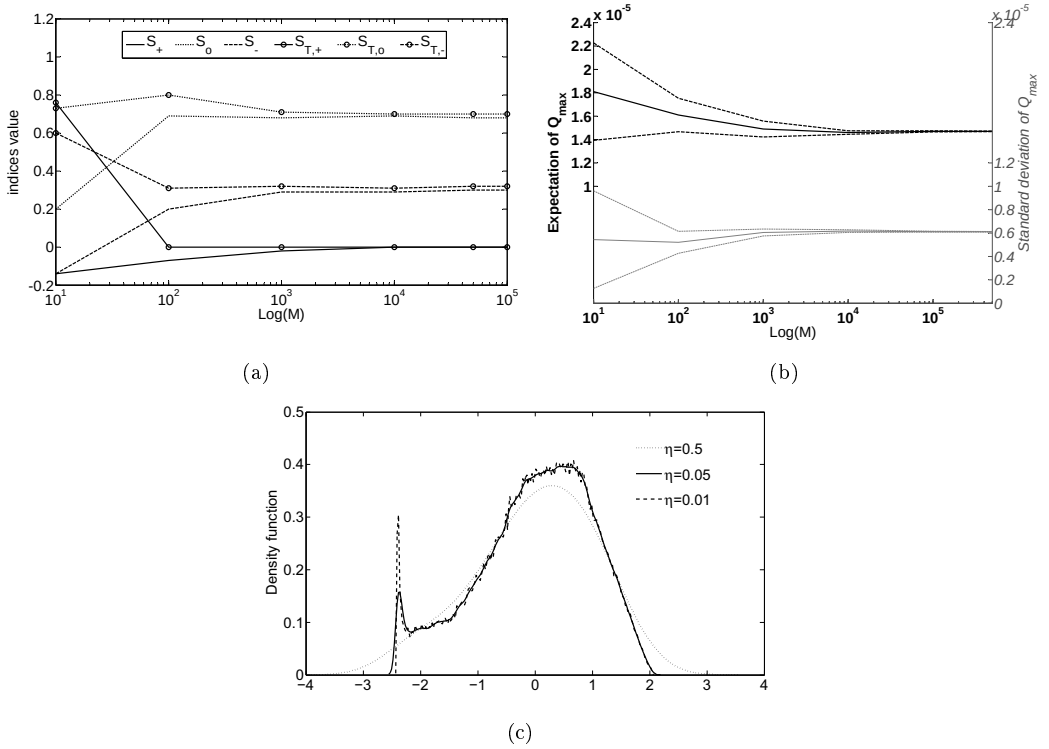
Figure 3: "Grass strip(s)" test case: initial configuration.

Event	R (m.s^{-1})	T_{R} (s)	T (s)
SRE	$8.33{\cdot}10^{-6}$	3,600	5,000
LRE	$1.11{\cdot}10^{-5}$	8,500	9,500

Table 3: "Grass strip(s)" test case, data for the two rainfall events: rainfall intensity R, rainfall duration T_{R}, and total simulation time T.

	field	GS
$K_{\mathrm{s}}^{\min}$	$3.57{\cdot}10^{-6}$	$2.22{\cdot}10^{-5}$
$K_{\mathrm{s}}^{\max}$	$6.35{\cdot}10^{-6}$	$3.33{\cdot}10^{-5}$

Table 4: "Grass strip(s)" test case: minimal and maximal values of K_{s} (m.s^{-1}).



(a) (b)

(c)

Figure 4: Numerical verification for the "Three-field" test case with the spatial localization $+ - o$: (a) convergence of the first-order and total sensitivity indices for $Q_{\max}$ as a function of sample dimension M; (b) convergence of the MC estimate of $Q_{\max}$ expectation and standard deviation with ± 3 standard error as a function of sample dimension M; (c) probability density estimation of $Q_{\max}$, using the kernel density estimator for different bandwidth values η.

5. Results and discussion

This study focuses on the effects of the uncertainties in K_{s} and of the spatial localization of the soil types on the peak runoff rate at the outlet, $Q_{\max}$, and the runoff coefficient, RC. The two test cases "Three-field" and "Grass-strip(s)", detailed in Section 4, are investigated. In particular, the different statistic estimators (moments, pdf's, and sensitivity indices) are calculated for the two rainfall events SRE and LRE, with a sample dimension equal to 100,000 and a uniform distribution for all RV's.

5.1. Three-field test case

Figures 5 and 6 present the 100,000 couples ($Q_{\max}$,RC) for the six possible spatial localizations of soil types and for SRE and LRE, respectively. An important point is that there is an important correlation between the two outputs $Q_{\max}$ and RC for each spatial localization. Indeed, all the correlation coefficients are greater than 0.9. Concerning SRE (Figure 5), the simulations even tend to line up in a curve. We observe that $Q_{\max} = 0$ when $K_{\mathrm{s},+}$ is downslope due to a complete infiltration of the rain and the upslope runoff.

The cloud of points for the spatial localization $+\ -\ o$ contains the one for $-\ +\ o$ which corresponds to the weakest discharges. The configurations where $K_{s,-}$ is located downslope are similar whatever the positions of $K_{s,+}$ and $K_{s,o}$ because the values of $K_{s,+}$ and $K_{s,o}$ are sufficiently important to infiltrate all the rain. Therefore, the clouds of points for SRE depend on the K_s located downslope. Concerning LRE (Figure 6), Q_{max} and RC increase since the rainfall duration is longer. As a result, the influence of $K_{s,+}$ is more pronounced for LRE and contributes more to the discharge at the outlet, whereas the influence of $K_{s,-}$ decreases. Therefore, the clouds of points for LRE depend essentially on the position of $K_{s,+}$ within the domain. Figures 5 and 6 stress the importance of the spatial distribution of the soil types since the outputs are influenced by the K_s located downslope for SRE and by the localization of $K_{s,+}$ for LRE. To better understand why this difference is observed by changing the rainfall duration, we focus on the infiltration process over the domain.

