Bevilacqua et al. (2019), Probabilistic forecasting of plausible debris flows from Nevado de Colima (México) using data from the Atenquique debris flow, 1955. https://doi.org/10.5194/nhess-19-791-2019

Refining the input space of plausible future debris flows using noisy data and multiple models of the physics.

Andrea Bevilacqua, Abani Patra, Marcus Bursik, E. Bruce Pitman, José Luis Macìas, Ricardo Saucedo, David Hyman

Pisa, 12 September 2019

See also the poster Bevilacqua et al. (2019) -this congress-

Acknowledgements:
This work is supported by National Science Foundation awards 1339765, 1521855, 1621853 and 1821311, and by Italian Ministry of Education, University, and Research, project FISR2017 - SOIR.

Prediction oriented modelling – overview of the problem

Hazard assessment of **geophysical mass flows**, such as landslides or pyroclastic flows, usually relies on the reconstruction of past flows that occurred in the region of interest using models of physics that have been successful in **hindcasting**.

While **physical models** relate inputs and outputs of the dynamical system of the mass flow this relation is **dependent on the choice of model** and **parameters initial range**, which is usually difficult for future events.

Choices based on limited data using **classical inversion** is often misleading since it does not reflect all potential event characteristics and even in a probabilistic setting can be error-prone, due to incorrectly limited event space.

In this work, we use a **multi-model** ensemble and a **plausible region** approach to provide a **prediction-oriented** probabilistic framework for hazard analysis.

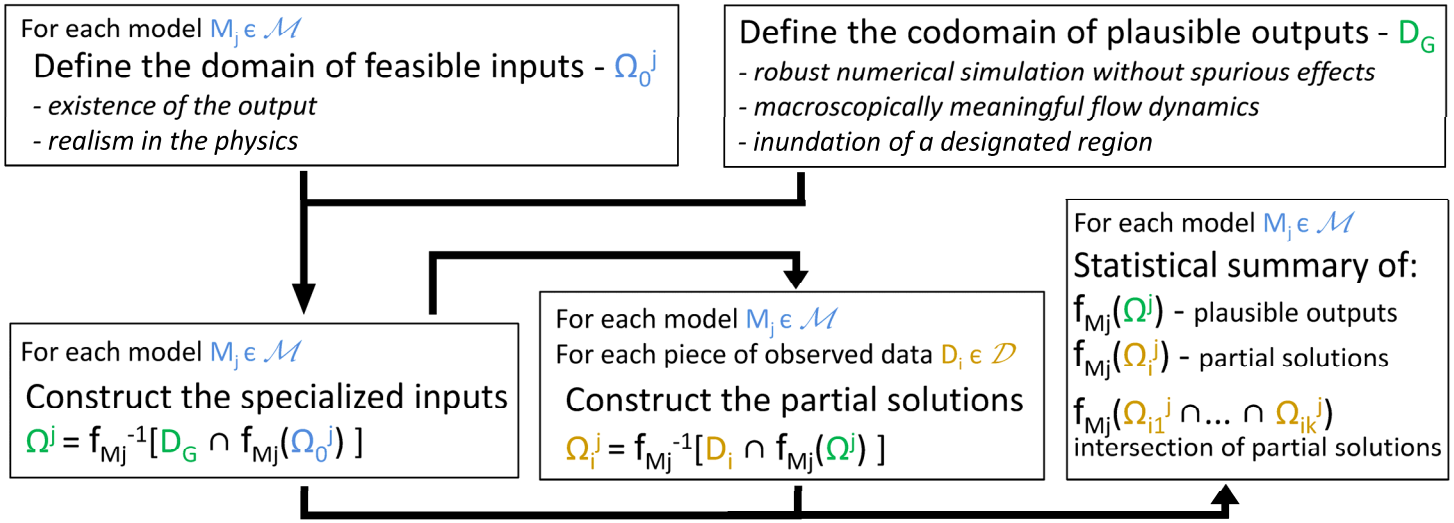
Prediction oriented modelling – input sets and subsets

We represent each model M_j with an operator: $f_{M_j} : \Omega_0^j \longrightarrow \mathbb{R}^d$

The set of feasible inputs: $\Omega_G := \bigsqcup_j \Omega_0^j$, is a natural meta-modeling framework.

Then, we characterize the codomain $D_G \subset \mathbb{R}^d$ of plausible outputs. It includes all the outputs consistent with the observed data, plus additional outputs which differ in arbitrary, but plausible ways.

Fig 2. Diagram of the steps of our meta-modeling approach.



We use the same notation for each piece of data D_i and the set of outputs consistent with it.

Then ∀j, the specialized input space is defined by: $\Omega^j = f_{Mj}^{-1}[D_G \cap f_{Mj}(\Omega_0^j)]$.

In a similar way, ∀i, the partial solutions to the inverse problem are: $\Omega_i^j = f_{Mj}^{-1}[D_i \cap f_{Mj}(\Omega^j)]$.

Prediction oriented modelling – the specialized input space

For each model M_j we define a probability measure P_M over the measurable parts of its input space Ω^j_0.
In the sequel we assume $P_M \sim \text{unif}(\Omega^j)$, where $\Omega^j \subseteq \Omega^j_0$ is called **specialized input space**.

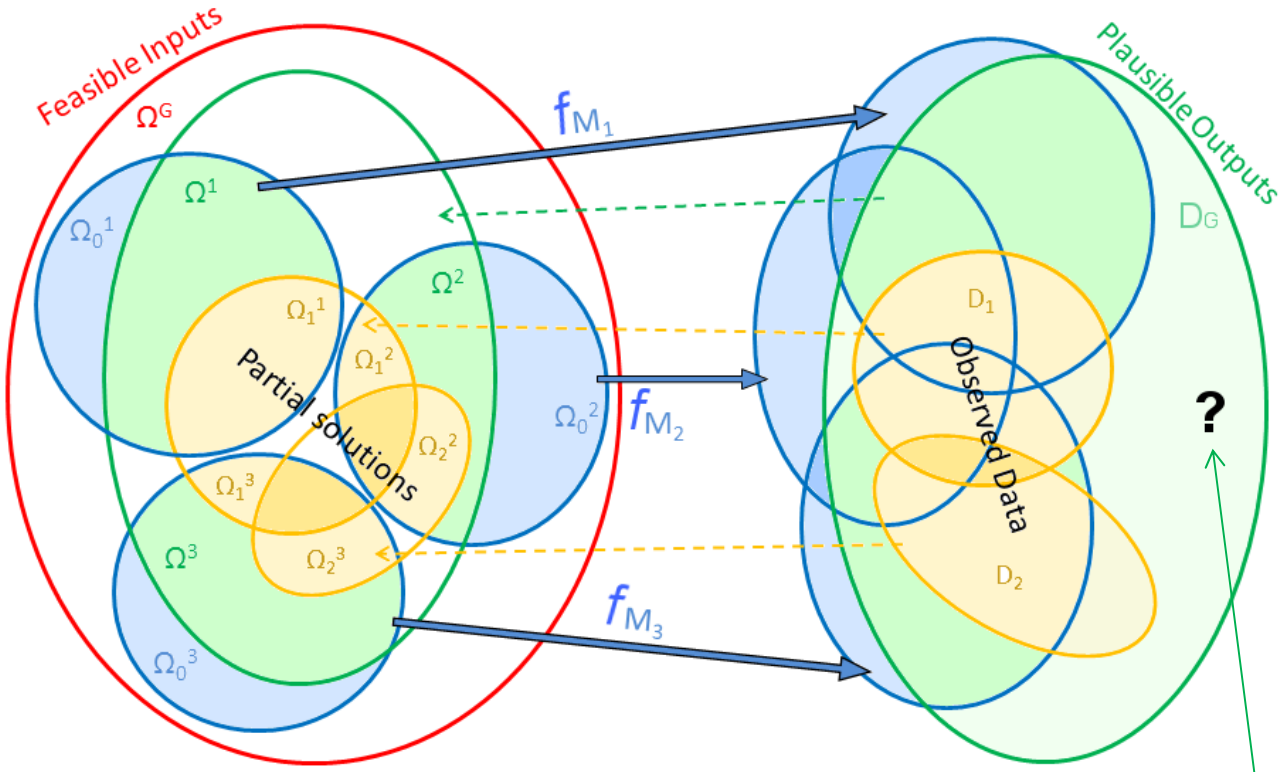


Figure. Venn Diagram:

- input spaces, output space, and model functions ,
- feasible inputs domain,
- plausible output codomain and specialized inputs,
- observed data and partial solutions.

The implementation of **multiple models** is a crucial aspect. Typically, a single model is not able to **entirely cover** D_G.

The **question mark** emphasizes that the covering of other plausible outputs could be enabled, adding more models if necessary.

Geophysical case study

The **Colima Volcanic Complex** is located in the western portion of the Trans-Mexican Volcanic Belt.

It consists of a N-S volcanic chain formed by three volcanoes. Nevado de Colima (4320 m.a.s.l.) occupies the central part of the volcanic complex, and is most voluminous of the three volcanoes.

The considered drainage starts at an elevation of **4000 m** on the eastern flank of Nevado, is occupied by **Atenquique river**, and ends at its junction with the Tuxpan River at **1040 m**.

On 16 October, 1955, at 10:45 am, a **8-9 m high** wave carrying mud, boulders and tree trunks reached the village of Atenquique. More than 23 people died and the flood leveled most buildings (Ponce Segura, 1983).

Deposits cover a minimal area of 1.2 km^2, and a minimum volume of 3.2 × 10^6 m^3 was estimated for the flow.

The diluted flow that inundated the village probably had a velocity in the range of **4 to 6 m/s**, based on the superelevation method (Pierson, 1985; Saucedo et al., 2008).

