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

Jeremy Rohmer¹*, Evelyne Foerster¹

¹BRGM, Natural Risks and Safety of CO₂ 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 $\{x_i\}_{i=1,...,n}$ so that $y = g(x)$. 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):

$$V_Y = \sum_i V_i + \sum_{i<j} V_{ij} + \sum_{i<j<k} V_{ijk} + \ldots + V_{ij...n}$$ (1)

where the partial variance $V_i$ and $V_{ij...n}$ read:
\[ V_i = \text{Var}[E[Y|X_i = x_i]] \]
\[ V_{g_{i\ldots n}} = \text{Var}[E[Y|X_i = x_i, X_j = x_j, \ldots, X_n = x_n]] - V_i - V_j - \ldots - V_n \]  \hspace{1cm} (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 \( x_j \), \ldots 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:

\[ g(x) = g_0 + \sum_{i=1}^{n} g_i(x_i) + \sum_{i<j}^{n} g_{ij}(x_i, x_j) + \ldots + g_{1\ldots n}(x_1, \ldots, x_n) \]  \hspace{1cm} (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:

\[ g_0 = E[Y] = \int_0^1 \int_0^1 \ldots \int_0^1 g(x_1, \ldots, x_n) dx_1 \ldots dx_n \]  \hspace{1cm} (4)

\[ g_i(x_i) = E[Y|X_i = x_i] - g_0 = \int_0^1 \int_0^1 \ldots \int_0^1 g(x) dx_{-i} - g_0 \]  \hspace{1cm} (5)
The total variance $V_y$ can then be expressed as:

$$V_y = \int_{0}^{1} \cdots \int_{0}^{1} g(x_1, \ldots, x_n)^2 dx_1 \cdots dx_n - g_0^2$$  \hspace{1cm} (7)$$

while the partial variances read as follows:

$$V_{i_{1} \ldots i_{s}} = V[g_{i_{1} \ldots i_{s}}] = \int_{0}^{1} \cdots \int_{0}^{1} g_{i_{1} \ldots i_{s}}(x_{i_{1}}, \ldots, x_{i_{s}})^2 dx_{i_{1}} \cdots dx_{i_{s}}$$  \hspace{1cm} (8)$$

with $1 \leq i_1 < \ldots < i_s \leq n$ 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:

$$S_i = \frac{V_i}{V_y}$$  \hspace{1cm} (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:

\[
\sum_{i=1}^{n} S_i + \sum_{i<j} S_{ij} + \ldots + S_{ij...n} = 1
\]

Defining \( S_{-i} \) as the sum of all the terms \( S_{i_{-i}} \) 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

\[
S_{Ti} = 1 - S_{-i}
\]

\( S_{Ti} = 0 \) 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 $\{x^{(j)}\}_{j=1}^{m}$ in the $n$-dimensional space of input factors:

$$
\begin{align}
\hat{g}_0 & \approx \frac{1}{m} \sum_{j=1}^{m} g(x^{(j)}) \\
\hat{V}_Y & \approx \frac{1}{m} \sum_{j=1}^{m} g^2(x^{(j)}) \hat{g}_0^2 \\
\hat{V}_i & \approx \frac{1}{m} \sum_{j=1}^{m} g(x^{-\{j\}, S_a, S_b}^{(j)}) g(x^{-\{j\}, S_a, S_b}^{(j)}) \hat{g}_0^2
\end{align}
$$

(12)

where $x_{-ij} = [x_{ij}, ..., x_{i-1j}, x_{i+1j}, ..., x_{nj}]$ 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 $Y(x)$, 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:
\[
Y(x) = f(x) + Z(x) \tag{13}
\]

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:

\[
f(x) = F(x) \cdot B = \sum_{i=0}^{n} b_i \cdot f_i(x) \tag{14}
\]

where \( B = \{b_i\}_{i=0}^{n} \) is the regression parameter vector; \( F(x) = [f_0(x), f_1(x), \ldots, f_n(x)] \) is the corresponding regression vector with \( f_i \) (\( i = 0, 1, \ldots, n \)), the basis functions. Assuming linearity for the mean \( f \), we have: \( f_0(x) = 1 \) and \( f_i(x) = x_i \) 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 \( \Sigma_S \) so that an element at the \( j \)th row and \( k \)th column of \( \Sigma_S \) is expressed as:

\[
\Sigma_{S,j,k} = \text{Cov}[Z(x^{(j)}), Z(x^{(k)})] = \sigma^2 \cdot \rho(x^{(j)}, x^{(k)}) \tag{15}
\]

