

# Comparative analysis of the structures and outcomes of geophysical flow models and modeling assumptions using uncertainty quantification

Abani K. Patra<sup>1,3</sup>, Andrea Bevilacqua<sup>2</sup>, Ali Akhavan-Safaei<sup>1</sup>, E. Bruce Pitman<sup>4</sup>, Marcus I. Bursik<sup>2</sup>, and David Hyman<sup>2</sup>

<sup>1</sup>*Dept. of Mech. and Aero. Eng., University at Buffalo, Buffalo NY 14260*

<sup>2</sup>*Dept. of Earth Sciences, University at Buffalo, Buffalo NY 14260*

<sup>3</sup>*Comp. Data Science and Eng., University at Buffalo, Buffalo NY 14260*

<sup>4</sup>*Dept. of Materials Design and Innovation, University at Buffalo, NY, 14260*

{abani, abevilac, aliakhav, pitman, mib, davidhym}@buffalo.edu

## Contents

<b>1 Models and assumptions</b>	<b>3</b>
<b>2 Analysis of modeling assumptions and models</b>	<b>4</b>
2.1 Analysis Process . . . . .	4
2.2 Monte Carlo Process and Statistical Analysis . . . . .	5
<b>3 Modeling of geophysical mass flows</b>	<b>5</b>
3.1 Overview of the models . . . . .	6
3.1.1 Mohr-Coulomb . . . . .	6
3.1.2 Pouliquen-Forterre . . . . .	7
3.1.3 Voellmy-Salm . . . . .	8
3.2 Latent variables . . . . .	9
<b>4 Small scale flow on inclined plane and flat runway</b>	<b>10</b>
4.1 Preliminary consistency testing of the input ranges . . . . .	10
4.2 Observable outputs . . . . .	11
4.2.1 Flow height . . . . .	11
4.2.2 Flow acceleration . . . . .	12
4.3 Statistical analysis of latent variables . . . . .	13
4.4 Flow extent and spatial integrals . . . . .	15
4.5 Power integrals . . . . .	16
<b>5 Large scale flow on the SW slope of Volcán de Colima</b>	<b>16</b>
5.1 Preliminary consistency testing of the input ranges . . . . .	18
5.2 Exploring Flow Limits . . . . .	19
5.3 Observable outputs . . . . .	20
5.3.1 Flow height in six locations . . . . .	21
5.4 Statistical analysis of latent variables . . . . .	23

5.4.1	Three locations proximal to the initial pile . . . . .	23
5.4.2	Three locations distal from the initial pile . . . . .	24
5.5	Flow extent and spatial integrals . . . . .	26
5.6	Power integrals . . . . .	27
<b>6</b>	<b>Discussion on the comparative performance of geophysical flow models and latent variables</b>	<b>28</b>
6.1	Characteristic features and their motivations . . . . .	28
6.1.1	Flow height and acceleration . . . . .	29
6.1.2	Statistical analysis of latent variables . . . . .	29
6.1.3	Flow extent and spatial integrals . . . . .	30
6.1.4	Power integrals . . . . .	30
6.2	Example of model performance calculation . . . . .	30
<b>7</b>	<b>Conclusions</b>	<b>32</b>
<b>A</b>	<b>Latin Hypercubes and orthogonal arrays</b>	<b>33</b>
<b>B</b>	<b>Conditional decomposition of expected contributions</b>	<b>34</b>

### Abstract

We develop a new statistically driven method for analyzing the modeling of geophysical flows. Many models have been advocated by different modelers for such flows incorporating different modeling assumptions. Limited and sparse observational data on the modeled phenomena usually does not permit a clean discrimination among models for fitness of purpose, and, heuristic choices are usually made, especially for critical predictions of behavior that has not been experienced. We advocate here a methodology for characterizing models and the modeling assumptions they represent using a statistical approach over the full range of applicability of the models. Such a characterization may then be used to decide the appropriateness of a model for use, and, perhaps as needed compositions of models for better predictive power. We present our method by comparing three different models arising from different rheology assumptions, and the data show unambiguously the performance of the models across a wide range of possible flow regimes. This comparison is facilitated by the recent development of the new release of our TITAN2D mass flow code that allows choice of multiple rheologies. The analysis of latent variables in the models is particularly illustrative of the impact of modeling assumptions. Knowledge of which assumptions dominate, and, by how much, is illustrated in two different case studies: a small scale inclined plane with a flat runway, and the large scale topography on the SW slope of Volcán de Colima (MX).

# 1 Models and assumptions

This paper presents a systematic approach to the study of models of complex systems, motivated by and applied to the case of geophysical mass flow models. (Gilbert, 1991) defines science as a process of constructing predictive conceptual models where these models represent consistent predictive relations in target systems. A simpler though not necessarily comprehensive definition of a model that is more appropriate for the context of this study is that:

*A model is a representation of a postulated relationship among inputs and outputs of a system, usually informed by observation and based on a hypothesis that best explains the relationship.*

The definition captures two of the most important characteristics

- models depend on a *hypothesis*, and,
- models use the *data from observation* to validate and refine the hypothesis.

Errors and uncertainty in the data and limitations in the hypothesis (usually a tractable and computable mathematical construct articulating beliefs like proportionality, linearity etc.) are immediate challenges that must be overcome, to construct useful and credible models.

A model is most useful in predicting the behavior of a system for unobserved inputs, and, interpreting or explaining of the system's behavior. Since, models require a hypotheses, it follows that the model is a formulation of a belief about the data. The immediate consequence of this is that the model may not be reliable about such prediction, since, the *subjectivity of the belief* can never be completely eliminated (Kennedy and O'Hagan, 2011; Higdon et al., 2004) even when sufficient care is taken to use all the available data and information. Secondly, the data at hand may not provide enough information about the system to characterize its behavior at the desired prediction. This *data inadequacy* is rarely characterized even for verified and validated models. What makes this problem even more acute is that we are often interested in modeling outcomes that are not observed and often not observable.

The consequence of this lack of knowledge and limited data is the multiplicity of beliefs about the complex system being modeled and a profusion of models based on different modeling assumptions and data use. These competing models lead to much debate among scientists. Principles like "Occam's razor" and Bayesian statistics (Farrell et al., 2015) provide some guidance but simple robust approaches that allow the testing of models for fitness need to be developed. We present in this paper a simple data driven approach to discriminate among models and the modeling assumptions implicit in each model, given a range of phenomena to be studied. We illustrate the approach by work on geophysical mass flows.

An assumption is a simple concept – any atomic postulate about relationships among quantities under study. Models are compositions of many such assumptions. The study of models is, thus, implicitly a study of these assumptions and their composability and applicability in a particular context. Sometimes a good model contains a useless assumption that may be removed, sometimes a good assumption could be added to a different model - these are usually subjective choices, not data driven. Moreover, the correct assumptions may change through time, making *model choice* more difficult.

The rest of the paper will define our approach and illustrate using three models for large scale mass flows incorporated in our large scale mass flow simulation framework TITAN2D (Patra et al., 2005, 2006; Yu et al., 2009; Aghakhani et al., 2016). So far, TITAN2D has been successfully applied to the simulation of different geophysical mass flows with specific characteristics (Sheridan et al., 2005; Rupp et al., 2006; Norini et al., 2009; Charbonnier and Gertisser, 2009; Procter et al., 2010; Sheridan et al., 2010; Sulpizio et al., 2010; Capra et al., 2011). Several studies involving TITAN2D were also directed towards a statistical study of geophysical flows, focusing on uncertainty quantification (Dalbey et al., 2008; Dalbey, 2009; Stefanescu et al., 2012a,b), or on the more efficient production of hazard maps (Bayarri et al., 2009; Spiller et al., 2014; Bayarri et al., 2015; Ogburn et al., 2016).

The 4<sup>th</sup> release of TITAN2D <sup>1</sup> offers multiple rheology options in the same code base. The availability of three distinct models for similar phenomena in the same tool provides us the ability to directly compare outputs and internal variables in all the three models and control for usually difficult to quantify effects like numerical solution procedures, input ranges and computer hardware. Given a particular problem for which predictive analysis is planned, the information generated here may be used to guide model choice and/or processes for integrating information from multiple models.

---

<sup>1</sup>available from vhub.org

## 2 Analysis of modeling assumptions and models

### 2.1 Analysis Process

Let us define  $(M(A), P_{M(A)})$ , where  $A$  is a set of assumptions,  $M(A)$  is the model which combines those assumptions, and  $P_M$  is a probability distribution in the parameter space of  $M$ . While the support of  $P_M$  can be restricted to a single value by solving an inverse problem for the optimal reconstruction of a particular flow, this is not desirable if we are interested in the general predictive capabilities of the model, where we are interested in the outcomes over a whole range.

**Stage 1: Parameter Ranges** In this study, we assume:

$$P_M(p_1, \dots, p_{N_M}) \sim \bigotimes_{i=1}^{N_M} \text{Unif}(a_{i,M}, b_{i,M}),$$

where  $N_M$  is the number of parameters of  $M$ . This is not restrictive, and in case of correlation between the commonly used parameters  $(\hat{p}_j)_{j=1, \hat{N}_M}$ , or non-uniform distributions, we can always define a function  $g$  such that  $g[(p_i)] = (\hat{p}_j)$ , and the  $(p_i)$  are independent and uniformly distributed. In particular, we choose these parameter ranges using information gathered from the literature about the physical meaning of those values, together with a preliminary testing for physical consistency of model outcomes and range of inputs/outcomes of interest. This step is critical, because if the total friction of the models does not cover a similar span, then the statistical comparison is dominated by trivial macroscopic differences, and cannot focus on the rheology details. In the preparation of hazard analysis, expert elicitation processes can be used to ensure that the studies correctly account for all anticipated and possible flow regimes.

**Stage 2 Simulations and Data Gathering** The simulation algorithms can be represented as:

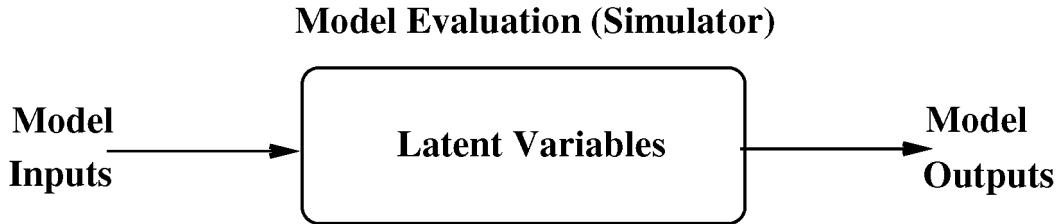


Figure 1: Models and variables

The *model inputs* are the parameters of  $M$ , the *latent variables* include quantities in the model evaluation that are ascribable to specific assumptions  $A_i$ . These are usually not observed as outputs from the model. For example in momentum balances of complex flow calculations these could be values of different source terms, dissipation terms and inertia terms. Finally, the *model outputs* include explicit outcomes e.g., for flow calculations these could be flow height, lateral extent, area, velocity, acceleration, and derived quantities such as Froude number  $Fr$ . In general, we use a Monte Carlo simulation, sampling the model inputs and obtaining a family of graphs plotting the expectation of the latent variables and model outputs. We also include their 5<sup>th</sup> and 95<sup>th</sup> percentiles. Our sampling technique of the input variables is based on the Latin Hypercube Sampling (LHS) idea, and in particular, on the improved space-filling properties of the orthogonal array-based Latin Hypercubes (see Appendix A). The volume of data generated is likely to be large but modern computing and data handling equipment readily available to most modeling researchers <sup>2</sup> in university and national research facilities are more than adequate.

**Stage 3 Results Analysis** These and other statistics can now be compared to determine the need for different modeling assumptions and the relative merits of different models. Thus, analysis of the data gathered over the entire range of flows for the state variables and outcomes leads to a quantitative basis for accepting or rejecting particular assumptions or models for specific outcomes.

---

<sup>2</sup>We thank the University at Buffalo Center for Computing Research

## 2.2 Monte Carlo Process and Statistical Analysis

In our study, the flow range is defined by establishing boundaries for inputs like flow volume or rheology coefficients characterizing the models. Latin Hypercube Sampling is performed over  $[0, 1]^3$  for the MC and VS input parameters, and  $[0, 1]^4$  for PF input parameters (see Appendix A). Those dimensionless samples are linearly mapped to fill the required intervals. Figure 11a,b,c displays examples of LHS design in the three models, with respect to their commonly used parameters.

Following the simulations, we generate data for each sample run and each outcome and latent variable  $f(\underline{\mathbf{x}}, t)$  calculated as a function of time on the elements of the computational grid. This analysis generates tremendous volume of data which must then be analyzed using statistical methods for summative impact. The latent variables in this case are the mass and force terms in the conservation laws defined above.

We devise many statistical measures for analyzing the data. For instance, let  $(F_i(\underline{\mathbf{x}}, t))_{i=1,\dots,4}$  be an array of force terms, where  $\underline{\mathbf{x}} \in \mathbf{R}^d$  is a spatial location, and  $t \in T$  is a time instant. The degree of contribution of those force terms to the flow dynamics can be significantly variable in space and time, and we define the *dominance factors*  $[p_j(\underline{\mathbf{x}}, t)]_{j=1,\dots,k}$ , i.e., the probability of each  $F_j(\underline{\mathbf{x}}, t)$  to be the dominant force. Those probabilities provide insight into the dominance of a particular source or dissipation term on the model dynamics. Each term is identified with a particular modeling assumption. We remark that we focus on the modulus of the forces and hence we cope with scalar terms. It is also important to remark that all the forces depend on the input variables, and they can be thus considered as random variables. Furthermore, these definitions are general and could be applied to any set of latent variables, and not only to the force terms.

**Definition 1 (Dominance factors)** *Let  $(F_i)_{i \in I}$  be random variables on  $(\Omega, \mathcal{F}, P_M)$ . Then,  $\forall i$ , the dominant variable is defined as:*

$$\Phi := \begin{cases} \max_i |F_i|, & \text{if not null;} \\ 1, & \text{otherwise.} \end{cases}$$

*In particular, for each  $j \in I$ , the dominance factors are defined as:*

$$p_j := P_M \{ \Phi = |F_j| \}.$$

Moreover, we define the *random contributions*, an additional tool that we use to compare the different force terms, following a less restrictive approach than the dominance factors. They are obtained dividing the force terms by the dominant force  $\Phi$ , and hence belong to  $[0, 1]$ .

**Definition 2 (Expected contributions)** *Let  $(F_i)_{i \in I}$  be random variables on  $(\Omega, \mathcal{F}, P_M)$ . Then,  $\forall i$ , the random contribution is defined as:*

$$C_i := \frac{F_i}{\Phi},$$

*where  $\Phi$  is the dominant variable. Thus,  $\forall i$ , the expected contributions are defined by  $E[C_i]$ .*

In particular, for a particular location  $x$ , time  $t$ , and parameter sample  $\omega$ , we have  $C_i(\underline{\mathbf{x}}, t, \omega) = 0$  if there is no flow or all the forces are null. The expectation of  $C_i$  is reduced by the chance of  $F_i$  being small compared to the other terms, or by the chance of having no flow in  $(\underline{\mathbf{x}}, t)$ . The meaning of the random variable  $\Phi(\underline{\mathbf{x}}, t)$  is explained by the dominance factors, and hence they can be used to define a further statistical decomposition of the random contributions, as detailed in the Appendix B.

## 3 Modeling of geophysical mass flows

Dense large scale granular avalanches are a complex class of flows with physics that has often been poorly captured by models that are computationally tractable. Sparsity of actual flow data (usually only a posteriori deposit information is available), and large uncertainty in the mechanisms of initiation and flow propagation, make the modeling task challenging, and a subject of much continuing interest. Models that appear to represent the physics well in certain flows, may turn out to be poorly behaved in others, due to intrinsic physical, mathematical or numerical issues. Nevertheless, given the large implications on life and property, many models with different modeling assumptions have been proposed. For example in (Iverson, 1997; Iverson and Denlinger, 2001; Denlinger and Iverson, 2001; Pitman et al., 2003b; Denlinger and Iverson, 2004; Iverson et al., 2004), the depth-averaged model was applied in the simulation of test

geophysical flows in large scale experiments. Several studies were specifically devoted to the modeling of volcanic mass flows (Bursik et al., 2005; Kelfoun and Druitt, 2005; Charbonnier and Gertisser, 2009; Kelfoun et al., 2009; Procter et al., 2010; Kelfoun, 2011; Charbonnier et al., 2013). In fact, volcanoes are great sources for a rich variety of geophysical flow types and provide field data from past flow events.

Modeling in this case proceeds by first assuming that the laws of mass and momentum conservation hold for properly defined system boundaries. The scale of these flows – very long and wide with small depth led to the first most generally accepted assumption – shallowness (Savage and Hutter, 1989). This allows an integration through the depth to obtain simpler and more computationally tractable equations. This is the next of many assumptions that have to be made. Both of these are fundamental assumptions which can be tested in the procedure we established above. Since, there is a general consensus and much evidence in the literature of the validity of these assumptions we defer analysis of these to future work.

The depth-averaged Saint-Venant equations that result are:

$$\begin{aligned} \frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(h\bar{u}) + \frac{\partial}{\partial y}(h\bar{v}) &= 0 \\ \frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}\left(h\bar{u}^2 + \frac{1}{2}kg_zh^2\right) + \frac{\partial}{\partial y}(h\bar{u}\bar{v}) &= S_x \\ \frac{\partial}{\partial t}(h\bar{v}) + \frac{\partial}{\partial x}(h\bar{u}\bar{v}) + \frac{\partial}{\partial y}\left(h\bar{v}^2 + \frac{1}{2}kg_zh^2\right) &= S_y \end{aligned} \quad (1)$$

Here the Cartesian coordinate system is aligned such that  $z$  is normal to the surface;  $h$  is the flow height in the  $z$  direction;  $h\bar{u}$  and  $h\bar{v}$  are respectively the components of momentum in the  $x$  and  $y$  directions; and  $k$  is the coefficient which relates the lateral stress components,  $\bar{\sigma}_{xx}$  and  $\bar{\sigma}_{yy}$ , to the normal stress component,  $\bar{\sigma}_{zz}$ . The definition of this coefficient depends on the constitutive model of the flowing material we choose. Note that  $\frac{1}{2}kg_zh^2$  is the contribution of hydrostatic pressure to the momentum fluxes.  $S_x$  and  $S_y$  are the sum local stresses: they include the gravitational driving forces, the basal friction force resisting to the motion of the material, and additional forces specific of rheology assumptions.

The final class of assumptions are the assumptions on the rheology of the flows – in particular in this context assumptions used to model different dissipation mechanisms embedded in  $S_x, S_y$  that lead to a plethora of models with much controversy on the most suitable model. We focus here on three models derived from different assumptions for essentially the same class of flows.

### 3.1 Overview of the models

In the three following sections, we briefly describe *Mohr-Coulomb* (MC), *Pouliquen-Forsterre* (PF) and *Voellmy-Salm* (VS) models. Models based on additional heterogeneous assumptions are possible, either more complex (Pitman and Le, 2005; Iverson and George, 2014) or more simple (Dade and Huppert, 1998). We decided to focus on these three because of their historical relevance. Moreover, if the degree of complexity in the models is significantly different, model comparison should take into account that, but this is outside the purpose of this study (e.g. (Farrell et al., 2015)).

#### 3.1.1 Mohr-Coulomb

Based on the long history of studies in soil mechanics (Rankine, 1857; Jaeger et al., 1989), the Mohr-Coulomb rheology (MC) was developed and used to represent the behavior of geophysical mass flows (Savage and Hutter, 1989).

Shear and normal stress are assumed to obey Coulomb friction equation, both within the flow and at its boundaries. In other words,

$$\tau = \sigma \tan \phi, \quad (2)$$

where  $\tau$  and  $\sigma$  are respectively the shear and normal stresses on failure surfaces, and  $\phi$  is a friction angle. This relationship does not depend on the flow speed.

We can summarize the MC rheology assumptions as:

- *Basal Friction* based on a constant friction angle.
- *Internal Friction* based on a constant friction angle.
- *Earth pressure coefficient* formula depends on the Mohr circle.

- Velocity based *curvature effects* are included into the equations.

Under the assumption of symmetry of the stress tensor with respect to the  $z$  axis, the earth pressure coefficient  $k = k_{ap}$  can take on only one of three values  $\{0, \pm 1\}$ . The material yield criterion is represented by the two straight lines at angles  $\pm\phi$  (the internal friction angle) relative to horizontal direction. Similarly, the normal and shear stress at the bed are represented by the line  $\tau = -\sigma \tan(\delta)$  where  $\delta$  is the bed friction angle.

**MC equations** As a result, we can write down the source terms of the Eqs. (1):

$$\begin{aligned} S_x &= g_x h - \frac{\bar{u}}{\|\bar{\mathbf{u}}\|} \left[ h \left( g_z + \frac{\bar{u}^2}{r_x} \right) \tan(\phi_{bed}) \right] - h k_{ap} \operatorname{sgn} \left( \frac{\partial \bar{u}}{\partial y} \right) \frac{\partial(g_z h)}{\partial y} \sin(\phi_{int}) \\ S_y &= g_y h - \frac{\bar{v}}{\|\bar{\mathbf{u}}\|} \left[ h \left( g_z + \frac{\bar{v}^2}{r_y} \right) \tan(\phi_{bed}) \right] - h k_{ap} \operatorname{sgn} \left( \frac{\partial \bar{v}}{\partial x} \right) \frac{\partial(g_z h)}{\partial x} \sin(\phi_{int}) \end{aligned} \quad (3)$$

Where,  $\bar{\mathbf{u}} = (\bar{u}, \bar{v})$ , is the depth-averaged velocity vector,  $r_x$  and  $r_y$  denote the radii of curvature of the local basal surface. The inverse of the radii of curvature is usually approximated with the partial derivatives of the basal slope, e.g.,  $1/r_x = \partial \theta_x / \partial x$ , where  $\theta_x$  is the local bed slope.