BARRANCA DE ATENQUIQUE

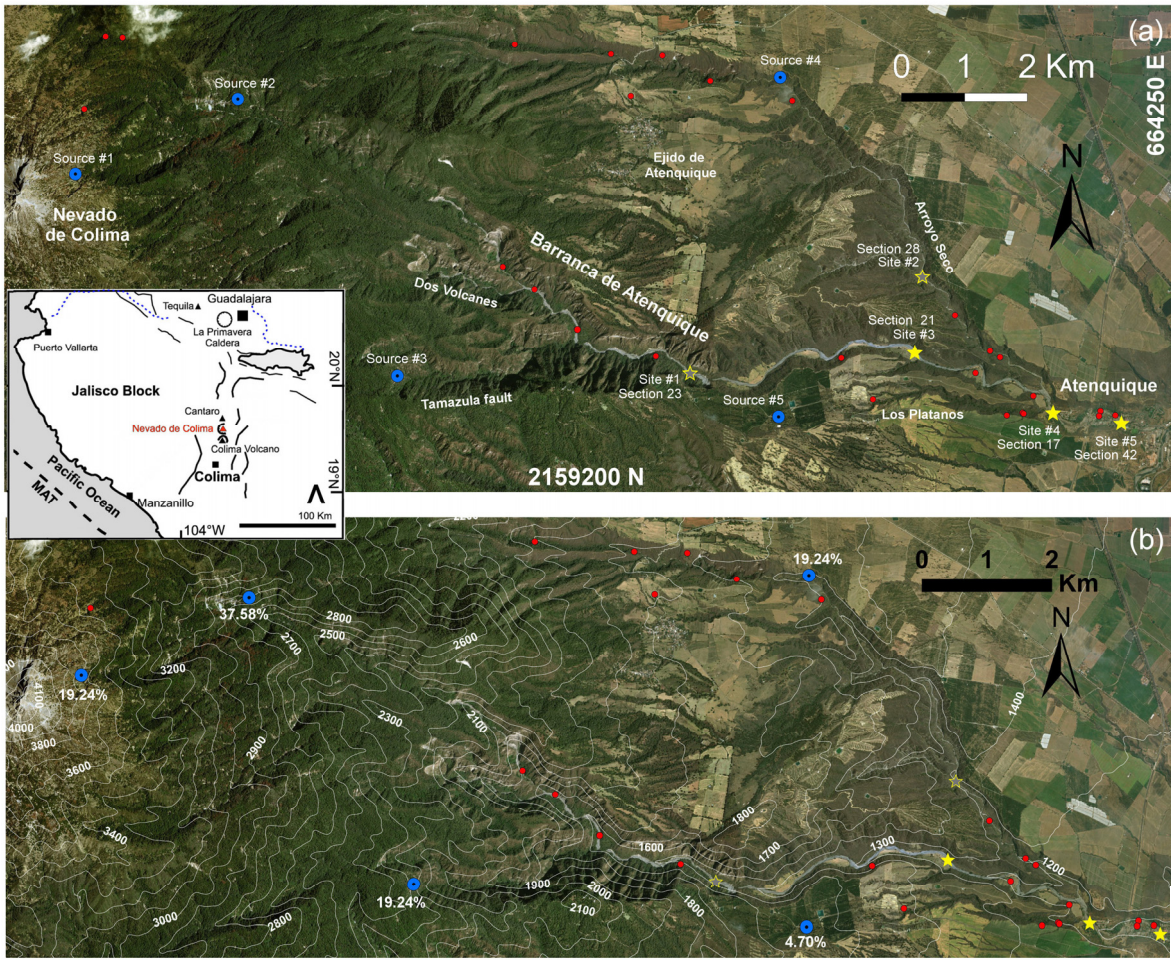


Figure. Barranca de Atenquique (MX) overview.
(a) sample sites of Saucedo et al., (2008) are marked with red dots, including 5 preferred locations (stars) and major ravines.
Initial source piles are marked by blue dots. UTM zone 13N WGS84.
(b) Digital elevation map including isolines (NASA, 2014). Volume partition percentage among sources is reported. A regional map is in a small box.

Geophysical models and input spaces - overview

We adopt and compare the three depth averaged models **Mohr-Coulomb** (MC) (Savage&Hutter, 1989), **Pouliquen-Forterre** (PF) (Pouliquen&Forterre, 2002), **Voellmy-Salm** (VS) (Voellmy, 1955; Salm, 1990).

Input spaces are explored by **Latin Hypercube sampling** (McKay, 1979; Owen, 1992b; Stein, 1987).

The models are incorporated in our large scale mass flow simulation framework **TITAN2D** (Patra et al., 2005).

Models are parameterized by:

$$\Omega_0^{\mathrm{PF}} = \left\{(\phi_1, \phi_2, L, V) \in \mathbb{R}_+^4\right\}$$

$$\Omega_0^{\mathrm{VS}} = \left\{[\arctan(\mu), \log_{10}(\xi), V] \in \mathbb{R}_+^3\right\}$$

$$\Omega_0^{\mathrm{MC}} = \left\{(\phi_{bed}, \phi_{int}, V) \in \mathbb{R}_+^3\right\}$$

The 4th release of TITAN2D available from **vhub.org** offers multiple rheology options in the same code base.

We enhance the sampling procedure by relying on orthogonal arrays (Owen, 1992a; Tang, 1993)

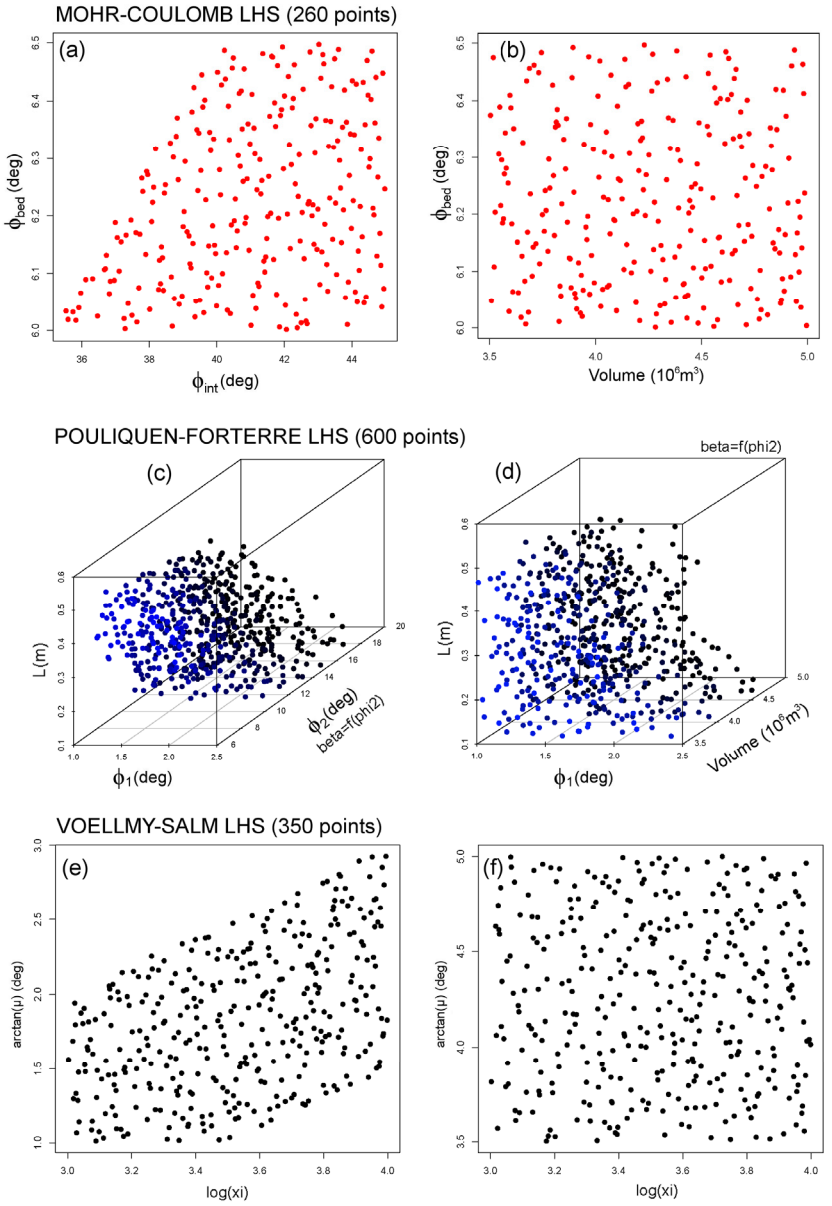


Figure. Overview of the specialized experimental design in (a-b) **MC**, (c-d) **PF**, (e-f) **VS** models.

(a-c-e) are projected along the V coordinate, and (b-d-f) along φ_{int}, φ_2 and ξ coordinates, respectively.

Geophysical models and input spaces – testing plausible outputs

- **Total Volume:** $V \in [3.5,\ 5] \times 10^6\ m^3$, i.e. $4.25 \pm 0.75 \times 10^6\ m^3$.
- **Input space constraints:**

Feasible inputs domain Ω_G

MC - $\phi_{bed} \geq 5°, \quad \phi_{int} \in [\phi_{bed},\ 45°]$.

PF - $\phi_1 \geq 1°, \quad \phi_2 \in [\phi_1 + 6°,\ \phi_1 + 18°], \quad L \in [0.1,\ 0.5]\ m$.

VS - $\arctan(\mu) \geq 1°, \quad \log_{10}(\xi) \leq 4$.

∀j the construction of Ω^j relies on **extensive testing** of the models over the input space.

We base our analysis on two properties that any **plausible output** is expected to have:

(i) the flow must reach the town of Atenquique in a reasonable time (< 1200 s)

(ii) the flow does not unphysically run-up and over-spill the ravine walls for >0.1 m depth.

The values in the table are elevation m.a.s.l. observed at the max. runout – the town of Atenquique starts below 1200 m.

