Surrogate-Based Parameter Inference in Debris Flow Model

Maria Navarro[a], Olivier P. Le Maître[b], Ibrahim Hoteit[a], David L. George[c], Kyle T. Mandli[d], Omar M. Knio[a]

[a] *King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
[b] *CMAP, CNRS, INRIA, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France*
[c] *U.S. Geological Survey, Vancouver, WA 98683*
[d] *Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027*

Abstract

This work tackles the problem of calibrating the unknown parameters of a debris flow model with the drawback that the information regarding the experimental data treatment and processing is not available. In particular, we focus on the the evolution over time of the flow thickness of the debris with dam-break initial conditions. The proposed methodology consists of establishing an approximation of the numerical model using a polynomial chaos expansion that is used in place of the original model, saving computational burden. The values of the parameters are then inferred through a Bayesian approach with a particular focus on inference discrepancies that some of the important features predicted by the model exhibit. We build the model approximation using a preconditioned non-intrusive method and show that a suitable prior parameter distribution is critical to the construction of an accurate surrogate model. The results of the Bayesian inference suggest that utilizing directly the available experimental data could lead to incorrect conclusions, including the over-determination of parameters. To avoid such drawbacks, we propose to base the inference on few significant features extracted from the original data. Our experiments confirm the validity of this approach, and shown that it does not lead to significant loss of information. It is further computationally more efficient than the direct approach, and can avoid the construction of an elaborate error model.

Keywords: Bayesian inference, Polynomial chaos expansion, debris flow, Uncertainty quantification

1. Introduction

The present work explores the possibility of exploiting experimental data resulting from measurements in [44] for the purpose of inferring selected parameters in the D-CLAW debris flow model [13, 23]. The inference focuses on a Bayesian approach that has the advantage of providing a complete characterization of the parameters' uncertainty through their resulting posterior distribution. While conceptually simple, performing a Bayesian inference raises several computational and practical difficulties at every one of its constitutive steps. One objective of the present work is to highlight these difficulties and underline issues and risk of applying the Bayesian procedure naively using the available measurements.

The key issue is related to the use of complex, non-linear models such as the one used here. In the context of this study, the use of D-CLAW makes the derivation of closed form expressions for the prediction of physical measurements difficult at best while also being computationally complex and costly; tackling this problem is highly non-trivial. In this context we (1) constructed a polynomial chaos expansion [30, 15] with the intention of using this expansion as a surrogate model for sampling the Bayesian posterior distribution [39, 40, 53]. Unfortunately, constructing this surrogate itself is not straightforward due to the complexity of the physical debris-flow model. This complexity calls for the use of advanced approximation methods, including preconditioned projections [1, 2, 37] coupled with regularized projection methods [27, 18]. This objective is further complicated due to the considered a

Email addresses: maria.navarroi@kaust.edu.sa (Maria Navarro), olivier.le-maitre@polytechnique.edu (Olivier P. Le Maître), ibrahim.hoteit@kaust.edu.sa (Ibrahim Hoteit), dgeorge@usgs.gov (David L. George), kyle.mandli@columbia.edu (Kyle T. Mandli), Omar.Knio@kaust.edu.sa (Omar M. Knio)

Article published in Computational Geosciences

priori parameter range and the multiple parameter regimes that can exist mathematically but inspection can rule out. Complicating even this problem the experimental data from [44] consists of multiple experiments with only derived quantities available. This leads to difficultly in constructing fair error models that adhere to classical assumptions (such as independent measurement noise) with the consequence that conclusions from the analysis can be misleading. Here we examine and demonstrate that alternative observables, while still physically relevant, can lead to significant improvements in surrogate model performance as well as Bayesian analysis results. In this particular case, integrated quantities, temporal points of interest, and particular features of the data can lead to an inference problem that is both more robust and more computationally tractable.

This paper is organized to follow both a description of the constructions used but also following the logical progression of the authors' realization that standard techniques would be insufficient at particular junctures. Section 2 describes the computational model and its uncertain parameters to be inferred, the available experimental measurements, together with a brief outline of the Bayesian inference framework. In Section 3.1, we present a brief discussion on the selection of the parameter priors distributions and simplifications of the physical model. In Section 3.2, we detail the preconditioned polynomial chaos (PC) expansion approach used for constructing surrogates of the observables to be compared with the measurements. In Section 3.3, we validate and analyze the accuracy of the surrogate model and perform an a priori global sensitivity analysis (GSA) of the observables. Section 4 discusses results of the Bayesian inference, comparing the use of different likelihood functions, involving the raw measurements or selected features of the model dynamics. We finally summarize our findings in Section 5.

2. Debris flows

Debris flows are highly mobile, fluidized mass-flows, composed of mixtures of saturated sediment with variably sized solid grains. Debris flows may evolve from a slope failure of a static mass of sediment, which may alternatively restabilize after an initial period of slumping. Therefore, the general class of landslide-type flows to which debris flows belong may behave like a slowly deforming solid, a surging fluid-like flow, and a range of intermediate behaviors. The initial bulk density and sediment properties of a landsliding mass that ultimately slumps and stabilizes, versus one that evolves into a fluidized debris flow, can be nearly undetectable from field measurements. Therefore, the physical mechanisms underlying these contrasting behaviors are subtle yet powerful. While these mechanisms are now widely understood, they are difficult to predict due to sensitivities and uncertainties regarding initial conditions and properties of a static mass of sediment.

Debris flows owe their exceptional mobility to the presence of high pore-fluid pressures which greatly exceed hydrostatic pressures and reduce the effective Coulomb frictional stresses of solid grains. These elevated pore pressures are due primarily to the contraction of the solid grain concentration upon shearing. Conversely, granular materials that tend to dilate upon shearing yield reduced pore-fluid pressures, which can encumber mobility or lead to restabilization (*e.g.,* [21]). In a real debris flow, the prevalence of these contrasting behaviors may evolve both spatially and temporally. Therefore, debris-flow models, or models developed to simulate more general forms of landslides, must contend with this evolution and variation of rheological-type attributes. Models that adopt a single fluid-like or solid-like rheological rule cannot generally reproduce this range of possible flow fates by simply varying initial conditions or model parameters [22].

For our inference problem, we focus on the debris flow model, D-Claw, ([24, 14]) briefly summarized in the following section. The parameters of the inference problem are subsequently introduced as well as the Bayesian framework adopted for their calibration.

2.1. The D-Claw numerical model

D-Claw is a depth-averaged two-phase flow model developed to simulate landslides and debris flows from initiation to deposition. In this model, there is a coupled coevolution of volume fractions, or porosity, and pore-fluid pressure. This coevolution is mediated by granular dilantancy, a principle developed in granular mechanics theory (*e.g.,* [46, 49, 22, 24, 14]). The dilantancy angle of a granular composition describes whether shearing will induce solid contraction or dilation, and hence fluid pressure elevation or reduction. For simple compositions, such as uniform spheres, the dilantancy angle can be precisely calculated. For poorly sorted geological debris, the angle must be based on phenomenological empiricisms. In the D-Claw model, the dilatancy is related to the difference of the ambient solid-volume fraction and an evolving equilibrium solid-volume fraction, which is a function of the ambient stress

state of the material and a fixed critical-state volume fraction (the quasi-static equilibrium solid-volume fraction of the sediment grain-size distribution). Therefore, an important fixed parameter of the model is the difference between the initial porosity and the critical-state porosity. Relatively loose materials will tend to initially contract and mobilize, and relatively dense materials will tend to initially dilate and stabilize. The feedback between contraction/dilation and the pore-fluid pressure ultimately determines mobility in the model. Important parameters regulating the timescale of pore-pressure evolution are the hydraulic permeability, related to the pore-pressure diffusivity of the mixture, and the pore-fluid viscosity (which includes the concentration of suspended silts and clays). These latter physical quantities are in theory measurable, but can be difficult to determine in practice and must be estimated for model runs.

For the experimental debris flows addressed in this work, the D-Claw equations can be written in a simplified form for flow on an inclined basal surface, with a bed-normal coordinate system, where x is tangent to the surface pointing in the highest slope direction and y in the transverse direction (see [23] for a detailed derivation). The model equations involve five independent variables: the debris height $h(x,y,t)$ (in the normal direction), the vector of depth- averaged velocities $\boldsymbol{u}$ with components $u(x,y,t)$ and $v(x,y,t)$ in the x and y directions, the depth-averaged solid volume fraction $m(x,y,t)$, and the pore-fluid pressure at the bed $p_b(x,y,t)$. The governing equations are then

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = \varphi_1, \tag{1}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}(hu^2) + \kappa \frac{\partial}{\partial x}(0.5 g_z h^2) + \frac{\partial (huv)}{\partial y} + \frac{h(1-\kappa)}{\rho}\frac{\partial p_b}{\partial x} = \varphi_2, \tag{2}$$

$$\frac{\partial (hv)}{\partial t} + \frac{\partial (huv)}{\partial x} + \frac{\partial}{\partial y}(hv^2) + \kappa \frac{\partial}{\partial y}(0.5 g_z h^2) + \frac{h(1-\kappa)}{\rho}\frac{\partial p_b}{\partial y} = \varphi_3, \tag{3}$$

$$\frac{\partial (hm)}{\partial t} + \frac{\partial (hum)}{\partial x} + \frac{\partial (hvm)}{\partial y} = \varphi_4, \tag{4}$$

$$\frac{\partial p_b}{\partial t} - \chi u \frac{\partial h}{\partial x} + \chi \frac{\partial (hu)}{\partial x} + u\frac{\partial p_b}{\partial x} - \chi v \frac{\partial h}{\partial y} + \chi \frac{\partial hu}{\partial y} + v \frac{\partial p_b}{\partial y} = \varphi_5. \tag{5}$$

