Contributions to the calibration and global sensitivity analysis of snow avalanche numerical models

Maria Belen Heredia Guzman

▶ **To cite this version:**

Maria Belen Heredia Guzman. Contributions to the calibration and global sensitivity analysis of snow avalanche numerical models. Earth Sciences. Université Grenoble Alpes [2020-..], 2020. English. NNT : 2020GRALU028 . tel-03163059

HAL Id: tel-03163059

https://theses.hal.science/tel-03163059v1

Submitted on 9 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Science de la Terre et de l'Univers et de l'Environment

Arrêté ministériel : 25 mai 2016

Présentée par

María Belén HEREDIA GUZMÁN

Thèse dirigée par **Nicolas ECKERT, chercheur**
et codirigée par **Clémentine PRIEUR, Professeur, UGA**

préparée au sein **du Laboratoire INRAE ETNA**
dans **l'École Doctorale Terre, Univers, Environment**

Contributions to the calibration and global sensitivity analysis of snow avalanche numerical models

Thèse soutenue publiquement le **10 décembre 2020**,
devant le jury composé de :

Véronique Maume-Deschamps
Professeur Université Lyon 1, Présidente et rapportrice
Bruno Sudret
Professeur ETH Zurich, Rapporteur
Jan-Thomas Fischer
Directeur de l'Unité de recherche Neige et Avalanche, BFW Innsbruck, Autriche, Examinateur
Sébastien Da Veiga
Expert Senior Statistique et Optimisation, Safran Tech, Examinateur
Eric Parent
Professeur AgroParisTech, Examinateur
Guillaume Evin
CR INRAE, Examinateur
Clémentine Prieur
Professeur Université Grenoble Alpes, Co-Directrice de thèse
Nicolas Eckert
ICPEF INRAE, Co-Directeur de thèse

I rarely end up where I was intending to go, but often I end up somewhere that I needed to be.

The Long Dark Tea-Time of the Soul, Douglas Adams

Abstract

Snow avalanche is a natural hazard defined as a snow mass in fast motion. Since the thirties, scientists have been designing snow avalanche models to describe snow avalanches. However, these models depend on some poorly known input parameters that cannot be measured. To understand better model input parameters and model outputs, the aims of this thesis are (i) to propose a framework to calibrate input parameters and (ii) to develop methods to rank input parameters according to their importance in the model taking into account the functional nature of outputs. Within these two purposes, we develop statistical methods based on Bayesian inference and global sensitivity analyses. All the developments are illustrated on test cases and real snow avalanche data.

First, we propose a Bayesian inference method to retrieve input parameter distribution from avalanche velocity time series having been collected on experimental test sites. Our results show that it is important to include the error structure (in our case the autocorrelation) in the statistical modeling in order to avoid bias for the estimation of friction parameters.

Second, to identify important input parameters, we develop two methods based on variance based measures. For the first method, we suppose that we have a given data sample and we want to estimate sensitivity measures with this sample. Within this purpose, we develop a nonparametric estimation procedure based on the Nadaraya-Watson kernel smoother to estimate aggregated Sobol' indices. For the second method, we consider the setting where the sample is obtained from acceptance/rejection rules corresponding to physical constraints. The set of input parameters become dependent due to the acceptance-rejection sampling, thus we propose to estimate aggregated Shapley effects (extension of Shapley effects to multivariate or functional outputs). We also propose an algorithm to construct bootstrap confidence intervals. For the snow avalanche model application, we consider different uncertainty scenarios to model the input parameters. Under our scenarios, the release avalanche position and volume are the most crucial inputs.

Our contributions should help avalanche scientists to (i) account for the error structure in model calibration and (ii) rank input parameters according to their importance in the models using statistical methods.

Keywords: Snow avalanche numerical models, Bayesian inference, aggregated Sobol' indices, aggregated Shapley effects, given data estimation method, dimension reduction

Résumé

Une avalanche de neige est un danger naturel défini comme une masse de neige en mouvement rapide. Depuis les années 30, scientifiques conçoivent des modèles d'avalanche de neige pour décrire ce phénomène. Cependant, ces modèles dépendent de certains paramètres d'entrée mal connus qui ne peuvent pas être mesurés. Pour mieux comprendre les paramètres d'entrée du modèle et les sorties du modèle, les objectifs de cette thèse sont (i) de proposer un cadre pour calibrer les paramètres d'entrée et (ii) de développer des méthodes pour classer les paramètres d'entrée en fonction de leur importance dans le modèle en tenant compte la nature fonctionnelle des sorties. Dans ce cadre, nous développons des méthodes statistiques basées sur l'inférence bayésienne et les analyses de sensibilité globale. Nos développements sont illustrés sur des cas de test et des données réelles des avalanches de neige.

D'abord, nous proposons une méthode d'inférence bayésienne pour récupérer la distribution des paramètres d'entrée à partir de séries chronologiques de vitesse d'avalanche ayant été collectées sur des sites de test expérimentaux. Nos résultats montrent qu'il est important d'inclure la structure d'erreur (dans notre cas l'autocorrélation) dans la modélisation statistique afin d'éviter les biais dans l'estimation des paramètres de frottement.

Deuxièmement, pour identifier les paramètres d'entrée importants, nous développons deux méthodes basées sur des mesures de sensibilité basées sur la variance. Pour la première méthode, nous supposons que nous avons un échantillon de données et nous voulons estimer les mesures de sensibilité avec cet échantillon. Dans ce but, nous développons une procédure d'estimation non paramétrique basée sur l'estimateur de Nadaraya-Watson pour estimer les indices agrégés de Sobol. Pour la deuxième méthode, nous considérons le cadre où l'échantillon est obtenu à partir de règles d'acceptation/rejet correspondant à des contraintes physiques. L'ensemble des paramètres d'entrée devient dépendant du fait de l'échantillonnage d'acceptation-rejet, nous proposons donc d'estimer les effets de Shapley agrégés (extension des effets de Shapley à des sorties multivariées ou fonctionnelles). Nous proposons également un algorithme pour construire des intervalles de confiance bootstrap. Pour l'application du modèle d'avalanche de neige, nous considérons différents scénarios d'incertitude pour modéliser les paramètres d'entrée. Dans nos scénarios, la position et le volume de départ de l'avalanche sont les entrées les plus importantes.

Nos contributions peuvent aider les spécialistes des avalanches à (i) prendre en compte

la structure d'erreur dans la calibration du modèle et (ii) proposer un classement des paramètres d'entrée en fonction de leur importance dans les modèles en utilisant des approches statistiques.

Mot-clés : Modèles numériques d'avalanche de neige, inférence bayésienne, indices de Sobol' agrégés, effets de Shapley agrégés, réduction de la dimension

Remerciements

La liste de gens que je voudrais remercier est longue. Chaque personne qui a fait partie de mon chemin de thèse a contribué à l'obtention de cet accomplissement. Je voudrais d'abord remercier Clémentine Prieur et Nicolas Eckert, mes encadrants de thèse, pour m'avoir guidé et accompagné durant ces trois années.

Je remercie les membres du jury, de leur curiosité scientifique et leur bienveillance. La soutenance de thèse a été un moment inoubliable et elle restera pour toujours comment un doux souvenir dans ma mémoire. Je remercie Véronique Maume-Deschamps et Bruno Sudret pour leurs rapports qui m'ont permis de prendre du recul et du plaisir sur mes travaux de thèse. Je remercie Jan Thomas Fischer, Sébastian Da Veiga, Eric Parent et Guillaume Evin pour leurs questions fortes intéressantes sur la thèse.

Je remercie les membres de mon comité de suivi de thèse, Eric Parent (de nouveau), Thierry Faug et Pascal Hagenmuller pour leurs conseils et le temps consacré au suivi de mon sujet de thèse. Leurs conseils ont été très significatifs pour améliorer le développement de ma thèse.

J'ai eu l'occasion unique de faire une thèse dans deux laboratoires et ça m'a permis de croiser des personnes très chouettes. Je remercie mes amis et collègues du labo INRAE Eduardo, Philomène, Ségolène, Manon et Guillaume, Cécile, Erwan et autres. Aussi, je remercie mes collègues du LJK, Anya, Victor, JB, Mussab, ... Merci pour les pauses-café au labo et les discussions sur des sujets si divers !

Je remercie aussi Clémentine Junquas, Thomas Condom et Clémentine Prieur (de nouveau), mes encadrants de master au LTHE, pour m'avoir appris les bonnes base sur la recherche scientifique. Je remercie aussi mes amis doctorants du labo LTHE Pati, Gabi, Magda, Claudio, Catherine. On est un si joli groupe multiculturel !

Je remercie toutes ces merveilleuses personnes que j'ai rencontré à Grenoble: Pablo, P0nce, Maêl, Louise, Raph, Nino, Cynthia, Nico, Orianne, Baptiste, Mathilde et autres. Et aussi, mes amis d'Equateur, Eve et David, même si la distance nous sépare, vous êtes toujours dans mon cœur.

Je remercie Sam Porter, Aloy, Nathan Drake et Geralt de Riv, des personnages des jeux vidèo, qui m'ont accompagné pendant la rédaction de mon manuscrite et qui m'ont permi voyager et rêver en restant confinée.

Finalement, je voudrais remercier ma famille et ma belle-famille. Je remercie ma très patiente mère Consuelo qui n'a pas imaginé (moi non plus) que ce voyage aurait été un peu plus long que ce qui avait été prévu. La vie est très incertaine et il y a des choses

qu'on ne peut pas prévoir comme se marier avec son collègue français du cours d'anglais. Je remercie mes frères Chalo, Pablo et Sebas et ma belle famille, Nicole, Bernard, Marion, Yann et Monique pour leur soutien, conseils et encouragement pendant la thèse. Je te remercie Thomas, mon chéri, une des plus belles choses que ce voyage m'a apporté. Cette thèse est à nous deux.

Contents

1 Introduction

In this chapter, we give a general overview of the thesis, of the scientific questions to be addressed in this work and of the way we handled them. The concepts and tools employed are detailed in Chapter 2.

1.1 Context

Snow covered mountains lie in apparent equilibrium till an avalanche suddenly occurs. Snow avalanche is a natural hazard defined as a snow mass in fast motion on a slope [Ancey, 2001]. In fact, avalanche velocities can reach 60 ms^{-1} [Schaer and Issler, 2001]. There exist many factors that interact to explain the release of an avalanche for example, the average inclination of starting zones (between 27° to 50°), the ground surface roughness, snowpack structure or temperature changes [Ancey, 2001]. But more important than its causes are its consequences: from building damage or destruction to casualties. Around 100 people die each year in the European Alps due to snow avalanches [Techel et al., 2016]. To try in someway to prevent these accidents, a better comprehension of avalanche phenomena might be crucial.

Since the thirties, scientists have been designing and improving snow avalanche models (e.g., physical models, probabilistic models) to describe and understand better snow avalanches. Models consists in a simplification of natural phenomena and they often depend on some poorly known input parameters that cannot be necessarily measured. To cite an example, friction parameters (representing the behavior of snow in motion) cannot be measured from in situ experiments. That's why friction parameter meaning and their link to physical properties of snow is still debated in the literature of the field Naaim et al. [2013].

To go further, the following questions may be addressed. First, if we are capable to measure some other attributes of an avalanche as e.g., its velocity, *how to calibrate the unobserved parameters?* Within this purpose, avalanche data (e.g, from photogrammetric techniques) with high temporal resolution (1 s) have been collected in experimental test sites to study snow avalanches [e.g., the French test site at Lautaret Pass Thibert et al., 2015]. One solution to calibrate model parameters given a set of observations is to solve an inverse problem, which consists into finding the parameter configuration(s) that approach the best reality, where reality is known through the observations. Second, *which parameters should be calibrated with high accuracy and which others might not?* The purpose of

this thesis is to answer these two questions using a set of advanced statistical techniques.

To answer the first question, we use Bayesian inference. Bayesian inference is a powerful tool to solve inverse problems. Its main advantage is its application even if sample size is small because prior knowledge, given by scientists, can be provided to improve parameter inference. Many scientists from different fields (e.g., social and environmental sciences) are turning to Bayesian inference to solve problems where samples are scarce. In particular, avalanche observations are scarce since avalanches are rare events. Thus, Bayesian inference is the ideal framework for avalanche models and indeed, it is becoming more and more popular in the avalanche community. In particular, the velocity avalanche profile we disposed has 21 observations. For this reason, we propose to answer the first question using a Bayesian framework to calibrate friction parameters for snow in motion. For the snow avalanche field, the novelty of our approach is that we consider the error (defined as observations − model simulations) structure in the calibration. Indeed, our observed velocities are a time series and the errors are autocorrelated. Therefore within our results, we demonstrate that it is important to consider the correct structure of the errors in the calibration framework because if errors are not properly modeled, estimation of friction parameters might be biased.

To answer the second question: *how to determine which input parameters contribute the most to model output uncertainty*, we use global sensitivity analysis. Global sensitivity analysis (GSA) is a probabilistic technique which consists in modeling unknown input parameters by a probability distribution which propagates through the model to the outputs. Input parameter distributions are often guided by practitioner's belief. Then, input parameters are ordered according to their contribution on the model outputs by computing sensitivity measures. By answering this second question, we gain (i) a better comprehension of avalanche model and, (ii) we might put more efforts to calibrate important input parameters and, fix the other inputs to literature nominal values without losing model output accuracy. In fact, a major drawback of Bayesian inference is the high number of model simulations required to get algorithms convergence. Indeed, at least 1000 simulations might be required to get convergence. Moreover, algorithm convergence could depend on the number of input parameters. Thus, the number of model simulations required could be reduced by identifying influential inputs.

From a GSA scientist's perspective, avalanche model outputs offer an interesting application framework because (i) they are functional (e.g., the functional flow velocity) and, (ii) their input/output samples could be selected by acceptance-rejection (AR) algorithms. For example, we could use only avalanche simulations with certain particular characteristics e.g., having a volume superior to a prescribed threshold (e.g., $7000m^3$) and, runout distance superior to a prescribed position. In that context, independence of input parameters is no longer true and the computation and interpretation of sensitivity measures is a difficult task. Thus, the functional output nature and dependence between input parameters were our guidelines for proposing a framework for GSA in our context of snow avalanches.

In GSA literature, there exist different sensitivity measures. We focus in two variance-based sensitivity measures: Sobol' indices [Sobol', 1993] and Shapley effects [Shapley, 1953]. Sobol' indices are sensitivity measures commonly used. They can be applied if inputs are independent. This sensitivity measure is easy to interpret because, (i) all the indices are normalized between 0 and 1 and (ii) sum to 1, (iii) a high index shows the input is relevant and a value close to zero shows that it is not. When inputs are dependent,

Sobol' indices interpretation is quite difficult because they can sum to greater than one and there could be some negative indices. In contrast to Sobol' indices, Shapley effects are still meaningful when input parameters are dependent. They have nice properties because they are non negative and they sum to one. We use these two sensitivity measures to determine important inputs in avalanche models. Moreover, we study Sobol' and Shapley effects for multivariate or functional outputs. These measures are called aggregated measures.

Regarding the estimation of Sobol' indices and Shapley effects, there are two kind of methods based on the *sampling structure design*: structured sampling design (such as pick-and-freeze or replicated designs) and not structured sampling design. The last ones are also called given data methods. To use structured sampling design methods, users might require access to the model. However, such sampling designs may be impossible in some applications. In particular, in our application where samples are constructed using AR sampling. Thus, we propose estimators of aggregated Sobol' indices and aggregated Shapley effects from a given sample.

1.2 Objectives

To summarize, the thesis objectives are to answer the two scientific challenges described in Section 1.1. More specifically we address the following questions:

Q1. *Given an avalanche velocity profile, how to calibrate friction parameters?*

We propose a framework based on Bayesian inference to calibrate an avalanche model. In more detail, we propose to solve the inverse problem of finding the most likely combinations of friction parameters given a set of avalanche velocities using Bayesian inference. To use Bayesian inference, certain assumptions on the model errors (observation-model simulations) have to be made. Traditionally, errors are assumed i.i.d. Gaussian centered variables. However, velocity trajectories are time series, thus inspired in works developed in hydrology by Kuczera [1983], Kuczera and Parent [1998], Evin et al. [2014], we propose to consider the error autocorrelation to calibrate the avalanche model. To determine the impact of including or not the autocorrelation errors in model calibration, we performed two calibrations: (i) considering independent errors and (ii) considering errors are autocorrelated. We compared the results of both calibrations and results are better if autocorrelation is included. This work was published in The Journal of Glaciology in January 2020 and the work is presented in Chapter 3:

Heredia MB, Eckert N, Prieur C, Thibert E (2020). Bayesian calibration of an avalanche model from autocorrelated measurements along the flow: application to velocities extracted from photogrammetric images. *Journal of Glaciology* 66(257), 373–385. https://doi.org/10.1017/jog.2020.11

Q2. *How to perform a global sensitivity analysis for avalanche models from a given data sample when input parameters are independent or dependent and outputs are functional?*

We focus here in the following setting to perform global sensitivity analysis. We are given a sample of simulations of moderate size and we want to estimate sensitivity measures with this given sample. Within this purpose and following Racine [2001], Da Veiga and Gamboa [2013] and Solís [2019], we develop a nonparametric estimation procedure based on the Nadaraya-Watson kernel smoother to estimate aggregated Sobol' indices. Input

distributions are chosen as the posterior of the Bayesian inference. This work was submitted to Reliability Engineering & System Safety journal in March 2020 and presented in Chapter 4:

Heredia MB, Prieur C, Eckert N. Nonparametric estimation of aggregated Sobol' indices: application to a depth averaged snow avalanche model. *Submitted in March 2020.*

Eventually, we consider the setting where the sample is obtained from an acceptance/rejection rule to physical constraints. The corresponding set of input parameters present dependencies thus, we propose to estimate Shapley effects extended to multivariate or functional model outputs. We call these new sensitivity measures aggregated Shapley effects. More precisely, we propose an algorithm to estimate aggregated Shapley effects in a given data framework based in Broto et al. [2020]. We also build bootstrap confidence intervals, adapting the ideas of Benoumechiara and Elie-Dit-Cosaque [2019]. This work has been submitted to the SIAM/ASA Journal on Uncertainty Quantification in July 2020 and presented in Chapter 5:

Heredia MB, Prieur C, Eckert N., Given data inference of aggregated Shapley effects, application to avalanche modeling. *Submitted in July 2020.*

1.3 Organization of the manuscript

The manuscript is organized as follows. In Chapter 2, we propose a short introduction to (i) 1-D avalanche modeling, (ii) Bayesian inference applied to parameter model calibration and (iii) global sensitivity analysis with a particular emphasis when inputs are independent or dependent, and multivariate or functional. This chapter can be seen as a toolbox to understand the developments we have made. Then, the manuscript is divided in two parts:

Part I: Bayesian Inference.
Part I is composed by Chapter 3. In Chapter 3, we propose a Bayesian framework to calibrate an avalanche model using velocities extracted from images.

Part II: Global Sensitivity Analysis
In Part II, we develop global sensitivity methods applied to dynamic models, more particular to a 1-D avalanche model. Part II is composed by Chapter 4 and Chapter 5. In Chapter 4, we proposed a non parametric method to estimate aggregated Sobol' indices from an i.i.d. sample and, in Chapter 5, we extend Shapley effects to multivariate and functional outputs, we also propose algorithms for the estimation and the computation of bootstrap confidence intervals.

Finally, conclusions and perspectives for this thesis are exposed in Chapter 6.

2

Avalanche models, Bayesian inference and global sensitivity analysis

Contents

Overview: In this chapter, we introduce the principal tools and concepts employed in the core chapters of this manuscript (Chapters 3 to 4). A short introduction to snow avalanche dynamic modeling is presented in Section 2.1. We put particular attention to previous studies of model calibration and sensitivity analysis in the avalanche field. Then, model calibration using Bayesian inference is explained in Section 2.2 and global sensitivity analysis methods are described in Section 2.3. We put emphasis on global sensitivity methods that can be applied to models with dependent input parameters and functional model outputs.

2.1 A short introduction to 1-D snow avalanche dynamic modeling

Avalanche numerical models provide a simplification of avalanche flow on the basis of physical conservation laws. To situate avalanche models in human history, one of the first avalanche models was probably proposed in the 1930s [Salm, 2004]. This first model had a dry friction coefficient (Coulomb friction) [Salm, 2004]. Then, Voellmy [1964] proposed a based on hydrodynamics model, that had two friction coefficients: a turbulent friction coefficient ξ and a Coulomb-like dry coefficient μ. Since then, different analytical and numerical models have been proposed based in Coulomb or Voellmy friction laws [e.g., Perla et al., 1980, Salm et al., 1990]. Analytical (e.g., sliding block models) and numerical models (e.g., Saint-Venant proposed by Naaim et al. [2004]) have advantages and disadvantages as for example their accuracy in test site experiments, computational cost, number of parameters, etc. Also, some of them may be only capable to output flow

velocities and runout distances[1] while others can also compute flow depths. Nevertheless, avalanche modeling is still challenging, due to lack of measurements and due to avalanche complexity (e.g., flow regime transitions, snow mass balance) [Harbitz et al., 1998]. However, depending on the application, analytical or numerical models (or both) might be used. A more detailed review of these models can be found in e.g., Harbitz et al. [1998], Salm [2004] and Jamieson et al. [2008].

