Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process meta-modelling

Jeremy Rohmer[1*], Evelyne Foerster[1]

[1]BRGM, Natural Risks and Safety of CO_2 geological Storage, 3, av. Claude Guillemin BP 36009, 45060 Orléans Cedex 2, FRANCE

e-mail: j.rohmer@brgm.fr, e.foerster@brgm.fr

Abstract

Large-scale landslide prediction is typically based on numerical modelling, with computer codes generally involving a large number of input parameters. Addressing the influence of each of them on the final result and providing a ranking procedure may be useful for risk management purposes. This can be performed by a variance-based global sensitivity analysis. Nevertheless, such an analysis requires a large number of computer code simulations, which appears impracticable for computationally-demanding simulations, with computation times ranging from several hours to several days. To overcome this difficulty, we propose a "meta–model"-based strategy consisting in replacing the complex simulator by a "statistical approximation" provided by a Gaussian-Process (GP) model. This allows computation of sensitivity measures from a limited number of simulations. For illustrative purposes, the proposed methodology is used to rank in terms of importance the properties of the

[*]Corresponding author: J. Rohmer, e-mail: j.rohmer@brgm.fr , Tel. + 33 2 38 64 30 92, Fax. +33 2 38 64 36 89

elastoplastic model describing the complex behaviour of the slip surface in the La Frasse landslide (Switzerland).

One limitation of the GP-based methodology is that the computation of sensitivity measures is associated with uncertainty as the simulator is approximated using a training sample of small size, *i.e.* a limited knowledge on the "true" simulator. This source of uncertainty can be taken into account by treating the GP model from a Bayesian perspective. This provides the full posterior probability distribution associated with the sensitivity measures, which can be summarized by a confidence interval to outline the regions where the GP model is "unsure". We show that this methodology is able to provide useful guidelines for the practical decision-making process and suggest further site investigations.

Keywords: Landslide model; computationally demanding code; global sensitivity analysis; Sobol' indices; meta-model; Gaussian Process.

1 Introduction

Landslides are very complex phenomena controlled by a range of processes. Geological history, lithology and structure, slope relief and shape, weather and climate, seismicity and human activity can be identified as the main causative factors (Crosta and Clague, 2009). The associated risk to communities can be high (*e.g.* Evans *et al*., 2002) and thus, predicting landslide behaviour is a major concern.

Due to the recent advances in computer modelling (e.g. in processor performance) and in particular in the finite element method (*e.g.* van den Ham *et al*., 2009), numerical models are commonly used in practice to get a better understanding of the landslide behaviour and to predict its evolution. The main drawback of such models is the high number of input factors required for analysis. Global sensitivity analysis of complex numerical models can then be

used to determine: (1) which input factors contribute the most to the output variability (within the "factors' prioritisation setting" as described by Saltelli *et al.*, 2008); (2) which input factors interact with each other; and (3) which input factors are insignificant and can be eliminated to "simplify" the model (within a "factors' fixing setting", Saltelli *et al.*, 2008). Such an analysis is useful in identifying which input factors require further investigations to reduce uncertainties in the computer code results, hence providing guidelines for risk management (Saltelli, 2002b).

Among the existing sensitivity methods, variance-based methods have proved to be effective (Saltelli *et al*., 2000). In this article, we focus on the method of Sobol' indices (Sobol', 1993; Archer *et al*., 1997; Sobol' and Kucherenko, 2005). Unlike traditional linear or rank regression-based methods, these indices allow representing the sensitivity of a general model without assuming any kind of linearity or monotonicity in the model (Saltelli and Sobol', 1995). In practice, the computation of Sobol' indices uses a Monte Carlo sampling strategy. An example of application in the field of landslide modelling with applications of moderate complexity is provided by Hamm *et al*. (2006). Such an approach, however, appears hardly applicable for more computationally demanding models, as it requires a large number of computer code evaluations. For instance, the study of Hamm *et al.*, (2006) required ten thousand model realisations, corresponding to about 20 hours of computation time (on a 2 GHz Pentium 4 PC). The same sensitivity analysis would require 208 days using a model that takes 30 minutes and 2500 days using a model that takes 6 hours to compute.

To overcome this difficulty, a first solution is to use a distributed parallel computing methodology, thus requiring an appropriate grid computing architecture and the optimization of computing resources (*e.g.* Dupros *et al*., 2006; Boulahya *et al*., 2007).

In this paper, an alternative is proposed using a limited number of computer code runs (also named "simulator", O' Hagan, 2006), which consists in replacing (i.e. approximating) the

simulator by a surrogate model with low computation time, also named a “meta-model”, to compute the Sobol’ indices (*i.e.* the sensitivity measures). Various “meta-models” exist (e.g. linear regression, nearest neighbour method, Multivariate Adaptative Regression Spline, neural network and Gaussian Process); see, for example, Storlie *et al.*, 2009 for a recent review.

The meta-model uses a limited number of simulator runs, *i.e.* input-output pairs (corresponding to the training sample), to infer the values of the complex simulator output given a yet-unseen input configuration. Such an approximation introduces a source of uncertainty referred to as “code uncertainty” associated with the meta-model (O’ Hagan, 2006), so that the sensitivity measures computed with the meta-model are “uncertain”.

In the present article, we choose to solve the described problem of approximation (and of inference) under the Bayesian formalism treating the simulator as an “unknown” function in the sense that the simulator output for any yet-unseen input configuration is unknown until the simulator is actually run for the considered configuration (Oakley and O’ Hagan, 2004). We choose to use the concept of an emulator corresponding to a statistical approximation so that a prior probability distribution is assigned to the simulator outputs and updated according to the usual Bayesian paradigm given the training sample. This approach returns not only the most likely value for the output given any input configuration, but also an entire probability distribution (O’ Hagan, 2006). This distribution can be used to estimate a level of confidence when the predictive quality of the meta-model is not high due to a small training data (see, for instance, Marrel *et al*., 2008 and 2009, Storlie *et al.,* 2009). A Gaussian Process (GP) is chosen as the prior model for the simulator. It has been widely used when designing computer experiments (Sacks *et al*., 1989; Kennedy and O’Hagan, 2001; Santner *et al*., 2003).

In the first section, the Sobol’ decomposition method is described in the general framework of the variance-based global sensitivity approach (Saltelli *et al.*, 2008).

Then, the GP model used as a meta-model of the computationally intensive simulator is described in the framework of the stochastic processes for computer code experiments under the Bayesian regression formalism. The methodology for computing the Sobol' indices using the GP model is described and illustrated in two applications. The first application is a simple analytical model based on "infinite slope analysis" (Hansen, 1984). This allows us to compare the sensitivity measures computed using the "true" model with those computed using the GP model. Finally, the application of this methodology to a La Frasse (Switzerland) landslide model (Laloui *et al*., 2004) is presented and we show how to use the sensitivity measures to guide the decision-making process for further site investigations.

2 Global sensitivity analysis by the Sobol' decomposition method

2.1 *Introduction on the variance-based sensitivity analysis*

Consider the simulator g and the scalar output y determined from a vector of n input factors $\mathbf{x} = \{x_i\}_{i=1,\ldots,n}$ so that $y = g(\mathbf{x})$.

Considering the n-dimensional vector as a random vector of independent random variable X_i, then the output Y is also a random variable (as a function of a random vector). A variance-based sensitivity analysis aims at determining the part of the total unconditional variance V_Y of the output Y resulting from each input random variable X_i. The total variance V_Y can be expressed as follows (Saltelli *et al*., 2000 & 2008):

$$V_Y = \sum_i V_i + \sum_{i<j} V_{ij} + \sum_{i<j<l} V_{ijl} + \ldots + V_{ij\ldots n} \qquad (1)$$

where the partial variance V_i and $V_{ij\ldots n}$ read:

$$V_i = Var[E[Y|X_i = x_i]]$$
$$V_{ij\ldots n} = Var[E[Y|X_i = x_i, X_j = x_j, \ldots, X_n = x_n]] - V_i - V_j - \ldots - V_n \quad (2)$$

with $E[Y|X_i = x_i]$, the expectation of Y given that the i^{th} input factor X_i has a fixed value x_i and $E[Y|X_i = x_i, X_j = x_j, \ldots, X_n = x_n]$ the conditional expectation of Y given that the i^{th} input factor X_i has a fixed value x_i, the j^{th} input factor X_j has a fixed value x_j, …etc.

The variance of the conditional expectation V_i represents the first order effect of the input factor X_i taken alone, whereas the higher order indices account for possible mixed influence of various input factors.

2.2 *The Sobol' decomposition method*

2.2.1 Presentation

To determine the partial variances of Y, Sobol' (1993) proposes the following decomposition of g into summands of increasing dimension provided that g is integrable:

$$g(\mathbf{x}) = g_0 + \sum_{i=1}^{n} g_i(x_i) + \sum_{i \le j}^{n} g_{ij}(x_i, x_j) + \ldots + g_{1,\ldots,n}(x_1, \ldots, x_n) \quad (3)$$

where g_0 corresponds to the mean constant value of the function g and each term can be evaluated through multidimensional integrals as follows:

$$g_0 = E[Y] = \int_0^1 \ldots \int_0^1 g(\mathbf{x}) d\mathbf{x} = \int_0^1 \ldots \int_0^1 g(x_1, \ldots, x_n) dx_1 \ldots dx_n \quad (4)$$

$$g_i(x_i) = E[Y|X_i = x_i] - g_0 = \int_0^1 \ldots \int_0^1 g(\mathbf{x}) d\mathbf{x}_{-i} - g_0 \quad (5)$$

$$g_{ij}(x_i, x_j) = E[Y|X_i = x_i, X_j = x_j] - g_0 - g_i - g_j = \int_0^1 \ldots \int_0^1 g(\mathbf{x}) d\mathbf{x}_{-(ij)} - g_0 - g_i(x_i) - g_j(x_j) \quad (6)$$

with $d\mathbf{x}_{-i}$ denotes the integration over all input factors except x_i and $d\mathbf{x}_{-(ij)}$, the integration over all input factors except both x_i and x_j. Similar formulae can be obtained for higher order terms.

The total variance V_Y can then be expressed as:

$$V_Y = \int_0^1 \ldots \int_0^1 g(x_1, ..., x_n)^2 dx_1, ...dx_n - g_0^2 \quad (7)$$

while the partial variances read as follows:

$$V_{i_1 \ldots i_s} = V[g_{i_1 \ldots i_s}] = \int_0^1 \ldots \int_0^1 g_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s})^2 dx_{i_1} \ldots dx_{i_s} \quad (8)$$

with $1 \le i_1 < \ldots < i_s \le n$ and $s = 1, \ldots, n$.