SPECIALIZED INPUT SPACES

(a)

Mohr-Coulomb

TABLE A		φ_{int} (deg)			
		30	35	40	45
φ_{bed} (deg)	5.5	1170	1150	1120	1100
	6	*1230*	**1200**	**1180**	**1130**
	6.5	*1260*	*1230*	**1200**	**1160**
	7		*1280*	*1250*	*1220*

(b)

Voellmy-Salm

TABLE C		$\log_{10}(\xi)$			
		2.5	3.0	3.5	4.0
atan(μ) (deg)	1	*1220*	**1020**	**980**	970
	1.5	*1260*	**1060**	**1010**	**990**
	2	*1330*	**1200**	**1060**	**1010**
	2.5	*1390*	*1280*	**1170**	**1050**
	3		*1360*	*1280*	**1200**
	3.5			*1370*	*1280*

(c)

Pouliquenne-Forterre (1) β=0.1 φ2=7 (deg)

TABLE B1		L (m)				
		0.1	0.2	0.3	0.4	0.5
φ_1 (deg)	1		1050	1080	**1110**	**1130**
	1.5		1090	1120	**1150**	**1170**
	2		1140	**1170**	**1200**	*1220*
	2.5		*1210*	*1230*	*1250*	*1260*
	3					

Pouliquenne-Forterre (2) β=0.3 φ2=11 (deg)

TABLE B2		L (m)				
		0.1	0.2	0.3	0.4	0.5
φ_1 (deg)	1	**990**	**1020**	**1070**	**1110**	**1150**
	1.5	**1020**	**1070**	**1120**	**1170**	**1200**
	2	**1070**	**1130**	**1190**	*1220*	*1250*
	2.5	**1150**	*1220*	*1260*	*1280*	
	3	*1230*	*1290*			

Pouliquenne-Forterre (3) β=0.5 φ2=15 (deg)

TABLE B3		L (m)				
		0.1	0.2	0.3	0.4	0.5
φ_1 (deg)	1	**990**	**1020**	**1090**	**1130**	**1170**
	1.5	**1020**	**1090**	**1150**	**1200**	*1230*
	2	**1070**	**1160**	*1220*	*1260*	*1290*
	2.5	**1170**	*1250*	*1280*		
	3	*1250*	*1330*			

Pouliquenne-Forterre (4) β=0.7 φ2=19 (deg)

TABLE B4		L (m)				
		0.1	0.2	0.3	0.4	0.5
φ_1 (deg)	1	**990**	**1050**	**1110**	**1170**	1220
	1.5	**1020**	**1100**	**1170**	1230	1280
	2	**1090**	**1190**	1250	1300	
	2.5	**1190**	1270	1340		
	3	1260	1360			

The existence of **multiple source areas** presents a challenge when attempting to model the considered flow (Saucedo et al., 2008).

Based on the work in (Rupp, 2004) we initiate the flow from **five major** source locations. Piles are paraboloids with unit aspect ratio.

In particular, ∀ k, $w_k = V_k/V$ is:
$w_1 = w_3 = w_4 =$ **19.24%,**
$w_2 =$ **37.58%,** $w_5 =$ **4.70%.**

This is equivalent to choosing pile radii of 80 m, 100 m and 50 m respectively.

Statistics of the plausible outputs – max. flow height

∀j we sample the model M_j input in a **Monte Carlo** simulation.

The output of each sample run is a function $f_j(\omega, \underline{x}, t)$, where ω is the input, t is the time and $\underline{x}$ is a spatial element of the computational grid.

f_j is a **random function** with respect to the probability distribution P_{Mj} over Ω^j.

The results are summarized in the figure by a family of spatial maps of **maximum flow depth** with respect to time, H.

We display the 5th and 95th percentiles with respect to P_{Mj}.

Local maxima of flow height are located in the ravine.

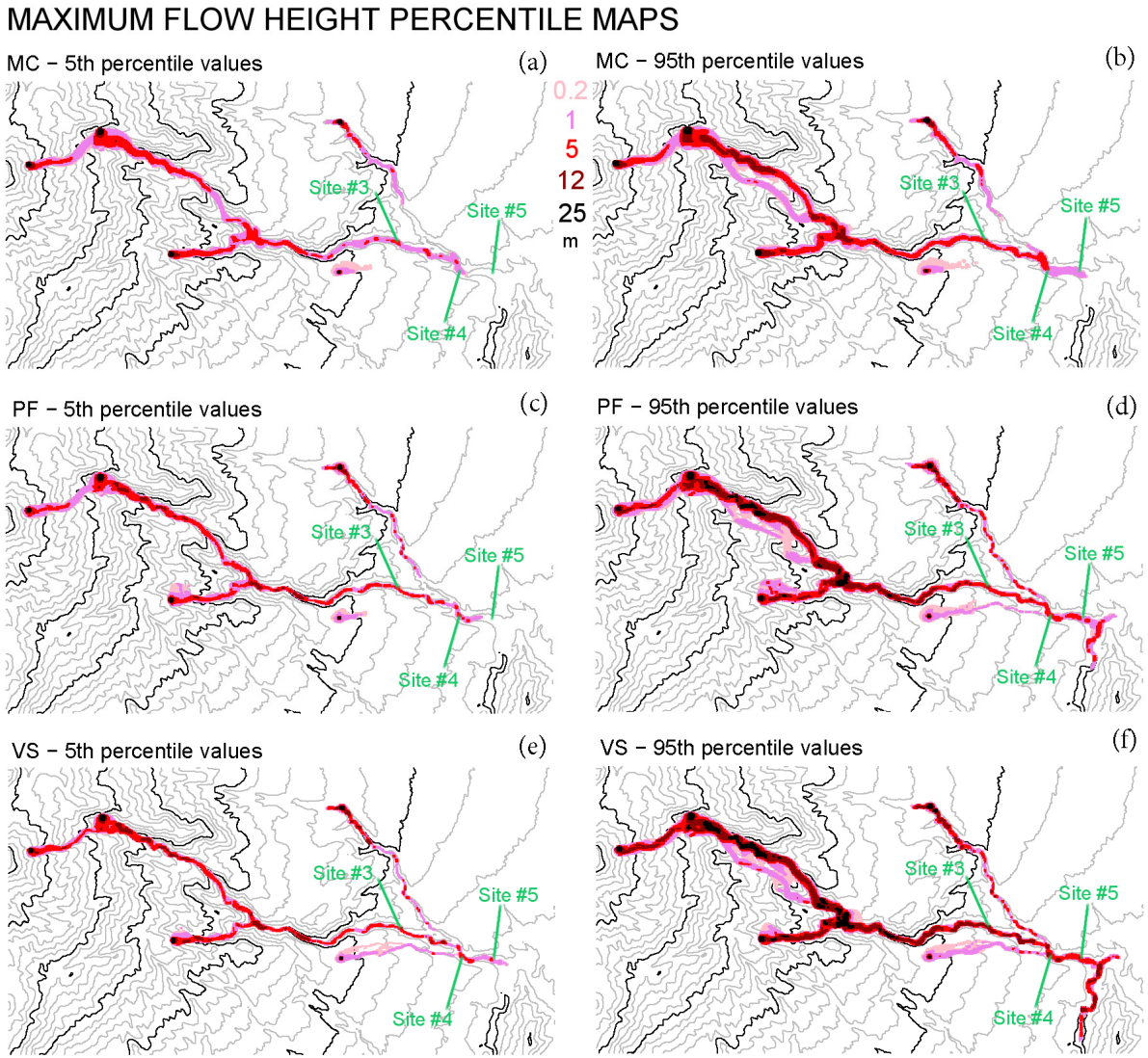


Figure. Maximum flow height H as a function of time in (a-b) MC, (c-d) PF, (e-f) VS model. (a-c-e) are the 5th and (b-d-f) are the 95th percentile values with respect to P^j.

Elevation contours are included at intervals of 100 m and 500 m (NASA, 2014). Sites #3, #4, #5 are displayed.

Close to Site #4 there are the supports of the new bridge of the freeway to the city of Colima.

Statistics of the plausible outputs – max. kinetic energy

MC shows the lowest values, while **VS the highest**, especially in the distal part of the flow.

In MC, the flow in the **tributaries** is not capable of reaching the village, while in the 95th percentile maps of PF and VS, it is.

In VS, the flow in Arroyo Plàtanos (the southernmost branch) joins the main ravine even in the 5th percentile map. Significant over-spill issues are absent.

This is formally the kinetic energy density per unit of surface area, for a mass with unit density.
Typical values of density of **mudflows** are above 2000 kg/m^3.

