Analyzing Complex Models using Data and Statistics

A.K. Patra, [1] A. Bevilacqua, [2] A. Akhavan Safaei [3]

[1]Dept. of Mech. and Aero. Eng.

[2]Dept. of Earth Sciences

[3]Comp. Data Science and Eng.

International Conference on Computational Sciences, 2018

Models and assumptions

What is a Model?

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.

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

Analysis Process - in *predictive mode*

We are interested in the general predictive capabilities of the models, related to the outcomes over a whole range.

- Stage 1: Parameter Ranges
  $P_M\,(p_1, \ldots, p_{N_M}) \sim \bigotimes_{i=1}^{N_M} Unif(a_{i,M}, b_{i,M})$.

- Stage 2: Simulations and Data Gathering

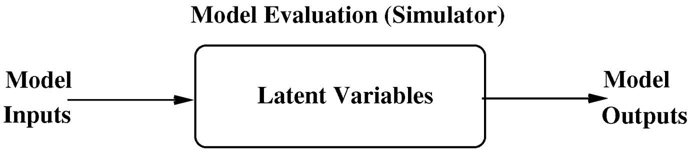

Figure: Models and variables

- Stage 3: Results Analysis

Statistics of latent variables - dominance factors

Dominance factors provide insight into the dominance of a particular latent variable over the others.

Definition (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:

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

Statistics of latent variables - expected contributions

Random contributions are obtained dividing the latent variables by Φ, and hence belong to $[0, 1]$.

Definition (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 Φ is the dominant variable. Thus, $\forall i$, the expected contributions are defined by $E\left[C_i\right]$.

Modeling of geophysical mass flows

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

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

Source terms S_x, S_y characterize *Mohr-Coulomb* (MC), *Pouliquen-Forterre* (PF) and *Voellmy-Salm* (VS) models.

Main assumptions - all the models include *curvature effects*.

Mohr-Coulomb

- *Basal Friction* based on a constant friction angle.
- *Internal Friction* based on a constant friction angle.

Pouliquen-Forterre

- *Basal Friction* is based on an interpolation of two different friction angles, based on the flow regime and depth.
- Normal stress is modified by a *hydrostatic pressure force* related to the flow height gradient.

Voellmy-Salm

- *Basal Friction* is based on a constant coefficient, similarly to the MC rheology.
- Additional *turbulent friction* is based on the local velocity by a quadratic expression.

Overview of the case studies

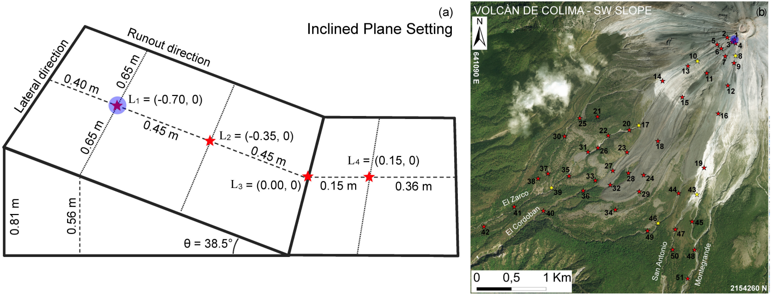

Figure: **[Left]** Inclined plane description, including local samples sites (red stars). Pile location is marked by a blue dot.**[Right]**(a) Volcán de Colima (México) overview, including 51 numbered local sample sites (stars) and four labeled major ravines channeling the flow. Pile location is marked by a blue dot. Reported coordinates are in UTM zone 13N. Background is a satellite photo.