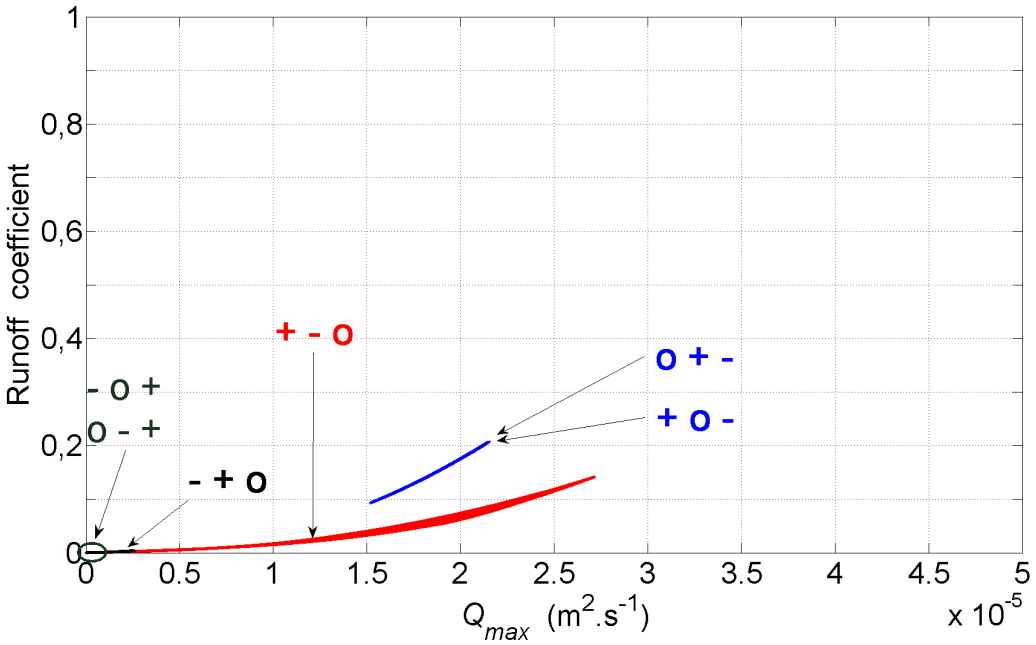


Figure 5: "Three-field" test case and Short Rainfall Event: peak runoff rates Q_{max} and runoff coefficients RC for the six possible spatial localizations of soil types.

Equation (6) implies that the ratio I_c/K_s tends to 1 when the infiltrated water volume tends to infinity (corresponding to a saturated soil). To study the effect of increasing the rainfall duration on soil saturation, Figure 7 presents the confidence interval (i.e. minimal and maximal values for the 100,000 model responses) of the ratio I_c/K_s at final time, as a function of spatial position, in grey for SRE and in black for LRE. As expected, the soil is more saturated for LRE and the ratio is closer to 1. Besides, the variability of the confidence interval is in general the highest for the subdomain where $K_{s,-}$ is located and the weakest for the subdomain where $K_{s,+}$ is located. The variability decreases as a function of soil saturation since the more saturated the soil, the smaller the variability, except for some limit cases where there is no runoff on the concerned subdomain.

Table 5 presents the mean μ, the standard deviation σ, the coefficient of variation $cov = \sigma/\mu$, the median P_{50}, and the 90th percentile P_{90} related to Q_{max} (white lines) and RC (grey lines). On the whole, there is more dispersion on the estimated values for SRE. For instance, cov varies between 10% when $K_{s,-}$ is located downslope and reaches 42% (for $+\ -\ o$) and 217% (for $-\ +\ o$) when $K_{s,o}$ is located downslope. Besides, the values are higher for SRE than for LRE. The increase of the rainfall duration leads to a decrease in the dispersion values, therefore dispersion depends on the state of soil saturation (as the variability of the ratio I_c/K_s observed previously in Figure 7). Furthermore, for SRE, the distribution is

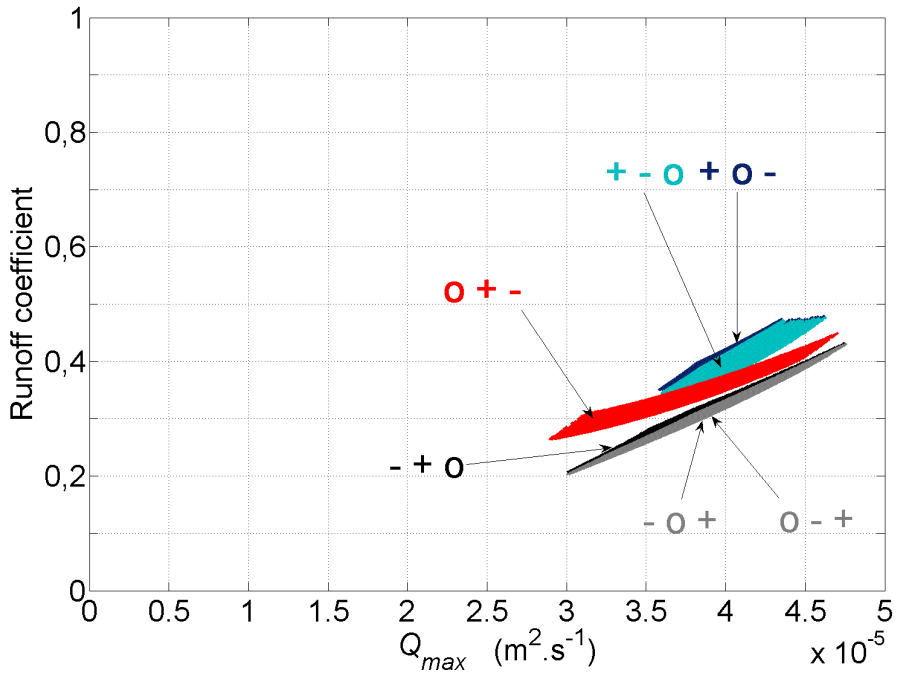


Figure 6: "Three-field" test case and Long Rainfall Event: peak runoff rates $Q_{\max}$ and runoff coefficients RC for the six possible spatial localizations of soil types.