In these equations g_z is the gravitational acceleration in the bed-normal direction, κ the lateral pressure coefficient (see [24]) and ρ the depth-averaged bulk density. The latter is given by $\rho = \rho_s m + (1-m)\rho_f$ where the intrinsic material densities of the solid and fluid components are fixed to $\rho_s = 2700$ and $\rho_f = 1100$, respectively. Finally, the coefficient $\chi = \rho g_z \frac{\rho_f + 3\rho}{4\rho}$ is introduced for notational convenience. The source terms appearing in governing equations are given by:

$$\varphi_1 = \frac{(\rho - \rho_f)}{\rho}\frac{-2k}{h\mu}(p_b - \rho_f g_z h), \tag{6}$$

$$\varphi_2 = h g_x + u\frac{(\rho - \rho_f)}{\rho}\frac{-2k}{h\mu}(p_b - \rho_f g_z h) - \frac{(\tau_{s,x} + \tau_{f,x})}{\rho}, \tag{7}$$

$$\varphi_3 = h g_y + v\frac{(\rho - \rho_f)}{\rho}\frac{-2k}{h\mu}(p_b - \rho_f g_z h) - \frac{(\tau_{s,y} + \tau_{f,y})}{\rho}, \tag{8}$$

$$\varphi_4 = \frac{2k}{hu}(p_b - \rho_f g_z h) m \frac{\rho_f}{\rho}, \tag{9}$$

$$\varphi_5 = \zeta \frac{-2k}{h\mu}(p_b - \rho_f g_z h) - \frac{3}{\alpha h}\|\boldsymbol{u}\| \tan(\psi), \tag{10}$$

where

$$\zeta = \frac{3}{2\alpha h} + \frac{g_z \rho_f (\rho - \rho_f)}{4\rho}, \quad \alpha = \frac{a}{m(\rho g_z h - p_b + \sigma_0)}. \tag{11}$$

The definition of the elastic compressibility of the debris, α, involves the compressibility constant $a > 0$ and the initial normal stress $\sigma_0 = 1000$. The hydraulic permeability, k, follows the empirical formula, $k = k_0 \exp(\frac{0.6-m}{0.04})$, where k_0 is the initial permeability. The basal shear tractions exerted on the solid and fluid phases, $\boldsymbol{\tau}_s$ and $\boldsymbol{\tau}_f$ respectively, are expressed according to:

$$\boldsymbol{\tau}_f = \frac{2\mu(1-m)}{h}\boldsymbol{u} \quad \text{and} \quad \boldsymbol{\tau}_s = (\rho g_z h - p_b)\tan(\phi + \psi)\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|},$$

where μ is the viscosity of the pore fluid, ϕ is the basal friction angle for dry sediment and ψ the granular dilatancy angle. Following [23], the dependence of ψ on m is expressed as

$$\tan(\psi) = m - \frac{m_{crit}}{1 + \sqrt{\frac{\mu\dot{\gamma}}{\rho_s\dot{\gamma}^2\delta^2 + (\rho g_z h - p_b)}}} \quad \text{with} \quad \dot{\gamma} = 2\|\boldsymbol{u}\|/h, \tag{12}$$

where $\delta = 0.001$ is the length scale associated with the grain collisions and $m_{crit} > 0$ is the static critical-state solid volume fraction. The hyperbolic character of the governing equations was proved in [13].

We rely on D-CLAW to compute numerical solutions of Eqs. (1)-(5). D-CLAW is an extension of the open source software package CLAWPACK (Conservation Laws Package [38]) dedicated to the simulation of hyperbolic systems of Partial Differential Equations (PDEs) using finite-volume, wave-propagation methods [34, 28, 32, 33].

The primary goal of this work is to estimate the five parameters of the model that is used in the flume example detailed in [12, 13]. These include the static critical- state solid volume fraction m_{crit}, the initial debris hydraulic permeability k_0, the pure-fluid viscosity μ, the steady-state friction angle in contact with the bed ϕ and the compressibility constant a as uncertain parameters. We shall rely on different Bayesian formulations to perform this inference, based on the a priori ranges provided in [12, 13] and using the experimental data described below.

2.2. Experimental data

For the purpose of calibrating the model, we consider the data reported by [44]. The data aggregates the results of 11 large-scale debris-flow flume experiments conducted by the United States Geological Survey (USGS) from 1994 to 2004. We focus on the experiments for a smooth basal bed (1 mm roughness) and debris formed with gravels without a significant fraction of mud grains. For each individual experiment, the static debris, saturated with water, are placed in a hopper behind a 2 m high vertical gate near the top of the flume and are released by suddenly opening the gate. After releasing the gate, the saturated mixture quickly liquefies and forms a debris flow that descends in the flume before emerging onto a nearly planar run-out pad to form deposits. The flume is a straight rectangular channel, 95 m long by 2 m wide and 1.2 m deep, with a slope of 31 degrees in the section of interest. At $x = 74$ m the flume bed begins to fatten, and beyond this point, it follows a catenary curve that drops 2.2 m vertically before reaching the flume mouth at $x = 82.5$ m. The run-out surface extends to $x = 107.5$ m.

In each repetition of the experiment the flow thickness (h) is measured in time at three cross sections located at $x = 2$ m, $x = 32$ m and $x = 66$ m. Data available in [44] consists of averages and variances of the time series data (depicted in Figure 6). The full data sets were truncated to provide equal length time series removing intervals that occurred before or after the primary debris flow with a resulting temporal resolution of 0.01 seconds.

2.3. Bayesian inference

To infer the value of the model parameters we apply a Bayesian approach. Let $\boldsymbol{q}$ be the generic vector of parameters, having for components the parameters of the analysis, and $\boldsymbol{d}$ the vector of the measured data used for the analysis. In the present Bayesian setting, $\boldsymbol{d}$ and $\boldsymbol{q}$ are real-valued random vectors. In the following, we denote $\pi(\boldsymbol{q})$ the *prior* probability distribution of the parameters that reflects the belief on values of $\boldsymbol{q}$ before the measurements are made available. From the measurements $\boldsymbol{d}$, Bayesian inference updates the prior distribution $\pi(\boldsymbol{q})$ to the *posterior* probability distribution, $\pi(\boldsymbol{q}|\boldsymbol{d})$, following Bayes' rule [11]:

$$\pi(\boldsymbol{q}|\boldsymbol{d}) = \frac{\mathcal{L}(\boldsymbol{d}|\boldsymbol{q})\pi(\boldsymbol{q})}{\int \mathcal{L}(\boldsymbol{d}|\boldsymbol{q})\pi(\boldsymbol{q})d\boldsymbol{q}}, \tag{13}$$

where $\mathcal{L}(\boldsymbol{d}|\boldsymbol{q})$ is the likelihood function, *i.e.* the probability of observing measurements $\boldsymbol{d}$ given $\boldsymbol{q}$. Typically, the likelihood compares the measurements with model predictions (functions of $\boldsymbol{q}$), and relies on a noise model to account for the measurement error; a model error contribution can also be included [26]. Specifically, the comparison can be made on the raw measurements ($\boldsymbol{d}$) or more generally on some derived quantities, say $\boldsymbol{m}(\boldsymbol{d})$. We shall denote $\boldsymbol{M}(\boldsymbol{q})$ the vector of model predictions using the forward model, *i.e.* the debris flow model in our case, such that

$$\boldsymbol{m}(\boldsymbol{d}) = \boldsymbol{M}(\boldsymbol{q}) + \epsilon_{\boldsymbol{m}},$$

where ϵ_m is a random vector accounting for measurement noise and model errors, whose statistical properties must be provided. In Section 4, the effect on the inferred parameters of selecting different error models and derived quantities is investigated.

The posterior distribution in Eq. (13) is usually not known in a closed form due to the complexity of the mapping $\boldsymbol{q} \mapsto \boldsymbol{M}(\boldsymbol{q})$ in addition to the noise and error models. Therefore sampling strategies, such as Markov Chain Monte Carlo (MCMC) methods [36], are needed to estimate the statistics of the posterior distribution of $\boldsymbol{q}$ (e.g. mean, moments, median and mode). In this work we use the Adaptive Metropolis (AM) algorithm [17], an extension of the Random-Walk Metropolis (RWM) algorithm, which adapts the proposal covariance matrix using the previously sampled points. We shall also consider the maximum a posteriori (MAP) point, $\boldsymbol{q}_{\mathrm{MAP}}$, defined by

$$\boldsymbol{q}_{\mathrm{MAP}} = \underset{\boldsymbol{q}}{\operatorname{argmax}}\, \pi(\boldsymbol{q}|\boldsymbol{d}). \tag{14}$$

The MAP can be estimated by the optimum point over the MCMC chain, or by solving Eq. (14) with a dedicated optimization procedure, such as gradient-based methods [43], simulated annealing (SA) [19], etc. To calculate the MAP we rely the MATLAB Optimization Toolbox [41] and use a genetic algorithm [9, 16]. The main drawback of both the MCMC and optimization approaches is that they require a large number of evaluations of the forward model $\boldsymbol{M}(\boldsymbol{q})$. In our case, this means solving the system of PDE in two dimensions thousands of times and as a result the approach would be computationally intractable. Instead we derive a surrogate model $\widehat{\boldsymbol{M}}(\boldsymbol{q})$, with low evaluation cost, that can be used in place of the original model $\boldsymbol{M}(\boldsymbol{q})$ in the likelihood, as proposed in [39]. If the surrogate model is accurate enough and can be constructed at a reasonable computational cost, the sampling of the posterior distribution can be orders of magnitude faster and cheaper than the "direct" sampling using the full forward problem. In the following section we detail the construction of this surrogate model.

3. Surrogate model

In this section we detail the construction of the model surrogate. To remain as general as possible the surrogate approximates the model predictions of the experimental data discussed in Section 2.2, which consists of measurements of the debris flow depth h at three spatial locations and for different times. We start by discussing the prior distribution of the parameters and some simplifications of the physical model. We then explain in detail the construction of the surrogate and its validation.

3.1. A priori parameter distributions

As a starting point, uniform or log-uniform distributions were used as priors, with ranges based on [12, 13] and assuming independent parameters. Specifically we use

$$m_{\mathrm{crit}} \sim \mathscr{U}[0.62,0.66], \quad \phi \sim \mathscr{U}[38,43], \quad a \sim \mathscr{U}[0.01,0.05],$$

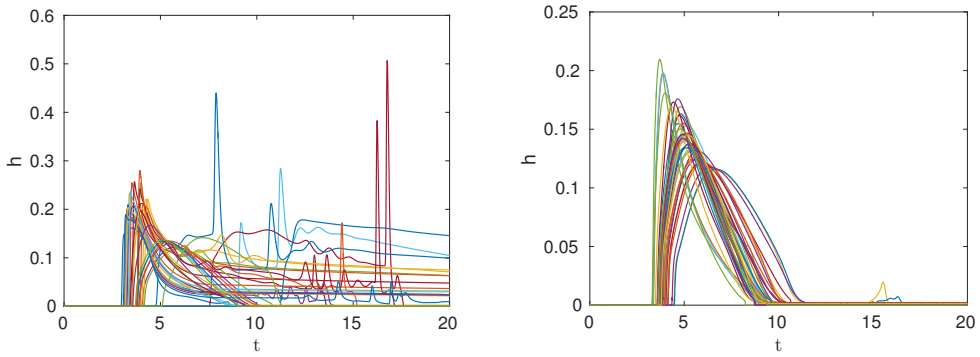
and

$$k_0 \sim \mathscr{U}_{\log}[10^{-12},10^{-8}], \quad \mu \sim \mathscr{U}_{\log}[0.005,0.05],$$

as the prior distributions, where $\mathscr{U}$ and $\mathscr{U}_{\log}$ denote the uniform and log-uniform distributions respectively.