For illustration purpose, Figure 2.1 shows a graphical simplification of a sliding block model. These kind of models represent an avalanche as a sliding rigid body experimenting frictional forces. The model is based in the motion equation:

$$\frac{du}{dt} = g\sin\theta - \frac{F}{m}, \tag{2.1}$$

where $u = ||\vec{u}||$ is the velocity norm, g is the gravity constant, θ is the local slope, h is the mean flow depth, m the mass and $F = ||\vec{F}||$ the friction force. For example, Coulomb friction law $F = \mu mg\cos\theta$ or Voellmy friction law $F = \mu mg\cos\theta + \frac{mg}{\xi h}u^2$ could be considered in Equation (2.1).

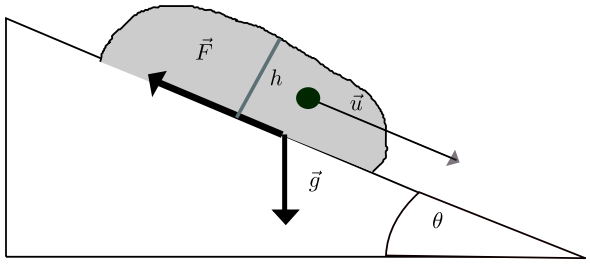

Figure 2.1: Avalanche block model.

2.1.1 Model input parameters

All of the aforementioned models, independently of their complexity, depend on unknown parameters that cannot be measured. More specifically, friction parameters cannot be measured from experiments. To guide practitioners, some authors have created parameter guidelines. For example, Salm et al. [1990] proposed reference values for μ according to the avalanche characteristics (e.g., wet snow, large avalanche) and for ξ according to the avalanche path characteristics (e.g., path local slope). Author guideline is called the Swiss guideline. For example, Christen et al. [2010b] applied the guideline recommendations for back calculation of the *In den Arlen*[2] avalanche using the 2-D RAMMS model. However, parameter in situ calibration is still necessary for a better comprehension of avalanche models. Deterministic inversion, which consists in writing friction coefficients as a function of observed quantities, has been employed to calibrate parameters from in situ experiments. In the work of Dent et al. [1998], friction coefficients were computed by measuring normal forces. Ancey et al. [2003] proposed a deterministic inversion framework to retrieve friction parameters from runout distances and the sum of snow fall amount during the three days previous to the avalanche event. Indeed, the snow fall during the days preceding the avalanche event is a key meteorological parameter in avalanche phenomena [Ancey et al., 2003]. In Ancey and Meunier [2004] a back analysis was performed to infer friction coefficients from velocities. Naaim et al. [2004] performed a back analysis to model validation, also the authors found an empirical relationship between the dry friction coefficient μ and runout distances. Recently, Oda et al. [2020] performed small scale

[1] The runout distance is the position where the avalanche stops flowing.

[2] Avalanche occurred in Switzerland the 27 January 1968. The event destroyed a farmhouse and killed four people.

experiments using a rotating drum device to measure flow velocities and flow depths in order to perform a deterministic inversion to retrieve friction parameters. Nevertheless, parameter point estimation does not take into account parameter uncertainty which is important for uncertainty propagation.

One of the principal difficulties of parameter calibration in avalanche problems is the low number of observations recorded. Therefore, more complex schemes of calibration based on Bayesian inference have been proposed. For example, Ancey [2005], Straub and Grêt-Regamey [2006] and Eckert et al. [2010] used runout distances to model calibration. Gauer et al. [2009] used the velocity at the lower avalanche track and runout distances to model calibration. Also, procedures based on optimization have been proposed. For example, Naaim et al. [2013] performed the calibration of friction parameters using runout distances from historical avalanche events. Regarding 3-D avalanche model calibration, in Fischer et al. [2014], the 3-D SamosAT (Snow Avalanche Modeling and Simulation Advanced technology) model is calibrated using flow depths and velocities deduced from Doppler radar measurements. Fischer et al. [2015] proposed a multivariate parameter optimization to calibrate SamosAT.

In the last two decades, the interest in Bayesian approaches is increasing in the avalanche community. To demonstrate this assertion, we cite some examples: the already described works of Ancey [2005], Straub and Grêt-Regamey [2006], Gauer et al. [2009], Eckert et al. [2010], Heredia et al. [2020] and other recent works, Schläppy et al. [2014] reconstructed runout distances from dendropomorphic historical data, then authors applied a Bayesian approach based on Eckert et al. [2010] to model calibration and to estimate return periods. Recently, Fischer et al. [2020] provided a comprehensive tool for snow avalanche simulation using Bayesian inference which is implemented in the open source software `r.avaflow`. The software `r.avaflow` was developed by Mergili et al. [2017].

2.1.2 Sensitivity analyses for avalanche models

As mentioned before, avalanche models depend on poorly known parameters (e.g., friction parameters and initial conditions corresponding to the avalanche released). Authors have studied the parameter sensitivity, for example, Barbolini and Savi [2001] used a Monte Carlo approach to analyze the sensitivity for the runout distances and impact pressures outputs of the VARA model. Authors found two interesting results: (i) for relatively frequent events, friction parameters and release depth and length are important for runout distances and, (ii) for extreme rare events, the friction coefficient has a higher importance than the two other parameters. Borstad and McClung [2009] developed a sensitivity analysis of an avalanche model with Coulomb friction law. Their results demonstrated that the friction coefficient is important for the runout positions. Bühler et al. [2011] studied the influence of the digital elevation model resolution (DEM) on the outputs of the RAMMS model. They found that DEM resolution and quality is critical for modeling and a spatial resolution of around 25m is sufficient for large-scale modeling. Fischer et al. [2014] performed a sensitivity analysis for RAMMS model by estimating the Spearman rank correlation between parameters and outputs. Buhler et al. [2018] made a sensitivity analysis of runout distance, volume and avalanche velocity in the runout zone with respect to the initial volume in the RAMMS model. They found that the released volume has a large effect in runout distance and avalanche velocity. However, even if all the previous

works provide useful information about avalanche models, none of them has used variance-based global sensitivity measures to quantify input parameter importance.

2.1.3 Few words about land-use planning

Finally, *why avalanche input parameters and thus, model outputs are important?* The reason is that model outputs (e.g., avalanche speed and runout distance) might be used in land use planning which typically consists into dividing a region in zones according to their potential risk [Salm, 2004]. In some countries (e.g., Austria, France, Switzerland), this assignation is done by colors: green (or white in France [Eckert et al., 2018b]), blue and red. The colors assignation are sometimes determined by runout return periods. A runout return period is the mean time in which a given runout distance is reached or exceeded at a given path's position [Eckert et al., 2007, 2008b, Schläppy et al., 2014]. For example in Switzerland, red zones have return periods lower than 300 years and avalanche pressures can exceed 30kNm^{-2} [WSL]. Constructions are forbidden in red zone [WSL]. Therefore, parameter estimation and associated uncertainty might be useful for computing more accurate avalanche simulations and as a consequence, for a better land use planning. The importance of avalanche model calibration and current issues for hazard mapping are nicely described in Jamieson et al. [2008].

To summarize, the two main axes of this thesis are model calibration using Bayesian inference and global sensitivity analysis applied to avalanche models. These two subjects are explained in Section 2.2 and Section 2.3.

2.2 Model calibration using Bayesian inference

Model calibration consists in finding the most likely combination(s) of parameters given some observations. We are not only interested in determining an optimal single combination of parameters, but also in the uncertainty associated to it [Kuczera and Parent, 1998]. Bayesian calibration offers a solution to this problem, because it allows to assess parameter uncertainty by deriving the *posterior distribution* of parameters from observations [Kennedy and O'Hagan, 2001]. Moreover, advantages of Bayesian calibration are twofold: (i) if model is used for prediction, all the sources of uncertainty are included, and (ii) since no model is perfect, the approach cope for any inadequacy between observations and model [Kennedy and O'Hagan, 2001].

Let us denote f a prediction model for a given phenomenon (e.g., an avalanche model which predicts the avalanche runout distance), f depends on parameters $\Theta \in \mathbb{R}^p$ and some forcing variables $\mathbf{x} \in \mathbb{R}^d$. Let us denote by $\mathbf{y}_{obs}$ a vector of n noisy observations:

$$y_{obs_i} = f(\Theta, \mathbf{x}) + \epsilon_i, \forall i \in \{1, \dots, n\} \tag{2.2}$$

where ϵ models observation errors. For example, $\mathbf{y}_{obs}$ could be a set of observed runout distances to calibrate an avalanche model. Note that within this formulation, we do not make distinction between the observation errors and the model inadequacy. The distinction has been studied in, e.g., Kennedy and O'Hagan [2001], Carmassi et al. [2018]. In a statistical framework to solve the inverse problem of retrieving the model parameters Θ, assumptions of the distribution of the residuals or errors ϵ might be required. For example, they could be assumed i.i.d. centered Gaussian random variables

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)\ \forall i \in \{1, \dots, n\}$. Let us denote γ the residual parameters. In our example, $\gamma = \{\sigma^2\}$. An advantage of Bayesian inference is that practitioner experience or beliefs could be plugged into the inference by the *prior distribution* denoted $\pi(\Theta, \gamma)$. Then, the posterior distribution is:

$$\pi(\Theta, \gamma | \mathbf{y}_{obs}, \mathbf{x}) = \frac{\mathcal{L}(\mathbf{y}_{obs} | \Theta, \gamma, \mathbf{x}) \pi(\Theta, \gamma)}{\int \mathcal{L}(\mathbf{y}_{obs} | \Theta, \gamma, \mathbf{x}) \pi(\Theta, \gamma) d\Theta d\gamma} \tag{2.3}$$

where $\mathcal{L}(\mathbf{y}_{obs} | \Theta, \gamma, \mathbf{x})$ is the likelihood of the observations which can be obtained from the error distribution. It is usually difficult to compute $\int \mathcal{L}(\mathbf{y}_{obs} | \Theta, \gamma, \mathbf{x}) \pi(\Theta, \gamma) d\Theta d\gamma$, except when there is an analytical expression of the posterior distribution, which is achieved by using conjugate priors. When there is no explicit posterior formulation, Monte Carlo Markov chain (MCMC) methods can be used to sample from the posterior distribution. MCMC methods consist into constructing a stationary and ergodic Markov chain that converges, under mild conditions, to the posterior distribution $\pi(\Theta, \gamma | \mathbf{y}_{obs}, \mathbf{x})$.

Among the MCMC methods, there is the Metropolis-Hasting algorithm [Metropolis et al., 1953, Hastings, 1970] which constructs sequentially a Markov chain by applying acceptance-rejection rules defined by a proposal distribution q. The convergence of the MH algorithm depends on q [Robert, 2015] and q is usually assumed as multinormal distribution.

If there is an analytical expression of all the parameter conditional distributions $\pi(\theta_\ell | \mathbf{y}_{obs}, \mathbf{x}, \theta_1, \dots \theta_{\ell-1}, \theta_{\ell+1}, \dots \theta_d)$, Gibbs sampling, a particular instance of MH algorithm, could be applied. The Hamiltonian Monte Carlo [Duane et al., 1987] algorithm, also known as hybrid Monte Carlo, has shown good results in applications. Nevertheless, the gradient of the distribution functions are required and their computations might be expensive.

When observations are time series, errors could be autocorrelated or/and suffer from heteroscedasticity. In the calibration of hydrological models, authors have studied the importance of considering the correlations and/or heteroscedasticity in residuals [e.g., Kuczera, 1983, Kuczera and Parent, 1998, Evin et al., 2014]. Indeed, they found that if autocorrelation is not included parameter estimation could be biased. In Chapter 3, given an avalanche velocity measured time series, we analyze the impact of considering or not the error autocorrelation in a Bayesian inference calibration framework. A block model based on the Voellmy friction law is considered.

2.3 Global sensitivity analysis

Saltelli et al. [2004] defined precisely the sensitivity analysis (SA) as "the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input"[3]. Under a SA framework, we consider a deterministic model f (computational code, black box model, etc) that depends on d inputs $\mathbf{X} = (X_1, \dots, X_d)$. First for simplicity, we consider the output Y as scalar $Y = f(\mathbf{X})$. The inputs are modeled by random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Thus, the output Y is also a random variable $(\Omega, \mathcal{F}, \mathbb{P})$ because it depends on the random vector $\mathbf{X}$. Let us denote $(\mathbf{x}, y)$ a random sample from $(\mathbf{X}, Y)$. Figure 2.2 illustrates the

[3]This definition is, in general, widely accepted in the SA community.

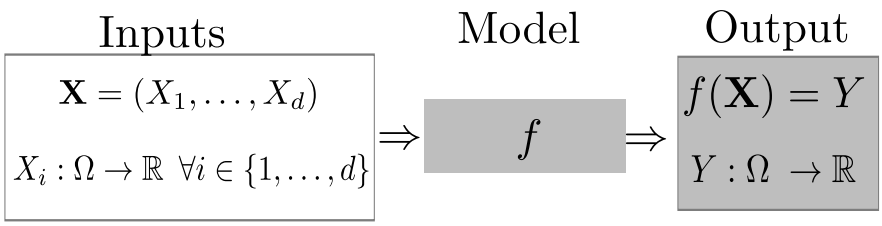

Figure 2.2: SA framework.

SA framework. The following exposition is inspired by the works of Iooss and Lemaître [2015] and Iooss and Saltelli [2017].

SA is divided in two classes: local and global methods. Local methods consist in studying the impact of small input perturbations around certain nominal values on the output. An example of local methods are partial derivatives, they consist into calculating $\frac{\partial f}{\partial x_i}$ at a given point $\mathbf{x}^0 = (x_1^0, \ldots, x_d^0)$ in the domain of $\mathbf{X}$. Local methods are sometimes used when model computational cost or number of parameters is high. For example, Castaings et al. [2009] proposed the adjoint state method to compute efficiently the derivatives of the output of a distributed hydrological model with respect to parameter inputs to perform local SA. Castillo et al. [2004] proposed a local SA method based on the duality property of mathematical programming to compute partial derivatives. In contrast to local methods, global methods explore all the input domain and not only variations around nominal values.

Global methods provide a more convincing tool for highly nonlinear models often encountered in environmental sciences. Indeed, global methods provide an exploration of the whole input space [Saltelli et al., 2008, Iooss and Lemaître, 2015]. Nonetheless, there is a price to pay to use global sensitivity methods in terms of number of model simulations. But there are some ways to reduce computational costs which will be explained in Subsection 2.3.5. Tang et al. [2006] compared local and global methods applied to a watershed model and they concluded that global methods yield robust results. [Saltelli et al., 2008, page 11] recommended strongly to use global methods rather than local ones. Hereafter, we focus on global sensitivity analysis methods.

Global sensitivity analysis (GSA) consists into identifying the input parameters that contribute the most (in a sense to be precised) to a given quantity of interest defined from the output of the model [Iooss and Lemaître, 2015]. The objectives of GSA are numerous: they are to provide a better comprehension of the numerical model, to identify non influential parameters to set them to nominal values, among others [Iooss and Lemaître, 2015]. To guide practitioners, Saltelli et al. [2004] proposed four settings of GSA objectives: factor prioritization, factor fixing, variance cutting and factor mapping. Before performing a GSA, it is important to have clear the objectives of the study [Iooss and Lemaître, 2015] and also, to select the adequate GSA method to avoid misinterpretations [Saltelli et al., 2019]. Then, parameter inputs might be ordered according to their importance or contributions by computing sensitivity measures.

GSA has been widely applied in many fields as for example, in environmental sciences and engineering. We cite some recent works in different application domains, Sarrazin et al. [2016] proposed a guide to apply GSA in environmental models, more particularly applied to hydrological models. Mavromatidis et al. [2018] developed a GSA for distributed energy systems. These models are important for an efficient and sustainable energy future. Prieur et al. [2019] performed a GSA for a marine biogeochemical model

with a high number of inputs (74 input parameters).

The GSA measures may be divided into: screening approaches, variance-based measures, moment free measures and, other measures. In the manuscript, we only focus on variance based measures which are explained in Subsection 2.3.2. Even if we do not use the other three GSA measures, they are briefly explained in Subsections 2.3.1, 2.3.3 and, 2.3.4 to provide a general overview of GSA measures.

2.3.1 Screening approaches

Screening approaches consist in the discretization of inputs $\mathbf{X}$ in levels [Iooss and Lemaître, 2015]. If the model has a large number of inputs $d \gg 1$, these sensitivity measures could help the user to identify non influential input parameters. Therefore, screening methods could be a preliminary step before computing other, in general more expensive, GSA measures when input dimension is high. One of the most popular screening methods is the OAT (one at a time) Morris method [Morris, 1991].

Here we describe briefly the Morris method. The d random inputs $\mathbf{X}$ are supposed *independent* and are modeled with uniform distributions on $[0,1]$, i.e. $X_i \sim \mathcal{U}[0,1]$ for all $i \in \{1,\dots,d\}$. The assumption of uniformity on $[0,1]$ is not constraining as for X_i with cumulative distribution function (cdf) F, we can apply the inverse transform sampling to come back to that setting.

The rectangular domain $[0,1]$ of each input is divided in a p-level grid, then each x_i may take values from $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$. Let $\Omega = \{0, 1/(p-1), 2/(p-1), \dots, 1\}^d$. Let Δ be a predetermined multiple of $1/(p-1)$. For a given value $\mathbf{x} \in \omega$, the elementary effect of the input X_i is defined according to:

$$d_i(\mathbf{x}) = \frac{f(x_1, x_2, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_d) - f(x_1, \dots, x_d)}{\Delta}$$

$\mathbf{x}$ must be such that $x_i + \Delta$ remains in $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$. We sample n points $\mathbf{x}^1, \dots, \mathbf{x}^n$ from $\mathbf{X}$ on Ω such that $x_i^\ell + \Delta$ is in $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$. Then, we compute $d_i(\mathbf{x}^\ell)$ for each $\ell \in \{1, \dots, n\}$. Finally, we can estimate the mean absolute value $\hat{\mu}_i$ and standard deviation s_i of the elementary effects. By calculating the indices $\hat{\mu}_i$ and s_i, the Morris method allows us to identify three kind of inputs (i) non influential inputs ($\hat{\mu}_i$ will have a low value), (ii) influential inputs with linear effects ($\hat{\mu}_i$ will be large and s_i will be small) and, (iii) influential inputs with non-linear and/or interactions effects ($\hat{\mu}_i$ and s_i will both be large). Graphical 2-D representation of $(\hat{\mu}_i, s_i)$ points might help the user to classify the inputs [see Iooss and Lemaître, 2015]. Here we will not explain the details of the Morris measure estimation (e.g., the structure of the design of experiment) and parameter tuning (e.g., Δ, p, the threshold for defining small or large $\hat{u}_i$ and s_i) since we do not use them in the core chapters of the manuscript, but more information can be found in the works of Morris [1991] or Campolongo et al. [2007], the last work proposed an improved sampling strategy.

Example

We illustrate the method in the Morris function [modified example from Morris, 1991]. The Morris function has $d = 20$ independent inputs $X_i \sim \mathcal{U}[0,1]$ $\forall i \in \{1, \dots, 20\}$ and is constructed as follows:

$$y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<\ell}^{20} \beta_{ij\ell} w_i w_j w_\ell + \sum_{i<j<\ell<s}^{20} \beta_{ij\ell s} w_i w_j w_\ell w_s, \tag{2.4}$$

where $w_i = 2(X_i - 1/2)$ except for $i = 3, 5$ and 7 where $w_i = 2(1.1X_i/(X_i + 0.1) - 1/2)$. The β coefficients are defined as:

$$\begin{aligned}
\beta_i &= 20, \qquad i = 1, \ldots, 7, \beta_8 = 25, \beta_9 = 30, \beta_{10} = 40, \\
\beta_{ij} &= -15, \qquad i, j = 1, \ldots, 6, \\
\beta_{ij\ell} &= -10, \qquad i, j, \ell = 1, \ldots, 5, \\
\beta_{ij\ell s} &= 5, \qquad i, j, \ell, s = 1, \ldots, 4,
\end{aligned}$$

the remainder coefficients are set to 0. Figure 2.3 displays $(\hat{\mu}_i, s_i)$ with $i \in \{1, \ldots, 20\}$. The standard error of the mean which is estimated as $s_i/\sqrt{n}$ (result from the central limit theorem) allow us to display confidence intervals for the mean corresponding to $\mu \pm 2s/\sqrt{n}$ (black line). Within this line, we can deduce the following results: inputs X_{11} to X_{20} are negligible because their means and standard deviations are 0. Inputs X_1 to X_{10} are influential. More particularly, inputs X_8, X_9 and X_{10} have large means and small standard deviations which suggest that they are influential with linear effects. And finally, the remaining inputs (X_3 to X_{10}) have non linear and/or interaction effects.

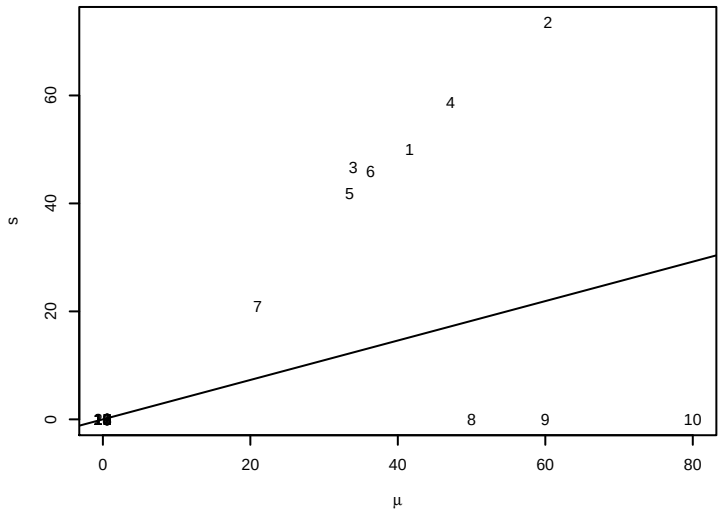

Figure 2.3: *Morris function.* $\hat{\mu}_i$ and s_i estimated with $n = 30$, $p = 8$, $\Delta = 4/7$ (a total of 630 model runs). Black line corresponds to $\mu = 2s/\sqrt{30}$.