2.2.2 **Definition of the Sobol' indices**

The Sobol' indices $S_{i_1 \ldots i_s}$ describe which amount of the total variance is due to the uncertainties of input factors in the set $\{i_1, \ldots, i_s\}$ and is expressed as the ratio between $V_{i_1 \ldots i_S}$ and V_Y, respectively the partial and total variances .

The first-order sensitivity index S_i for input factor X_i is expressed as follows:

$$S_i = \frac{V_i}{V_Y} \quad (9)$$

The sensitivity measure S_i is referred to as "the main effect of X_i" and can be interpreted as the expected reduction in the total variance of the output Y (*i.e.* representing the uncertainty in Y) if the true value of the input factor X_i was known. This index provides a measure of importance useful to rank the input factors (Saltelli *et al.*, 2000 & 2008).

The main effect and the higher order Sobol' indices satisfy the following property:

$$\sum_{i=1}^{n} S_i + \sum_{i<j} S_{ij} + \ldots + S_{ij\ldots n} = 1 \qquad (10)$$

Defining S_{-i} as the sum of all the terms $S_{i_1,\ldots,i_s}$ but the i^{th} index, the total effect index S_{Ti} of X_i is defined as the total contribution of the i^{th} input factor to the total variance. It reads as follows (using eq. 10):

$$S_{Ti} = 1 - S_{-i} \qquad (11)$$

$S_{Ti} = 0$ means that the input factor X_i has no effect. Thus, it can be fixed at any value over its uncertainty range (Saltelli *et al.*, 2008).

As the total number of sensitivity indices reaches 2^n-1 (Saltelli *et al.*, 2000), hence representing a high computational cost, the sensitivity analysis is generally limited, in practice, to the pair of indicators corresponding to the main effect S_i and to the total effect S_{Ti} of X_i (Saltelli *et al.*, 2008).

2.2.3 **Numerical implementation**

The evaluation of the Sobol' indices can be carried out through a Monte Carlo sampling strategy (Saltelli *et al.,* 2000), which remains an approximation of the true value of the

sensitivity indices. Thus, the quality of the approximation directly depends on the sample size. Let us consider m sampled elements $\left\{\mathbf{x}^{(j)}\right\}_{j=1,\ldots,m}$ in the n-dimensional space of input factors:

$$\begin{cases} \hat{g}_0 \approx \dfrac{1}{m}\sum_{j=1}^{m} g\left(\mathbf{x}^{(j)}\right) \\ \hat{V}_Y \approx \dfrac{1}{m}\sum_{j=1}^{m} g^2\left(\mathbf{x}^{(j)}\right) - \hat{g}_0^2 \\ \hat{V}_i \approx \dfrac{1}{m}\sum_{j=1}^{m} g\left(\mathbf{x}_{S_a}^{-(ij)}, \mathbf{x}_{S_a}^{(ij)}\right) \cdot g\left(\mathbf{x}_{S_b}^{-(ij)}, \mathbf{x}_{S_a}^{(ij)}\right) - \hat{g}_0^2 \end{cases} \quad (12)$$

where $\mathbf{x}_{-ij} = [x_{1j},\ldots,x_{(i-1)j},x_{(i+1)j},\ldots,x_{nj}]$ and the subscripts S_a and S_b in eq. (12) indicate that two sampling data matrices are being used.

The main and total effects can be estimated using the sampling strategy proposed by Saltelli (2002a) at a computation time cost of $m\times(n+2)$ model evaluations. Additional computational efficiency can be achieved by making best use of sampling designs, for instance Sobol' quasi-random sequences, and estimators, for instance Jansen's estimator (Saltelli *et al.*, 2010). However, the computational effort for simulators with computation time ranging from several hours to several days may still be high and the present work focuses on a strategy based on Gaussian Process meta-modelling to reduce this effort.

3 Gaussian Process (GP) modelling

3.1 *Description of the stochastic process framework*

First, the deterministic response y(**x**) of the simulator is treated as a realization of a random variable $Y(\mathbf{x})$, which can be decomposed into a deterministic function f, which represents the mean (*i.e.* expectation of Y), and a stochastic function Z as proposed by Sacks *et al.* (1989) and reads as follows:

$$Y(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}) \quad (13)$$

Note that the case of multiple outputs is beyond the scope of the present article and the interested reader is advised to refer to Le and Zidek (2006).

Without prior information between the modelling inputs and outputs, f is chosen as a multivariate linear regression model (Martin and Simpson, 2005) so that:

$$f(\mathbf{x}) = \mathbf{F}(\mathbf{x}) \cdot \mathbf{B} = \sum_{i=0}^{n} b_i \cdot f_i(\mathbf{x}) \quad (14)$$

where $\mathbf{B} = \{b_i\}_{i=0,\ldots,n}$ is the regression parameter vector; $\mathbf{F}(\mathbf{x}) = [f_0(\mathbf{x}), f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})]$ is the corresponding regression vector with f_i (i=0,1,…,n), the basis functions. Assuming linearity for the mean f, we have: $f_0(\mathbf{x})=1$ and $f_i(\mathbf{x})=x_i$ for $i=1,\ldots,n$.

The stochastic part Z can be seen as a confidence measure on the model output mean. It represents a zero mean random process, characterized by its $n \times n$ covariance matrix $\mathbf{\Sigma_S}$ so that an element at the j^{th} row and k^{th} column of $\mathbf{\Sigma_S}$ is expressed as:

$$\Sigma_{S\,j,k} = \mathrm{Cov}[Z(\mathbf{x}^{(j)}), Z(\mathbf{x}^{(k)})]_{j,k=1\ldots N_S} = \sigma^2 \cdot \rho\left(\mathbf{x}^{(j)}, \mathbf{x}^{(k)}\right)_{j,k=1\ldots N_S} \quad (15)$$

where $\mathbf{x}^{(j)}$ is the j^{th} input vector (with $j=1\ldots N_S$) and $\mathrm{Var}[Z]=\sigma^2$. The correlation function ρ provides the interpolation and spatial correlation properties.

Treating the simulator g within the stochastic process framework allows the deterministic part f to account for the global behaviour of g, whereas the correlation terms allow the meta-model to "locally" interpolate the known data by introducing a strong correlation in the

neighbourhood of these points. Complex input-output behaviours are hence better represented (*e.g.* see Langewisch and Apostolakis, 2010).

3.2 *Description of the Gaussian correlation model*

Various authors (*e.g.* Stein, 1999; Le and Zidek, 2006) have discussed different types of correlations functions. For our purposes, the study is restricted to the Gaussian correlation model so that the value of the correlation matrix only depends on the normalised distance between the input vectors $\mathbf{x}^{(j)}$ and $\mathbf{x}^{(k)}$. Assuming the correlation model is invariant to any translation in the input space (*e.g.* Rasmussen and Williams, 2006), the Gaussian correlation function reads as follows:

$$\rho\left(\mathbf{x}^{(j)},\mathbf{x}^{(k)}\right)=\prod_{i=1}^{n}\rho_i\left(x_i^{(j)},x_i^{(k)}\right)=\exp\left\{-\sum_{i=1}^{n}\cdot\frac{\left\|x_i^{(j)}-x_i^{(k)}\right\|^2}{\theta_i}\right\} \quad (16)$$

where $\boldsymbol{\theta}=\{\theta_i\}_{i=1,\ldots,n}$ are the correlation lengths, also referred to as "hyper-parameters" (*e.g.* see Rasmussen and Williams, 2006). θ_i parameter quantifies the rate at which the output varies as i^{th} input factor is changed.

Note that in case of data measurements errors or non deterministic computer code, a constant regularization term referred to as "nugget effect" may be defined, hence introducing a white noise.

3.3 *Principle and prediction under the Bayesian formalism*

In this paper, we focus on the stationary GP model which fits the stochastic framework and has been broadly used in designing computer experiments (Sacks *et al.*, 1989; Kennedy and O'Hagan, 2001; Santner *et al.*, 2003).

Let us define the training sample as the N_S training data pairs $\{\mathbf{X}_S, \mathbf{Y}_S\}$, which represent a mapping between the spaces of input factors $\mathbf{X}_S = \left\{\mathbf{x}^{(j)}\right\}_{j=1...N_S}$, with $\mathbf{x}^{(j)} = \left\{x_i^{(j)}\right\}_{i=1,...,n}$, and the outputs $\mathbf{Y}_S = \left\{y^{(j)}\right\}_{j=1...N_S}$, obtained through the N_S selected simulator runs so that $y^{(j)} = \left\{f(\mathbf{x}^{(j)})\right\}_{j=1,...,N_S}$.

In a first step, constructing the GP model implies considering the simulator output Y as a random variable, which is assumed to follow a multivariate Gaussian distribution (denoted G) for any random vector of input factors $\mathbf{X}$. This assumption represents our prior belief on the simulator. Using the training sample $\{\mathbf{X}_S, \mathbf{Y}_S\}$, the Bayes theorem is used to refine the mentioned prior information in order to yield the posterior distribution of the output, known as the "emulator" (e.g. O' Hagan, 2006). This latter not only provides an expected value for any "yet-unseen" input configuration, but it also gives an entire posterior distribution given the observed data.

Formally, the probability $p(\mathbf{Y}_S|\mathbf{X}_S)$ of obtaining $\mathbf{Y}_S$ given $\mathbf{X}_S$, is expressed under the GP assumption, as follows:

$$p(\mathbf{Y}_S|\mathbf{X}_S) \sim G(F(\mathbf{X}_S).\mathbf{B}, \mathbf{\Sigma}_S) \tag{17}$$

Considering a new vector $\mathbf{x}^{(N_S+1)}$ of input factors and the associated output $Y^{(N_S+1)} = Y(\mathbf{x}^{(N_S+1)})$, the joint probability distribution of the random variables $\left(\mathbf{Y}_S, Y^{(N_S+1)}\right)$ reads as follows:

$$p\left(\mathbf{Y}_S, Y^{(N_S+1)} \middle| \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right) \sim G\left(\begin{bmatrix} \mathbf{F}_S \\ F\left(\mathbf{x}^{(N_S+1)}\right) \end{bmatrix} \cdot \mathbf{B}, \begin{bmatrix} \mathbf{\Sigma}_S & k\left(\mathbf{x}^{(N_S+1)}\right) \\ k\left(\mathbf{x}^{(N_S+1)}\right)^T & \sigma^2 \end{bmatrix}\right) \tag{18a}$$

where $\mathbf{F}_S = [F(\mathbf{x}^{(1)}), F(\mathbf{x}^{(2)}), ..., F(\mathbf{x}^{(N_S)})]$ corresponds to the regression matrix.