where \( x^{(j)} \) is the \( j \)th input vector (with \( j = 1 \ldots N_S \)) and \( \text{Var}[Z] = \sigma^2 \). The correlation function \( \rho \) 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 $x^{(j)}$ and $x^{(k)}$. 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:

$$\rho(x^{(j)}, x^{(k)}) = \prod_{i=1}^{n} \rho_{i}(x_{i}^{(j)}, x_{i}^{(k)}) = \exp \left( - \sum_{i=1}^{n} \frac{\|x_{i}^{(j)} - x_{i}^{(k)}\|^{2}}{\theta_{i}} \right)$$  \hspace{1cm} (16)

where $\Theta = \{\theta_{i}\}_{i=1,...,n}$ are the correlation lengths, also referred to as “hyper-parameters” (e.g. see
Rasmussen and Williams, 2006). $\theta_{i}$ 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 \{x_{S}, y_{S}\}, which represent a mapping between the spaces of input factors $x_{S} = \{x^{(j)}\}_{j=1...N_{S}}$, with $x^{(j)} = \{x_i^{(j)}\}_{i=1...n}$, and the outputs $y_{S} = \{y^{(j)}\}_{j=1...N_{S}}$, obtained through the $N_S$ selected simulator runs so that

$$y^{(j)} = f(x^{(j)})_{j=1...N_{S}}.$$ 

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 $X$. This assumption represents our prior belief on the simulator. Using the training sample $\{x_{S}, y_{S}\}$, 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 $p(y_{S}|x_{S})$ of obtaining $y_{S}$ given $x_{S}$, is expressed under the GP assumption, as follows:

$$p(y_{S}|x_{S}) \sim G(F(x_{S})B, \Sigma_{S})$$ \hspace{1cm} (17)

Considering a new vector $x^{(N_{S}+1)}$ of input factors and the associated output $y^{(N_{S}+1)} = y(x^{(N_{S}+1)})$, the joint probability distribution of the random variables $\{y_{S}, y^{(N_{S}+1)}\}$ reads as follows:

$$p(y_{S}, y^{(N_{S}+1)}|x_{S}, B, \sigma, \theta) \sim G\left(\begin{bmatrix} F_S \cr F(x^{(N_{S}+1)}) \end{bmatrix}, \begin{bmatrix} \Sigma_S & k(x^{(N_{S}+1)}) \\ k(x^{(N_{S}+1)})^T & \sigma^2 \end{bmatrix} \right)$$ \hspace{1cm} (18a)

where $F_S = [F(x^{(1)}), F(x^{(2)}), ..., F(x^{(N_{S})})]$ corresponds to the regression matrix.

$k(x^{(N_{S}+1)})$ represents the vector of correlation functions between each of the $N_S$ training input vectors $x^{(j)}$ and the new element $x^{(N_{S}+1)}$. It can be written as:
Within the Bayesian framework, the posterior distribution of the computed output random variable \( \mathbf{Y}^{(N+1)} \) 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 \( \mathbf{x}^{(N+1)} \), and follows a multivariate Gaussian distribution (Von Mises, 1964):

\[
p \left( \mathbf{Y}^{(N+1)} \mid \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right) = \frac{p \left( \mathbf{Y}^{(N+1)} \mid \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right)}{p \left( \mathbf{Y}_S \mid \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right)} \sim G \left[ \mathbf{E} \left[ \mathbf{Y}^{(N+1)} \mid \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right] \right] Var \left[ \mathbf{Y}^{(N+1)} \mid \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right]
\]

where:

\[
\mathbf{E} \left[ \mathbf{Y}^{(N+1)} \mid \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right] = \mathbf{F} \left( \mathbf{X}^{(N+1)} \right) \mathbf{B} + \mathbf{k} \left( \mathbf{x}^{(N+1)} \right) \mathbf{\Sigma}_S^{-1} \left( \mathbf{Y}_S - \mathbf{F}_S \mathbf{B} \right)
\]

and:

\[
Var \left[ \mathbf{Y}^{(N+1)} \mid \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \mathbf{0} \right] = \sigma^2 - \mathbf{k} \left( \mathbf{x}^{(N+1)} \right) \mathbf{\Sigma}_S^{-1} \mathbf{k} \left( \mathbf{x}^{(N+1)} \right)
\]

with \( \mathbf{\Sigma}_S^{-1} \), 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 \( \mathbf{x}^{(N+1)} \) is far away from the training input data \( \mathbf{X}_S \), the term

\[
\mathbf{k} \left( \mathbf{x}^{(N+1)} \right) \mathbf{\Sigma}_S^{-1} \mathbf{k} \left( \mathbf{x}^{(N+1)} \right)
\]

will be small so that the predicted variance will be large.