As we have seen in this example, the Morris method gives useful insights about input importance. However, to rank unambiguously the inputs according to their importance, we will need other sensitivity measures and they are explained in the following section.

2.3.2 Variance-based measures

Variance-based measures analyze how uncertainty in inputs impacts the *variance* of the output(s). The two main variance-based measures are Sobol' indices [Sobol', 1993] and Shapley effects [Owen, 2014]. Sobol' indices are based on a functional ANOVA decomposition [Efron and Stein, 1981], which is unique if inputs are independent, and Shapley effects are based on the cooperative game theory concept of Shapley value [Shapley, 1953]. The main difference between both measures is that Shapley effects are easier to interpret if the inputs are not independent.

Sobol' indices

If the output of a model is scalar, its total variance can be split into partial variances by using the Hoeffding decomposition [Hoeffding, 1948]. If the inputs are independent, each of the partial variances is associated to an input or to an interaction between inputs through the functional ANOVA decomposition [Efron and Stein, 1981]. Then, the Sobol' indices [Sobol', 1993] are calculated as the ratio of each partial variance to the total variance. In mathematical terms, this is expressed as follows: let us suppose the random variables $X_1, \ldots, X_d$ are *independent* from each other and $f \in L^2(\mathbb{P}_{\mathbf{X}})$ where $\mathbb{P}_{\mathbf{X}} = \prod_{i=1}^{d} \mathbb{P}_{X_i}$. Thus, f can be decomposed as [Hoeffding, 1948]:

$$f(\mathbf{X}) = f_{\emptyset} + \sum_{i=1}^{d} f_i(X_i) + \sum_{1 \leq i < \ell \leq d} f_{i\ell}(X_i, X_\ell) + \ldots + f_{\{1,\ldots,d\}}(X_1, \ldots, X_d), \tag{2.5}$$

where $f_{\emptyset} = \mathbb{E}(Y)$ and $f_{\mathfrak{u}}(X_{\mathfrak{u}}) = \mathbb{E}(Y|X_{\mathfrak{u}}) - \sum_{\mathfrak{v} \subsetneq \mathfrak{u}} f_{\mathfrak{v}}(X_{\mathfrak{v}}) = \sum_{\mathfrak{v} \subseteq \mathfrak{u}} (-1)^{|\mathfrak{u}|-|\mathfrak{v}|} \mathbb{E}(f(\mathbf{X})|X_i, i \in \mathfrak{v})$. This decomposition is the so-called functional ANOVA decomposition. From (2.5) and from the fundamental independence assumption we get [Efron and Stein, 1981]:

$$\mathrm{Var}(Y) = \sum_{i=1}^{d} \mathrm{Var}(f_i(X_i)) + \sum_{1 \leq i < \ell \leq d} \mathrm{Var}(f_{i\ell}(X_i, X_\ell)) + \ldots + \mathrm{Var}(f_{\{1,\ldots,d\}}(X_1, \ldots, X_d)). \tag{2.6}$$

Then, Sobol' indices are defined:

- First-order Sobol' index of input X_i:

$$S_i = \frac{\mathrm{Var}(f_i(X_i))}{\mathrm{Var}(f(\mathbf{X}))} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(f(\mathbf{X}))}. \tag{2.7}$$

- Second-order Sobol' index of input X_i and X_j, to quantify interactions:

$$S_{i\ell} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_\ell))}{\mathrm{Var}(f(\mathbf{X}))} - S_i - S_\ell. \tag{2.8}$$

- Total Sobol' index of input X_i to quantify the effect of X_i and all its interaction with the other inputs:

$$S_i^T = 1 - \frac{\mathrm{Var}(\mathbb{E}(Y|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d))}{\mathrm{Var}(f(\mathbf{X}))}. \tag{2.9}$$

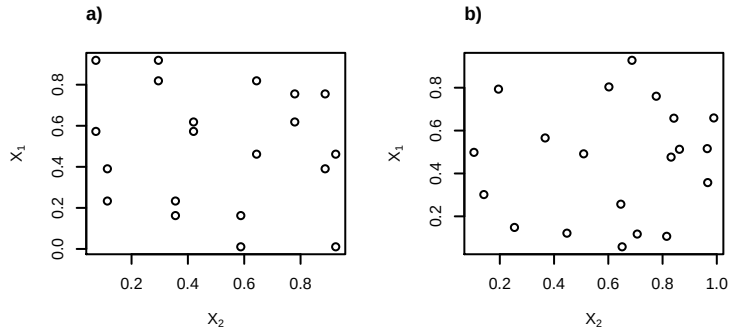

Figure 2.4: $n = 20$ samples of X_1, X_2 random variables uniformly distributed on $[0, 1]$ a) structured sample design required to estimate Sobol' indices using Tissot and Prieur [2015] method. b) A random sample.

Higher order indices might be also defined (e.g., S_{i_1,i_2,i_3} with $i_1, i_2, i_3 \in \{1, \ldots, d\}$) but in general, only the aforementioned indices are computed in applications. Sobol' indices interpretation is easy (if inputs are independent, otherwise it becomes challenging) because they take values in $[0, 1]$, they sum to one and, a high value means the input is influential.

Estimation methods

The estimation of Sobol' indices is a challenging issue because, except for rare cases, a high number of simulations is required. Indeed, to estimate a single Sobol' index with an uncertainty of 10%, it could be required to perform 10^4 model runs [Iooss and Lemaître, 2015]. For explanation purpose, it is convenient to divide the estimation techniques according to the *sampling structure*: structured sample design and given data methods. The structured sample design techniques can be applied if we have access to the model f but, in some applications we only dispose of a random i.i.d. n sample from $(\mathbf{X}, Y)$. Figure 2.4 shows a structured sample and a non structured sample designs.

Among the structured sample methods, we can mention the Fast Amplitude Sensitivity Test (FAST) introduced by Cukier et al. [1978] [see, e.g., Saltelli et al., 1999, and references therein], Sobol' pick-freeze schemes [Sobol', 1993], FAST mixed with Random Balance Designs [Tarantola et al., 2006]. Tissot and Prieur [2015] proposed a pick-freeze procedure based on replicated sampling which reduce the estimation cost. In Prieur and Tarantola [2017], the most efficient estimation methods are well presented. Moreover, depending on the estimation method, confidence intervals or significance test could be deduced from asymptotic results of the central limit theorem [see more details in Janon et al., 2014a]. However, bootstrap intervals can be also built, under some conditions, to construct error bars.

To overcome the drawback of structured sample methods (computational cost, access to the model), authors propose given data methods. Among the given data methods, we can cite the EASI spectral method of Plischke [2010], Plischke et al. [2013] method relies on the notion of class-conditional densities, where a class is a sub-sample stemming from a suitable partition of the dataset, the nonparametric estimation methods of Da Veiga and Gamboa [2013] or Solís [2019], the fully Bayesian given data procedure proposed by Antoniano-Villalobos et al. [2019]. And more recently, the Gamboa et al. [2020] estimators based on rank statistics. It is worth to mention that the EASI method gives good results in test cases even with a small bunch of observations (e.g., 200 samples). The two precursory given data methods, the

EASI method and NSD by Plischke et al. [2013] method, are explained in more detail and compared to the structured sample method of Tissot and Prieur [2015] in Appendix A.

In Chapter 4, we propose a given data method to estimate first-order Sobol' indices, based on a nonparametric Nadaraya-Watson [Nadaraya, 1964, Watson, 1964] kernel regression with bias corrected [Racine, 2001] to quantify the importance of the inputs of an avalanche model.

Dependent inputs

Independence assumption on input parameters is crucial although unrealistic in many applications for a comprehensive interpretation of Sobol' indices. In fact, if input parameters are dependent, the variance decomposition in Equation (2.6) used for the interpretation of Sobol' indices is not true anymore. Some authors have proposed strategies to estimate Sobol' indices if input parameters are dependent. Jacques et al. [2006] propose to separate the dependent inputs in independent groups and to compute Sobol' indices for each group. In other words, they propose to organize the inputs in s independent groups $\mathbf{X} = (X_{\mathfrak{u}_1}, \ldots, X_{\mathfrak{u}_s})$ with $\mathfrak{u}_\ell \subset \{1, \ldots, d\}$ and then compute, the Sobol' index $S_{\mathfrak{u}_\ell} = \mathrm{Var}(\mathbb{E}(Y|X_{\mathfrak{u}_\ell}))/\mathrm{Var}(Y)$ for all $\ell \in \{1, \ldots, s\}$. This method has two main limitations: it assumes all variables are not correlated and, we cannot recognize the influence of single inputs. Da Veiga et al. [2009] proposed to estimate Sobol' indices using local polynomials. Broto et al. [2019] studied the Sobol' indices for independent groups in the particular case of block-additive models.

Other works focused on alternative ANOVA based indices when inputs are dependent. Xu and Gertner [2008] [see also Zhang et al., 2015] proposed to decompose the contribution in two parts: the correlated and uncorrelated contributions, the method assumes that the linearity of both the model and dependence structure which limits its application. Li et al. [2010] proposed three sensitivity indices to reflect the total, structural (linked to the system structure) and correlative contributions (linked to the correlated input distribution) extending the results of Xu and Gertner [2008] to more general models using the concept of High Dimensional Model representation (HDMR), Mara and Tarantola [2012] proposed a first step consisting into decorrelating the inputs using the Gram-Schmidt procedure and then, to perform the ANOVA-HDMR decomposition and compute the sensitivity measures. However, the decorrelation procedure depends on the variable order which makes intricate the index interpretation [Chastaing et al., 2015]. Chastaing et al. [2012] define new sensitivity measures based on the Generalized Hoeffding decomposition proposed by Stone [1994], these new indices are called the generalized Sobol' indices. However, the index construction requires specific assumptions on the joint distribution function of the inputs [Chastaing et al., 2012]. Also, their interpretation is not trivial because even if they sum to one, they could be negative [Chastaing et al., 2015]. Kucherenko et al. [2012] proposed a copula-based approach to sample multivariate distributions but, it requires a priori knowledge of the conditional densities. Moreover, index interpretation may not be clear because if input correlation is high, first order index are higher than the total one. Mara et al. [2015] established a link between Kucherenko et al. [2012] and Mara and Tarantola [2012] works proposing the estimation of four indices, namely full and independent first-order and total indices. Recently, Hart and Gremaud [2018] reformulated total Sobol' indices in terms of approximation errors rather than variance analysis. More precisely, they proposed to see total Sobol' indices as a measure of discrepancy between the model and a surrogate using less variables, i.e. the error to approximate $f(\mathbf{X})$ by $f(g(X_{-u}), X_{-u})$ where $g(X_{-u})$ is an approximation of X_u. However, all the afore-

mentioned works have drawbacks and specially they do not provide an univocal way of partitioning the influence of input parameters on the output. For these reasons, Shapley effects seems more adapted in a dependent framework. Sensitivity measures when inputs are dependent is a still ongoing research subject which has been recently studied in the Master thesis of Wiederkehr [2018] and Ph. D. thesis of Broto [2020].

Shapley effects

In contrast to Sobol' indices, Shapley effects [Owen, 2014] are meaningful in the framework of dependent input parameters [Owen and Prieur, 2017]. This measure is based on the Shapley value which is a cooperative game theory concept. Briefly speaking, Shapley value ensures a fair distribution of a gain among team players according to their individual contributions. As a sensitivity measure, Owen [2014] adapted the Shapley value [Shapley, 1953] into the Shapley effects by considering model input parameters as players and the gain function as the output variance. The main advantage of such an approach is that it is possible to attribute a non negative sensitivity index to each parameter, and the sum of the indices is equal to one [Broto et al., 2020, Iooss and Prieur, 2019].

The definition of Shapley effects is based on Shapley value concept which is next explained. Given a set of d players in a coalitional game and a characteristic function $\text{val} : \mathcal{S} \to \mathbb{R}$, $\text{val}(\emptyset) = 0$, where $\mathcal{S}$ is the set of all subsets of $\{1, \dots, d\}$, the Shapley value $(\phi_1, \dots, \phi_d)$ is the only distribution of the total gains $\text{val}(\{1, \dots, d\})$ to the players satisfying the desirable properties listed below:

1. (Efficiency) $\sum_{i=1}^{d} \phi_i = \text{val}(\{1, \dots, d\})$.
2. (Symmetry) If $\text{val}(\mathfrak{u} \cup \{i\}) = \text{val}(\mathfrak{u} \cup \{\ell\})$ for all $\mathfrak{u} \subseteq \{1, \dots, d\} - \{i, j\}$, then $\phi_i = \phi_\ell$.
3. (Dummy) If $\text{val}(\mathfrak{u} \cup \{i\}) = \text{val}(\mathfrak{u})$ for all $\mathfrak{u} \subseteq \{1, \dots, d\}$, then $\phi_i = 0$.
4. (Additivity) If val and val' have Shapley values ϕ and ϕ' respectively, then the game with characteristic function val + val' has Shapley value $\phi_i + \phi'_i$ for $i \in \{1, \dots, d\}$.

It is proved in Shapley [1953] that according to the Shapley value, the amount that player i gets given a coalitional game (val, d) is:

$$\phi_i = \frac{1}{d} \sum_{\mathfrak{u} \subseteq -\{i\}} \binom{d-1}{|\mathfrak{u}|}^{-1} (\text{val}(\mathfrak{u} \cup \{i\}) - \text{val}(\mathfrak{u})) \ \ \forall i \in \{1, \dots, d\}. \tag{2.10}$$

The Shapley effects are defined by considering the characteristic function of the game as:

$$\text{val}(\mathfrak{u}) = \frac{\text{Var}(\mathbb{E}(Y|\mathbf{X}_\mathfrak{u}))}{\text{Var}(Y)}, \ \mathfrak{u} \subseteq \{1, \dots, d\}$$

in Equation (2.10). Thus, the scalar Shapley effect of input i is defined as:

$$Sh_i = \frac{1}{d \, \text{Var}(Y)} \sum_{\mathfrak{u} \subseteq -\{i\}} \binom{d-1}{|\mathfrak{u}|}^{-1} (\text{Var}(\mathbb{E}(Y|\mathbf{X}_{\mathfrak{u} \cup i})) - \text{Var}(\mathbb{E}(Y|\mathbf{X}_\mathfrak{u}))). \tag{2.11}$$

Estimation methods

Song et al. [2016] mentioned that Shapley effects estimation is costly because one has to consider all possible subsets of players (inputs) which is 2^d. In their work, two estimators are proposed: a first estimator constructed by enumerating all possible players permutations, the permutations set is denoted $\mathcal{S}_d$ and a second estimator, which is a modified and improved version of Castro et al. [2009] algorithm, which selects randomly m permutations from $\mathcal{S}_d$. Note that m should be chosen as large as possible. Benoumechiara and Elie-Dit-Cosaque [2019] proposed bootstrap confidence intervals for Song et al. [2016] estimators. Song et al. [2016] algorithms assume that it is possible to sample from the distribution of all subsets of the input parameters conditionally to the complementary set of input parameters which might be complicated in some applications. Plischke et al. [2020] proposed an estimation algorithm based on Song et al. [2016] algorithm and the Möbious inverse to reduce estimation computational cost. Their algorithm also requires the explicit knowledge of the conditional distributions. Finally, Broto et al. [2020] proposed given data estimators based on nearest-neighbor, i.e. estimators that can be computed from an i.i.d. sample, which is in general more convenient for real applications. In Chapter 5, we extend Broto et al. [2020] work to functional outputs (see Subsection 2.3.6) and we construct bootstrap confidence intervals for these estimators.

Example

To illustrate the difficulty of ranking dependent inputs using Sobol' indices, we consider a linear model with three Gaussian inputs when two of them are highly correlated [example from Section 3.4 Iooss and Prieur, 2019]. Let us consider the linear model:

$$Y = f(\boldsymbol{X}) = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{X}$$

with $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = (1, 1, 1)^T$, $\boldsymbol{X} = (X_1, X_2, X_3)^T$, $\boldsymbol{X} \sim \mathcal{N}(\mu, \Sigma)$ with $\mu = (0, 0, 0)^T$ and $\Sigma \in \mathbb{R}^{3\times 3}$ a positive-definite matrix. X_1 is independent from X_2 and X_3, and X_2 and X_3 are highly correlated (correlation $\rho = 0.9$). We set $\sigma_1 = \sigma_2 = 1$ and $\sigma_3 = 2$. The covariance matrix is then:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1.8 \\ 0 & 1.8 & 4 \end{pmatrix}$$

The variance of the output is $\sigma^2 = \mathrm{Var}[f(\boldsymbol{X})] = \sum_{j=1}^3 \beta_j^2\sigma_j^2 + 2\rho\beta_2\beta_3\sigma_2\sigma_3$.

There exists analytical expressions for Sobol' and total indices, and Shapley effects for this test case and they are the following:

- Shapley effects [analytical expression from Iooss and Prieur, 2019]:

$$\begin{aligned} Sh_1 &= \beta_1^2\sigma_1^2/\sigma^2, \\ Sh_2 &= [\beta_2^2\sigma_2^2 + \rho\beta_2\beta_3\sigma_2\sigma_3 + \frac{\rho^2}{2}(\beta_3^2\sigma_3^2 - \beta_2^2\sigma_2^2)]/\sigma^2, \\ Sh_3 &= [\beta_3^2\sigma_3^2 + \rho\beta_2\beta_3\sigma_2\sigma_3 + \frac{\rho^2}{2}(\beta_2^2\sigma_2^2 - \beta_3^2\sigma_3^2)]/\sigma^2. \end{aligned}$$

- Sobol' indices [analytical expression from Kucherenko et al., 2012]:

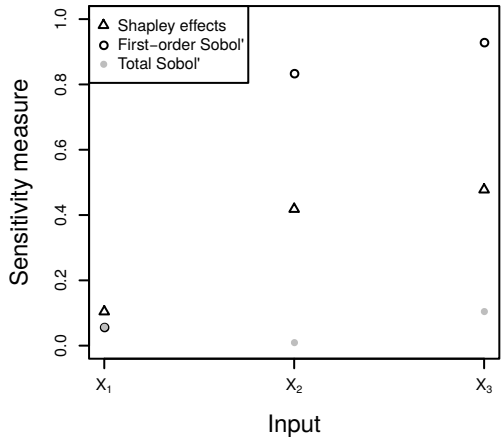

Figure 2.5: Comparison between Shapley effects and Sobol' indices when inputs are correlated on a linear model with three Gaussian inputs. X_1 is independent from X_2 and X_3, and X_2 and X_3 are highly correlated (correlation $\rho = 0.9$).

$$
\begin{aligned}
S_1 &= \frac{1}{2+\sigma^2+2\rho\sigma}, \\
S_2 &= \frac{(1+\rho\sigma)^2}{2+\sigma^2+2\rho\sigma} \\
S_3 &= \frac{(\sigma+\rho)^2}{2+\sigma^2+2\rho\sigma}.
\end{aligned}
$$

- Total Sobol' indices [analytical expression from Kucherenko et al., 2012]:

$$
\begin{aligned}
S_1^T &= \frac{1}{2+\sigma^2+2\rho\sigma}, \\
S_2^T &= \frac{1-\rho^2}{2+\sigma^2+2\rho\sigma} \\
S_3^T &= \frac{\sigma^2(1-\rho^2)}{2+\sigma^2+2\rho\sigma}.
\end{aligned}
$$

Figure 2.5 shows the analytical sensitivity measures for the linear model example. This example illustrates how difficult it is to interpret Sobol' indices if inputs are correlated since first-order indices are higher than total indices (see the indices of X_2 and X_3). On the contrary to Sobol' indices, Shapley effects are easy to interpret.

2.3.3 Moment-free measures

In this section, we describe briefly other useful global sensitivity measures that could be applied when inputs are correlated. These measures are not used in the manuscript core chapters but they offer an interesting approach to GSA when inputs are dependent. Moment-free sensitivity measures were early proposed by Borgonovo [2007] [see also Borgonovo et al., 2011]. These sensitivity measures are useful because they analyze the influence of input uncertainty on the whole output distribution, this is why they are called moment-free, and they can be defined also in the presence of correlated inputs

[Borgonovo, 2007]. Indeed, Borgonovo [2007] affirms that the influence of an input on the model output is represented by the entire uncertainty distribution and not only its variance.

The Borgonovo's δ sensitivity measure for input X_i is defined according to:

$$\delta_i = \frac{1}{2}\mathbb{E}_{X_i}(s(X_i)),$$

where $s(X_i) = \int |f_Y(y) - f_{Y|X_i}(y)|dy$, f_Y is the probability density function of Y, $f_{Y|X_i}$ is the conditional density function of Y given X_i. Borgonovo' δ quantifies the shift area between f_Y and $f_{Y|X_i}$. Borgonovo's δ has some nice properties: (i) δ_i lies between 0 and 1 for all $i \in \{1, \ldots, p\}$, and (ii) it is equal to 0, if Y is independent of X_i. Borgonovo's δ can also be applied to group of inputs [see Equation 21 of Borgonovo, 2007].