In our study, sampled input parameters are  $\phi_{bed}$ , and  $\Delta\phi := \phi_{int} - \phi_{bed}$ . In particular, the range of  $\phi_{bed}$  depends on the case study, while  $\Delta\phi \in [2^\circ, 10^\circ]$  (Dalbey et al., 2008).

### 3.1.2 Pouliquen-Forterre

The scaling properties for granular flows down rough inclined planes led to the development of the Pouliquen-Forterre rheology (PF), assuming a variable frictional behavior as a function of Froude Number and flow depth (Pouliquen, 1999; Forterre and Pouliquen, 2002; Pouliquen and Forterre, 2002; Forterre and Pouliquen, 2003).

PF rheology assumptions can be summarized as:

- *Basal Friction* is based on an interpolation of two different friction angles, based on the flow regime and depth.
- *Internal Friction* is neglected.
- *Earth pressure coefficient* is equal to one.
- Normal stress is modified by a *hydrostatic pressure force* related to the flow height gradient.
- Velocity based *curvature effects* are included into the equations.

Two critical slope inclination angles are defined as functions of the flow thickness, namely  $\phi_{start}(h)$  and  $\phi_{stop}(h)$ . The function  $\phi_{stop}(h)$  gives the slope angle at which a steady uniform flow leaves a deposit of thickness  $h$ , while  $\phi_{start}(h)$  is the angle at which a layer of thickness  $h$  is mobilized. They define two different basal friction coefficients.

$$\mu_{start}(h) = \tan(\phi_{start}(h)) \quad (4)$$

$$\mu_{stop}(h) = \tan(\phi_{stop}(h)) \quad (5)$$

An empirical friction law  $\mu_b(\|\bar{\mathbf{u}}\|, h)$  is then defined in the whole range of velocity and thickness. The expression changes depending on two flow regimes, according to a parameter  $\beta$  and the Froude number  $Fr = \|\bar{\mathbf{u}}\| / \sqrt{hg_z}$ .

**Dynamic friction regime -  $Fr \geq \beta$**

$$\mu(h, Fr) = \mu_{stop}(h\beta/Fr) \quad (6)$$

**Intermediate friction regime -**  $0 \leq Fr < \beta$

$$\mu(h, Fr) = \left(\frac{Fr}{\beta}\right)^{\gamma} [\mu_{stop}(h) - \mu_{start}(h)] + \mu_{start}(h), \quad (7)$$

where  $\gamma$  is the power of extrapolation, assumed equal to  $10^{-3}$  in the sequel (Pouliquen and Forterre, 2002).

The functions  $\mu_{stop}$  and  $\mu_{start}$  are defined by:

$$\mu_{stop}(h) = \tan \phi_1 + \frac{\tan \phi_2 - \tan \phi_1}{1 + h/\mathcal{L}} \quad (8)$$

and

$$\mu_{start}(h) = \tan \phi_3 + \frac{\tan \phi_2 - \tan \phi_1}{1 + h/\mathcal{L}} \quad (9)$$

The critical angles  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  and the parameters  $\mathcal{L}, \beta$  are the parameters of the model.

In particular,  $\mathcal{L}$  is the characteristic depth of the flow over which a transition between the angles  $\phi_1$  to  $\phi_2$  occurs, in the  $\mu_{stop}$  formula. In practice, if  $h \ll \mathcal{L}$ , then  $\mu_{stop}(h) \approx \tan \phi_2$ , and if  $h \gg \mathcal{L}$ , then  $\mu_{stop}(h) \approx \tan \phi_1$ .

**PF equations** The depth-averaged Eqs. (1) source terms thus take the following form:

$$\begin{aligned} S_x &= g_x h - \frac{\bar{u}}{\|\bar{\mathbf{u}}\|} \left[ h \left( g_z + \frac{\bar{u}^2}{r_x} \right) \mu_b(\|\bar{\mathbf{u}}\|, h) \right] + g_z h \frac{\partial h}{\partial x} \\ S_y &= g_y h - \frac{\bar{v}}{\|\bar{\mathbf{u}}\|} \left[ h \left( g_z + \frac{\bar{v}^2}{r_y} \right) \mu_b(\|\bar{\mathbf{u}}\|, h) \right] + g_z h \frac{\partial h}{\partial y} \end{aligned} \quad (10)$$

In our study, sampled input parameters are  $\phi_1$ ,  $\Delta\phi_{12} := \phi_2 - \phi_1$ , and  $\beta$ . In particular, the range of  $\phi_1$  depends on the case study, whereas  $\Delta\phi_{12} \in [10^\circ, 15^\circ]$ , and  $\beta \in [0.1, 0.85]$ . Moreover,  $\phi_3 = \phi_1 + 1^\circ$ , and  $\mathcal{L}$  is equal to  $1dm$  and  $1mm$  in the two case studies, respectively (Pouliquen and Forterre, 2002; Forterre and Pouliquen, 2003).

### 3.1.3 Voellmy-Salm

The theoretical analysis of dense snow avalanches led to the VS rheology (VS) (Voellmy, 1955; Salm et al., 1990; Salm, 1993; Bartelt et al., 1999). Dense snow or debris avalanches consist of mobilized, rapidly flowing ice-snow mixed to debris-rock granules (Bartelt and McArdell, 2009). The VS rheology assumes a velocity dependent resisting term in addition to the traditional basal friction, ideally capable of including an approximation of the turbulence-generated dissipation. Many experimental and theoretical studies were developed in this framework (Gruber and Bartelt, 2007; Kern et al., 2009; Christen et al., 2010; Fischer et al., 2012).

The following relation between shear and normal stresses holds:

$$\tau = \mu \sigma + \frac{\rho \|\underline{\mathbf{g}}\|}{\xi} \|\bar{\mathbf{u}}\|^2, \quad (11)$$

where,  $\sigma$  denotes the normal stress at the bottom of the fluid layer and  $\underline{\mathbf{g}} = (g_x, g_y, g_z)$  represents the gravity vector. The two parameters of the model are the bed friction coefficient  $\mu$  and the velocity dependent friction coefficient  $\xi$ .

We can summarize VS rheology assumptions as:

- *Basal Friction* is based on a constant coefficient, similarly to the MC rheology.
- *Internal Friction* is neglected.
- *Earth pressure coefficient* is equal to one.
- *Additional turbulent friction* is based on the local velocity by a quadratic expression.
- Velocity based *curvature effects* are included into the equations, following an alternative formulation.

The effect of the topographic local curvatures is addressed with terms containing the local radii of curvature  $r_x$  and  $r_y$ . In this case the expression is based on the speed instead of the scalar components of velocity (Pudasaini and Hutter, 2003; Fischer et al., 2012).

**VS equations** Therefore, the final source terms take the following form:

$$\begin{aligned} S_x &= g_x h - \frac{\bar{u}}{\|\tilde{\mathbf{u}}\|} \left[ h \left( g_z + \frac{\|\tilde{\mathbf{u}}\|^2}{r_x} \right) \mu + \frac{\|\mathbf{g}\|}{\xi} \|\tilde{\mathbf{u}}\|^2 \right], \\ S_y &= g_y h - \frac{\bar{v}}{\|\tilde{\mathbf{u}}\|} \left[ h \left( g_z + \frac{\|\tilde{\mathbf{u}}\|^2}{r_y} \right) \mu + \frac{\|\mathbf{g}\|}{\xi} \|\tilde{\mathbf{u}}\|^2 \right]. \end{aligned} \quad (12)$$

In our study, sampled input parameters are  $\mu$ , and  $\xi$ , on ranges depending on the case study. In particular,  $\xi$  uniform sampling is accomplished in log-scale. In fact, values of  $\xi$  between 250 and 4,000  $m/s^2$  have been described for snow avalanches (Salm, 1993; Bartelt et al., 1999; Gruber and Bartelt, 2007).

### 3.2 Latent variables

For analysis of modeling assumptions we need to record and classify the results of different modeling assumptions. In our case study, we focus on the right-hand side terms in the momentum equation and we call them RHS forces. They are latent variables since internal to the computation and rarely visible as a system output.

$$\mathbf{RHS}_1 = [g_x h, g_y h], \quad (13)$$

it is the gravitational force term, it has the same formulation in all models.

The expression of **basal friction force**  $\mathbf{RHS}_2$  depends on the model:

$$\begin{aligned} \mathbf{RHS}_2 &= -hg_z \tan(\phi_{bed}) \left[ \frac{\bar{u}}{\|\tilde{\mathbf{u}}\|}, \frac{\bar{v}}{\|\tilde{\mathbf{u}}\|} \right], \text{ in MC model.} \\ \mathbf{RHS}_2 &= -hg_z \mu_b(\|\tilde{\mathbf{u}}\|, h) \left[ \frac{\bar{u}}{\|\tilde{\mathbf{u}}\|}, \frac{\bar{v}}{\|\tilde{\mathbf{u}}\|} \right], \text{ in PF model.} \\ \mathbf{RHS}_2 &= -hg_z \mu \left[ \frac{\bar{u}}{\|\tilde{\mathbf{u}}\|}, \frac{\bar{v}}{\|\tilde{\mathbf{u}}\|} \right], \text{ in VS model.} \end{aligned} \quad (14)$$

The expression of the force related to the **topography curvature**,  $\mathbf{RHS}_3$ , also depends on the model:

$$\begin{aligned} \mathbf{RHS}_3 &= -h \tan(\phi_{bed}) \left[ \frac{\bar{u}^3}{r_x \|\tilde{\mathbf{u}}\|}, \frac{\bar{v}^3}{r_y \|\tilde{\mathbf{u}}\|} \right], \text{ in MC model.} \\ \mathbf{RHS}_3 &= -h \mu_b(\|\tilde{\mathbf{u}}\|, h) \left[ \frac{\bar{u}^3}{r_x \|\tilde{\mathbf{u}}\|}, \frac{\bar{v}^3}{r_y \|\tilde{\mathbf{u}}\|} \right], \text{ in PF model.} \\ \mathbf{RHS}_3 &= -h \mu \left[ \frac{\bar{u} \|\tilde{\mathbf{u}}\|}{r_x}, \frac{\bar{v} \|\tilde{\mathbf{u}}\|}{r_y} \right], \text{ in VS model.} \end{aligned} \quad (15)$$

All the three models have an additional force term, having a different expressions and different meaning in the three models:

$$\begin{aligned} \mathbf{RHS}_4 &= -hk_{ap} \sin(\phi_{int}) \left[ \operatorname{sgn}\left(\frac{\partial \bar{u}}{\partial y}\right) \frac{\partial(g_z h)}{\partial y}, \operatorname{sgn}\left(\frac{\partial \bar{v}}{\partial x}\right) \frac{\partial(g_z h)}{\partial x} \right], \text{ in MC model.} \\ \mathbf{RHS}_4 &= g_z h \left[ \frac{\partial h}{\partial x}, \frac{\partial h}{\partial y} \right], \text{ in PF model.} \\ \mathbf{RHS}_4 &= -\frac{\|\mathbf{g}\|}{\xi} \|\tilde{\mathbf{u}}\|^2 \left[ \frac{\bar{u}}{\|\tilde{\mathbf{u}}\|}, \frac{\bar{v}}{\|\tilde{\mathbf{u}}\|} \right], \text{ in VS model.} \end{aligned} \quad (16)$$

These latent variables can be analyzed locally and globally for discriminating among the different modeling assumption.

Finally, we also study the spatial integrals defined by  $F(t) = \int_{\mathbb{R}^k} f(\mathbf{x}, t) d\mathbf{x}$ , where  $d\mathbf{x}$  is the area of the mesh elements. This provides a global view of the results and is complementary to the observations taken locally. For instance, by integrating the scalar product of source terms in the momentum balance and velocity we can compare the relative importance of modeling assumptions when we seek accuracy on global quantities.

## 4 Small scale flow on inclined plane and flat runway

Our first case study assumes very simple boundary conditions, and corresponds to a laboratory experiment fully described in (Webb, 2004; Bursik et al., 2005; Webb and Bursik, 2016). It is a classical flow down an inclined plane set-up, including a change in slope to an horizontal plane (Fig. 2). Modeling flow of granular material down an inclined plane was explored in detail by several studies, both theoretically and experimentally (e.g. (Ruyer-Quil and Manneville, 2000; Silbert et al., 2001; Pitman et al., 2003a)).

In our setting, four locations are selected among the center line of the flow to accomplish local testing. These are: the initial pile location  $L_1 = (-0.70, 0)$  m, the middle of the inclined plane  $L_2 = (-0.35, 0)$  m, the change in slope  $L_3 = (0, 0)$  m, the middle of the flat runway  $L_4 = (0.15, 0)$  m.

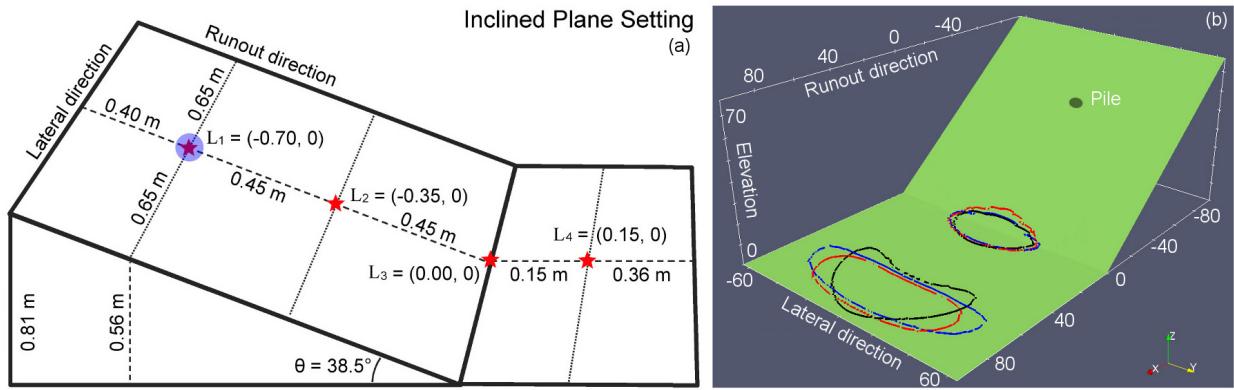


Figure 2: (a) Inclined plane overview, including samples sites (red stars). Pile location is marked by a blue dot. (b) Contours of  $h = 1.0$  mm at last simulated snapshot ( $t = 1.5$  sec) for simulated flows with *minimum runout obtained from min. volume – max. resistance*, and *maximum runout obtained from max. volume – min. resistance*. — : MC, — : PF, — : VS.

### 4.1 Preliminary consistency testing of the input ranges

Addressing a similar case study, (Dalbey et al., 2008) assumed  $\phi_{bed} = [15^\circ, 30^\circ]$ , while (Webb and Bursik, 2016) performed a series of laboratory experiments and found  $\phi_{bed} = [18.2^\circ, 34.4^\circ]$ . We relied on those published parameter choices to select a comprehensive parameter range. Figure 2b displays the screenshots of flow height observed in the extreme cases tested. The Digital elevation Map (DEM) has a 1mm cell size. Simulation options are - `max_time = 2 s`, `height/radius = 1.34`, `length_scale = 1 m`, `number_of_cells_across_axis = 10`, `order = first`, `geoflow_tiny = 1e-4` (Patra et al., 2005; Aghakhani et al., 2016). Initial pile geometry is cylindrical. We remark that small changes in the parameter ranges did not change significantly the results.

- **Material Volume:**  $[449.0, 607.0] \text{ cm}^3$ , i.e. average of  $528.0 \text{ cm}^3$  and uncertainty of  $\pm 15\%$ .
- **Rheology models' parameters:**
  - MC -  $\phi_{bed} \in [18^\circ, 30^\circ]$ .
  - PF -  $\phi_1 \in [10^\circ, 22^\circ]$ .
  - VS -  $\mu \in [0.22, 0.45]$ ,  $\log(\xi) \in [3, 4]$ .

Even if maximum and minimum runout are both matching, the shape and lateral extent of the flow are different between the three models. In particular, MC model can produce the largest lateral extent, and VS model displays an accentuated bow-like shape, due to the increased friction in the lateral margins.

## 4.2 Observable outputs

We express the flow height and acceleration as a function of time, measured in the four locations  $L_1, \dots, L_4$  displayed in Fig. 2a. Uncertainty quantification (UQ) is performed, accordingly to the parameter ranges described above. We always show the mean values and the corresponding 5<sup>th</sup> and 95<sup>th</sup> percentile values, defining an uncertainty range. Estimates of the Froude Number are included in Supporting Information S1.

### 4.2.1 Flow height

Figure 3 displays the flow height,  $h(L, t)$ , at the points  $(L_i)_{i=1, \dots, 4}$ . Given a particular type of flow and collected data we can clearly distinguish model skill in capturing not only that flow but also other possible flows. Past work (Webb, 2004) allowed us to conclude that MC rheology is adequate for modeling simple dry granular flows. We must note the effect of neglecting the flow when its height is  $< 1$  mm, which is at the scale of the smallest granular size (Aghakhani et al., 2016). In fact, continuity assumption would not be valid below this scale, and the 5<sup>th</sup> and 95<sup>th</sup> percentile plots are vertically cut to zero when they decrease over that threshold. The mean plot is not cut to zero but it is dulled by this cutoff.

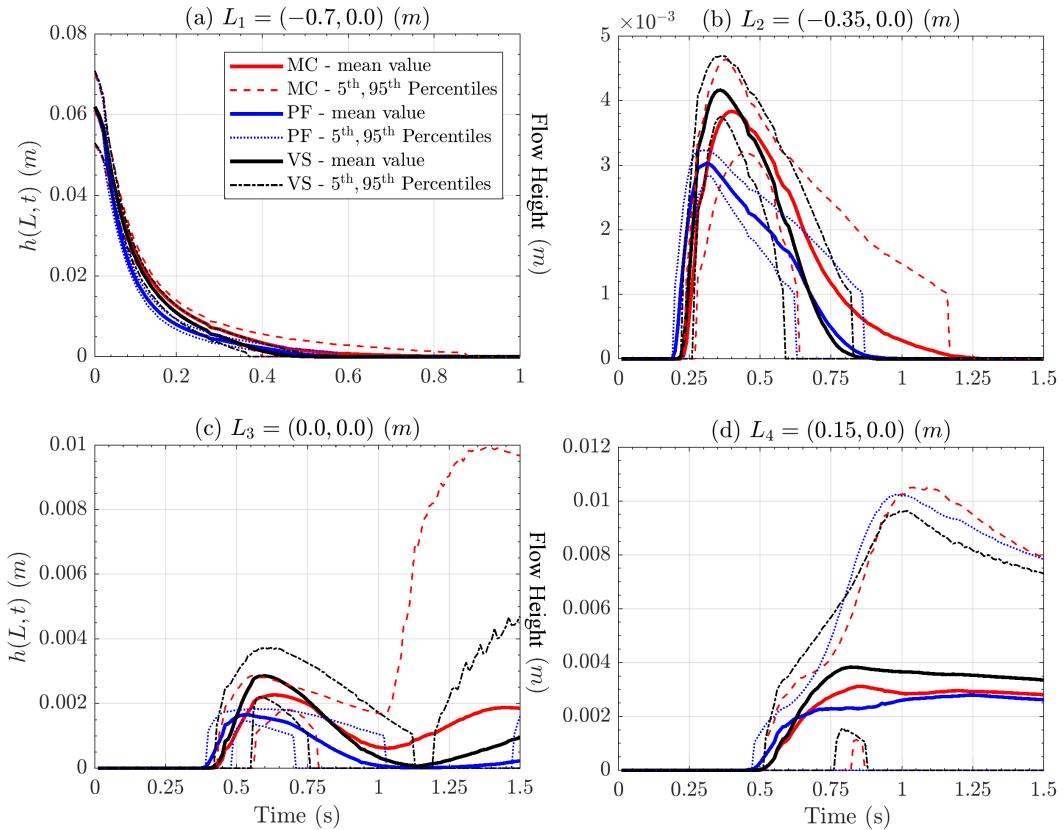


Figure 3: Flow height in four locations. Bold line is mean value, dashed/dotted lines are 5<sup>th</sup> and 95<sup>th</sup> percentile bounds. Different models are displayed with different colors. Plots are at different scale, for simplifying exposition.

In plot 3a, related to point  $L_1$  placed on the initial pile, the values of  $\sim 6 \pm 1$  cm are equal and express the assigned pile height. The flow height decreases slightly faster in PF model, and slower in MC, compared to VS. Differences are more significant in plot 3b, related to point  $L_2$ , placed in the middle of the slope. Maximum flow height on average is greater in VS,  $4.1 \pm 0.2$  mm, but more uncertain in MC,  $3.9 \pm 0.4$  mm, and generally smaller in PF model,  $3.0 \pm 0.1$  mm. After the peak, PF decreases significantly slower than the other models. These height values are about 15 times smaller than initial pile height. None of the models leaves a significant material deposit in  $L_1$  or  $L_2$ , and hence the 95<sup>th</sup> percentile of the height is null at the ending-time. In contrast, a deposit is left at points  $L_3$  and  $L_4$ , i.e. plot 3c placed at the change in slope, and plot 3d in the middle of the flat runway. At  $L_3$  MC's deposit, 2 mm with uncertainty [-2,+8] mm, is higher than the other models' deposits. The plot profile is bimodal, showing a first peak at  $\sim 0.6$  s, and then a reduction until 1 s, before the final accumulation. At  $L_4$ , deposit it is not significantly different between the three models. It measures  $\sim 3$  mm on average, slightly more than this in VS, with uncertainty [-3,+7] mm.