$$\kappa := \frac{1}{2}\frac{(h\bar{u})^2 + (h\bar{v})^2}{h}$$

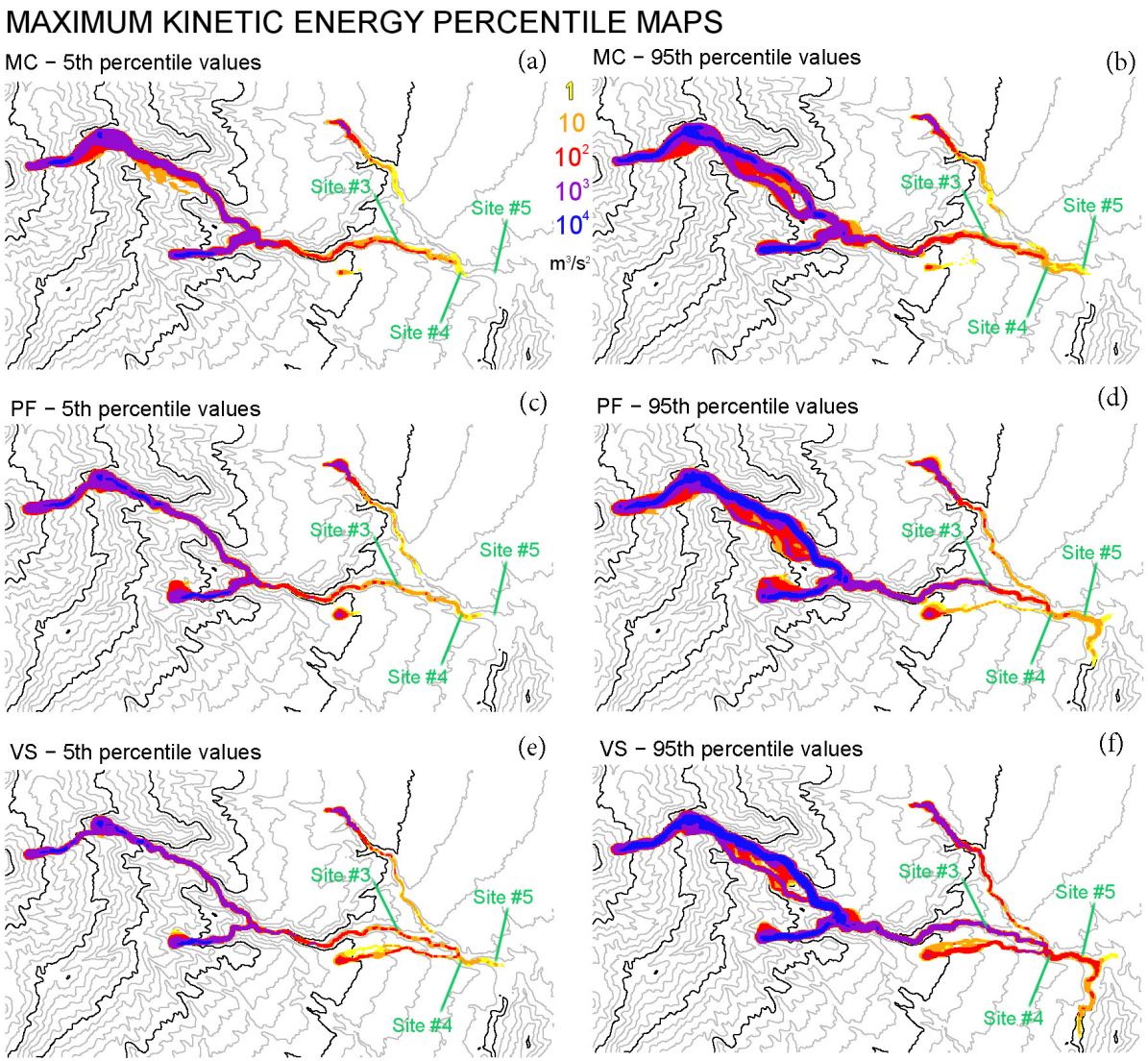


Figure. Maximum kinetic energy K as a function of time in (a-b) MC, (c-d) PF, (e-f) VS model. (a-c-e) are the 5th and (b-d-f) are the 95th percentile values with respect to P^j.

Elevation contours are included at intervals of 100 m and 500 m (NASA, 2014). Sites #3, #4, #5 are displayed.

Statistics of the plausible outputs – local histograms

We focus our analysis on the **sites #3,4,5**, all placed in proximity to the town of **Atenquique**.

The figure shows the flow height after 2400 s of simulation, and the max. height and speed. Data of the 1955 flow is included.

[Site #3] ~2 km upstream from the town;

[Site #4] Immediately upstream from the town;

[Site #5] In the town, ~1 km downstream from site #4.

The deposit thickness is inferred from the closest geological sections:
[3.7, 5.5] m at Site #3,
[1.7, 3] m at Site #4,
[1.4, 3.8] m at Site #5.

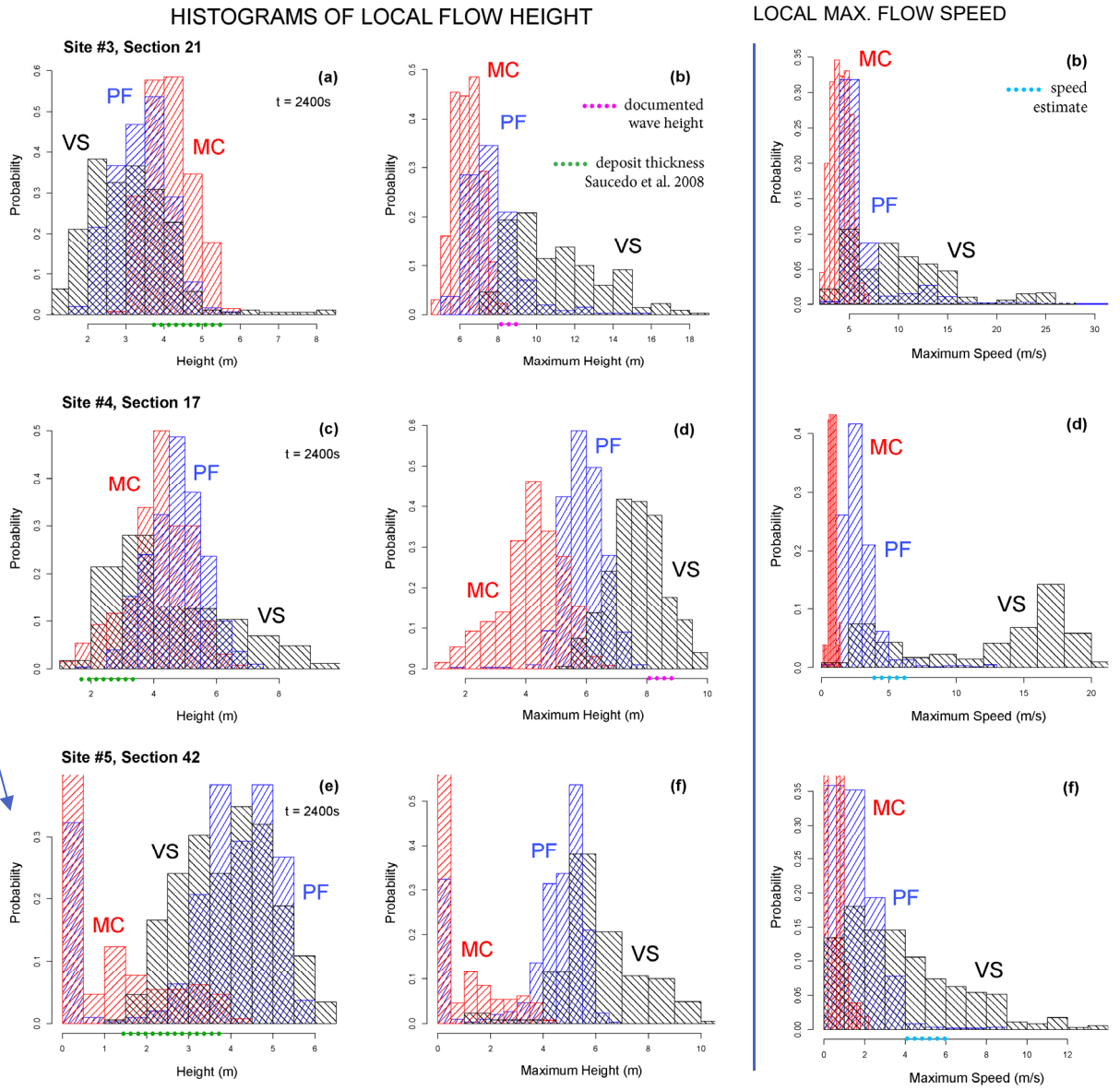


Figure. Histograms of local flow height and speed in Sites #3, #4, #5. (a,d,g) show height at t=2400 s, (b,e,h) maximum height, (c,f,i) maximum speed.

The models are displayed with different colors. Dots on the height axis show the uncertainty interval of data.

Local height and speed as a function of time

Figure. Local flow properties at Site #3 and #4.

(a, e, i) show flow height, (b, f, j)} flow speed.

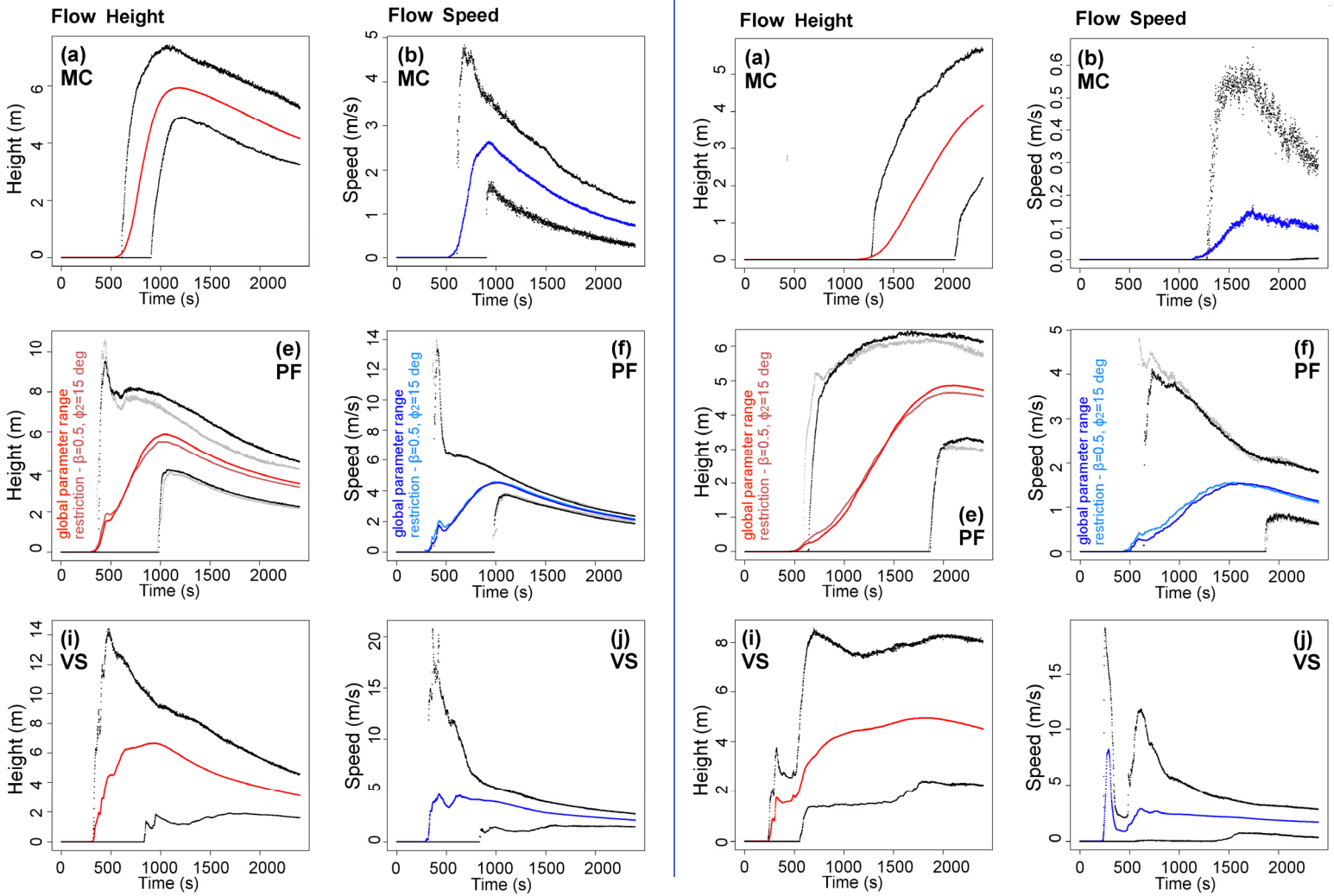
Different models are plotted separately: (a-d)} assume MC, (e-h) assume PF. (i-l) assume VS.