Da Veiga [2015] extended the work of Borgonovo [2007] by proposing a general GSA framework based on the concept of dissimilarity measures (e.g. the Csiszàr f-divergence, integral probability metrics) and dependence measures (e.g., Hilbert-Schmidt independence criterion HSIC). More precisely, the impact of X_i on Y could be given by [see also Baucells and Borgonovo, 2013]:

$$D_i = \mathbb{E}_{X_i}(d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})),$$

where $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X_i}$ are the probability measures of Y and $Y|X_i$, respectively and, d is a dissimilarity measure between two probability measures. Borgonovo's δ is a particular case where the dissimilarity measure is the Csiszàr f-divergence. The first-order Sobol' index is also a particular case if $d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i}) = (\mathbb{E}(Y) - \mathbb{E}(Y|X_i))^2$.

It is worth to mention that HSIC measure developed for GSA by Da Veiga [2015] [see also Meynaoui et al., 2019] can be applied to perform target sensitivity analysis (TSA) which consists into measuring the input importance on a restricted or critical domain of the model output [see more details in Marrel and Chabridon, 2020]. TSA has a wild spectrum of applications e.g., in reliability and risk management where the critical output domain is sometimes associated to a low probability [Marrel and Chabridon, 2020]. Also, HSIC measure can be applied to optimization problems with variable selection. The variable selection is performed through GSA. Indeed, in optimization problems with large number of parameters, identifying the influential inputs might be crucial to reduce the problem dimension and therefore, the computational cost of such complex applications [see more detail in Spagnol et al., 2019].

2.3.4 Other measures

There are other global sensitivity measures, to mention some of them we can cite the derivative based global sensitivity measures (DGSM) proposed by Sobol' and Kucherenko [2009]. DGSM extend local sensitivity measures based on derivatives by computing integrals of the square derivatives on the whole input domain i.e. $\mathbb{E}\left(\frac{\partial f}{\partial X_i}\right)^2$ (see also Lamboni et al. [2013]). DGSM requires less model evaluations in comparison to other GSA measures [Lamboni et al., 2013]. However to apply DGSM, regularity assumptions might be required. Also, they are not recommended for highly nonlinear functions because the ranking obtained may suggest false conclusions [Sobol' and Kucherenko, 2009]. Global sensitivity metrics from active subspaces, called activity scores, were recently proposed by Constantine and Diaz [2017]. Active subspaces are emerging tools based on spectral

decomposition for dimension reduction [Constantine and Diaz, 2017]. Finally, one could fit a linear model to obtain some information about the model, and then compute Pearson correlation or partial correlation coefficients, these methods are called the sampling-based GSA methods [Helton et al., 2006]. However, these kind of methods require linear and/or monotony hypotheses which limit their application [see Iooss and Lemaître, 2015, Section 3.1.]. Nevertheless, scatterplots will always give some preliminary information about the input-output dependence structure (e.g., linear, non linear) thus, they should not be neglected if the number of parameters is moderate[4]. This exposition was not exhaustive but it provides good insights of the current works developed in the GSA community. Other surveys of GSA can be found in e.g., Iooss and Lemaître [2015] and Borgonovo et al. [2016].

2.3.5 A little bit about metamodels

As already mentioned, the estimation of GSA measures, require large number of model simulations. If a model is computationally expensive, the construction of a metamodel or surrogate could be necessary to compute GSA measures. A metamodel is a mathematical construction to approximate f with a low computational cost. There exists a lot of metamodels (e.g., linear regression model, polynomials, splines). In particular, some metamodels provide Sobol' indices analytical expressions which means, they could be directly deduced without any additional cost. For example, Sobol' indices can be deduced from Gaussian process metamodels [Marrel et al., 2009] [see also Durrande et al., 2013, for a special case of ANOVA kernels], polynomial of chaos [Blatman and Sudret, 2010].

2.3.6 Multivariate and/or functional model outputs

Until here, we supposed Y scalar but some avalanche model outputs are functional. For example, Figure 2.6 shows the functional high density region (FHDR) bloxplots [see FHDR in Hyndman and Shang, 2010] of the functional outputs (panel a flow depth and panel b velocity) and scalar output (panel c runout distance) of $n = 5000$ random samples from the Saint-Venant [Naaim et al., 2004] avalanche model. If the output is multivariate $f(\mathbf{X}) = (Y_1, \dots, Y_p)$ or functional $Y = f_s(\mathbf{X})$ with $s \in \tau \subset \mathbb{R}$ where τ is the temporal or spatial domain of the phenomenon, it is possible to apply sensitivity analysis to each component of the multivariate output but this could lead to redundancies in the results, particularly in the setting of discretized functional outputs. Lamboni et al. [2009] and Gamboa et al. [2013] extended Sobol' indices to multivariate or functional outputs, see also Alexanderian et al. [2020] whose work extended Sobol' indices to time-dependent outputs.

The aggregated Sobol' indices are defined as:

$$GS_i = \frac{\sum_{j=1}^{p} \mathrm{Var}(Y_j) S_i^j}{\sum_{j=1}^{p} \mathrm{Var}(Y_j)},$$

where S_i^j is the first-order Sobol' index of the output Y_j with respect to the input X_i given in Equation (2.7).

[4] If the number of inputs is high (e.g., GPT-3 autoregressive model which has 175 billion of parameters), it might be better to use Morris method first.

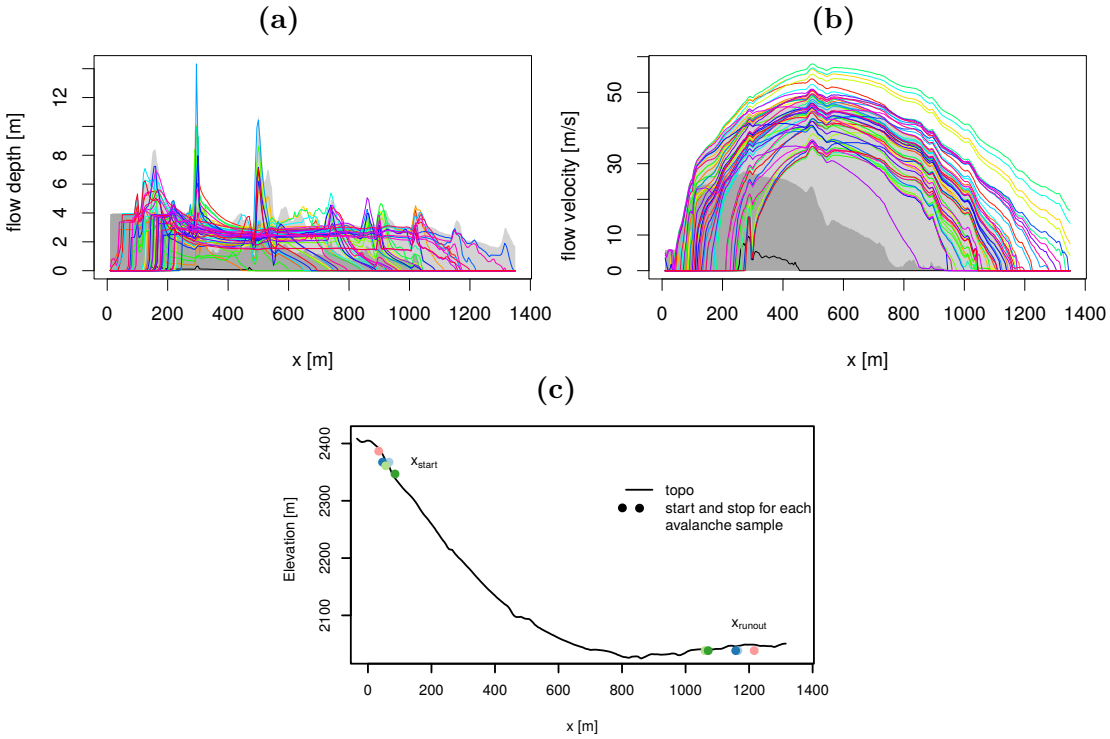

Figure 2.6: FHDR boxplots of a random $n = 5000$ sample of avalanche runs a) flow depth, b) flow velocity. c) Some examples (5 selected runs) of released and runnout positions. The topography (black line) corresponds to an avalanche path at the Lautaret test site.

Shapley effects can be naturally extended to multivariate output models by following the ideas presented by Lamboni et al. [2009], Gamboa et al. [2013]. Thus, aggregated Shapley effect of an input i is then defined as:

$$GSh_i = \frac{\sum_{j=1}^{p} \mathrm{Var}(Y_j) Sh_i^j}{\sum_{j=1}^{p} \mathrm{Var}(Y_j)}, \tag{2.12}$$

where Sh_i^j is the scalar Shapley effect of input X_i on output Y_j given in Equation (2.11). In Chapter 5, we introduce more precisely the extension of Shapley effects to multivariate or functional outputs. Note that aggregated HSIC indices proposed by Da Veiga [2015] could be also used for multivariate outputs.

If the output dimension is high (as it is the case, e.g., when considering the discretization of a functional output), a dimension reduction should be applied as a preliminary step to estimate efficiently aggregated Sobol' or Shapley effects. Some examples of dimension reduction techniques are presented in Section 2.4.

2.4 Dimension reduction techniques for functional outputs

Campbell et al. [2006] suggested that if model output is functional, sensitivity analysis should be carried out in two steps: first, to expand the functional output in terms of a set of basis functions and then, to perform sensitivity analysis on the coefficients of the

expansion. In general, dimension reduction consists into projecting the output matrix $\mathbf{Y} \in \mathbb{R}^{n \times N_s}$, composed by n discretized output curves in a grid of size N_s, in a basis of lower dimension $\Phi_{q \times N_s}$ with $q \leq N_s$:

$$Y - \bar{Y} \approx H\Phi^T,$$

where $\bar{Y} = n^{-1}\mathbf{1}\mathbf{1}^T Y$ with $\mathbf{1}$ a n-vector of ones, the columns of Φ codes the basis functions and, H the coefficients of the expansion. Thus, each curve $\ell \in \{1, \ldots, n\}$ can be approximated as:

$$y_\ell(s) - \bar{y}(s) \approx \sum_{k=1}^{q} h_{\ell k}\phi_k(s) \text{ for all } s \in \tau$$

Principal component analysis (PCA) is one of the most common techniques employed for dimension reduction. PCA consists in the spectral decomposition, i.e. eigenvalues and vector decomposition, of the variance-covariance matrix of the output $\mathbf{Y}$. Note that if the output is functional, functional PCA (fPCA) might be preferred (see Ramsay and Silverman [2005] and Yao et al. [2005]). Legendre polynomials, Haar functions or wavelet bases are other bases that could be used for dimension reduction [Campbell et al., 2006]. We choose to use PCA in Chapter 4 and fPCA in Chapter 5, whose use in GSA is common (e.g., Lamboni et al. [2009, 2011], Alexanderian et al. [2020]).

Part I

Model calibration using Bayesian inference

3

Bayesian calibration of an avalanche model from autocorrelated measurements along the flow: application to velocities extracted from photogrammetric images

María Belén Heredia, Nicolas Eckert, Clémentine Prieur, Emmanuel Thibert, *Bayesian calibration of an avalanche model from autocorrelated measurements along the flow: application to velocities extracted from photogrammetric images*

Contents

Abstract: Physically-based avalanche propagation models must still be locally calibrated to provide robust predictions, e.g. in long-term forecasting and subsequent risk assessment. Friction parameters cannot be measured directly and need to be estimated from observations. Rich and diverse data are now increasingly available from test-sites, but for measurements made along flow propagation, potential autocorrelation should be explicitly accounted for. To this aim, this work proposes a comprehensive Bayesian calibration and statistical model selection framework. As a proof of concept, the framework was applied to an avalanche sliding block model with the standard Voellmy friction law and high rate photogrammetric images. An avalanche released at the Lautaret test-site and a synthetic data set based on the avalanche are used to test the approach and to

illustrate its benefits. Results demonstrate (i) the efficiency of the proposed calibration scheme, and (ii) that including autocorrelation in the statistical modelling definitely improves the accuracy of both parameter estimation and velocity predictions. Our approach could be extended without loss of generality to the calibration of any avalanche dynamics model from any type of measurement stemming from the same avalanche flow.

3.1 Introduction

The complexity of snow avalanche physics is related to the variability and changing nature of snow [Schweizer et al., 2003, Ancey, 2006, Castebrunet et al., 2012, Steinkogler et al., 2014]. Evidence obtained at full-scale experimental slopes [Sovilla et al., 2008, Vriend et al., 2013, Prokop et al., 2015, Faug et al., 2018] shows a myriad of avalanche propagation and stopping regimes [Köhler et al., 2016], and numerical propagation models can reproduce these observations with increasing realism [Bartelt et al., 2016, Gaume et al., 2018]. However, knowledge concerning the mechanical behaviour of snow during motion and associated processes (granulation, erosion/deposition, etc.) remains incomplete [Steinkogler et al., 2015, Ancey, 2016, Truong et al., 2018]. From a macroscopic point of view, experimental approaches [Casassa et al., 1989, Rognon et al., 2008, Kern et al., 2009] and the proposal of Voellmy [1964] suggest rheological behaviours which remain ad-hoc. This renders on-site calibration on the basis of local data unavoidable [Ancey and Meunier, 2004, Salm, 2004, Eckert et al., 2012] to, e.g. predict high-return-period avalanches in land use planning and assess the related risk [Keylock et al., 1999, Meunier et al., 2004, Favier et al., 2014b,a]. This is all the more true given that studies have shown that avalanche propagation models are highly sensitive to their friction parameter values [see e.g., Borstad and McClung, 2009, Fischer, 2013].

After deterministic inversion methods had shown their limits [Dent and Lang, 1980, Dent et al., 1998, Ancey et al., 2003], and following progress made in many fields where accurate numerical model calibration is now recognised as a crucial issue [e.g., Oakley and O'Hagan, 2002, Carmassi et al., 2018], the Bayesian framework has become an appealing avenue in snow science over the past years, especially in the frequent case of small data samples [e.g., Ancey, 2005, Straub and Grêt-Regamey, 2006, Eckert et al., 2007, 2008b, 2009, 2010, Schläppy et al., 2014]. Specifically, Gauer et al. [2009] used a Bayesian framework to calibrate the friction parameters of three avalanche sliding block models and Fischer et al. [2014] proposed a method to evaluate simulations compared to Doppler radar observations. Ultimately, Naaim et al. [2013] could establish empirical links between friction parameters and physical properties of snow.

However, most of these existing approaches remain limited to rather coarse field data (e.g., samples of runout distances supplemented by input conditions), and when more comprehensive data sets have been considered, improper likelihood formulations have been used more often than not [Fischer et al., 2015]. For instance, little attention has been given so far to the specific difficulty induced by potential autocorrelation between the data used for calibration. Whereas this is not a matter for example, a sample of runout corresponding to distinct avalanche events, the assumption of independent observations is much more questionable in the current context of the increasingly diverse and rich measurements made on test-sites within the same avalanche. Many environmental applications have indeed demonstrated that neglecting potential autocorrelation between different measurements used in a calibration scheme can lead to biases in parameter esti-

mation and/or to lower predictive performances [see e.g., Kuczera, 1983, McInerney et al., 2017, Schaefli and Kavetski, 2017, Sun et al., 2017]. If this happens, both the physical interpretation of parameter estimates and the operational use of model predictions can be questioned.

On this basis, in this paper, we propose a Bayesian approach to calibrate the friction parameters of an avalanche propagation model using data of high temporal resolution (of the order of 1 s). Inspired by studies done for the calibration of hydrological models [e.g., Kuczera and Parent, 1998, Evin et al., 2014], the main novelty of our work is to explicitly account for potential autocorrelation between measurements made along the avalanche flow within the calibration framework. In what follows, we demonstrate that doing so results in a model better supported by the data that improves the accuracy of friction parameter estimates and of velocity predictions. Application is made on a well documented avalanche event from the Lautaret full-scale test-site [Thibert et al., 2015]. The used velocity data were obtained from high rate positioning from photogrammetric images. Such data were already applied by various authors to avalanche simulations performed with reference friction parameter values [e.g., Dreier et al., 2016, Gauer, 2014, Turnbull and Bartelt, 2003], but without including them within an explicit calibration scheme. That taking into account potential autocorrelation between measurements within the calibration is unavoidable to get unbiased estimates is further demonstrated with synthetic data analogue to the case-study.

In this work, as a proof of concept, we use the sliding block avalanche model, also known as 1-D Voellmy model instead of state of the art depth-averaged models [Naaim et al., 2004, Christen et al., 2010a, Bartelt et al., 2012]. We are aware of its limitations, notably that it makes it impossible to depict, e.g., flow depth variations in space and time. Also, some authors have shown that it may underestimate avalanche velocities [see e.g., Ancey and Meunier, 2004, Gauer, 2014]. However, for hazard mitigation, simple models with few parameters remain useful [Salm, 2004] and have the advantage to allow fast computation in comparison to more complex ones. A very simple avalanche propagation model is therefore a good choice for developing a calibration approach which could be in the future be applied to any other, more advanced, avalanche propagation model as soon as autocorrelation in measurements series is suspected.

3.2 Avalanche model calibration principle

3.2.1 Sliding block propagation model

Our model considers the avalanche as a rigid body sliding over a bidimensional curvilinear profile starting from the top of the path. The mass m and body shape variations of the avalanche are neglected. Under these assumptions, the motion equation of the avalanche mass center is:

$$\frac{du}{dt} = g \sin\theta - \frac{\mathrm{F}}{m}, \tag{3.1}$$

where $u = ||\vec{u}||$, $\frac{du}{dt}$ is the acceleration, g is the gravity constant, θ is the local slope angle and $\mathrm{F} = ||\vec{F}||$ is the frictional force. In this study, we consider the classical Voellmy friction law [Voellmy, 1964]. This means that the friction force is:

$$\mathrm{F} = \mu m g \cos\theta + \frac{mg}{\xi h} u^2, \tag{3.2}$$

depending on two friction parameters $\Theta = \{\mu, \xi\}$. Often it is assumed that μ evolves with the physical properties of snow whereas ξ may correspond to the geometry of the avalanche path and to terrain roughness [Ancey et al., 2003]. However, whereas this interpretation is sound or not is not our debate here. We do not make any further assumption regarding the linkages between (μ, ξ) and snow and topographical variables and simply search for the best couple on the basis of the data. The propagation model also depends on three forcing variables $x = \{T, h, x_{\text{start}}\}$ where T defines the topography of the terrain, h the mean flow depth of the avalanche and x_{start} the release abscissa (the protected runout length) of the avalanche mass center [Ancey and Meunier, 2004, Eckert et al., 2007].

3.2.2 Statistical model formulation

Let us denote f our avalanche propagation model. The model f predicts the avalanche speed (m s^{-1}) and position along the slope (m) at a time t (s). The model depends on parameters $\Theta \in \mathbb{R}^p$ and forcing variables $x \in \mathbb{R}^d$. The observed velocity, collected on the field, at time t is noted by v_t and we denote $v_{\text{obs}} = \{v_1, \dots v_n\}$ the set of observations where n denotes the number of observations.

The aim of model calibration is i) to find the optimal combination of parameters Θ that minimizes the discrepancy between the observations v_{obs} and the model simulation $f(\Theta, x)$ and ii) to rigorously quantify the associated uncertainty. To this end, we use the generic statistical model:

$$\mathcal{M} : v_t = f_t(x, \Theta) + \epsilon_t, \quad \forall t \in \{1, \dots n\}, \tag{3.3}$$

where $f_t(\Theta, x)$ denotes the simulation of the avalanche velocity at time t and ϵ_t is the model error. With this classical additive formulation, propagation model errors and observation errors are modelled altogether in the residuals ϵ_t.

In nearly all existing avalanche model calibration approaches, model errors are implicitly or explicitly assumed as independent and identically normally distributed (i.i.d.). In other words, $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, $\forall t \in \{1, \dots n\}$, where $\mathcal{N}$ denotes the Normal distribution and σ^2 is the common variance of the errors. However, this may be a too strong assumption for different measurements made along the same avalanche. For instance, the errors ϵ_t are likely to present a non-negligible correlation between two consecutive time steps (in other words, ϵ_t and ϵ_{t-1} may be correlated). To include the errors' autocorrelation in the calibration we propose to model them as an autoregressive (AR) process. Specifically, we consider an AR model of order 1 only (AR1). AR models of higher order could be used but this would imply the estimation of additional parameters, a non-trivial task in our case of a limited velocity sample. In addition, results obtained for the application indeed suggest that an AR(1) is sufficient to accurately account for the data variability (see Section 3.4). Thus, the errors ϵ_t can be expressed as:

$$\epsilon_t = \phi \epsilon_{t-1} + \eta_t, \quad \eta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \ \forall t \in \{1, \dots n\}, \tag{3.4}$$

where $\phi \in \mathbb{R}$ and η_t are the coefficient and innovations of the AR(1) process, respectively. If $|\phi| < 1$, $(\epsilon_t)_{t \in \mathbb{N}}$ is defined as the unique stationary solution of equation (3.4). In

Model	Errors	Parameters
$\mathcal{M}_0$	Independent and identically distributed $\mathcal{N}(0, \sigma^2)$	Θ, σ^2
$\mathcal{M}_1$	AR(1)	Θ, ϕ, σ^2

Table 3.1: Considered statistical models and corresponding parameters.

our study, $|\phi| < 1$ (see Section 3.4) that guarantees the stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$. Note also that with this model the innovations η_t are normally independent distributed but not the errors ϵ_t.