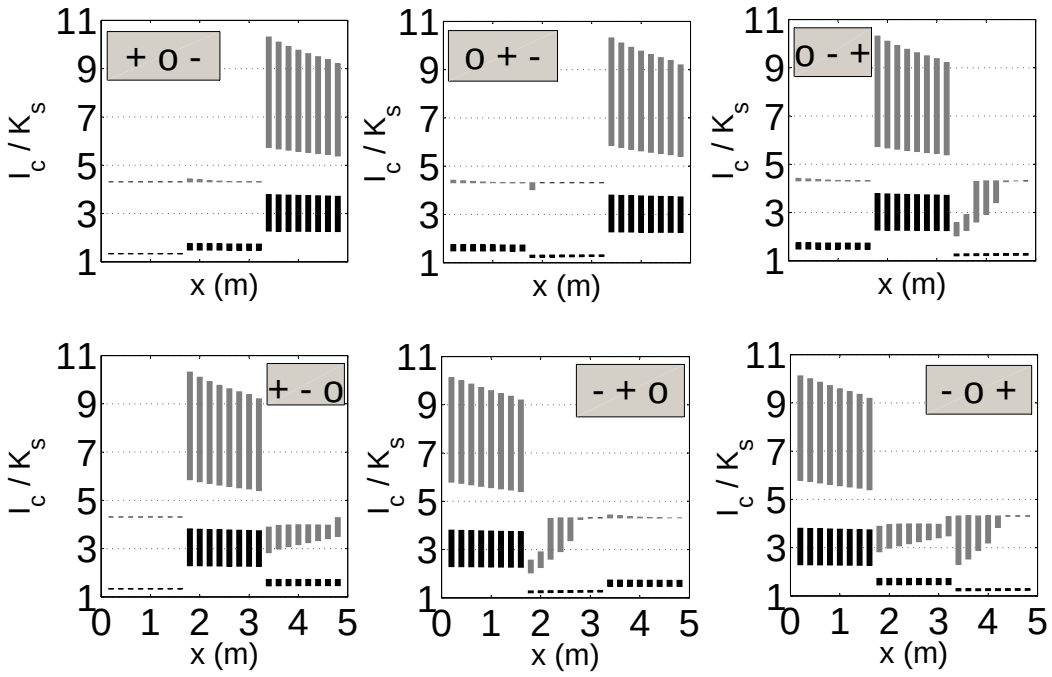


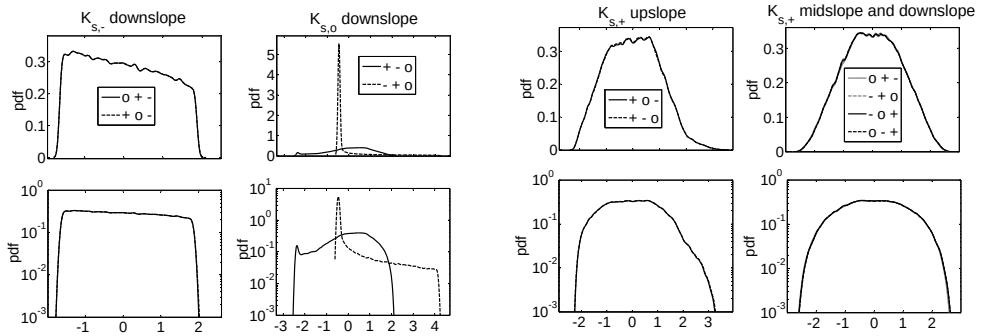
Figure 7: "Three-field" test case: confidence interval of the ratio I_c/K_s at final time as a function of spatial position and for the six possible localizations of soil types; Short Rainffal Event (in grey) and Long Rainfall Event (in black).

not uniform when $K_{s,o}$ is downslope since the median and the mean are different. Finally, similar conclusions can be drawn from the statistical values associated with RC. Moreover,

we observe that for both SRE and LRE, *cov* takes higher values for RC than for $Q_{\max}$.

	Short Rainfall Event				Long Rainfall Event					
	$K_{s,-}$ downslope		$K_{s,o}$ downslope		$K_{s,+}$ upslope		$K_{s,+}$ midslope		$K_{s,+}$ downslope	
	$+\ o\ -$	$o\ +\ -$	$+\ -\ o$	$-\ +\ o$	$+\ o\ -$	$+\ -\ o$	$o\ +\ -$	$-\ +\ o$	$o\ -\ +$	$-\ o\ +$
μ	$1.8{\cdot}10^{-5}$	$1.8{\cdot}10^{-5}$	$1.5{\cdot}10^{-5}$	$2.6{\cdot}10^{-7}$	$4.0{\cdot}10^{-5}$	$4.0{\cdot}10^{-5}$	$3.8{\cdot}10^{-5}$	$3.9{\cdot}10^{-5}$	$3.9{\cdot}10^{-5}$	$3.9{\cdot}10^{-5}$
	$2.8{\cdot}10^{-4}$	$2.8{\cdot}10^{-4}$	$8.7{\cdot}10^{-5}$	$5.9{\cdot}10^{-7}$	$8.2{\cdot}10^{-3}$	$8.1{\cdot}10^{-3}$	$6.9{\cdot}10^{-3}$	$6.2{\cdot}10^{-3}$	$6.2{\cdot}10^{-3}$	$6.2{\cdot}10^{-3}$
σ	$1.8{\cdot}10^{-6}$	$1.8{\cdot}10^{-6}$	$6.1{\cdot}10^{-6}$	$5.6{\cdot}10^{-7}$	$1.8{\cdot}10^{-6}$	$1.8{\cdot}10^{-6}$	$3.4{\cdot}10^{-6}$	$3.3{\cdot}10^{-6}$	$3.3{\cdot}10^{-6}$	$3.3{\cdot}10^{-6}$
	$6.3{\cdot}10^{-5}$	$6.3{\cdot}10^{-5}$	$6.1{\cdot}10^{-5}$	$1.5{\cdot}10^{-6}$	$5.2{\cdot}10^{-4}$	$5.4{\cdot}10^{-4}$	$6.6{\cdot}10^{-4}$	$8.2{\cdot}10^{-4}$	$8.3{\cdot}10^{-4}$	$8.3{\cdot}10^{-4}$
cov	10%	10%	42%	217%	5%	5%	9%	9%	9%	8%
	22%	22%	70%	250%	6%	7%	10%	13%	13%	13%
P_{50}	$1.8{\cdot}10^{-5}$	$1.8{\cdot}10^{-5}$	$1.5{\cdot}10^{-5}$	0	$4.0{\cdot}10^{-5}$	$4.0{\cdot}10^{-5}$	$3.8{\cdot}10^{-5}$	$3.9{\cdot}10^{-5}$	$3.9{\cdot}10^{-5}$	$3.9{\cdot}10^{-5}$
	$2.8{\cdot}10^{-4}$	$2.8{\cdot}10^{-4}$	$7.8{\cdot}10^{-5}$	0	$8.2{\cdot}10^{-3}$	$8.1{\cdot}10^{-3}$	$6.8{\cdot}10^{-3}$	$6.2{\cdot}10^{-3}$	$6.1{\cdot}10^{-3}$	$6.2{\cdot}10^{-3}$
P_{90}	$2.1{\cdot}10^{-5}$	$2.1{\cdot}10^{-5}$	$2.2{\cdot}10^{-5}$	$1.1{\cdot}10^{-5}$	$4.2{\cdot}10^{-5}$	$4.2{\cdot}10^{-5}$	$4.3{\cdot}10^{-5}$	$4.3{\cdot}10^{-5}$	$4.3{\cdot}10^{-5}$	$4.3{\cdot}10^{-5}$
	$3.8{\cdot}10^{-4}$	$3.8{\cdot}10^{-4}$	$1.8{\cdot}10^{-4}$	$2.3{\cdot}10^{-6}$	$8.9{\cdot}10^{-3}$	$8.8{\cdot}10^{-3}$	$7.8{\cdot}10^{-3}$	$7.4{\cdot}10^{-3}$	$7.3{\cdot}10^{-3}$	$7.3{\cdot}10^{-3}$

Table 5: "Three-field" test case: mean $\mu = \hat{E}[Q_{\max}]$, standard deviation $\sigma = \sqrt{\hat{V}(Q_{\max})}$, coefficient of variation $cov = \sigma/\mu$, median P_{50}, and 90th percentile P_{90} for the peak runoff rate $Q_{\max}$ (white lines) and the runoff coefficient RC (grey lines).