$k\left(\mathbf{x}^{(N_S+1)}\right)$ represents the vector of correlation functions between each of the N_S training input vectors $\mathbf{x}^{(j)}$ and the new element $\mathbf{x}^{(N_S+1)}$. It can be written as:

$$k(\mathbf{x}^{(N_S+1)}) = \sigma^2.\left[\rho\left(\mathbf{x}^{(1)}, \mathbf{x}^{(N_S+1)}\right),..., \rho\left(\mathbf{x}^{(N_S)}, \mathbf{x}^{(N_S+1)}\right)\right]^T \quad (17b)$$

Within the Bayesian framework, the posterior distribution of the computed output random variable $Y^{(N_S+1)}$ is conditioned on the "observed" (i.e. actually calculated) values corresponding to the training sample $\{\mathbf{X}_S, \mathbf{Y}_S\}$, given the new element $\mathbf{x}^{(N_S+1)}$, and follows a multivariate Gaussian distribution (Von Mises, 1964):

$$\begin{aligned} p\left(Y^{(N_S+1)} \middle| \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right) &= \frac{p\left(\mathbf{Y}_S, Y^{(N_S+1)} \middle| \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right)}{p\left(\mathbf{Y}_S \middle| \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right)} \\ &\sim G\left(E\left[Y^{(N_S+1)} \middle| \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right], Var\left[Y^{(N_S+1)} \middle| \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right]\right) \end{aligned} \quad (18a)$$

where:

$$E\left[Y^{(N_S+1)} \middle| \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right] = F\left(\mathbf{x}^{(N_S+1)}\right) \cdot \mathbf{B} + k\left(\mathbf{x}^{(N_S+1)}\right)^T \cdot \boldsymbol{\Sigma}_S^{-1} \cdot \left(\mathbf{Y}_S - \mathbf{F}_S \cdot \mathbf{B}\right) \quad (18b)$$

and:

$$Var\left[Y^{(N_S+1)} \middle| \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}\right] = \sigma^2 - k\left(\mathbf{x}^{(N_S+1)}\right)^T \cdot \boldsymbol{\Sigma}_S^{-1} \cdot k\left(\mathbf{x}^{(N_S+1)}\right) \quad (18c)$$

with $\boldsymbol{\Sigma}_S^{-1}$, the inverse of the covariance matrix associated to the training input data $\mathbf{X}_S$.

The conditional mean of eq. (18b) is used as a predictor and the conditional variance in eq. (18c) corresponds to the mean square error of the predictor term. Provided that the new candidate $\mathbf{x}^{(N_S+1)}$ is far away from the training input data $\mathbf{X}_S$, the term $k\left(\mathbf{x}^{(N_S+1)}\right)^T \cdot \boldsymbol{\Sigma}_S^{-1} \cdot k\left(\mathbf{x}^{(N_S+1)}\right)$ will be small so that the predicted variance will be large.

In a more general manner, if we consider two new test candidates **u** and **v**, the general expression of the conditional GP model can be written as:

$$\left(Y|\mathbf{Y}_S,\mathbf{X}_S,\mathbf{B},\sigma,\boldsymbol{\theta}\right)\sim GP\left(E\left[Y(\mathbf{u})|\mathbf{Y}_S,\mathbf{X}_S,\mathbf{B},\sigma,\boldsymbol{\theta}\right],\mathrm{Cov}\left[Y(\mathbf{u}),Y(\mathbf{v})|\mathbf{Y}_S,\mathbf{X}_S,\mathbf{B},\sigma,\boldsymbol{\theta}\right]\right) \quad (19)$$

The conditional mean is used as a predictor and is expressed as in eq. (18b).

The conditional covariance provides the confidence on the prediction and reads as follows:

$$\mathrm{Cov}\left[Y(\mathbf{u}),Y(\mathbf{v})|\mathbf{Y}_S,\mathbf{X}_S,\mathbf{B},\sigma,\boldsymbol{\theta}\right]=\sigma^2\,\rho(\mathbf{u},\mathbf{v})-k(\mathbf{u})^T\cdot\boldsymbol{\Sigma}_S^{-1}\cdot k(\mathbf{v}) \quad (20)$$

The regions of the input space, where few data are available will be underlined by higher variance.

The main difficulty in constructing a conditional GP model given a training sample resides in determining the parameters corresponding to the regression coefficient vector **B**, the hyper-parameters $\boldsymbol{\theta}$ and the variance σ^2. A first approach consists in estimating them as solutions of the optimization problem using the maximum likelihood method, *e.g.* implemented in the GEM-SA software (O'Hagan, 2006) and the MATLAB toolbox DACE (Lophaven, 2002). However, the optimisation algorithms used for the parameters identification may show limitations, especially in case of high dimension problem (*e.g.* see Marrel *et al.*, 2008). Besides, such an approach may underestimate the variance in the predictions of new observations (Cressie, 1993).

In this paper, an approach based on the Bayesian framework (*e.g.* Rasmussen 1996) is chosen so that the hyper-parameters are given prior distributions $p(\boldsymbol{\theta})$. In the Bayesian framework, the Markov Chain Monte Carlo methods (Gilks *et al.*, 1996) are used to integrate over the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{X}_S,\mathbf{Y}_S)$ associated with the GP parameters *i.e.* the training sample is used to update the GP parameters. The posterior distribution of the hyper-parameters will be hence concentrated on values that are consistent with actually observed data. This procedure is implemented in the package named "TGP" of the "R" software ("R" Development Core

317 Team, 2009) by Gramacy and co-workers (Gramacy, 2007; Gramacy and Taddy, 2010).
318 Further theoretical details can be found in Gelman *et al.* (1995) and Gramacy and Herbert
319 (2009). Though computationally more intensive (Storlie *et al.*, 2009), this approach presents
320 the attractive feature to incorporate the uncertainty related to the construction of the GP model,
321 so that the level of confidence associated with the "meta-model"-computed sensitivity
322 measures also takes this source of uncertainty into account (see step 4 of section 3.4.).

323 ## 3.4 *A "GP-based" methodology for sensitivity analysis*

324 In this section, we describe the methodology to compute the sensitivity measures (*i.e.* the
325 Sobol' indices) using a GP model as a surrogate model of the computationally intensive
326 simulator.

327 ### *3.4.1* Step 1: representation of the input factor uncertainty

328 The first step is to characterize and mathematically represent the uncertainty (range and form
329 of the probability distribution) on each of the input factors. This representation can have a
330 strong influence on the final sensitivity results, hence on risk management decision making
331 (Saltelli, 2002b).
332 Representing the uncertainty through empirical probability distributions requires a large
333 amount of data (laboratory or *in situ* measurements), which may not be practical in many
334 situations. Thus, knowledge on the range of uncertainty is commonly evaluated either based
335 on physical reasoning, on analogies with similar cases or simply from expert opinions,
336 whereas the mathematical representation of the probability distribution may either be
337 theoretically known or assumed. In a situation where "sparse, vague and incomplete" data are
338 available, a common approach consists in assigning a uniform probability distribution based
339 on the "maximum entropy" approach (Gzyl, 1995).

3.4.2 Step 2: setting training data

The objective then is to run the simulator for a limited number of times N_S in order to create a mapping between the input factor and the computer code output domain. The number N_S should be defined as a compromise between the minimization of the computation time cost and the maximization of the input factor domain exploration (directly linked with the accuracy and reliability of the GP model, see step 3).

In this view, we propose to use the Latin hypercube sampling method (McKay *et al.*, 1979) in combination with the "maxi-min" space-filling design criterion (Koehler and Owen, 1996). More sophisticated strategies exist mainly based on sequentially adaptive design of experiments adding new training candidates where the predictive uncertainty is high (*e.g.* Gramacy and Herbert, 2009). The use of such approaches is beyond the scope of this paper.

3.4.3 Step 3: constructing the GP model

Using the GP model instead of the simulator introduces an additional source of uncertainty referred to as "code uncertainty" (O' Hagan, 2006). In the regions where the true simulator is not evaluated, we are uncertain about what the "true" simulator would introduce. This sort of uncertainty can be reduced by increasing our knowledge of the true simulator, *i.e.* by increasing the training sample size.

Except when a "nugget" effect is included, the GP model is an exact interpolator, so that residuals of the training data cannot be directly used to validate the approximation (Marrel *et al.*, 2008). The key aspect for validating the "statistical" approximation is to estimate the expected level of fit (*i.e.* predictive quality) of the GP model to a data set that is independent of the data (*i.e.* "yet-unseen" data) that were used to train the GP model.

As additional simulator runs are costly, using a test sample of new data might be impractical and cross-validation procedures such as the "*k*-fold" cross-validation technique (Hastie, 2002) should be used. In this cross-validation procedure, the initial training sample is randomly

365 partitioned into *k* subsets. In a first step, a single subset is used as the validation data for
366 testing the GP model, and the remaining *k*-1 subsets are used as training data for the
367 construction of the GP model. For each step, the *k* validation data are estimated and the
368 coefficient of determination R^2 for the procedure is computed as follows:

369 $$R^2(\mathbf{y},\hat{\mathbf{y}}) = 1 - \frac{\sum_{i=1}^{k}(\hat{y}_i - y_i)}{\sum_{i=1}^{k}(y_i - \bar{\mathbf{y}})} \quad (21)$$

370 where $\mathbf{y}$ represents the vector of observations in the validation set; $\bar{\mathbf{y}}$ is the mean of the
371 corresponding sample and $\hat{\mathbf{y}}$, the vector of predicted values using the GP model.
372 The coefficient R^2 provides a metric of the predictive quality so that a value close to 100 %
373 indicates that the GP model is successful in matching the validation data. A typical threshold
374 of 80 % is commonly used to qualify the predictive quality as "satisfactory" (*e.g.* Marrel et
375 al., 2008).
376 The cross-validation process is then repeated *k* times using each of the *k* subsets as validation
377 samples. For small training sets, the cross validation procedure with *k*=1 is usually used
378 corresponding to the so-called "leave-one-out" cross validation procedure.