Hereafter, the model with the assumption of normally distributed and independent errors is denoted $\mathcal{M}_0$ and the model with AR(1) errors is denoted $\mathcal{M}_1$. Depending on the $\mathcal{M}_i$ model with $i \in \{0, 1\}$, the whole set of parameters to be estimated is different. Table 3.1 summarizes the parametrization of the competing statistical models considered. Hence, model $\mathcal{M}_1$ with $\phi = 0$ corresponds to $\mathcal{M}_0$.

3.2.3 Bayesian framework

The probability of the data can be maximised with respect to the model parameters [Fisher, 1922]. Bayesian statistics is an useful framework to estimate model parameters from scarce data. Within this approach, the quantification of uncertainty in parameter estimation is straightforward. In fact, the advantage of the Bayesian approach is that the uncertainty on parameters is assessed through credibility intervals by contrast to traditional methods (confidence intervals) [Bayes, 1763, Bernardo and Smith, 2009]. Hence, Bayesian statistics is now widely accepted as a reasonable option in environmental sciences [Berliner, 2003, Clark, 2005] and we use this framework in what follows.

For simplicity, let us denote γ, the set of parameters related to the errors, it means $\gamma = \{\sigma^2\}$ for $\mathcal{M}_0$ and $\gamma = \{\phi, \sigma^2\}$ for $\mathcal{M}_1$, and $x = \{T, h, x_{\text{start}}\}$ the forcing variables. Under a Bayesian framework, the joint posterior distribution of the parameters is the following one:

$$\pi(\Theta, \gamma | v_{\text{obs}}, x) \propto \mathcal{L}(v_{\text{obs}} | \Theta, \gamma, x)\pi(\Theta, \gamma), \tag{3.5}$$

where $\mathcal{L}(v_{\text{obs}}|\Theta, \gamma, x)$ is the likelihood of the observations and $\pi(\Theta, \gamma)$ is the joint prior distribution. Thus, the likelihood $\mathcal{L}(v_{\text{obs}}|\Theta, \gamma, x)$ is required. Under the $\mathcal{M}_0$ model, the observations follow a Gaussian distribution and the likelihood $\mathcal{L}(v_{\text{obs}}|\Theta, \gamma, x, \mathcal{M}_0)$ writes as:

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(v_t - f_t(\Theta, x))^2\right]. \tag{3.6}$$

Under the $\mathcal{M}_1$ model, the likelihood $\mathcal{L}(v_{\text{obs}}|\Theta, \gamma, x, \mathcal{M}_1)$ writes as (for more detail, see Appendix A):

$$\sqrt{\frac{1-\phi^2}{2\pi\sigma^2}} \exp\left[-\frac{1-\phi^2}{2\sigma^2}(v_1 - f_1(\Theta, x))^2\right] \times \frac{1}{(2\pi\sigma^2)^{(n-1)/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=2}^{n}\eta_t^2\right]. \tag{3.7}$$

Note that, in this study, we do not calibrate the quantities x_{start}, h, T because they were inferred from the data to put the effort on the friction law calibration. The sensitivity to these quantities in avalanche models has been studied in the works of e.g., Barbolini and Savi [2001], Borstad and McClung [2009], Buhler et al. [2018]. The authors found that changes in avalanche volume have a larger effect on both runout distance and avalanche velocity and that the friction coefficient μ (in a Coulomb model) is of high importance. However, this question should be analyzed deeper in a formal statistical framework and it is out of the scope of this work.

3.2.4 Metropolis-Hastings algorithm

The main practical difficulty in Bayesian inference is how to compute the normalizing constant in Bayes theorem. We overcome it by implementing a sequential Metropolis-Hastings algorithm, hereafter denoted MH [Metropolis et al., 1953, Hastings, 1970]. The MH algorithm proposes a generic way to construct a stationary and ergodic Markov chain that converges, under mild conditions, to the posterior distribution $\pi(\Theta, \gamma | v_{\text{obs}}, x)$. The Markov chain returned by the algorithm can be considered as a sample from $\pi(\Theta, \gamma | v_{\text{obs}}, x)$.

The following description of the MH algorithm is obtained from Robert [2015]. Let us denote, for simplicity, $\psi = \{\Theta, \gamma\}$ the set of the error and model parameters. For the application of the MH algorithm, it is needed an initial value $\psi^{(0)}$ and a proposal distribution q. Each iteration k of the algorithm consists in:

1. Generating $\psi' \sim q(.|\psi^{(k-1)})$.
2. Calculating $u \sim \mathcal{U}(0, 1)$.
3. Taking

$$\psi^{(k)} = \begin{cases} \psi' & \text{if} \quad u \leq \alpha, \\ \psi^{(k-1)} & \text{otherwise} \end{cases} \tag{3.8}$$

with

$$\alpha = \min\left(\frac{\mathcal{L}(v_{\text{obs}}|\psi', x)\pi(\psi')}{\mathcal{L}(v_{\text{obs}}|\psi^{(k-1)}, x)\pi(\psi^{(k-1)})} \frac{q(\psi^{(k-1)}|\psi')}{q(\psi'|\psi^{(k-1)})}, 1 \right) \tag{3.9}$$

where we recall that $\mathcal{L}(v_{\text{obs}}|\psi, x)$ and $\pi(\psi) = \pi(\Theta, \gamma)$ stand for the likelihood of velocity observations and the joint prior distribution, respectively. As mentioned in Robert [2015], the performance of the algorithm depends on the choice of q. For example, a standard choice for the proposal distribution q is a multinormal distribution centered in $\psi^{(k-1)}$ and with a given covariance Σ_q, which defines a random walk. For our application, Σ_q was tuned according to the optimal acceptance rates that grant fast convergence of the MH algorithm [Gelman and Rubin, 1992].

3.2.5 Model selection using Bayes Factor

To compare the accuracy of the two competing models $\mathcal{M}_0$ and $\mathcal{M}_1$, we use the Bayes factor. The Bayes factor is a criterion of the evidence provided by the data (in our case v_{obs}) to reject a model ($\mathcal{M}_0$) compared to another one ($\mathcal{M}_1$). The Bayes factor is defined as the ratio of marginal probabilities [see more details in Kass and Raftery, 1995]:

$$B_{10} = \frac{\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_1, x)}{\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_0, x)}, \tag{3.10}$$

where

$$\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_i, x) = \int \mathcal{L}(v_{\mathrm{obs}}|\psi_i, \mathcal{M}_i, x)\pi(\psi_i|\mathcal{M}_i, x)d\psi_i, \quad i \in \{0, 1\}. \tag{3.11}$$

The conditioning by $\mathcal{M}_i$ highlights that likelihood, prior and posterior distributions depend on the considered model. The Bayes factor can be estimated by applying importance sampling and using the MH sample drawn from the posterior density $\pi(\psi_i|v_{\mathrm{obs}}, x, \mathcal{M}_i)$ (see more details in Appendix B Section 3.7). An interpretation of the numerical value obtained to determine if there is evidence provided by the data to reject the model $\mathcal{M}_0$ compared to $\mathcal{M}_1$ according to the value of $\log_{10} B_{10}$ is proposed in the work of Kass and Raftery [1995], where $\log_{10}$ denotes the common logarithm. Notably, a $\log_{10} B_{10}$ value between 1 and 2 indicates strong evidence in favour of rejecting $\mathcal{M}_0$ compared to $\mathcal{M}_1$, and $\log_{10} B_{10} > 2$ a decisive evidence in favour of $\mathcal{M}_1$.

3.3 Application data

3.3.1 Avalanche data from the Lautaret test-site

The case study is an avalanche released at the Lautaret full-scale test-site [Thibert et al., 2015]. This test-site, located in the French Alps, holds a succession of avalanche paths. Here, the path referred as "path number 2" was used to artificially trigger an avalanche on 13 February 2013 (Figure 3.1). This path is 450 m long, dropping from 2400 to 2100 m a.s.l. (Figure 3.2). Its upper part part, where the acceleration of the flow occurs, is steep with an average inclination of 37° and a maximal slope around 45° in the starting area. This part of the path is steep-sided and around 10 m-wide, so that the flow is channelized. The lower part of the path is mainly the run-out zone, a large and naturally open slope. At the transition between the two parts, a road (Col du Galibier road, open in summer only) crosses the avalanche path, which generates a local slope rupture in the avalanche track.

The avalanche released on February 13 2013 was composed of a dense part with a limited saltation layer on top. Properties of the snow involved in the flow were characterized with a density, temperature and hardness profile of the released snow layer. Snow grain types and dimensions were also characterized. The avalanche was released at 11h58 AM when air temperature was about -10° Celsius. A 0.25 m thick layer of fragmented and decomposing snow particles was released. The mean density was 250 $kg.m^{-3}$, ranging between 270 and 225 $kg.m^{-3}$ from the bottom to the surface of the released snow cover. In the released layer, particles diameter is less than 0.5 mm. Snow temperature was between -4.7 and -5° Celcius. Hardness was "fist" (hand index) and measured as 20 N in Ram Resistance Equivalents [Fierz et al., 2009].

Avalanche front positions were determined using a high rate photogrammetric system that was specifically developed [Soruco et al., 2011, Thibert et al., 2015]. We used a low-cost non-metric imagery system (numerical reflex Nikon D2Xs cameras in DX format, CMOS sensor with Nikon 85 mm f/1.4 AF fixed focal lenses). An advanced ad hoc tuning was performed to account for the radial distortion of the lenses, the decentration

of the principal point (principal point shift) and the exact focal length of the lenses required for the correct scaling of the images [Faig et al., 1990]. The resulting error in positioning avalanche fronts was estimated to be less than 25 cm after image orientation on ground control points and a comparison of direct photogrammetric measurements to laser scanning on a test area. Synchronization between the two cameras was achieved within a precision of 6.10^{-6} s, therefore the error associated with the time sequence is also negligible. Eventually, terrestrial laser scanning was used to retrieve terrain elevation before and after avalanche triggering [Prokop, 2008, Prokop et al., 2015], so as to quantify the snow mass transfer by the avalanche (3.1). In Appendix D, Figure 3.10 shows some examples of images used in the photogrammetric process.

The sliding block model is a one dimensional model representing the successive positions of the center of mass of the avalanche, thus we used the following procedure to determine the velocity vector v_{obs}. At each time t, we computed the center of mass position of the avalanche using a set of front points $\mathcal{D}_t \in \mathbb{R}^3$ from the delineated front. $\mathcal{D}_t$ excludes the lateral front points because they are less active in the avalanche flow [Pulfer et al., 2013]:

$$c_t = \frac{\sum_{p_t \in \mathcal{D}_t} p_t}{n_t}, \tag{3.12}$$

where n_t is the number of elements in $\mathcal{D}_t$. Note that, from the data available, it is not possible to calculate with certainty the position of the mass center of each avalanche front. In this study, we estimated each of them using equation (3.12), arguably a rough estimation but sufficient for obtaining an approximation of the avalanche velocity consistent with the modelling framework we are using. Figure 3.1 shows the front of the avalanche and the center of mass at each five seconds. Then, the velocity at time t is estimated as the mean velocity of the center of mass between two successive images:

$$v_t = \frac{c_t - c_{t-1}}{\Delta t}. \tag{3.13}$$

In our application, $\Delta t = 1s$ and the data set is composed of 21 observations. We consider the first 21 seconds before the avalanche splits in sub-avalanches to ensure the validity of the application of a sliding block model. The mean flow depth h value was calculated as the mean of the difference between the snow depth of the digital elevation model taken after the release of the avalanche and the snow depth registered at the mass center locations during the avalanche motion. The value calculated was $h = 2.19m$ with a standard deviation of $0.78m$. This estimation is rough but it is a reasonable value given the characteristics of the avalanche studied.

The topography T was constructed as follows: the site under study has a simple geometry. It is a rather straight avalanche path starting with a cornice and of limited and rather constant width. The chosen 2-D topography is the main flow path starting from the cornice in the middle of the path (see more details in Thibert et al. [2015]). From a fine-scale Digital Elevation Model, it was extracted as a grid of horizontal resolution of $1.4m$. This is largely small enough to make the impact of the numerical approximation on the computed velocities negligible. Indeed, Bühler et al. [2011] showed that a spatial resolution of $25m$ is sufficient for avalanche modelling.

Figure 3.2 shows the topography T and the points where the fronts were recorded.

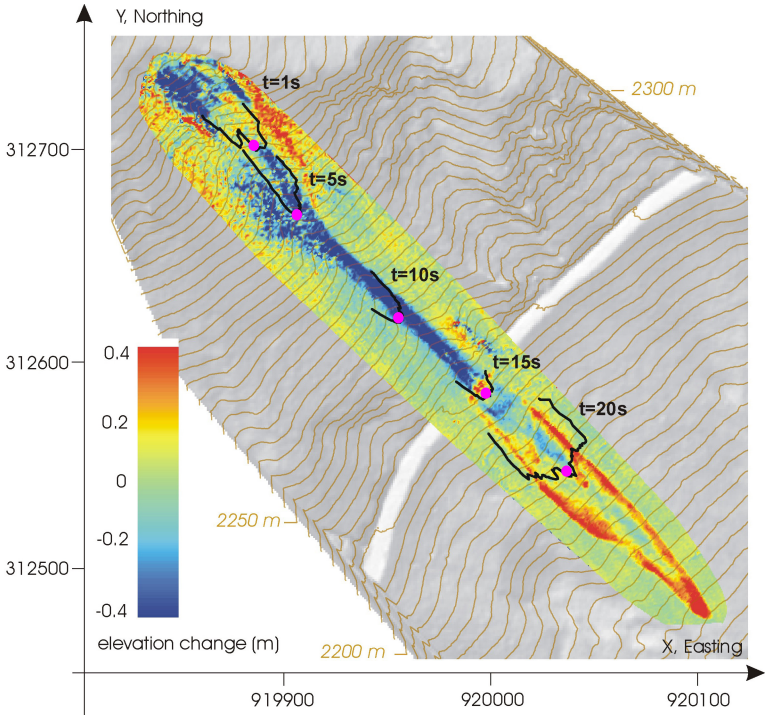

Figure 3.1: 3D-topography of path number 2 of the Lautaret test-site, some front positions (black lines) of the avalanche released on 13 February 2013 as determined from photogrammetric measurements and changes in snow depth before/after the avalanche as inferred from terrestrial laser scanning.

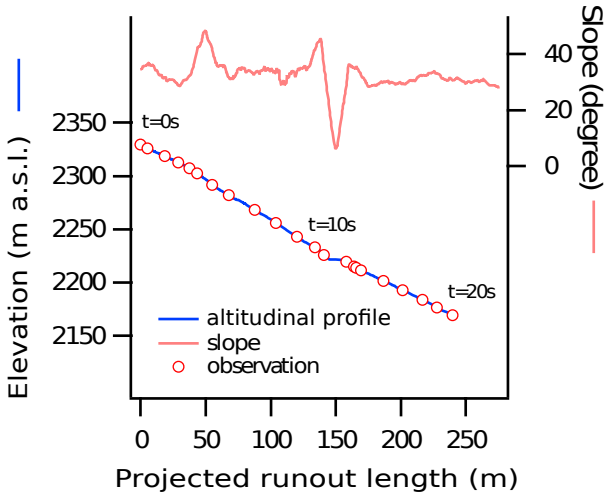

Figure 3.2: 2D-topography of the considered avalanche path (path number 2 of the Lautaret test-site), and position of the avalanche mass center each second for the studied avalanche released on 13 February 2013 (red circles).

Parameter	MAP estimate
μ_{MAP}	0.32
$\xi_{\text{MAP}}[m.s^{-2}]$	495.73
$\sigma^2_{\text{MAP}}[m^2.s^{-2}]$	8.07
ϕ_{MAP}	0.65

Table 3.2: Parameter values used for the generation of the synthetic data set. They correspond to the maximum (MAP) obtained with model $\mathcal{M}_1$ for the avalanche of the 13 February 2013.

3.3.2 Synthetic data generation

In order to further evaluate the accuracy of parameter estimation with the two models, we used synthetic data. This standard approach in statistical developments allows validation because "truth" (i.e. the parameter values used for the synthetic data generation) is known. Any parameter values could have been used at this stage. To resemble the most possible to the measured avalanche velocities, we chose the combination of parameter estimates corresponding to the measured case study. Also, to highlight that taking into account autocorrelation between measurements within the calibration is crucial to get unbiased estimates when such autocorrelation actually exists, our synthetic data were generated with model $\mathcal{M}_1$.

Specifically, synthetic velocity time series were generated from the maximum a posteriori (MAP) resulting from the application of our calibration approach to the avalanche described above with model $\mathcal{M}_1$. The maximum a posteriori is the mode of the posterior distribution, namely the most plausible value given the data (Table 3.2). In detail, after the MAP estimators of the μ, ξ, σ^2 and ϕ parameters of $\mathcal{M}_1$ model were determined for the avalanche, we proceed as follows:

1. An avalanche model simulation was conducted using the μ_{MAP} and ξ_{MAP} parameters. The result of this simulation is a velocity time series denoted by $f(\Theta_{\text{MAP}}, x)$.
2. An AR(1) error of parameters σ^2_{MAP} and ϕ_{MAP} was simulated.
3. The AR(1) error simulated was added to $f(\Theta_{\text{MAP}}, x)$. If negative velocity was obtained, the AR(1) error was resampled.

We generated 100 samples of synthetic data following this procedure and, for each of these, the 21 velocity values which correspond to the same location of the measurement were kept (to stay with the same data size as for the avalanche). Figure 3.3 shows some examples of the velocity samples generated versus the avalanche data.

3.4 Application, results and discussion

3.4.1 The avalanche

Prior distributions

Our Bayesian framework was used to calibrate the parameters of the $\mathcal{M}_0$ and $\mathcal{M}_1$ models (see Table 3.1). Marginal prior distributions used are described in Table 3.3. Parameters were assumed marginally independent a priori. This assumption is not strong because the

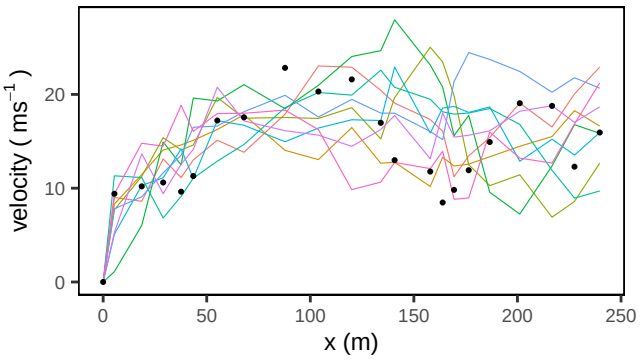

Figure 3.3: Samples of the generated synthetic longitudinal velocity profiles (colour lines) and observations corresponding to the studied avalanche (dots).

$\mu \sim \Gamma(9, 0.03)$	$\sigma^2 \sim \text{InvGamma}(3, 20)$
$\xi \sim \Gamma(25, 40)$	$\phi \sim \mathcal{U}(-1, 1)$

Table 3.3: Parameters marginal prior distributions. The prior distribution of ϕ applies only to $\mathcal{M}_1$ model. Γ represents the Gamma distribution and $\mathcal{U}$ the uniform distribution.

dependence of the parameters is reflected in their joint posterior distribution [Gilks et al., 1995, Eckert et al., 2010]. Informative marginal priors were used for all parameters but one, ϕ, for which a vague (poorly informative) prior was used instead, namely an uniform distribution $\mathcal{U}(-1, 1)$. This latter choice was done because prior knowledge for ϕ was unavailable. Instead, informative marginal prior distributions for other parameters could be determined on the basis of expert knowledge and well known reference values from Salm et al. [1990]. In addition, a comprehensive prior sensitivity analysis was conducted (see Appendix C in Section 3.8). It demonstrates that our results are highly robust to the prior choice so that this question is no longer further considered in what follows.

Convergence of the MH algorithm

For both models, we generated 50,000 MH samples and the last 25,000 were kept. The first 25,000 iterations were discarded as a burn-in period. This step avoids dependence on initial values. To further assess the convergence of the MH algorithm, we computed standard diagnoses [Kuczera and Parent, 1998, Torre et al., 2001, Robert and Casella, 2009, Eckert et al., 2010]. The R Core Team [2017] package `coda` created by Plummer et al. [2006] was used to calculate some of these. Specifically, three parallel chains were generated starting from different initial points and it was graphically checked that they were mixing well enough (not displayed). Also, the Kolmogorov-Smirnov test (H_0: the two samples where drawn from the same continuous distribution) was applied to the samples from the different chains and results showed that these were indeed drawn from the same distribution. In addition, for both models and all parameters, it was checked that acceptance rates (defined as the number of candidates accepted over the total number of iterations of the MH algorithm) were close enough to 0.25, the optimal value according to Robert [2015] to reach fast convergence in a random walk. Finally, the Gelman-Rubin convergence criterion was computed [Gelman and Rubin, 1992]. It is based on the difference between a weighted estimator of the variance and the variance of the estimators from

the different chains [Robert and Casella, 2009]. In our case, it equals 1 for all parameters, a perfect result. This all demonstrates that convergence was clearly reached for both models applied to the case study, leading to meaningful numerical approximations of the joint posterior distributions and, hence, reliable posterior estimates of the parameters of both models and of the Bayes factor.