We start by checking that the debris flow dynamics is sufficiently smooth when the parameters $\boldsymbol{q}$ change in the prior range, as smoothness is important to obtain an accurate polynomial surrogate model at a reasonable cost [30, 8]. Sampling the parameters according to their probability distribution and recording the corresponding change in flow thickness h can reveal the smoothness of the forward model dependence on the input parameters. Figure 1(a) demonstrates the complexity of the dynamics that can be expected as the parameters are varied. This includes long-time behaviors and "surge" like behavior (spikes) in the time signals.

Several self convergence studies for the spatial and time discretizations were carried out in order to confirm that the solutions were accurately computed, and to dismiss the possibility of complex dynamics induced by numerical errors. Further investigations of the numerical solutions revealed that the complex dynamics were inherent to the physical model, due to the emergence of fragmentation of the debris plume and development of transverse waves for some values of the parameters.



(a) Original distribution, 2-D model

(b) Modified distribution, 1-D model

Figure 1: Flow thickness versus time for selected values of the parameters and location $x = 32$ m; (a) corresponds to the original parameters range and 2-D model, whereas (b) corresponds to the reduced range for k_0 and 1D model.

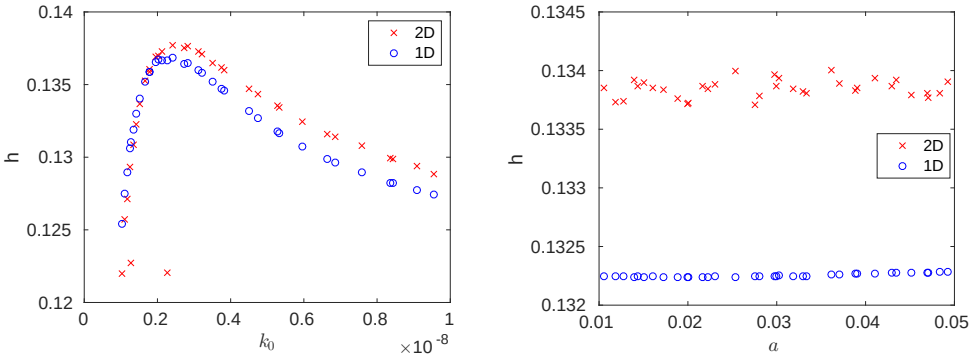
To investigate the fragmentation process, we performed a "one at a time" variation of the parameters around their nominal values given in [13, 12], corresponding to the median values of the distributions above. From these numerical experiments it was concluded that the initial debris hydraulic permeability, k_0, was predominantly responsible for the emergence of the fragmentation and associated peaks in the time signals (something supported by the physics). It was further found that the fragmentation is associated with the lowest values of k_0 in its prior range. In addition, observing that such peaks do not appear in the averaged experimental measurements of the debris elevation (see Figure 6), it was decided to rule out a priori lowest values of k_0 below 10^{-9}, responsible for the fragmentation-induced peaks and to only consider the following prior distributions of the parameters:

$$m_{\text{crit}} \sim \mathscr{U}[0.62,0.66], \quad k_0 \sim \mathscr{U}_{\log}[10^{-9},10^{-8}], \tag{15}$$

$$\mu \sim \mathscr{U}_{\log}[0.005,0.05], \quad \phi \sim \mathscr{U}[38,43], \quad a \sim \mathscr{U}[0.01,0.05]. \tag{16}$$

The transverse waves mentioned previously also lead to issues due to loss of uniqueness in the solution. This is due to the fact that transverse waves can propagate asymmetrically along the y direction. As the propagation direction is highly variable, the meaning or representativity of a particular numerical prediction is questionable. Furthermore, the construction of the surrogate model raises questions regarding the definition of the model prediction and its use in the likelihood function. To avoid these issues, the spatial discretization of the model was reduced to a single cell in the cross-wise y direction, effectively removing the y-dependence of the solution while retaining all other characteristics of the flow. With this dimension reduction we not only recover uniqueness of the solution, but also rely on simulations that are faster and cheaper to produce. The effect of the dimension reduction on the model predictions were carefully analyzed and it was found that the reduction has no impact on the long-time solution. Differences between the 1D and 2D predictions are observed at early times, particularly when the debris height is close to its maximum, as reported in Figure 2, showing the "one at a time" sampling for parameters k_0 and a. The plots shows that the trends and dependencies are correctly captured, but in contrast with the 2D case the dependence of the 1D model solution with respect to the parameter is smooth. Finally, the differences between the 1D and 2D models are smaller than or comparable to the standard deviation of the experimental data (Figure 6) and should not have a significant impact on the inference.

Finally, Figure 1(b) shows the time evolution of the predicted debris elevation at $x = 32$ m for a random sample set of parameters, drawn from their reduced range in Eq. (15) and (16), using the 1D model. One can see that the time signals are much smoother than those of the original setting shown in Figure 1(a). In fact, except for few remaining low-amplitude secondary peaks at later times ($t \approx 15$), associated to extreme parameters values and corresponding to a fragmentation of the debris mass in the wake of the flow, the dependencies with the parameters can now be well approximated by a surrogate model, as discussed below.



(a) Sampling k_0 (b) Sampling a

Figure 2: Comparison of the 1-D and 2-D model predictions. One at a time variation of parameters k_0 (left) and a (right) for the predicted debris height h at $t = 5.04$ and $x = 32$ m.

3.2. PC surrogate

For the surrogate model construction we rely on Polynomial Chaos (PC) expansion methods [30, 54] using a preconditioned, non-intrusive approach [1]. A non-intrusive approach is preferred in this work because of the complexity of the forward model that prevents using other alternatives, such as a stochastic Galerkin projection method for conservation laws [15, 31]. The non-intrusive approach also allows us to precondition the construction, exploiting the self-similarities between the dynamics for different values of the parameters (see Figure 1(b)), and improve the accuracy of the surrogate model as shown below. For convenience, we build the surrogate in terms of canonical independent random variables $\boldsymbol{\xi}$ having uniform distributions in the unit interval [0,1]. Here, five random variables ξ_i are used to parametrize the physical parameters in $\boldsymbol{q}$ through a simple transformation. Abusing notation, we shall write $\boldsymbol{q}(\boldsymbol{\xi})$ to stress the parametrization of the random parameter vector with the canonical variables, and use the short-hand notation $h(x,t,\boldsymbol{\xi})$ in lieu of $h(x,t,\boldsymbol{q}(\boldsymbol{\xi}))$.

3.2.1. Preconditioning

The objective of preconditioning is to reduce the variance of the quantity being represented and is typically achieved by introducing an *invertible* transformation Φ of the model prediction [1]. As an example, consider the case of approximating $h(t,\boldsymbol{\xi})$ for some fixed spatial location x. We seek a transformation of the form

$$h(t,\boldsymbol{\xi}) \mapsto \Phi[h(t,\boldsymbol{\xi})] = g(\tau(t,\boldsymbol{\xi}),\boldsymbol{\xi}),$$

where $g(\tau(t,\boldsymbol{\xi}),\boldsymbol{\xi}) \in L_2([0,1]^5)$ (*i.e.* a second-order stochastic process). Following [1], the variance of $g(\tau,\boldsymbol{\xi})$ is minimized by transforming h using scaling factors designed to synchronize the realizations when the parameters change. In light of the results in Figure 1, we propose to consider of the following scaling factors: the arrival time of the debris t_{arr}, the time t_{max} and value h_{max} at the signal maximum, and the time t_{dec} when the signal has decayed to 40% of its maximum. The scaling factors are summarized in Table 1.

Scaling factor	Symbol	Definition
Arrival time	$t_{\text{arr}}(\boldsymbol{\xi})$	First time $h(t,\boldsymbol{\xi})$ exceeds $\varepsilon \ll 1$
Height at maximum	$h_{\text{max}}(\boldsymbol{\xi})$	$h_{\text{max}}(\boldsymbol{\xi}) = \max_t h(t,\boldsymbol{\xi})$
Time at maximum	$t_{\text{max}}(\boldsymbol{\xi})$	$t_{\text{max}}(\boldsymbol{\xi}) = \arg\max_t h(t,\boldsymbol{\xi})$
Decay time	$t_{\text{dec}}(\boldsymbol{\xi})$	$t_{\text{dec}}(\boldsymbol{\xi}) > t_{\text{max}}(\boldsymbol{\xi})$ such that $h(t_{\text{dec}},\boldsymbol{\xi}) = 0.4 h_{\text{max}}(\boldsymbol{\xi})$

Table 1: Definition of the scaling factors for the preconditioning.

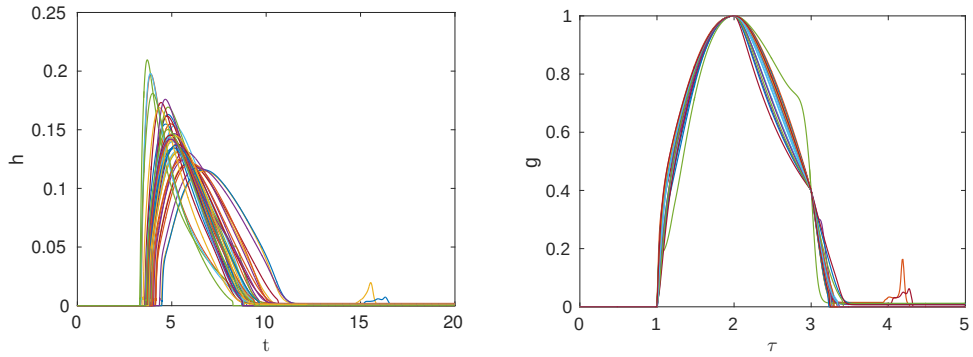
We define, for $t \in [0,20]$, the transformation of $h(t,\boldsymbol{\xi})$ to its preconditioned version $g(\tau,\boldsymbol{\xi})$ according to

$$\Phi\left[h(t,\boldsymbol{\xi})\right] = g(\tau(t,\boldsymbol{\xi}),\boldsymbol{\xi}) \doteq \frac{h(t,\boldsymbol{\xi})}{h_{\max}(\boldsymbol{\xi})},$$

where $\tau(t,\boldsymbol{\xi}) \in [0,5]$ is the scaled time defined by

$$\tau(t,\boldsymbol{\xi}) = \begin{cases} \dfrac{t}{t_{\text{arr}}(\boldsymbol{\xi})} & 0 \le t < t_{\text{arr}}(\boldsymbol{\xi}), \\ 1 + \dfrac{t - t_{\text{arr}}(\boldsymbol{\xi})}{t_{\max}(\boldsymbol{\xi}) - t_{\text{arr}}(\boldsymbol{\xi})} & t_{\text{arr}}(\boldsymbol{\xi}) \le t < t_{\max}(\boldsymbol{\xi}), \\ 2 + \dfrac{t - t_{\max}(\boldsymbol{\xi})}{t_{\text{dec}}(\boldsymbol{\xi}) - t_{\max}(\boldsymbol{\xi})} & t_{\max}(\boldsymbol{\xi}) \le t < t_{\text{dec}}(\boldsymbol{\xi}), \\ 3 + 2\dfrac{(t - t_{\text{dec}}(\boldsymbol{\xi}))}{20 - t_{\text{dec}}(\boldsymbol{\xi})} & t_{\text{dec}}(\boldsymbol{\xi}) \le t \le 20. \end{cases} \tag{17}$$