Colored line is mean value, and black and gray lines are 5th and 95th perc. bounds.

(e, f) include estimates on a hyperplanar restriction of the input domain

SITE #3
~2 km upstream from the town

SITE #4
Immediately upstream from the town



Force dominance factors and expected contributions as a function of time

FrontUQ19

Let $F_n(\underline{x}, t)_{n=1,\dots,N}$ be an array of force terms, where $\underline{x} \in \mathbb{R}^d$ is a spatial location, and $t \in T$ is a time instant.

We define the dominance factors $P_n(\underline{x}, t)_{n=1,\dots,N}$, i.e., the probability of each $F_n(\underline{x}, t)$ being the greatest force in $(\underline{x}, t)$.

We define the force contributions $C_n(\underline{x}, t)_{n=1,\dots,N}$ by dividing the force terms $F_n(\underline{x}, t)_{n=1,\dots,N}$ by the greatest force term in $(\underline{x}, t)$.

The expected contributions are $E[C_n(\underline{x}, t)_{n=1,\dots,N}]$.

Expected contributions and dominance factors provide insight into the dominance of a particular source or dissipation term on the model dynamics.

More details about this can be found in Patra et al. (2018), Bevilacqua et al. (2019).

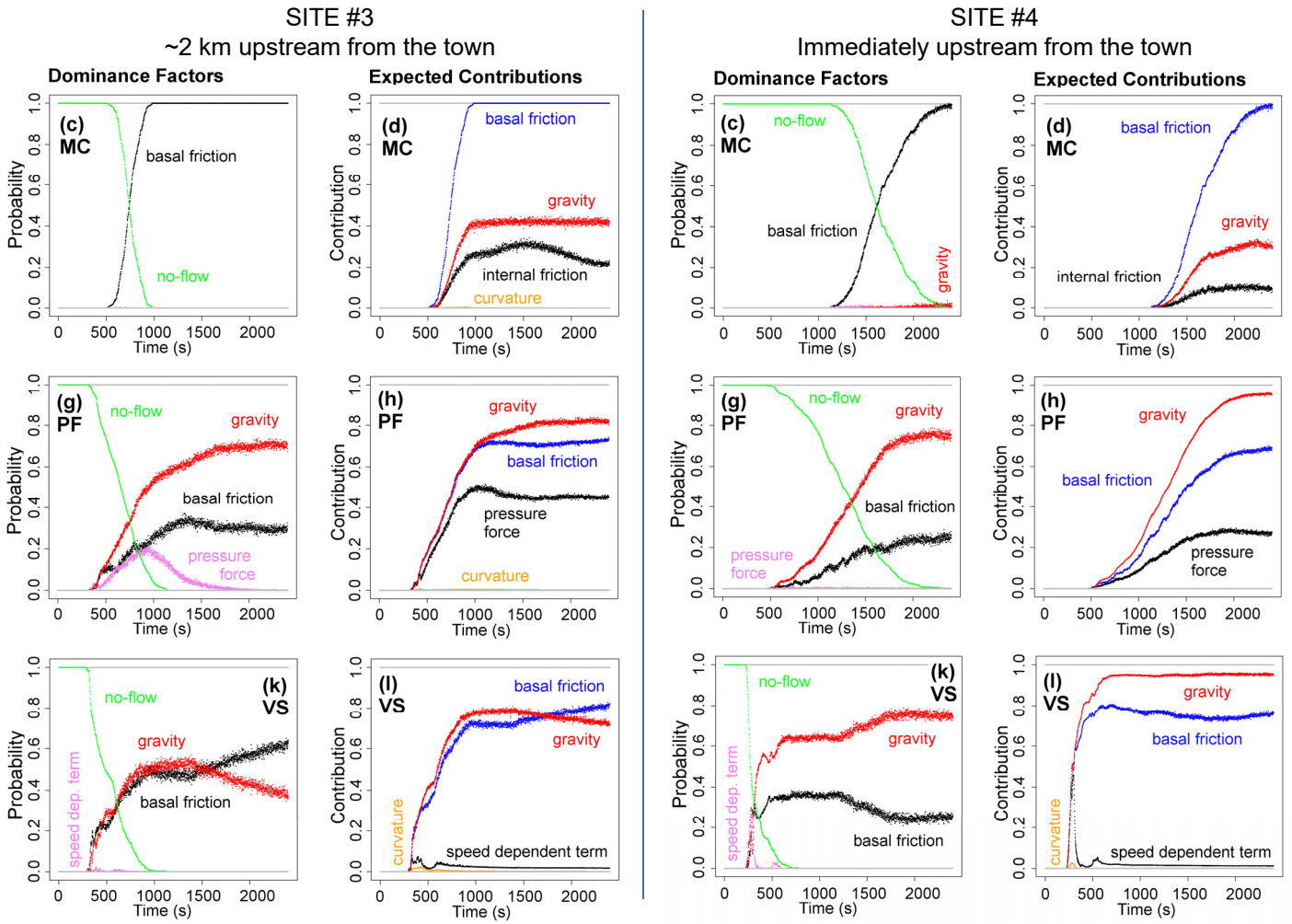


Figure. Local flow properties at Sites #3 and #4.

(c, g, k) show force dominance factors, (d, h, l) force expected contributions.

Different models are plotted separately: (a-d)} assume MC, (e-h) assume PF. (i-l) assume VS.

A. Bevilacqua, A. Patra, M. Bursik, E.B. Pitman, J.L. Macìas, R. Saucedo, D. Hyman

Partial solutions in the input space – model likelihoods

The figure shows the barplots of **data likelihood**.
Model performance depends on the selected **type of data and site**.

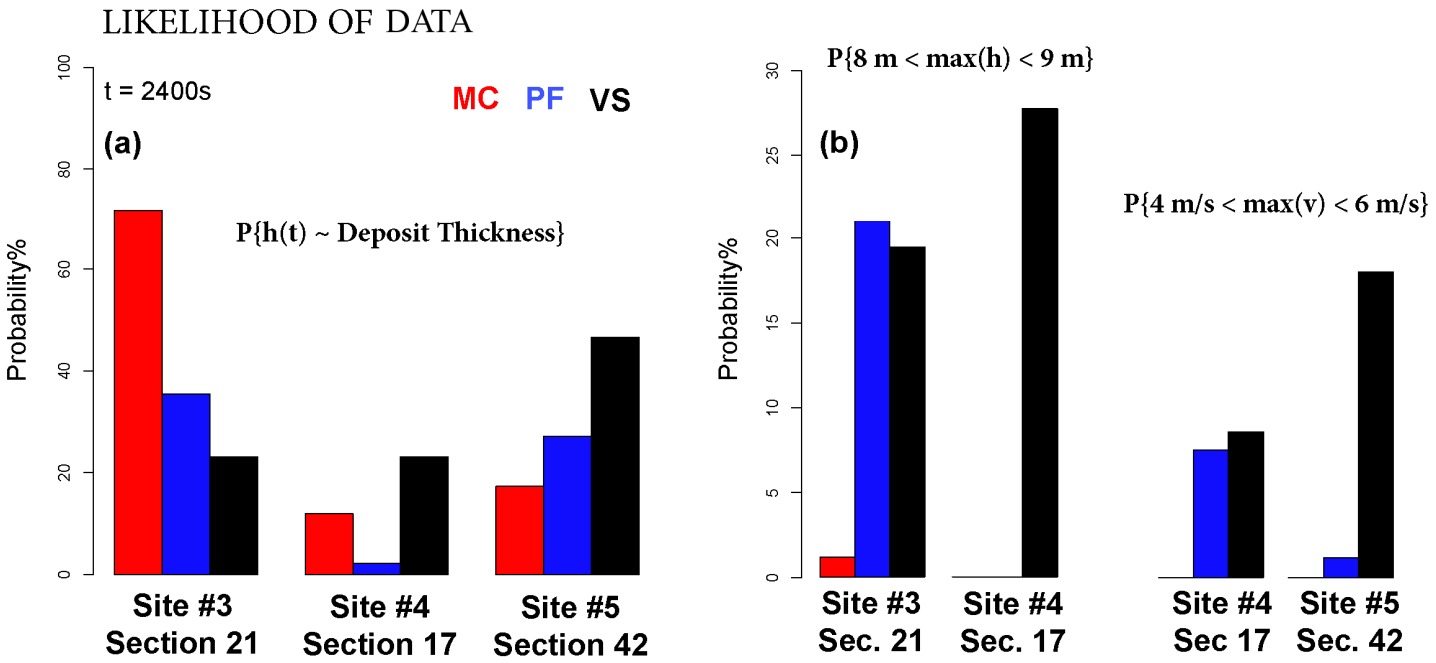
- [**deposits height**] MC performs well at Site #3, while VS at Site #5.
- [**maximum flow height**]
  PF and VS can replicate the values at Site #3, and only VS can replicate the values at Site #4.
- [**maximum flow speed**] PF and VS perform moderately well at Site #4, while only VS at Site #5.

