AEx-GRANIER: Global Sensitivity Analysis of Rockfall Trajectory

Ritesh Gupta, Emilie Rouzies, Franck Bourrier, Vincent Acary

AEx-GRANIER: Global Sensitivity Analysis of Rockfall Trajectory

Ritesh Gupta[a,*], Emilie Rouzies[a], Franck Bourrier[a,b], Vincent Acary[a]

[a] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, Institute of Engineering, LJK, 38000 Grenoble, France*
[b] *Univ. Grenoble Alpes, INRAE, CNRS, IRD, Grenoble INP, IGE, 38000 Grenoble, France*

Abstract

The present work report is a part of the project AEx-GRANIER (Action Exploratoire - GRAvitatioNal hazards in mountaIns in the contExt of Risks prediction — Nonsmooth modeling and simulation with data in Geomechanics). The Non-smooth contact dymanics (NSCD) method based numerical model, implemented in Siconos software package, is presented for the simulation of gravitational natural hazards in the mountain environments. The fundamental idea is to narrow down the existing gap between the real event data and the numerical models towards hazard prevention and risk prediction. In this work, the NSCD model simulations are combined with the so-called 'data-driven modelling' techniques to access the descriptive statistics of the phenomenon. The characteristics of the rockfall process in a given terrain are highlighted by statistically presenting a correlation and quantification of the uncertainty between input and output parameters.

Keywords: Prediction, Prevention, Data driven modelling, Uncertainty Quantification

1. Introduction

Climate change has many consequences on natural gravity hazards in mountain environments, both in terms of intensity and frequency. The objective of this exploratory action concerns the application of digital sciences for the prediction and mitigation of the risks associated with these hazards. The mechanical modeling of gravitational natural hazards in mountain environments is a crucial subject for the quantification of the risks associated with these hazards because the mechanical hazard models make it possible to quantify the zones affected by the hazard and the intensity of the hazard at the level of these areas. The data to calibrate these models are few, spatially poorly resolved, do not concern all the output quantities of the models. In particular, only the affected areas are generally documented.

The complexity of the physical phenomena involved, the diversity of the spatial scales to be treated (from the material scale to that of the slope) and the variability of these hazards are still only imperfectly translated by the many existing approaches for this type of modelization. Furthermore, the question of the use of event data for these hazards within mechanical models to better quantify the uncertainties associated with the forecasting and predetermination of these risks constitutes a major challenge. For example, the numerical simulation of gravitational flows (falling blocks, rock flow, debris flows, snow avalanches, ice avalanches, etc.) has certainly reached a certain level of maturity, but its use in a prediction and prevention framework is still in its infancy, mainly for lack of direct confrontation with data.

This work explores the rockfall trajectory through 'data- driven modeling' methods to explore the gravity force induced trajectory of the falling rock. It involves making the most of laboratory and observatory data

*Corresponding author
Email address: ritesh.gupta@inria.fr (Ritesh Gupta)

in order to be able to build and calibrate models, assess their sensitivity, improve their predictive nature, *i.e.*, control and take into account uncertainties, thanks to variational, statistical and AI methods. In turn, it is hoped that this provide guidelines to improve data generation and sustainability. These methods, which have already been tested in the context of climate modelling, have hardly been used for gravitational flows and complex rheologies of the frictional cohesive type, which are intrinsically non-smooth. This constitutes the highly exploratory nature of this research.

For the study of the falls of unit blocks and the physical phenomena involved during the propagation of rockslides, the methods making it possible to model explicitly the constituent particles of granular materials (in particular the Methods with Discrete Elements) are preferred. In addition, the extension of existing simulation methods in mechanics of complex flows in a non-smooth framework certainly makes it possible to greatly improve the robustness of the models while reducing the number of physical parameters, but many questions remain open as to the feasibility of such improvements in the context of non-smooth mechanics.

Sensitivity analysis aims at identifying which part of the output uncertainty can be associated to each input factor (or groups of input factors). Such a study can have several objectives: (1) Get insight into the model functioning, (2) Simplify the model if some parameters are identified as non-influential (screening), (3) Be able to focus calibration efforts on most influencing parameters in order to reduce uncertainty (ranking). Following this idea and relying on Dobre et al. [1], sensitivity analysis may bring information about how identifiable is one parameter in a data assimilation process (null Sobol' indexes may indicate that a parameter is not identifiable).

In this report, first the framework of NSCD numerical model development is presented. Further the explored terrain models for the rockfall simulations are presented. The statistical correlation between input and output parameters are established using descriptive statistical data visualisation methods. Lastly, the sensitivity analysis of the input parameters is presented. The work is concluded with a brief discussion on the challenges, applicability and the limitations in the adopted approaches.

2. Numerical Model development

2.1. NSCD framework

- Non-smooth contact dynamics: The Non-Smooth Contact Dynamics (NSCD) method developed by M. Jean and J.J. Moreau [2, 3] is a modelling and numerical framework for the dynamics of multi-body systems in interaction through frictional contact interfaces.

- Siconos software features: An NSCD-based model of the falling rock in a terrain is developed using the Siconos software [4]. Siconos is an open-source scientific software primarily targeted at modelling and simulating non-smooth dynamical systems in C++ and in Python, including mechanical systems, switched electrical circuits, sliding mode control systems, and biology. The NSCD method guarantees that Signorini's condition at the velocity level is satisfied without introducing contact stiffnesses and that the coulomb friction, especially in the sticking mode is satisfied without resorting to viscous friction. This peculiarity of NSCD is enforced with the use of a specific implicit scheme (Moreau–Jean scheme) and yields a robust numerical method, which consistently models the threshold phenomena (friction, contact) and the dissipation properties of the model in discrete time, in particular impact dissipation and energy properties [5]. In this work, the Coulomb friction with rolling resistance [6] is implemented to take into account the rotational movement impact of the falling rock block downhill a terrain.

3. Descriptive Statistics

The event of falling rock block downhill a terrain is modelled in the NSCD framework. Courtesy of the NSCD model computational speed, a large number (in thousands) of these events are simulated within one hour. The rock block in each simulation is characterised by nine input parameters presented in the following section. The distribution of the stopped rock block position in the terrain domain is considered the basis to observe descriptive statistics on input-output parameters correlation and sensitivity.

The estimation of the uncertainty and the sensitivity of the rockfall trajectory on the input parameter is expected to be complicated due to the chaotic nature of the rockfall process. By definition, a phenomena is said to be of chaotic nature when even if it is expressed in a deterministic setting yet still be inherently unpredictable over long periods of time.

This is because a specific trajectory of the falling rock is an accumulation of multiple contacts with the terrain profile. (see Appendix B for reference) Thereby, a slight change in any one or more rockfall parameter(s) magnitude renders each new contact response being different. This influences the trajectory and ultimately results into very different stopped positions of the blocks.

Hence, the usage of typical sensitivity analysis methods such as Sobol indices (see Appendix C for reference) is rendered not suitable to observe and interpret the rockfall phenomenon. Notably, these methods were initially tested for the rock fall simulations in the context of this project, whose details can be referred to from Appendix E.

Therefore, the descriptive statistics is rather carried out to observe a global significance of model parameters in the output response. This strategy enables a practitioner to exploit the following questions,'*Which parameter should be tuned to let the block fall in one or the other location/zone of interest.*' and '*what is the relative distribution of the blocks in different locations in the terrain.*'

3.1. Input parameters definition and distributions

The uncertainty both due to a lack of knowledge on parameters (epistemic uncertainty) and due to natural variability (aleatory uncertainty) are considered in order to be as exhaustive as possible. Input factors considered are linked to parameters : friction coefficient μ, rolling friction coefficient μ_r, coefficient of restitution e_n, but also to initial orientation defined by three rotation parameters θ_x, θ_y, θ_z and to block shape. The Sneed and Folk ternary diagram [7] provide a basis to characterize rockfall shape. Block shapes are classified into 10 categories (see Table 1) that mix compact (cubic), platy (tanular) and elongated (rod shaped) shapes.

Shape index	Class name	$\frac{C}{A}$	$\frac{B}{A}$	Probability of occurence
1	Compact	1	1	0.5
2	Compact platy	2	2	0.1
3	Compact bladed	1.5	2	0.1
4	Compact elongated	1	2	0.1
5	Platy	4	4	0.055
6	Bladed	3	4	0.055
7	Elongated	1	4	0.055
8	Very platy	12.5	12.5	0.01
9	Very bladed	4	12.5	0.015
10	Very elongated	1	12.5	0.01

Table 1: Description of shape classes adapted from Sneed and Folk ternary diagram [7].

The different classes can be referred to at Bonneau et al. [8]. The y-ratio and z-ratio values are taken as a continuous variable to establish a continuity in the rock shape parameter. In addition to shape class, a block is also characterized by its volume, which is defined as a convex mesh with a fixed number of vertices normalised by the aspect ratio and the desired rock volume parameters.

Distribution types and bounds are chosen according to expert knowledge. While distribution shapes are the same for both site, bounds must be chosen according to local properties. A truncated normal distribution is assigned to the friction parameter μ while a lognormal distribution is assigned to rolling friction parameter μ_r and a triangular distribution is assigned to the restitution coefficient e. Initial rotations and volume are associated to uniform distributions. As the shape index is a discrete value and the different shape categories are not equiprobable, a categorical probability is associated to each class. The input factors considered in this study and associated distribution for both sites are described in Table 2. Notably, a correlation coefficient of 0.5 is assigned to the y-ratio and z-ratio parameters such that the latter is always equal or inferior to the former respectively.

Parameters	Input factor	Symbol	Distribution
Mechanical	friction coefficient	μ	TN(0.7, 0.1, 0.4, 1)
	rolling friction coefficient	μ_r	LN(-1.29, 0.12)
	coefficient of restitution	e_n	T(0, 0, 0.2)
Block geometry	y-ratio [-]	Y_r	EX(0.75,1)
	z-ratio [-]	Z_r	EX(0.75,1)
	volume [m^3]	V	U(0.25, 10)
Initial orientation	x-axis initial rotation	θ_x	U(0, 2π)
	y-axis initial rotation	θ_y	U(0, 2π)
	z-axis initial rotation	θ_z	U(0, 2π)

Table 2: Input factors considered in the global sensitivity analysis and corresponding distributions. Statistics of the assigned distributions are defined as follows: uniform: U (min, max), triangular: T (min, mean, max), log-normal: LN (μ, σ), normal: N (μ, σ), truncated-normal: TN (μ, σ, min, max), exponential: EX (λ, γ)

The sets of input parameters for each rockfall event are sampled using Latin hypercube sampling (LHS) method. The statistical sampling method provides a random selection of the input parameters within their

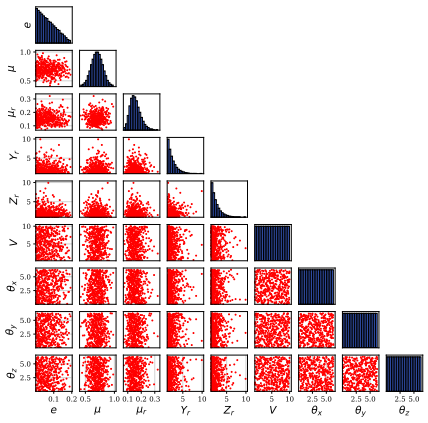


Figure 1: Input sample distribution for the rock block parameters. Here, a sample size of 400 is presented for readability.

predefined ranges such that the input space is well-filled. The distribution of these parameters within their bounds is presented in Figure 1. Here, all parameters except Y_r and Z_r are not correlated as visible by the random distribution of the scattered input points (red dots) in the sample domain. Notably, the defined correlation between Y_r and Z_r can be seen in the scatter plot where all red dots (except one) are below the virtual diagonal in the box. Notably, the variability of the numerical parameters corresponding to the SICONOS model definition were also examined (see Appendix A). However, their global influence is reported to be limited in comparison to the model parameters (stated above) and hence not considered further in the present work. The details on the numerical parameters variability can be referred to in .