It is observed that the mapping $t \mapsto \tau$ is invertible for given $\boldsymbol{\xi}$. Figure 3 depicts the effect of the preconditioning on a random sample set of $h(t,\boldsymbol{\xi})$ at $x = 32$ m. It is seen that $g(\tau,\boldsymbol{\xi})$ indeed shows a reduced level of variability compared to its untransformed version, and by construction a vanishing variability for $\tau = 1,2$ and 3. Except for the remaining small peaks for $\tau \approx 4$, the approximation of the preconditioned signal will be easier to represent than the original one, due to the reduced variability [1].



(a) Unscaled realizations (b) Scaled realizations

Figure 3: Selected realizations of the original $h(t,\boldsymbol{\xi})$ and preconditioned version $g(\tau,\boldsymbol{\xi})$ at location $x = 32$ m.

3.2.2. Preconditioned surrogate

We now return to the construction of the surrogate model. We assume that a sample set of m observations of $h(\cdot,\boldsymbol{\xi})$ has been computed for a sample set of realizations $\boldsymbol{\xi}_i$. Different strategies can be used in the selection of the set of sampling points [3, 6]. Here we use a Latin hypercube sampling [35] to generate the set of the random parameters with m =1,471 points. Based on this sample set, we derive corresponding sample sets for the scaling parameters and denote $t_{\text{arr}}^{(i)} = t_{\text{arr}}(\boldsymbol{\xi}_i)$ the i-th realization of the arrival time.

We first build PC approximations of the scaling factors. To this end, we consider a PC basis $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \boldsymbol{\alpha} \in \mathscr{A}\}$, where $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ denotes the tensorized orthonormal Legendre polynomial basis. Let $P = \text{card}(\mathscr{A})$ be the basis dimension resulting from a truncation based on the total order p. Provided the scaling factors are second-order random variables, they have the convergent PC approximation [7]:

$$\widehat{h}_{\max}(\boldsymbol{\xi}) = \sum_{k=1}^{P} h_{\max}^k \Psi_k(\boldsymbol{\xi}) \quad \widehat{t}_{\text{arr}}(\boldsymbol{\xi}) = \sum_{k=1}^{P} t_{\text{arr}}^k \Psi_k(\boldsymbol{\xi}) \tag{18}$$

$$\widehat{t}_{\max}(\boldsymbol{\xi}) = \sum_{k=1}^{P} t_{\max}^k \Psi_k(\boldsymbol{\xi}) \quad \widehat{t}_{\text{dec}}(\boldsymbol{\xi}) = \sum_{k=1}^{P} t_{\text{dec}}^k \Psi_k(\boldsymbol{\xi}).$$

The convergence is in the mean-squared sense as the polynomial degree $p \to \infty$. The expansion coefficients of the PC expansion are computed using regression methods [52, 20, 4], which are more robust to discontinuous dependencies and noisy observations compared to alternative non-intrusive methods [10, 45]. In these methods, the expansion coefficients are the solution of an optimization problem which aims at minimizing the prediction error, in some prescribed norm, while possibly considering some regularization terms. In the present work, the basis dimension P is small enough compared to the number of observations, such that a regularization was not necessary. To construct the PC approximations, we tested both L_2 and L_1 norms corresponding to a least squared (LS) residual and a least absolute deviation (LAD) optimization problem. For example, the optimization problem for arrival time aims at minimizing for its PC coefficients $t^1_{\text{arr}}, \ldots, t^P_{\text{arr}}$ the distance function

$$R^\ell(t^1_{\text{arr}}, \ldots, t^P_{\text{arr}}) = \sum_{i=1}^{m} \left| t_{\text{arr}}(\boldsymbol{\xi}_i) - \sum_{k=1}^{P} t^k_{\text{arr}} \Psi_k(\boldsymbol{\xi}_i) \right|^\ell , \tag{19}$$

with $\ell = 1$ (resp. 2) for the LAD (resp. LS) method. As shown below, the LAD method is more robust than LS [48] in presence of non-smooth dependences. In practice we use the Iterative Re-weighted Least Squares Algorithm (IRLS) [29] to minimize the residuals in the LAD case.

Once the PC approximation of the scaling factors have been computed, the sample set of model observations $h(t,\boldsymbol{\xi}_i)$ can be transformed to obtain the sample set of preconditioned variables $g(\tau,\boldsymbol{\xi}_i) = \Phi\left[h(t,\boldsymbol{\xi}_i)\right]$. In practice, $g(\tau,\boldsymbol{\xi}_i)$ is evaluated on a fine grid in the scaled times $\tau_j \in [0,5]$, through simple time interpolations [1]. Then, for every scaled time τ_j, we consider the PC expansion of the preconditioned variable,

$$\widehat{g}(\tau_j,\boldsymbol{\xi}) = \sum_{k=1}^{P} g_{j,k} \Psi_k(\boldsymbol{\xi}). \tag{20}$$

Again, the PC coefficients of the preconditioned variable at scaled time τ_j can be computed by regression, to complete the construction of the surrogate model.

Finally, the evaluation of the surrogate model $\hat{h}(t,\boldsymbol{\xi})$ for some given $\boldsymbol{\xi}$ is obtained as follows. First, the preconditioned variables $\widehat{g}(\tau_j,\boldsymbol{\xi})$ are evaluated using Eq. (20) for all τ_j. This is computationally effective as it essentially amounts to evaluating P polynomials. Using the PC approximation of the scaling variables, one can determine unscaled time t_j corresponding to the $\tau_j(\boldsymbol{\xi})$, by simply inverting the time transformation in (17). Subsequent time interpolations between the t_j and a final scaling by $\hat{h}_{\max}(\boldsymbol{\xi})$ completes the evaluation of $\hat{h}(t,\boldsymbol{\xi})$ for any $t \in [0,20]$ given $\boldsymbol{\xi}$.

3.3. Surrogate validation and a priori sensitivity analysis

3.3.1. Surrogate error

For the validation of the surrogate model we consider the following estimator of the L_2 error norm,

$$\varepsilon^2(t) = \left\| \hat{h}(t,\cdot) - h(t,\cdot) \right\|_2^2 \approx \frac{1}{N} \sum_{i=1}^{N} (\hat{h}(t,\boldsymbol{\xi}_i) - h(t,\boldsymbol{\xi}_i))^2, \tag{21}$$

where h and $\hat{h}$ are the true (D-Claw) and surrogate model predictions respectively. The approximation of the norm uses a secondary sample set of N random samples, distinct from the Latin hypercube sample set used for the construction of the surrogate model. In the following tests, we used N =2,500 samples, a value that was found sufficient to obtain satisfactorily converged error estimates.

Figure 4 reports the evolution with t of the error $\varepsilon(t)$ at $x = 32$ m (similar results are obtained at other locations), contrasting different surrogate constructions. Specifically, for comparison purposes, the errors are shown for a direct PC projection of $h(t,\boldsymbol{\xi})$, that is using $g(\tau,\boldsymbol{\xi}) = h(t,\boldsymbol{\xi})$ with $\tau(\boldsymbol{\xi}) = t$, and the LAD or LS distance in the regression problems, and for the corresponding preconditioned surrogates labeled P-LAD and P-LS, respectively. These error curves are obtained for a fixed PC basis, using order $p = 3$, so the basis dimension is $P = 321$. The PC order was selected carefully to ensure sufficient surrogate accuracy. The reported errors show that the preconditioned surrogates, regardless of the regression method used, are significantly more accurate than their direct PC projection counterparts.

The errors also show that the two regression approaches perform similarly, in the preconditioned case, in terms of resulting L_2 error norms. However, a throughout inspection of the preconditioned surrogate models reveals that the LAD regression yields a significantly lower error in L_∞ norm (maximum of absolute error); this approach is thus preferred over the LS regression. Consequently, all results presented in the following sections are based on the preconditioned surrogate constructed with LAD regression.

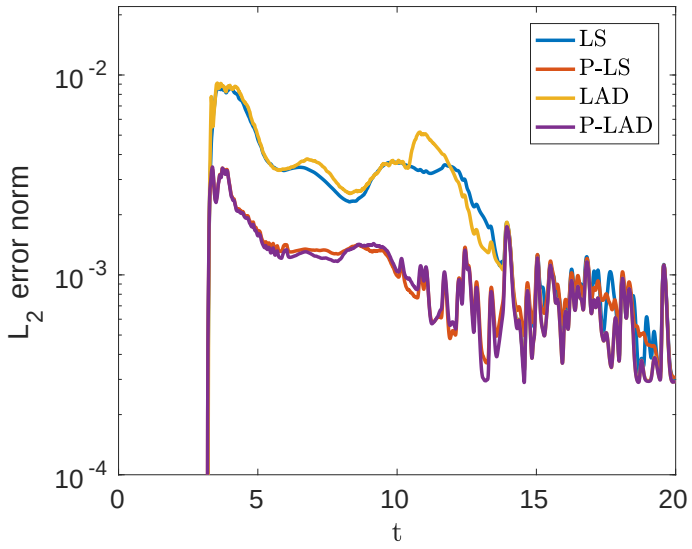
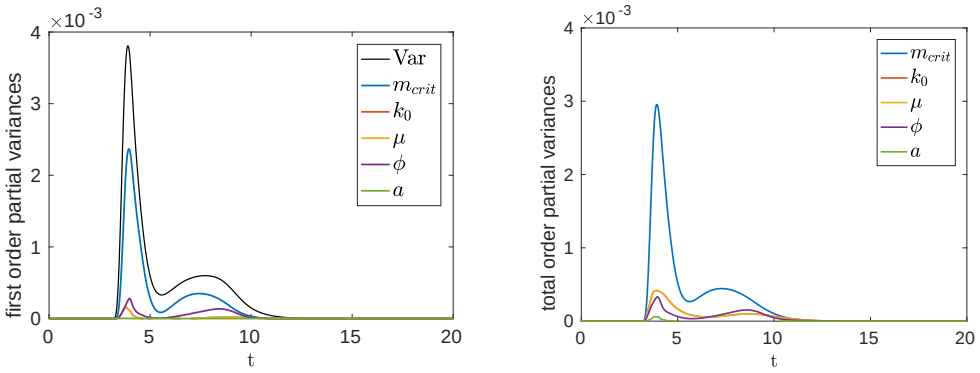


Figure 4: Errors for different surrogate constructions based on a sample of size 2,500.