Posterior distributions and posterior estimates

Figure 3.4 shows the parameter prior and posterior densities according to the MH samples. In Figure 3.4 panels **a)** to **c)** (resp., Figure 3.4 panels **d** to **g**), the $\mathcal{M}_0$ densities are shown (resp., $\mathcal{M}_1$) and Table 3.4 sums up the corresponding descriptive statistics. Eventually, posterior correlations between model parameters are shown in Figures 3.5 and 3.6 for $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively. For the two models, the marginal posterior distributions have lower variance than their priors, meaning that the observations have conveyed information into the analysis (see Figure 3.4), an expected result.

The variances of the posterior distribution of the friction parameters μ and ξ in the $\mathcal{M}_0$ model are lower than in $\mathcal{M}_1$ model. This result (arguably the sole undesirable one with $\mathcal{M}_1$ compared to $\mathcal{M}_0$) could be a consequence of the interactions between the friction coefficients and error parameters in the more parametrized $\mathcal{M}_1$ model (four free parameters instead of three with $\mathcal{M}_0$). Indeed, there is strong correlation between the parameters ξ and ϕ (0.49) for the model $\mathcal{M}_1$, which may preclude reaching sharp estimates of friction parameters with model $\mathcal{M}_1$. Conversely, even if there is a high correlation between the parameters μ and ξ with both models, switching from $\mathcal{M}_0$ to $\mathcal{M}_1$ reduces it from 0.85 to 0.50. The high correlation between the two parameters of the Voellmy friction law is known since the first calibration approaches of, e.g. Dent and Lang [1980]. This usually limits robust interpretation of obtained estimates, so that reducing it thanks to $\mathcal{M}_1$ should be seen as advantageous. This correlation reduction may be a collateral effect of having one more free parameters, allowing a bit more flexibility to fit the data. Similar effect is reflected in the lower correlation between σ^2 and the friction parameters with $\mathcal{M}_1$ than with $\mathcal{M}_0$. Note by the way that fairly assessing such correlations is a real strong point of our formal Bayesian calibration approach.

Remarkably, posterior estimates of friction parameters are very contrasted under both models. Even if posterior densities are not significantly different at the 95% credibility level (notably because of the higher a posteriori variance with model $\mathcal{M}_1$), differences in posterior estimates reaches 13% for μ (relative difference between 0.34 with $\mathcal{M}_1$ instead of 0.3 with $\mathcal{M}_0$), and 70% for ξ (relative difference between 696 $m^2.s^{-1}$ with model $\mathcal{M}_1$ and 410 $m^2.s^{-1}$ with model $\mathcal{M}_0$). This indicates that using either $\mathcal{M}_0$ or $\mathcal{M}_1$ makes a huge difference if one aims at interpreting the value of posterior estimates, for, e.g., relating avalanche friction characteristics to snow and topographical conditions.

Another remarkable result is the diminution of the error variance σ^2 from model $\mathcal{M}_0$ to model $\mathcal{M}_1$ (Table 3.4). Indeed, the errors variance σ^2 is higher if the autocorrelation ϕ is not included. Specifically, the standard deviation of the errors decreases from 3.6 to 3.1 $m.s^{-1}$ indicating that with model $\mathcal{M}_1$ velocity predictions may be seen as 13% more accurate (relative difference between both estimates), another desirable property for practical use in snow science. Eventually, the autocorrelation parameter ϕ is largely positive, with a posterior mean of 0.71. Also, its posterior 95% credibility interval whose lover bound is 0.33 firmly excludes the zero value corresponding to $\mathcal{M}_0$. This pleads for a high and significant autocorrelation between velocities along flow propagation.

Model	$\mathcal{M}_0$			$\mathcal{M}_1$			
	μ	$\xi[m.s^{-2}]$	$\sigma^2[m^2.s^{-2}]$	μ	$\xi[m.s^{-2}]$	$\sigma^2[m^2.s^{-2}]$	ϕ
Mean	0.3	409.74	12.71	0.34	696.12	9.79	0.71
sd.	0.07	101.12	3.53	0.08	305.71	2.62	0.17
$q_{0.025}$	0.16	264.30	7.66	0.18	306.97	5.93	0.33
$q_{0.50}$	0.3	393.34	12.17	0.34	627.36	9.33	0.73
$q_{0.975}$	0.43	649.64	20.98	0.50	1475.67	16.15	0.96

Table 3.4: Posterior distribution characteristics: posterior mean, posterior standard deviation (sd.), median (50% percentile) and 95% credibility interval (2.5% and 97.5% percentiles).

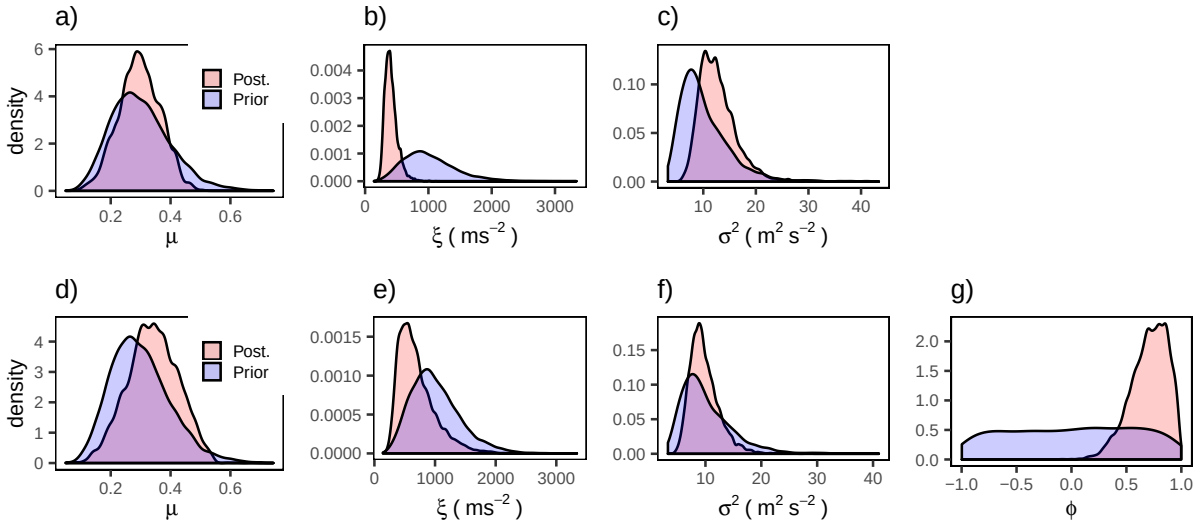

Figure 3.4: Prior and posterior densities of model parameters. Panels **a)** to **c)** model $\mathcal{M}_0$ and panels **d)** to **g)** model $\mathcal{M}_1$.

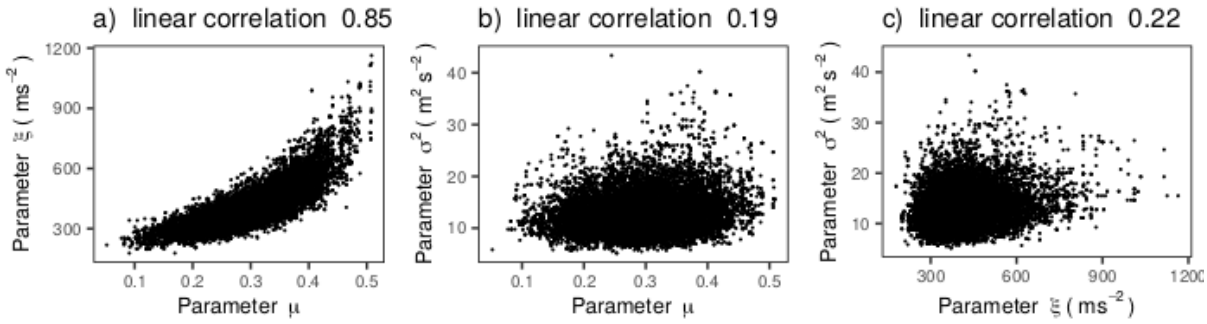

Figure 3.5: Joint posterior distribution of parameters of $\mathcal{M}_0$ model highlighting inter-parameter correlations.

Model posterior estimates

We calculated for model $\mathcal{M}_0$ the posterior estimate of Voellmy model simulations $f(\Theta, x)$ from the posterior estimates of parameters μ and ξ (Table 3.4). Resulting posterior estimates of model errors ϵ_t were analyzed by applying the Ljung-Box test (H_0: the data are independently distributed). We found a p-value lower than 0.05 indicating that there is a significant autocorrelation of model errors. In other words, $\text{cor}(\epsilon_t, \epsilon_{t-1})$ for all $t > 1$ is significant. Hence, one of the assumptions underlying statistical model $\mathcal{M}_0$ is not fulfilled by the data. On the other hand, applying standard statistical tests to posterior estimates of model errors shows no evidence of non-Gaussian errors (Shapiro test H_0: the sample is normally distributed, p-value > 0.05) or heteroscedasticity. The assumptions of Gaussian errors of common variance underlying statistical model $\mathcal{M}_0$ are thus fulfilled.

We proceeded similarly for statistical model $\mathcal{M}_1$, applying the same statistical tests to the posterior estimates of the innovations η_t of the AR(1) model. Remember that, in the $\mathcal{M}_1$ model, the independence assumption is on the innovations η_t and not on the errors ϵ_t. Test results show that innovations are indeed independent (p-value of the Ljung-Box $>$ 0.05), and normally distributed (p-value of Shapiro test > 0.05) with common variance. These results show that $\mathcal{M}_1$ represents correctly the autocorrelation of the errors, and, more widely, that contrary to model $\mathcal{M}_1$ all of its underlying assumption are fulfilled, which promotes its use from a strict statistical point of view.

Predictive velocity distributions

To analyze how the statistical model choice affects velocity estimation, we propagated parameter uncertainty up to model predictions. Two sets of posterior predictive simulations were performed, the first integrating over the posterior distribution of friction parameters only, leading the posterior predictive distribution $p(v_x|v_{\text{obs}}, x, \mathcal{M}_i)$ of the avalanche propagation model, and the second integrating over the distribution of both friction parameters and model error parameters leading the full posterior predictive distribution of avalanche velocities for the case study $p'(v_x|v_{\text{obs}}, x, \mathcal{M}_i)$. More precisely, the first writes:

$$\text{p}(v_x|v_{\text{obs}}, x, \mathcal{M}_i) = \int f(\Theta, x)\text{p}(\Theta|v_{\text{obs}}, x, \mathcal{M}_i)d\Theta, \tag{3.14}$$

and, the second one :

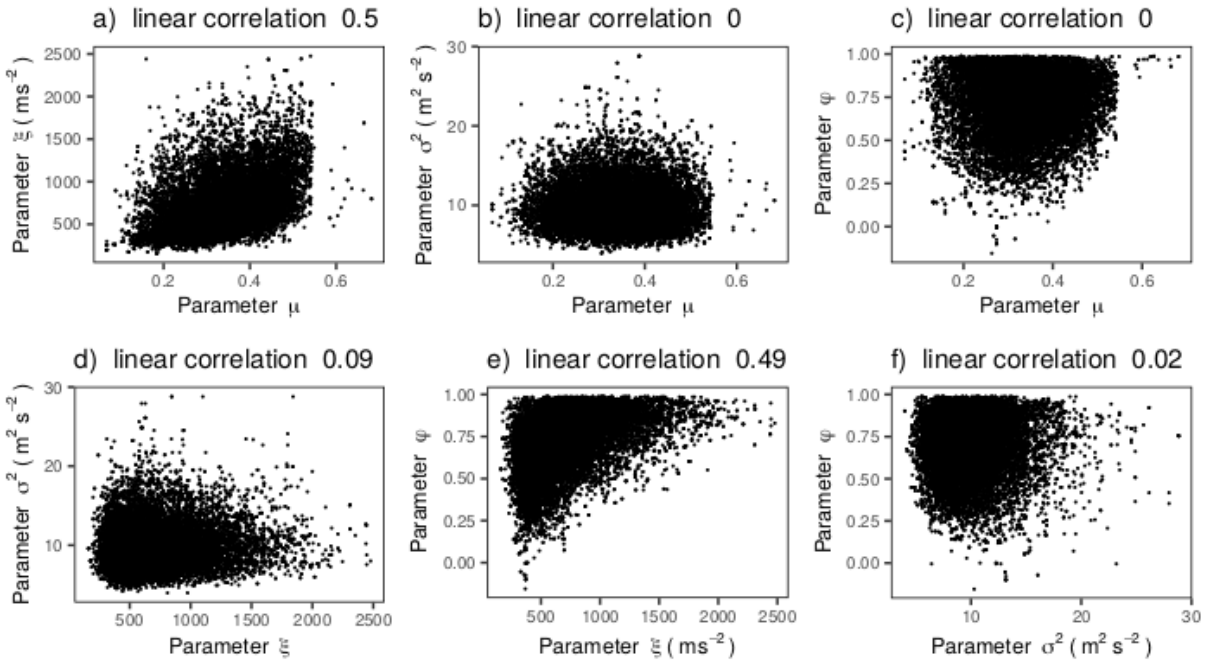

Figure 3.6: Joint posterior distribution of parameters of $\mathcal{M}_1$ model highlighting inter-parameter correlations.

$$\mathrm{p}'(v_x|v_{\mathrm{obs}}, x, \mathcal{M}_i) = \int (f(\Theta, x) + \epsilon)\mathrm{p}(\Theta, \gamma|v_{\mathrm{obs}}, x, \mathcal{M}_i)d\Theta d\gamma. \tag{3.15}$$

Eventually, to quantify the added value of parameter inference, the prior distribution of the velocity was also calculated by integrating over the prior distribution on the friction and model error parameters:

$$\pi(v_x|x, \mathcal{M}_i) = \int (f(\Theta, x) + \epsilon)\pi(\Theta, \gamma|\mathcal{M}_i)d\Theta d\gamma. \tag{3.16}$$

By using a Bayesian approach, we obtained a sample of parameters (μ, ξ, σ^2) for $\mathcal{M}_0$ (and also of ϕ for $\mathcal{M}_1$). Then, we computed avalanche model simulations using these two samples to obtain several curves of velocity. Finally, we drawn the percentile of each curve in Figure 3.7. Also, this figure shows the resulting 90% credibility intervals. For both models, comparison between prior and posterior credible intervals shows that predicted velocities are logically shifted towards observations. Uncertainty reduction (change in the width of the 90% credible intervals) is stronger with $\mathcal{M}_0$, but also exists with $\mathcal{M}_1$. This simply reflects the larger a posterior variance of model $\mathcal{M}_1$ parameters highlighted above.

In detail, almost all the elements of v_{obs}, except two with model $\mathcal{M}_0$ and one with model $\mathcal{M}_1$, are inside the 90% posterior credibility intervals drawn on $p'(v_x|v_{\mathrm{obs}}, x, \mathcal{M}_i)$ (blue dotted lines Figure 3.7), so that from this perspective the added value of $\mathcal{M}_1$ is not obvious. However, why $\mathcal{M}_1$ model is advantageous clearly appears if one focuses on the deterministic Voellmy model component (dotted orange lines in Figure 3.7): predictions come closer to observations and notably the velocity underestimation generally attributed to the Voellmy sliding block model [see e.g., Ancey and Meunier, 2004, Gauer, 2014] is reduced. Indeed, only 33% of the observations are inside the 90% posterior credibility

intervals of the sliding block model simulations $p(v_x|v_{\text{obs}}, x, \mathcal{M}_0)$, compared to 62% with $\mathcal{M}_1$ model.

Another quantification is provided by the systematic difference between both models that we evaluated through the mean difference and the mean quadratic difference between the medians of predicted velocities under $\mathcal{M}_0$ (Figure 3.7 panel **a)**, black line) and $\mathcal{M}_1$ (Figure 3.7 panel **b)**, black line). If we denote $\text{q50}_{v_x,\mathcal{M}_i}$ the median (50% percentile) of the posterior predictive distribution of velocities under model $\mathcal{M}_i$, the first writes:

$$\Delta_1 = \frac{1}{240}\int_0^{240} (q50_{v_x,\mathcal{M}_1} - q50_{v_x,\mathcal{M}_0})dx$$

and the second:

$$\Delta_2 = \sqrt{\frac{1}{240}\int_0^{240} (q50_{v_x,\mathcal{M}_1} - q50_{v_x,\mathcal{M}_0})^2 dx}$$

Obtained values are $\Delta_1 = 1.31$ m.s^{-1} and $\Delta_2 = 1.72$ m.s^{-1}. The positive value of the mean difference and the fact that it is not that much lower than the mean quadratic difference clearly shows that the inclusion of autocorrelation into the modelling truly leads to systematically higher velocity predictions. Predicted velocities are rather constant between $x = 50m$ and $x = 250m$ (plateau phase), around $15m.s^{-1}$ with $\mathcal{M}_0$. The underestimation with $\mathcal{M}_0$ with regards to $\mathcal{M}_1$ can be estimated by the ratio between Δ_2 and this velocity to 11.5%.

All in all, the analysis of predictive velocity distributions further confirms that model $\mathcal{M}_1$ should be preferred to get sharper, less underestimated, point estimates of avalanche velocities and to assess the related uncertainty fairly (credibility intervals lager but more likely to realistically describe the range of possible results).

Model selection

Finally, the logarithm of the Bayes Factor $\log_{10} \text{B}_{10}$ was calculated to compare the $\mathcal{M}_0$ model to the $\mathcal{M}_1$ model and we obtained a value of 2.02. According to the interpretation suggested in [Kass and Raftery, 1995, page 777] evidence, given by the data, to reject the $\mathcal{M}_0$ model compared to the $\mathcal{M}_1$ model is thus decisive. Indeed, $\log_{10} \text{B}_{10} > 2$ simply implies that given the data model $\mathcal{M}_1$ is more than 100 times more likely than model model $\mathcal{M}_0$.

3.4.2 Synthetic data

We then applied our Bayesian framework to calibrate the parameters of the 100 synthetic velocity profiles generated using the MAP estimators of model $\mathcal{M}_1$. We used the same marginal prior distributions as before (see Table 3.3). For each synthetic avalanche, we generated 20,000 MH samples and the last 10,000 were kept. Convergence was verified with the same diagnoses as for the measured data. We could then determine the ability of the $\mathcal{M}_0$ and $\mathcal{M}_1$ models to retrieve the true model parameter values used for the synthetic data generation. For this, we calculated the 90% coverage rates for the different parameters, this means, the number of 90% credibility intervals recovering the true value of Table 3.2. According to the results of Table 3.5, $\mathcal{M}_1$ model has a much better ability to determine the true model parameter values compared to the $\mathcal{M}_0$ model. This especially

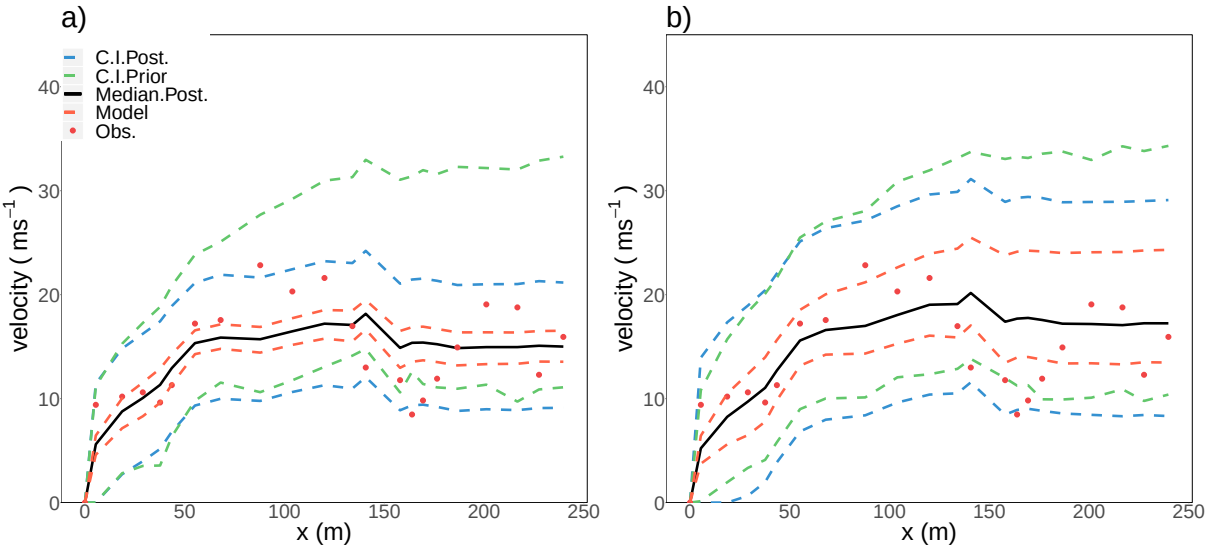

Figure 3.7: Predictive velocity distributions versus data: **a)** model $\mathcal{M}_0$ and **b)** model $\mathcal{M}_1$. 90% posterior credibility intervals (CI) are computed according to Equations (3.14) and (3.15) for the Voellmy propagation model (orange dotted lines), and the complete propagation and error model (blue dotted lines), respectively. Black plain line denotes the posterior median of the complete model. 90% prior credibility intervals are computed according to Eq. (3.16) (green dotted lines). Observations used for calibration appear as red points.

	$\mathcal{M}_0$			$\mathcal{M}_1$			
Parameter	μ	ξ	σ^2	μ	ξ	σ^2	ϕ
Coverage rate	64	84	0	98	82	98	93

Table 3.5: The 90% coverage rates for the synthetic data sample. For each parameter and both models, the 90% coverage rate corresponds to the number of times over the sample of 100 synthetic avalanches for which the 90% posterior credibility interval includes the true value used for data generation. In other words, 90% is the perfect validation score. Parameter ϕ applies to model $\mathcal{M}_1$ only.

applies to σ^2 for which $\mathcal{M}_0$ model is fully unable to identify the true value and, to a lower extent, to μ, for which the true value is in the posterior 90% credibility interval less than two times over three. Conversely, 90% coverage rates with $\mathcal{M}_1$ model are rather fair, varying between 82% and 98% for all the four parameters.