3.2. Explored terrain models

A total of four terrain profiles are examined to explore the trajectory of the falling rock namely Biplane', 'Dole P1', 'Dole P2' and 'Vaujany' terrain model. The first terrain 'bi-planer' represents a user-defined simple profile a higher inclination of the first slope which changes to 10° for the second slope. The last three terrain profiles represent the actual on-site digital elevation model (DEM) profile retrieved from the on-site campaigns (in Alpes mountain range). Notably, the starting point of the falling rocks for any explored terrain is the same. It is identified by the site investigations for the last three terrain profiles, while it is considered at a 2% (of the total x-length of the slope) to the right of the starting point of the slope.

3.3. Biplane terrains

The 'Biplane' is a user-defined terrain model with an initial slope where the rockfall starts followed by a 10° slope of sufficient length and width to stop nearly all blocks within the terrain model dimensions. This terrain model presents the simplest case to simulate the trajectory of the falling rock having varying geometrical, orientation and mechanical characteristics within the ranges reported in the Table 2.

The biplaner terrain, due to its simple profile, enables to explore the influence of the terrain in the rockfall trajectory (*i.e.*, considering it as an independent variable. For the same set of input parameters, eight different incident slopes are explored varied between 25° and 60°. The second slope is kept constant at 10° to keep the variability due to the terrain profile to a minimum. Also, the height of fall in the upper

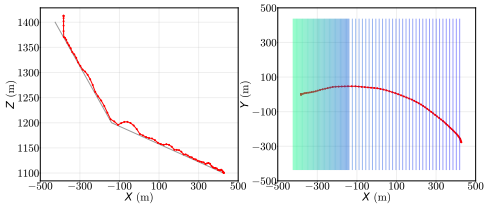


Figure 2: The test run of a rockfall on 35°-10° biplaner terrain. Here, the dotted red line presents the trajectory of the test rock block.

slope is kept constant at 200m for all eight biplaner terrain models. The design of the biplaner terrain and the trajectory of the falling rock are illustrated for a test rockfall in Figure 2.

Further, multiple biplaner terrains with different slopes are simulated to access the effect of terrain profile as a input parameter. The test rockfall run on different terrain profiles is presented in Figure 3.

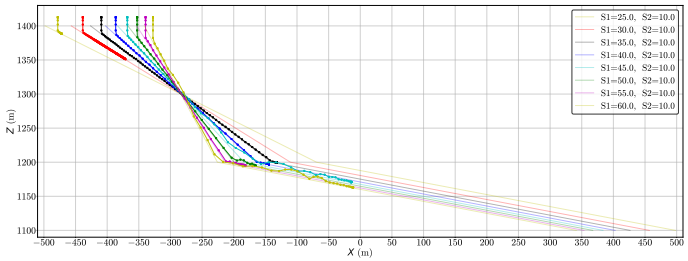


Figure 3: The test run of a rockfall on multiple biplaner terrains with varying incident slope between 25° to 60°. Here, the dotted line presents the trajectory of the same test rock block in different terrains.

The stopped block positions of 9000 different sets of input parameters (obtained from the LHS statistical sampling method) for all eight terrain profiles are presented in Figure 4. Here, all blocks outside the terrain profile are marked in red and their count is presented in the respective label.

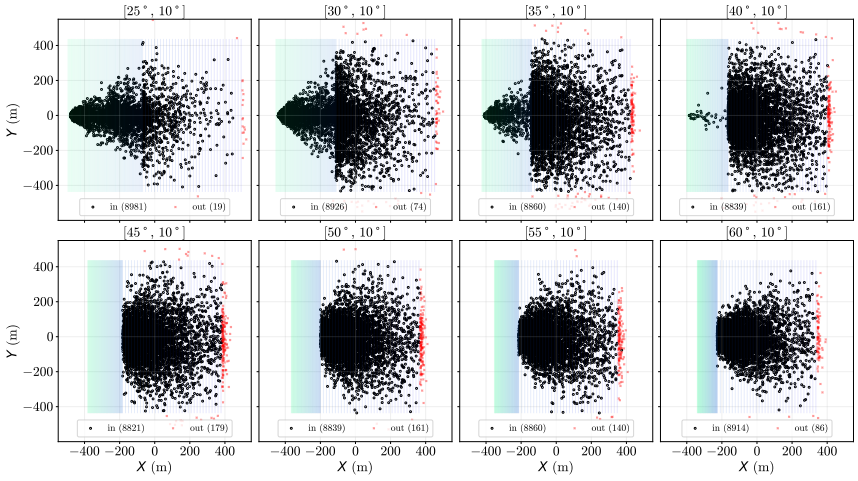


Figure 4: The stopped block position for 9000 sets of input parameters for eight biplaner terrains with varying incident slopes

The relative distribution of the stopped blocks in the terrain presents the qualitative indication to the effect of terrain profile. Further, it is of the practitioner's interest to observe the relative position of a specific

stopped block in different terrains. For this, as established above, tracking a single block is cumbersome due to chaotic nature of the followed trajectory of a falling rock. Therefore, the descriptive statistics are exploited to address the practitioner's concerns. In other words, the whole process of the stopped blocks position distribution is exploited as a 'classification' problem instead of a 'regression' problem.

This presents the interest of the sensitivity analysis is local to specific component of the whole terrain. This is to understand that how the stopped blocks are distributed in amongst themselves and which input parameter to tune such that the falling block falls into one or the other cluster.

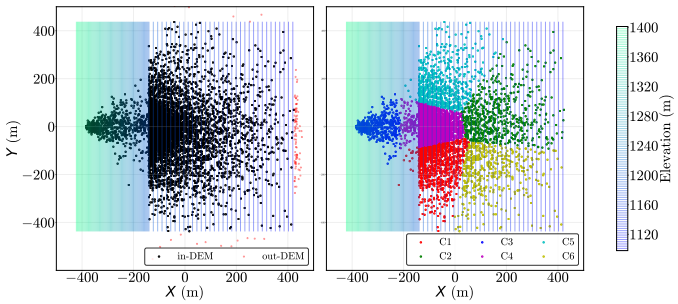


Figure 5: Blocks stopped positions for biplane terrain of 35°-10° and corresponding clustered positions obtained from the k-means clustering method

Now, there are some statistical methods available to divide the scattered cloud of points into different clusters, *e.g.*, k-means clustering. This is illustrated for 35°-10° terrain model in Figure 5. However, such clustering method is based on the relative distribution of the scattered points and therefore is dependent on the number of simulations. Besides, such method does not aid the practitioner as the location or zone of interest is typically already known. For instance, a civil occupied zone in a valley is of interest for the practitioner to ensure the falling rock block either does not enter such zone or is well intercepted beforehand through installed protection structures (such as flexible barriers).

Therefore, the manual clustering practice is opted for the biplaner terrains where clusters are divided with a geometrical correlation with the incident slope angle and height of all. The manual clustering practice designed in this work, keeping the variability of the terrain is demonstrated in Figure 6.

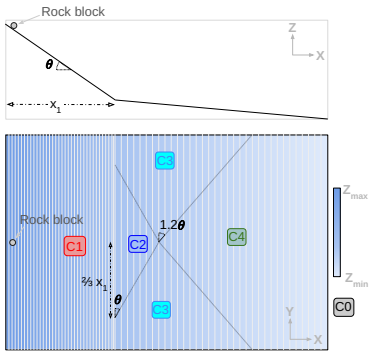


Figure 6: Biplaner terrains manual clustering (C) where the rock blocks stopping in steeper slope are classified in C1, in the triangular wedge at the start of gentle slope are classified in C2, in the left and right sides of the gentle terrain are classified in C3, in the lower part of the gentle slope are classified in C4 and the ones falling out of DEM are classified in C0.

Following the clustering division format, the 9000 simulated input set of falling rock parameter for each

158 of the eight biplaner terrains are classified in different clusters, as presented in Figure 7.

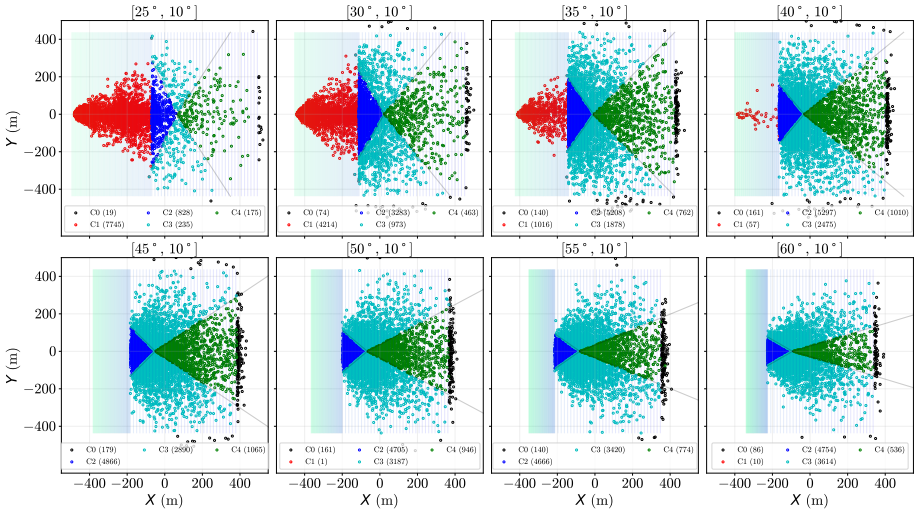


Figure 7: All Biplaner slopes with 5 manual clusters based on 9000 sets of input parameters

159 Here, the blocks are reported to stop in C1 (*i.e.*, in the steeper slope) up to the 40° angle. This is
160 consistent with the tangent of the maximum of friction coefficient $\mu \approx 1$. The highest number of blocks
161 stop in the C2 (*i.e.*, in the triangular wedge). The cluster C3 is divided into two symmetric sub-regions
162 accounting for the blocks with transversal trajectory characteristics. The cluster C4 takes into account the
163 blocks stopping at the lower part of the gentle slope segment characterising the blocks with highest trajectory
164 lengths. Lastly, the cluster C0 accounts for the blocks which fall out of the DEM on either side or up-front
165 characterising the possibility to take into account block input parameters which shall be later removed from
166 the full input set to conduct a global sensitivity analysis for a particular terrain model.

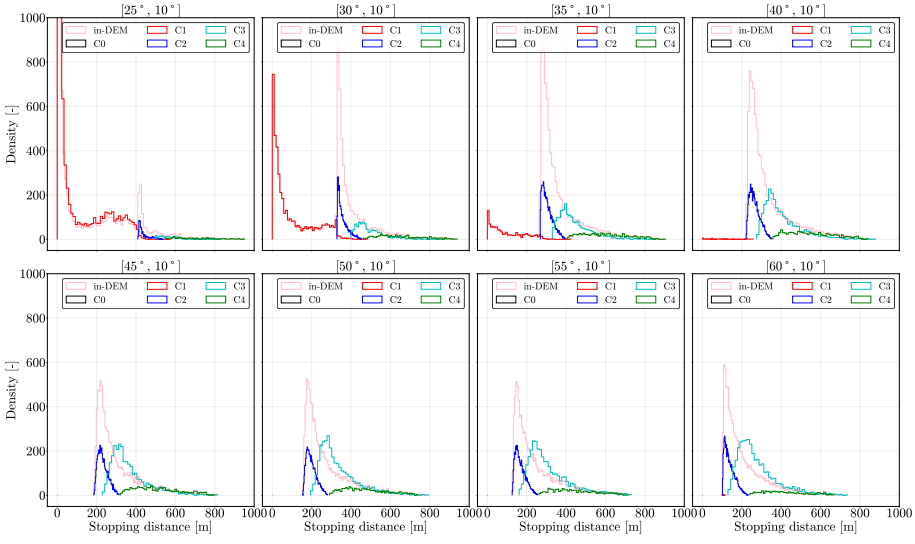


Figure 8: All Biplaner slopes with 5 manual clusters stopping distance

167 Further, the distribution of the stopping distances for the full terrain model as well for each cluster are

presented in Figure 8. The stopping distance is computed as the 2D relative difference between the block's final and initial positions expressed as:

$$l_{stop} = \sqrt{(x_N - x_0)^2 + (y_N - y_0)^2} \tag{1}$$

Notably, the blocks falling out of DEM are recorded with an arbitrary stopping distance of -9999 m, hence not included in the plot. Here, each cluster present similar qualitative distribution for different terrain models, emphasising the generality of using the correlated clustering division technique for different terrain models.