3.3.2. A priori variability

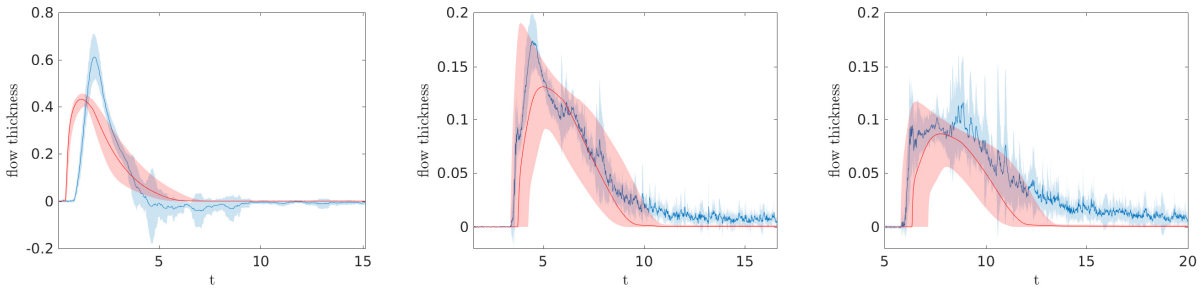
Before addressing the inference of the parameters, the surrogate model is used to conduct an a priori Global Sensitivity Analysis (GSA) of the wave thickness h. To this end, we rely on a variance decomposition method, or Sobol-Hoeffding (SH) decomposition, to derive sensitivity indices [51]. The surrogate model being preconditioned, explicit expressions of the sensitivity coefficients from the PC expansions [52] can not be used; we then rely on Monte-Carlo sampling of the surrogate, following the approach proposed in [47] for the first-order partial variances and the Jansen formula [25] for the estimation of the total order partial variances. Figure 5 summarizes the results of the GSA at location $x = 32$ m as, again, similar findings are obtained at the other locations. Figure 5(a) shows the first order partial variances, measuring the variability due to individual parameters only, whereas Figure 5(b) depicts the total order partial variances, measuring the variability due to a parameter including its interactions with the others [51]. It is seen that most of the variability in h is due to the static critical-state solid volume fraction, m_{crit}, specially during the initial stages after the arrival of the debris at the considered location. The figures also demonstrate moderate, but non-negligible, interactions between parameters as the first order and total order partial variances are similar.

To complete the analysis of the surrogate model, we present in Figure 6 the resulting a priori predictive range of h as a function of time, for the three measurement locations x. The a priori predictive range is defined by the 5th to 95th percentiles values and is depicted with the red shaded area in the plots. The percentiles were estimated from a random sampling of the surrogate model with 15,000 points. The experimental data are also reported using the mean values with their plus and minus one standard deviation bounds, plotted in blue. The plots highlight significant discrepancies between the model predictions and the measurements. In particular, the arrival times at locations $x = 2$ m and the long-term behavior are clearly not well predicted by the model, with a faster decay of h in the model.



(a) first order partial variances (b) total order partial variances

Figure 5: Global sensitivity analysis of the wave height at $x = 32$ m.



(a) $x = 2$ m (b) $x = 32$ m (c) $x = 66$ m

Figure 6: A priori predictive range of the surrogate for the debris flow thickness h (red shaded area) and mean plus and minus one standard deviation bounds of the experimental measurements (blue shaded area). The a priori predictive range is defined by the 5th percentile to 95th percentile interval.

4. Inference of debris flow parameters

In this section, we report the results of the Bayesian inference of the model parameters, and contrast the cases of different likelihood functions. In particular, we contrast the inference results for the case of a likelihood comparing directly the reported experimental data to the model predictions, in Section 4.1, with the case of an inference based on quantities derived from the available data, in Sections 4.2 and 4.3.

We recall that the experimental measurements concerned the wave height h at different times and at three different locations, see Section 2.2. In fact, the available data consist of the debris height $\bar{h}_i$ averaged over 11 repetitions of the experiment and its standard deviations σ_i, as depicted in Figure 6. We stress that the measurements of individual experiments are not provided. In addition, we have retained 1,501 data at each location, so a total of $m_d = 1501 \times 3$ data are available for the inference.

4.1. Direct inference from the experimental data

We begin by performing the inference directly on the available experimental data. Using the notation of Section 2.3, this corresponds to using $\boldsymbol{m}(\boldsymbol{d}) = \boldsymbol{d}$. The posterior distribution of the parameter is fully determined by the likelihood function $\mathcal{L}(\boldsymbol{d}|\boldsymbol{q})$. Recalling that $\boldsymbol{q}$ is parametrized by the canonical variables $\boldsymbol{\xi}$ we shall also write the likelihood $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi})$ in the following.

Gaussian models are often considered to model likelihood functions assuming that the discrepancy between the observed and predicted values is Gaussian. Following this path, we write

$$\bar{h}_i = \widehat{h}(t_i, x_i, \boldsymbol{\xi}) + \epsilon_i, \quad i = 1, \ldots, m_d, \tag{22}$$

where t_i and x_i denote here the time and location of the i-th experimental data. One still has to specify the structure of the Gaussian vector $\boldsymbol{\epsilon} = (\epsilon_1 \cdots \epsilon_{m_d})$. Based on the available data, the simplest model for $\boldsymbol{\epsilon}$ is a vector of independent components having zero mean and a variance given by the estimated experimental variability, that is $\epsilon_i \sim \mathcal{N}(0,\sigma_i^2)$. In this case, the likelihood expression becomes

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi}) = \prod_{i=1}^{m_d} \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(\bar{h}_i - \widehat{h}_i(\boldsymbol{\xi}))^2}{2\sigma_i^2}\right], \tag{23}$$

where $\widehat{h}_i(\boldsymbol{\xi}) \doteq \widehat{h}(t_i,\boldsymbol{\xi})$ is the surrogate model prediction of the i-th experimental data.

When expressed in terms of the canonical variables $\boldsymbol{\xi}$, which are uniformly distributed, the MAP determination reduces to the maximization of the likelihood because $\pi(\boldsymbol{\xi})$ is constant. For convenience, the log-likelihood is maximized in practice, using a genetic algorithm as discussed before. In addition to the MAP computation, we also sample the posterior distribution by means of Markov-Chain Monte-Carlo method, to estimate the posterior density of the parameters (or canonical random variables) as well as the posterior predictive distribution of the debris height. In all computations presented hereafter, the convergence of the MCMC chains have been carefully checked, through visual inspection and computation of the empirical auto-correlation at lags [5]. Typically, 60,000 steps after the burn-in period were used to estimate the statistics. A Standard Kernel Density Estimation (KDE) method [42, 50] was subsequently applied on the MCMC samples to obtained the posteriori densities.

Figure 7(a) compares the posterior predictive range using the likelihood definition in Eq. 23 (labeled original, shaded in red) and the experimental data. The posterior predictive range is again defined based on the 5th to 95th percentiles. It is seen that the posterior predictive range remains tight around the MAP prediction, suggesting a high confidence in the inferred values for the parameters. This is confirmed by the highly concentrated marginal posteriors of the canonical variables reported in Figure 8(a).

This impressive reduction of the prior distribution to a highly concentrated posterior should be contrasted with the actual remaining discrepancies between the model predictions and the assumptions used to derive the likelihood function. First, it should be noted that the posterior predictive range does not contain the experimental data. In addition, inspection of the residuals for the MAP solution, $\epsilon_i = \bar{h}_i - \widehat{h}_i(\boldsymbol{\xi}_{MAP})$ reported in ref in Figure 7(b), reveals that the ϵ_i are clearly not independent, but quite correlated in time, and assume values inconsistent with the experimental values σ_i.

Clearly, a poor discrepancy model is responsible for the over determination of the posterior of the parameter vector. To mitigate this issue, it would be necessary to introduce more elaborate statistical models for the ϵ_i, involving bias and correlations, and possibly using hyper-parameters to be inferred for as well. As provided, the variances σ_i^2 used in the definition (23) of the likelihood function are questionable. Indeed, the values σ_i^2 characterize a variability in the experimental procedure and apparatus, and not a distance to the model prediction. Also, the available information does not give access to any correlation structure as individual measurements are not reported. In fact, the large and erratic changes in the value of σ_i^2 from one time to the next, seen in Figure 6, suggest at least a poorly converged estimation of the experimental variability. This observation leads us to smooth the variances σ_i^2, discarding the idea that noisy variances estimates could result in the over-determination of the parameters. To this end, we use a moving average filter to smooth out short-term fluctuations. The smoothed variances, at a given location x, are defined generically as

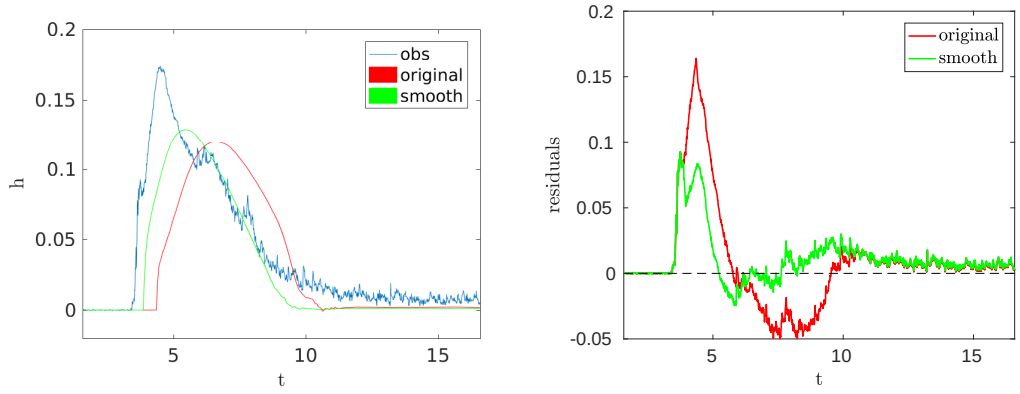
$$\bar{\sigma}_i^2 = \frac{1}{2s+1} \sum_{j=-s}^{s} \sigma_{i+j}^2, \tag{24}$$

where $2s$ is the window size of the filter, and the time series of σ_i^2 at the considered location is padded outside of their original support in time.