Figure. Barplots of data likelihood in Sites #3, #4, #5.

(a) compares flow height at t=2400 s with observed deposit thickness (Saucedo et al., 2008).

(b) compares maximum height and maximum speed with observed wave height (Ponce Segura et al., 1983) and peak flow speed (Pierson, 1985).

Models are displayed with different colors.



Partial solutions in the input space – example #1

The figure displays two examples of **partial solutions** in the specialized experimental design.

For each example n=1,2 we select a **subfamily of empirical data** $(D_i)_{i \in I_n}$ and define, $\forall$ j:

$$\Theta_n^j := \bigcap_{i \in I_n} \Omega_i^j$$

[Example #1]
We consider the deposit thickness at Site #5.
Then, Example #1 evaluates maximum flow height at Site #4, and the maximum flow speed at Site #5.

In MC and PF the maximum flow height from the data is never reproduced. Thus, $\Theta_1^j \neq \emptyset$ only if j corresponds to **VS**.

MC model is not capable of reproducing the required maximum flow height and speed in the village.

Bevilacqua et al., (2019) includes an additional example focusing on **deposit thickness** in the three sites. That is without a solution. Deposit at Site #3 is inconsistent with deposit downstream in all models.

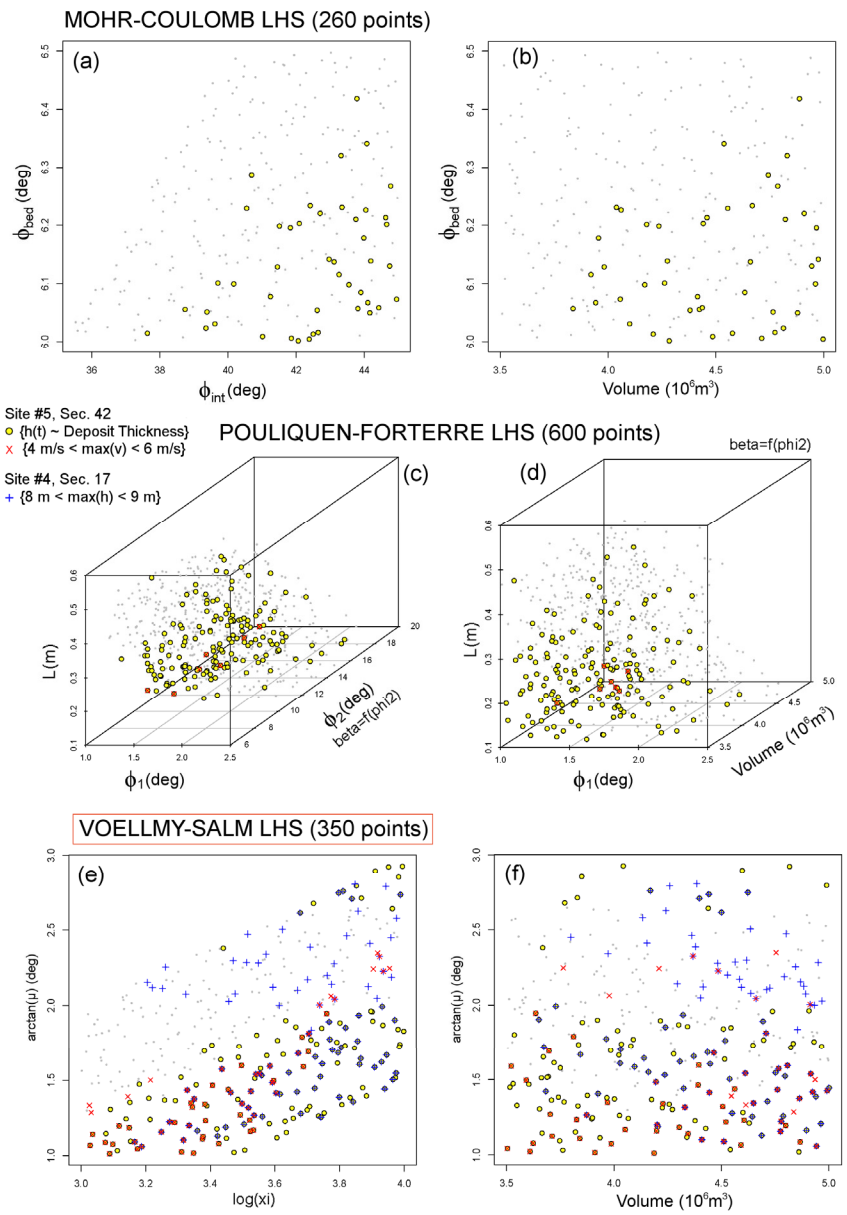


Figure. Example #1 of partial solution inputs in (a-b) MC, (c-d) PF, (e-f) VS.

The color expresses the considered data:
yellow is deposit thickness in Site #5,
blue is wave height in Site #4,
red is flow speed in Site #5.

(a-c-e) and (g-i-k) are projected along the V coordinate,
(b-d-f) and (h-j-l) along φ_{int}, φ_2 and ξ coordinates, respectively.

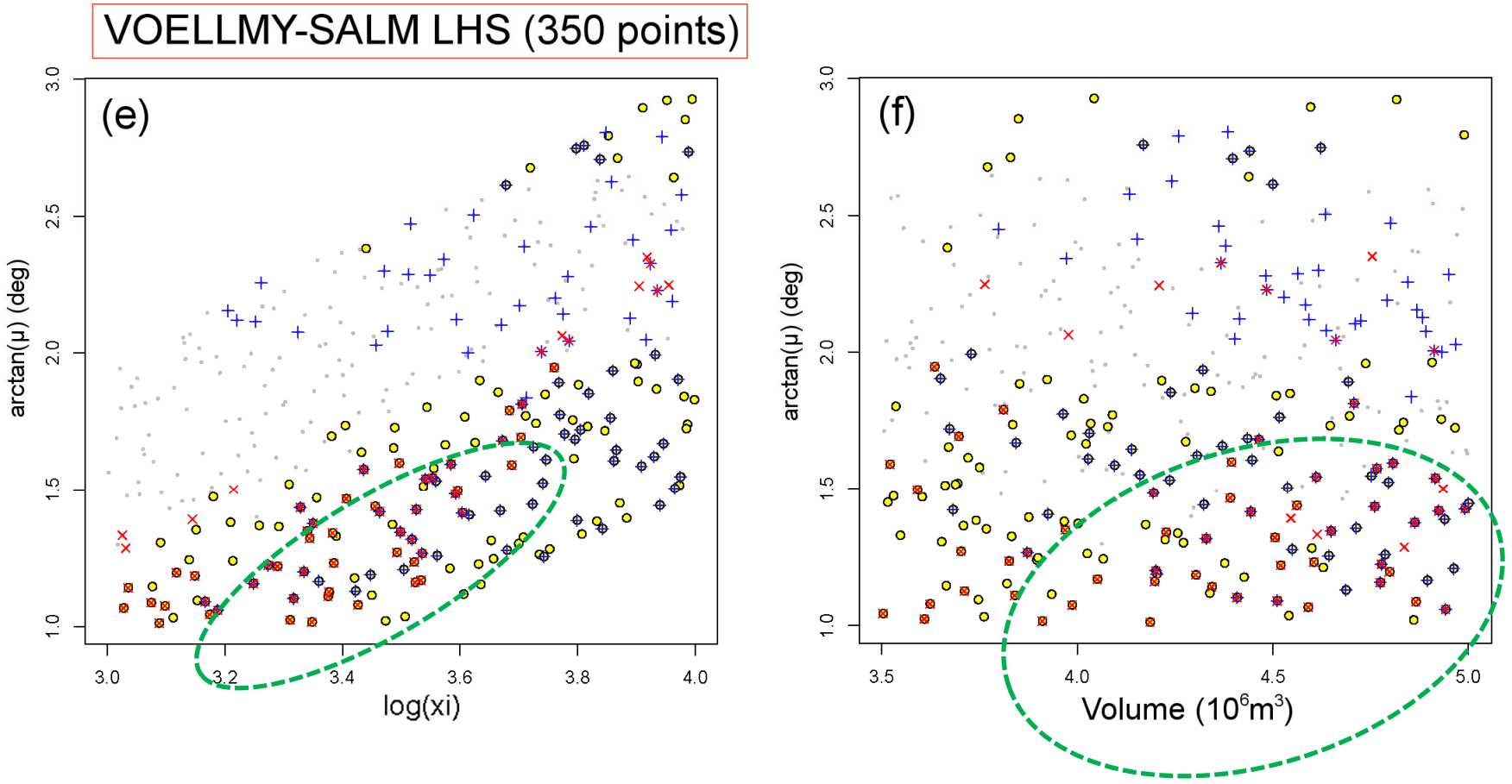
Partial solutions in the input space – example #1

Thus, $\Theta_1^j \neq \emptyset$ only if j corresponds to **VS**, and that set is bounded by:
$\arctan(\mu) \in [1.0, 1.8]$, $\xi \in [3.1, 3.7]$, $V \in [3.8, 5.0] \times 10^6$ m^3.

Figure. Example #1 of partial solution inputs in (e-f) VS.

The color expresses the considered data:
yellow is deposit thickness in Site #5, blue is wave height in Site #4, red is flow speed in Site #5.
A green ellipse bounds the partial solution set.

(e) is projected along the V coordinate and (f) along the ξ coordinate, respectively.



Partial solutions in the input space – example #2

[Example #2]
We consider the deposit thickness at Site #5.
In addition, we evaluate the max. flow height at Site #3, and the max. flow speed at Site #4.