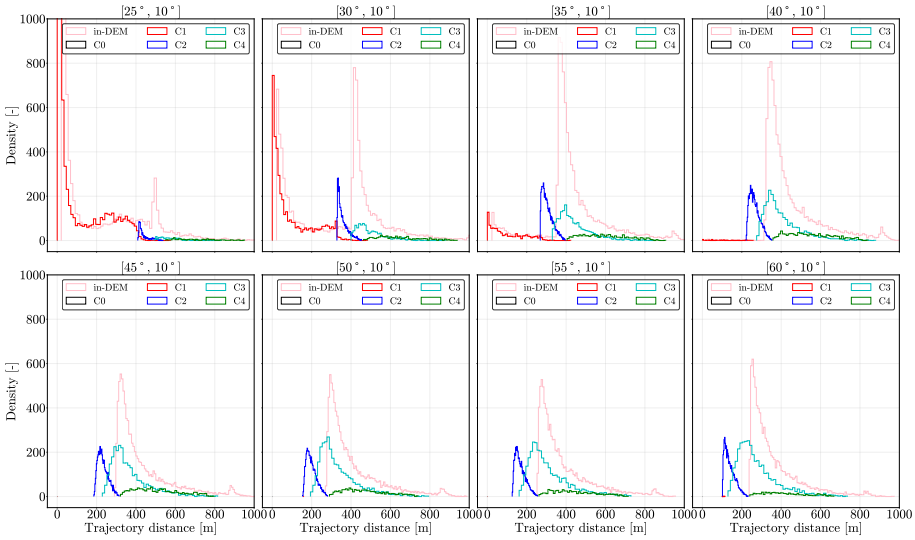


Figure 9: All Biplaner slopes with 5 manual clusters trajectory distance

Similar to the stopping distance, the distribution of the trajectory length of each block amongst the set of 9000 blocks is presented in Figure 9. Here, the trajectory length is computed by summing the distance travelled at each time step as:

$$l_{traj} = \sum_{t=1}^{N} \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2 + (z_t - z_{t-1})^2} \tag{2}$$

Here, the trajectory length estimate for the blocks falling out of the terrain is also added. This however poses a potentially misleading distribution for the full terrain model as the finite estimate of the trajectory length from the block falling out of terrain is mixed within the ones stopping within the terrain domain. Nonetheless, the cluster-wise trajectory length distributions enable the clear identification of such trajectory lengths. Besides, similar to the stopping distance distributions, the trajectory length distributions are qualitatively similar for all terrains. This additionally emphasises the generality of using the correlated clustering division technique for different terrain models.

Lastly, the falling rocks are characterised by the energy line computation. It takes into account the tangent of the ratio between the elevation lost and 2D planer distance travelled between final and the initial position of the block, expressed as:

$$m_{el} = \arctan\left(\frac{z_N - z_0}{\sqrt{(x_N - x_0)^2 + (y_N - y_0)^2}}\right) \tag{3}$$

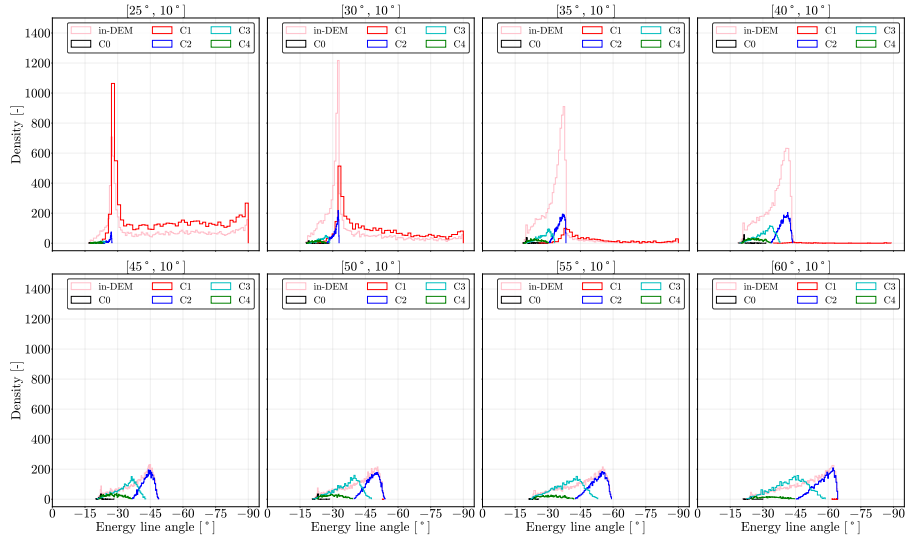


Figure 10: All Biplaner slopes with 5 manual clusters energy line

Overall, a qualitative similarity in the distribution of clusterwise response is observed in all three Figures above. However, the distribution of the input parameters corresponding to one specific trend, for instance log-normal distribution of the stopping distance in cluster 2 for all terrains or the clusterwise input parameter distribution for the corresponding output distribution cannot be conclusively inferred to be the similar as well.

For this, the comparative histograms are plotted first for one specific terrain, presented in Figure 11 and then for all terrains altogether for each cluster, presented in Figure 12.

Besides, the chaotic nature of the rockfall process makes it difficult to carry out the sensitivity analysis using for example Sobol' indices. However, the clusterwise input and output parameter distribution enables to access the sensitivity of the process. This is demonstrated in Figure 11 for 35°-10° terrain model.

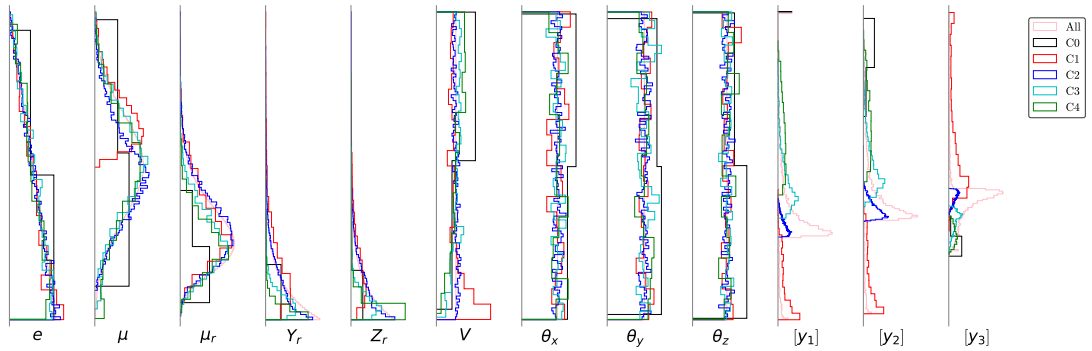
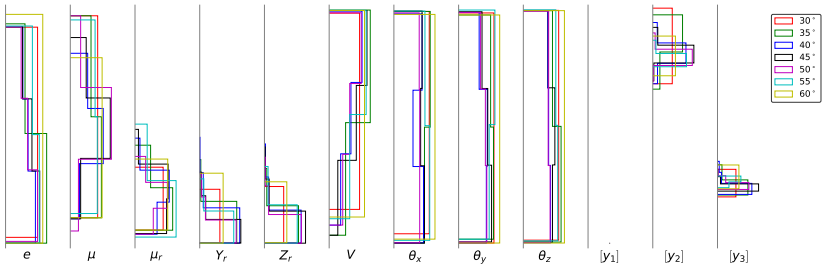


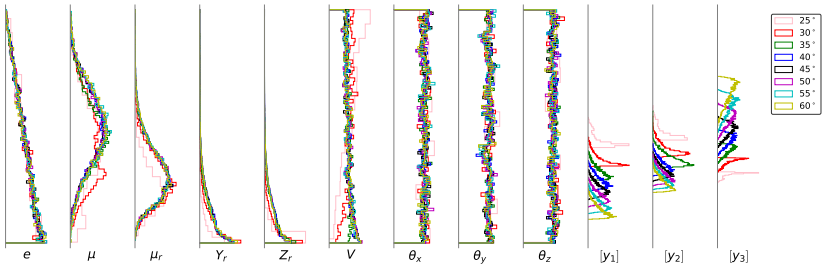
Figure 11: Histogram for all and cluster-wise input-output correlation for 35°-10° terrain model

Here, the number of simulated cases in each bin are kept constant which is also observed through the relative width of each step in the histogram corresponding to different cluster output. Also, the non-clustered distribution is also presented to illustrate the significance of carrying out a clusterwise analysis for this physical system.

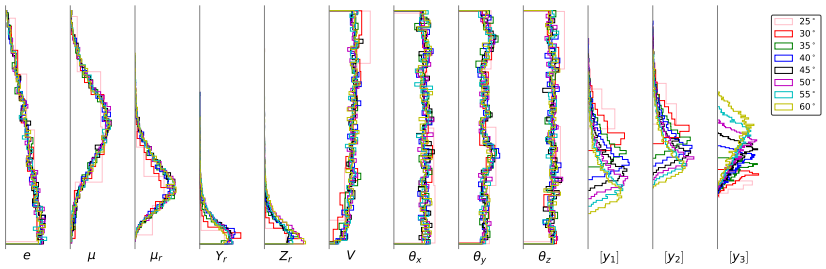
The presentation of histogram is also extended to understand the effect of the terrain model in each cluster distribution of the input-output parameters. This is presented for all clusters (except C1 because of its absence in half of the explored terrain models) in Figure 12.



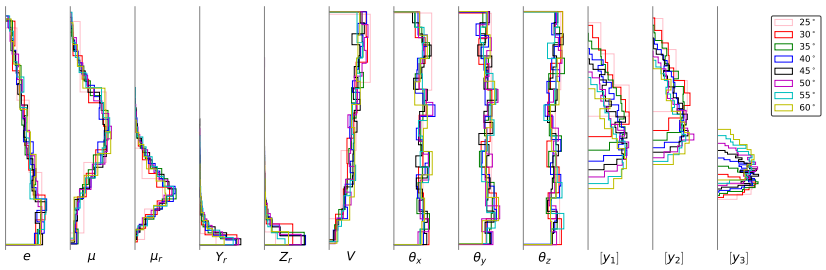
(a) Cluster 0



(b) Cluster 2



(c) Cluster 3



(d) Cluster 4

Figure 12: Clusterwise histogram for input output correlation for all clusters (except C1). Here, output $[y_1]$ refers to the stopping distance, $[y_2]$ refers to the trajectory length, and $[y_3]$ refers to the energy line

Additionally, the data visualisation with Cobweb plots, [9], allow to access the input-output correlation as discrete observation combined together for multiple rockfall events simulations to exploit the descriptive statistics. The Cobweb plot is first demonstrated for 35°-10° terrain model with a limited number of model simulations, presented Figure 13.

Here, a limited number of simulations are presented for visualisation. The solid line output ($[y]$) screen

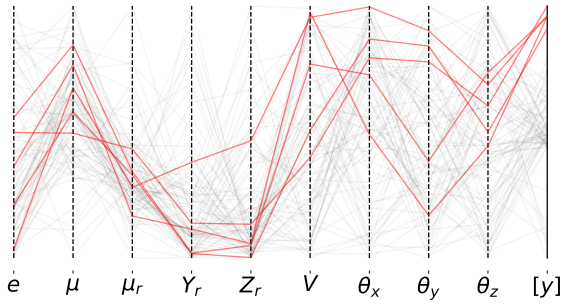


Figure 13: Demonstration of Cobweb plot for 90 sets of input parameter correlation with the output of the stopping distance for 35°-10° terrain model. Here, the top 5% quantile stop distance outputs and the chain of input parameters connected to those outputs are highlighted in red.

represents the stopping distance magnitudes for the simulated cases. The back tracking of each line to extreme left passes through the corresponding input parameter magnitude collectively linking input to the output of interest. As per user's choice, a subset of such simulations can be extracted by assigning the relevant filtering condition. Here, it is demonstrated for top 5% quantile of the stopping distance.