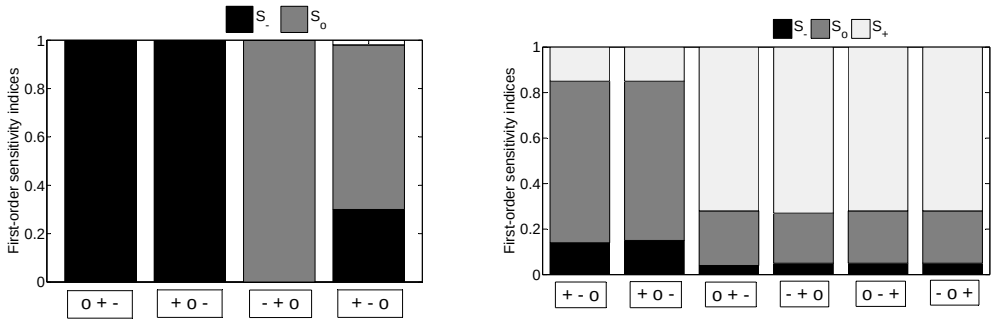


(a) SRE (b) LRE

Figure 8: "Three-field" test case: probability density function of the peak runoff rate $Q_{\max}$ estimated with a bandwidth $\eta = 0.05$; linear scale (top) and logarithmic scale (bottom); (a) Short Rainfall Event (SRE); (b) Long Rainfall Event (LRE).

Figures 8(a) and 8(b) present the pdf's of $Q_{\max}$ estimated by the Parzen–Rozenblatt method and plotted by groups in function of the K_s which influences the most the discharge at the outlet. A first important point is that the pdf shape shows that the process studied is not Gaussian. Concerning SRE, the spatial localizations of soil types yielding clouds of points that are correlated and uniformly distributed in Figure 5 (the ones where $K_{s,-}$ is downslope), generate a spread pdf looking like a rectangular function. The pdf resulting from the configuration $-\ +\ o$ has a marked peak owing to the numerous null discharges observed. This marked peak does not have the expected form on the left part because it is difficult to approximate accurately such a pdf (resembling a Dirac function) by a Gaussian kernel. Concerning LRE, the six curves on Figure 8(b) are very close and have the form of a flat bell on top and are almost symmetrical with respect to zero (i.e, with respect to the mean because the samples are standardized). So, contrary to SRE, the spatial distribution of K_s does not influence the distribution of $Q_{\max}$ for LRE. The estimated pdf's for RC lead to the same conclusions. We can conclude that the dispersions calculated in Table 5 for SRE are confirmed by the non-uniform distribution obtained in the pdf curves.

Figures 9(a) and 9(b) present the first-order sensitivity indices estimated for the 100,000 standardized samples and for the two rainfall events. The sensitivity indices related to $K_{s,+}$, $K_{s,o}$, and $K_{s,-}$ are respectively denoted by S_+, S_o, and S_-. Concerning SRE (Figure 9(a)),



(a) SRE (b) LRE

Figure 9: "Three-field" test case: sensitivity indices of the peak runoff rate $Q_{\max}$; (a) Short Rainfall Event (SRE); (b) Long Rainfall Event (LRE).

the highest index corresponds to the parameter K_{s} located downslope, thus corroborating the previous conclusions on the most influent K_{s}. For instance, in Figure 5, for the spatial localizations where $K_{\mathrm{s},-}$ is located downslope, the clouds of points are similar. Switching $K_{\mathrm{s},o}$ and $K_{\mathrm{s},+}$ does not impact the outlet discharge, meaning that only $K_{\mathrm{s},-}$ influences this quantity, and indeed the indices S_- are equal to 1. Concerning LRE, since the rainfall duration is longer, more runoff is generated in the most infiltrating subdomain because of the decreasing of the infiltration capacity. In Figure 9(b), the most influent parameter is either $K_{\mathrm{s},+}$ (with $S_+ \approx 72\%$) or $K_{\mathrm{s},o}$ (with $S_o \approx 70\%$) when $K_{\mathrm{s},+}$ is located upslope. Moreover, $K_{\mathrm{s},-}$ is not very influent, and contrary to Figure 5, the three parameters K_{s} are not negligible in the sensitivity analysis. Moreover, the total sensitivity indices are equal to the corresponding first-order index, that is, $S_{\mathrm{T},i} \approx S_i$. These equalities mean that there is no significant interaction between the input parameters. Concerning the runoff coefficient RC, the sensitivity analysis leads to the same conclusions. In practice, in case of soils with low levels of saturation (for SRE), it is important to focus the measurements on the parameter closer to the outlet. For more saturated soils (e.g., for LRE), the measurements should focus on the most infiltrating parts of the domain.

To study the effect of the length L of the domain, we have also tested the case where $L = 48$ m with LRE. It is interesting to notice that the length of the domain does not affect the results. The clouds of points and the pdf's have the same shape, and the most influent sensitivity index is the same, i.e, $K_{\mathrm{s},o}$ in cases where $K_{\mathrm{s},+}$ is located upslope or $K_{\mathrm{s},+}$ in other cases. The only significant difference is that the sensitivity index for $K_{\mathrm{s},+}$ vanishes when $K_{\mathrm{s},+}$ is located upslope. This result can be explained by the fact that longer domains lead to an augmentation of the distance between the upslope subdomain and the outlet.

5.2. Grass strip(s) test case

Figure 10 presents the couples ($Q_{\max}$,RC) corresponding to the 100,000 model responses for the four configurations (1 or 3 GS; SRE or LRE). In each configuration, the clouds of points are well correlated and, as previously, the values of $Q_{\max}$ and RC are higher for LRE than for SRE. Concerning $Q_{\max}$, for both SRE and LRE, the values are contained approximately in the same intervals whatever the spatial localization of the GS. Concerning RC, its values are slightly higher for the spatial configuration with 1 GS, and this effect is more significant for SRE. We conclude that the spatial localization of the GS has very little influence on the variability of the runoff, and almost none on that of the flow at the outlet.