In MC, the required maximum flow speed is, again, not reproduced.
We have that, $\Theta_2^j \neq \emptyset$ if M_j is **PF or VS**.

In VS, the partial solutions set is small and **close to the frontiers** of the uncertainty ranges (see the arrows in the figure):
$\arctan(\mu) \approx 1.2$, $\xi \approx 3.1$, $V \approx 3.6 \times 10^6$ m^3.
We do not detail it further.

Thus,
PF reproduces the required height and speed **when impacting** the village (Example #2), but only **VS** maintains them in the **downstream** part of the village (Example #1).

Solving the partial inverse problems enables us to **select a model**, which nevertheless depends on the type of data and site:
- **VS** in Example #1,
- **PF** in Example #2.

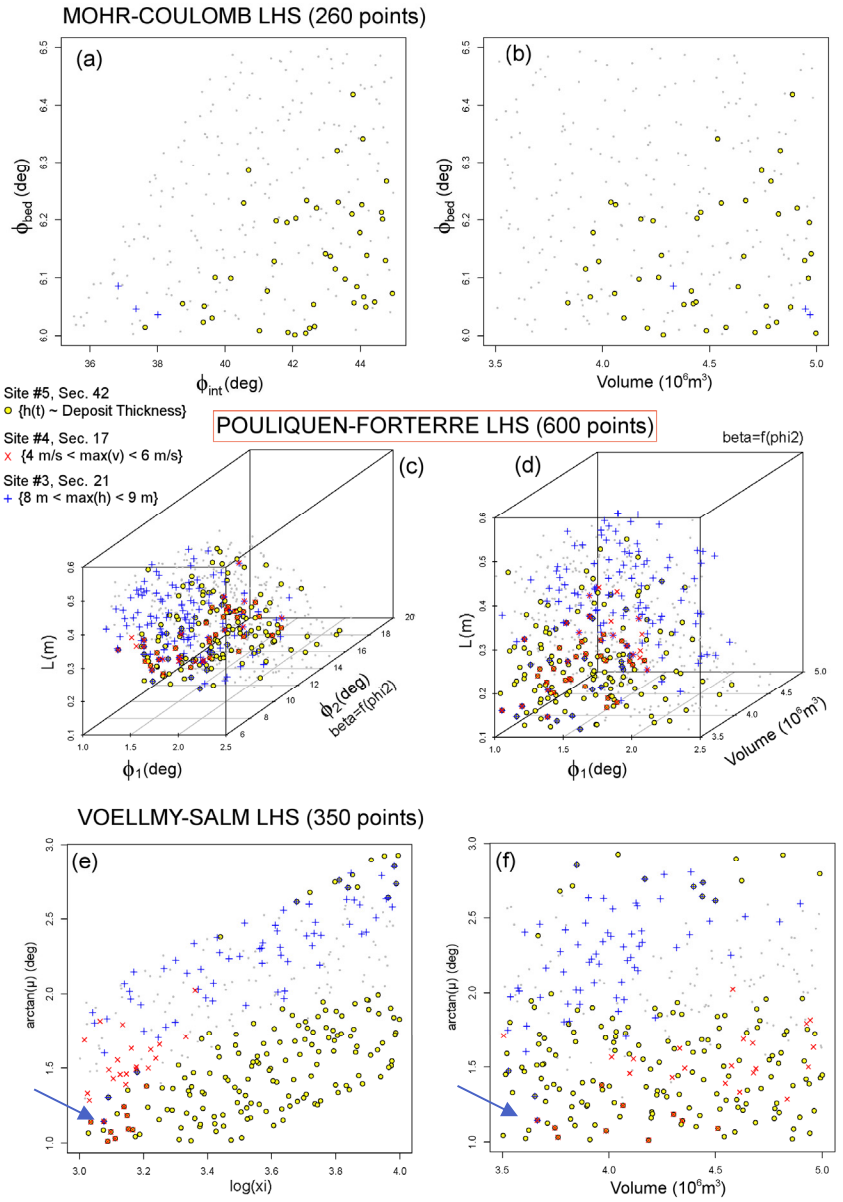


Figure. Example #2 of partial solution inputs in (a-b) MC, (c-d) PF, (e-f) VS.
Yellow is deposit thickness in Site #5,
blue is wave height in Site #3,
red is flow speed in Site #4.

(a-c-e) and (g-i-k) are projected along the V coordinate,
(b-d-f) and (h-j-l) along φ_{int}, φ_2 and ξ coordinates, respectively.

Partial solutions in the input space – example #2

In PF the set is bounded by:
$\varphi_1 \in [1.0, 1.6]$, $L \in [0.12, 0.25]$ m, $V \in [3.9, 4.9] \times 10^6$ m^3.

Figure. Example #2 of partial solution inputs in (c-d) PF.

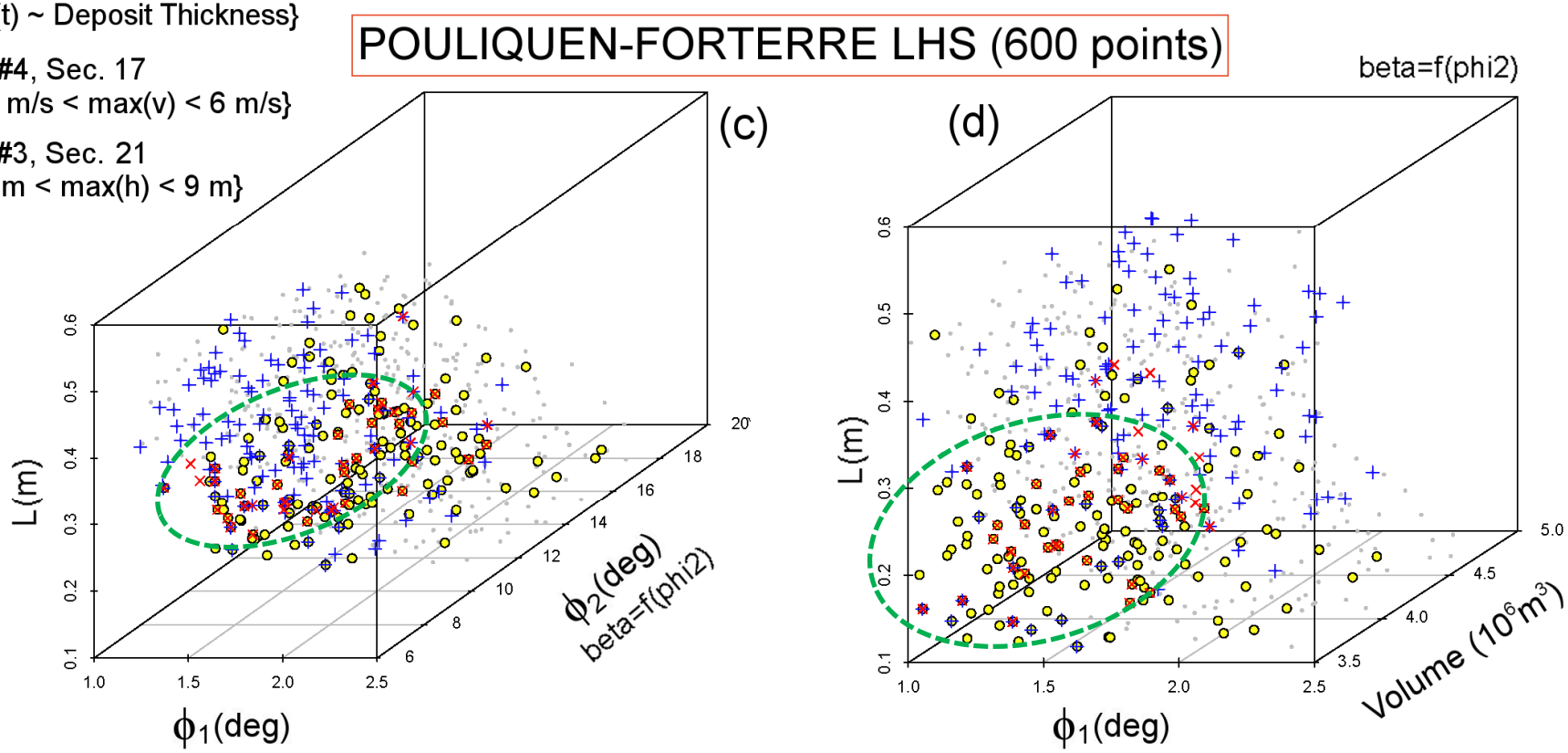
The color expresses the considered data:
yellow is deposit thickness in Site #5, blue is wave height in Site #3, red is flow speed in Site #4.
A green ellipse bounds the partial solution set.

(e) is projected along the V coordinate and (f) along the φ_2 coordinate, respectively.

Site #5, Sec. 42
○ {h(t) ~ Deposit Thickness}

Site #4, Sec. 17
x {4 m/s < max(v) < 6 m/s}

Site #3, Sec. 21
+ {8 m < max(h) < 9 m}



Examples of conditional results – example #1

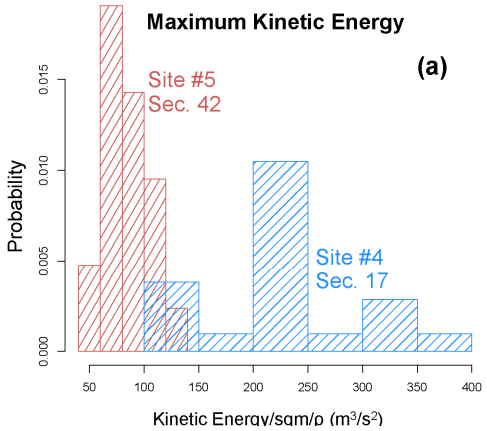
We obtain conditional spatial maps of mean of maximum in **flow height**, H, and **kinetic energy** κ.