As an extension, the Cobweb plot, presented in Figure 14 for top 5% quantile of the trajectory distance in each cluster enables observation of the sensitivity of each input parameter. Notably, here, the trajectory length output is selected instead of stopping distance to enable accessing a finite measurement for the blocks falling out of DEM.

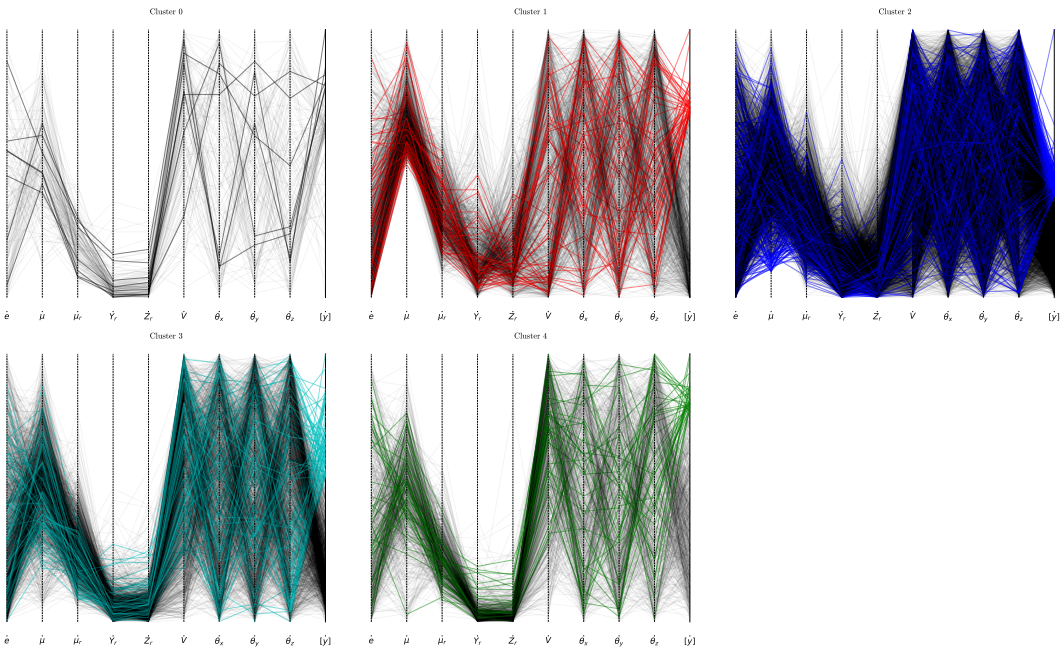


Figure 14: Cobweb plot for all and cluster-wise input correlation with top 5% quantile of the stopping distance for 35°-10° terrain model

Here, despite a random distribution of the input/output parameters correlation, some observations can be inferred. For instance, the top 5% quantile of blocks stopping in the Cluster-1 have relatively elongated or higher class (see Table 1) shape. The blocks stopping in Cluster-4 predominately have compact shape and low magnitudes of μ and μ_r. The blocks falling out of DEM (Cluster-0) can be expected to have high e to felicitate high post-contact velocity. However, the Cobweb plot reveals even the low values of e resulting in the same response. Also, the block volume (V) shows limited influence except for Cluster-0 where large volume of blocks contribute to it's falling out of DEM. Lastly, the block orientation in all three directions is observed to show no clear trend to distinguish it's individual influence on the overall process. Such type of Cobweb plots can also be presented for other terrain models.

Further, the Cobweb plots can also be used to access the relative distribution of different input parameters for blocks stopping at a specific zone in all terrain models. This is demonstrated in Figure 15 for all terrain models. Here, multiple output screens are used, identified by the x and y coordinates (x_{pos} and y_{pos}) of the zone of interest.

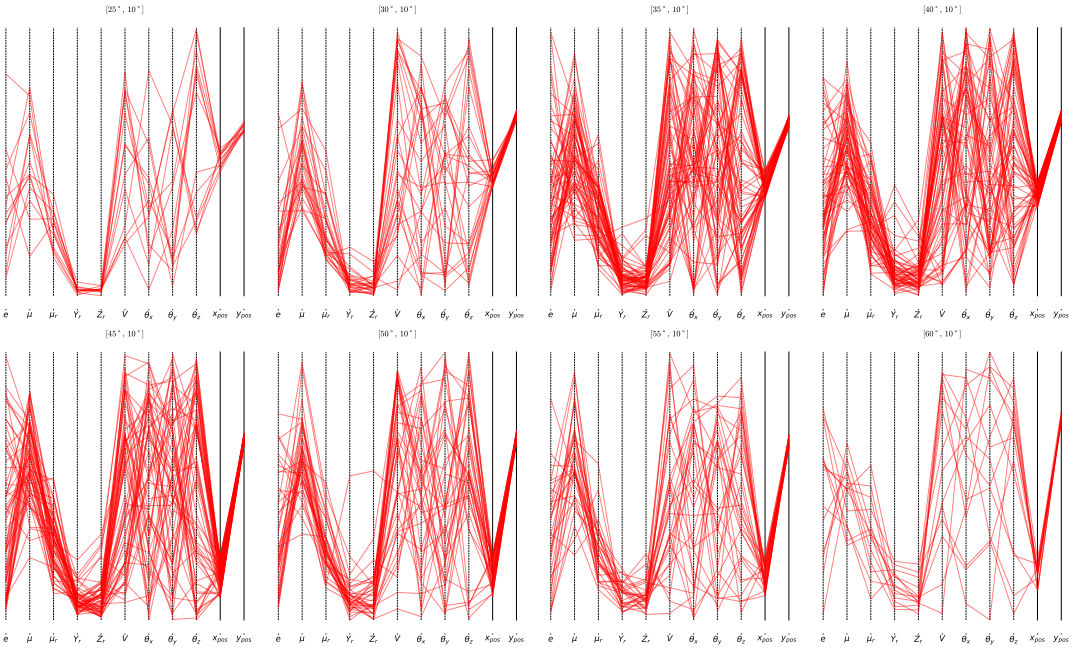


Figure 15: Cobweb plot for block stopped in a zone of interest, explored though the implementation of multiple screens for model output

The zone within x = [x_1+50, x_1+150] (see Figure 6 for x_1) and y = [150, 200] is selected for the illustration purpose. Here, only the blocks residing in the zone of interest are plotted for better visualisation. Here, μ and μ_r parameters provide a noticeable variability for different terrains while other parameter distribution remains random. Overall, the combined data visualisation with histogram and Cobweb can further provide insights to interpret the response collectively in discrete and continuous domains.

3.4. Real terrains models

In addition, three real world terrains models are also explored namely Dole-1, Dole-2 and Vaujany. The DEM of these terrains provide information on the profile which is modelled as a contact surface for their

modelling in SICONOS. The starting point of the rock block fall is identified for these terrains and a large number of model simulations are carried out to access model response.

Here, k-means clustering method is used to classify the blocks stopping at different clusters. The stopped blocks positions and the clustered blocks from 5500 simulations for first two terrains and last one with 11000 simulations from the same set of input parameters as used for the biplaner terrains are presented in Figures 16, 18 and 20 respectively for the aforementioned three terrains.

3.4.1. Dole P1 terrain

The Dole P1 terrain DEM is extracted from the 'Dole' commune in the eastern French department of Jura. This terrain DEM segment comprises of two cliffs and two relatively gentle slopes. The NSCD model computations on the 5500 sets of input parameters are carried out and the block stopping positions are presented in Figure 16.

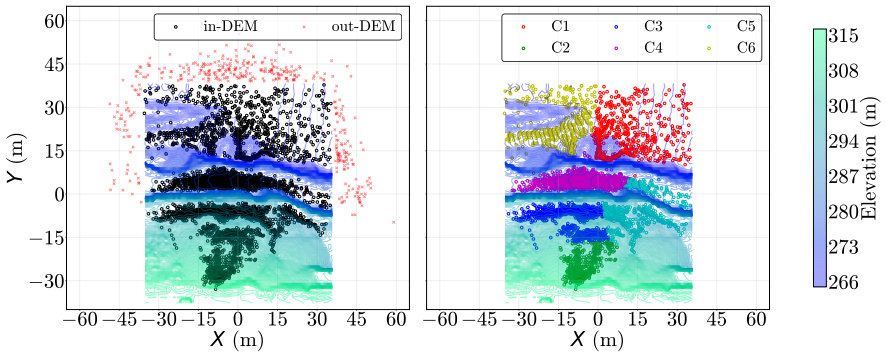


Figure 16: Block stopping distances in DEM un-clustered and clustered position of stopped block for Dole P1 terrain

Here, a sufficient number of blocks are observed to fall out of the DEM terrain from left, right and up front. These observations are removed from the scattered points cloud before carrying out the k-means clustering computations. Similar to what stated above for the biplaner terrain, the k-means clustering method is observed to be relatively misrepresentation due to it's dependence in the relative distribution of the cloud of points. For instance, the stopped blocks in-between two cliffs shall ideally be placed in the same cluster. However, they are divided in C4 and C5 as observed from Figure 16.

Further, the histogram of the stopping distance output and its correlation with the input parameters histogram is also explored, as presented in Figure 17. The rightward skew of the distribution of μ for C2 is clearly identified stating that the contact friction coefficient is a dominant parameter for a block to stop at the upper part of the terrain. Additional observations can also be inferred based on user's requirement.

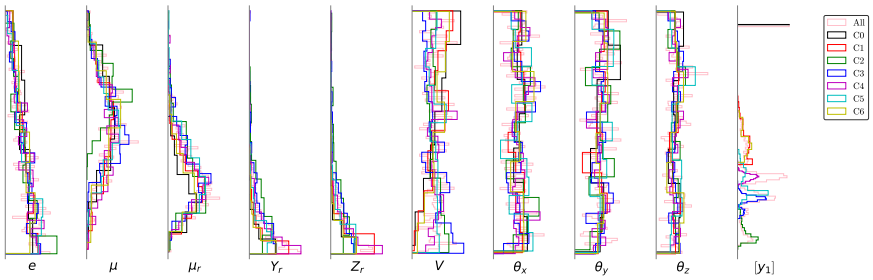


Figure 17: k-means clustered zones of the deposited blocks in DEM of Dole P1 site

Notably, the maunal clustering method is not explored for this terrain due to the absence of the knowledge of the susceptible areas and the interest of a practitioner. Nonetheless, upon availability of such knowledge, the aforementioned procedure for manual clustering, histograms and Cobweb plots can be conveniently carried out.

3.4.2. Dole P2 terrain

Another terrain segment from Dole region is exploited following the similar guidelines as opted for the Dole P1 terrain. The unclustered and clustered stopped block positions in the contour map of the terrain are presented in Figure 18.