Figure 11 presents the confidence interval (for the 100,000 model responses) of the ratio $I_{\mathrm{c}}/K_{\mathrm{s}}$ at final time, as a function of spatial position. As expected, because of the duration of the rainfall events, the ratios are closer to 1 for LRE. Besides, compared to Figure 7, the values taken by the ratio are very close to 1, meaning that the soil is almost saturated. For each rainfall event, the values taken by $I_{\mathrm{c}}/K_{\mathrm{s}}$ for the two spatial configurations (1 GS and 3 GS) are very close. Furthermore, we observe that for SRE, the variability of the ratio $I_{\mathrm{c}}/K_{\mathrm{s}}$ is between 2 and 4 times higher for the GS than for the field. Conversely, for

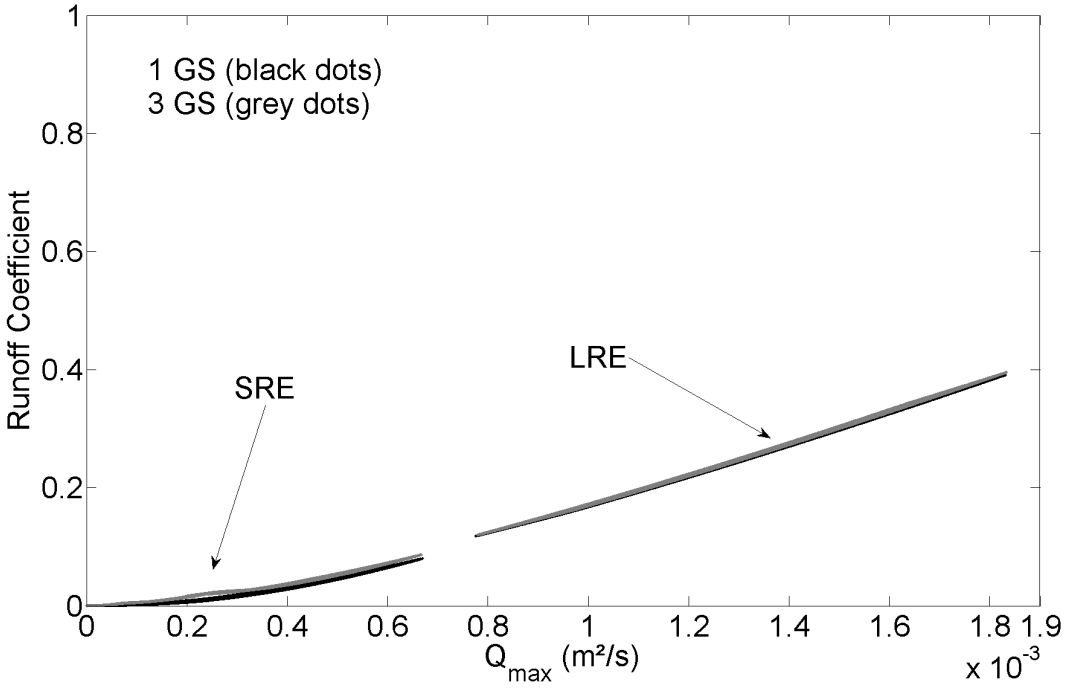


Figure 10: "Grass strip(s)" test case with Short and Long Rainfall Events (resp. SRE and LRE): peak runoff rates $Q_{\max}$ and runoff coefficients RC for the two configurations (one large grass strip downslope (1 GS), or three narrow grass strips (3 GS)).

LRE, the variability is approximately 3 times more important for the field than for the GS. An interesting result concerning SRE is that the effect of having 3 GS instead of 1 GS downslope, is to somehow homogenize the level of variability of I_c/K_s along the field.

Statistical values ($\mu, \sigma, cov, \mathrm{P}_{50}$, and P_{90}), not presented here, confirm that the spatial configurations with 1 GS and 3 GS are similar regarding Q_{max} for LRE, and very close for SRE. Concerning LRE, in agreement with the almost essentially flat shape of the pdf's (Figure 12), we report the same values for the model outputs with the mean values of the parameters, the mean estimation, and the median. Concerning SRE, highly marked peaks are observed with significantly different values (4.2 for 1 GS versus 3.1 for 3 GS). These peaks explain the difference between the mean and the median. Moreover, the mean values of the model outputs differ from the model outputs with the mean parameters. This underlines the importance of non-linear processes. The statistical values and the estimated pdf's for RC lead to the same conclusions.

Figures 13 and 14 present the three statistic estimators μ, P_{50}, and P_{90} for the peak runoff rate $\max_t Q(x,t)$ as a function of spatial position, and the two deterministic values of this quantity (taking $K_s = K_s^{\min}$ and $K_s^{\max}$). The curves for P_{50} almost coincide with those for $\max_t Q(x,t)$ calculated with the value $K_s = K_s^{\mathrm{mean}}$. Contrary to LRE where equality is obtained, the median is inferior to the mean for SRE. Both for the 1 GS and 3 GS configurations, the distribution is not uniform in space. Moreover, for both SRE and LRE, RC is slightly higher with the 3 GS configuration. Although the runoff volumes are comparable for 1 GS and 3 GS, the spatial distribution of maximal discharges varies. Indeed, both in Figures 13 and 14, the discharges along the spatial domain are weaker for 3 GS, owing to the presence of the three GS which slow down the flow. Moreover, this effect is more significant for the SRE because of the saturation of the soil. Therefore, for processes like soil erosion, which are influenced by the maximal discharge, the main result of Figure 13 is that the 3 GS configuration reduces (especially for SRE) the occurrence of high values for $\max_t Q$. Moreover a relevant information obtained with the stochastic approach is that, for SRE, (resp. LRE) the 90th percentile is 33% (resp. 11%) lower with the 3 GS configuration