Small scale flow - observable outputs

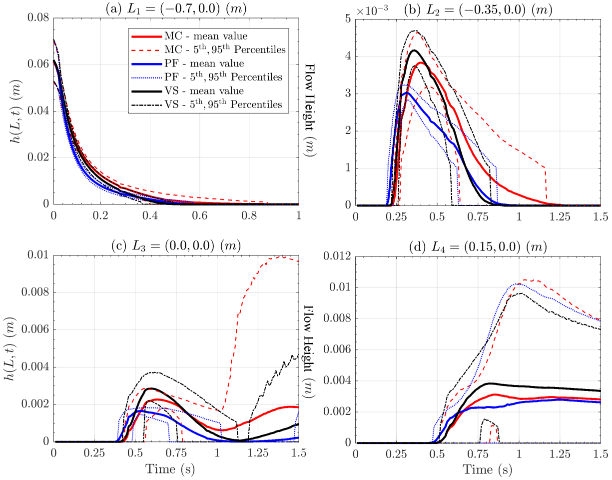

Figure: Flow height in four locations. Bold line is mean value, dashed/dotted lines are 5^{th} and 95^{th} percentile bounds. Different models are displayed with different colors. Plots are at different scale, for simplifying exposition.

Small scale flow - power integrals

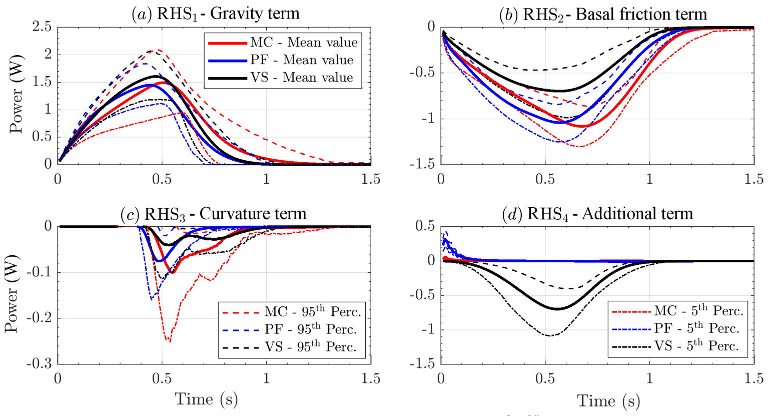

Figure: Spatial integral of the RHS powers. Bold line is mean value, dashed lines are 5^{th} and 95^{th} percentile bounds. Different models are displayed with different colors.

Large scale flow - proximal to the initial pile

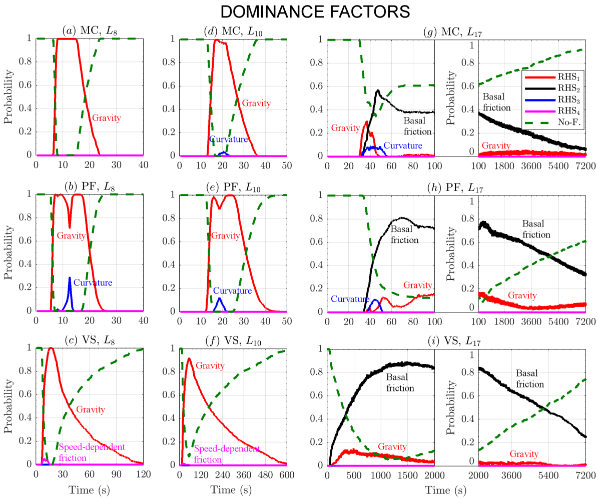

Figure: Dominance factors of **RHS** forces in three locations in the first km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms. No-flow probability is displayed with a green dashed line.

Large scale flow - proximal to the initial pile

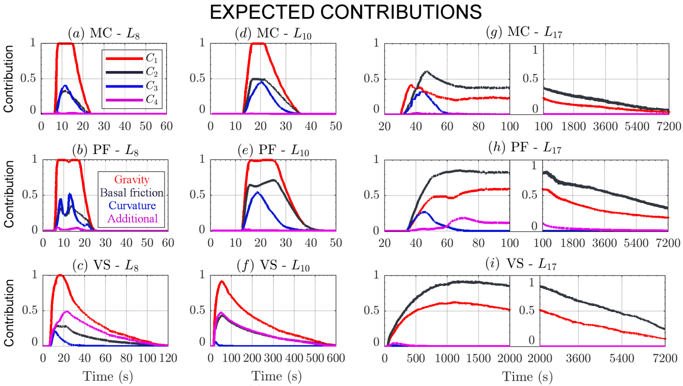

Figure: Expected contributions of **RHS** forces in three locations in the first km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms.