We repeated the inference task considering the likelihood function (23) with the smoothed variance instead of the original σ_i^2, and considering time windows with increasing sizes ranging from $s = 25$ to $s = 750$. We also perform the experiments using $\sigma_i^2 = \bar{\sigma}^2$, the average over the three locations. Concerning the sequence of MAP points obtained for increasing s, it is found to converge to the case of the constant variance $\bar{\sigma}^2$. Obviously, this case also corresponds to the optimum in terms of resulting residuals norm at the MAP prediction, as measured by

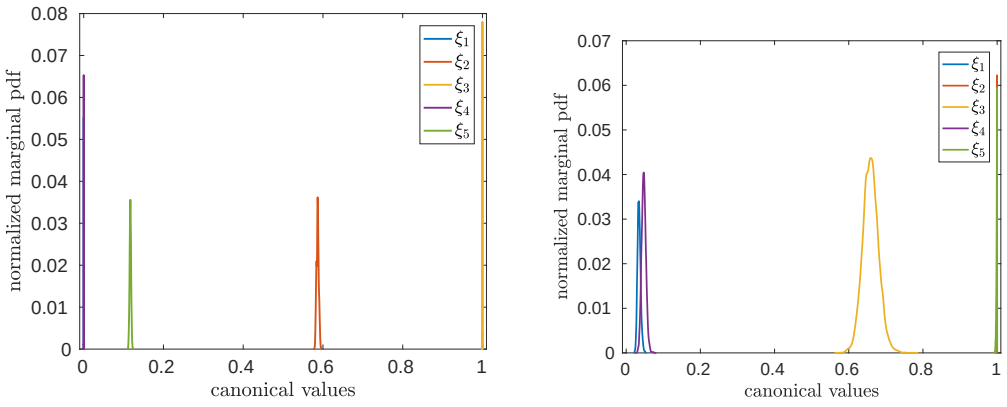
$$\|\boldsymbol{\epsilon}_{\mathrm{MAP}}\|_2^2 = \sum_{i=1}^{m_d} |\bar{h}_i - \widehat{h}_i(\boldsymbol{\xi}_{\mathrm{MAP}})|^2. \tag{25}$$

The improvement of the residuals for the MAP solution using $\bar{\sigma}^2$ can be appreciated from Figure 7(b), while the effect of the smoothing on the posterior predictive range can be seen in Figure 7(a). Comparing the original and smoothed MAP, it is seen that the agreement with the experimental data has improved for the smoothed case, with a better (though not perfect) fit of the arrival time and during the early decay phase. However, even for the highest smoothing, it is seen that the posterior predictive range remains extremely tight, with a highly concentrated posterior, as can be seen in Figure 8(b). Again we claim that a rough and too simplistic discrepancy model is causing an overly narrow estimation of the model parameters. To overcome this issue we pursue in the following an entirely different definition of the likelihood function.



(a) Predictive ranges and data (b) Residuals

Figure 7: Directed inference: (a) Comparison between the data and the a posteriori predictive ranges for the original and smoothed variances ($\bar{\sigma}^2$), at $x = 32$ m. (b) Corresponding residuals of the MAP prediction.



(a) Original variances (b) Smoothed variances

Figure 8: Direct inference: Marginal posterior distributions of the parameters, for the original (a) and the smoothed variances ($\bar{\sigma}^2$).

4.2. Calibration from the scaling variables

To avoid the complicated task of designing an appropriate discrepancy model and based on the observation that we mostly focus on the discrepancies in the arrival time, location of the maximum height and decay behavior, we propose to consider a likelihood function based on the scaling factors involved in the construction of the preconditioned surrogate model. Specifically, we base the inference approach on the arrival time, time at maximum, debris height

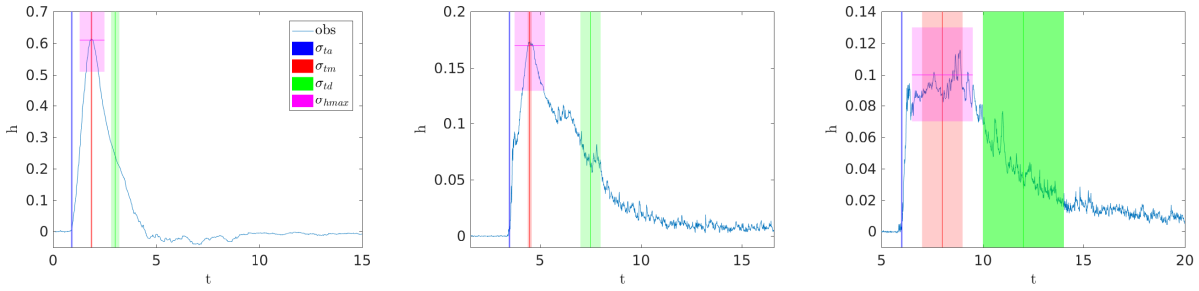
at maximum and time to decay, as defined in Table 1. By doing so, we aim at focusing the parameter inference on the important features predicted by the model, thus mitigating the impact of noisy details present in the experimental data. In addition, defining the likelihood from these few scaling factors alleviates the computational burden of having to evaluate the preconditioned surrogate for a large number of h_i, with greater computational efficiency as a result.

For the definition of the likelihood, we again assume the error on the scaling factors to have independent components corresponding to normally distributed random variables with zero mean. The log-likelihood at a given location x is expressed as

$$\begin{aligned}\ln(\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi})) \quad \propto \quad & -\left(\frac{t_{\text{arr}} - \widehat{t}_{\text{arr}}(\boldsymbol{\xi})}{2\sigma_{arr}}\right)^2 - \left(\frac{t_{\text{max}} - \widehat{t}_{\text{max}}(\boldsymbol{\xi})}{2\sigma_{t_{\text{max}}}}\right)^2 - \\ & - \left(\frac{t_{\text{dec}} - \widehat{t}_{\text{dec}}(\boldsymbol{\xi})}{2\sigma_{\text{dec}}}\right)^2 - \left(\frac{h_{\text{max}} - \widehat{h}_{\text{max}}(\boldsymbol{\xi})}{2\sigma_{h_{\text{max}}}}\right)^2 . \end{aligned} \tag{26}$$

In Eq. (26), t_{arr}, t_{max}, t_{dec} and h_{max} are experimental estimates of the scaling factors at the considered location, and $\sigma^2_{t_{\text{arr}}}$, $\sigma^2_{t_{\text{max}}}$, $\sigma^2_{t_{\text{dec}}}$ and $\sigma^2_{h_{\text{max}}}$ are their associated variances. These quantities must be estimated from the data at each spatial location, and the log-likelihoods of the different locations are summed-up to form the full log-likelihood.

The experimental scaling factors are graphically estimated from the data at each location, as illustrated in Figure 9. The experimental values are selected as the best graphical estimates, with associated standard deviations reflecting their probable ranges, adjusted to account for the experimental noise and fluctuations. In the plots, the shaded areas are centered on the best estimates, with extents corresponding to the ±1 standard deviation uncertainty ranges on the experimental estimates. It is seen that the experimental arrival times are all well determined (with low variances), while the experimental decay times exhibit higher standard deviations, especially at the largest downstream location, where it cumulates significant uncertainty in the value of the maximum h_{max} and large fluctuations in the experimental data during the decay phase. The estimates of the scaling factors used for the inference problem and corresponding to the graphical representation in Figure 9 are listed in Table 2.



(a) $x = 2$ m (b) $x = 32$ m (c) $x = 66$ m

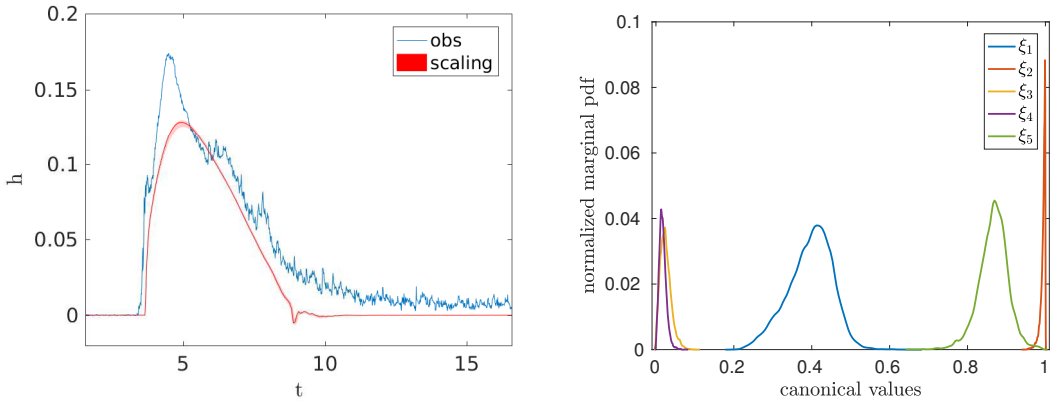
Figure 9: Estimation of the experimental scaling factors. The shaded areas are centered on the best estimates, while they extend over ±1 standard deviation range, horizontally for the scaling times and vertically for h_{max}.

$x = 2m$		$x = 32m$		$x = 66m$	
$t_{\text{arr}} = 0.9$	$\sigma_{t_{\text{arr}}} = 0.05$	$t_{\text{arr}} = 3.5$	$\sigma_{t_{\text{arr}}} = 0.05$	$t_{\text{arr}} = 6$	$\sigma_{t_{\text{arr}}} = 0.05$
$t_{\text{max}} = 1.85$	$\sigma_{t_{\text{max}}} = 0.05$	$t_{\text{max}} = 4.5$	$\sigma_{t_{\text{max}}} = 0.1$	$t_{\text{max}} = 8$	$\sigma_{t_{\text{max}}} = 1$
$t_{\text{dec}} = 3$	$\sigma_{t_{\text{dec}}} = 0.2$	$t_{\text{dec}} = 7.5$	$\sigma_{t_{\text{dec}}} = 0.5$	$t_{\text{dec}} = 12$	$\sigma_{t_{\text{dec}}} = 2$
$h_{\text{max}} = 0.61$	$\sigma_{h\text{max}} = 0.1$	$h_{\text{max}} = 0.17$	$\sigma_{h_{\text{max}}} = 0.04$	$h_{\text{max}} = 0.1$	$\sigma_{h_{\text{max}}} = 0.03$

Table 2: Estimates of the experimental scaling factors used for the inference problem using the log-likelihood function (26).