379 *3.4.4* Step 4: estimating the sensitivity measures

380 The most likely value $\mu(S_i)$ for the sensitivity measures is computed using the conditional
381 mean of the GP model in eq. 18b. Additional useful information for risk management
382 purposes is the level of confidence (or accuracy) related to the sensitivity analysis based on
383 the GP model. A confidence interval $CI(S_i)$ can be defined with bounds corresponding to the
384 5% and to the 95 % quantile of the full posterior distribution of the sensitivity measures. This
385 confidence interval both summarizes the "code uncertainty" associated with the meta-model
386 (O' Hagan, 2006) and the uncertainty on the estimation of the GP model parameters (see
387 section 3.3).

4 Illustrative analytical model

In this section, we consider the infinite slope analytical model (*e.g.* Hansen, 1984) in order to illustrate the methodology described in section 3.4. This model is of course not a computationally demanding function, but we imagine it as representing a calculation that may take several minutes or even hours of computation to evaluate. Besides, using this analytical model also allows us to compare the results of the sensitivity analysis using the "true" model with those using the GP model.

4.1 *Description of the analytical model*

The stability of the infinite slope model as depicted in Fig. 1 is evaluated by deriving the factor of safety *FS*, which corresponds to the ratio between the resisting and the driving forces acting on the slope (eq. 22):

$$FS = \frac{C + (\gamma - m \cdot \gamma_w) \cdot z \cdot \cos(\theta) \cdot \cos(\theta) \cdot \tan(\phi)}{\gamma \cdot z \cdot \sin(\theta) \cdot \cos(\theta)} \quad (22)$$

[Fig. 1 about here]

where C is the cohesion of the soil material; ϕ, the friction angle; θ, the slope angle; γ, the soil unit weight; γ_w, the water unit weight; z, the thickness of slope material above the slip plane; and m, the ratio between thickness of surficial saturated slope material and z. If FS is lower than 1.0 the potential for failure is high.

4.2 *Sensitivity analysis of the analytical model*

For illustrative purposes, we only considered the thickness z and the slope angle θ as uncertain input factors. The other input factors were assumed fixed: C=10 kPa, ϕ=25°, γ=22 kN.m^{-3}, γ_w= 9.81 kN.m^{-3} and m=90 %.

The objective was to identify whether z or θ contributes the most to the FS variability within a "factors' prioritisation setting". It is assumed that very sparse data are available to characterize the uncertainty on these input factors so that z uniformly varies between 5 and 25 m and θ uniformly varies between 25° and 35° (step 1). We generated two different training samples of respectively 6 and 20 training data of the form $\{z ; \theta ; FS\}$, using the Latin hypercube sampling approach (step 2) and for each training sample, a GP model was constructed.

[Fig. 2 about here]

Fig. 2 (top) shows the comparison between the values of FS obtained from direct simulations on a grid of 10×10 in the input factor domain [5 ; 25]×[25 ; 35] (straight line) and from the prediction on the same grid using the GP model (dashed line) for both training samples (Fig. 2, left for 6 training data and right for 20 training data). The coefficient of determination or goodness of fit (eq. 22) estimated for both GP models was equal to 90.9 % for the first training sample and to 98.8 % for the second one, hence showing a very good match for both meta-models. The quality of the approximation was then estimated through a "leave-one-out" cross validation procedure (step 3): we obtained a coefficient of determination of 96.2 % for the first sample and 99.7 % for the second one, hence indicating a "high" predictive quality. The estimated FS using both GP models (Fig. 2, middle) were compared to the "true" observed FS. The closer the dots to the straight black line, the better the approximation.

The results for the computation of the main effects required (step 4) $m\times(n+2)=2500\times(2+2)=10000$ model evaluations using the sampling strategy of Saltelli (2002a). The most likely of the main effects calculated with both GP models (blue dots in Fig. 2, bottom) were compared to the main effects obtained from direct simulations (red dots on Fig. 2, bottom) by means of the R package "sensitivity" and the function referred to as Sobol2002 (available at http://cran.r-project.org/web/packages/sensitivity/index.html). These results are summarized in Table 1.

[Table 1 about here]

We see that differences are larger for the GP model constructed with the lower training sample size but, however, the "true" values for the main effects still lie within the confidence interval bounded by the 5 % and the 95 % quantile (black cross-type marker in Fig. 2, bottom). Not surprisingly, increasing the number of training samples (*i.e.* our knowledge of the true function) decreases the range of code uncertainty as well as the differences between the true values and estimates (Table 1).

5 Computationally intensive numerical model

In this section, we present the application of the proposed GP-based sensitivity analysis methodology (section 3.4.) to the landslide finite-element model originally used for the simulation of the La Frasse (Switzerland) landslide during the 1994 crisis period (Laloui *et al.*, 2004).

5.1 *General description of the landslide model*

The La Frasse landslide covers a total area of roughly 1000×1000 m², and represents an average thickness of 80 m in its upper part and 40 m in its lower part. The total volume of the

La Frasse landslide reaches 73 million m^3. Since 1975, a constant movement has been observed in its upper and central parts, varying between 10.10^{-2} and 15.10^{-2} m per year (Noverraz and Bonnard, 1988; Bonnard *et al.*, 1995). The evolution of the groundwater table is considered to be at the origin of the sliding and the instabilities were mainly observed during the 1994 crisis (over a period of nearly 300 days). Therefore, in order to assess the effect of the hydraulic regime on the geomechanical behaviour of the landslide, finite-element simulations considering a 2D cross-section through the centre of the landslide were performed by Laloui *et al.* (2004) using the finite element program GEFDYN (Aubry *et al.*, 1986).
The model is composed of 1694 nodes, 1530 quadrangular elements, and six soil layers derived from the geotechnical investigations. Fig. 3 gives an overview of the model, as well as the boundary conditions used for analysis. Instabilities observed in 1994 were triggered by pore pressure changes occurring at the base of the slide (see Laloui *et al.*, 2004 for further details).

[Fig. 3 about here]

The general behaviour of the landslide is strongly correlated to the properties of the slip surface. The complex behaviour of the slip surface material was modelled using the Hujeux elastoplastic multi-mechanism constitutive model (Aubry *et al.*, 1982; Hujeux, 1985; Lopez-Caballero *et al.*, 2007; Lopez-Caballero and Modaressi Farahmand-Razavi, 2008) and the Mohr-Coulomb criterion was assumed for the other soil materials.
The Hujeux constitutive model permits coverage of a large range of deformation and takes into account: (1) the influence of confinement and stress path on the moduli; (2) the effects of over-consolidation; and (3) the influence of the void ratio. It can be used for granular as well as clayey soil behaviours and it is based on a Coulomb type failure criterion and the critical

state concept. The volumetric and deviatoric hardening regimes implemented in the Hujeux model lead to a dependence on the consolidation pressure as in the Cam-Clay family models, and to the evolution of the plastic yield surface with the deviatoric and volumetric plastic strains. Moreover, the model accounts for dilatancy/contractance of soils and non-associated flowing behaviour with evolution of the plastic strain rate through a Roscoe-type dilatancy rule.

As outlined by Laloui *et al.* (2004), the main parameters for the slip surface materials are: (1) the bulk (K) and shear (G) elastic modules, which are assumed to depend on the mean effective stress through a power-type law of exponent n_e; (2) the critical state and plasticity parameters, essentially the friction angle ϕ at perfect plasticity, the plastic compressibility β; and (3) the dilatancy angle Ψ, appearing in the flow rule and defining the limit between soil dilatancy and contractance.

Note that these parameters can be directly measured from either *in situ* or laboratory test results (Lopez-Caballero *et al.*, 2007; Lopez-Caballero and Modaressi Farahmand-Razavi, 2008). The other Hujeux law parameters, appearing in the flow rule, the hardening and the threshold domains definition are categorized as “not-directly measurable” (Lopez-Caballero and Modaressi Farahmand-Razavi, 2008) and are estimated through numerical calibration techniques between the observed/experimental data and the simulated ones.

5.2 *Sensitivity analysis using the GP-based methodology*

The sensitivity analysis using the GP-based methodology (see section 3.4.) was carried out to assess the importance of the input factors of the Hujeux constitutive model describing the slip surface behaviour within a “factors’ prioritisation setting”, so that the main effects (first order Sobol’ indices) were used for ranking.

The quantity of interest was chosen as the horizontal displacement calculated at two observation points, namely in the upper (observation point 1, Fig. 3), and lower parts of the

landslide (observation point 2, Fig. 3). The sensitivity analysis was carried out in a dynamic manner at each step of the 300 days long crisis period (decomposed into a hundred time steps). It was focused on the main measurable parameters of the Hujeux constitutive model (total number of seven input factors), the others being kept constant *i.e.* treated with "no uncertainty". The properties of the other soil layers were assumed to be constant as well.

5.2.1 Step 1: representation of the input factor uncertainty

In this illustrative study, our objective was to explore the situation where the same "level of uncertainty" is associated with the parameters of Hujeux model: a 25 % variation around the original values identified by Laloui and co-authors (Laloui *et al.*, 2004) was affected to each of the seven input factors (Table 2). Considering no further information on the uncertainty, a uniform probability distribution was assigned to each of these input factors (see section 3.4.1.).

[Table 2 about here]

5.2.2 Step 2: setting training data

A total number of 30 input parameter configurations was generated. The resulting horizontal displacements computed over the crisis period are shown on Figure 4 for the observation points 1 and 2. For a given input configuration, a simulator run required ≈96 hours on a computer unit (CPU) with a 2.6 GHz dual core processor and 1 GB of RAM. The training sample was generated using a grid computing architecture or computer cluster composed of 30 CPU, so that all simulations were performed in parallel.

[Fig. 4 about here]

5.2.3 Step 3: constructing the GP model

At each step of the 1994 crisis period, a GP model was constructed using the 30 training data to approximate the horizontal displacements at the observation points 1 and 2.

A "leave-one-out cross-validation" procedure was carried out for each step in order to assess the predictive quality of the GP models. Fig. 5 depicts the temporal evolution of the coefficient of determination R^2 for the cross-validation procedure.