Large scale flow - distal from the initial pile

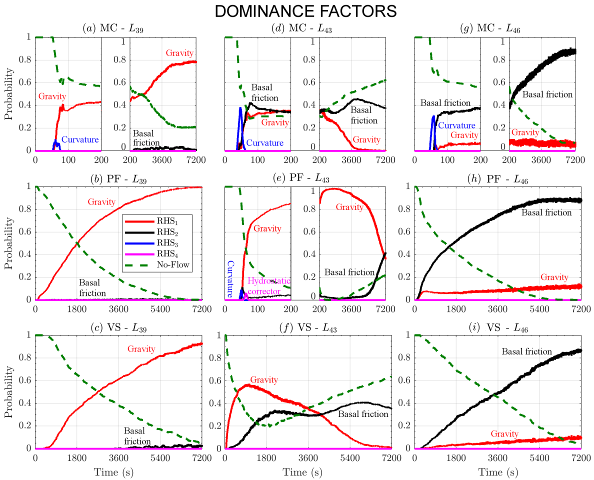

Figure: Dominance factors of **RHS** forces in three locations after 2 km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms. No-flow probability is displayed with a green dashed line.

Large scale flow - distal from the initial pile

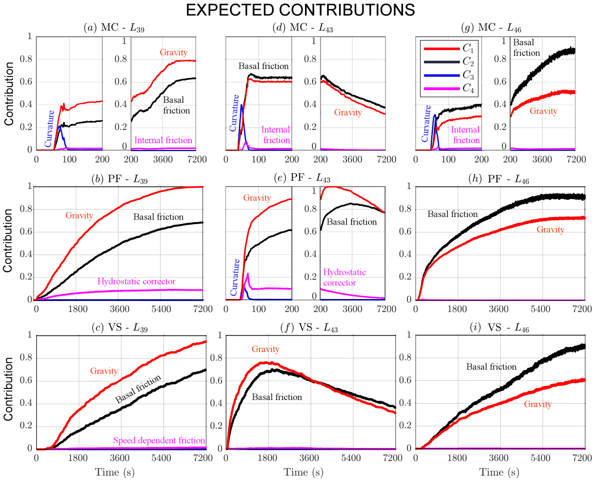

Figure: Expected contributions of **RHS** forces in three locations after 2 km of runout. Different models are plotted separately: (a,d,g) assume MC; (b,e,h) assume PF; (c,f,i) assume VS. Different colors correspond to different force terms.

Large scale flow - flow extent and spatial integrals

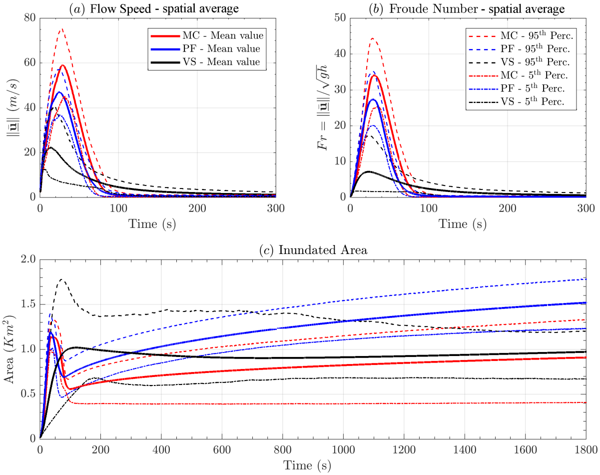

Figure: Comparison between spatial averages of (a) flow speed, and (b) Froude Number, in addition to the (c) inundated area, as a function of time. Bold line is mean value, dashed/dotted lines are 5^{th} and 95^{th} percentile bounds. Different models are displayed with different colors.

Conclusions

- our approach evaluates the statistics of a range of simulations, produced by the couple (M, P_M).
- the new statistical framework processes the mean and the uncertainty range of either observable or latent variables in the simulation.
- analysis is performed at selected sites, and spatial integrals were also performed, illustrating the characteristics of the entire output.
- the new concepts of dominance factors and expected contributions, enable a simplified description of the local dynamics.