#### 4.2.2 Flow acceleration

Figure 4 shows the flow acceleration,  $\|\underline{\mathbf{a}}\|(L, t)$ , at the points  $(L_i)_{i=1,\dots,4}$ . Acceleration is the link between force terms and observable motion. We calculated it from the left-hand-side of the dynamical equation, but using the RHS terms produces very similar results.

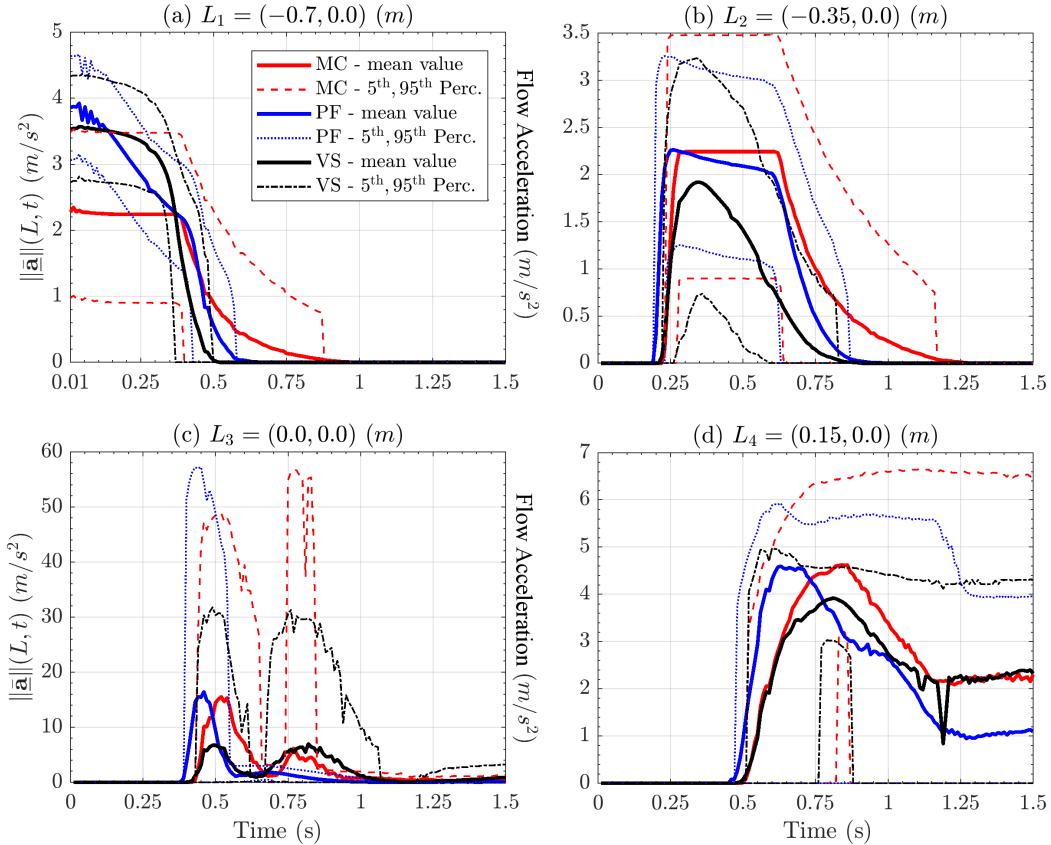


Figure 4: Flow acceleration modulus in four locations. Bold line is mean value, dashed lines are 5<sup>th</sup> and 95<sup>th</sup> percentile bounds. Different models are displayed with different colors. Plots are at different scale.

In plot 4a, related to point  $L_1$ , MC and VS show a plateau before  $\sim 0.4$  sec, at  $\sim 2.5$   $m/s^2$  and  $\sim 3.5$   $m/s^2$ , respectively, while PF linearly decreases between those same values. Instead in plot 4b, related to point  $L_2$ , MC and PF show a plateau, at  $\sim 2.2$   $m/s^2$ , while VS has a more bell-shaped profile.

UQ tells us that PF is affected by a smaller uncertainty than the other models. In plot 4c, related to point  $L_3$ , all the models show a bimodal profile, with peaks at  $\sim 0.5$  sec and 0.8 sec. This is more accentuated in MC and VS, whereas the second peak is almost absent from PF's profile. The second peak is motivated by an increase of velocity in the down-slope direction after its reduction due to lateral spreading of material. At the first peak, acceleration values are significant, with average peaks in MC and PF both at  $\sim 15 \text{ m/s}^2$ , and 95<sup>th</sup> percentile plot reaching  $\sim 50 \text{ m/s}^2$  and  $\sim 55 \text{ m/s}^2$ , respectively. VS shows about halved acceleration peak values. At the second peak, average acceleration values are similar in MC and VS, at  $\sim 5 \text{ m/s}^2$ . In contrast, 95<sup>th</sup> percentile plot is  $> 50 \text{ m/s}^2$  for MC, while  $\sim 30 \text{ m/s}^2$  in VS. In plot 4d, related to point  $L_4$ , the acceleration has a first peak at  $\sim 4 \text{ m/s}^2$ , and a final asymptote at  $\sim 2 \text{ m/s}^2$  in MC and VS,  $\sim 1 \text{ m/s}^2$  for PF. These values indicate flow deceleration, and uncertainty is more relevant in MC and PF than in VS.

### 4.3 Statistical analysis of latent variables

Figure 5 shows the dominance factors  $(P_i)_{i=1,\dots,4}$ , obtained from the modulus of the RHS terms in the slope direction.

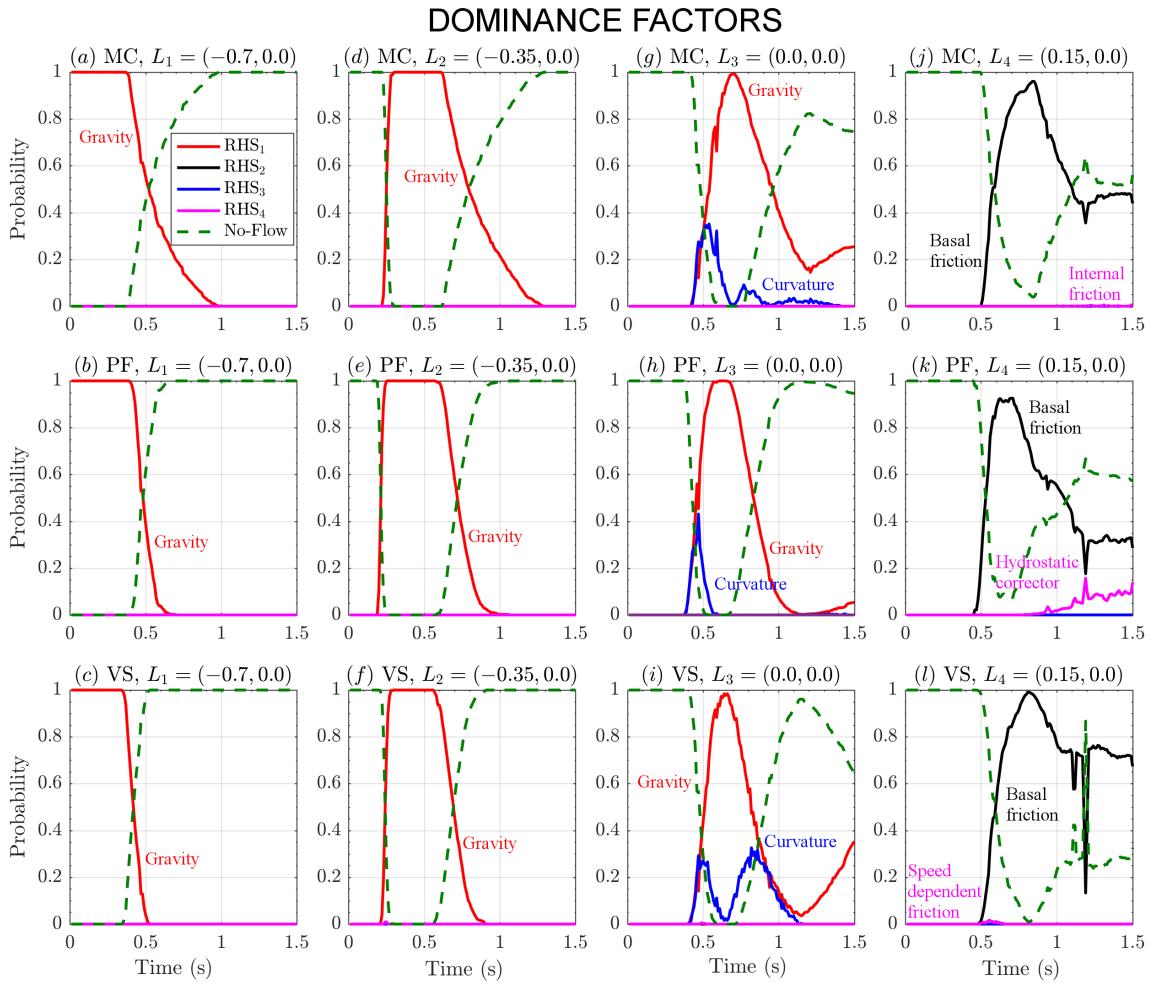


Figure 5: Dominance factors of **RHS** forces on the slope direction in four locations. Different models are plotted separately: (a,d,g,j) assume MC; (b,e,h,k) assume PF; (c,f,i,l) assume VS. Different colors correspond to different force terms. No-flow probability is also displayed with a green dashed line.

Each dominance factor is the probability of a force term to be the greatest one, and hence belongs to  $[0, 1]$ . The plots include also the probability of no-flow being observed at the considered point.

The plots 5a,b,c are related to point  $L_1$ , placed on the initial pile. Only  $\mathbf{RHS}_1$  can be the dominant variable, and no-flow probability is  $(1 - P_1)$ . Same thing in the plots 5d,e,f related to point  $L_2$ , placed in the middle of the slope. Then, plots 5g,h,i are related to point  $L_3$ , placed at the change in slope. In  $L_3$ ,  $\mathbf{RHS}_3$  can be the dominant term for a short time, with a peak probability of  $\sim 30\%$ . Plots 5j,k,l are related to point  $L_4$ , placed in the middle of the flat runway. In  $L_4$  only  $\mathbf{RHS}_2$  can be the dominant term, except in PF where there is a  $\sim 10\%$  chance that  $\mathbf{RHS}_4$  is the dominant term at the ending-time.

Dominance factors describe the main dynamics of the flow, but they cannot tell anything on the not-dominant variables. The expected contributions complete the statistical description of the latent variables. They scale the force terms by the dominant dynamics, and represent the degree of relevance of the secondary assumptions with respect to the dominant one. They are averaged with respect to  $P_M$  on the parameter range, and change as a function of time.

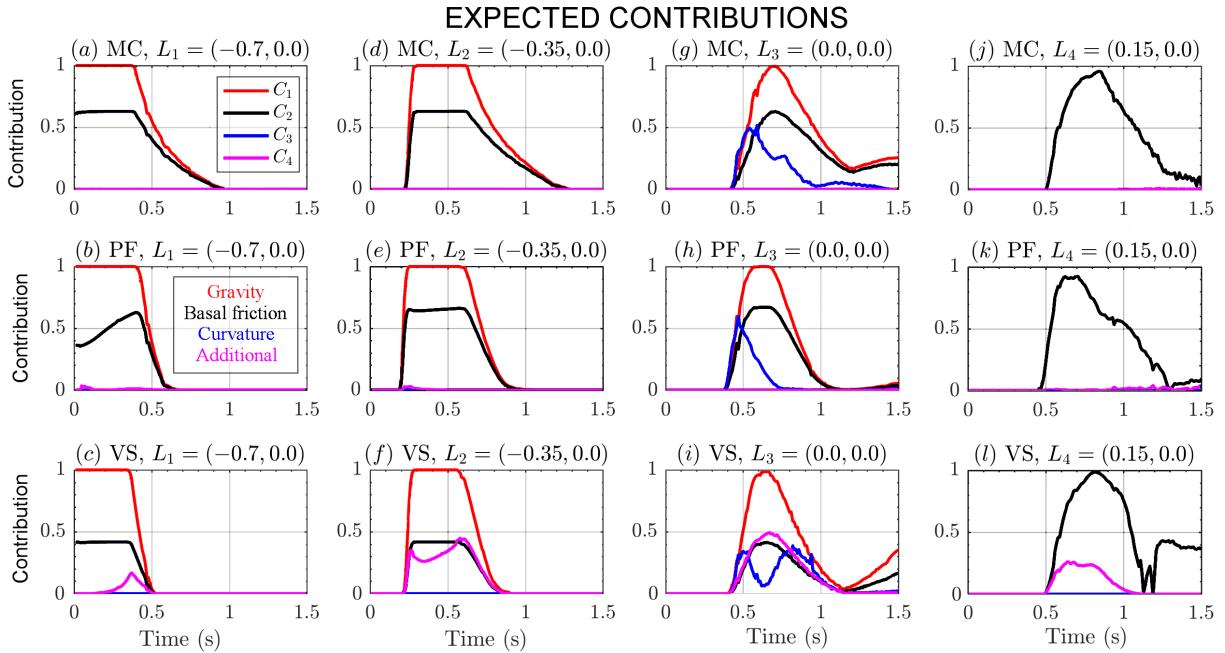


Figure 6: Expected contributions of  $\mathbf{RHS}$  forces on the slope direction in four locations. Different models are plotted separately: (a,d,g,j) assume MC; (b,e,h,k) assume PF; (c,f,i,l) assume VS. Different colors correspond to different force terms.

Figure 6 shows  $\mathbb{E}[C_i]_{i=1,\dots,4}$ , for the three rheology models. The plots 6a,b,c are related again to point  $L_1$ .  $C_1$  and  $C_2$  give the major contributions, with a minor contribution from  $C_4$  in VS. Contributions profiles are flat plateaus that start to wane after  $0.4s$ . The plots 6d,e,f are related to point  $L_2$ . The major contributions are  $C_1$  and  $C_2$ , and have a trapezoidal profile. In VS,  $C_4$  resembles  $C_2$ , but it is bimodal instead than trapezoidal. The plots 6g,h,i are related to point  $L_3$ .  $C_1$  and  $C_2$  are still the largest, but their profiles are bell-shaped. In VS,  $C_4$  is almost equal to  $C_2$ . In all the models,  $C_3$  becomes significant, with a peak similar to  $C_2$ . It shows different profiles - triangular for MC and PF, bimodal for VS. In MC the decrease occurs in two stages. Due to the presence of deposit, all the contributions are small (particularly small in PF), but not zero at the ending time. The plots 6j,k,l are related to point  $L_4$ . Only  $C_2$  has a major role, with a bell shaped profile faster to wax than to wane. Contribution  $C_4$  has a minor role in VS and PF.

#### 4.4 Flow extent and spatial integrals

Figure 7 shows the volumetric average of speed and Froude Number. Moreover, it shows the lateral extent and inundated area of flow, as a function of time. These global quantities have smoother plots than the local measurements described above. However, most of the details observed in local measurements are not easy to discern anymore.

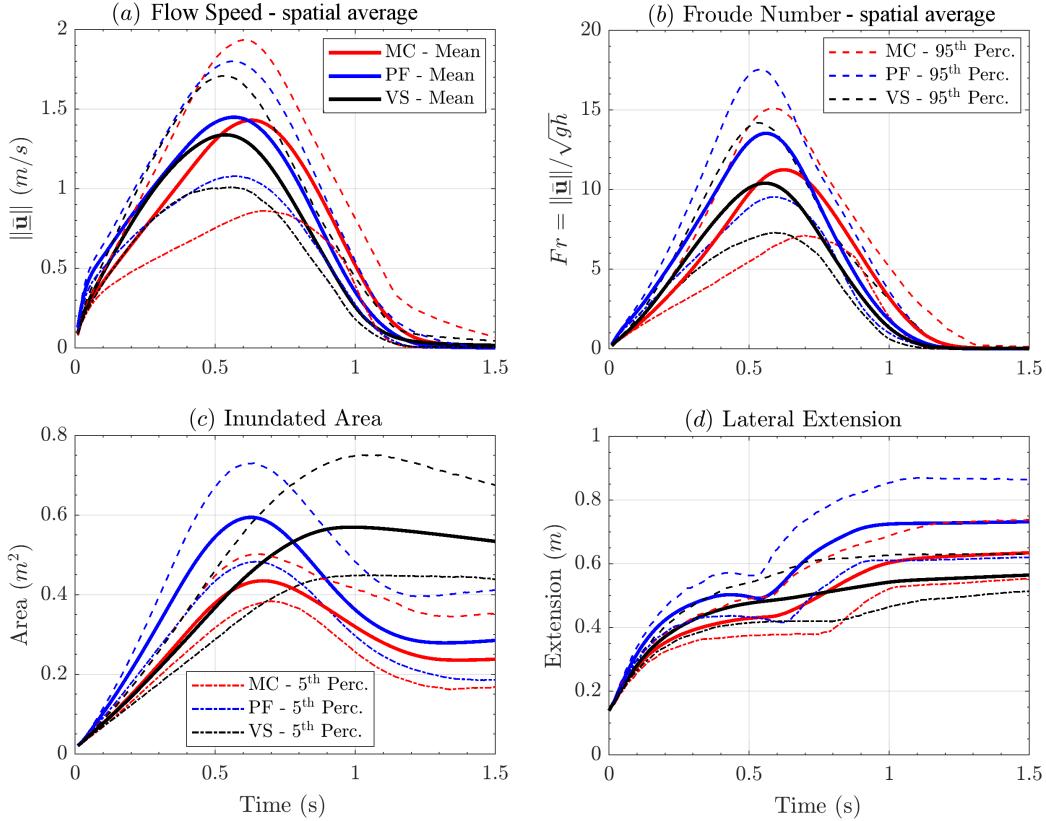


Figure 7: Comparison between spatial averages of (a) flow speed, and (b) Froude Number in addition to the flow (c) lateral extent, and (d) inundated area, as a function of time. Different models are displayed with different colors.

In plot 7a the speed has a bell-shaped profile in all the models, with an average peak at  $\sim 1.4m/s$  and uncertainty range of  $\pm 0.4m/s$  for PF and VS. VS is slightly slower, reaching  $\sim 1.3m/s$  on average. MC shows a larger uncertainty range, of  $\pm 0.6m/s$ . The maximum speed is reached first by VS and PF at  $\sim 0.55s$ , and last by MC at  $\sim 0.65s$ . In plot 7b, also the Froude Number has a bell-shaped profile.  $Fr$  peaks are temporally aligned with speed peaks, and are  $\sim 10$  in VS,  $\sim 11$  in MC,  $\sim 13.5$  in PF, on average. Uncertainty range is about  $\pm 4$  in all models. In plot 7c inundated area shows similar maximum values in PF and VS, at  $\sim 0.6m^2$  on average, and uncertainty of  $\pm 0.15m^2$ . MC is lower, at  $\sim 0.45m^2$  on average, and less uncertain,  $\pm 0.10m^2$ . VS does not decrease significantly after reaching the peak, whereas the other models contract their area to approximately half its maximum extent. In plot 7d the lateral extent starts equal to the pile diameter  $\sim 15cm$ , and then rises in two stages in MC and PF. The second and greater rise starts at  $\sim 0.6s$ , and corresponds with the time of arrival at the change in slope (see Fig. 3c). In contrast, VS rises without showing two phases. At  $\sim 0.6s$ , average lateral extent is  $\sim 50cm$  in PF and VS, and  $\sim 43cm$  in MC. Uncertainty range is  $\pm 7cm$  for all models at that time. Final extent is  $\sim 75cm$  in PF,  $\sim 65cm$  in MC,  $\sim 55cm$  in VS. Uncertainty range is  $\pm 5cm$  in VS, but rises to  $\pm 10cm$  in MC and PF.

## 4.5 Power integrals

Figure 8 shows the spatial integral of powers (scalar product of force and velocity terms). The spatial integration is performed on half spatial domain, due to the symmetry with respect to the flow central axis. In particular, the power estimates assume a material density  $\rho = 805\text{kg/m}^3$ . This a constant scaling factor, and the plots are not further affected by its value. Corresponding plots of the force terms are included in Supporting Information S2.

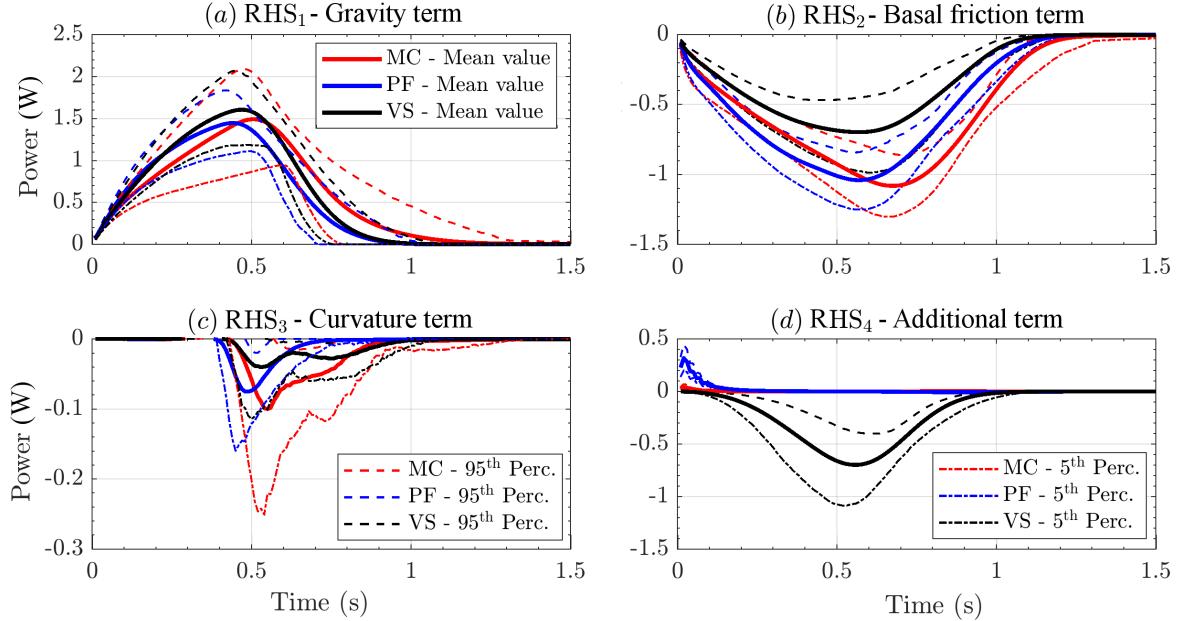


Figure 8: Spatial integral of the RHS powers. Bold line is mean value, dashed lines are 5<sup>th</sup> and 95<sup>th</sup> percentile bounds. Different models are displayed with different colors.