During the first half of the crisis period (first 150 days), R^2 decreases over time for both observation points between 99.9 % and ≈95 %, hence indicating that the predictive quality is "high" over this period. During the second half of the crisis period, the quality is still satisfactory if we consider observation point 2 (R^2 varying between ≈80 % and ≈95 %, see Marrel *et al.*, 2008), whereas it can be qualified as "low to moderate" for observation point 1 (R^2 steeply decreasing from ≈95 % to ≈62 %), hence indicating possibly high uncertainty on the GP model.

[Fig. 5 about here]

5.2.4 Step 4: estimating the sensitivity measures

The main effects were calculated using the sampling strategy of Saltelli (2002a), hence requiring $m\times(n+2)=1000\times(7+2)=9000$ GP model evaluations. Preliminary convergence tests were carried out for m=250, 500, 1000 and 2000: they showed that m=1000 yields satisfactory convergence of the sensitivity measures to two decimal places (+/- 0.025).

The total computation time of the GP-based sensitivity analysis reached a total of 108 hours (4.5 days), including the generation of the training sample (≈4 days), the construction of a GP

554 model at each step of the crisis period (≈3 hours) and the cross-validation procedure (≈3
555 hours).
556 If the same analysis had been undertaken by direct simulations, the total computation time
557 would have reached 9000/30×96=28800 hours (1200 days) using the same 30 CPU cluster. To
558 achieve a computation time of 108 hours, a computer cluster composed of 8000 CPU would
559 have been required.

560 ### *5.2.5* Analysis of the temporal evolution of the main effects

561 Fig. 6 (top) depicts the temporal evolution of the "first most important" input factor (straight
562 green line) at the observation point 1 in the upper part of the landslide (Fig. 6, left) and at the
563 observation point 2 in the lower part of the landslide (Fig. 6, right). Similarly, Fig. 6 (bottom)
564 provides the temporal evolution of the "second most important" input factor. The input factors
565 (Table 2) were ranked in terms of importance based on the mean of the main effect (blue
566 straight line, Fig. 6) computed with the GP models constructed at each instant of the crisis
567 period.
568
569 *[Fig. 6 about here]*
570
571 This preliminary ranking of the input factors, only based on the mean of the main effect, was
572 assessed again in a second step taking into account the range of uncertainty associated to the
573 sensitivity measures *i.e.* using the 5% and to the 95 % quantile of the posterior probability
574 distribution associated to the main effects (black dashed line, Fig. 6). The procedure consisted
575 in qualifying the GP model as "unsure" with respect to the sensitivity measures in regions
576 where the confidence intervals of the first and second most important input factors intersect.
577 Considering the observation point 1, Fig. 6 (left) shows that for the first 150 days, coefficient
578 n_e can be identified as the "first most important" input factor with a mean of the main effect

579 constant at ≈20 %, whereas the dilatancy angle Ψ can be identified as the "second most
580 important" input factor with a mean of the main effect constant at ≈10 %. For the second
581 crisis period, the confidence intervals intersect and the ranking is "unsure". Fig. 7 (left) gives
582 the mean of the main effects and the associated confidence intervals at three different steps of
583 the crisis period, namely 30 days (Fig. 7, top), 150 days (Fig. 7, middle) and 210 days (Fig. 7,
584 bottom). At 30 days, n_e can clearly be identified as the first most important input factor, but
585 the ranking of the other input factors is hardly feasible considering the intersecting confidence
586 intervals. Over time (at 150 and 210 days), the confidence intervals for all input factors
587 intersect so that the ranking is "unsure". This result is in agreement with the low coefficient of
588 determination of the cross-validation procedure over the second half of the crisis period (Fig.
589 5, black dashed line). As a conclusion, the knowledge on the "true" simulator should be
590 increased for the second crisis time period in order to increase the predictive quality of the GP
591 model, hence to narrow the width of the confidence interval.
592
593 *[Fig. 7 about here]*
594
595 Considering the observation point 2, Fig. 6 (right) shows that before ≈50 days, the confidence
596 intervals intersect and the ranking is "unsure". Over the time period after ≈50 days,
597 coefficient n_e can be identified as the "first most important" input factor with a mean of the
598 main effect increasing from ≈20 % to ≈45 %, whereas the plastic compressibility β can be
599 identified as the "second most important" input factor with a mean of the main effect
600 approximately constant and equal to 15 %. As for point 1, Fig. 7 (right) gives the mean of the
601 main effects and the associated confidence intervals for steps 30 days (Fig. 7, top), 150 days
602 (Fig. 7, middle) and 210 days (Fig. 7, bottom). It shows that over time, n_e and β can be
603 identified "with certainty" as the "first and the second most important input factors" for steps

604 150 and 210 days, but the ranking of the other input factors is hardly feasible considering the
605 intersecting confidence intervals.
606 Despite the limited number of simulator runs (30) *i.e.* the limited knowledge on the "true"
607 simulator, several conclusions can still be drawn to guide future investigations. The sensitivity
608 analysis based on the GP modelling emphasizes coefficient n_e as the "most important" *i.e.* as
609 the input factor requiring further investigations over the crisis period, whatever the part of the
610 landslide (upper or lower). In practice, the estimation of this parameter is strongly dependent
611 on the availability of lab tests at small strains, where the behaviour is truly elastic (e.g. strains
612 lower than 10^{-4}). This condition is not realized for classical triaxial tests where the accuracy is
613 not better than 10^{-3} (*e.g.* Biarez and Hicher, 1994) so that this parameter is usually deduced
614 using standard values estimated for analogous types of soil. Nevertheless, such an analogy-
615 based approach is hardly achievable in the La Frasse landslide case as the considered soil
616 material, being on the slip surface, is inherently heterogeneous.
617 The sensitivity analysis also outlines the plastic compressibility β as "important" for further
618 investigations in the lower part of the landslide *i.e.* where the evolution of pore pressures was
619 the most important. In practice, this parameter can be obtained from oedometer tests. No
620 further conclusions can be drawn without increasing the knowledge on the "true" simulator,
621 for the third (or lower) "most important input factor" due to the uncertainty on the GP model.
622 These conclusions are valid for the considered illustrative case especially regarding the
623 assumptions on the range of uncertainty assigned to all input factors (variation in a range of
624 25 % around the original values). Within a procedure aiming at calibrating the observed
625 displacements with the simulated ones, the uncertainty on each input factor should be
626 adequately represented making use of any kind of information related to the measurement
627 procedure of the constitutive model parameters (number of samples, estimation of

measurement error, possibility to construct empirical probability distribution, error of calibration between observed and simulated curves, etc.).

Concluding remarks and further works

Landslide numerical modelling involves a large number of input factors, whose influence and importance should be assessed to guide risk management and possible further investigations (laboratory or *in situ*). A variance-based global sensitivity analysis (Saltelli *et al.*, 2008) using the calculation of Sobol' indices (Sobol', 1993; Archer *et al.*, 1997; Sobol' and Kucherenko, 2005) can provide such guidelines. Nevertheless, such an analysis requires a large number of direct simulations (*i.e.* simulator runs), which can be unfeasible in practice for computationally intensive models (*i.e.* those characterized by computation times ranging from several hours to several days). In this paper, we proposed a methodology based on Gaussian Process meta-modelling to perform such an analysis using a limited number of training samples.

The construction of the training sample is based on a space-filling approach using Latin Hypercube sampling. The possible correlation between input factors is not tackled in this paper and this can be further developed using, for instance, the works of Hamm and co-workers (Hamm *et al.,* 2006). We presented the construction of the meta-model and how to combine it with a strategy to verify the predictive quality based on a cross-validation procedure. This methodology is demonstrated on a numerical model of La Frasse (Switzerland) landslide (Laloui *et al.*, 2004), where the importance of the main constitutive model parameters describing the slip surface material behaviour is assessed. Due to high computational costs, the GP model is constructed only using 30 simulator runs, *i.e.* with a limited knowledge of the "true" simulator. This induces an additional source of uncertainty (referred to as code uncertainty of the meta-model) on the sensitivity measures, which is

652 tackled by treating the GP model from a Bayesian perspective: the full posterior probability
653 distribution associated with the sensitivity measures is computed and summarized by a
654 confidence interval used to outline the regions where the GP model is “unsure” with respect
655 to the sensitivity measures. When a large number of input factors (> 30) are present, the
656 Bayesian treatment of the GP model may show limitations as it is more computationally
657 demanding compared to other meta-model techniques (Storlie *et al.*, 2009). However, recent
658 works (e.g. Marrel *et al.*, 2009) pertaining to variable selection for GP model can be used to
659 overcome this difficulty. In the identified “unsure” regions, further simulator runs should be
660 carried out and the choice of the new input configurations can be guided by taking advantage
661 of the recent advances in adaptive design of experiments (*e.g.* Gramacy and Herbert, 2009),
662 which constitutes a possible future direction.

663 # Acknowledgements

664 This work was funded under the BRGM’s Directorate of Research project VULNERISK. The
665 application is based on the finite-element landslide model built by the LMS of EPFL. We are
666 grateful to B. Gramacy for useful discussion on the hyper-parameters estimations of the GP
667 model. We would like to thank Dr. J. Douglas for proofreading and the two anonymous
668 reviewers for their detailed and constructive reviews.

669 ## *References*

670 Archer, G.E.B., Saltelli, A., Sobol’, I.M., 1997. Sensitivity measures, ANOVA like
671 techniques and the use of bootstrap. Journal of Statistical Computation and Simulation
672 58, 99–120.
673 Aubry, D., Chouvet, D., Modaressi, A., Modaressi, H., 1986. Gefdyn software - Logiciel
674 d'analyse du comportement mécanique des sols par éléments finis avec prise en
675 compte du couplage sol-eau-air (Gefdyn software, Finite element analysis of soil

mechanical behaviour taking into account the soil-water-air coupling). Report of Ecole Centrale Paris (in French).

Aubry, D., Hujeux, J.-C., Lassoudière, F., Meimon, Y., 1982. A double memory model with multiple mechanisms for cyclic soil behaviour. In: Proceedings of the International symposium on numerical models, Balkema, pp. 3–13.

Biarez, J., Hicher, P., 1994. Elementary mechanics of. soil behaviour. Balkema, Rotterdam, The Netherlands, 208 pp.

Bonnard, C., Noverraz, F., Lateltin, O., Raetzo, H., 1995. Large Landslides and Possibilities of Sudden Reactivation. In: Proceedings 44th Geomechanics Colloquy, Salzburg., Austria, pp. 401-407.