In a more general manner, if we consider two new test candidates \( \mathbf{u} \) and \( \mathbf{v} \), the general expression of the conditional GP model can be written as:
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:

\[
\text{Cov}\left[\hat{y}(u), \hat{y}(v) \mid Y_S, X_S, B, \sigma, \theta\right] = \sigma^2 \rho(u, v) - k(u)^T \cdot \Sigma^{-1} \cdot k(v)
\]

(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 \( \theta \) and the variance \( \sigma^2 \). A first approach consists in estimating them as solutions of the optimization problem using the maximum likelihood method, \textit{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 (\textit{e.g.} see Marrel \textit{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 (\textit{e.g.} Rasmussen 1996) is chosen so that the hyper-parameters are given prior distributions \( p(\theta) \). In the Bayesian framework, the Markov Chain Monte Carlo methods (Gilks \textit{et al.}, 1996) are used to integrate over the posterior distribution \( p(\theta \mid X_S, Y_S) \) associated with the GP parameters \textit{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...
Further theoretical details can be found in Gelman et al. (1995) and Gramacy and Herbert (2009). Though computationally more intensive (Storlie et al., 2009), this approach presents the attractive feature to incorporate the uncertainty related to the construction of the GP model, so that the level of confidence associated with the “meta-model”-computed sensitivity measures also takes this source of uncertainty into account (see step 4 of section 3.4.).

3.4 A “GP-based” methodology for sensitivity analysis

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

3.4.1 Step 1: representation of the input factor uncertainty

The first step is to characterize and mathematically represent the uncertainty (range and form of the probability distribution) on each of the input factors. This representation can have a strong influence on the final sensitivity results, hence on risk management decision making (Saltelli, 2002b).

Representing the uncertainty through empirical probability distributions requires a large amount of data (laboratory or in situ measurements), which may not be practical in many situations. Thus, knowledge on the range of uncertainty is commonly evaluated either based on physical reasoning, on analogies with similar cases or simply from expert opinions, whereas the mathematical representation of the probability distribution may either be theoretically known or assumed. In a situation where “sparse, vague and incomplete” data are available, a common approach consists in assigning a uniform probability distribution based 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...
partitioned into $k$ subsets. In a first step, a single subset is used as the validation data for
testing the GP model, and the remaining $k-1$ subsets are used as training data for the
construction of the GP model. For each step, the $k$ validation data are estimated and the
coefficient of determination $R^2$ for the procedure is computed as follows:

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

(21)

where $y$ represents the vector of observations in the validation set; $\bar{y}$ is the mean of the
corresponding sample and $\hat{y}$, the vector of predicted values using the GP model.

The coefficient $R^2$ provides a metric of the predictive quality so that a value close to 100 %
indicates that the GP model is successful in matching the validation data. A typical threshold
of 80 % is commonly used to qualify the predictive quality as “satisfactory” (e.g. Marrel et
al., 2008).

The cross-validation process is then repeated $k$ times using each of the $k$ subsets as validation
samples. For small training sets, the cross validation procedure with $k=1$ is usually used
corresponding to the so-called “leave-one-out” cross validation procedure.

**3.4.4 Step 4: estimating the sensitivity measures**

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

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

[Fig. 1 about here]

where $C$ is the cohesion of the soil material; $\phi$, the friction angle; $\theta$, the slope angle; $\gamma$, the soil unit weight; $\gamma_w$, 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 $\theta$ as uncertain input factors. The other input factors were assumed fixed: $C=10$ kPa, $\phi=25^\circ$, $\gamma=22$ kN.m$^{-3}$, $\gamma_w=9.81$ kN.m$^{-3}$ and $m=90\%$.

The objective was to identify whether $z$ or $\theta$ 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 $\theta$ uniformly varies between $25^\circ$ and $35^\circ$ (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 (top) shows the comparison between the values of $FS$ obtained from direct simulations on a grid of $10 \times 10$ in the input factor domain $[5 ; 25] \times [25 ; 35]$ (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.

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 $1000 \times 1000$ m$^2$, 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 $10 \times 10^{-2}$ and $15 \times 10^{-2}$ m per year (Novraz 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).

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 $\phi$ at perfect plasticity, the plastic compressibility $\beta$; and (3) the dilatancy angle $\Psi$, 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.).

\[\text{Table 2 about here}\]

### 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 \(\approx 96\) 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.

\[\text{Fig. 4 about here}\]
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.