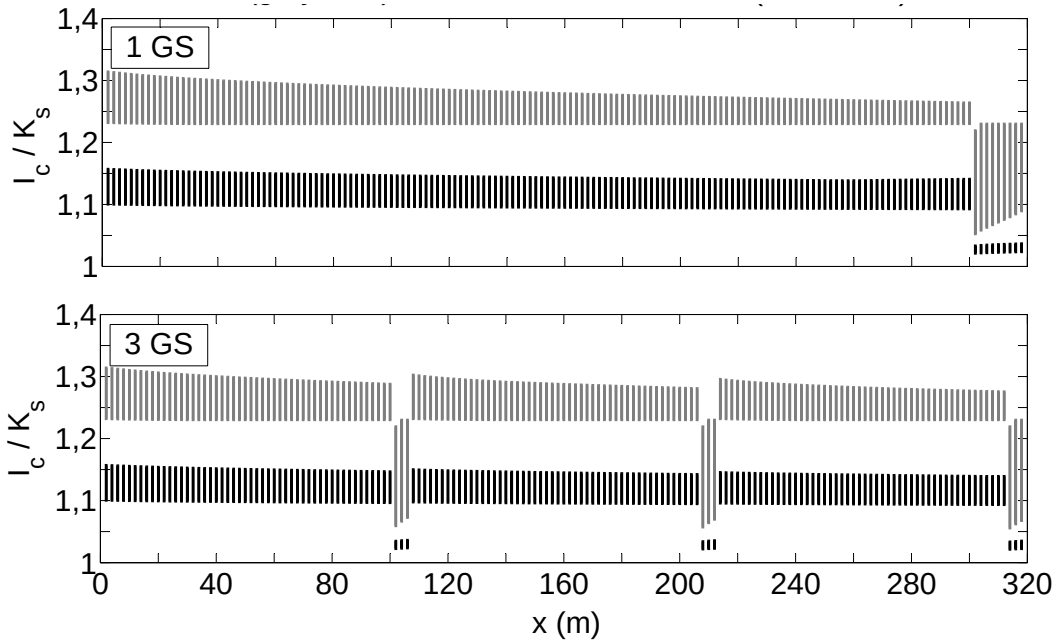


Figure 11: "Grass strip(s)" test case: confidence interval of the ratio I_c/K_s at final time as a function of spatial position for the two configurations (one large grass strip downslope (1 GS), or three narrow grass strips (3 GS)); Short Rainfall Event (in grey) and Long Rainfall Event (in black).

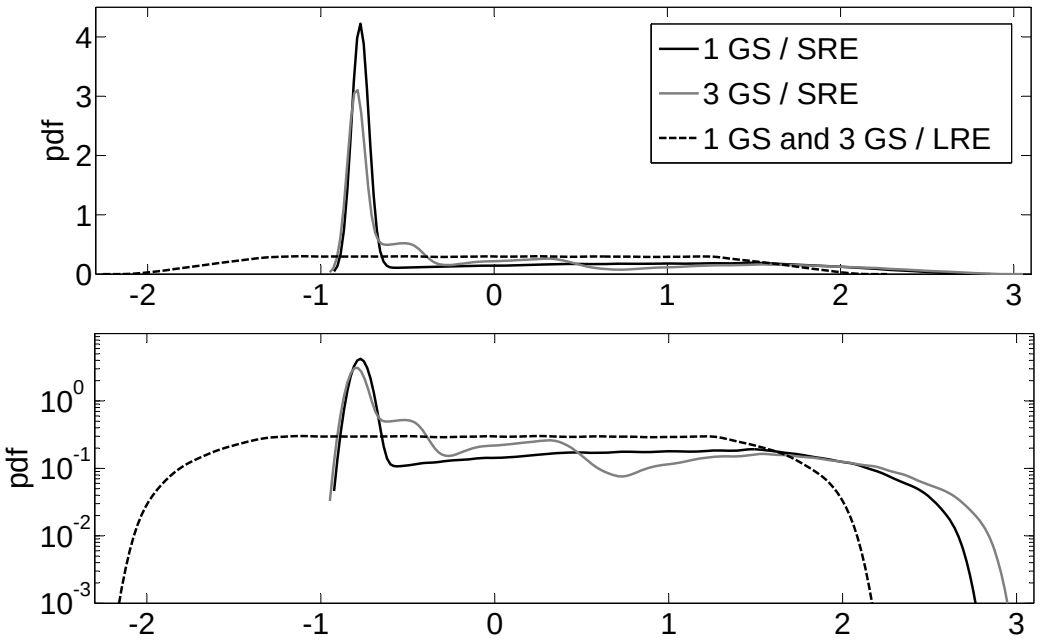


Figure 12: "Grass strip(s)" test case: probability density function of the peak runoff rate $Q_{\max}$ for the Short and Long Rainfall Events (SRE and LRE), estimated with a bandwidth $\eta = 0.05$; linear scale (top) and logarithmic scale (bottom).

than with the 1 GS configuration.

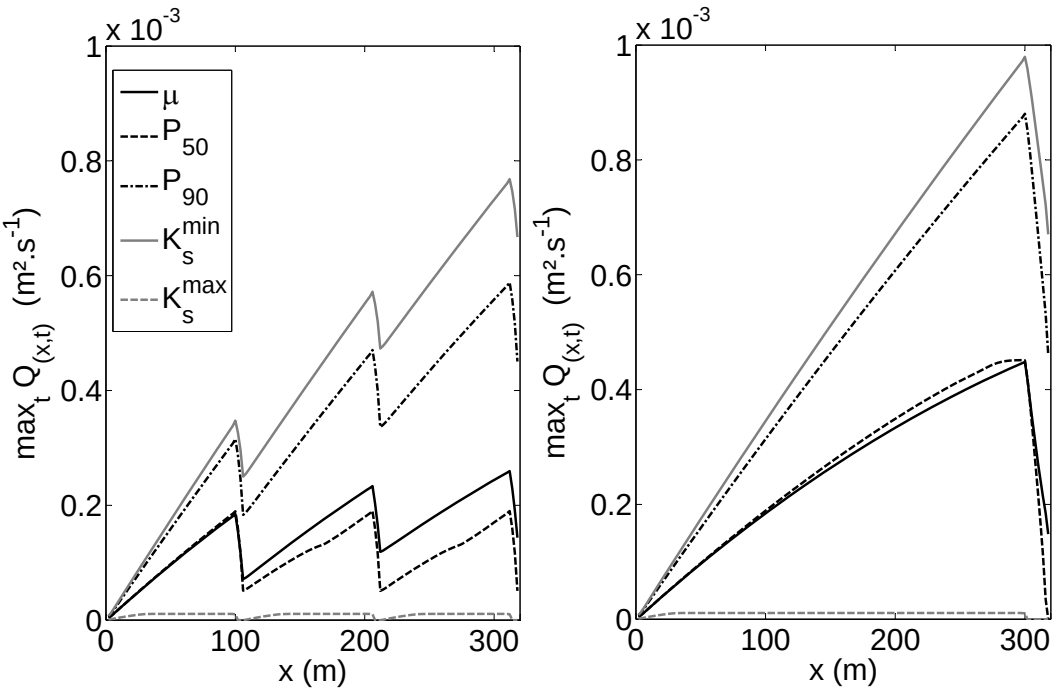


Figure 13: "Grass strip(s)" test case: statistical estimations of the peak runoff rate $\max_t Q(x,t)$ as a function of spatial position (mean $\mu = \hat{E}[\max_t Q(x,t)]$), median P_{50}, and 90th percentile P_{90}), and some deterministic values of this quantity (taking $K_\mathrm{s} = K_\mathrm{s}^\mathrm{min}$ or $K_\mathrm{s}^\mathrm{max}$) for the Short Rainfall Event.