Boulahya, F., Dubus, .I.G., Dupros, F., Lombard, P., 2007. Footprint@work, a computing framework for large scale parametric simulations: application to pesticide risk assessment and management, In: Forum EGEE Enabling Grids for E-sciencE, Manchester, UK, pp. 160.

Cressie, N.A.C., 1993. Statistics for Spatial Data, Wiley, New York, 900 pp.

Crosta, G.B., Clague, J.J., 2009. Dating, triggering, modelling, and hazard assessment of large landslides. Geomorphology 103, 1-4.

Dupros, F., Boulahya, F., Vairon, J., Lombard, P., Capit, N., Méhaut, J.F., 2006. IGGI, a computing framework for large scale parametric simulations: application to uncertainty analysis with toughreact. In: Tough symposium 2006, Berkeley, 6 pp.

Evans, S.G., Scarascia Mugnozza, G., Strom, A., Hermanns, R.L., 2002. Landslides from massive rock slope failure. In: Proceedings NATO Advanced Research Workshop on Massive Rock Slope Failure, Celano, Italy, 662 pp.

van den Ham, G., Rohn, J., Meier, T., Czurda, K., 2009. Finite Element simulation of a slow moving natural slope in the Upper-Austrian Alps using a visco-hypoplastic constitutive model. Geomorphology 103, 136-142.

Hujeux, J.-C., 1985. Une loi de comportement pour le chargement cyclique des sols. In: Davidovici V. (Ed.) Génie Parasismique, Presses ENPC, France, p. 278–302.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. 2004. Bayesian Data Analysis, 2nd edn., Chapman and Hall/CRC, London, UK, 698 pp.

Gilks, W., Richardson, S., Spiegelhalter, D. 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC, London, UK, 512 pp.

Gramacy, R. B., 2007. tgp: An R Package for Bayesian Nonstationary, Semiparametric Nonlinear Regression and Design by Treed Gaussian Process Models. Journal of Statistical Software 19.

Gramacy, R.B., Herbert K.H.L., 2009. Adaptive Design and Analysis of Supercomputer Experiments. Technometrics 51(2), 130-145.

Gramacy, R.B., Taddy, M. 2010. Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with tgp Version 2, an R Package for Treed Gaussian Process Models. Journal of Statistical Software, 33, 1–48.

Gzyl, H., 1995. The Method of Maximum Entropy. In: Bellomo, F., Brezzi, N. (Eds.),Series on Advances in Mathematics for Applied Sciences. World Scientific Publishing Co, p. 29.

Hamm, N.A.S., Hall, J.W., Anderson, M.G., 2006. Variance-based sensitivity analysis of the probability of hydrologically induced slope instability. Computers and Geosciences 32, 803-817.

Hansen, A., 1984. Landslide Hazard Analysis. In: Brunsden, D., Prior, D.B. (Eds.), Slope Instability. Wiley and Sons: New York, pp. 523-602.

724 Hastie, T., Tibshirani, R., Friedman, J., 2002. The Elements of Statistical Learning. Springer :
725 Data Mining, Inference, and Prediction, Springer-Verlag, New York, 552pp.
726 Kennedy, M., O'Hagan, A., 2001. Bayesian calibration of computer models (with discussion).
727 Journal of the Royal Statistical Society 63 (Series B), 425-464.
728 Koehler, J.R., Owen, A.B., 1996. Computer experiment, In: Ghosh S. and Rao C.R. (Eds.),
729 Handbook of Statistics. Elsevier Science, New York, USA, 13, p. 261-308.
730 Laloui, L., Tacher, L., Moreni, M., Bonnard, C., 2004. Hydromechanical modeling of crises
731 of large landslides: application to the La Frasse Landslide, In: Proceedings of the 9th
732 International Symposium on Landslides, Rio de Janeiro, Brazil, pp. 1103-1110.
733 Langewisch, D.R., Apostolakis, G.E., 2010. A comparison of polynomial response surfaces
734 and Gaussian rocesses as metamodels for uncertainty analysis with long-running
735 computer codes. In: Proceedings of the 10th International Probabilistic Safety
736 Assessment & Management Conference, Seattle, Washington USA, 12 pp.
737 Le, N.D., Zidek, J.V., 2006. Statistical Analysis of Environmental Space Time Processes.
738 Springer Series in Statistics, New York, 341 pp.
739 Lopez-Caballero, F., Modaressi Farahmand-Razavi, A., Modaressi, H., 2007. Nonlinear
740 numerical method for earthquake site response analysis —elastoplastic cyclic model
741 and parameter identification strategy. Bulletin of Earthquake Engineering 5(3), 303-
742 323.
743 Lopez-Caballero, F., Modaressi Farahmand-Razavi, A., 2008. Numerical simulation of
744 liquefaction effects on seismic SSI, Soil Dynamics and Earthquake Engineering 28(2)
745 85-98.
746 Lophaven, S.N., Nielsen, H.B., Sondergaard J., 2002. DACE—A Matlab kriging toolbox,
747 version 2.0. Technical Report IMM-TR-2002-12, Informatics and Mathematical

748 Modelling, Technical University of Denmark, 28 pp.
749 http://www.immm.dtu.dk/_hbn/dace.
750 Marrel, A., Iooss, B., Van Dorpe, F., Volkova, E., 2008. An efficient methodology for
751 modeling complex computer codes with Gaussian processes. Computational Statistics
752 and Data Analysis 52, 4731-4744.
753 Marrel, A., Iooss, B., Laurent B., Roustant O., 2009. Calculations of Sobol indices for the
754 Gaussian process metamodel, Reliability Engineering and System Safety 94, 742-751.
755 Martin, J.D., Simpson, T.W., 2005. On the use of kriging models to approximate deterministic
756 computer models. Journal of American Institute of Aeronautics and Astronautics 43,
757 853–863.
758 McKay, M. D., Beckman R. J. and Conover W. J., 1979. A comparison of three methods for
759 selecting values of input variables in the analysis of output from a computer code.
760 Technometrics 21, 239–245.
761 Noverraz, F., Bonnard, Ch., 1990. Technical note on the visit of La Frasse Landslide. In:
762 Proceeding 5th International Symposium on Landslides 3, pp. 1549-1554.
763 Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a
764 Bayesian approach. Journal of Royal Statistical Society 66 (Series B), 751–769.
765 O'Hagan, A., 2006. Bayesian analysis of computer code outputs: A tutorial. Reliability
766 Engineering and System Safety 91, 1290–1300.
767 R Development Core Team, 2009. R: A Language and Environment for Statistical
768 Computing. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-
769 900051-07-0, http://www.R-project.org/
770 Rasmussen, C. E., 1996. Evaluation of Gaussian Processes and other Methods for Non-linear
771 Regression. PhD thesis, Department of Computer Science, University of Toronto, 132
772 pp.

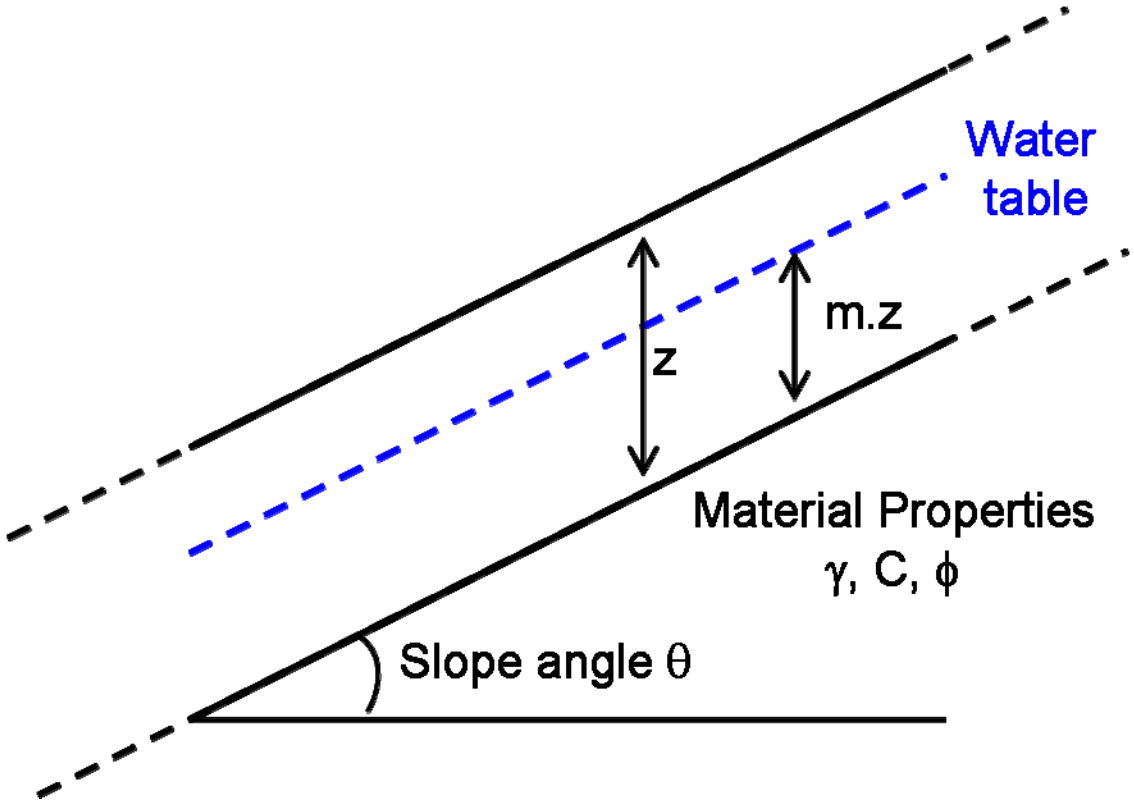
773 Rasmussen, C.E., Williams, C.K.I, 2006. Gaussian processes for machine learning, MIT Press,
774 Cambridge, UK, 266 pp.
775 Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and analysis of computer
776 experiments. Statistical Science 4, 409–435.
777 Saltelli A., Sobol' IM., 1995. About the use of rank transformation in sensitivity of model
778 output. Reliability Engineering and System Safety 50, 225–239.
779 Saltelli A., Chan K., Scott E. M. (Eds.) 2000. Sensitivity Analysis. In: Probability and
780 Statistics Series. Chichester: John Wiley and Sons, 475 p.
781 Saltelli, A., 2002a. Making best use of model evaluations to compute sensitivity indices.
782 Computer Physics Communications 145, 280-297.
783 Saltelli, A., 2002b. Sensitivity Analysis for Importance assessment. Risk analysis 22(3), 579-
784 590.
785 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M.,
786 Tarantola, S., 2008. Global sensitivity analysis: The Primer. Wiley, Chichester, UK,
787 304 pp.
788 Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance
789 based sensitivity analysis of model output. Design and estimator for the total
790 sensitivity index. Computer Physics Communications 181, 259–270.
791 Santner, T., Williams, B., Notz, W., 2003. The Design and Analysis of Computer
792 Experiments, Springer Verlag, New York, 308 pp.
793 Sobol', I.M., 1993. Sensitivity estimates for non linear mathematical models. Mathematical
794 Modelling and. Computational Experiments 1, 407–414.
795 Sobol', I.M. Kucherenko, S.S., 2005. Global sensitivity indices for nonlinear mathematical
796 models, Review of Wilmott Magazine 1, 56–61.