The scalar product with velocity imposes a bell-shaped profile, as observed in Fig. 7a. In plot 8a the power of **RHS<sub>1</sub>** represents the effect of the gravity in all the models. It starts from zero and rises up to  $\sim 1.5W$  at  $\sim 0.55s$ , then decreases to zero after the material crosses the change in slope. Uncertainty range of  $\pm 0.5W$  affects the peak values. MC decreases slower than the other models, and has a more significant uncertainty after the change in slope. PF decreases faster. In plot 8b the power of **RHS<sub>2</sub>** represents the basal friction. It is negative and peaks to  $\sim 1.1 \pm 0.2W$  in MC,  $\sim 1.0 \pm 0.2W$  PF,  $\sim 0.7 \pm 0.3W$  in VS. A similar bell-shaped profile is shared by the three models, and basal friction power becomes negligible at the ending-time. In plot 8c the power of **RHS<sub>3</sub>** is related to the curvature effects, and it is not null only at the change in slope. It is always dissipative, i.e. opposed to flow velocity, because it is equivalent to the friction due to the additional weight generated by centrifugal forces. It is weaker than  $-0.1W$  on average, ten times smaller than the previous powers, although MC lower percentile reaches  $\sim -0.25W$ . VS displays a bimodal profile, with a second and weaker peak at  $\sim 0.75s$ . In plot 8d the power of **RHS<sub>4</sub>** is related to the additional forces of the models, differently characterized. This term is really relevant in VS, although also in PF has a very short lasting positive peak up to  $0.3W$  before to become null at  $\sim 0.1s$ . This power in VS is a speed dependent term, always dissipative. It is bell shaped and null before  $\sim 0.1s$  and after  $\sim 1s$ . At the time of change in slope it is  $\sim -0.7W$ ,  $\pm 0.3W$ .

## 5 Large scale flow on the SW slope of Volcán de Colima

Our second case study is a geophysical flow down the SW slope of Volcán de Colima (MX) - an andesitic stratovolcano that rises to 3,860 m above sea level, situated in the western portion of the Trans-Mexican

Volcanic Belt (Fig. 9).

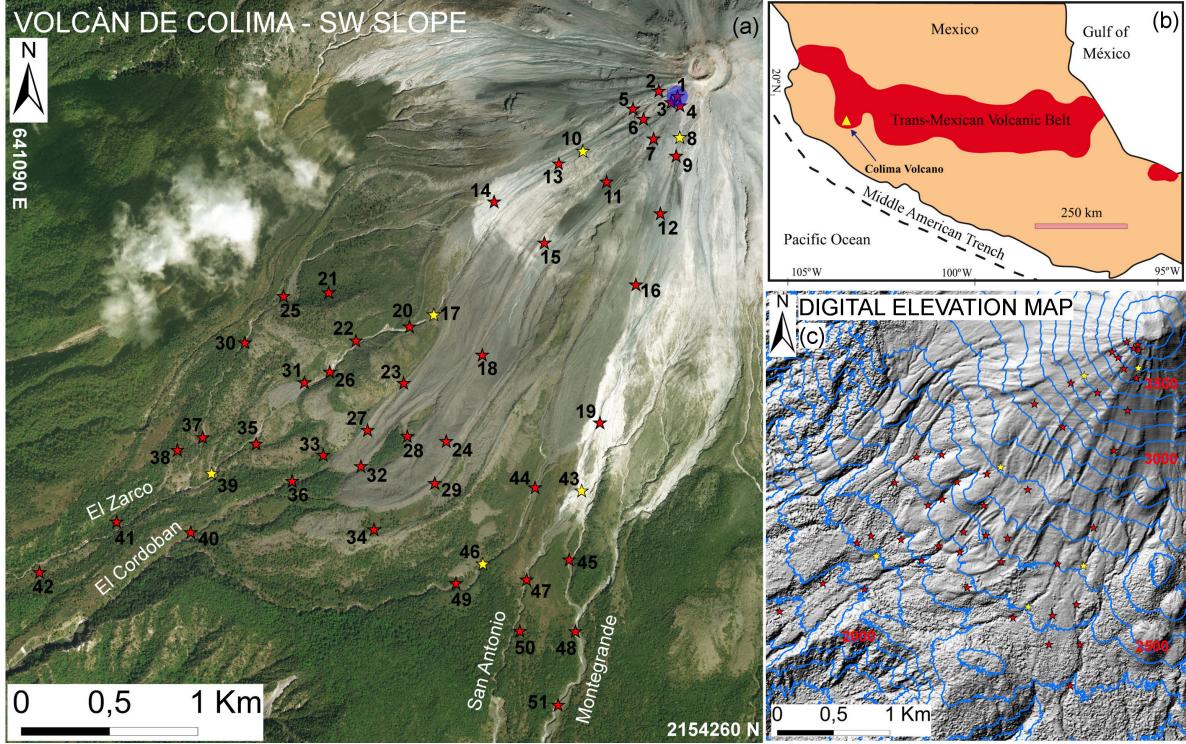


Figure 9: (a) Volcán de Colima (México) overview, including 51 numbered locations (stars) and major ravines. Initial pile is marked by a blue dot. Coordinates are in UTM zone 13N. (b) Regional geology map. (c) Digital elevation map. Six preferred locations are colored in yellow. Elevation isolines are displayed in blue, elevation values in red.

Volcán de Colima has historically been the most active volcano in México (la Cruz-Reyna, 1993; Zobin et al., 2002; González et al., 2002). The modeling of pyroclastic flows generated by explosive eruptions and lava dome collapses of Volcán de Colima is a well studied problem (Martin Del Pozzo et al., 1995; Sheridan and Macías, 1995; Saucedo et al., 2002, 2004, 2005; Sarocchi et al., 2011; Capra et al., 2015). The presence of a change in slope followed by multiple ravines characterize the SW slope of the volcano. Volcán de Colima has been used as a case study in several research papers involving the Titan2D code (Rupp, 2004; Rupp et al., 2006; Dalbey et al., 2008; Yu et al., 2009; Sulpizio et al., 2010; Capra et al., 2011; Aghakhani et al., 2016). Remarkably, on July 10<sup>th</sup>-11<sup>th</sup>, 2015, the volcano underwent its most intense eruptive phase since its Subplinian-Plinian 1913 AD eruption (Saucedo et al., 2010; Zobin et al., 2015; Reyes-Dávila et al., 2016; Capra et al., 2016; Macorps et al., 2018).

We assume the flow to be generated by the gravitational collapse of a material pile placed close to the summit area - at 644956N, 2157970E UTM13 (Rupp et al., 2006; Aghakhani et al., 2016). The volcano has produced several pyroclastic flows generated by lava dome collapse, called Merapi style flows (Macorps et al., 2018). A dome collapse occurs when there is a significant amount of recently-extruded highly-viscous lava piled up in an unstable configuration around a vent. Further extrusion and/or external forces can cause the still hot dome of viscous lava to collapse, disintegrate, and avalanche downhill (Bursik et al., 2005; Wolpert et al., 2016; Hyman and Bursik, 2018). The hot, dense blocks in this “block and ash” flow (BAF) will typically range from centimeters to a few meters in size. Our computations were performed on a DEM of 5m-pixel resolution, obtained from Laser Imaging Detection and Ranging (LIDAR) data acquired in 2005 (Davila et al., 2007; Sulpizio et al., 2010). We placed 51 locations along the flow inundated area to accomplish local testing. Six of them are then adopted as preferred locations, being representative of different flow regimes.

## 5.1 Preliminary consistency testing of the input ranges

In this same setting, (Dalbey et al., 2008) assumed  $\phi_{bed} = [15^\circ, 35^\circ]$ , while (Capra et al., 2011) adopted  $\phi_{bed} = 30^\circ$ . Moreover, (Spiller et al., 2014; Bayarri et al., 2015; Ogburn et al., 2016) found a statistical correlation between flow size and effective basal friction inferred from field observation of geophysical flows. A BAF at the scale of our simulations would possess  $\phi_{bed} = [13^\circ, 18^\circ]$  according to their estimates. Small changes in the parameter ranges do not change significantly the results.

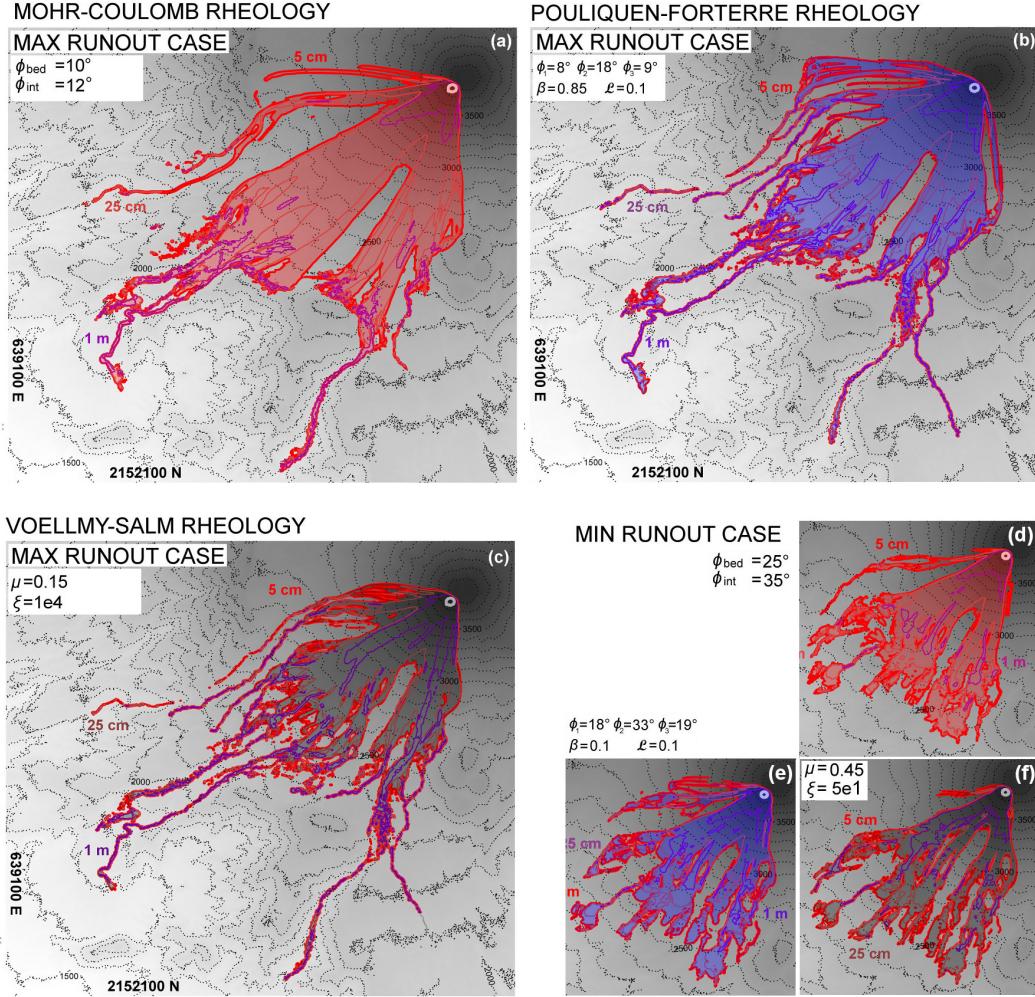


Figure 10: Volcán de Colima - comparison between *max flow height* maps of simulated flow, assuming MC (a),(d), PF (b),(e), and VS (c),(f) models. Extreme cases - (a),(b),(c) **max. volume – min. resistance** and (d),(e),(f) **min. volume – max. resistance**.

Figure 10 displays the maps of maximum flow height observed in the extreme cases tested. Simulation options are - `max_time = 7200 s` (2 hours), `height/radius = 0.55`, `length_scale = 4e3 m`, `number_of_cells_across_axis = 50`, `order = first`, `geoflow_tiny = 1e-4` (Patra et al., 2005; Aghakhani et al., 2016). Initial pile geometry is paraboloid. The models display different macroscopic features. In particular, VS lateral spreading is significant lower and the material reaches higher thickness, whereas PF model seems to stop more gradually than MC with more complex inundated area boundary lines.

- **Material Volume:**  $[2.08, 3.12] \times 10^5 m^3$ , i.e. average of  $2.6 \times 10^5 m^3$  and uncertainty of  $\pm 20\%$ .

- Rheology models' parameters:
  - MC -  $\phi_{bed} \in [10^\circ, 25^\circ]$ .
  - PF -  $\phi_1 \in [8^\circ, 18^\circ]$ .
  - VS -  $\mu \in [0.15, 0.45]$ ,  $\log(\xi) \in [1.7, 4]$ .

## 5.2 Exploring Flow Limits

Figure 11d,e,f,g,h,i show examples of the contributions obtained assuming parameter values at the extremes of their range.

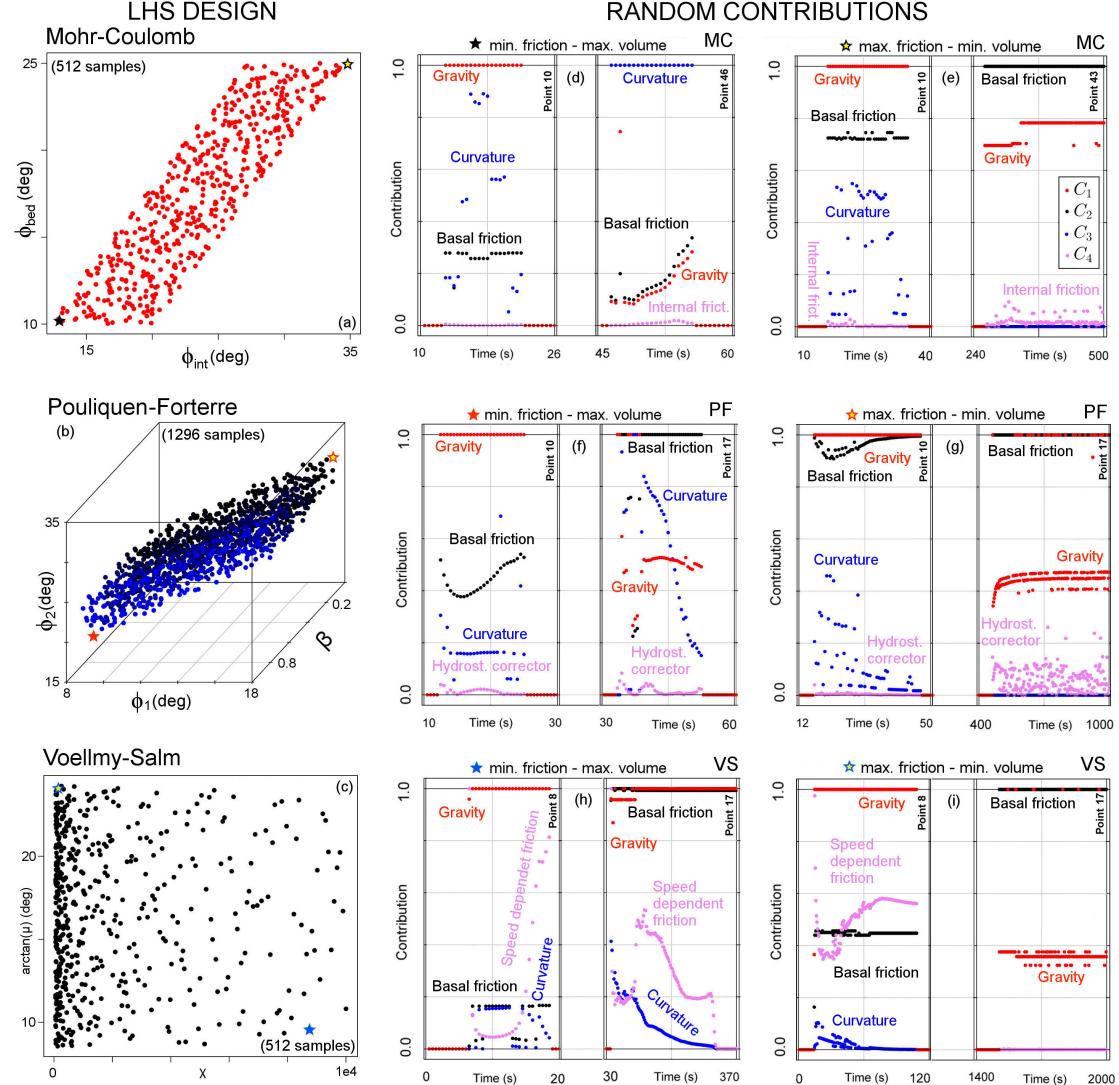


Figure 11: (a,b,c) Example of LHS design, Volcán de Colima case study. Colored stars mark the values producing the minimum and maximum friction. The parameter values are projected with respect to their volume value. Plots of random contributions are included for (d,e) MC, (f,g) PF, and (h,i) VS model. Each plot includes two graphs. One refers to a proximal location to the initial pile (left), and the other to a more distal location (right). Point numbers refer to Fig.13. Different colors correspond to different force terms.

The dominant variable is expressed by the dots on the top line,  $C_i = 1$ . Data is inherently discontinuous due to the mesh modification, and it is reported with colored dots. If the mesh element which contains the considered spatial location changes, then the force term is calculated on a different region and suddenly changes too. However, it is evident that the dynamics and its temporal scale is evolving, and that the contributions can reveal a large amount of information about it. We remark that the calculation of  $\mathbb{E}[C_i]$  with respect to  $P_M$  removes the effects of data discontinuity, and hence this is a fundamental step in our further analysis. We note that the above choices are easily changed, and if we are interested for instance in the performance of the models for very large or very small flows, a suitable volume range can be chosen and the procedure re-run.

### 5.3 Observable outputs

The number of spatial locations is significantly high. We placed 51 points to span the entire inundated area, in search of different flow regimes, as displayed in Fig. 9. We remark that all the distances reported in the following are measured in vertical projection, thus without considering the differences in elevation.

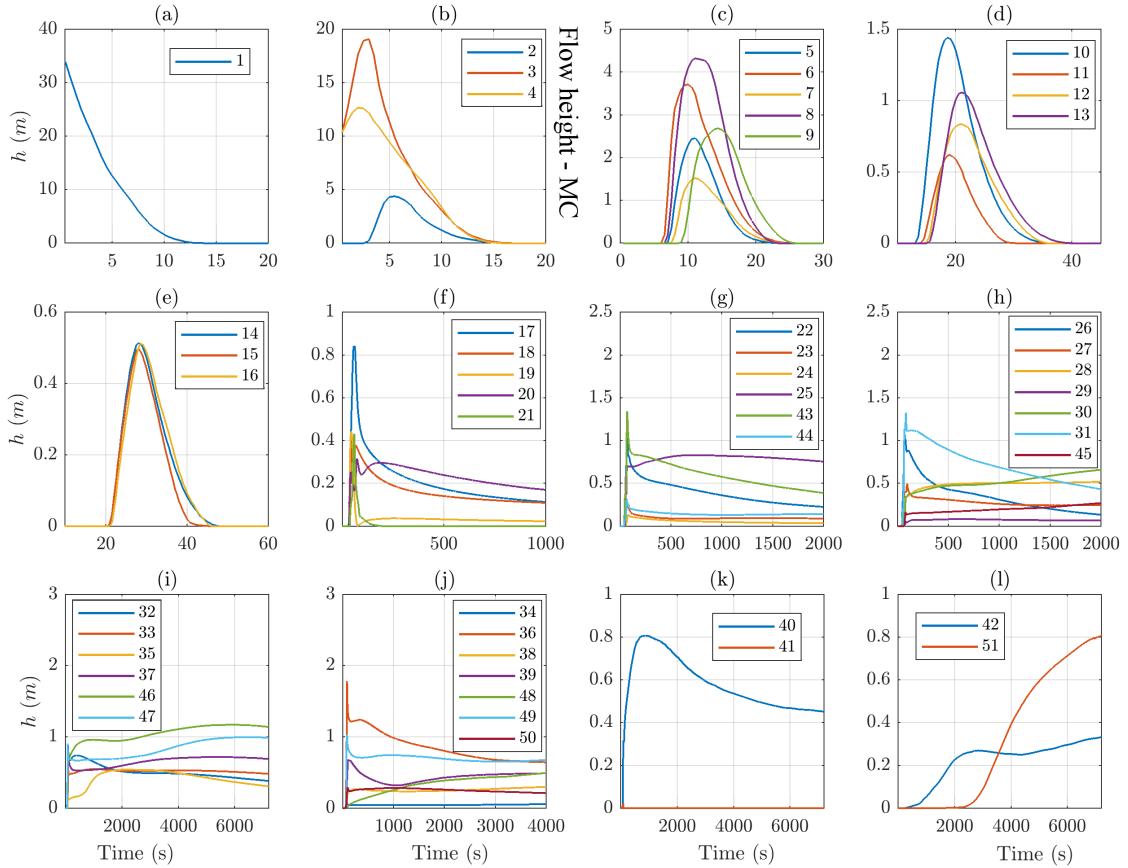


Figure 12: MC model, mean flow height  $h(L, t)$  in 51 numbered locations (Fig. 9). Different plots have different scales on either time and space axes.