Concerning the sensitivity analysis, for the four configurations (1 GS or 3 GS; SRE or LRE), the first-order sensitivity indices related to the field (in the range 92% to 96%) are much higher than those related to the GS. This shows that only the K_s of the field is an influent parameter, owing to the very important infiltration capacity of the GS.

To study the effect of the values for K_s, we have also tested the "Grass strip(s)" test case with less infiltrating GS. The obtained results corroborate the previous observations. There is no significant difference in terms of runoff and discharge at the outlet, but the presence of three GS slows down the flow and diminishes the occurrence of extreme values for the flow rates.

6. Conclusion

In this work, we have studied the impact of the variability in soil properties on overland flows caused by rainfall events. We have considered the soil saturated hydraulic conductivity K_s as the most uncertain input parameter in the framework of the Green–Ampt infiltration model. To model uncertainties, the flow domain has been split into subdomains reflecting the spatial organization of the landscape (e.g., agricultural fields, grass strips), and the saturated hydraulic conductivity has been described by statistically independent and uniformly distributed random variables, with one random variable assigned to each subdomain. Concerning output quantities, we have focused on the discharges at the outlet (peak runoff rate and runoff coefficient). Two test cases, named "Three-field" and "Grass strip(s)", have been investigated.

The "Three-field" test case aims at understanding the role of spatial organization in uncertainty propagation. The conclusions depend on the level of soil saturation. For long rainfall events leading to highly saturated soils, the variability of model outputs remains moderate. Moreover, the most influent input parameter is the K_s taking the highest values, except when the most infiltrating subdomain is located upslope, in which case the most

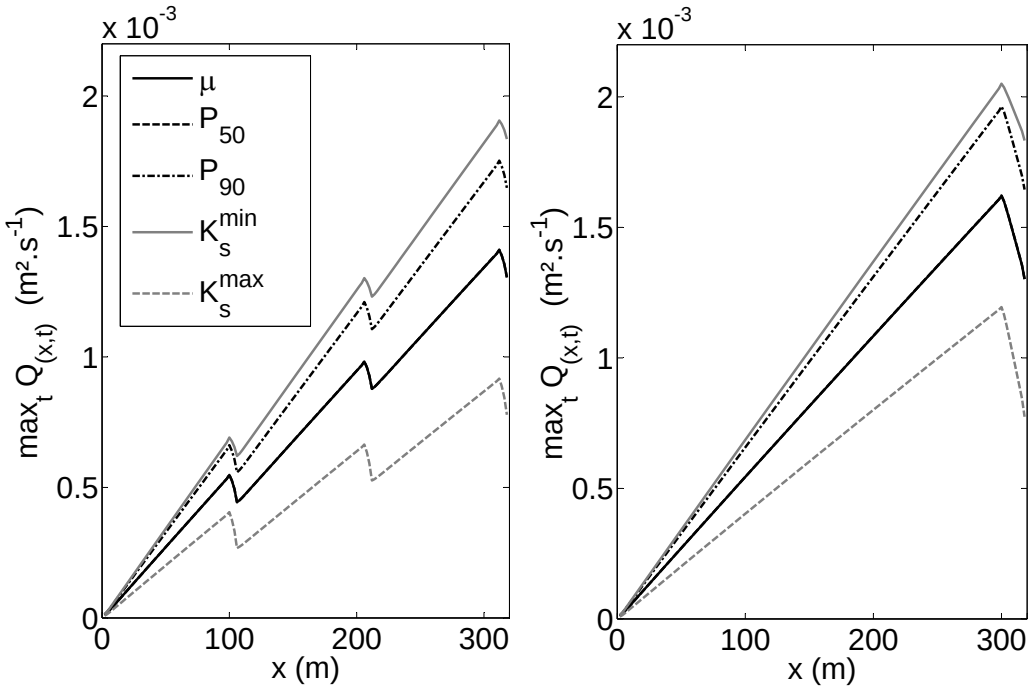


Figure 14: "Grass strip(s)" test case: statistical estimations of the peak runoff rate $\max_t Q(x,t)$ as a function of spatial position (mean $\mu = \hat{E}[\max_t Q(x,t)]$), median P_{50}, and 90th percentile P_{90}), and some deterministic values of this quantity (taking $K_\mathrm{s} = K_\mathrm{s}^{\min}$ or $K_\mathrm{s}^{\max}$) for the Long Rainfall Event.

influent input parameter is the K_s taking intermediate values. For short rainfall events with moderately saturated soils, the most influent input parameter, regardless of its value, is the K_s located downslope, that is, the closest to the outlet. Moreover, since the variability of model responses is high, a practical conclusion in this situation is that additional measurements near the outlet should be the most effective to help reduce output uncertainties.

The "Grass strip(s)" test case aims at comparing, in view of possible land management issues, runoff uncertainties obtained with two possible spatial localizations of grass strips within a field, namely three narrow, equally-spaced grass strips versus one large grass strip located at the field outlet. The first conclusion is that the duration of the rainfall event substantially impacts the shape of the probability density function (pdf) of the model outputs. Specifically, highly peaked pdf's are obtained for short rainfall events (and moderately saturated soils), while relatively flat pdf's are obtained for long rainfall events (and highly saturated soils). The second conclusion is that the localization of the grass strips does not impact the variability of the model outputs. However, one important difference concerns the spatial distribution of maximal discharges since the configuration with three grass strips leads to less probable extreme values, as reflected by the lower values taking by the 90th percentile for the configuration with three grass strips. This observation is relevant in view of assessing erosion risks, since the detachment of soil particles is very sensitive to the peak discharge.

To sum up, the present methodology can be used as a tool to determine where additional efforts should be concentrated when collecting input parameters to reduce output uncertainties, and also as a tool to assess risks due to extreme values, keeping in mind that quantitative results depend on the probabilistic model selected for the input parameters. Possible further applications can target soil erosion based on sediment transport models and groundwater contamination based on pollutant transport models.

[1] E. Audusse. *Modélisation hyperbolique et analyse numérique pour les écoulements en eaux peu profondes.* PhD thesis, Université Pierre et Marie Curie - Paris VI, 2004.