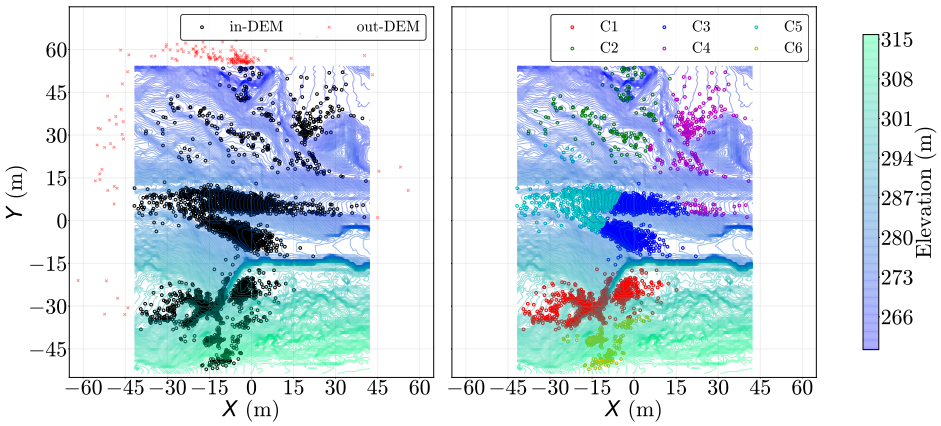


Figure 18: Block stopping distances in DEM un-clustered and clustered position of stopped block for Dole P1 terrain

Here, the relative number of blocks falling out of DEM are limited, however, the left and right side out of DEM dispersion is still present. The majority of rock blocks stop in the middle region of the terrain where two gentle slope surfaces are present. A limited number of blocks reach downhill in the relatively flat part of the terrain which are identified in C4. Likewise as in Dole P1, the k-means clustering algorithm does not provide satisfactory observation and hence manual clustering efforts are advised.

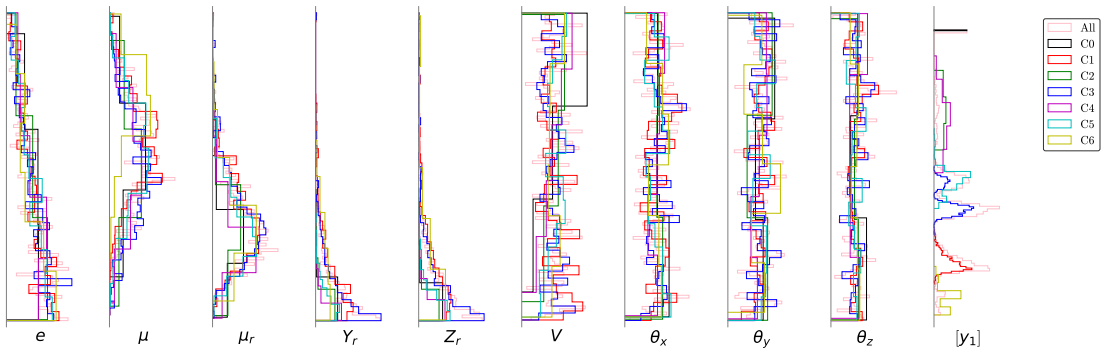


Figure 19: k-means clustered zones of the deposited blocks in DEM of Dole P2 site

Moreover, the correlation of histogram (presented in Figure 19) of input parameters with the output stopping distance reveals higher μ responsible for early stopping of the block. Despite many overlaps, the clusterwise distributions of output stopping distance is observed to be well segmented. This shall aid in the exploration from a practitioner point of view.

3.4.3. *Vaujany terrain*

The last explored terrain in from the Vaujany commune in the canton of Oisans-Romanche, in the Isère department in southeastern France. It's a slender terrain patch encompassing two forest roads in the middle of the terrain. There is a flat valley component downhill the terrain beyond which another uphill slope initiates close to the the elevation level of second forest road. The 11000 sets of input parameters are sampled using LHS method and launched from the point of fall observed on site. The stopped block positions in the terrain are presented in Figure 20.

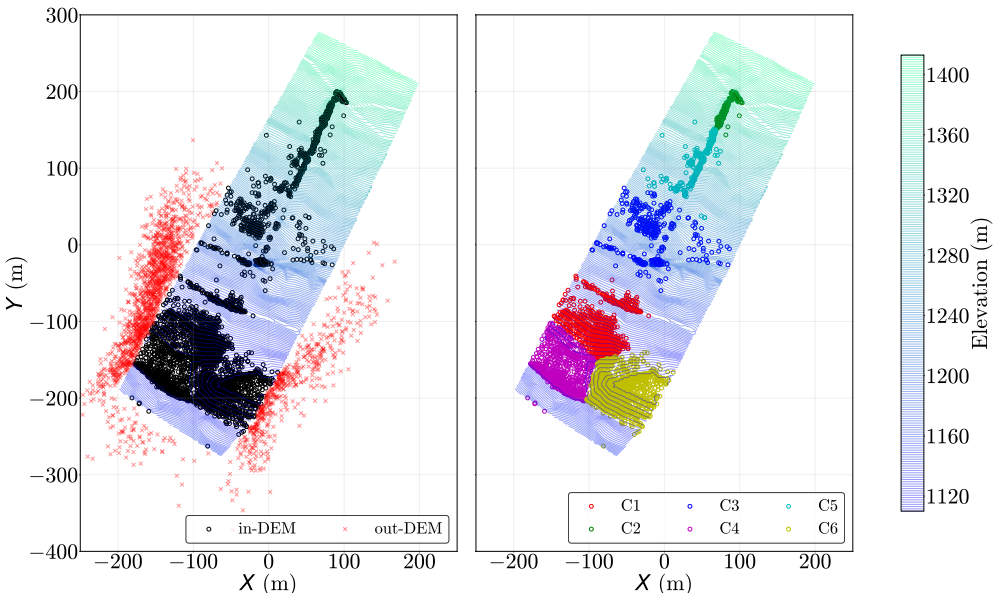


Figure 20: Block stopping distances in DEM un-clustered and clustered position of stopped block for Vaujany terrain

Here, the distinct divisions of the blocks position in different segments of the terrain are clearly visible. The first trench along the downhill just after the point of rockfall start stops a large number of blocks. Following this, the two forest roads stop the further movement of sufficient number of blocks which enables the practitioner to identify the pertinent risk to road safety and functioning. Further, nearly all blocks moving past the second forest road stop in the valley region. Notably, the analysis of block stop position (and also stop distance) in this terrain doesn't distinguish between a block which might have fallen back to valley after surpassing towards upward elevation post valley and the block which simply stopped at the valley. Nonetheless, the trajectory length of the falling rock can enable to simply distinguish between the two assuming that their overall trajectory was similar up to the point they approached the valley.

On a counter side, the slenderness of this terrain model resulted in nearly 18% of blocks falling out of terrain majorly from the left and right sides. This induces a significant loss in statistical variability of the terrain effect on incoming rockfall. This is nonetheless taken care of by selecting a largest sample size in comparison to the other terrain to compensate for the loss.

Further, k-means clustering presents the same conclusion as observed for the other modelled terrains. The spatial characteristics of this terrain model however present an advantage to carry out manual clustering practice.

The input/output correlation histogram, presented in Figure 21, indicates a large majority of blocks stopping very close to the rockfall start point. However, the absence of clear variability observation in the input parameters for this region (C2 from k-means clustering) indicates that it is prominently due to the

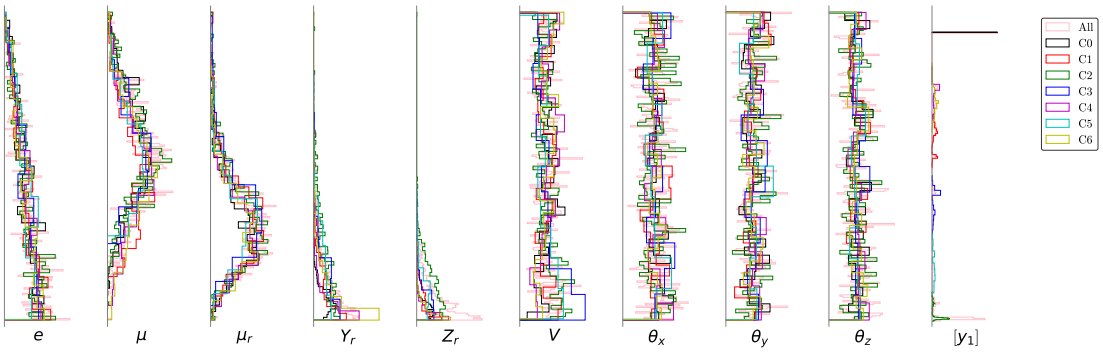


Figure 21: k-means clustered zones of the deposited blocks in DEM of Vaujany site

terrain characteristics rather than model parameters. The exploration of this zone using Cobweb plot (similar to Figure 15) shall aid in clear observation on the contribution of the input parameters.

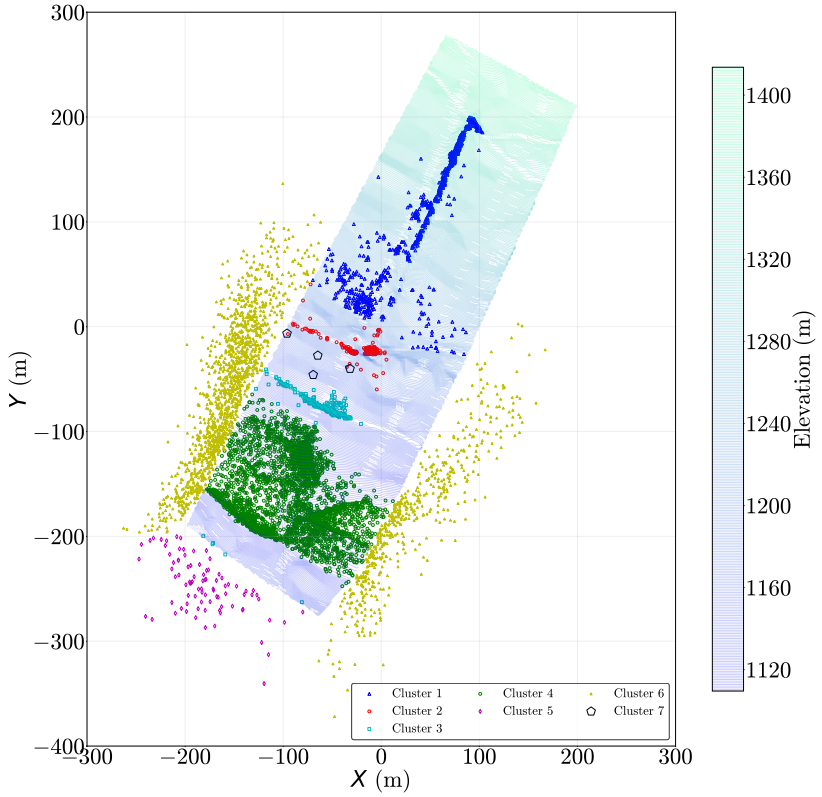


Figure 22: Manually clustered zones of the deposited blocks in DEM of Vaujany site

Lastly, the manual clustering of Vaujany terrain is presented in Figure 22 based on the elevation range of stopped blocks in different locations. Here, the blocks falling out of DEM from either left or right side of the terrain boundaries are considered a part of the same cluster. Moreover, the blocks falling from the upfront of the terrain (after reascending from the valley region) are considered in a separate cluster to identify the high kinetic energy blocks.

The input-output correlation is established through the histograms, presented in Figure 23. Here, some

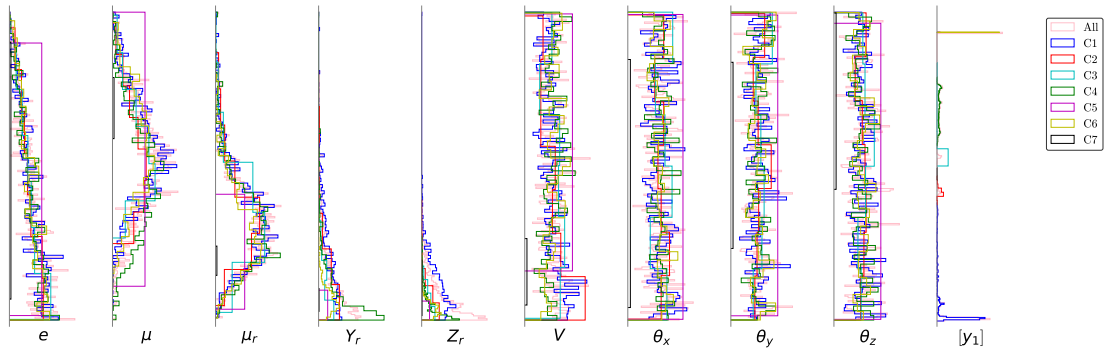


Figure 23: Manually clustered zones of the deposited blocks in DEM of Vaujany site

distinct observations are visible, for instance, the stopped blocks in the valley region (cluster 4) have a peak for low values of Y_r and Z_r emphasising that the compact shape of the block is dominant to long travel length. The blocks stopped in the upper forest road (cluster 2) have comparatively volume. The different peaks in the rolling friction distribution μ_r for blocks stopped in two forest roads enables to distinguish the pertinent block characteristics.