Figure 12 shows the mean flow height,  $h(L, t)$ , at the 51 spatial locations of interest, according to MC. Plots of the Froude Number at the 51 points are included in Supporting Information S3. In plot 12a, the only location is set on the center of the initial pile, and the profile is similar to what observed in point  $L_1$  of the inclined plane case study, in Fig.3a. In this case the height decreases from the initial value to zero in  $\sim 15$ s. In plots 12b,c,d,e, the locations are set at less than  $\sim 1$  km radius from the initial pile. Their profiles are similar to point  $L_2$  in Fig.3b. The height profile is bell-shaped, starting

from zero and then waning back to zero in  $\sim 20$  s. All the dynamics occurs during the first minute. In plots 12f,g,h,i,j, points are set where the slope reduces, and the flow can channelize, and typically leaves a deposit. The distance from the initial pile is  $\sim 2 - 3$  km. The profiles are sometimes similar to  $L_3$  of Fig.3c, other times to  $L_4$  of Fig.3d, in a few cases showing intermediate aspects. In general is either observed an initial short-lasting bulge followed by a slow decrease lasting for several minutes and asymptotically tending to a positive height, or a steady increase of material height tending to a positive height. In both cases it is sometimes observed a bimodal profile in the first 5 minutes. Finally, plots 12k,l focus on three points set at about the runout distance of the flow, in the most important ravines, at  $\sim 4 - 5$  km from the initial pile. Profiles are similar to what observed in point  $L_4$  of Fig.3d.

### 5.3.1 Flow height in six locations

We select six preferred locations, illustrative of a range of flow regimes. They are  $[L_8, L_{10}, L_{17}, L_{39}, L_{43}, L_{46}]$ , as displayed in Figure 13. The first two points,  $L_8$  and  $L_{10}$ , are both proximal to the initiation pile. Points  $L_{17}$  and  $L_{43}$  are placed where the slope is reducing and the ravines start, and  $L_{39}$  and  $L_{46}$  are placed in the channels, further down-slope. In particular,  $L_8$ ,  $L_{43}$ , and  $L_{46}$  are at the western side of the inundated area, whereas  $L_{10}$ ,  $L_{17}$ , and  $L_{39}$  are at the eastern side.

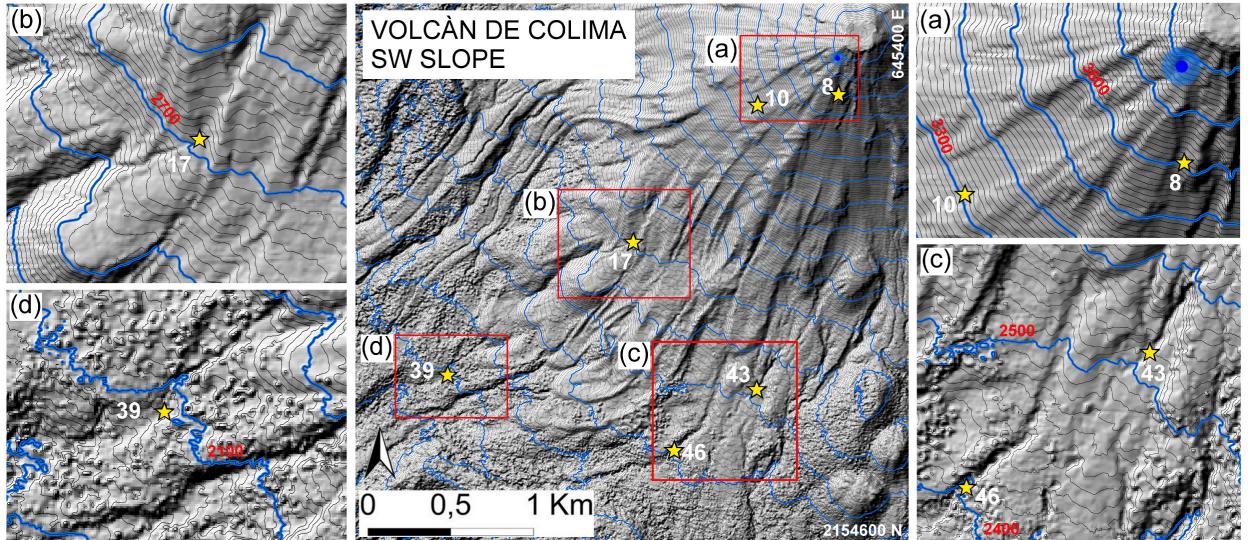


Figure 13: Volcán de Colima (Méjico) overview, including six numbered locations (stars). In (a), (b), (c), (d) are enlarged the proximal topographic features to those locations. Initial pile is marked by a blue dot. Reported coordinates are in UTM zone 13N. Elevation isolines at every 10m are displayed in black, at every 100m in bold blue. Elevation values in red.

Figure 14 shows the flow height,  $h(L, t)$ , at the points  $(L_i)_{i=8,10,17,39,43,46}$ . Froude Number and flow acceleration are included in Supporting Information S4 and S5. Distances from the initial pile are in vertical projection. In plots 14a,b, we show the flow height in points  $L_8$  and  $L_{10}$ ,  $\sim 200$ m and  $\sim 500$ m from the initial pile, respectively. Models MC and PF display similar profiles, positive for less than 15s and bell-shaped. VS requires a significantly longer time to decrease, particularly in point  $L_{10}$ , where the average flow height is still positive after  $\sim 200$ s. Peak average values in  $L_8$  are 3.4m in PF, 4.3m in MC, 4.7m in VS. Uncertainty is about  $\pm 2$ m, halved on the lower side in MC, and PF. In  $L_{10}$ , models MC and PF are very similar, with peak height at 1.4m and uncertainty  $\pm 0.5$ m. Model VS, in contrast, has a maximum height of 1.1m lasting for 50s, and 95<sup>th</sup> percentile reaching 3.7m.

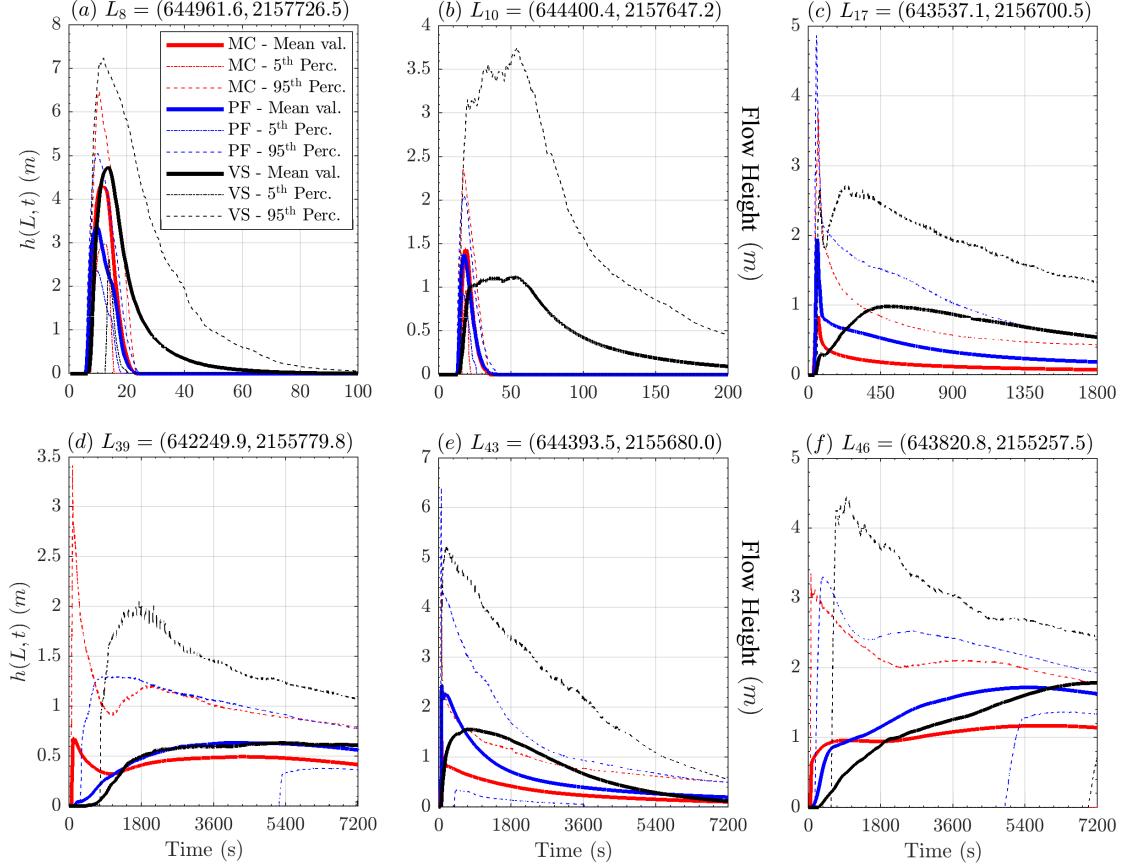


Figure 14: Flow height in six locations. Bold line is mean value, dashed/dotted lines are 5<sup>th</sup> and 95<sup>th</sup> percentile bounds. Different models are displayed with different colors. Plots are at different scale, for simplifying lecture.

In plots 14c,e, we show the flow height in points  $L_{17}$  and  $L_{43}$ , both at  $\sim 2\text{km}$  from the initial pile. All the models show a fast spike during the first minute, followed by a slow decrease. There is still material after 1800s. VS has a secondary rise peaking at  $\sim 450\text{s}$ , which is not observed in the other models. This produces higher values for the most of the temporal duration, but similar deposit thickness after more than 1 hour. Maximum values are 1m for MC, 2m for PF, and 1.5m for VS, in both locations. The 5<sup>th</sup> percentile is zero in all the models, meaning that the parameter range does not always allow the flow to reach these locations. The 95<sup>th</sup> percentile is above 5m, except for VS in point  $L_{17}$ . In plots 14d,f, we show the flow height in points  $L_{39}$  and  $L_{46}$ , both placed at more than 3km from the initial pile. The three models all show a monotone profile except for MC in point  $L_{39}$ , which instead displays an initial spike and a decrease before to rise again. A similar thing is observed in the 95<sup>th</sup> percentiles of all the models. It is significant that the 5<sup>th</sup> percentile of PF becomes positive after  $\sim 5400\text{s}$ , meaning that almost surely the flow has reached that location. Deposit thickness in point  $L_{39}$  is  $\sim 0.5\text{m}$  for all the models, whereas in point  $L_{46}$  it is 1.7m in VS, 1.6m in PF, and 1.2m in MC.

## 5.4 Statistical analysis of latent variables

### 5.4.1 Three locations proximal to the initial pile

Figure 15 shows the dominance factors  $(P_i)_{i=1,\dots,4}$  of the RHS terms modulus, in the three proximal points  $L_8$ ,  $L_{10}$ , and  $L_{17}$ , all closer than 1 km to the initial pile.

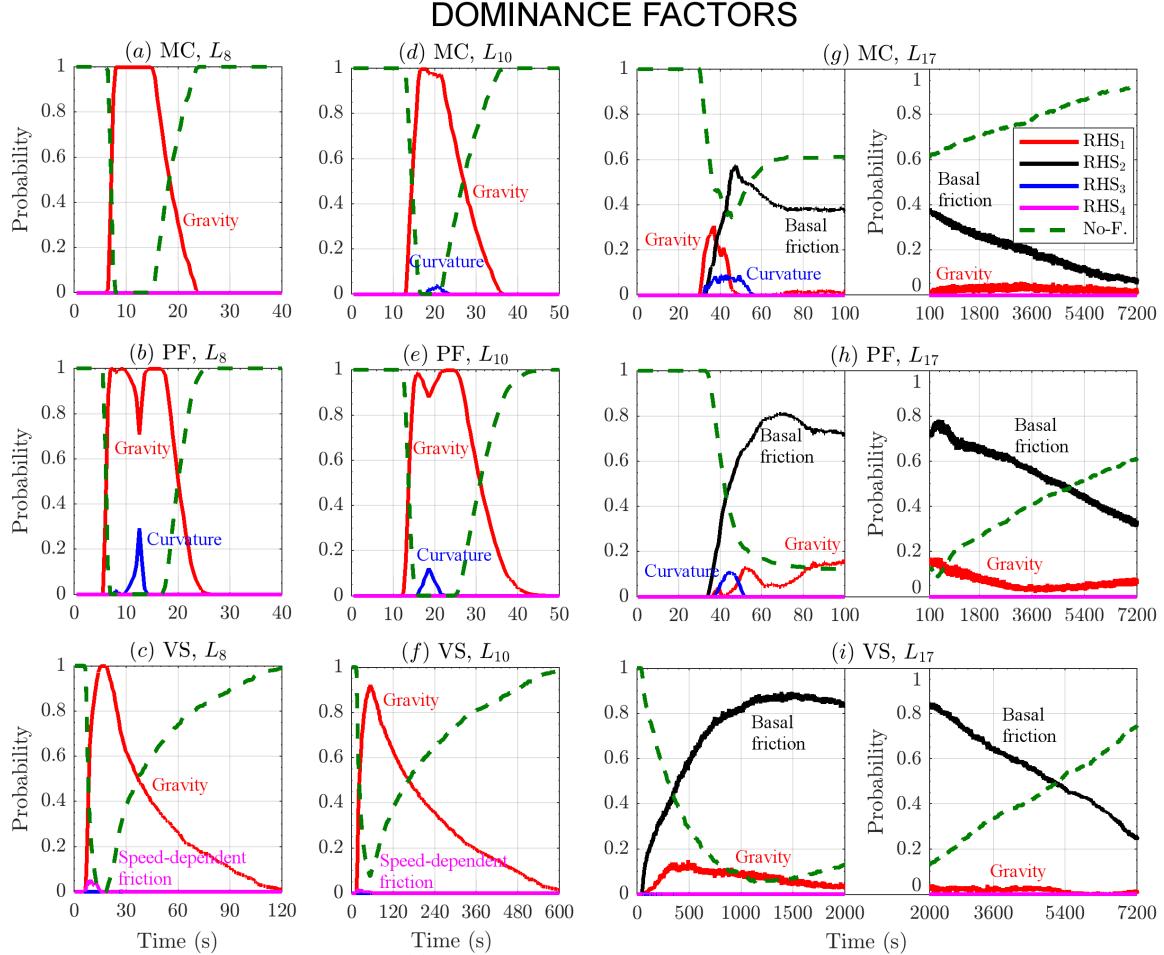


Figure 15: Dominance factors of **RHS** forces in three locations in the first km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms. No-flow probability is displayed with a green dashed line.

The plots 15a,b,c and 15d,e,f are related to point  $L_8$  and  $L_{10}$ , respectively. They are significantly similar. The gravitational force **RHS<sub>1</sub>** is the dominant variable with a very high chance,  $P_1 > 90\%$ . In MC and PF there is a small probability,  $P_3 = 5\% - 30\%$ , of **RHS<sub>3</sub>** being the dominant variable for  $\sim 5s$ . In VS it is observed a  $P_4 = 5\%$  chance of **RHS<sub>4</sub>** being dominant, just for a few seconds. Plots 15g,h,i concern the relatively more distal point  $L_{17}$ . They are split in two sub-frames at different time scale. In all the models, **RHS<sub>2</sub>** is the most probable dominant variable, and its dominance factor has a bell-shaped profile. In all the models, also **RHS<sub>1</sub>** has a small chance of being the dominant variable. In MC this chance is more significant, at most  $P_1 = 30\%$  for  $\sim 20s$ , and again  $P_1 = 5\%$  in  $[100, 7200]s$ . In PF  $P_1 = 15\%$  in two peaks, one short lasting at about  $55s$ , and the second extending in  $[100, 500]s$ . Also in VS,  $P_1 = 15\%$  at  $[300, 500]s$ . Its profile is unimodal in time and becomes  $P_1 < 2\%$  after 2000s. In MC and PF, **RHS<sub>3</sub>** has a dominance factor  $P_3 = 10\%$  at  $[30, 50]s$  and  $[40, 50]s$ , respectively.

Figure 16 shows the corresponding expected contributions  $\mathbb{E}[C_i]_{i=1,\dots,4}$ . The contributions in points  $L_8$  and  $L_{10}$  are shown in 16a,b,c and 16d,e,f, respectively. The plots related to the same model are similar. In all the models  $C_1$  is significantly larger than  $C_2$  and  $C_3$ , which are almost equivalent in MC and PF, while  $C_2 > C_3$  in VS.  $C_4$  always gives a negligible contribution, except in VS, where it is comparable to  $C_2$ . In  $L_8$ , following PF,  $C_3$  is bimodal, whereas it is unimodal in MC and VS. This is not observed in  $L_{10}$ . In  $L_8$ ,  $C_3$  is greater than in  $L_{10}$ , compared to the other forces. VS always shows a slower decrease of the plots. In plots 16g,h,i are shown the expected contributions in  $L_{17}$ . The plots are split in two sub-frames at different time scale. Initial dynamics is dominated by  $C_2$ , except for in MC, and only for a short time, [30, 35]s. In MC there is an initial peak of  $C_2$  which is not observed in the other models.  $C_3$  has a significant size, in MC and PF, and unimodal profile. In PF, after  $C_3$  wanes, at about 60s also  $C_4$  becomes not negligible for  $\sim 40$ s. The second part of the temporal domain is characterized by a slow decrease of  $C_2 > C_1$ .

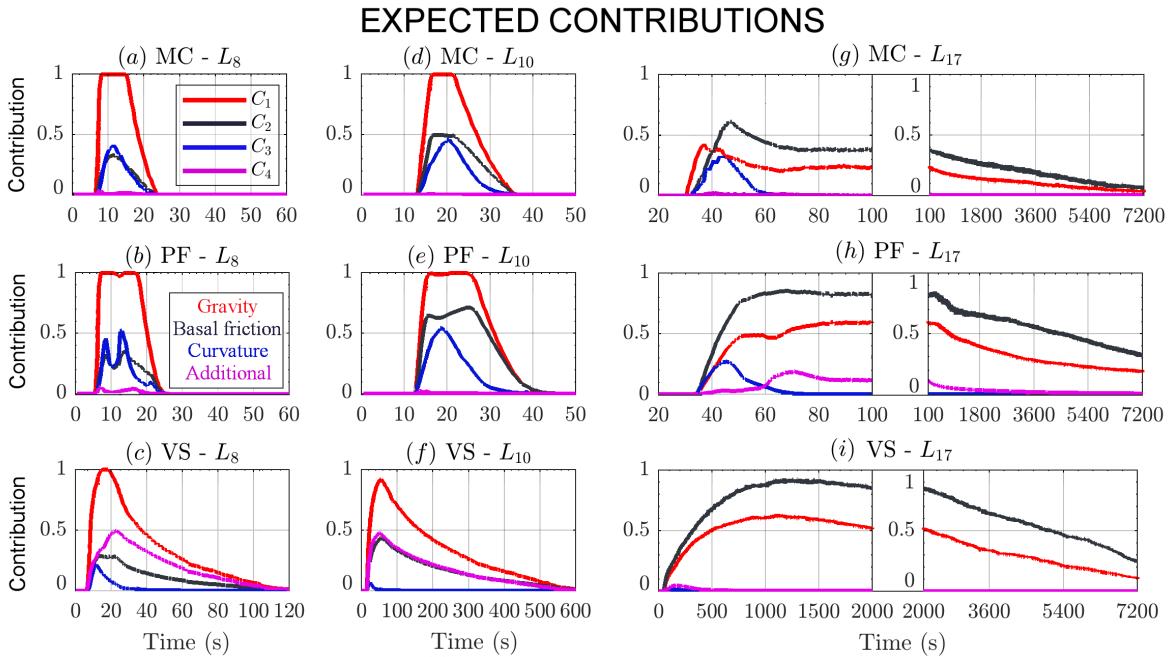


Figure 16: Expected contributions of **RHS** forces in three locations in the first km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms.

#### 5.4.2 Three locations distal from the initial pile

Figure 17 shows the dominance factors  $(P_i)_{i=1,\dots,4}$  in the three distal points  $L_{39}$ ,  $L_{43}$ , and  $L_{46}$ , all more than 2 km far from the initial pile. Plots 17a,b,c and  $L_{39}$  are dominated by **RHS<sub>1</sub>**. In all the models  $P_1$  is increasing and  $P_1 > 90\%$  at the end of the simulation. In MC,  $P_1$  shows a plateau at  $\sim 40\%$  in [90, 2000]s preceded and followed by steep increases, while in the other models it rises gradually.  $P_2 > 0$  after  $\sim 500$ s and 3600s, respectively, but is never greater than 2%. In MC  $P_3 = \sim 10\%$  at [50, 70]s. No-flow probability becomes zero in PF and VS, while stops at 20% in MC. Plots 17d,e,f are related to point  $L_{43}$ , and are remarkably complex. In MC, either  $P_1$  and  $P_2$  are  $\sim 35\%$  in the first 200s. Then,  $P_2$  increases, and **RHS<sub>2</sub>** becomes the only dominant variable after 3600s. The no-flow probability is never below 30%.  $P_3 = 35\%$  in [40, 60]s. Instead, in PF  $P_1 > 90\%$  until 3600s, and  $P_2$  rises only in the very last amount of time, reaching  $P_2 = P_1 = 40\%$ . The no-flow probability is very low during the most of the temporal window, rising at 20% only at 7200s. Both  $P_3$  and  $P_4$  show short peaks,  $\sim 10\%$ , at [50, 60]s. In VS the no-flow probability is never below 20%, and the dominance factors are broadly equivalent to MC, although  $P_1$  is the greatest up to 4000s, and  $P_3 \equiv 0$ . Plots 17g,h,i are related to point  $L_{46}$  and they

are similar to those recorded at point  $L_{17}$ , but  $P_2 > 90\%$  and the no-flow probability decreases to zero in the second half of the simulation. Moreover, in all the models  $P_1$  does not show any initial peak and instead increases slowly, reaching  $P_1 = 10\%$  after more than 3600s.