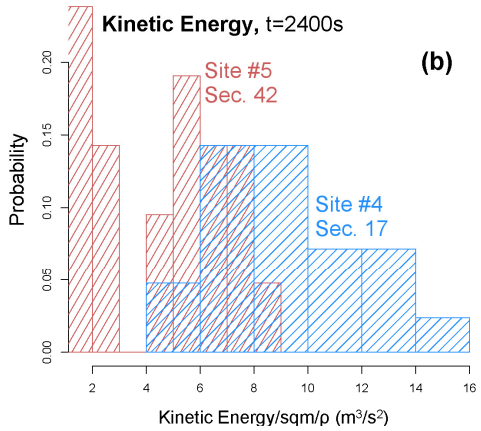
We also show maximum **dynamic pressure** Q.

$$\kappa := \frac{1}{2}\frac{(h\bar{u})^2 + (h\bar{v})^2}{h} \qquad Q := \max_{t\in T}\frac{\kappa}{h}$$

The dynamic pressure Q and the kinetic energy κ are scaled for a mass with unit density.
Typical values of density of **mudflows** are above 2000 kg/m^3.

CONDITIONAL RESULTS - VOELLMY SALM





The [4, 6] m/s constraint on maximum flow speed has an **immediate** effect on Q.

Imposing it at Site #5, as in the Example #1, or Site #4, as in the Example #2, radically changes the results.

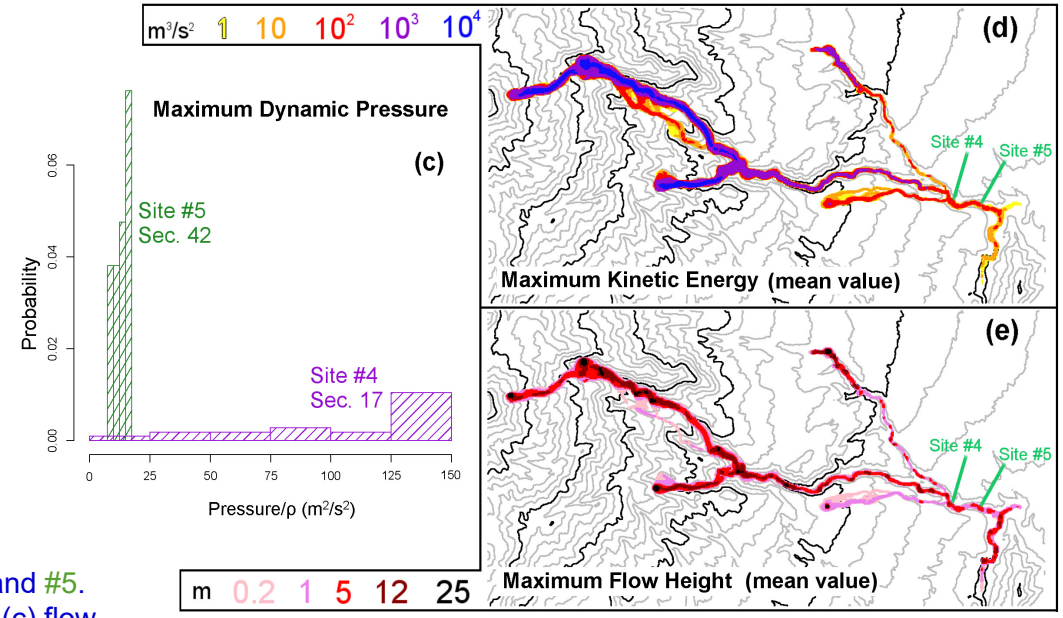


Figure. (a-b-c) Flow properties of VS model, in Example #1.

(a-d) Histograms of maximum dynamic pressure in Sites #4 and #5. Mean values of the maps of (b) maximum kinetic energy and (c) flow height as a function of time. Colors are related to their values.

Elevation contours are at intervals of 100 m and 500 m (NASA, 2014).

Examples of conditional results – example #2

In Example #1, Q∈ [0, 150] m^2/s^2 at Site #4, and Q∈ [7.5, 17.5] m^2/s^2 at #5.

In Example #2, Q∈ [8, 13] m^2/s^2 at Site #4, Q∈ [4, 7] m^2/s^2 at #5.

In the spatial maps **PF** shows slightly lower maximum flow height, and significantly lower energy, than **VS**.

The flow in the **tributaries** are able to reach the town of Atenquique.

These conditional results focus on the replication of specific properties of the 1955 flow and are not hazard maps for every potential scenario.

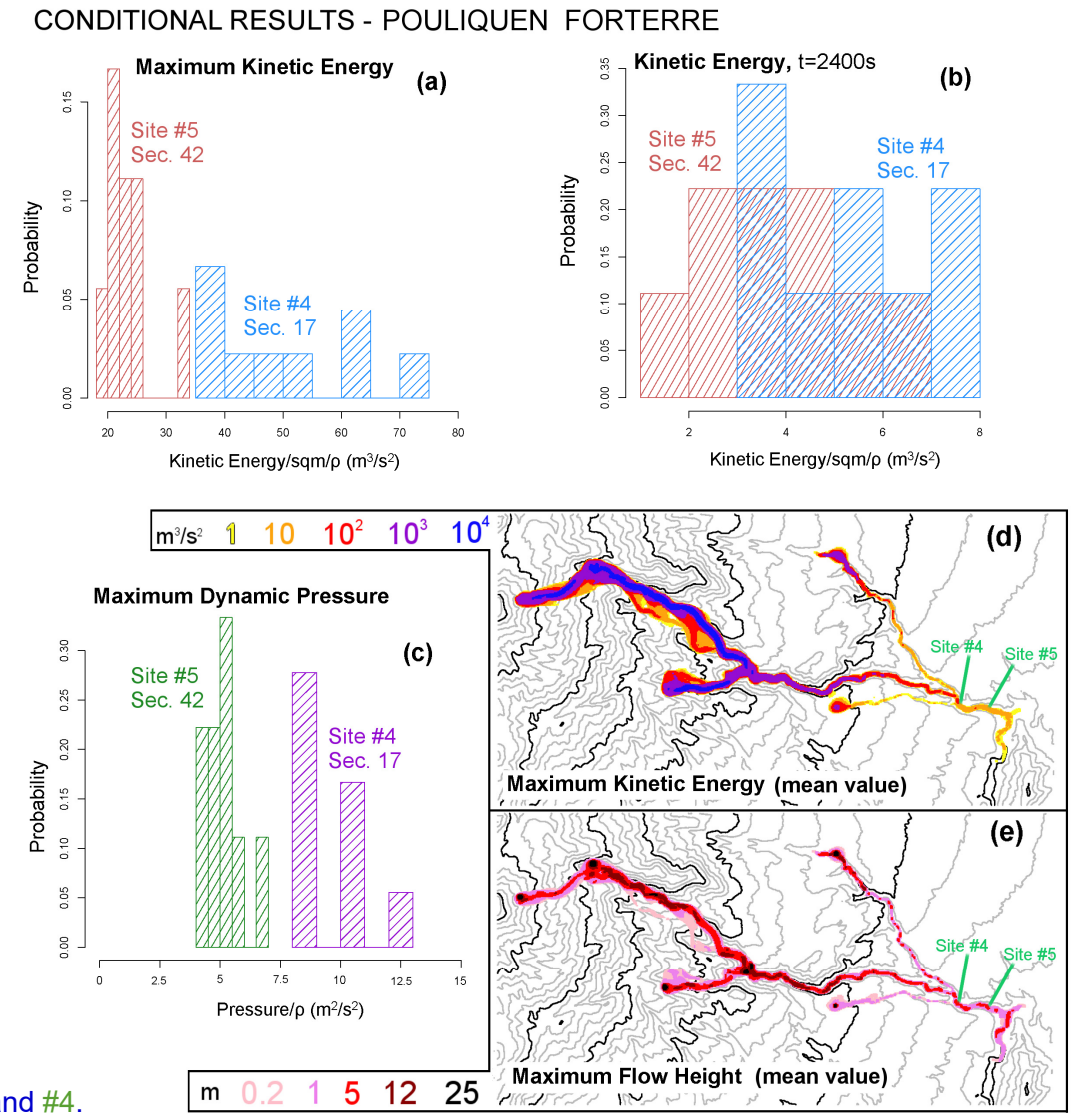


Figure. (a-b-c) Flow properties of VS model, in Example #2.

(a-d) Histograms of maximum dynamic pressure in Sites #3 and #4. Mean values of the maps of (b) maximum kinetic energy and (c) flow height as a function of time. Colors are related to their values.

Elevation contours are at intervals of 100 m and 500 m (NASA, 2014).

Conclusions

We introduced a prediction-oriented method for hazard assessment of debris flows.

In summary:

- We defined a **specialized experimental design** after assuming: the realism of the underlying physics, the numerical simulation robustness, and the meaningfulness of flow dynamics and inundation output. This contains valuable information for hazard assessment and this is a first step towards the development of an objective and **partially automated experimental design**.

- We calculated the **likelihood** that different model realizations reasonably represented the 1955 Atenquique flow, given **multiple pieces of field data**. The exercise provided useful information in either model selection or data inversion.

  The connection of inverse problems and model uncertainty is a challenge in the future development of multi-model solvers

- We constructed **partial solutions to the inverse problem**, conditioning the specialized experimental design to be consistent with subsets of the observed data.
  We found **model selection** to be inherently linked to the inversion problem.

  Additional tests at a finer resolution in the experimental design could be performed to achieve a more accurate characterization of the **conditional input spaces**, if required.

- Our results are consistent with **evolution of flow rheology** downstream in the vicinity of the village, from MC above the village, to either PF or VS within and downstream from the village. The meaning may reflect an evolution from inertial to macroviscous debris flow behavior near Atenquique.

Bevilacqua et al. (2019), Probabilistic forecasting of plausible debris flows from Nevado de Colima (México) using data from the Atenquique debris flow, 1955. https://doi.org/10.5194/nhess-19-791-2019

Acknowledgements:
This work is supported by National Science Foundation awards 1339765, 1521855, 1621853 and 1821311, and by Italian Ministry of Education, University, and Research, project FISR2017 - SOIR.