4. Conclusions

The descriptive statistics of rock block falling event is presented with insights to the complexity of the phenomenon. The observed response in the overall terrain domain is sub-divided into clusters to understand the local variance and establish a link between input and output parameters. This way, the uninformative overall variance is replaced by informative assessment of the rock block characteristics for it's stoppage at a particular zone.

PCE based surrogate models fail to accurately identify the relation between input and output parameters due to the chaotic nature of the model response. Therefore, alternative data processing techniques, namely histograms and Cobweb plots are used to understand the input/output relation. The potential use of the terrain model as an input parameter is also tested through a simple biplaner terrain to enhance the variability of the phenomenon. Besides, given the practitioners interest and expertise to identify the risk prone zones, a manual clustering practice specific to a terrain model is encouraged.

A detailed README_New file is available in the git project of this work for user's understanding of the python scripts for model design, post-processing and data analysis. The presented work opens door for further improvement as expressed through following perspective ideas:

- The descriptive statistics strategy is concluded to be appropriate to interpret the presented phenomenon. It also provides the basis for the design and implementation for the classification based machine learning algorithms. This shall enable generalisation of the model response for a given terrain (or beyond) to create metamodels for probabilistic statistics and implementation of advanced sensitivity analysis and uncertainty quantification algorithms.

- Kernel density smoothing for histogram updates can be implemented to overcome the limitation in the interpretation of the presented results albeit to the bin width.

- Given the observed correlated qualitative trends in the biplaner terrain models, further exploration becomes evident to introduce the terrain slope as an input variable. This can further enable for model

training, or hyperparamter tuning in the real terrain models (*e.g.*, by subdividing the terrain DEM as a number of biplaner terrain representation). Also, the functions like cross-entropy loss, a popular loss function used in machine learning to measure the performance of a classification model can be useful. It measures the difference between the predicted probability distribution and the true distribution.

- The Cobweb plot data visualisation can also be carried out for a specific range of input parameters to see the relative distribution of the output of interest. This can aid in the parameter tuning corresponding to a specific terrain profile. For instance, opting the current simulations scheme for a green terrain profile (say) would be interesting to explore for a lower bandwidth of the restitution coefficient.

- The secondary campaign shall be explored for the effect of the geometrical parameters only while keeping the mechanical parameters (*i.e.*, e, μ and μ_r) constant at their mean value. Notably, the other way around effect is also tested and reported having limited variability. Moreover, the experimental data available from the site investigation shall be used to do the needful on the model performance validation and calibration against different types of blocks.

- The input parameters are not correlated in the beginning of the sample definition (except for parameters Y_r and Z_r. However, the post clustering observations of the similar input parameter distribution can reveal insights to the correlation between them. It can also be observed through data visualisation methods such as 'seaborn' python package.

- The efforts on manual clustering in the real terrain models can be advanced by the usage of different search algorithms (for instance the ones used for image segmentation), which enable identification of non-linear zonal boundary in the terrain.

- The user might also consider updating the simulation stopping criteria to avoid some cases where the block despite being falling out of terrain, records a finite stopping distance computation. Even though such observations are one or two in a set of 9000 simulations, it would still be worthwhile to update it. The reason behind the presence of such rarity is the 'mod' of the recorded output frequency and number of simulation steps.

- Another quantity of interest could be velocity at given points or screens along the block downhill trajectory, for instance as reported in Bourrier et al. [10]. This shall enable to potentially access the variability in the response of incoming blocks irrespective of the observation of stopping nearly at the same distance.

- The potential construction of the metamodel can be made efficient by the selection of optimum sample size. This selection can be examined following the guidelines reported by [11], where authors report how to minimize the size of a test sample, while obtaining a good estimate of the metamodel prediction. Also, 'lazy predict' python package can be used to access the accuracy indices various metamodels applicable both for regression and classification problems.

- The clusterwise sensitivity analysis can be integrated into a global sensitivity analysis framework, following the guidelines reported in Roux et al. [12].

- The usage of Ishigami function (see Appendix D for reference) is recommended to rapidly test the applicability and the accuracy of the formulated metamodels and sensitivity analysis methods.

- In the present work, the mesh size is assigned a constant magnitude of 2.5m for biplaner and Vaujany terrain models and 1m for Dole P1 and P2 terrain models. The effect of terrain mesh size could therefore be interesting to explore given the fast computation time for each rock block fall simulation.

Appendix A. Test on Numerical parameters variability

The possible influence of the variability in the numerical model parameters which control the simulation run and the contact engagement responses is tested by selecting set of numerical input parameters, presented in Table A.3.

Parameters	Input factor	Distribution
Run options	time step h	U($1e^{-6}$,$1e^{-2}$)
	Newton tolerance	U($1e^{-9}$,$1e^{-6}$)
	Constraint activation threshold	U(0, 0.02)
Solver options	Tolerance	U($1e^{-6}$,$1e^{-2}$)
Bullet options	Contact breaking threshold	U(0.01, 0.2)

Table A.3: Input factors considered in the Siconos model of Vaujany terrain to represent the numerical parameters variability and corresponding distributions. Statistics of the assigned distributions are defined as follows: uniform: U (min, max) and categorical.

Here, a uniform distribution is assigned for all in the absence of any prejudice of user's knowledge. Figure A.24 shows the distribution of stopping distances calculated for Vaujany terrain using the statistical sample incorporating model, numerical and model+numerical variability. Notably, for the first two scenarios, the other mean values for the other component were assigned.

Simulations identified as leaving the DEM before the end of the simulation (altitude inferior to minimum of DEM) have been removed to improve interpretation. The distribution is bimodal, showing two main areas of deposit: part of the blocks stop near the intial position while others stop at the end of the DEM, near river bed. To be noted that maximum distance before leaving the DEM is 499.65 m, Figure A.24 then shows that no block leaves the DEM through the extreme part situated after the river.

Here, globally the numerical parameters variability is observed to present not significant influence on the overall distribution of the stopping distances. Therefore, the numerical parameters variability is removed from the analysis enabling a clear focus on the rock block geometrical and contact characteristics and providing aid for the uncertainty quantification.

Notably, these presented results were tested on the old set of input parameters. where block shape parameters were restricted to the classes presented in Table 1, and hence were considered a categorical variable. Besides, angularity was also user defined parameter which controlled the mesh design. After observing relatively less influence of numerical parameter variability in the overall process (see Figure A.24, further computations are carried out using the model input parameters only. Notably, the shape index categorical variable was replaced by two continuous variables Y_r and Z_r. Also, the angularity is assigned as a fixed to 100 points which enables to reduce the variability due to mesh design while considering sufficient number of nodes to design the rock block shape as triangulated element constituting a convex shape.

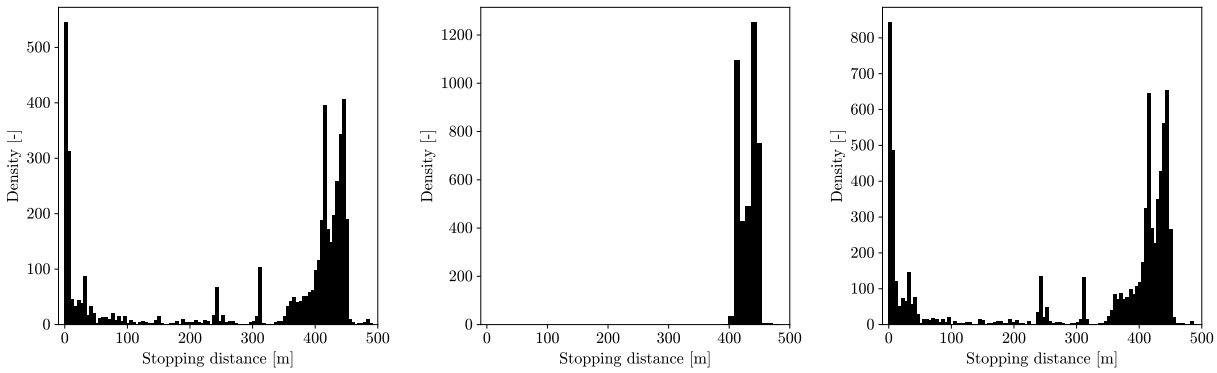


Figure A.24: Distribution of stopping distances for a 500-size Sobol sample for (a) model parameters only, (b) numerical parameters only, (c) model and numerical parameters together. Notably, sample truncated to exclude -9999 values.

Appendix B. Test on simultaneous rockfall for Siconos model stochastic behaviour

Given the highly computational efficiency of the rockfall simulation, it could be useful to simulate multiple blocks falling together having different input characteristics. In Siconos modelling, there is an inherent feature, where the definition of contact detection between two bodies is assigned through collision group. It means, in its absence, two physical bodies would be invisible to each other even if they are placed at the same point. In simple words, if the contact between block and terrain is not explicitly assigned, the falling block would pass through the terrain.

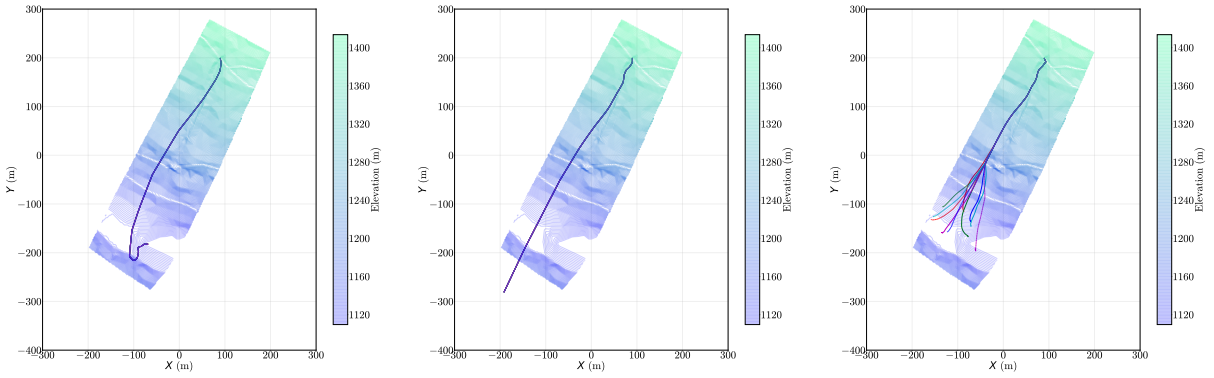
This inherent characteristic enables user to model multiple rock blocks falling together. Nonetheless, despite being a suitable idea to execute to optimise the model output database (for example), this idea is superseded due to the observation presented hereafter.

The simulation of the trajectory of a block includes randomness that is very difficult to control. Generally speaking, the trajectories of rigid mechanical systems with contact and impact stresses are not continuous with respect to the initial conditions, and naturally to some parameters. The consequence of this discontinuity is that a computation on a given block can be random for a fixed set of conditions and parameters.

This is tested by simulating 10 blocks falling together where all of them have identical input parameters. Such configuration is tested on the Vaujany terrain model and the results from 10 blocks falling together with three distinct sets of input parameters are presented for the demonstration in Figure B.25.

Here, 10 blocks are observed to follow the same trajectory when launched together, which is inline with the user expectation. However, such similarity is observed to disappear for the last set of input parameters where despite following the same computational schema, the trajectories deviate beyond a certain location as they travel together downhill. This inherent uncertainity in the Siconos model prevents to examine multiple blocks falling together.