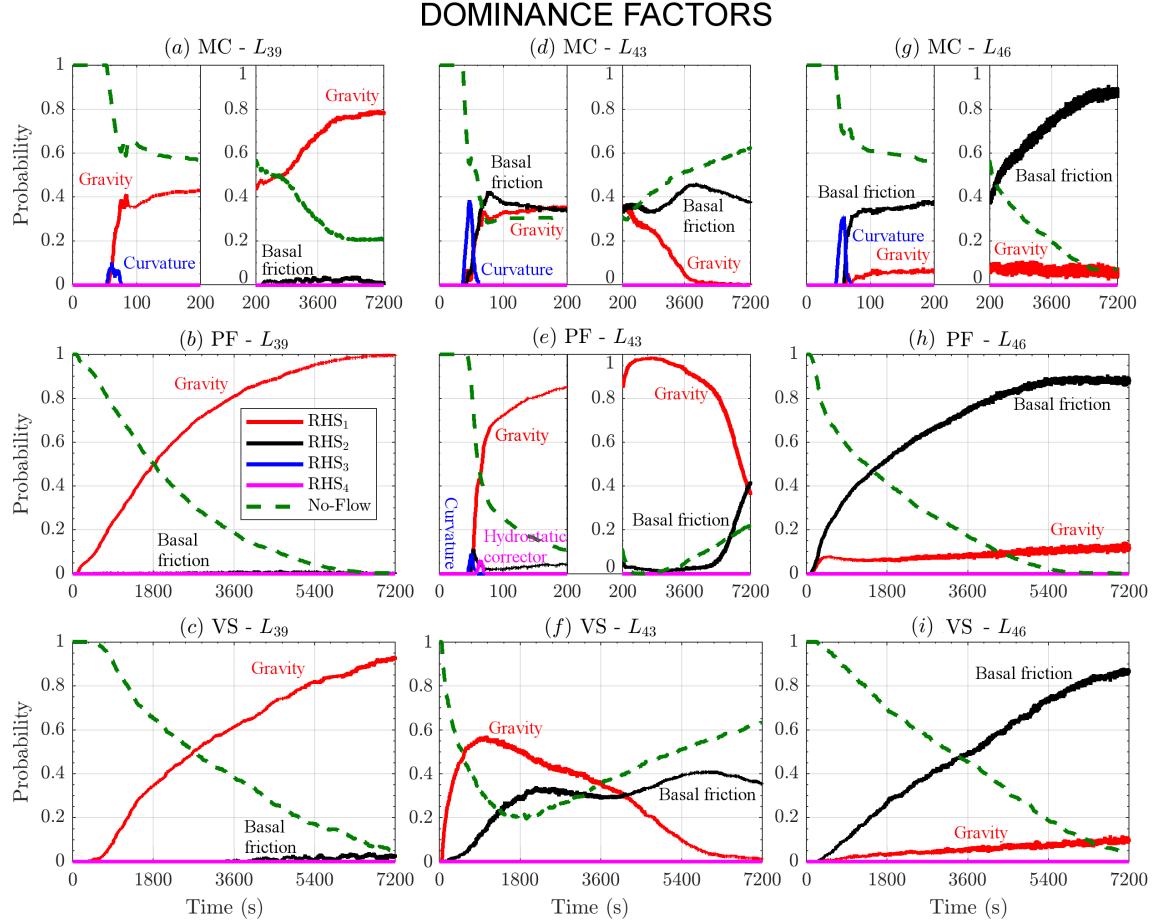


Figure 17: Dominance factors of **RHS** forces in three locations after 2 km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms. No-flow probability is displayed with a green dashed line.

Figure 18 shows the expected contributions in the distal points. All the plots are dominated by  $C_1$  and  $C_2$ , and the remarkable differences observed in the dominance factors depend on which contribution is the greatest. In general these two contributions have similar profiles. Plots 18a,b,c are related to point  $L_{39}$  and  $C_1 > C_2$ . In MC, also is  $C_3 > 0$  for a short time. In MC and PF also  $C_4 > 0$ , but it is significantly lower than the previous contributions, almost negligible in MC. Plots 18d,e,f concern point  $L_{43}$ . In MC  $C_1 < C_2$ , in PF  $C_1 > C_2$ , in VS they  $C_1$  decreases and crosses  $C_2$  at  $\sim 3600$ s. The two contributions form a plateau in MC, in  $[90, 200]$ s. In MC and PF  $C_3 > 0$  for a few seconds, and also  $C_4 > 0$  with an initial spike at  $\sim 60$ s. In PF it shows a long lasting plateau, while becomes negligible in MC. Plots 18g,h,i are related to point  $L_{46}$ . In all the models  $C_1 < C_2$ , and these force contributions are monotone increasing. Only in MC  $C_3 > 0$  shortly, and  $C_4 > 0$ , but almost negligible.

## EXPECTED CONTRIBUTIONS

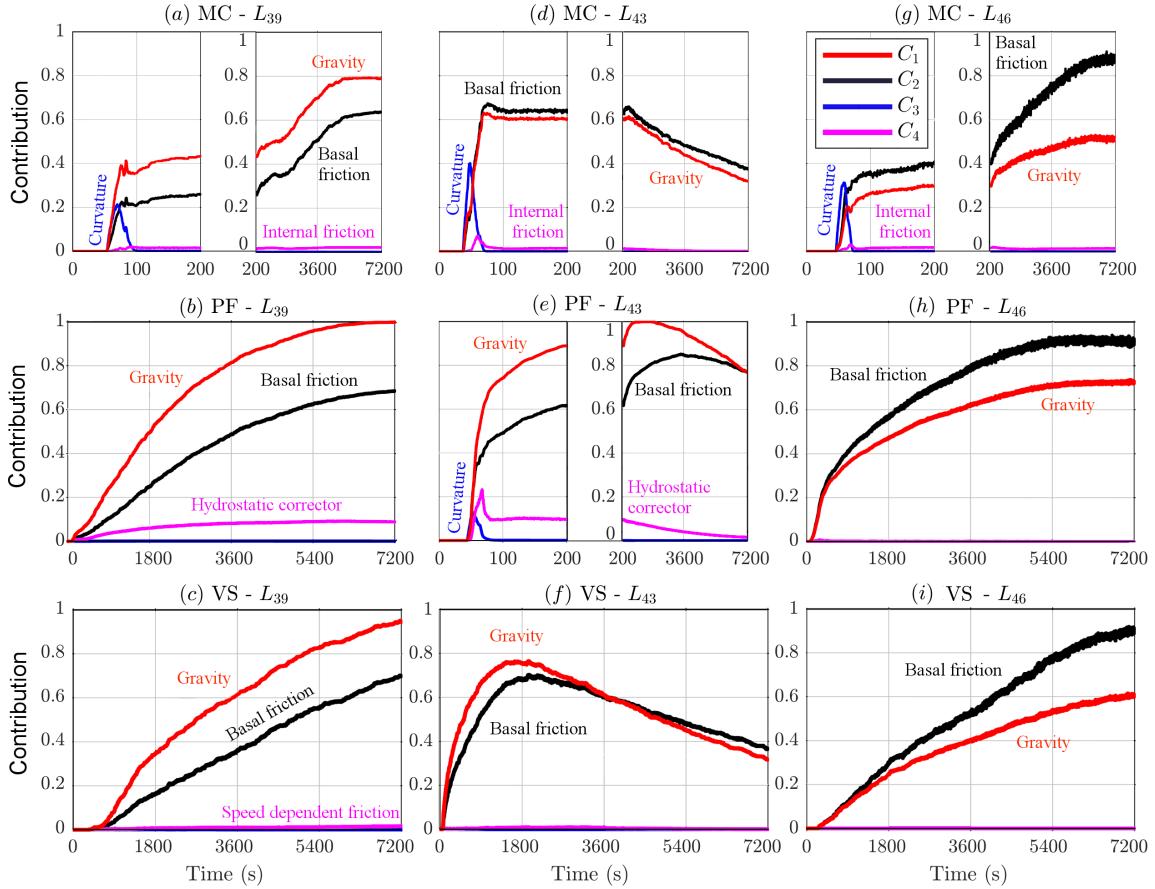


Figure 18: Expected contributions of **RHS** forces in three locations after 2 km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms.

### 5.5 Flow extent and spatial integrals

Figure 19 shows the volumetric average of speed and Froude Number. It also shows the inundated area as a function of time. Like in Fig.7, spatial averages and inundated area have smoother plots than local measurements, and most of the details observed in local measurements are not easy to discern. In plot 19a the speed shows a bell-shaped profile in all the models, but whereas the all values were close in the inclined plane experiment, in this case the maximum speed is  $\sim 60m/s$  in MC,  $\sim 50m/s$  in PF,  $\sim 20m/s$  in VS, on average. Uncertainty is  $\pm 18m/s$  in MC, similar, but skewed, in VS, and  $\pm 10m/s$  in PF. In plot 19b, the  $Fr$  profile is very similar to the speed, but the difference between VS and the other models is accentuated. Maximum values are  $\sim 50$  in MC,  $\sim 38$  in PF,  $\sim 5$  in VS, whereas uncertainty is  $\pm 10$  in MC,  $\pm 7$  in PF, and skewed  $[-5, +10]$  in VS. In plot 19c, inundated area has a first peak in MC and PF, both at  $\sim 1.15km^2$ , followed by a decrease to  $0.55km^2$  and  $0.7km^2$ , respectively, and then a slower increase up to a flat plateau at  $0.9km^2$  and  $1.5km^2$ , respectively. Uncertainty is  $\sim \pm 0.2km^2$  in both MC and PF until  $\sim 100s$ , and then it increases at  $\pm 0.3km^2$  and  $[-0.5, +0.4]km^2$ , respectively. In MC this increase in uncertainty is concentrated at  $\sim 100s$ , while it is more gradual in PF. VS has a different profile. The initial peak is only significant in the 95<sup>th</sup> percentile values, and occurs later, at  $\sim 100s$ . The peak is of  $\sim 1km^2$  on the average, but up to  $\sim 1.8km^2$  in the 95<sup>th</sup> percentile. The decrease after the peak is very slow and the average inundated area never goes below  $0.85km^2$ , and eventually reaches back to  $\sim 1km^2$ . Uncertainty is  $[-0.3, +0.2]km^2$ .

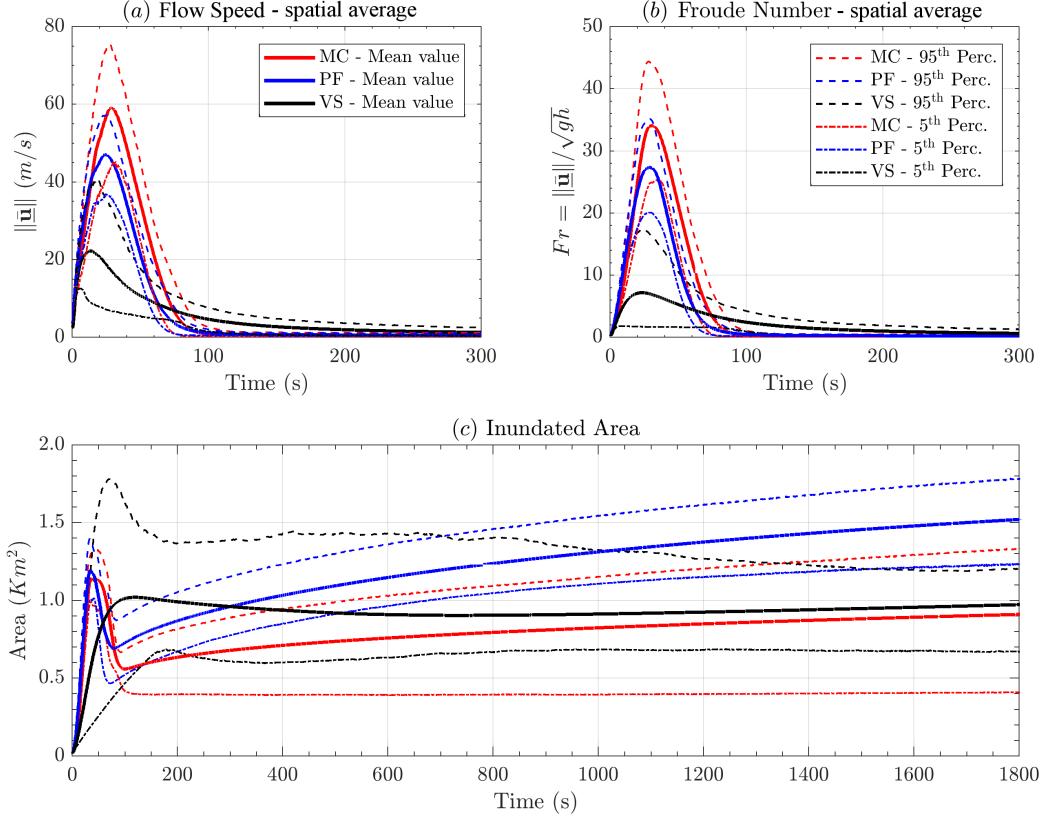


Figure 19: Comparison between spatial averages of (a) flow speed, and (b) Froude Number, in addition to the (c) inundated area, as a function of time. Bold line is mean value, dashed/dotted lines are 5<sup>th</sup> and 95<sup>th</sup> percentile bounds. Different models are displayed with different colors.

## 5.6 Power integrals

Figure 20 shows the spatial sum of the powers. The estimates in this section assume  $\rho = 1800 \text{ kg/m}^3$  as a constant scaling factor. Corresponding plots of the force terms are included in Supporting Information S6. The scalar product of force with velocity imposes the bell-shaped profile already observed in Fig. 8a. In plot 20a the power of  $\mathbf{RHS}_1$  starts from zero and rises up to  $\sim 1.4e11 \text{ W}$  in MC,  $\sim 1.2e11 \text{ W}$  in PF,  $\sim 6.5e10 \text{ W}$  in VS. Uncertainty is  $\pm 4e10 \text{ W}$  in MC,  $\pm 3e10 \text{ W}$  in PF,  $[-4e10, +5e10] \text{ W}$  in VS. The decrease of gravitational power is related to the slope reduction, and this decrease is more gradual in VS than in the other models. In plot 20b the power of  $\mathbf{RHS}_2$  is always negative and peaks to  $\sim -6.5e10 \text{ W}$  in MC,  $\sim -5e10 \text{ W}$  in PF,  $\sim -2e10 \text{ W}$  in VS. In VS this dissipative power is significantly more flat than in the other models. MC and PF show negligible powers after  $\sim 100 \text{ s}$ , VS after  $\sim 200 \text{ s}$ . Uncertainty is  $\pm 2e10 \text{ W}$  in MC,  $\pm 1.5e10 \text{ W}$  in PF,  $[-2e10, +1e10] \text{ W}$  in VS. In PF, the plot starts from stronger values than in the other models, but it is also the faster to wane. In plot 20c the power of  $\mathbf{RHS}_3$  shows a negative peak at  $\sim -7e10 \text{ W}$  in MC,  $\sim -4.5e10 \text{ W}$  in PF,  $\sim -5e9 \text{ W}$  in VS. Uncertainty on the peak value is  $[-4.5e10, +3.5e10] \text{ W}$  in MC,  $[-2.5e10, +2e10] \text{ W}$  in PF,  $[-1e10, +5e9] \text{ W}$  in VS. The three models all show a bell-shaped profile, MC and PF waning to zero at 90s, VS at  $\sim 30 \text{ s}$ . In plot 20d the power of  $\mathbf{RHS}_4$  has a different meaning in the three models. In MC it is the internal friction term, and it only has almost negligible ripple visible in the first second. In PF it is a depth averaged correction in the hydrostatic pressure, and has an almost negligible effect only during the first second of simulation, at  $5e9 \text{ W}$ . It becomes null at  $\sim 10 \text{ s}$ . In VS, instead, it is a speed dependent term, and has a very relevant effect. The plot shows a bell-shaped profile, with a peak of  $\sim -3.5e10 \text{ W}$ ,  $[-2e10, +1e10] \text{ W}$ . After that, this dissipative power gradually decreases, and becomes negligible at 200s.

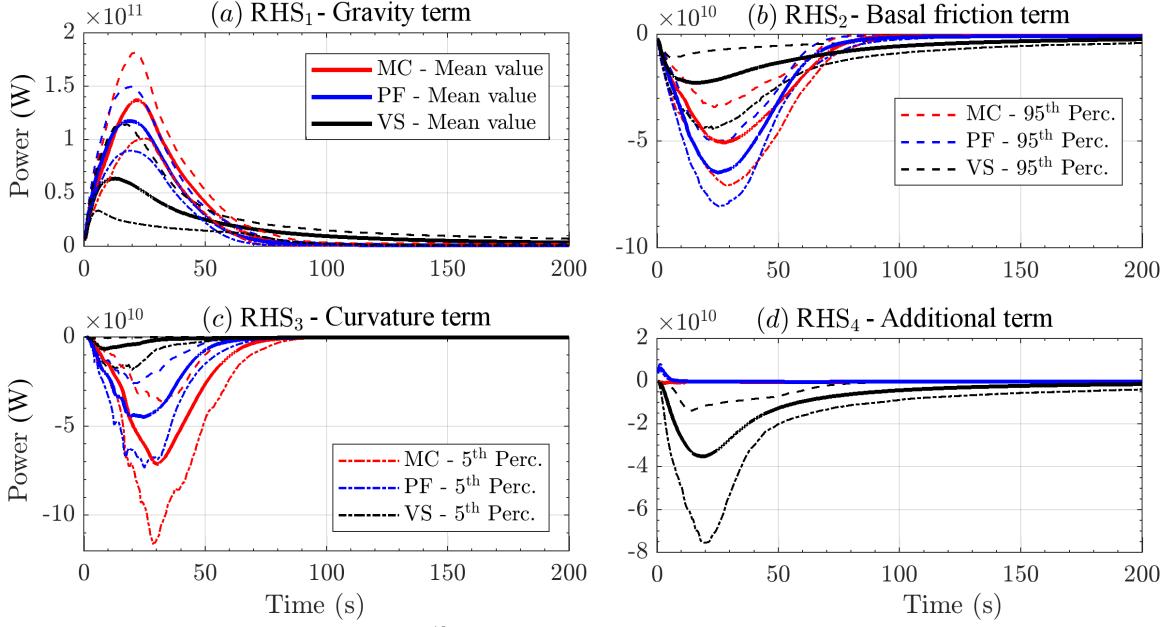


Figure 20: Spatial integrals of the RHS powers. Bold line is mean value, dashed lines are 5<sup>th</sup> and 95<sup>th</sup> percentile bounds. Model comparison on the mean value is also displayed. Different models are displayed with different colors.

## 6 Discussion on the comparative performance of geophysical flow models and latent variables

The statistically driven method introduced in this study for analyzing complex models provides extensive and quantitative information. Geophysical flow modeling usually compares simulation to observation, and fits the model parameters coping with inverse problems. Nevertheless, this is not always sufficient to solve forecasting problems, in which the range of possible flows might not be limited to a single type and scale of flow. Our approach is different, and evaluates the statistics of a range of flows, produced by the couple  $(M, P_M)$  - i.e. a model  $M$  and a probability distribution for its parameters  $P_M$ .

New quantitative information can solve classical qualitative problems, either model-model or model-observation comparison. The mean plot represents the average behavior of flows in the considered range, and provides the same type of data that is provided by a single simulation. Moreover, the uncertainty ranges stem additional pieces of information that often exalt the differences between the models.

In the following, we summarize first the general features which differ between the model outputs. This is a traditional type of analysis, but performed by us in a new probabilistic framework that is likely more useful for extrapolating and forecasting. After that, we describe and compare the general features of latent variables. This is a new type of analysis enabled by our approach and allows us to evaluate modeling assumptions and their relative importance. Finally, we give an example of model performance evaluation of the couple  $(M, P_M)$  according to a specific observation.

### 6.1 Characteristic features and their motivations

The different features in the models macroscopically affect the shape of the inundated area, both in the small flow on the inclined plane, and in the large scale geophysical flow. In Fig. 2b the flow runout displays a larger lateral extent in PF, and a bent shape in VS - the lateral wings remain behind the central section of the flow. Also in Fig. 10, the three models look clearly different. Even if the maximum runout is matched between the models, MC displays a further distal spread before entering the ravines, PF shows a larger angle of lateral spread at the initiation pile, VS is less laterally extended and generally

looks significantly channelized, and displays several not-inundated spots due to minor topographical coulées.