Stein, M.L., 1999. Interpolation of Spatial Data. Springer: Some Theory for Kriging, New York, 247 pp.

Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C. J., 2009. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliability Engineering and System Safety 94, 1735–1763.

Von Mises, R., 1964. Mathematical Theory of Probability and Statistics. Mathematical Theory of Probability and Statistics, Academic Press, New York, 694 pp.

805 **FIGURES CAPTIONS**



806

807 Fig. 1: Schematic representation of the infinite slope model (adapted from Hansen, 1984).

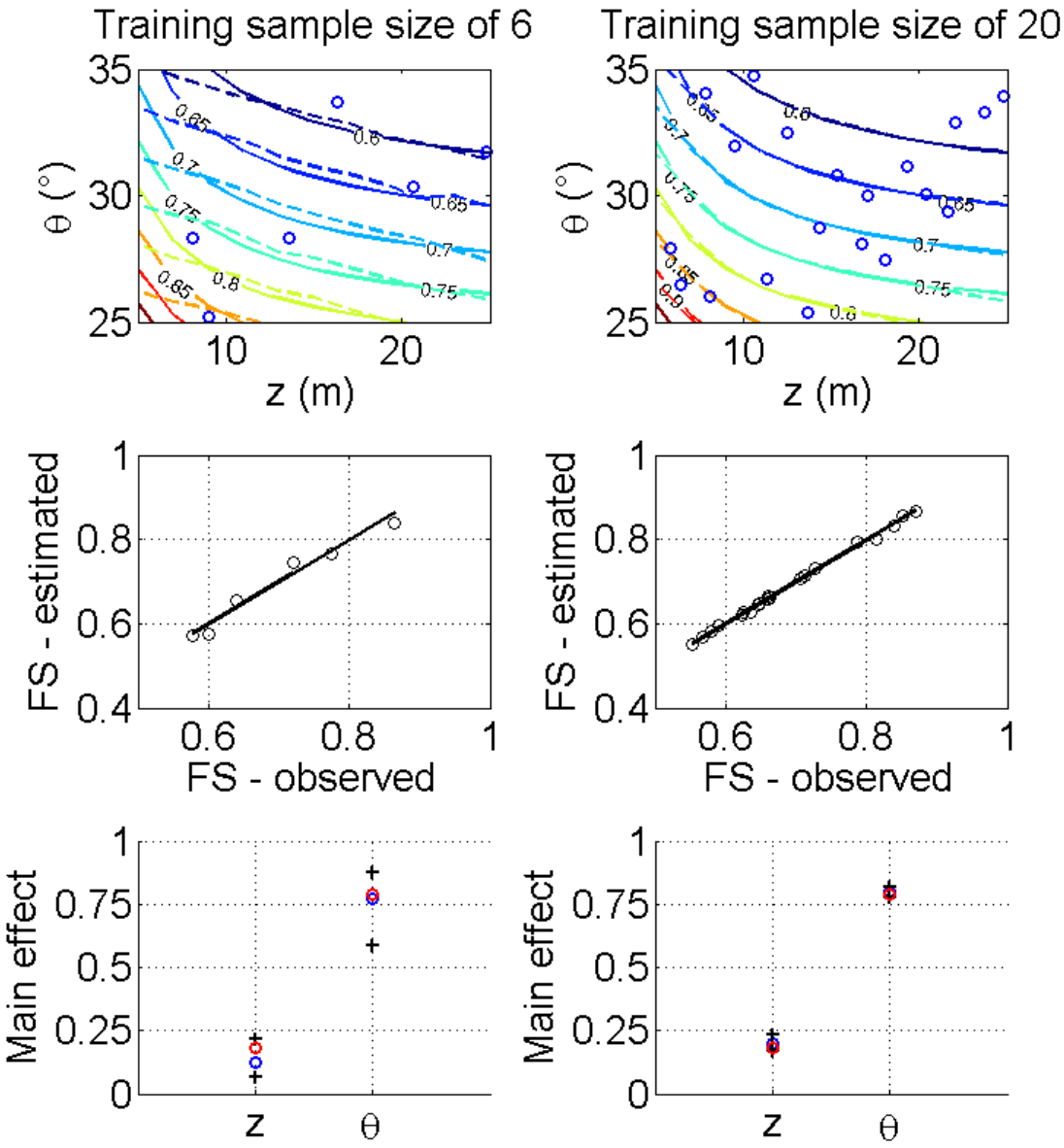
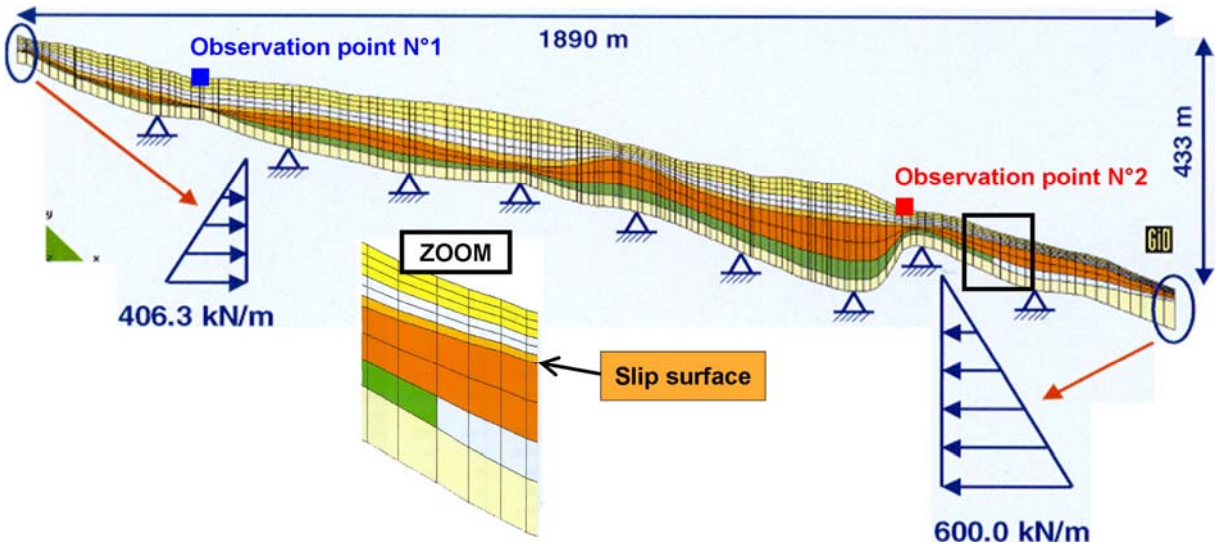
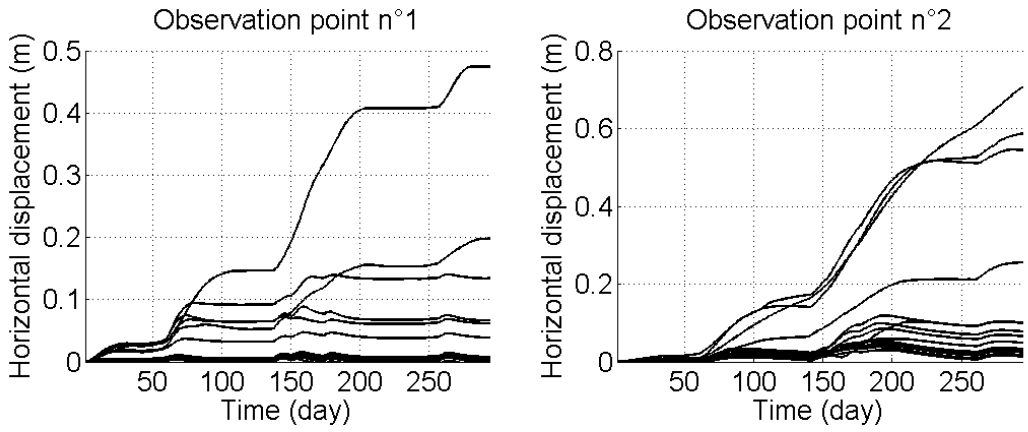


Fig. 2: Top: comparison between the true values and the estimates of the factors of safety *FS* using the GP model constructed with a training sample size of 6 (left) and of 20 (right). The training input configurations are represented by blue dots. Middle: comparison between the observed *FS* and the estimates within a "leave-one-out" cross validation procedure. Bottom: comparison between the true values (red dots) and the estimates of the main effects for the slope thickness z and of the slope angle θ (blue dots). The bounds of the confidence intervals associated with both GP models are represented by black cross-type markers.