It can certainly be argued that such deviation might not be present if the launched blocks possess different input parameter characteristics. However, this proposition is still superseded due to following two reasons: (1) since the overall process stops once the last block stops, it would be waste of computational data for the block which stopped in the beginning instances of the simulations, (2) since a unidentified connect exists in the detected contact processing, one will continue to remain uncertain on whether the observed trajectory is reproducible or not if the number of blocks falling together are changed.



(a) (b) (c)

Figure B.25: Trajectory of 10 blocks with identical input parameters falling together

Also, the machine level accuracy is also observed to play a role in the trajectory of the block as same script running in two different computers might not provide the same trajectory. The *bullet* package that is used to detect contact in the SICONOS model framework includes randomness which cannot controlled, thereby, it can be a potential source of the issue demonstrated in Figure B.25. Considering such intrinsic randomness, the model is no longer deterministic and should rather be considered as a stochastic model. A stochastic model is one which yields different outputs even when same input set of variables is used. Such a computer model is inherently stochastic because the simulator uses random numbers. The number of simulations required to incorporate this would multi fold equal to number of blocks considered in each fall. Notably, a python script to visualise the variability of each simulation set with multiple blocks falling together as a box plot is available in this project files.

Keeping these issues in mind, each rock block fall simulation is carried out individually and only one machine is used for all computations to reduce inherent variability to a possible minimum (as SICONOS software update might also influence the trajectory under similar line of explanation). The usage of any user defined randomness (*e.g.*, via 'random' python package) is completely avoided in all (deterministic) computations presented in this work.

Nonetheless, if further examination on the stochastic interpretation is desired for sensitivity analysis, the user may consider following directions.

The analysis of variance (ANOVA) decomposition of the output variance does not exist for stochastic model but several approaches can be considered to circumvent this difficulty. The reader may refer to Kouye [13] or Zhu and Sudret [14] for a review on Sobol' indices for stochastic models.

The first option consists in defining a deterministic QoI that summarizes the distribution such as the mean, the variance, any moment of higher order or probabilities to stop in given areas, or even the probability of leaving the DEM. However, it can result in a loss of information [14].

Another option consists in including the internal source of randomness in the set of input factors. A Sobol' index associated to the code randomness can thus be calculated. In our case, we could then calculated Sobol' index associated bullet randomness.

Finally, a third option consists in considering the stochastic model as a random field. For each value of internal randomness, the stochastic model becomes a deterministic function and Sobol' indexes are well defined. However, they remain random variables as they depend on the internal randomness. For this, a Gaussian process model, proposed by Marrel et al. [15] for the stochastic computer codes can be used.

This model is applicable for the scenarios where the result of each code run is itself random. They use the framework of the joint modeling of the mean and dispersion of hetroscedastic data. This study can help us to characterise the variability that originates when a lot of blocks having each parameter same, fall down the hill together.

In this study, option 2 and 3 are difficult to apply as bullet randomness is not controllable. First option that consists in scrutinizing specific statistic moments is therefore preferable.

Appendix C. Global sensitivity analysis approaches

Appendix C.1. Sobol' indexes

For any scalar output Y and d input factors $[\mathrm{X}_1, ..., \mathrm{X}_d] = \mathbf{X}$ such as $\mathrm{Y} = f(\mathbf{X})$, the total variance Var[Y] can be split into partial variances through the functional ANOVA decomposition [16]. Assuming independent input factors, each partial variance can thus be associated to an input factor or to the interactions between inputs. The Sobol' indexes [17] result from the ANOVA decomposition and are calculated as the ratio of such partial variances to the total variance.

Among Sobol' indexes, Sobol' first order index associated to the input factor X_i is defined as :

$$S_i = \frac{\mathrm{Var}(\mathbb{E}[\mathrm{Y}|\mathrm{X_i}])}{\mathrm{Var}(\mathrm{Y})}. \tag{C.1}$$

It measures the portion of output variance that can be attributed to the variation of input factor X_i only without any interaction with other input factors (main effect). They can be interpreted as the decrease in the total output variance that could be obtained when removing the uncertainty about X_i when setting X_i to a fixed value [18].

Second order index associated to input factors X_i and X_j is defined as :

$$S_{i,j} = \frac{\mathrm{Var}(\mathbb{E}[\mathrm{Y}|(\mathrm{X_i}, \mathrm{X_j})])}{\mathrm{Var}(\mathrm{Y})} - S_i - S_j. \tag{C.2}$$

$S_{i,j}$ quantifies the impact of the interactions between X_i and X_j on the output variance with no consideration of main effect of each input. Indexes of any higher order can be calculated this way.

Finally, one defines Sobol' total order index associated to input factor X_i as :

$$S_i^T = 1 - \frac{\mathrm{Var}(\mathbb{E}(\mathrm{Y}|\mathrm{X}_1, ..., \mathrm{X}_{i-1}, \mathrm{X}_{i+1}, ..., \mathrm{X}_d))}{\mathrm{Var}(\mathrm{Y})}. \tag{C.3}$$

S_i^T quantifies the total impact of the input factor X_i through main effects and interactions. It stands for the portion of total output variance that remains as long as X_i stays unknown [18].

Appendix C.2. HSIC dependence measures

The HSIC dependence measure [19] analyzes the dependence between the random variables X_i and Y by computing the distance between the joint distribution $P_{X_i,Y}$ and the product of the marginal distributions $P_{X_i}P_Y$. The higher this distance is, the more dependent Y and X_i are (meaning a high level of influence of X_i).

Considering $\mathcal{X}_i, \forall i \in \{1, ..., d\}$ and $\mathcal{Y}$ some measurable space, let $\mathcal{F}_i : \mathcal{X}_i \to \mathbb{R}$ and $\mathcal{G} : \mathcal{Y} \to \mathbb{R}$ be two Reproducing Kernel Hilbert Spaces (RKHS) defined by their characteristic kernel functions, respectively $k_{x_i}(.,.)$ and $k_y(.,.)$ and equipped with their product scalar $< .,. >_{\mathcal{F}_i}$ and $< .,. >_{\mathcal{G}}$.
Let also consider $\mathcal{H}_i$ a RKHS over $\mathcal{X}_i \times \mathcal{Y}$ equipped with its characteristic kernel $h_i(.,.)$. Then, we can define the mean embedding of $P_Y P_{X_i}$ and P_{Y,X_i} in $\mathcal{H}_i$:

$$\mu[P_Y P_{X_i}] = \mathbb{E}_Y \mathbb{E}_{X_i}[h_i((Y, X_i), .)],$$
$$\mu[P_{Y,X_i}] = \mathbb{E}_{Y,X_i}[h_i((Y, X_i), .)].$$

The HSIC dependence measure then corresponds to the square distance between the mean embeddings of $P_Y P_{X_i}$ and P_{Y,X_i}:

$$HSIC(X_i, Y)_{\mathcal{F}_i,\mathcal{G}} = ||\mu[P_{Y,X_i}] - \mu[P_Y P_{X_i}]||^2. \tag{C.4}$$

Considering (X_i', Y') as an independent and identically distributed copy of (X_i, Y) and assuming that $h_i((Y, X_i), (Y', X_i')) = k_{x_i}(X_i, X_i')k_y(Y_i, Y_i')$, Gretton et al. [19] shows that:

$$HSIC(X_i, Y)_{\mathcal{F}_i,\mathcal{G}} = \mathbb{E}[k_{x_i}(X_i, X_i')k_y(Y_i, Y_i')] + \mathbb{E}[k_{x_i}(X_i, X_i')]\mathbb{E}[k_y(Y_i, Y_i')] - 2\mathbb{E}[\mathbb{E}[k_{x_i}(X_i, X_i')|X_i]\mathbb{E}[k_y(Y_i, Y_i')|Y_i]]. \tag{C.5}$$

In practice, HSIC direct computation is not possible as it depends of the unknown distribution of (X_i, Y). Then, an estimator must be used. In this study, we use the following estimator which is unbiased for small sample sizes [20]:

$$\widehat{HSIC(X_i, Y)} = \frac{1}{n(n-3)}[\mathrm{Tr}(\tilde{\mathbf{L}}_i\tilde{\mathbf{L}}) + \frac{\mathbf{1}^\top\tilde{\mathbf{L}}_i\mathbf{1}\mathbf{1}^\top\tilde{\mathbf{L}}\mathbf{1}}{(n-1)(n-2)} - \frac{2}{n-2}\mathbf{1}^\top\tilde{\mathbf{L}}_i\tilde{\mathbf{L}}\mathbf{1}], \tag{C.6}$$

where $\tilde{L}_{i_{j,k}}$ and $\tilde{L}_{j,k}$ are computed as:

$$\tilde{L}_{i_{j,k}} = (1 - \delta_{j,k})k_{x_i}(X_i^{(j)}, X_i^{(k)}),$$
$$\tilde{L}_{j,k} = (1 - \delta_{j,k})k_y(Y^{(j)}, Y^{(k)}).$$

When using a universal kernel, the HSIC indices can be statistically used for screening purpose [21]. A statistical test can be set with the null hypothesis "X_i *and* Y *are independent*" . Considering an experimental design of N points $(x_i^1, ...x_i^N)$ and the associated output points $(y^1, ..., y^N)$, an estimator $\widehat{HSIC(\mathrm{X}_i, \mathrm{Y})}$ of the dependence measure $HSIC(\mathrm{X}_i, \mathrm{Y})$ is firstly computed. Then B bootstrap versions $\underline{\mathbf{Y}}^{[1]}$,...,$\underline{\mathbf{Y}}^{[B]}$ are resampled from the original output sample $(y_1, ..., y_N)$ with replacement so as to contain the same number of points N. For each $\underline{\mathbf{Y}}^{[B]}$ the input points associated to X_i are not resampled. Indeed, under the independence hypothesis, any values of Y can be associated to X_i. For each bootstrap version b, an estimator $\widehat{HSIC}^{[b]}(\mathrm{X}_i, \mathrm{Y})$ is computed. Then, the associated bootstrapped p-value is given by :

$$p\text{-val}_B = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\widehat{HSIC}^{[b]}(\mathrm{X}_i,\mathrm{Y}) > \widehat{HSIC}(\mathrm{X}_i,\mathrm{Y})} \tag{C.7}$$

Finally, denoting α the significance level, if $p\text{-val}_B < \alpha$, the independence hypothesis is rejected, otherwise it is accepted.

In addition to screening, HSIC measures can also be used for ranking. Da Veiga [22] proposes the following sensitivity index that allows to compare levels of influence of different input factors X_i:

$$\widehat{R^2_{HSIC_i}} = \frac{\widehat{HSIC(X_i, Y)}}{\sqrt{\widehat{HSIC(X_i, X_i)}\widehat{HSIC(X_i, Y)}}} \tag{C.8}$$

Appendix D. Sensitivity analysis test on the Ishigami function

The Ishigami function is a standard practice method to test the applicability of a particular SA and uncertainly processes. It is an artificial analytical model with three input variables.

$$Y = \sin(X_1) + a\sin(X_2)^2 + b(X_3)^4 \sin(X_1) \tag{D.1}$$

Here, 'a' and 'b' are coefficient with typically used values 7 and 0.1 respectively. X_1, X_2 and X_3 are the input variables and Y is the output of interest. Here, the interaction between input variables is assigned through the last term. Besides, both linear and non-linear relationships are used between inputs and output.

Configuration: The three input variables are assigned uniform distribution as presented in Table D.4.

	Input factor	Distribution
Parameters	X_1	U(-π, π)
	X_2	U(-π, π)
	X_3	U(-π, π)

Table D.4: Input factors considered in the global sensitivity analysis and corresponding distributions. Statistics of the assigned distribution is defined as follows: uniform: U (min, max).

Design of experiment: The Sobol sampling method is used to generate a sample of size 1000 which accounts for 1000*(3+2) = 5000 computations. Figure D.26 presents the experimental design. Other sampling methods can also be used as per the study requirement.