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 model.
model at each step of the crisis period (≈3 hours) and the cross-validation procedure (≈3 hours).

If the same analysis had been undertaken by direct simulations, the total computation time would have reached \(9000/30 \times 96 = 28800\) hours (1200 days) using the same 30 CPU cluster. To achieve a computation time of 108 hours, a computer cluster composed of 8000 CPU would have been required.

5.2.5 Analysis of the temporal evolution of the main effects

Fig. 6 (top) depicts the temporal evolution of the “first most important” input factor (straight green line) at the observation point 1 in the upper part of the landslide (Fig. 6, left) and at the observation point 2 in the lower part of the landslide (Fig. 6, right). Similarly, Fig. 6 (bottom) provides the temporal evolution of the “second most important” input factor. The input factors (Table 2) were ranked in terms of importance based on the mean of the main effect (blue straight line, Fig. 6) computed with the GP models constructed at each instant of the crisis period.

This preliminary ranking of the input factors, only based on the mean of the main effect, was assessed again in a second step taking into account the range of uncertainty associated to the sensitivity measures i.e. using the 5% and to the 95 % quantile of the posterior probability distribution associated to the main effects (black dashed line, Fig. 6). The procedure consisted in qualifying the GP model as “unsure” with respect to the sensitivity measures in regions where the confidence intervals of the first and second most important input factors intersect. Considering the observation point 1, Fig. 6 (left) shows that for the first 150 days, coefficient \(n_e\) can be identified as the “first most important” input factor with a mean of the main effect...
constant at \( \approx 20\% \), whereas the dilatancy angle \( \Psi \) can be identified as the “second most important” input factor with a mean of the main effect constant at \( \approx 10\% \). For the second crisis period, the confidence intervals intersect and the ranking is “unsure”. Fig. 7 (left) gives the mean of the main effects and the associated confidence intervals at three different steps of the crisis period, namely 30 days (Fig. 7, top), 150 days (Fig. 7, middle) and 210 days (Fig. 7, bottom). At 30 days, \( n_e \) can clearly be identified as the first most important input factor, but the ranking of the other input factors is hardly feasible considering the intersecting confidence intervals. Over time (at 150 and 210 days), the confidence intervals for all input factors intersect so that the ranking is “unsure”. This result is in agreement with the low coefficient of determination of the cross-validation procedure over the second half of the crisis period (Fig. 5, black dashed line). As a conclusion, the knowledge on the “true” simulator should be increased for the second crisis time period in order to increase the predictive quality of the GP model, hence to narrow the width of the confidence interval.

[Fig. 7 about here]

Considering the observation point 2, Fig. 6 (right) shows that before \( \approx 50 \) days, the confidence intervals intersect and the ranking is “unsure”. Over the time period after \( \approx 50 \) days, coefficient \( n_e \) can be identified as the “first most important” input factor with a mean of the main effect increasing from \( \approx 20\% \) to \( \approx 45\% \), whereas the plastic compressibility \( \beta \) can be identified as the “second most important” input factor with a mean of the main effect approximately constant and equal to 15\%. As for point 1, Fig. 7 (right) gives the mean of the main effects and the associated confidence intervals for steps 30 days (Fig. 7, top), 150 days (Fig. 7, middle) and 210 days (Fig. 7, bottom). It shows that over time, \( n_e \) and \( \beta \) can be identified “with certainty” as the “first and the second most important input factors” for steps
150 and 210 days, but the ranking of the other input factors is hardly feasible considering the intersecting confidence intervals.

Despite the limited number of simulator runs (30) \textit{i.e.} the limited knowledge on the “true” simulator, several conclusions can still be drawn to guide future investigations. The sensitivity analysis based on the GP modelling emphasizes coefficient \( n_e \) as the “most important” \textit{i.e.} as the input factor requiring further investigations over the crisis period, whatever the part of the landslide (upper or lower). In practice, the estimation of this parameter is strongly dependent on the availability of lab tests at small strains, where the behaviour is truly elastic (e.g. strains lower than \( 10^{-4} \)). This condition is not realized for classical triaxial tests where the accuracy is not better than \( 10^{-3} \) (\textit{e.g.} Biarez and Hicher, 1994) so that this parameter is usually deduced using standard values estimated for analogous types of soil. Nevertheless, such an analogy-based approach is hardly achievable in the La Frasse landslide case as the considered soil material, being on the slip surface, is inherently heterogeneous.