### 6.1.1 Flow height and acceleration

Flow height gives additional insight on the model features. In Fig. 3, MC is more distally stretched, but starts to deposit material earlier and closer to the initial pile compared to the other models. PF height is generally shorter, and displays a small temporal anticipation in its arrival at the sample points. These features are probably due to the hydrostatic correction term, which additionally pushes forward the material during the pile collapse. A linear cut in the flow height profile of PF is also observed when the flow thins on the slope. That is probably generated by the interpolation between the two basal friction angles as a function of flow height and speed. VS tends to be higher than the other models, if observed at the same instant, because of the reduced lateral spreading of material.

In Fig. 14 the flow height plots allow us to classify the local flow regimes according to their similarity with what observed in the four sample points of the small scale case study. There is a new feature which was not present in the small scale flow - VS is temporally stretched, and material arrives later and stays longer in all the sample points. This is a consequence of the speed dependent term reducing flow velocity.

Flow acceleration brings our analysis from the mechanics to the dynamics of the flow. In Fig. 4, at the point on the slope, MC has a flat plateau, while PF is linearly decreasing, and VS is bell shaped. These differences are a consequence of the assumptions behind the models - double bed friction angle in PF, and speed dependent term in VS. At the slope change point, VS and MC display a bimodal profile in the acceleration. This is not a statistical effect, and it is also observed in single simulations. The first maximum is when the head of the flow hits the ground, while the second maximum is when the accumulating material in the tail arrives there. In VS the maxima are equal, because the tail is not laterally spread and hits the ground compactly. In contrast, PF does not show such a second peak, due to the accentuated lateral spreading in the tail. On the flat runway, the average acceleration in PF is lower than in the other models, probably because the material stops more suddenly.

### 6.1.2 Statistical analysis of latent variables

The new concepts of dominance factors and expected contributions are adopted in the study of the latent variables. We focus on the force decomposition, but a similar procedure can be applied to any set of latent variables. The dominance factors illustrate the predominant force term through time, while the expected contributions provide a full comparison of the mean force terms.

In the inclined plane case study, Fig. 5 shows minor differences in the dominance factors between the models. In general, there is always a single dominant variable, and its profile is complementary with the no-flow probability. In the slope change point the differences between the models are more significant. Curvature term dominance probability is bimodal in MC and VS, and in VS the two peaks are equivalent. On the flat runway, in PF the hydrostatic correction can be the dominant variable with a small chance. In Fig. 6 the expected contributions illustrate the relative scale of all the forces, and show that the second strongest force is never above 60% of the dominant.

In the Volcán de Colima case study, the dominance factors and expected contributions are more complex. In the proximal points to the initiation, Fig. 15, gravitational force is dominant with a very high chance until the no-flow probability becomes large. In MC and PF curvature related forces can also be dominant for a short time. In VS gravitational force is dominant for a larger time span than in the other models, because of the longer presence of the flow. The speed dependent friction can be dominant with a small probability at the beginning of the dynamics. The expected contributions in Fig. 16 confirm these features.

In the distal points, 17, the gravity or the basal friction are dominant with high probability. Some of the points have a deposit at the end with a high chance, some other not, depending on the slope. In general, in MC the no-flow probability tends to be larger than in the other models, because some flow samples stops earlier, or completely leaves the site. Again, curvature can have a small chance to be dominant in MC and PF, particularly when the speed is high. The expected contributions, Fig. 18, show that the remarkable diversity between the different locations is a consequence of small unbalance between gravity and basal friction.

Point  $L_{43}$  is particularly interesting. It is not proximal to the initiation, but the no-flow probability is increasing at the end, meaning that all the material tends to leave the site. Moreover, the dominating

force can be the gravity or the basal friction depending on the time and the model. In MC and VS both the two forces have similar chances to be dominant for the most of the time of the simulation. In PF, only the gravitational force is dominant with a high chance. This is probably because point  $L_{43}$  is situated downhill of a place where a significant amount of material stops.

### 6.1.3 Flow extent and spatial integrals

In Fig. 7 the spatially averaged speed and Froude Number are significantly similar between the models. The differing features appear to be mostly localized in space. However, VS is confirmed to be significantly slower than the other models after the initial collapse. Moreover, it is the only model which presents a significant amount of long lasting and slowly moving material. Inundated area in PF has a greater maximum value, because of the accentuated lateral spread. In VS the inundated area almost does not decrease from its peak, because of the strictly increasing lateral spread. Vice versa, lateral spread in MC and PF has a temporary stop when the bulk of the flow hits the ground. This is a consequence of the interplay of accumulating material and the push of new material, which is stronger in the middle than in the lateral wings. In PF lateral extent weakly decreases for a short time. The hydrostatic correction term may generate the pull reducing the lateral extent. In Fig. 19 average speed and Froude Number display that VS slower speed is not a local feature. Inundated area is again significantly larger in PF than in the other models.

### 6.1.4 Power integrals

Figure 8 and 20 produce a clear explanation of the source (gravity) and the dissipation of power. In both cases the dissipation lags the source (gravity) slightly as expected and additional terms  $RHS_4$  only has a major impact in VS. Similarly the curvature based term  $RHS_2$  has a minimal impact on VS.

Spatial integrals of power terms have several features in common with the corresponding forces, and provide a decomposition of the acceleration sources. Main dissimilarity between forces and powers is that gravitational and basal friction powers have a profile starting from zero when the flow initiates, because the flow speed starts from zero. In Fig. 8 the differences between models are particularly relevant in term  $\mathbf{RHS}_4$ . Speed dependent power in VS is at least one order of magnitude larger than the maximum values of the corresponding terms in MC and PF. Those are decreasing to zero after a short time from the initiation. Hydrostatic correction in PF is clearly positive in the speed direction, and hence contributing to push the flow ahead. The effects of internal friction in MC are almost negligible, and initially positive, then negative. This is motivated by an initial compression of the material during the pile collapse, followed by its stretching. It is worth remarking that  $\mathbf{RHS}_2$  and  $\mathbf{RHS}_3$  are both smaller in VS, due to the lower basal friction angles involved. In Fig. 20, the differences are accentuated, because of the topography complexity. Gravity term is larger in VS, because a portion of the flow lingers on the higher slopes for a long time. Basal friction has a higher peak in PF compared to the other models, due to the interpolation of the two basal friction angles.

## 6.2 Example of model performance calculation

Analysis of model suitability have been conducted. In past work (Patra et al., 2005), MC rheology was tuned to match deposits for known block and ash flows, but *a priori* predictive ability was limited by inability to tune without knowledge of flow character. The new procedure developed in this study enables an enhanced quantification of model performance, i.e. the similarity of the outputs and real data. We remark that the measured performance refers to the couple  $(M, P_M)$ , and that different parameter ranges can produce different performances. Our example concerns the Volcán de Colima case study, and in particular we compare the inundated region in our simulations to the deposit of a BAF occurred 16 April 1991 (Saucedo et al., 2004; Rupp, 2004; Rupp et al., 2006). The inundated region is defined as the points in which the maximum flow height  $H$  is greater than 10cm. A similar procedure may be applied to any observable variable produced by the models, if specific data become available. Let  $\mathcal{M} : \mathcal{P}(\mathbb{R}^2) \rightarrow [0, 1]$  be a similarity index defined on the subsets of the real plane. An equivalent definition can be based on the pseudo-metric  $1 - \mathcal{M}$ .

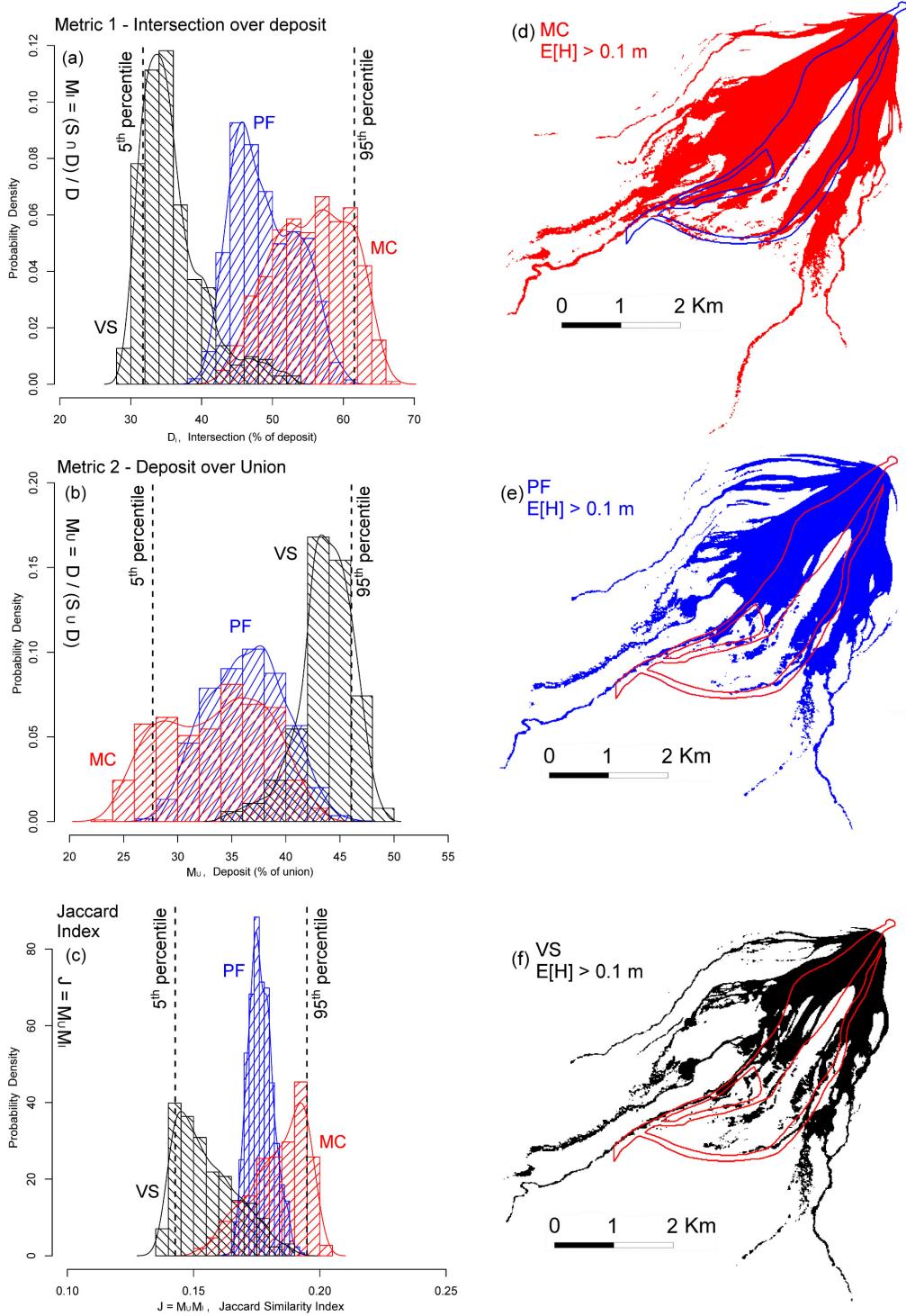


Figure 21: Volcán de Colima. Pdf of the similarity index of inundated regions and a real BAF deposit. (a) is based on  $\mathcal{M}_I$ , (b) on  $\mathcal{M}_U$ , (c) on  $\mathcal{J}$ . The models are MC (red), PF (blue), VS (black). Data histograms are displayed in the background. Global 5<sup>th</sup> and 95<sup>th</sup> percentile values are indicated with dashed lines. Plots (d,e,f) display the average inundated region  $\{\mathbf{x} : E[H(\mathbf{x})] > 10\text{cm}\}$ . The boundary of the real deposit is marked with a colored line.

For example, we define

$$\mathcal{M}_I := \frac{\int_{\mathbb{R}^2} 1_{S \cap D}(\mathbf{x}) d\mathbf{x}}{\int_{\mathbb{R}^2} 1_D(\mathbf{x}) d\mathbf{x}}, \quad \mathcal{M}_U := \frac{\int_{\mathbb{R}^2} 1_D(\mathbf{x}) d\mathbf{x}}{\int_{\mathbb{R}^2} 1_{S \cup D}(\mathbf{x}) d\mathbf{x}}, \quad \mathcal{J} := \mathcal{M}_I \cdot \mathcal{M}_U,$$

where  $S \subset \mathbb{R}^2$  is the inundated region, and  $D \subset \mathbb{R}^2$  is the recorded deposit. In particular,  $\mathcal{M}_I$  is the area of the intersection of inundated region and deposit over the area of the deposit,  $\mathcal{M}_U$  is the area of the deposit over the area of the union of inundated region and deposit,  $\mathcal{J}$  is the product of the previous, also called Jaccard Index (Jaccard, 1901).

Figure 21 shows the probability distribution of the similarity indices, according to the uniform probability  $P_M$  on the parameter ranges defined in this study. Different metrics can produce different performance estimates, for example MC inundates most of the deposit, but overestimates the inundated region, while VS relatively reduces the inundated region outside of the deposit boundary, but also leaves several not-inundated spots inside it.

Let  $g : [a, b] \rightarrow [0, 1]$  be a score function defined over the percentile range of the similarity index. The global 5<sup>th</sup> and 95<sup>th</sup> percentile values  $[a, b]$  are defined assuming to select the model randomly with equal chance, and are also shown in Fig.21a,b,c. Then the performance score  $G_g$  of model  $(M, P_M)$  is defined as:

$$G_g(M, P_M) = \int_{[a, b]} g(x) df_M(x),$$

where  $f_M$  is the pdf related to the model. Possible score functions include a step function at the global median, a linear or quadratic function, a sigmoid function. Supporting Information S7 reports a Table of alternative performance scores, according to changing similarity indices and score functions.

## 7 Conclusions

In this study, we have introduced a new statistically driven method for analyzing complex models. We have used three different models arising from different rheology assumptions. The data shows unambiguously the performance of the models across a wide range of possible flow regimes and topographies. The analysis of latent variables is particularly illustrative of the impact of modeling assumptions. Knowledge of which assumptions dominate, and, by how much, will allow us to construct efficient models for desired inputs. Such model composition is the subject of ongoing and future work, with the purpose to bypass the search for a unique best model, and yet to go beyond a simple mixture of alternative models.

In summary, our new method enabled us to break down the effects of the different physical assumptions in the dynamics, providing an improved understanding of what characterizes each model. The procedure was applied to two different case studies: a small scale inclined plane with a flat runway, and a large scale DEM on the SW lobe of Volcán de Colima (MX). In particular, we presented:

- a short review of the assumptions characterizing three commonly used rheologies of Mohr-Coulomb, Pouliquen-Forterre, Voellmy-Salm. This included a qualitative list of such assumptions, and a breaking down of the different terms in their differential equations.
- a new statistical framework, processing the mean and the uncertainty range of either observable or latent variables in the simulation. The new concepts of dominance factors and expected contributions enabled a simplified description of the local dynamics. These quantities were analyzed at selected sites, and spatial integrals were also performed, illustrating the characteristics of the entire flow.
- a final discussion, explaining all the observed features in the results on the light of the known physical assumptions of the models, and the evolving flow regime in space and time. This included an example of model performance estimation method, depending on the metric and the cost function adopted.

Our statistical analysis based on UQ depicted the following main features:

- Compared to a classical MC model, PF lacks of internal friction and this produces an accentuated lateral spread. This is increased by the hydrostatic correction, which briefly pushes the flow ahead and laterally during the initial collapse. That force can also have some minor effects in the final deposit accumulation. The interpolation of the smaller bed friction angle  $\phi_1$  with the larger value

$\phi_2$ , suddenly stops the flow if it thin compared to its speed. This mechanism does not allow for large speed peaks.

- Instead in VS, the speed dependent friction has a great effect in reducing lateral spread and producing channeling features even due to minor ridges in topography. The flow tends to be significantly slower and more stretched in the slope direction. The effects of the different formulation of the curvature term are less impacting than the effects of the lower basal friction and speed.

Additional research concerning other case studies, and different parameter ranges, might reveal other flow regimes, and hence differences in the consequences of the modeling assumptions under new circumstances.

## A Latin Hypercubes and orthogonal arrays

The Latin Hypercube Sampling (LHS) is a well established procedure for defining pseudo-random designs of samples in  $\mathbb{R}^d$ , with good properties with respect to the uniform probability distribution on an hypercube  $[0, 1]^d$  (McKay et al., 1979; Owen, 1992a; Stein, 1987; Ranjan and Spencer, 2014; Ai et al., 2016). In particular, compared to a random sampling, a LHS: (i) enhances the capability to fill the  $d$ -dimensional space with a finite number of points, (ii) in case  $d > 1$ , avoids the overlapping of point locations in the one dimensional projections, (iii) reduces the dependence of the number of points necessary on the dimensionality  $d$ .

**Definition 3 (Latin hypercube sampling)** Let  $\Xi = \{\xi_i : i = 1, \dots, N\}$  be a set of points inside the  $d$ -dimensional hypercube  $C = [0, 1]^d$ . Let  $[0, 1] = \bigcup_{j=1}^N I_j$ , where  $I_j = [\frac{(j-1)}{N}, \frac{j}{N}]$ . Let  $\xi_i = (\xi_i^1, \dots, \xi_i^d)$ , and for each  $k \in \{1, \dots, d\}$ , let  $\Xi^k = \{\xi_i^k : i = 1, \dots, N\}$ . Let  $\lambda^d$  be the uniform probability measure supported inside  $C$ , called Lebesgue measure. Then  $\Xi$  is a latin hypercube w.r.t.  $\lambda^d \iff \forall j \in \{1, \dots, N\}, \forall k \in \{1, \dots, d\}, |I_j \cap \Xi^k| = 1$ .

The procedure is simple: once the desired number of samples  $N \in \mathbb{N}$  is selected, and  $[0, 1]$  is divided in  $N$  equal bins, then each bin will contain one and only one projection of the samples over every coordinate. The LHS definition is trivially generalized over  $C = \prod_i^d [a_i, b_i]$ , i.e. the cartesian product of  $d$  arbitrary intervals. That will be applied in this study, defining LHS over the parameter domain of the flow models.

There are a large number of possible designs, corresponding the number of permutations of the bins in the  $d$ -projections, i.e.  $d \cdot N!$ . If the permutations are randomly sampled there is a high possibility that the design will have good properties. However, this is not assured, and clusters of points or regions of void space may be observed in  $C$ . For this reason, we base our design on the orthogonal arrays (OA) (Owen, 1992b; Tang, 1993).

**Definition 4 (Orthogonal arrays)** Let  $S = \{1, \dots, s\}$ , where  $s \geq 2$ . Let  $Q \in S^{n \times m}$  be a matrix of such integer values. Then  $Q$  is called an OA( $n, m, s, r$ )  $\iff$  each  $n \times r$  submatrix of  $Q$  contains all possible  $1 \times r$  row vectors with the same frequency  $\lambda = n/s^r$ , which is called the index of the array. In particular,  $r$  is called the strength,  $n$  the size, ( $m \geq r$ ) the constraints, and  $s$  the levels of the array.

Orthogonal arrays are very useful for defining latin hypercubes which are also forced to fill the space (or its  $r$ -dimensional subspaces) in a more robust way, at the cost of potentially requiring a larger number of points than a traditional LHS.

**Proposition 5** Let  $Q$  be an OA( $n, m, s, r$ ). Then let  $U \in \mathbb{R}^{n \times m}$  be defined as follows:

$$\forall k \in \{1, \dots, s\}, \forall j \in \{1, \dots, m\}, \{Q[\cdot, j] : Q[i, j] = k\} = \Pi(\{(k-1)\lambda s^{r-1}, \dots, k\lambda s^{r-1}\}),$$

where  $\Pi$  is a random permutation of  $\lambda s^{r-1}$  elements. Then  $\Xi = \{\xi_i = U[i, \cdot] : i = 1, \dots, n\}$  is a LHS w.r.t to  $\lambda^m$  over  $C = [0, 1]^m$ . Moreover, let  $[0, 1]^r = \bigcup_{(h_1, \dots, h_r)=1}^s I_{(h_i)}$ , where  $I_{(h_i)} = \prod_i^r [\frac{(h_i-1)}{s}, \frac{h_i}{s}]$ . Then  $\forall D = (d_1, \dots, d_r) \subseteq \{1, \dots, m\}$ , let  $\Xi^D = \{(\xi^{d_1}, \dots, \xi^{d_r}) : i = 1, \dots, n\}$ . We have that

$$\forall k \in \{1, \dots, s\}, \forall (h_i : i = 1, \dots, r) \in \{1, \dots, d\}^r, |I_{(h_i)} \cap \Xi^D| = \lambda.$$

For each column of  $Q$  we are replacing the  $\lambda s^{r-1}$  elements with entry  $k$  by a random permutation of  $((k-1)\lambda s^{r-1} + h)_{h \in 1, \dots, \lambda s^{r-1}}$ . After the replacement procedure is done, the newly obtained matrix  $U$  is equivalent to a LHS which inherits from  $Q$  the property of fully covering  $s^r$  equal r-dimensional hypercubes in every r-dimensional projection. Each hypercube contains  $\lambda$  points. In other words, inside each r-dimensional projection, the design associated to  $U$  fills the space like a regular grid at the scale of those  $s^r$  hypercubes, but it is still an LHS at a finer scale, i.e. the  $\lambda s^{r-1}$  one dimensional bins. A complete proof can be found in (Tang, 1993) and it is a straightforward verification of the required properties.

However, even in an LHS based on an  $OA(n, m, s, r)$ , if  $r < m$  what happens in the projections with dimension  $r' > r$  is not controlled, and randomizing procedures are made more difficult by the additional structure imposed by the  $OA$ . Moreover, the total number of points necessary to achieve a full design increases with  $r$ , and hence is affected by dimensionality issues.