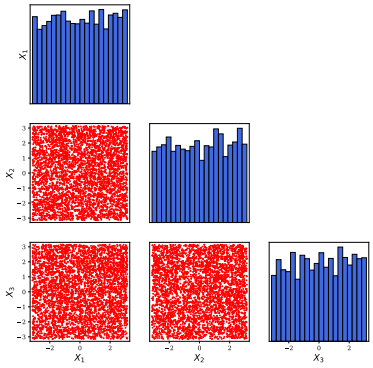


Figure D.26: Experimental design for the Ishigami function

Results: First the output and input variables relation is presented in Figure D.27 from the data-set of 5000 simulations. The unique shapes of input/output parameter scatter plot present the functional significance of the Ishigami function.

Second the output expectancy is presented through the histogram presented in Figure D.28. It presents the overall distribution of the model output to enable user to access the need of using component wise post-processing or global post-processing to access the variance of the model response.

Sensitivity analysis: The Sobol indices are estimated for the Ishigami function, presented in Figure D.29. Here, first order indices present the main effect of each input parameter while the total order indices present the overall effect inclusive of the interaction between input variables.

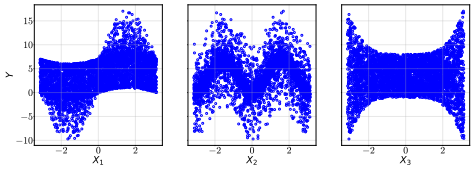


Figure D.27: Output-input relation for Ishigami function

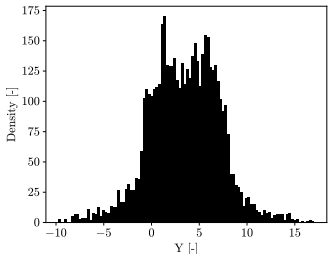


Figure D.28: Output histogram for the Ishigami function. Here, 100 bins are used.

Overall, in the present work, the Sobol index computation can be carried out on clusterwise distributions. Nonetheless, the attempts to establish a function between input and output parameters, for example, through PCE based meta-models have not been successful. Therefore, the Sobol index computation using the sample data only does not allow user to have confidence in the observed indices.

This can be further explored using highly sophisticated machine learning tools to establish the metamodel of the response. From current user experience, the clusterwise metamodel construction is likely to be more accurate than overall response domain. On the other hand, the classification based approach can also be used, *i.e.*, constructing a metamodel which captures the cluster in which a new rock block will fall. This can help user to shift the focus from predicting the individual block performance to a more general performance assessment of the range of input parameters.

Appendix E. First SA results from Emilie's report - text amended

For preliminary analysis and HSIC indexes calculation, a 1000-point Latin Hypercube Sample (LHS) is first generated. In a second time, a Sobol' sample [17] is also generated. Recommended size for Sobol' sample is $N \times (d+2)$ where d is the number of input parameters and N an integer commonly chosen between 500 and 1000 [23]. We choose $N = 1000$.

Appendix E.1. Screening based on HSIC indices

Still based on the 1000-point LHS, a screening step is first performed with an independence test based on HSIC measure. Some permutations are used for the statisitical test and the acceptance level α is set to 0.05. Screening results are presented in Figure E.30. Reminding that the higher the p-value is, the higher are the chances that the two variables are actually independent, results show that the coefficient of restitution e so as the angularity and the initial rotations along the y-axis and the z-axis are unlikely to have any influence on the stopping distance. The model can thus be simplified by fixing the value for these factors before performing ranking.

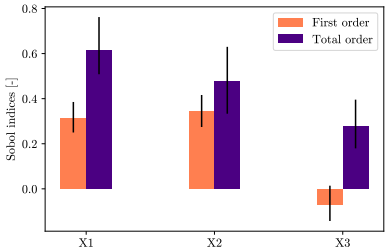


Figure D.29: First and total order Sobol indices having values [0.310667, 0.341578, -0.0709682] and [0.616315, 0.476504, 0.277471] respectively for the Ishigami function.

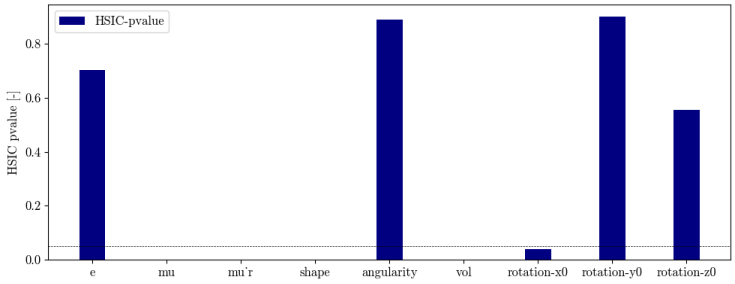


Figure E.30: p-values from HSIC independence test (HSIC_indices.py).

Appendix E.2. Ranking

According to the screening results, ranking is performed considering the following input factors: μ, μ_r, shape category, block volume and initial rotation along the x-axis.

- Sobol indices calculated from a Sobol sample (sample available at https://mybox.inria.fr/library/ecf449e5-a52e-4ad2-9efc-76c42b665c69/Rockfall/output_sobol_after_screening_new)
- HSIC indices calculated from a 500-point LHS sample (sample available at https://mybox.inria.fr/library/ecf449e5-a52e-4ad2-9efc-76c42b665c69/Rockfall/output_LHS500_after_screening)

Notably, these files are not accessible in this link and must be asked from the author.

Sobol' indices are presented in Figure E.31 while HSIC indices are presented in Figure E.32. Rankings from both methods show a very limited influence of the initial rotation along the x-axis (which was already shown by the independence test, see Figure E.30). However, the rankings for the remaining parameters are roughly opposite. While the shape category is captured as poorly influential and the friction parameters are captured as preponderant according to Sobol' indices, HSIC indices come to the opposite conclusion. Several hypothesis can explained such finding:

- Sobol and HSIC indices intrinsically define the notion of sensitivity in different ways, meaning that the sensitivity indices cannot be really compared.
- The shape variable is a catogerical variable and may be treated differently.

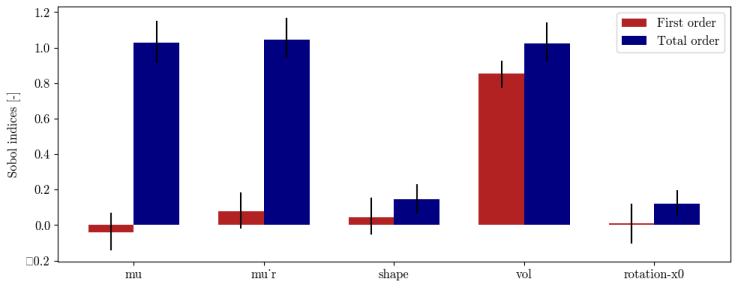


Figure E.31: Sobol indices for the set of screened parameters (Sobol_indices.py).

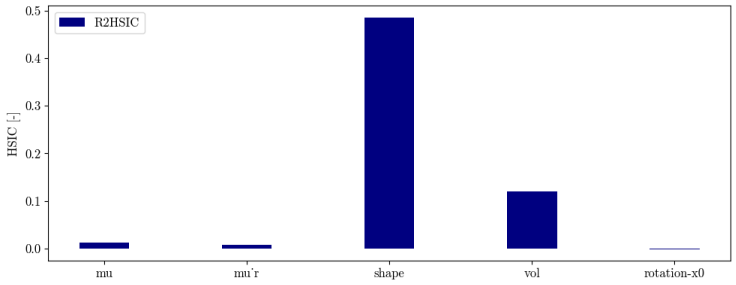


Figure E.32: R2HSIC values for set of screened parameters (HSIC_indices.py).

References

[1] S. Dobre, T. Bastogne, A. Richard, Global sensitivity and identifiability implications in systems biology, in: 11th IFAC Symposium on Computer Applications in Biotechnology, CAB 2010, Leuven, Belgium, 2010, p. CDROM. URL: https://hal.science/hal-00501632.

[2] M. Jean, J. J. Moreau, Dynamics in the presence of unilateral contacts and dry friction: a numerical approach, in: G. Del Pietro, F. Maceri (Eds.), Unilateral problems in structural analysis. II, CISM 304, Springer Verlag, 1987, pp. 151–196.

[3] M. Jean, The non smooth contact dynamics method, Computer Methods in Applied Mechanics and Engineering 177 (1999) 235–257. Special issue on computational modeling of contact and friction, J.A.C. Martins and A. Klarbring, editors.

[4] V. Acary, F. Perignon, Siconos: A Software Platform for Modeling, Simulation, Analysis and Control of Nonsmooth Dynamical Systems, SIMULATION NEWS EUROPE, ArgeSIM/ASIM 17 (2007) 19–26.

[5] V. Acary, Energy conservation and dissipation properties of time-integration methods for nonsmooth elastodynamics with contact, Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik 96 (2016) 585–603.

[6] V. Acary, F. Bourrier, Coulomb friction with rolling resistance as a cone complementarity problem, European Journal of Mechanics - A/Solids 85 (2021) 104046.

[7] E. D. Sneed, R. L. Folk, Pebbles in the lower colorado river, texas a study in particle morphogenesis, The Journal of Geology 66 (1958) 114–150.

[8] D. A. Bonneau, D. J. Hutchinson, P.-M. DiFrancesco, M. Coombs, Z. Sala, Three-dimensional rockfall shape back analysis: methods and implications, Natural Hazards and Earth System Sciences 19 (2019) 2745–2765.

[9] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, Uncertainty management in simulation-optimization of complex systems: algorithms and applications (2015) 101–122.

[10] F. Bourrier, L. Dorren, F. Nicot, F. Berger, F. Darve, Toward objective rockfall trajectory simulation using a stochastic impact model, Geomorphology 110 (2009) 68–79.

[11] B. Iooss, L. Boussouf, V. Feuillard, A. Marrel, Numerical studies of the metamodel fitting and validation processes, 2010. URL: https://arxiv.org/abs/1001.1049. arXiv:1001.1049.

[12] S. Roux, S. Buis, F. Lafolie, M. Lamboni, Cluster-based gsa: Global sensitivity analysis of models with temporal or spatial outputs using clustering, Environmental Modelling & Software 140 (2021) 105046.

[13] H. M. Kouye, Sensitivity analysis approaches for stochastic models. Application to compartmental models in epidemiology, Ph.D. thesis, Université Paris-Saclay, 2022.

[14] X. Zhu, B. Sudret, Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models, Reliability Engineering & System Safety 214 (2021) 107815.

[15] A. Marrel, B. Iooss, S. Da Veiga, M. Ribatet, Global sensitivity analysis of stochastic computer models with joint metamodels, Statistics and Computing 22 (2012) 833–847.

[16] B. Efron, C. Stein, The Jackknife Estimate of Variance, The Annals of Statistics 9 (1981) 586 – 596.

[17] I. M. Sobol, Sensitivity analysis for non-linear mathematical models, Math. Modeling Comput. Exp. 1 (1993) 407–414.

[18] S. Tarantola, N. Giglioli, J. Jesinghaus, A. Saltelli, Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making?, Stochastic Environmental Research and Risk Assessment 16 (2002) 63–76.

[19] A. Gretton, O. Bousquet, A. Smola, B. Schölkopf, Measuring statistical dependence with hilbert-schmidt norms, in: International conference on algorithmic learning theory, Springer, 2005, pp. 63–77.

[20] M. Baudin, A. Dutfoy, B. Iooss, A.-L. Popelin, Open turns: An industrial software for uncertainty quantification in simulation, arXiv preprint arXiv:1501.05242 (2015).

[21] M. De Lozzo, A. Marrel, New improvements in the use of dependence measures for sensitivity analysis and screening, Journal of Statistical Computation and Simulation 86 (2014).

[22] S. Da Veiga, Global sensitivity analysis with dependence measures, Journal of Statistical Computation and Simulation 85 (2015) 1283–1305.

[23] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, John Wiley & Sons, 2008.