Figure 10(a) shows the posterior predictive range for the proposed likelihood together with the experimental data at $x = 32$ m. It is seen that the predictions are very close to the posterior prediction obtained with the smoothed

variances above, with comparable residuals norms for the two MAPs as measured using Eq. 25. Further, inspecting the marginal distributions of the canonical variables shows that, in the inference based on the scaling factors, the posterior distribution is slightly less concentrated around the MAP point but is still very tight, see Figure 10(b). Obviously, the experimental estimates considered in the present analysis can be questioned, regarding both the best and standard deviation values. However, further numerical experiments (not shown) with lower and higher variances have revealed that the outcome of the inference problem is robust with respect to the experimental estimates used. In particular the MAP point is only weakly affected by the variances affecting the experimental scaling factors, with mostly an impact on the spread of the posterior distribution.



(a) Posterior predictive range

(b) Marginal posteriors

Figure 10: Inference based on scaling factors: (a) Posterior predictive range for the inference based on the scaling factors, at location $x = 32$ m. (b) Marginal posterior distributions of the canonical random variables.

4.3. Inference using phases durations

The results of the previous section support the idea of extracting a few important features from the original data, with no significant loss of information and comparable posterior predictions, while being computationally much cheaper. In addition, the approach avoids requiring the proposal of a complex error model. On the other hand, the posterior prediction is still poorly fitting the experimental data with a very tight posterior predictive range. This is not so surprising as the a priori sensitivity analysis shown in Section 3.3.2 had revealed the limitation of the physical model, in particular when it comes to the arrival time prediction whereas the experimental data show little or no variability in t_{arr}; the numerical model is not capable of reproducing it for any value of the parameters in the a priori range. In fact, increasing the a priori range may not necessarily help in improving the model predictions regarding arrival times. This leads us to wonder whether the individual experimental measurements may have been synchronized before being averaged, with an arbitrary time shift as a result. In the absence of information regarding the experimental data treatment and processing, we opted to design a formulation of the inference problem that is independent to absolute time reference. Instead, we propose to base the inference on the durations of the two main phases of the debris flow dynamics, and not on their starting and ending times. For instance, we do not attempt to fit the arrival time and time at maximum, separately, but instead we consider the duration of the growth phase, spanning from the arrival to the maximum debris height. Similarly, we set the duration of the decay phase as the time span between the time at maximum and the time to decay, as defined previously. The definitions of these two characteristic durations are summarized in Table 3.

Having selected these two characteristic durations, we again assume a Gaussian error model with independent components, which leads to the following log-likelihood function at a spatial location:

$$\ln(\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi})) \propto -\left(\frac{T_{\text{grw}} - \widehat{T_{\text{grw}}}(\boldsymbol{\xi})}{2\sigma_{T_{\text{grw}}}}\right)^2 - \left(\frac{T_{\text{dec}} - \widehat{T_{\text{dec}}}(\boldsymbol{\xi})}{2\sigma_{T_{\text{dec}}}}\right)^2 - \left(\frac{h_{\text{max}} - \widehat{h}_{\text{max}}(\boldsymbol{\xi})}{2\sigma_{h_{\text{max}}}}\right)^2, \tag{27}$$

Phase name	*Symbol*	*Definition*
Growth duration	T_{grw}	$T_{\text{grw}} = t_{\max} - t_{\text{arr}}$
Decay duration	T_{dec}	$T_{\text{dec}} = t_{\max} - t_{\text{dec}}$

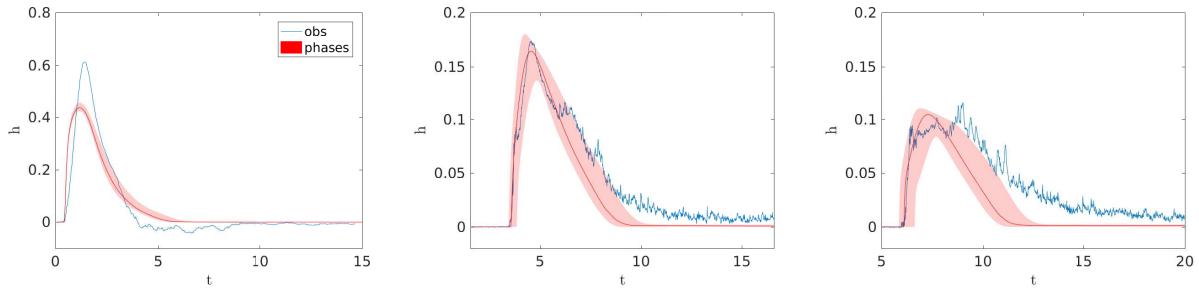
Table 3: Definitions of the characteristic durations of the debris flow dynamics.

where $\widehat{T_{\text{grw}}} = \widehat{t}_{\max} - \widehat{t}_{\text{arr}}$ and $\widehat{T}_{\text{dec}} = \widehat{t}_{\text{dec}} - \widehat{t}_{\max}$ are the surrogate models for the growth and decay durations, respectively. The best experimental estimates of T_{grw} and T_{dec} are computed from the estimates of t_{arr}, $t_{\max}$ and t_{dec} reported in Table 2, while their standard deviations $\sigma_{T_{\text{grw}}}$ and $\sigma_{T_{\text{dec}}}$ are estimated graphically from the experimental data, proceeding similarly as for the standard deviations of the experimental scaling factors. The estimates of $h_{\max}$ and $\sigma_{h_{\max}}$ are kept the same as in the previous case. Table 4 summarizes the experimental estimates used for the inference based on durations.

$x = 2$ m		$x = 32m$		$x = 66m$	
$T_{\text{grw}} = 0.95$	$\sigma_{T_{\text{grw}}} = 0.1$	$T_{\text{grw}} = 1$	$\sigma_{T_{\text{grw}}} = 0.25$	$T_{\text{grw}} = 2$	$\sigma_{T_{\text{grw}}} = 1$
$T_{\text{dec}} = 1.15$	$\sigma_{T_{\text{dec}}} = 0.25$	$T_{\text{dec}} = 3$	$\sigma_{T_{\text{dec}}} = 0.6$	$T_{\text{dec}} = 4$	$\sigma_{T_{\text{dec}}} = 2$
$h_{\max} = 0.61$	$\sigma_{h_{\max}} = 0.1$	$h_{\max} = 0.17$	$\sigma_{h_{\max}} = 0.04$	$h_{\max} = 0.1$	$\sigma_{h_{\max}} = 0.03$

Table 4: Experimental estimates of the characteristic durations of the debris flow.

Figure 11 plots the posterior predictive ranges at the three locations, as red shaded areas, corresponding to inference using durations. The solid red line inside the shaded area corresponds to the posterior expected predictions of $h(t)$. Since the estimation is time-independent, the experimental data have been shifted in time such that the arrival times of both the predictions and the data match. A striking difference with the inference results presented above is the much larger posterior predictive range obtained in the present case. Indeed, although the a priori predictive ranges (see Figure 6) have been noticeably reduced, the posterior predictions of the debris height are now significantly uncertain. The overlap of the posterior prediction with the experimental data is also significantly improved, compared to the cases of the previous likelihood considered. Besides the impact on the spread of the posterior distribution, the inference based on durations is also seen to drastically improve the prediction of the maximum debris height, which was systematically underestimated for the other likelihoods tested previously (see Figures 7 and Figure 10). Note also that the MAP prediction (not shown) is not contained within the posterior predictive range as defined by the 5th to 95th percentiles.



(a) $x = 2m$ (b) $x = 32m$ (c) $x = 66m$

Figure 11: Inference based on durations: Posterior predictive ranges for the inference based on durations. The experimental data (labeled Obs) are also reported for comparison, and have been shifted in time to match the expected arrival times.

To better appreciate the impact of the proposed formulation of the inference problem on its outcome, we present in Figure 12 the marginal posterior distributions of the canonical random variables. In contrast to the previous cases, the marginals are not concentrated anymore and exhibits several local maxima. Note that the canonical random

variable ξ_1 presents a posterior marginal that differs the most from its prior distribution with essentially no posterior probability for values lower than 0.2. This is consistent with the findings of the a priori sensitivity analysis shown in Section 3.3.2, since ξ_1 parametrizes m_{crit} that was found to be responsible for most of the variability in the model output. The other canonical random variable associated to the other model parameters exhibit in contrast much flatter posterior marginals reflecting that the inference process has not resulted in a significant update of their prior. In our opinion, this result is much more satisfying than the previous ones and is consistent with the a priori GSA where the parameter values had little impact on the prediction.

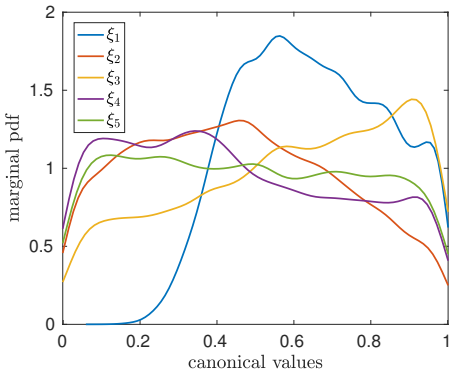
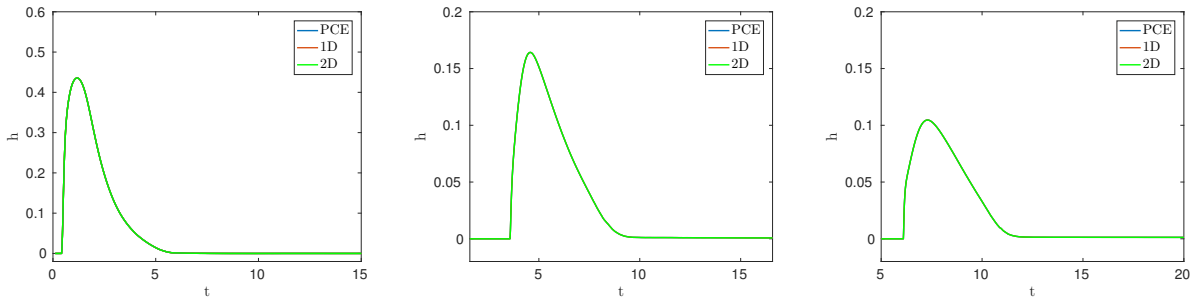


Figure 12: Marginal posterior distributions of the canonical random variables.