The sensitivity analysis also outlines the plastic compressibility \( \beta \) as “important” for further investigations in the lower part of the landslide \textit{i.e.} where the evolution of pore pressures was the most important. In practice, this parameter can be obtained from oedometer tests. No further conclusions can be drawn without increasing the knowledge on the “true” simulator, for the third (or lower) “most important input factor” due to the uncertainty on the GP model. These conclusions are valid for the considered illustrative case especially regarding the assumptions on the range of uncertainty assigned to all input factors (variation in a range of 25 \% around the original values). Within a procedure aiming at calibrating the observed displacements with the simulated ones, the uncertainty on each input factor should be adequately represented making use of any kind of information related to the measurement 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
tackled by treating the GP model from a Bayesian perspective: the full posterior probability distribution associated with the sensitivity measures is computed and summarized by a confidence interval used to outline the regions where the GP model is “unsure” with respect to the sensitivity measures. When a large number of input factors (> 30) are present, the Bayesian treatment of the GP model may show limitations as it is more computationally demanding compared to other meta-model techniques (Storlie et al., 2009). However, recent works (e.g. Marrel et al., 2009) pertaining to variable selection for GP model can be used to overcome this difficulty. In the identified “unsure” regions, further simulator runs should be carried out and the choice of the new input configurations can be guided by taking advantage of the recent advances in adaptive design of experiments (e.g. Gramacy and Herbert, 2009), which constitutes a possible future direction.

Acknowledgements

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

References


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 $\theta$ (blue dots). The bounds of the confidence intervals associated with both GP models are represented by black cross-type markers.
Fig. 3: Overview of the landslide numerical model (adapted from Laloui et al., 2004). The slip surface is outlined by the light coloured orange surface. The observation point 1 (in the upper part of the landslide) and observation point 2 (in the lower part of the landslide) used for the sensitivity analysis of the horizontal displacements are respectively outlined by a blue and a red coloured square-type marker.

Fig. 4: Temporal evolution of the training samples corresponding to the horizontal displacements (m) calculated for 30 different input configurations of the Hujeux law parameters (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)).
Fig. 5: Temporal evolution of the coefficient of determination $R^2$ for the “leave-out-out” cross validation procedure of the GP models constructed at each instant of the crisis period at the observation point 1 in the upper part of the landslide (black dashed line) and at the observation point 2 in the lower part of the landslide (black straight line). The threshold of $80\%$ indicating a “satisfactory” predictive quality is outlined by a horizontal red straight line.
Fig. 6: Temporal evolution during the crisis period of the mean of the main effects (blue straight line) at the observation point 1 in the upper part of the landslide (left) and at the
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.
Table 1: Comparison between the “true” and the estimates of the main effects for the infinite slope analytical model. $\mu$ corresponds to the mean of the main effect computed with the GP model. $CI$ corresponds to the confidence interval defined by the 5 % and the 95 % quantile computed with the GP model.

<table>
<thead>
<tr>
<th>Input factor</th>
<th>True model</th>
<th>GP model constructed with 6 training samples</th>
<th>GP model constructed with 20 training samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thickness $z$ (m)</td>
<td>18.41 %</td>
<td>$\mu=12.74 %$ $CI=[6.48 ; 21.73] %$</td>
<td>$\mu=20.02 %$ $CI=[16.90 ; 23.29] %$</td>
</tr>
<tr>
<td>Slope angle $\theta$ (°)</td>
<td>78.76 %</td>
<td>$\mu=77.01 %$ $CI=[58.41 ; 87.43] %$</td>
<td>$\mu=79.77 %$ $CI=[77.90 ; 81.68] %$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td>K</td>
<td>G</td>
<td>$n_e$</td>
<td>$\phi$</td>
<td>$\Psi$</td>
<td>$\beta$</td>
<td>$pc_0$</td>
</tr>
<tr>
<td>Unit</td>
<td>MPa</td>
<td>MPa</td>
<td>-</td>
<td>-°</td>
<td>-°</td>
<td>-</td>
<td>MPa</td>
</tr>
<tr>
<td>Lower value</td>
<td>180</td>
<td>83.25</td>
<td>0.225</td>
<td>19.125</td>
<td>14.25</td>
<td>20.625</td>
<td>0.375</td>
</tr>
<tr>
<td>Upper value</td>
<td>300</td>
<td>138.75</td>
<td>0.375</td>
<td>31.875</td>
<td>23.75</td>
<td>34.375</td>
<td>0.625</td>
</tr>
</tbody>
</table>