To further compare the precision of the estimates led by the two models, the MAP estimators corresponding to the full synthetic sample are presented in Figure 3.8 in panels from **a)** to **d)**. In mean, both models result in an underestimation of the μ parameter, but the underestimation with $\mathcal{M}_1$ is much smaller than with $\mathcal{M}_0$. In addition, the true value is well within the range of variability of the different estimates corresponding to the 100 synthetic avalanches with $\mathcal{M}_1$ whereas it is clearly outside with $\mathcal{M}_0$. Similarly, parameter σ^2 is much better estimated by $\mathcal{M}_1$ model. Parameter ξ is slightly underestimated by $\mathcal{M}_0$ model and slightly overestimated by $\mathcal{M}_1$ model, but both models perform reasonably. Finally, $\mathcal{M}_1$ model estimates correctly the autocorrelation parameter ϕ. Overall, this analysis confirms that only $\mathcal{M}_1$ model is able to retrieve the true parameter values as soon as autocorrelation between measurements actually exists. In other words, not accounting

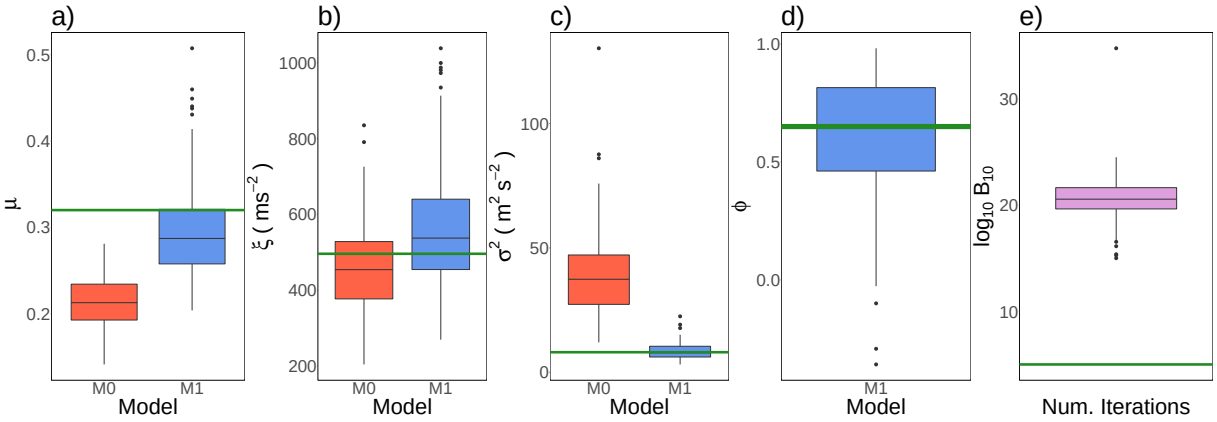

Figure 3.8: Panels **a)** to **d)**: Boxplots of the MAP estimators under models $\mathcal{M}_0$ (red colour) and $\mathcal{M}_1$ (blue colour) corresponding to the 100 synthetic avalanches. The true parameter values used for synthetic data generation are shown with a green colour line. Panel **e)** Boxplot of $\log_{10} B_{10}$ obtained for the 100 synthetic avalanches. The value of 2 in green corresponds to the reference value of Kass and Raftery [1995] above which evidence in favour of $\mathcal{M}_1$ is decisive.

for autocorrelation within model calibration when autocorrelation actually exists carries high risk to lead to biased estimates. This provides another very strong argument in favour of $\mathcal{M}_1$ since, as evidenced by the case study, significant autocorrelation may indeed exist between measurements made along the same avalanche flow.

Eventually, the common logarithm of the Bayes factor calculated for each synthetic avalanche (see Figure 3.8 panel **e)** indicates for all synthetic velocity series decisive evidence to reject the $\mathcal{M}_0$ model compared to the $\mathcal{M}_1$ model. This result is all the more logical given that the Bayes factor asymptotically selects the true model when it is included within a sample of competing models. However, this can be seen as a last strong point to advise using model $\mathcal{M}_1$ as soon as autocorrelation is suspected.

3.5 Conclusion and outlook

In this work, a Bayesian calibration of an avalanche flow dynamics model from data of high temporal resolution [1s] was developed. The objective of this work was to show how potential autocorrelation between measurements made along an avalanche flow can be considered within the calibration of an avalanche model, and to demonstrate that this improves the accuracy of friction parameters estimation and velocity predictions.

Two statistical models representing the discrepancies between observations and simulations were proposed: the first one, $\mathcal{M}_0$, classically considered the errors as independent and identically normally distributed, and the second one, $\mathcal{M}_1$, modelled the errors as an autoregressive process of order 1. The latter accounts for potential autocorrelation between measures made along the same avalanche flow, a question which has been poorly addressed in the snow science literature so far. Our objective was to determine the accuracy of parameter estimation when the autocorrelation is included or not in the modelling. More generally, we wanted to develop a framework able to link in a more mathematically consistent way, the rich and diverse data now increasingly available from avalanche test-sites with numerical propagation models.

Application was made on a well documented avalanche event and on synthetic data. The avalanche was released at the Lautaret full-scale test-site on February 13 2013 and the corresponding velocity time series was obtained from high rate positioning photogrammetry. A synthetic data set of 100 velocities time series mimicking the avalanche was created in order to further test our approach and to illustrate its benefits. Results for the avalanche showed strong and significant autocorrelation between predicted velocities. It also appeared that the velocity time series was correctly modelled by the statistical model $\mathcal{M}_1$ only (Ljung-Box test p-values > 0.05) so that there was decisive evidence to reject $\mathcal{M}_0$ compared to $\mathcal{M}_1$. In addition, resulting posterior estimates for friction parameters μ and ξ and velocities along the path were shown to be different and arguably more realistic (for the estimated velocities) with $\mathcal{M}_1$ than those obtained with $\mathcal{M}_0$. Eventually, our synthetic data confirmed that in presence of autocorrelation between measurements, not accounting for it may lead to biased estimates. This all demonstrates the necessity and usefulness to explicitly account for autocorrelation within the calibration, both to realistically predict avalanche characteristics and to get friction parameter estimates that can be related to snow and topographical characteristics in a meaningful way, two points of crucial importance in snow science. Hence, since avalanche velocities are time-series, as a good practice of modelling, potential autocorrelation of the errors should always be envisaged, since we clearly demonstrate that neglecting this effect may lead to undesirable consequences.

To develop and illustrate our approach, we chose the simple sliding block model with the Voellmy friction law because it is faster to run than more computationally intensive state of the art avalanche models. As it was mentioned in the Introduction, we used this simple model as a proof of concept but our approach could be generalized to other friction laws (e.g., the Coulomb one which has the advantage of having a single parameter (μ)), and to other ways of describing avalanche flows (e.g., depth-averaged equations or a sliding block with a PCM formulation, which would avoid the estimation of the flow depth) after suitable adaptation of the algorithm. Also, on the existing basis, it should be possible to include within the calibration other quantities which can be measured simultaneously as for example, the snow depth in the release area, the runout distance, the deposited volume, etc. This would better constrain the calibration, leading to potentially less uncertain posterior estimates. In such a case, an appropriate likelihood would need to be proposed and a higher number of observations would be required to calibrate the new likelihood parameters. To overcome this shortcoming using several avalanches altogether could be an option which was out of the scope of this work.

One of the main drawbacks of a Bayesian approach is the high number of simulations that are needed (at least 1000 simulations) to reach convergence in a Markov Chain Monte Carlo setting. Also, a compromise between the number of parameters that can be estimated and the size of the data must be found. Hence, whether or not our approach can be practically implemented with the most computationally intensive avalanche model and how many unknown parameters can be inferred as function of the available data sets remain questions to be investigated.

Finally, it is worth to mention that the avalanche size and velocities we found are typical of a medium size avalanche (class 3 on the CAA international avalanche scale). Specifically, friction parameter values correspond to the range of values that can be found in the literature. In future work, it could therefore be informative to apply our approach to a much large sample of avalanches to exploit its potential for inferring relevant physics

(e.g., to find the relationship between friction parameters and snow conditions). Ultimately, this all should i) lead to a more accurate evaluation of the highest-return-period events required for avalanche risk assessment, and ii) simultaneously improve our knowledge of the relevant physics by providing sharper quantification of underlying processes.

3.6 Appendix A: Likelihood of an AR(1) process

Since $|\phi| < 1$, the ϵ_t error has a representation in terms of the innovations η_t (also known as MA(∞) representation, see Gouriéroux and Monfort [1995] for more details):

$$\epsilon_t = \phi^t \epsilon_0 + \sum_{i=0}^{t-1} \phi^i \eta_{t-i}, \tag{3.17}$$

where we recall that ϕ and η_{t-i} are the coefficient and the innovations of the AR(1) process, respectively (see equation 3.4).

From the last equation, the unique stationary solution is obtained for ϵ_0 with zero mean and variance equal to $\frac{\sigma^2}{1-\phi^2}$, independent from $(\eta_1, \ldots, \eta_t)$, and we get that:

$$\mathrm{E}(\epsilon_t) = \mathrm{E}(\epsilon_0) + \sum_{i=0}^{t-1} \phi^i \mathrm{E}(\eta_{t-i}) \quad = \quad 0, \tag{3.18}$$

$$\mathrm{Var}(\epsilon_t) = \phi^{2t}\mathrm{Var}(\epsilon_0) + \sum_{i=0}^{t-1} \phi^{2i} \mathrm{Var}(\eta_{t-i}) \quad = \quad \frac{\sigma^2}{(1-\phi^2)}. \tag{3.19}$$

The marginal distribution of ϵ_t is Gaussian because it is the sum of independent and identically normally distributed η_t. In particular the first error term ϵ_1 is Gaussian:

$$\epsilon_1 \sim \mathcal{N}\left(0, \frac{\sigma^2}{(1-\phi^2)}\right). \tag{3.20}$$

Considering that $\epsilon_2 = \phi\epsilon_1 + \eta_2$, the conditional distribution of ϵ_2 given ϵ_1 is $\epsilon_2|\epsilon_1 \sim \mathcal{N}(\phi\epsilon_1, \sigma^2)$. In a more general way:

$$\epsilon_t|\epsilon_{t-1} \sim \mathcal{N}(\phi\epsilon_{t-1}, \sigma^2). \tag{3.21}$$

Then, the joint distribution of errors is:

$$\mathrm{p}(\epsilon_1, \cdots, \epsilon_n|\phi, \sigma^2) = \mathrm{p}(\epsilon_1) \prod_{i=2}^{n} \mathrm{p}(\epsilon_i|\epsilon_{i-1}). \tag{3.22}$$

From this equation, the likelihood expression of equation (3.7) is obtained.

3.7 Appendix B: Numerical evaluation of the Bayes Factor

Thanks to the Bayes theorem, we can write:

$$\frac{1}{\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_i,x)} = \frac{\pi(\psi_i|v_{\mathrm{obs}},\mathcal{M}_i,x)}{\mathcal{L}(v_{\mathrm{obs}}|\psi_i,\mathcal{M}_i,x)\pi(\psi_i|\mathcal{M}_i,x)}, \tag{3.23}$$

where $i \in \{0,1\}$.

If g is a density function defined on an ensemble Ω:

$$\begin{aligned}\frac{1}{\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_i,x)} &= \frac{1}{\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_i,x)}\int_{\psi_i\in\Omega} g(\psi_i)d\psi_i \\ &= \int_{\psi_i\in\Omega} \frac{g(\psi_i)\pi(\psi_i|v_{\mathrm{obs}},\mathcal{M}_i,x)}{\mathcal{L}(v_{\mathrm{obs}}|\psi_i,\mathcal{M}_i,x)\pi(\psi_i|\mathcal{M}_i,x)}d\psi_i.\end{aligned}$$

From Monte Carlo simulations, $\frac{1}{\mathrm{p}(v_{\mathrm{obs}}|\mathcal{M}_i,x)}$ can thus be estimated as:

$$\frac{1}{p(v_{\mathrm{obs}}|\mathcal{M}_i,x)} = \frac{1}{N}\sum_{j=1}^{N}\frac{g(\psi_i^{(j)})}{\mathcal{L}(v_{\mathrm{obs}}|\psi_i^{(j)},\mathcal{M}_i,x)\pi(\psi_i^{(j)}|\mathcal{M}_i,x)}. \tag{3.24}$$

where $\{\psi_i^{(j)}; j = 1,\cdots,N\}$ is a sample from the posterior distribution of $\mathcal{M}_i$ (see Kass and Raftery [1993] page 19 equation 12).

For this application, the g function was chosen as a multinormal distribution with mean equal to the empirical mean and covariance matrix equal to the estimated covariance matrix of our Metropolis-Hastings sample, respectively.

3.8 Appendix C: Prior sensitivity analysis

To study the robustness of our results, we conducted a prior sensitivity analysis by varying prior information for model parameters (μ,ξ,σ^2). We explored the range of priors from the marginal informative priors used in the paper core up to vague (poorly informative) priors as much as possible. In the case of $\mathcal{M}_0$ model, marginal vague priors could be tested for all the three parameters instead of informative ones. For $\mathcal{M}_1$ model, however, the informative prior should be always kept for ξ, because, if for this parameter a vague prior was used, the MH algorithm convergence could not be achieved. This is explained by the high correlation between the ξ and ϕ parameters (close to 0.5, see Figure 3.6). For the ϕ parameter, a uniform distribution $\mathcal{U}(-1,1)$ was used in all the analysis. The different combinations of informative and non-informative priors tested are shown in Table 3.6. Corresponding posterior distributions are shown in Figure 3.9.

Even if slight differences between the different posteriors can be noted as one replaces one prior by another, overall, for both models, results remain quite similar, leading in all cases to posterior estimates very close to the ones obtained with the informative priors used in the paper core (see Table 3.4). Also, the common logarithm of the Bayes Factors was calculated between all combinations of $\mathcal{M}_1$ and $\mathcal{M}_0$ posterior samples and, in nearly all cases, there is strong to decisive evidence to reject $\mathcal{M}_0$ compared to $\mathcal{M}_1$ model, as demonstrated in the paper core with informative priors. Indeed, the obtained $\log_{10}(\mathrm{B}_{10})$ values are between 0.57 and 2.49 with a mean of 1.38 and a standard deviation of 0.61. Hence, all in all, posterior inference and model selection show little sensitivity to prior specification. Since it is advised to use priors with a finite domain as much as possible to avoid spurious results [Kuczera and Parent, 1998] we kept the informative priors in the paper core.

Table 3.6: Marginal prior distributions used for the sensitivity analysis. Priors 3 and 5 are used with $\mathcal{M}_0$ model only, and prior 6 with $\mathcal{M}_1$ model only. Γ denotes the Gamma distribution and InvΓ denotes the inverse Gamma distribution. Prior1 is the one used in the paper core. "-" denotes vague marginal priors.

Parameter	μ	ξ	σ^2
Prior1	$\Gamma(9, 0.03)$	$\Gamma(25, 40)$	Inv$\Gamma(3,\ 20)$
Prior2	-	$\Gamma(25, 40)$	Inv$\Gamma(3,\ 20)$
Prior3	$\Gamma(9, 0.03)$	-	Inv$\Gamma(3,\ 20)$
Prior4	$\Gamma(9, 0.03)$	$\Gamma(25, 40)$	-
Prior5	-	-	-
Prior6	-	Inv$\Gamma(3,\ 20)$	-

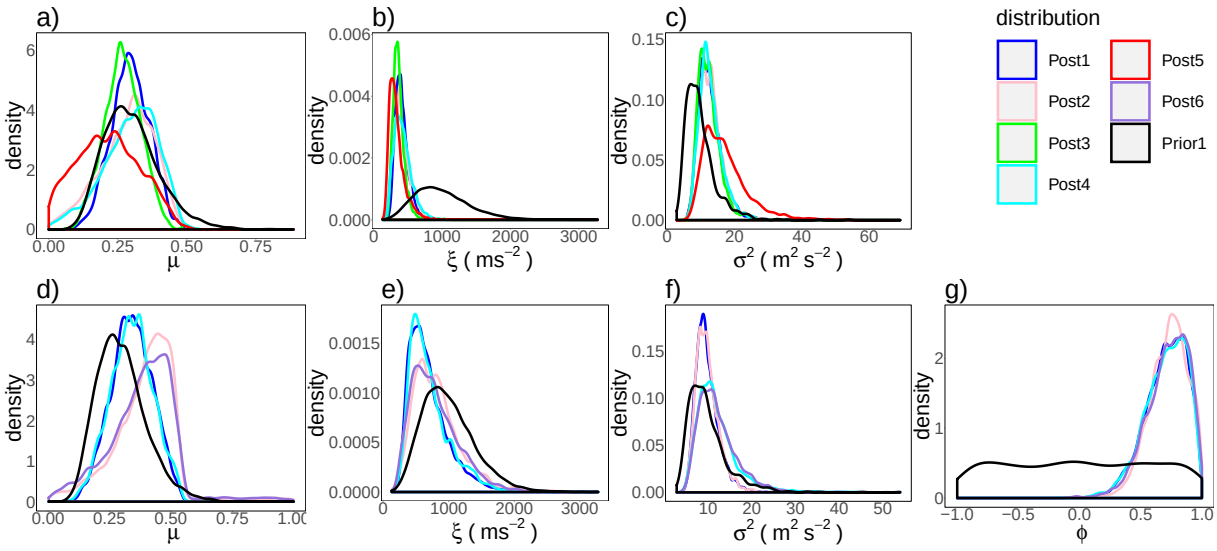

Figure 3.9: Prior sensitivity analysis. Postk, where k ranges for 1 to 6, are the posterior distributions obtained with the priors 1 to 6 of Table 3.6. Prior1 is the informative prior used in the paper core. Panels **a)** to **c)**: $\mathcal{M}_0$ model; panels **d)** to **g)**: $\mathcal{M}_1$ model.

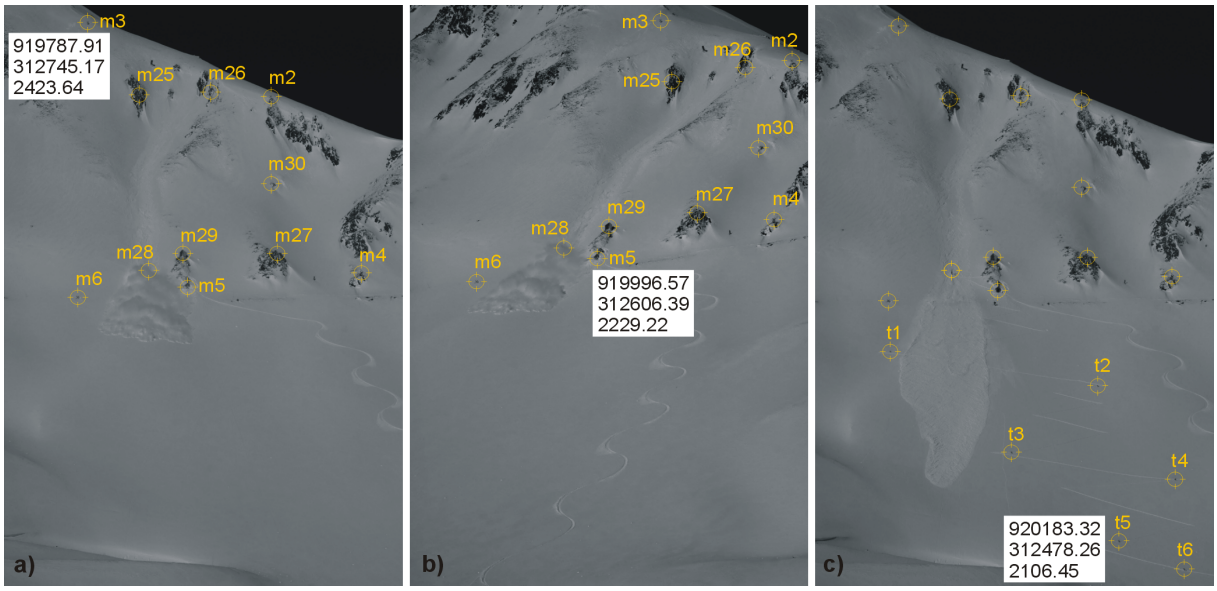

Figure 3.10: Images used in the photogrammetric process. **a)** and **b)** are the left and right images for the avalanche released on 13 February 2013 for t=20s after triggering. This couple is used for restitution and to map the location of the avalanche head. Markers plot the permanent ground control points used for image orientation. To illustrate the spatial extension, the coordinates and elevations (in meters) are indicated for the upper and lower control points $(m3, t6)$, and for control point $m5$ at location of Col du Galibier road. The right image **c)** has been taken after the avalanche stops. Temporary control points $t1$ to $t6$ are setup after the avalanche in the runout area to improve image orientation in this area according to the avalanche deposit.

3.9 Appendix D: Images used in the photogrammetric process

In this appendix, some examples of images used in the photogrammetric process are shown.