As a final check, we assess the quality of the surrogate model for the posterior expected value of the parameters. To this end, we compare in Figure 13 the predictions of the surrogate model (PCE) at the three measurement locations with the predictions of the detailed D-Claw predictions using the one-dimensional approximation (1D) and the original two-dimensional problem (2D). It is seen that at the three locations all models are in excellent agreement, confirming the validity of the approach.



(a) $x = 2m$ (b) $x = 32m$ (c) $x = 66m$

Figure 13: Comparison of the surrogate model (PCE) and one-dimensional (1D) and two-dimensional (2D) D-Claw model prediction for the posterior expected value of the parameters. Shown are the evolution in time of the debris height at the three measurement locations.

5. Summary and conclusions

We developed a computational approach to propagate uncertainty in the unknown parameters of the D-CLAW debris-flow model [23, 13] and calibrated these parameters based on experimental data reported in [44]. The methodology consisted of establishing an approximation of the quantities of interest (QoIs) involved in the calibration using polynomial chaos (PC) expansions, and using the resulting representations to accelerate the Bayesian calibration

process. Due to the lack of sufficient information regarding the treatment and preprocessing of the available measurements, as well as the complexity of the solution for certain ranges of the uncertain parameters, the calibration was primarily based on relevant features predicted by the model.

To build the model approximation, we considered a preconditioned non-intrusive approach, and relied on regression methods to compute the expansion coefficients. Due to the appearance of transverse waves and debris fragmentation, a parametric study was first conducted in order to refine the scope of the simulations. This led us to consider one-dimensional computations and to restrict the prior range of the hydraulic permeability. We analyzed the resulting simplified model, and we verified that the restrictions resulting in efficient calculations and smoother signals, but did not otherwise impact the broad features of the solution. This highlighted the importance of defining suitable priors, and of carefully selecting smooth QoIs. We also analyzed the quality of the surrogate models in terms of L_2 metrics, and showed that for the present applications, regardless of the regression method used, the preconditioned PC representation leads to more accurate representation than direct PC representation of the same QoI. The analysis also showed that the use of LAD regression produces a significant improvement when comparing the errors in terms of the L_∞ norm.

Before performing the inference, we exploited the PC surrogates to quantify the sensitivity of the selected QoIs to the uncertain parameters. In particular, we estimated the a priori predictive range and performed a Global Sensitivity Analysis (GSA). Whereas for the present setup the predictive ranges highlighted significant discrepancies between the model predictions and the measurements, the GSA showed that most of the variability in the flow thickness is due to the static critical-state solid volume fraction. We have also demonstrated the usefulness of this information, namely by using it as an indicator of the quality of the parameter calibration.

We first performed the calibration by using the reported experimental data directly in a Bayesian inverse setting. Both the originally reported and smoothed variances were considered in this initial exercise. Analysis of the results showed that this approach led to overly confident estimates in the value of the parameters. This was not surprising in light of the results of the sensitivity analysis, which revealed large discrepancies between the model predictions and the reported data. The analysis also showed that a poor discrepancy model was responsible for the observed over-determination of the parameters.

To mitigate this difficulty, we focused on performing the calibration based on the important features predicted by the model. Specifically, we considered a likelihood based on arrival time, time at maximum, debris height at maximum and time to decay. For the present setup, the results indicated that the feature-based inference approach is computationally cheaper than the direct approach, that it does not involve significant loss of information and avoids the need to define a complex error model. However, we observed that whereas the predictions were consistent with those obtained using the direct approach, the posterior marginal distributions of the canonical random variables used to parametrize the uncertain inputs were still very tight.

Further analysis of the predictions led us to postulate the occurrence of time synchronization issues between the numerical model and the experimental data, which prevents us from correctly capturing the arrival time and its variability. Consequently, we explored the possibility of performing the inference based only on ("time-independent") integral features of the model. Alternatively, this formulation can be interpreted as the introduction of an improved error model, permitting a constant time-shift between the predictions and the observations, following a non-informative prior. This methodology resulted in two main differences with respect to previous approaches: a higher posterior variability was observed, as well as more accurate predictions of the arrival time of the peak wave height. In addition, the predictions were also consistent with the results of the GSA, indicating a more reliable calibration process.

References

[1] A. Alexanderian, O. P. Le Maître, H.N. Najm, M. Iskandarani, and O. Knio. Multiscale stochastic preconditioners in non-intrusive spectral projection. *Journal of Scientific Computing*, 50(2):306–340, 2012.

[2] A. Alexanderian, F. Rizzi, M. Rathinam, O.P. Le Maître, and O. Knio. Preconditioned bayesian regression for stochastic chemical kinetics. *Journal of Scientific Computing*, 58(3):592–626, 2014.

[3] H. L. Anderson. *Metropolis, Monte Carlo and the MANIAC*. Los Alamos Science, 1986.

[4] M. Berveiller. *Stochastic finite elements: intrusive and non intrusive methods for reliability analysis*. PhD thesis, Universite Blaise Pascal, Clermont-Ferrand, 2005.

[5] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control. 3rd ed.* Englewood Cliffs, NJ: Prentice Hall,, 1994.

[6] R. E. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta Numerica*, pages 1–49, 1998.

[7] R. H. Cameron and W. T. Martin. The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals. *Annals of Mathematics*, 48:385–392, 1947.

[8] C. Canuto, M. Y. Hussaini, A. Quateroni, and T. A. Zang. *Spectral methods : Fundamentals in single domain*. Springer, Berlin, Heidelberg, 2006.

[9] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572, 1991.

[10] T. Crestaux, O. P. Le Maître, and J. M. Martinez. Polynomial chaos expansion of sensitivity analysis. *J. Rel. Eng. Syst. Safety*, 94(7):1161–1182, 2009.

[11] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, third edition edition, 2013.

[12] D. L. George. Flume problems. https://github.com/dlgeorge/flume/tree/master/GateRelease1/, 2016.

[13] D. L. George and R. M. Iverson. A depth-averaged debris-flow model that includes the effects of evolving dilatancy. ii. numerical predictions and experimental tests. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 470(2170), 2014.

[14] D. L. George and R. M. Iverson. A depth-averaged debris-flow model that includes the effects of evolving dilatancy. II. Numerical predictions and experimental tests. *Proc. R. Soc. A*, 470(2170):20130820, October 2014.

[15] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A spectral approach*. Springer New York, 1991.

[16] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[17] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[18] Jerrad Hampton and Alireza Doostan. Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression. *Computer Methods in Applied Mechanics and Engineering*, 290:73 – 97, 2015.

[19] G. Host. *Simulated Annealing - Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2014.

[20] S. Isukapalli. *Uncertainty analysis of transport-transformation models*. PhD thesis, The State University of New Jersey, 1999.

[21] R. M. Iverson. The physics of debris flows. *Rev. Geophys.*, 35(3):245–296, 1997.

[22] R. M. Iverson. Regulation of landslide motion by dilatancy and pore pressure feedback. *J. Geophys. Res.*, 110(F2):F02015, 2005.

[23] R. M. Iverson and D. L. George. A depth-averaged debris-flow model that includes the effects of evolving dilatancy. i. physical basis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 470(2170), 2014.

[24] R. M. Iverson and D. L. George. A depth-averaged debris-flow model that includes the effects of evolving dilatancy. I. Physical basis. *Proc. R. Soc. A*, 470(2170):20130819, October 2014.

[25] M. J. W. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117:35–43, 1999.

[26] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *J. R. Statistical Society B*, 63:425 – 464, 2001.

[27] J. V. Langenhove, D. Lucor, and A. Belme. Robust uncertainty quantification using preconditioned least-squares polynomial approximations with l1-regularizations. *International Journal for Uncertainty Quantification*, 6:57–77, 2016.

[28] J. O. Langseth and R. J. LeVeque. A wave-propagation method for three-dimensional hyperbolic conservation laws. *J. Comput. Phys.*, 165:126–166, 2000.

[29] C. L. Lawson. *Contribution to the Theory of Linear Least Maximum Approximations*. PhD thesis, University of California, 1961.

[30] O. P. Le Maître and O .M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, New York, NY, 2010.

[31] O. P. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. A stochastic projection method for fluid flow. II. Random process. *Journal of Computational Phyics*, 181:9–44, 2002.

[32] R. J. LeVeque. High-resolution conservative algorithms for advection in incompressible flow. *SIAM J. Numer. Anal.*, 33:627–665, 1996.

[33] R. J. LeVeque. Wave propagation algorithms for multi-dimensional hyperbolic systems. *J. Comput. Phys.*, 131:327–353, 1997.

[34] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.

[35] Wei-Liem Loh. On latin hypercube sampling. *Annals of Statistics*, 24(5):2058–2080, 1996.

[36] N. Madras. *Lectures on Monte Carlo methods*. American Mathematical Society, Providence, RI, 2001.

[37] C. V. Mai and B. Sudret. Surrogate models for oscillatory systems using sparse polynomial chaos expansions and stochastic time warping. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):540–571, 2017.

[38] Kyle T Mandli, Aron J Ahmadia, Marsha Berger, Donna Calhoun, David L George, Yiannis Hadjimichael, David I Ketcheson, Grady I Lemoine, and Randall J. LeVeque. Clawpack: building an open source ecosystem for solving hyperbolic PDEs. *PeerJ Computer Science*, 2(3):e68, August 2016.

[39] Y. M. Marzouk, H. N. Najm, and L. A. Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560 – 586, 2007.

[40] Youssef M. Marzouk and Habib N. Najm. Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, 2009.

[41] Matlab optimization toolbox, 2016a. The MathWorks, Natick, MA, USA.

[42] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[43] E. Polak. *Optimization: Algorithms and Consistent Approximations*. Applied mathematical sciences, 9780387949710, Springer-Verlag, 1997.

[44] Richard R. M. Iverson, M. Logan, R. LaHusen, and M. Berti. The perfect debris flow? aggregated results from 28 large-scale experiments. *Journal of Geophysical Research: Earth Surface*, 115(F3), 2010.

[45] M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132:545–555, 2003.

[46] P. W. Rowe. The stress-dilatancy relation for static equilibrium of an assembly of particles in contact. *Proc. Roy. Soc., London*, A269:500–527, 1962.

[47] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280 – 297, 2002.

[48] J. A. Scales and A. Gersztenkorn. Robust methods in inverse theory. *Inverse Problems*, 4(4):1071, 1988.

[49] A. N. Schofield and C. P. Wroth. *Critical State Soil Mechanics*. McGraw Hill, 1968.

[50] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.
[51] I.M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling & Comput. Exp.*, 1:407–414, 1993.
[52] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.
[53] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005.
[54] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.