[2] E. Audusse, F. Bouchut, M.O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.
[3] N. Baghdadi, O. Cerdan, M. Zribi, V. Auzet, F. Darboux, M. El Hajj, and R. Bou Keir. Operational performance of current synthetic aperture radar sensors in mapping soil surface characteristics: application to hydrological and erosion modelling. *Hydrological Processes*, 22:9–20, 2008.
[4] G. Baroni, A. Facchi, C. Gandolfi, B. Ortuani, D. Horeschi, and J.C. van Dam. Uncertainty in the determination of soil hydraulic parameters and its influence on the performance of two hydrological models of different comple. *Hydrology and Earth System Sciences*, 14:251–270, 2010.
[5] K. Beven and A. Binley. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298, 1992.
[6] E Boegh, M Thorsen, M.B Butts, S Hansen, J.S Christiansen, P Abrahamsen, C.B Hasager, N.O Jensen, P van der Keur, J.C Refsgaard, K Schelde, H Soegaard, and A Thomsen. Incorporating remote sensing data in physically based distributed agro-hydrological modelling. *Journal of Hydrology*, 287(1-4):279–299, 2004.
[7] F. Bouchut. *Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws and Well-Balanced Schemes for Sources*. Birkhäuser Basel, 2004.
[8] J. Bouma. Using soil survey data for quantitative land evaluation. *Advances in Soil Science*, pages 177–213, 1989.
[9] J. Bouma and H.A.J. van Lanen. Transfer functions and threshold values: from soil characteristics to land qualities. pages 106–110, 1987.
[10] M.O. Bristeau and B. Coussin. Boundary Conditions for the Shallow Water Equations solved by Kinetic Schemes. Research Report RR-4282, INRIA, 2001. Projet M3N.
[11] O. Cerdan, N. Baghdadi, J.P. Denux, J.F. Desprats, M. Gay, C. Albergel, I. Dubus, F. Dupros, N. Holah, and M. El Hajj. Apibar : Appui à la prévision des innondations cas des bassins rapides du sud de la france. Rapport Final BRGM/RP-54218-FR, BRGM, 2006.
[12] V.T Chow. *Open Channel Hydraulics*. McGraw-Hill College, 1959.
[13] J.F. Desprats, D. Raclot, M. Rousseau, O. Cerdan, M. Garcin, Y. Le Bissonnais, A. Ben Slimane, J. Fouche, and D. Monfort-Climent. Mapping linear erosion features using high and very high resolution satellite imagery. *Land Degradation and Development*, 2011.
[14] Q. Duan, S. Sorooshian, and V. Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4):1015–1031, 1992.
[15] M. Esteves, X. Faucher, S. Galle, and M. Vauclin. Overland flow and infiltration modelling for small plots during unsteady rain: numerical results versus observed values. *Journal of Hydrology*, 228(3-4):265–282, 2000.
[16] F.R. Fiedler and J.A. Ramirez. A numerical method for simulating discontinuous shallow flow over an infiltrating surface. *International Journal for Numerical Methods in Fluids*, 32:219–239, 2000.
[17] J.F. Gerbeau and B. Perthame. Derivation of Viscous Saint-Venant System for Laminar Shallow Water; Numerical Validation. Research Report RR-4084, INRIA, 2000. Projet M3N.
[18] R. Ghanem and P. Spanos. *Stochastic finite elements: a spectral approach*. Dover, 2003.
[19] W.H. Green and G. Ampt. Studies on soil physics: 1, flow of air and water through soils. *Journal of Agricultural Science*, 4:1–24, 1911.
[20] A. Harten and S. Osher. Uniformly high-order accurate nonoscillatory schemes. I. *SIAM J. Numer. Anal.*, 24(2):279–309, 1987.
[21] W.K. Hastings. Monte carlo samping methods using markov chains and their applications. *Biometrika*, pages 97–109, 1970.
[22] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.
[23] D. Huard and A. Mailhot. Calibration of hydrological model gr2m using bayesian uncertainty analysis. *Water Resources Research*, 44:1–19, 2008.
[24] R.K. Jhorar, J.C. van Dam, W.G.M Bastiaanssen, and R.A. Feddes. Calibration of effective soil hydraulic parameters of heterogeneous soil profiles. *Journal of Hydrology*, 285:233–247, 2004.
[25] G Kuczera and E Parent. Monte carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. *Journal of Hydrology*, 211(1-4):69–85, 1998.
[26] E. Laloy and C. L. Bielders. 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*, 60:1005–1021(17), December 2009.
[27] F. Marche. *Theoretical and Numerical Study of Shallow Water Models. Applications to Nearshore Hydrodynamics*. PhD thesis, Université de Bordeaux, France, 2005.
[28] R.G. Mein and C.L. Larson. Modeling infiltration during a steady rain. *Water Resources Research*, 9(2):384–394, 1973.
[29] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
[30] H.O. Sharif, F.L. Ogden, W.F. Krajewski, and M. Xue. Numerical simulations of radar rainfall error propagation. *Water Resour. Res*, 38:10–1029, 2002.
[31] V. Sheikh, E. van Loon, R. Hessel, and V. Jetten. Sensitivity of lisem predicted catchment discharge to initial soil moisture content of soil profile. *Journal of Hydrology*, 393(3-4):174–185, 2010.
[32] I.M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414, 1993.

[33] J.J. Stoker. *Water Waves: The Mathematical Theory with Applications*. Wiley-Interscience, 1992.
[34] M. Thiemann, M. Trosset, H. Gupta, and S. Sorooshian. Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, 37(10):2521–2535, 2001.
[35] O. Tietje and V. Hennings. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within fao textural classes. *Geoderma*, 69(1-2):71–84, 1996.
[36] O. Tietje and O. Richter. Stochastic modeling of the unsaturated water flow using auto-correlated spatially variable hydraulic parameters. *Modeling Geo-Biosphere Processes*, 1:163–183, 1992.
[37] M.T. Van Genuchten. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44:892–898, 1980.
[38] P.L. Viollet, J.P. Chabard, P. Esposito, and D. Laurence. *Mécanique des fluides appliquée*. Presses de l'Ecole Nationale des Ponts et Chaussées, 1998.
[39] J.A. Vrugt, H.V. Gupta, W. Bouten, and S. Sorooshian. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8), 2003.
[40] J.A. Vrugt, C.J.F. ter Braak, M.P. Clark, J.M. Hyman, and B.A. Robinson. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. *Water Resources Research*, 44, Dec 2008.
[41] P. Wang and D. M. Tartakovsky. Probabilistic predictions of infiltration into heterogeneous media with uncertain hydraulic parameters. *Int. J. Uncert. Quant.*, 1(1):35–47, 2011.
[42] P. Wang and D. M. Tartakovsky. Reduced complexity models for probabilistic forecasting of infiltration rates. *Adv. Water Resour.*, 34:375–382, 2011.
[43] J.H.M Wösten, Ya. A. Pachepsky, and W.J. Rawls. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, 251(3-4):123–150, 2001.