Dealing with relatively small  $d$ , i.e.  $d \in \{3, 4\}$ , we adopt a LHS  $U$  created by a  $OA(s^d, d, s, d)$ . The strength is equal to the dimension  $d$ , hence the design fills the entire space like a  $d$  dimensional grid, but it is a LHS as well. In this case there is one point in each hypercube, and  $\lambda = 1$ . We take  $s = 8$  for the 3-dimensional designs over the parameter space of Mohr-Coulomb and Voellmy-Salm models, i.e. 512 points; we took  $s = 6$  for the 4-dimensional designs over the more complex parameter space of the Pouliquen-Forterre model, i.e. 1296 points.

## B Conditional decomposition of expected contributions

Expected contributions are obtained after diving the force terms by the dominant variable  $\Phi$ , which is an unknown quantity depending on time, location, and input parameters. Thus we provide an additional result, further explaining the meaning of those contributions through the conditional expectation.

**Proposition 6** *Let  $(F_i)_{i \in I}$  be random variables on  $(\Omega, \mathcal{F}, P)$ . For each  $i$ , let  $C_i$  be the random contribution of  $F_i$ . Then we have the following expression:*

$$E[C_i] = \sum_j p_j \mathbb{E} \left[ \frac{F_i}{|F_j|} \mid \Phi = |F_j| \right],$$

where  $p_j := P\{\Phi = |F_j|\}$ .

**Proof.** Let  $Z$  be a discrete random variable such that, for each  $j \in \mathbb{N}$ ,  $(Z = j) \iff (\Phi = |F_j|)$ . Then, by the rule of chain expectation:

$$\begin{aligned} E[C_i] &= \mathbb{E} \left[ \frac{F_i}{\Phi} \right] = \mathbb{E} \left[ \mathbb{E} \left[ \frac{F_i}{\Phi} \mid Z = j \right] \right] = \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{F_i}{|F_j|} \mid Z = j \right] \right] = \sum_j P\{Z = j\} \mathbb{E} \left[ \frac{F_i}{|F_j|} \mid Z = j \right]. \end{aligned}$$

Moreover, by definition,  $p_j = P\{Z = j\}$ . This completes the proof.  $\square$

Consequently, we define the conditional contributions.

**Definition 7 (Conditional contributions)** *Let  $(F_i)_{i=1, \dots, k}$  be random variables on  $(\Omega, \mathcal{F}, P)$ . Then,  $\forall (i, j)$ , the conditional contribution  $C_{i,j}$  is defined as:*

$$C_{i,j} := \mathbb{E} \left[ \frac{F_i}{|F_j|} \mid \Phi = |F_j| \right],$$

where  $\Phi$  is the dominant variable.

## References

- Aghakhani, H., Dalbey, K., Salac, D., and Patra, A. K. (2016). Heuristic and Eulerian interface capturing approaches for shallow water type flow and application to granular flows. *Computer Methods in Applied Mechanics and Engineering*, 304:243–264.
- Ai, M., Kong, X., and Li, K. (2016). A general theory for orthogonal array based latin hypercube sampling. *Statistica Sinica*, 26:761–777.
- Bartelt, P. and McArdell, B. (2009). Granulometric investigations of snow avalanches. *Journal of Glaciology*, 55(193):829–833(5).
- Bartelt, P., Salm, B., and Gruber, U. (1999). Calculating dense-snow avalanche runout using a Voellmy-fluid model with active/passive longitudinal straining. *Journal of Glaciology*, 45(150):242–254.
- Bayarri, M. J., Berger, J. O., Calder, E. S., Dalbey, K., Lunagomez, S., Patra, A. K., Pitman, E. B., Spiller, E. T., and Wolpert, R. L. (2009). Using Statistical and Computer Models to Quantify Volcanic Hazards. *Technometrics*, 51(4):402–413.
- Bayarri, M. J., Berger, J. O., Calder, E. S., Patra, A. K., Pitman, E. B., Spiller, E. T., and Wolpert, R. L. (2015). Probabilistic Quantification of Hazards: A Methodology Using Small Ensembles of Physics-Based Simulations and Statistical Surrogates. *International Journal for Uncertainty Quantification*, 5(4):297–325.
- Bursik, M., Patra, A., Pitman, E., Nichita, C., Macias, J., Saucedo, R., and Girina, O. (2005). Advances in studies of dense volcanic granular flows. *Reports on Progress in Physics*, 68(2):271.
- Capra, L., Gavilanes-Ruiz, J. C., Bonasia, R., Saucedo-Giron, R., and Sulpizio, R. (2015). Re-assessing volcanic hazard zonation of Volcán de Colima, México. *Natural Hazards*, 76(1):41–61.
- Capra, L., Macís, J., Cortés, A., Dávila, N., Saucedo, R., Osorio-Ocampo, S., Arce, J., Gavilanes-Ruiz, J., Corona-Chávez, P., García-Sánchez, L., Sosa-Ceballos, G., and Vázquez, R. (2016). Preliminary report on the July 10–11, 2015 eruption at Volcán de Colima: Pyroclastic density currents with exceptional runouts and volume. *Journal of Volcanology and Geothermal Research*, 310(Supplement C):39 – 49.
- Capra, L., Manea, V. C., Manea, M., and Norini, G. (2011). The importance of digital elevation model resolution on granular flow simulations: a test case for Colima volcano using TITAN2D computational routine. *Natural Hazards*, 59(2):665–680.
- Charbonnier, S., Germa, A., Connor, C., Gertisser, R., Preece, K., Komorowski, J.-C., Lavigne, F., Dixon, T., and Connor, L. (2013). Evaluation of the impact of the 2010 pyroclastic density currents at Merapi volcano from high-resolution satellite imagery, field investigations and numerical simulations. *Journal of Volcanology and Geothermal Research*, 261(Supplement C):295 – 315. Merapi eruption.
- Charbonnier, S. J. and Gertisser, R. (2009). Numerical simulations of block-and-ash flows using the Titan2D flow model: examples from the 2006 eruption of Merapi Volcano, Java, Indonesia. *Bulletin of Volcanology*, 71(8):953–959.
- Christen, M., Kowalski, J., and Bartelt, P. (2010). RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Regions Science and Technology*, 63:1–14.
- Dade, W. B. and Huppert, H. E. (1998). Long-runout rockfalls. *Geology*, 26(9):803–806.
- Dalbey, K., Patra, A. K., Pitman, E. B., Bursik, M. I., and Sheridan, M. F. (2008). Input uncertainty propagation methods and hazard mapping of geophysical mass flows. *Journal of Geophysical Research: Solid Earth*, 113:1–16.
- Dalbey, K. R. (2009). *Predictive Simulation and Model Based Hazard Maps*. PhD thesis, University at Buffalo.

- Davila, N., Capra, L., Gavilanes-Ruiz, J., Varley, N., Norini, G., and Vazquez, A. G. (2007). Recent lahars at Volcán de Colima (Mexico): Drainage variation and spectral classification. *Journal of Volcanology and Geothermal Research*, 165(3):127 – 141.
- Denlinger, R. P. and Iverson, R. M. (2001). Flow of variably fluidized granular masses across three-dimensional terrain: 2. Numerical predictions and experimental tests. *Journal of Geophysical Research*, 106(B1):553–566.
- Denlinger, R. P. and Iverson, R. M. (2004). Granular avalanches across irregular three-dimensional terrain: 1. Theory and computation. *Journal of Geophysical Research (Earth Surface)*, 109(F1):F01014.
- Farrell, K., Tinsley, J., and Faghihi, D. (2015). A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems. *Journal of Computational Physics*, 295.
- Fischer, J., Kowalski, J., and Pudasaini, S. P. (2012). Topographic curvature effects in applied avalanche modeling. *Cold Regions Science and Technology*, 74-75:21–30.
- Forsterre, Y. and Pouliquen, O. (2002). Stability analysis of rapid granular chute flows: formation of longitudinal vortices. *Journal of Fluid Mechanics*, 467:361–387.
- Forsterre, Y. and Pouliquen, O. (2003). Long-surface-wave instability in dense granular flows. *Journal of Fluid Mechanics*, 486:21–50.
- Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1):73–79.
- González, M. B., Ramírez, J. J., and Navarro, C. (2002). Summary of the historical eruptive activity of Volcán De Colima, Mexico 1519-2000. *Journal of Volcanology and Geothermal Research*, 117(1):21 – 46.
- Gruber, U. and Bartelt, P. (2007). Snow avalanche hazard modelling of large areas using shallow water numerical models and GIS. *Environmrntal Modelling & Software*, 22:1472–1481.
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466.
- Hyman, D. and Bursik, M. (2018). Deformation of volcanic materials by pore pressurization: analog experiments with simplified geometry. *Bulletin of Volcanology*, 80(3):19.
- Iverson, R. M. (1997). The physics of debris flows. *Reviews of Geophysics*, 35:245–296.
- Iverson, R. M. and Denlinger, R. P. (2001). Flow of variably fluidized granular masses across three-dimensional terrain: 1. Coulomb mixture theory. *Journal of Geophysical Research*, 106(B1):537.
- Iverson, R. M. and George, D. L. (2014). 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).
- Iverson, R. M., Logan, M., and Denlinger, R. P. (2004). Granular avalanches across irregular three-dimensional terrain: 2. Experimental tests. *Journal of Geophysical Research: Earth Surface*, 109(F1):n/a–n/a. F01015.
- Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272.
- Jaeger, H. M., Liu, C., and Nagel, S. R. (1989). Relaxation at the Angle of Repose. *Physical Review Letters*, 62(1):40–43.
- Kelfoun, K. (2011). Suitability of simple rheological laws for the numerical simulation of dense pyroclastic flows and long-runout volcanic avalanches. *Journal of Geophysical Research*, 116:B08209.

- Kelfoun, K. and Druitt, T. H. (2005). Numerical modeling of the emplacement of Socompa rock avalanche, Chile. *Journal of Geophysical Research: Solid Earth*, 110(B12).
- Kelfoun, K., Samaniego, P., Palacios, P., and Barba, D. (2009). Testing the suitability of frictional behaviour for pyroclastic flow simulation by comparison with a well-constrained eruption at Tungurahua volcano (Ecuador). *Bulletin of Volcanology*, 71:1057–1075.
- Kennedy, M. C. and O'Hagan, A. (2011). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.
- Kern, M., Bartelt, P., Sovilla, B., and Buser, O. (2009). Measured shear rates in large dry and wet snow avalanches. *Journal of Glaciology*, 55(190):327–338.
- la Cruz-Reyna, S. D. (1993). Random patterns of occurrence of explosive eruptions at Colima Volcano, Mexico. *Journal of Volcanology and Geothermal Research*, 55(1):51 – 68.
- Macorps, E., Charbonnier, S. J., Varley, N. R., Capra, L., Atlas, Z., and Cabré, J. (2018). Stratigraphy, sedimentology and inferred flow dynamics from the July 2015 block-and-ash flow deposits at Volcán de Colima, Mexico. *Journal of Volcanology and Geothermal Research*, 349:99 – 116.
- Martin Del Pozzo, A. M., Sheridan, M. F., Barrera, M., Hubp, J. L., and Selem, L. V. (1995). Potential hazards from Colima Volcano, Mexico. *Geofisica Internacional*, 34:363–376.
- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.
- Norini, G., De Beni, E., Andronico, D., Polacci, M., Burton, M., and Zucca, F. (2009). The 16 November 2006 flank collapse of the south-east crater at Mount Etna, Italy: Study of the deposit and hazard assessment. *Journal of Geophysical Research: Solid Earth*, 114(B2). B02204.
- Ogburn, S. E., Berger, J., Calder, E. S., Lopes, D., Patra, A., Pitman, E. B., Rutarindwa, R., Spiller, E., and Wolpert, R. L. (2016). Pooling strength amongst limited datasets using hierarchical Bayesian analysis, with application to pyroclastic density current mobility metrics. *Statistics in Volcanology*, 2:1–26.
- Owen, A. B. (1992a). A Central Limit Theorem for Latin Hypercube Sampling. *Journal of the Royal Statistical Society.*, 54(2):541–551.
- Owen, A. B. (1992b). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2(2):439–452.
- Patra, A., Nichita, C., Bauer, A., Pitman, E., Bursik, M., and Sheridan, M. (2006). Parallel adaptive discontinuous Galerkin approximation for thin layer avalanche modeling. *Computers & Geosciences*, 32(7):912 – 926. Computer Simulation of natural phenomena for Hazard Assessment.
- Patra, A. K., Bauer, A. C., Nichita, C. C., Pitman, E. B., Sheridan, M. F., Bursik, M., Rupp, B., Webber, A., Stinton, A. J., Namikawa, L. M., and Renschler, C. S. (2005). Parallel adaptive numerical simulation of dry avalanches over natural terrain. *Journal of Volcanology and Geothermal Research*, 139(1-2):1–21.
- Pitman, E., Nichita, C., Patra, A., Bauer, A., Bursik, M., and Webb, A. (2003a). A model of granular flows over an erodible surface. *Discrete and Continuous Dynamical Systems - Series B*, 3(4):589–599.
- Pitman, E. B. and Le, L. (2005). A two-fluid model for avalanche and debris flows. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 363(1832):1573–601.
- Pitman, E. B., Nichita, C. C., Patra, A., Bauer, A., Sheridan, M., and Bursik, M. (2003b). Computing granular avalanches and landslides. *Physics of Fluids*, 15(12):3638.
- Pouliquen, O. (1999). Scaling laws in granular flows down rough inclined planes. *Physics of Fluids*, 11(3):542–548.

- Pouliquen, O. and Forterre, Y. (2002). Friction law for dense granular flows: application to the motion of a mass down a rough inclined plane. *Journal of Fluid Mechanics*, 453:133–151.
- Procter, J. N., Cronin, S. J., Platz, T., Patra, A., Dalbey, K., Sheridan, M., and Neall, V. (2010). Mapping block-and-ash flow hazards based on Titan 2D simulations: a case study from Mt. Tarawera, NZ. *Natural Hazards*, 53(3):483–501.
- Pudasaini, S. P. and Hutter, K. (2003). Rapid shear flows of dry granular masses down curved and twisted channels. *Journal of Fluid Mechanics*, 495:193–208.
- Ranjan, P. and Spencer, N. (2014). Space-filling Latin hypercube designs based on randomization restrictions in factorial experiments. *Statistics and Probability Letters*, 94(Supplement C):239 – 247.
- Rankine, W. J. M. (1857). On the Stability of Loose Earth. *Phil. Trans. R. Soc. Lond.*, 147(2):9–27.
- Reyes-Dávila, G. A., Arámbula-Mendoza, R., Espinasa-Perena, R., Pankhurst, M. J., Navarro-Ochoa, C., Savov, I., Vargas-Bracamontes, D. M., Cortés-Cortés, A., Gutiérrez-Martínez, C., Valdés-González, C., Domínguez-Reyes, T., González-Amezcua, M., Martínez-Fierros, A., Ramírez-Vázquez, C. A., Cárdenas-González, L., Castaneda-Bastida, E., de los Monteros, D. M. V. E., Nieto-Torres, A., Champion, R., Courtois, L., and Lee, P. D. (2016). Volcan de Colima dome collapse of July, 2015 and associated pyroclastic density currents. *Journal of Volcanology and Geothermal Research*, 320(Supplement C):100 – 106.
- Rupp, B. (2004). An analysis of granular flows over natural terrain. Master's thesis, University at Buffalo.
- Rupp, B., Bursik, M., Namikawa, L., Webb, A., Patra, A. K., Saucedo, R., Macías, J. L., and Renschler, C. (2006). Computational modeling of the 1991 block and ash flows at Colima Volcano, México. *Geological Society of America Special Papers*, 402(11):223–237.
- Ruyer-Quil, C. and Manneville, P. (2000). Improved modeling of flows down inclined planes. *Eur. Phys. J. B*, 15:357–369.
- Salm, B. (1993). Flow, flow transition and runout distances of flowing avalanches. *Annals of Glaciology*, 18:221–226.
- Salm, B., Burkard, A., and Gubler, H. (1990). Berechnung von Fliesslawinen: eine Anleitung für Praktiker mit Beispielen. *Mitteilungen des Eidgenössische Institutes für Schnee- und Lawinenforschung*, 47.
- Sarocchi, D., Sulpizio, R., Macías, J., and Saucedo, R. (2011). The 17 July 1999 block-and-ash flow (BAF) at Colima Volcano: New insights on volcanic granular flows from textural analysis. *Journal of Volcanology and Geothermal Research*, 204(1):40 – 56.
- Saucedo, R., Macías, J., Bursik, M., Mora, J., Gavilanes, J., and Cortes, A. (2002). Emplacement of pyroclastic flows during the 1998–1999 eruption of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, 117(1):129 – 153.
- Saucedo, R., Macías, J., Gavilanes, J., Arce, J., Komorowski, J., Gardner, J., and Valdez-Moreno, G. (2010). Eyewitness, stratigraphy, chemistry, and eruptive dynamics of the 1913 Plinian eruption of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, 191(3–4):149 – 166.
- Saucedo, R., Macías, J., Sheridan, M., Bursik, M., and Komorowski, J. (2005). Modeling of pyroclastic flows of Colima Volcano, Mexico: implications for hazard assessment. *Journal of Volcanology and Geothermal Research*, 139(1):103 – 115. Modeling and Simulation of Geophysical Mass Flows.
- Saucedo, R., Macías, J. L., and Bursik, M. (2004). Pyroclastic flow deposits of the 1991 eruption of Volcán de Colima, Mexico. *Bulletin of Volcanology*, 66(4):291–306.
- Savage, S. B. and Hutter, K. (1989). The motion of a finite mass of granular material down a rough incline. *Journal of Fluid Mechanics*, 199:177.

- Sheridan, M., Stinton, A., Patra, A., Pitman, E., Bauer, A., and Nichita, C. (2005). Evaluating Titan2D mass-flow model using the 1963 Little Tahoma Peak avalanches, Mount Rainier, Washington. *Journal of Volcanology and Geothermal Research*, 139(1):89 – 102. Modeling and Simulation of Geophysical Mass Flows.
- Sheridan, M. F. and Macías, J. L. (1995). Estimation of risk probability for gravity-driven pyroclastic flows at Volcán Colima, Mexico. *Journal of Volcanology and Geothermal Research*, 66(1):251 – 256. Models of Magnetic Processes and Volcanic Eruptions.
- Sheridan, M. F., Patra, A. K., Dalbey, K., and Hubbard, B. (2010). *Probabilistic digital hazard maps for avalanches and massive pyroclastic flows using TITAN2D*, volume Stratigraphy and geology of volcanic areas, 464, chapter 14, page 281. Geological Society of America.
- Silbert, L. E., Ertas, D., Grest, G. S., Halsey, T. C., Levine, D., and Plimpton, S. J. (2001). Granular flow down an inclined plane: Bagnold scaling and rheology. *Physical Review E*, 64(5):051302.
- Spiller, E. T., Bayarri, M. J., Berger, J. O., Calder, E. S., Patra, A. K., Pitman, E. B., and Wolpert, R. L. (2014). Automating Emulator Construction for Geophysical Hazard Maps. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):126–152.
- Stefanescu, E. R., Bursik, M., Cordoba, G., Dalbey, K., Jones, M. D., Patra, A. K., Pieri, D. C., Pitman, E. B., and Sheridan, M. F. (2012a). Digital elevation model uncertainty and hazard analysis using a geophysical flow model. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 468(2142):1543–1563.
- Stefanescu, E. R., Bursik, M., and Patra, A. K. (2012b). Effect of digital elevation model on Mohr-Coulomb geophysical flow model output. *Natural Hazards*, 62(2):635–656.
- Stein, M. (1987). Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics*, 29(2):143–151.
- Sulpizio, R., Capra, L., Sarocchi, D., Saucedo, R., Gavilanes-Ruiz, J., and Varley, N. (2010). Predicting the block-and-ash flow inundation areas at Volcán de Colima (Colima, Mexico) based on the present day (February 2010) status. *Journal of Volcanology and Geothermal Research*, 193(1):49 – 66.
- Tang, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397.
- Voellmy, A. (1955). Über die Zerstörungskraft von Lawinen. *Schweiz Bauzeitung*, 73:159–165, 212–217, 246–249, 280–285.
- Webb, A. (2004). Granular flow experiments to validate numerical flow model, TITAN2D. Master's thesis, University at Buffalo.
- Webb, A. and Bursik, M. (2016). Granular flow experiments for validation of numerical flow models. Technical report, University at Buffalo.
- Wolpert, R., Ogburn, S., and Calder, E. (2016). The longevity of lava dome eruptions. *Journal of Geophysical Research: Solid Earth*, 121(2):676–686. 2015JB012435.
- Yu, B., Dalbey, K., Webb, A., Bursik, M., Patra, A. K., Pitman, E. B., and Nichita, C. (2009). Numerical issues in computing inundation areas over natural terrains using Savage-Hutter theory. *Natural Hazards*, 50(2):249–267.
- Zobin, V., Arámbula, R., Bretón, M., Reyes, G., Plascencia, I., Navarro, C., Téllez, A., Campos, A., González, M., León, Z., Martínez, A., and Ramírez, C. (2015). Dynamics of the January 2013–June 2014 explosive-effusive episode in the eruption of Volcán de Colima, México: insights from seismic and video monitoring. *Bulletin of Volcanology*, 77(4):31.
- Zobin, V., Luhr, J., Taran, Y., Bretón, M., Cortés, A., De La Cruz-Reyna, S., Domínguez, T., Galindo, I., Gavilanes, J., Muniz, J., Navarro, C., Ramírez, J., Reyes, G., Ursúa, M., Velasco, J., Alatorre, E., and Santiago, H. (2002). Overview of the 1997–2000 activity of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, 117(1):1 – 19.