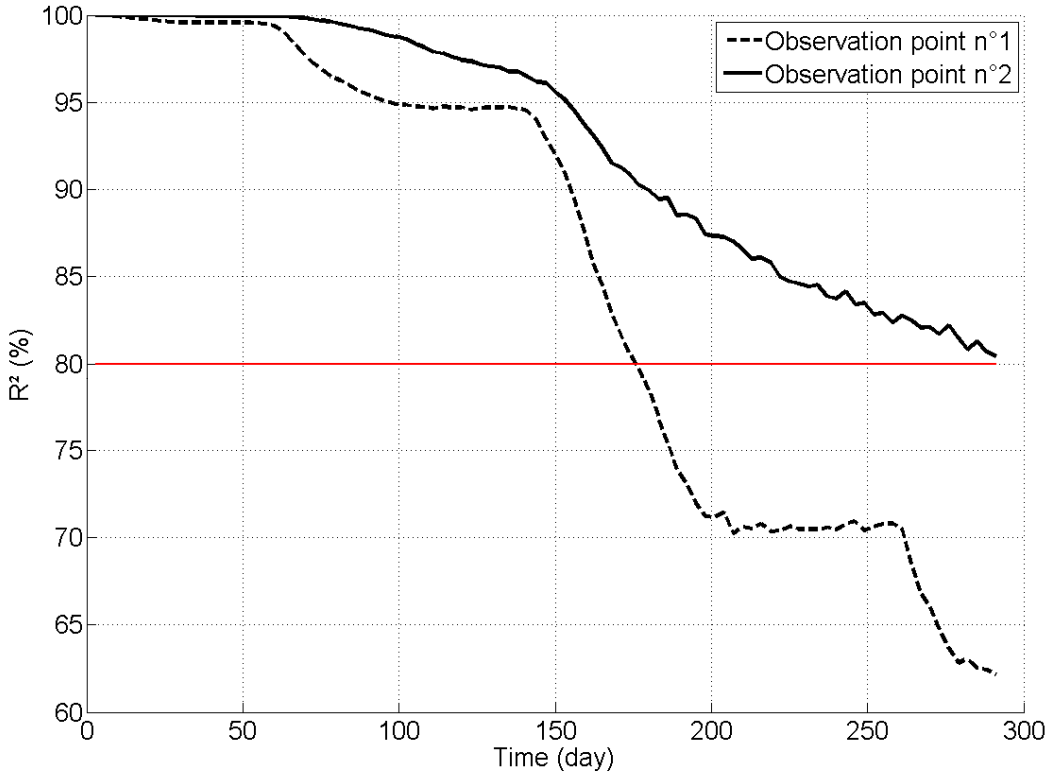
816

817 Fig. 3: Overview of the landslide numerical model (adapted from Laloui *et al.*, 2004). The
818 slip surface is outlined by the light coloured orange surface. The observation point 1 (in the
819 upper part of the landslide) and observation point 2 (in the lower part of the landslide) used
820 for the sensitivity analysis of the horizontal displacements are respectively outlined by a blue
821 and a red coloured square-type marker.



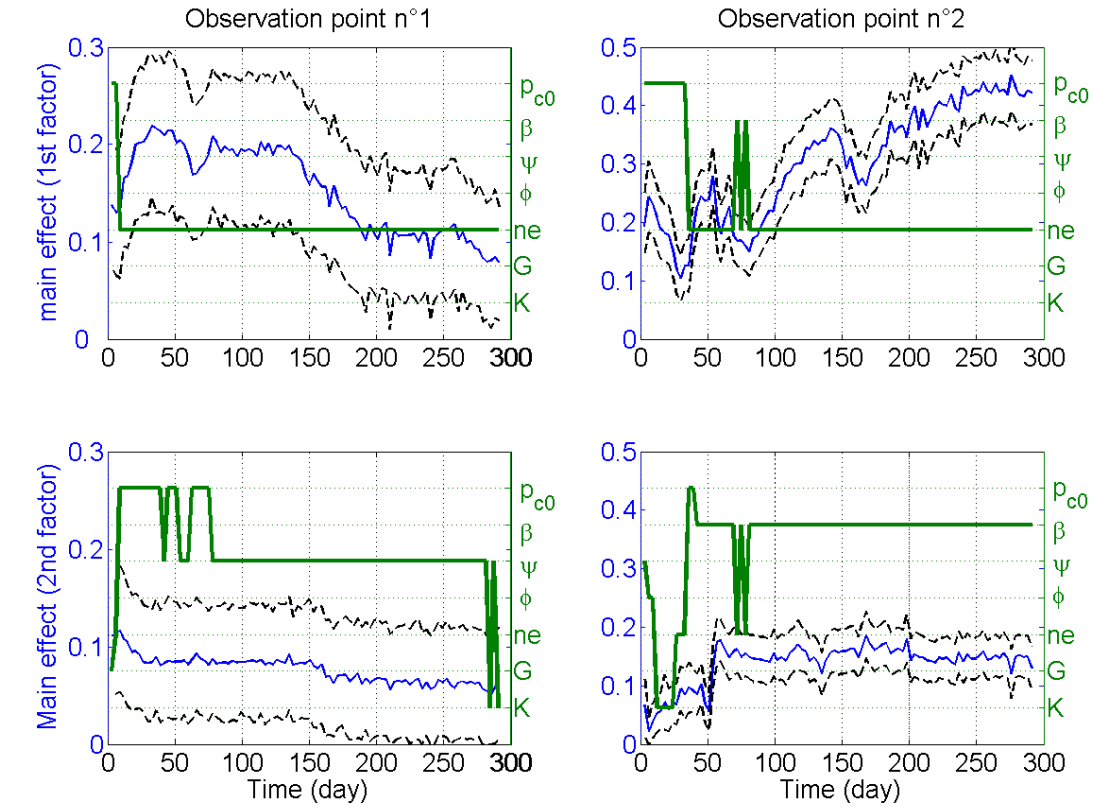
822

823 Fig. 4: Temporal evolution of the training samples corresponding to the horizontal
824 displacements (m) calculated for 30 different input configurations of the Hujeux law
825 parameters (at the observation point 1 in the upper part of the landslide (left) and at the
826 observation point 2 in the lower part of the landslide (right)).

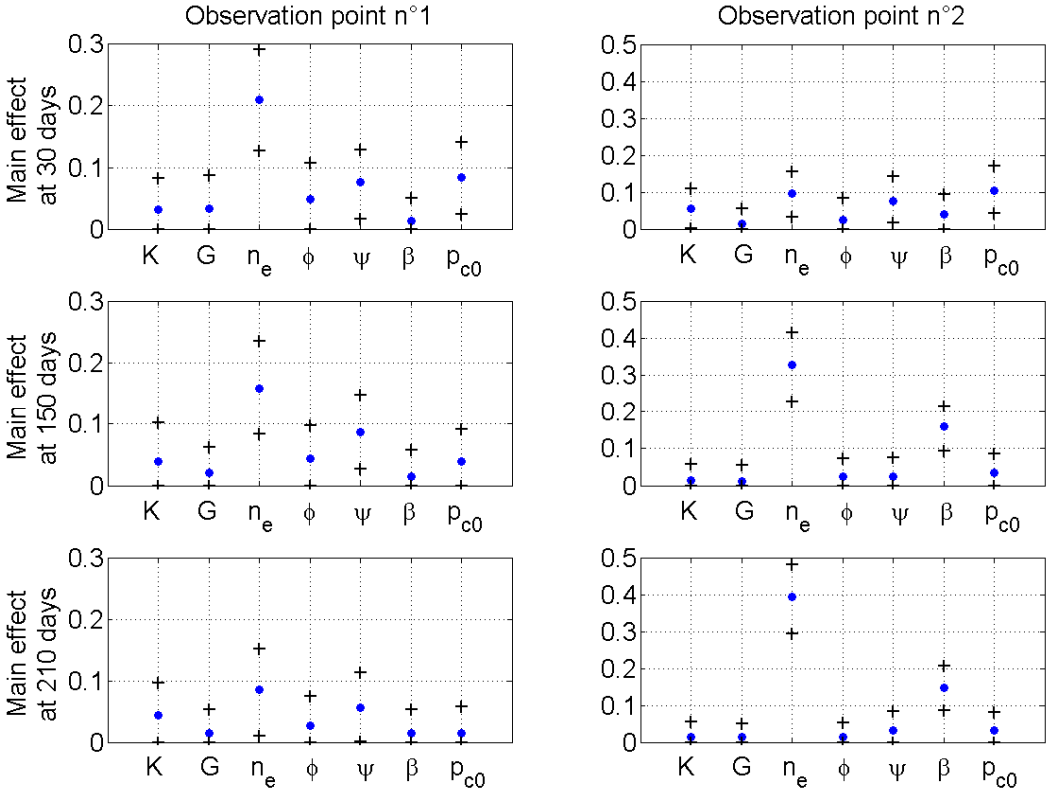


827

828 Fig. 5: Temporal evolution of the coefficient of determination R^2 for the “leave-out-out” cross
829 validation procedure of the GP models constructed at each instant of the crisis period at the
830 observation point 1 in the upper part of the landslide (black dashed line) and at the
831 observation point 2 in the lower part of the landslide (black straight line). The threshold of
832 80 % indicating a “satisfactory” predictive quality is outlined by a horizontal red straight line.



833



834

835 Fig. 6: Temporal evolution during the crisis period of the mean of the main effects (blue

836 straight line) at the observation point 1 in the upper part of the landslide (left) and at the

observation point 2 in the lower part of the landslide (right) for the first (green straight line, top) and the second (green straight line, bottom) “most important” input factor. The black dashed lines represent the 5% and the 95 % quantile.

Fig.7: Mean of the main effect (blue dots) for each input factor of the slip surface constitutive law at different instants of the crisis period (30 days (top), 150 days (middle) and 210 days (bottom)) at the observation point 1 in the upper part of the landslide (left) and at the observation point 2 in the lower part of the landslide (right). The bounds of the confidence intervals (5% and 95 % quantile) are outlined by black cross-type markers.

847 **TABLES**

848

849 Table 1: Comparison between the “true” and the estimates of the main effects for the infinite
850 slope analytical model. μ corresponds to the mean of the main effect computed with the GP
851 model. *CI* corresponds to the confidence interval defined by the 5 % and the 95 % quantile
852 computed with the GP model.

Input factor	True model	GP model constructed with 6 training samples	GP model constructed with 20 training samples
Thickness z (m)	18.41 %	μ=12.74 % *CI*=[6.48 ; 21.73] %	μ=20.02 % *CI*=[16.90 ; 23.29] %
Slope angle θ(°)	78.76 %	μ=77.01 % *CI*=[58.41 ; 87.43] %	μ=79.77 % *CI*=[77.90 ; 81.68] %

853

854 Table 2: Range of values for the slip surface properties of the La Frasse landslide (variation in
855 a range of 25 % around the original values given in Laloui *et al.*, 2004)

Input factor	Vol. comp. mod.	Shear mod.	Non-linearity coeff.	Internal friction angle	Dilatancy angle	Plastic comp.	Initial critical pressure
Symbol	K	G	n_e	ϕ	Ψ	β	p_{c0}
Unit	MPa	MPa	-	°	°	-	MPa
Lower value	180	83.25	0.225	19.125	14.25	20.625	0.375
Upper value	300	138.75	0.375	31.875	23.75	34.375	0.625

856