Part II

Global sensitivity analysis

4

Nonparametric estimation of aggregated Sobol' indices: application to a depth averaged snow avalanche model

María Belén Heredia, Clémentine Prieur, Nicolas Eckert, *Nonparametric estimation of aggregated Sobol' indices: application to a depth averaged snow avalanche model*

Contents

Abstract: Avalanche models are increasingly employed for elaborating land-use maps and designing defense structures, but they rely on poorly known parameters. Careful uncertainty assessment is thus required but difficulty arises from the nature of the outputs of these models, which are commonly both functional and scalar. Hence, so far in the avalanche field, few sensitivity analyses have been performed. In this work, we propose to determine the most influential inputs of an avalanche model by estimating aggregated first-order Sobol' indices. We propose a nonparametric estimation procedure based on the Nadaraya-Watson kernel smoother, which allows to estimate the aggregated Sobol' indices from a given random sample of small to moderate size. Due to the limited size of the sample, the kernel estimation is biased. Therefore, we propose a bootstrap based bias correction before selecting the bandwidth by cross-validation. To estimate the aggregated Sobol' indices, we reduce the dimension of the output using principal components analysis. After different test cases showing the efficiency of our approach, it is applied to a real avalanche case. Results show that the friction parameters and the snow depth in the release zone are the most influential parameters determining the avalanche characteristics.

4.1 Introduction

In mountain environments, snow avalanches are a constant threat for settlements and their inhabitants [McClung and Schaerer, 2006, Ancey, 2006]. This creates conflict between development and, safety and land-use planning is an efficient way to reduce death tolls [Gruber and Margreth, 2001, Eckert et al., 2018a]. To this aim, avalanche flow dynamic models are increasingly employed for elaborating land-use maps and designing defense structures [Naaim et al., 2010, Favier et al., 2014b]. However, all existing physics-based snow avalanche models remain based on some parameters which are poorly known [Ancey and Meunier, 2004, Eckert et al., 2008b]. This applies, for instance, to friction parameters representing the behavior of snow in motion and to initial conditions corresponding to the avalanche release [Jamieson et al., 2008, Eckert et al., 2010, Naaim et al., 2013]. Thus, a careful assessment of the impact of the uncertainty of the inputs on the outputs should be carried out [Fischer, 2013]. Difficulty arises from the nature of the outputs of these models, which are commonly both functional (e.g., the velocity and flow depth of the avalanche as function of space and time) and scalar (e.g., the runout distance, aka the point of further reach on the 2D topography of the avalanche flow path).

One of the main purposes of sensitivity analysis is to determine the inputs which are the most influential on the output or outputs of a model where the model can be of any nature, for example a black-box model or a complex computational code [Saltelli et al., 2000]. More specifically, Saltelli et al. [2004] proposed four settings as a guide of objectives of a sensitivity analysis: factor prioritization, factor fixing, variance cutting and factor mapping [Saltelli et al., 2004, Iooss and Prieur, 2017]. It depends on the analyst to define the objectives of its sensitivity analysis. In the global sensitivity (GSA) framework, the input parameters are modeled by random variables, which will be assumed independent in this paper. The probability distribution chosen to model the input vector is often guided by practitioner's belief. The output is then random as it depends on the inputs through the model. In this work, our objective is factor prioritization which consists in identifying which inputs or factors once fixed would reduce the variance of the output [Saltelli et al., 2008] at most.

There exist different sensitivity measures in the literature for quantifying the influence of each input on the output: variance based indices, also known as Sobol' indices [Sobol', 1993], density based measures [Borgonovo et al., 2016], entropy-based sensitivity measures [Auder and Iooss, 2008], etc. A detailed review of sensitivity measures can be found, e.g., in Iooss and Lemaître [2015] and Borgonovo and Plischke [2016].

We focus our study on variance based first-order Sobol' indices [Sobol', 1993] commonly used for factor prioritization. In short, Sobol' indices are constructed as follows: if the output of a model is scalar, its total variance can be split into partial variances by using the Hoeffding decomposition [Hoeffding, 1948]. If the inputs are independent, each of the partial variances is associated to an input or to an interaction between inputs through the ANOVA decomposition [Efron and Stein, 1981]. Then, the Sobol' indices [Sobol', 1993] are calculated as the ratio of each partial variance to the total variance. Moreover, Sobol' indices generalize the linear correlation coefficient R^2 for any kind of models.

Sobol' indices are easy to interpret: a high index value shows the input is relevant and a value close to zero shows that it is not, all the indices are normalized between 0 and 1 and sum to 1. Moreover, Sobol' indices have a natural extension to multivariate or functional outputs [see, e.g., Lamboni et al., 2009, Gamboa et al., 2013] which is the

framework of the application that motivated this study.

If the output is multivariate, it is possible to apply sensitivity analysis to each component of the multivariate output but this could lead to redundancies in the results, particularly in the setting of discretized functional outputs. It seems therefore interesting to turn to the so-called aggregated indices first introduced in Lamboni et al. [2009] which summarize the information. A preliminary step of output dimension reduction such as, e.g., principal component analysis (PCA) or partial least-squares may also be applied first [see, e.g., Campbell et al., 2006]. For our real application in the avalanche field, dimension reduction was performed by simultaneous PCA introduced in Ramsay and Silverman [2005] [see also Nanty et al., 2017, and references therein].

We now discuss the estimation of aggregated Sobol' indices. There are many methods to estimate Sobol' indices: the Fast Amplitude Sensitivity Test (FAST) [see, e.g., Saltelli et al., 1999, and references therein], Random Balance Design [Tarantola et al., 2006], Sobol' pick-freeze schemes [Sobol', 1993]. The main drawback of all the aforementioned methods [Plischke, 2010] is that they are based on sampling designs of particular type. Nevertheless, for many memory and time consuming real applications, the cost (in terms of number of model evaluations) of these approaches is prohibitive. For example, to estimate a single Sobol' index with an uncertainty of 10%, it could be required to perform 10^4 model runs [Iooss and Lemaître, 2015].

To overcome the drawback of the aforementioned methods, authors use approaches not based on structured sampling designs to estimate the indices. These methods are known as given data or one sample methods. They also correspond to *green* sensitivity analysis because available data from previous model runs can be reused. Among these approaches, we can cite the effective algorithm for computing global sensitivity indices (EASI) method proposed by Plischke [2010], which is a spectral method based on the Fast Fourier Transform. We also mention the work in Plischke et al. [2013], which relies on the notion of class-conditional densities, where a class is a sub-sample stemming from a suitable partition of the dataset. Much more recently, Antoniano-Villalobos et al. [2019] proposed a fully Bayesian given data procedure.

In the present paper, we propose a new given data method to estimate first-order Sobol' indices, based on a nonparametric Nadaraya-Watson [Nadaraya, 1964, Watson, 1964] bias corrected kernel regression [Racine, 2001]. Our approach is close to the one introduced in Solís [2019]. The main difference is in the formula we apply to correct the bias, which seems to be more efficient, at least in the results of our simulation study. We then extend this estimation procedure to the estimation of the aggregated Sobol' indices [Lamboni et al., 2009]. To our knowledge, it is the first time a nonparametric given data procedure is proposed in this framework. It must be noted that the estimation of higher-order and total Sobol' indices may not be possible using kernel estimation with moderate sample sizes due to the curse of dimensionality Da Veiga et al. [2009], Hence, the estimation method proposed in this work focuses on aggregated first-order Sobol' indices. Kernel regression with bandwidth selection via cross-validation was already used in Doksum and Samarov [1995] to estimate the relative measure of the importance of a subset of input variables relative to the full set of input variables. It is also of interest to cite Sparkman et al. [2016], in which the authors proposed a sample weighting scheme similar to kernel regression for estimating the moments of the conditional distributions in the calculation of first-order and total Sobol' indices. Da Veiga et al. [2009] also proposed a methodology based on local linear regression for the estimation of first-order Sobol'

indices. The Nadaraya-Watson estimator can be seen as a particular case of the wider class of nonparametric estimators, called local polynomial estimators used in Da Veiga et al. [2009]. Specifically, it corresponds to performing a local constant fit. However the procedure in Da Veiga et al. [2009] is clearly different from ours, as it consists in splitting the sample in two disjoint sub-samples, the first one used to emulate the conditional expectation, the second one used to estimate each first-order Sobol' index, replacing the conditional expectation with its corresponding emulator. Note that metamodels have been widely used to estimate Sobol' indices (e.g., linear regression models, polynomials, splines). Most of the time, the metamodel is built to emulate f rather than the conditional expectation. Some of these metamodels, such as polynomial chaos (see, e.g., Sudret [2008], Le Gratiet et al. [2016]), lead to an analytical expression of Sobol' indices (using, e.g., Parseval equality if polynomial chaos are considered). In Marrel et al. [2009], the authors derive an analytical expression of Sobol' indices based on Gaussian process regression. In Durrande et al. [2013], the authors prone the use of ANOVA kernels.
In the framework of sensitivity analysis with a nonparametric estimator based on nearest neighbors, Broto et al. [2020] proposed to estimate Shapley effects introduced by Owen [2014].

Within the avalanche field, only a few studies have analyzed the influence of inputs on the outputs of dynamic propagation models. For example, Barbolini and Savi [2001] used a Monte Carlo approach to analyze the sensitivity of runout distances and impact pressures in the VARA model. Jamieson et al. [2008] discussed the main sources of uncertainty in the inputs of avalanche models. Borstad and McClung [2009] developed a sensitivity analysis of an avalanche model with a Coulomb-type friction law. Bühler et al. [2011] explored the influence of the digital elevation models resolution on the outputs of the RAMMS avalanche model. Eventually, Buhler et al. [2018] developed a sensitivity analysis of released volumes, runout distances and avalanche velocities in the runout zone with respect to the initial released volume with the RAMMS avalanche model. However, the previous studies did not apply formal statistical methods to quantify the respective importance of the inputs. Moreover, authors considered only scalar outputs in their approaches. By contrast, in our paper, we focus on the avalanche model proposed by Naaim et al. [2004] and we quantify the importance of its inputs on its outputs by estimating the aggregated Sobol' indices with our new nonparametric procedure. The uncertainty on the input parameters was defined based on data obtained from an avalanche released on 13 February 2013 at the Lautaret full-scale test-site [Thibert et al., 2015].

In summary, the aims of this study are: (i) to propose a nonparametric estimation method for the aggregated Sobol' indices in a given data framework when sample size is small (ii) to quantify the input importance in avalanche models, having complex outputs (e.g., a mix of functional and scalar outputs). The approach can be easily adapted to other avalanche models and more widely in environmental sciences in the frequent case of complex models with outputs which are both functional and scalar. This work is organized as follows: the aggregated Sobol' indices and the estimation method are described in Section 4.2. In Section 4.3, we test the estimation method on toy functions. Then, in Section 4.4, the avalanche model is described and the results are presented. Finally, in Section 4.5, the conclusions and perspectives are discussed.

4.2 Aggregated Sobol' indices

Let us denote by f the model which takes as inputs the vector $\mathbf{X} = (X_1, \ldots, X_d)$. The inputs $X_1, \ldots, X_d$ are modeled by random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and valued in a measurable space $E = E_1 \times E_2 \ldots \times E_d$. The output of the model f is the p multivariate vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^T$. It means, we have:

$$f(\mathbf{X}) = f(X_1, \ldots, X_d) = \mathbf{Y} = (Y_1, \ldots, Y_p)^T.$$

In the following, we assume that $\mathbb{E}(Y_1^2 + \ldots + Y_p^2) < \infty$ and that the random variables $X_1, \ldots, X_d$ are independent from each other. This condition guarantees the uniqueness of the ANOVA decomposition [Sobol', 1993]. To summarize the importance of each input X_i on the multivariate output $\mathbf{Y}$, we aim at computing the aggregated Sobol' index GS_i introduced in Lamboni et al. [2009] [see also Gamboa et al., 2013] defined as:

$$GS_i = \frac{\sum_{j=1}^{p} \mathrm{Var}(Y_j) S_i^j}{\sum_{j=1}^{p} \mathrm{Var}(Y_j)},$$

where S_i^j is the first-order Sobol' index of the output Y_j with respect to the input X_i namely

$$S_i^j = \frac{\mathrm{Var}(\mathbb{E}(Y_j|X_i))}{\mathrm{Var}(Y_j)} \tag{4.1}$$

Notice that $GS_i \in [0, 1]$. As in the case of the scalar indices, the main advantage of the aggregated indices is their easy interpretation: a high value means the input is important, a value close to zero means it is not, and $\sum_{i=1}^{d} GS_i = 1$.

In the following, we propose to estimate nonparametrically GS_i from given data, proposing a new procedure based on the Nadaraya-Watson kernel smoother [Nadaraya, 1964, Watson, 1964].

4.2.1 Nonparametric estimation procedure

Let $(X_i^\ell, Y_j^\ell)_{1 \leq \ell \leq n}$ be an independent identically distributed random sample of (X_i, Y_j). The Nadaraya-Watson kernel smoother consists in estimating the conditional mean of $\mathbb{E}(Y_j|X_i)$ at a point x in the domain of X_i by:

$$\widetilde{m}_h(x) = \sum_{\ell=1}^{n} Y_j^\ell W_{\ell,h}(x), \tag{4.2}$$

where

$$W_{\ell,h}(x) = \frac{K\left(\frac{x - X_i^\ell}{h}\right)}{\sum_{\ell'=1}^{n} K\left(\frac{x - X_i^{\ell'}}{h}\right)}, \tag{4.3}$$

with K a kernel function and h a bandwidth. In Equation (4.3), different weights $W_{\ell,h}$ can be used, as for example, local linear regression proposed in [Da Veiga et al., 2009] and nearest neighbor smoothers as proposed by Broto et al. [2020].

When the sample size is small, the selection of the bandwidth h is critical. Indeed, to avoid undersmoothing or oversmoothing, an optimal value of h which balances bias

and variance must be found [Tsybakov, 2008, page 17]. An option widely used to select h is cross-validation but if the sample size is small, the estimator will suffer of finite-sample bias [see Racine, 2001, Solís, 2019]. Therefore, if the sample size is small, it is recommended to perform a bias correction to the kernel smoother estimator preliminary to the bandwidth selection.

It is important to underline that the criterion we introduce here to select the bandwidth is different from the one proposed in Solís [2019], which was based on bagging minimization. The selection of the bandwidth we propose to apply in our paper is described in the next section.

Once the bandwidth h has been chosen, the estimation of the aggregated Sobol' index is straightforward: given a n sample $(X_i^k, Y_j^k)_{1\leq k\leq n}$ of (X_i, Y_j),

$$\widehat{GS}_i = \frac{\sum_{j=1}^{p} \widehat{\mathrm{Var}}\big[(Y_j^k)_{1\leq k\leq n}\big]\hat{S}_i^j}{\sum_{j=1}^{p} \widehat{\mathrm{Var}}\big[(Y_j^k)_{1\leq k\leq n}\big]}, \tag{4.4}$$

where $\hat{S}_i^j$ is the estimation of the first-order Sobol' index:

$$\hat{S}_i^j = \frac{\widehat{\mathrm{Var}}\big[\big(\widehat{m}_{\mathrm{h},-k}(X_i^k)\big)_{1\leq k\leq n}\big]}{\widehat{\mathrm{Var}}\big[(Y_j^k)_{1\leq k\leq n}\big]}, \tag{4.5}$$

with $\widehat{\mathrm{Var}}$ denoting empirical variance and

$$\widehat{m}_{\mathrm{h},-k}(X_i^k) = \sum_{\ell\neq k} y_j^\ell \frac{K\left(\frac{X_i^k - X_i^\ell}{h}\right)}{\sum_{\ell'\neq k} K\left(\frac{X_i^k - X_i^{\ell'}}{h}\right)}. \tag{4.6}$$

Note that the same sample $(X_i^k, Y_j^k)_{1\leq k\leq n}$ is used for both the estimation of $\mathbb{E}(Y_j|X_i)$ and the outer variance. This is the reason why we use the so-called leave-one-out estimate $\widehat{m}_{\mathrm{h},-k}$ and not the estimate $\widetilde{m}_h$ defined in (4.2).

4.2.2 Bias correction and bandwidth selection

As mentioned previously, cross-validation is one option to select the bandwidth when using the Nadaraya-Watson kernel smoother defined in (4.2). In more details, given a n sample $\mathbf{X} = (X^1, \ldots, X^n)$ of X_i and the corresponding evaluations $\mathbf{Y} = (Y^1, \ldots, Y^n)$ of Y_j, the bandwidth can be selected as (see, e.g., Takezawa [2005]):

$$h_{\mathrm{cv}} = \mathrm{argmin}_{h>0} \frac{1}{n}\sum_{k=1}^{n} \frac{\big(Y^k - \widetilde{m}_{\mathrm{h}}(X^k)\big)^2}{1 - W_{k,\mathrm{h}}(X^k)} \tag{4.7}$$

with $W_{k,\mathrm{h}}(X^k)$ defined by (4.3).

For notation simplicity, $K_h(.)$ denotes $K\left(\frac{\cdot}{h}\right)$ and we dropped all the i and j subindices, but keep in mind that we refer to a sample of X_i and Y_j.

Even if the kernel smoother with cross-validation is asymptotically unbiased, it is well known that it has a relatively large finite-sample bias. To tackle this issue, Racine [2001] proposed a bootstrap bias correction as a preliminary step to bandwidth selection. In this paper, we adapt the procedure introduced in Racine [2001] to the framework of Sobol' index estimation. The steps of our procedure are described hereafter:

1. Given a n sample $\mathbf{X} = (X^1, \ldots, X^n)$ of X_i, the corresponding evaluations $\mathbf{Y}_j = (Y^1, \ldots, Y^n)$ of Y_j and a bandwidth $h > 0$ (for example, an initial bandwidth value could be $h = h_{\text{cv}}$ given by Equation (4.7)), we calculate the leave-one-out Nadaraya-Watson kernel regression estimate $\widehat{m}_{\text{h},-k}(X^k)$ defined in (4.6). If $\sum_{\ell' \neq k} K_{\text{h}}(X^k - X^{\ell'}) = 0$, we set $\widehat{m}_{\text{h},-k}(X^k) = 0$.

2. We compute the errors:

$$\hat{\epsilon}^k = Y^k - \widehat{m}_{\text{h},-k}(X^k), \qquad k = 1, \ldots, n.$$

   These errors may be heteroscedastic because Y_j depends also on $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d$, therefore, a standardization of the errors must be applied before bootstraping.

3. The errors $\{\hat{\epsilon}^k\}_{1 \leq k \leq n}$ are standardized as:

$$\hat{v}^k = \frac{\hat{\epsilon}^k - \bar{\epsilon}}{r(X^k)}, \qquad k = 1, \ldots, n$$

   where $\bar{\epsilon}$ is the empirical mean of $\{\hat{\epsilon}^k\}_{1 \leq k \leq n}$ and $r(X^k)$ is the estimate of the conditional standard deviation of the random variable $\epsilon | X_i$ at X^k. Note that $r(X^k)$ is estimated using nonparametric kernel smoother estimation. More precisely, we estimate $r^2(x)$ by applying the Nadaraya-Watson kernel smoother as defined in (4.2) with $Y^\ell = \left(\epsilon^\ell\right)^2$, $\ell = 1, \ldots, n$ and selecting the bandwidth via cross-validation defined in (4.7).

4. A bootstrap sample $\mathbf{Y}^{(b)}$ of $\mathbf{Y}$ is created as:

$$Y^{(b)k} = \widehat{m}_{\text{h},-k}(X^k) + r(X^k)\hat{v}^{(b)k}, \qquad k = 1, \ldots, n,$$

   where $\hat{v}^{(b)k}$ is a bootstrap sample of $\{\hat{v}^k\}_{1 \leq k \leq n}$.

5. B bootstrap samples of $\mathbf{Y}$ are created using Step 4. For each bootstrap sample $(\mathbf{X}, \mathbf{Y}^{(b)})$, we estimate the leave-one-out kernel smoother $\widehat{m}^{(b)}_{\text{h},-k}$ as in (4.6).

6. The bias of $\widehat{m}_{\text{h},-k}(X^k)$ is estimated as:

$$\frac{1}{B} \sum_{b=1}^{B} \widehat{m}^{(b)}_{\text{h},-k}(X^k) - \widehat{m}_{\text{h},-k}(X^k).$$

   Then, the corrected kernel smoother $\widehat{m}^c_{\text{h},-k}(X^k)$ is computed as:

$$\widehat{m}^c_{\text{h},-k}(X^k) = \widehat{m}_{\text{h},-k}(X^k) - \left(\frac{1}{B} \sum_{b=1}^{B} \widehat{m}^{(b)}_{\text{h},-k}(X^k) - \widehat{m}_{\text{h},-k}(X^k) \right) \tag{4.8}$$

$$= 2\widehat{m}_{\text{h},-k}(X^k) - \frac{1}{B} \sum_{b=1}^{B} \widehat{m}^{(b)}_{\text{h},-k}(X^k). \tag{4.9}$$

   Note that the last two equations correspond to the standard bootstraping bias correction proposed by Efron and Gong [1983].

7. Finally, the bandwidth h is selected as

$$\mathrm{h}_{\mathrm{boot}} = \mathrm{argmin}_{h>0} \frac{1}{n} \sum_{k=1}^{n} \left(Y^k - \widehat{m}^c_{\mathrm{h},-k}(X^k)\right)^2.$$

We then estimate the aggregated Sobol' index GS_i with (4.4) by replacing $\widehat{m}_{\mathrm{h},-k}(X_i^k)$ by $\widehat{m}^c_{\mathrm{h}_{\mathrm{boot}},-k}(X_i^k)$ in (4.5). Racine [2001] also proposed an iterative version of the bootstrap bias corrections. In this work, we used the non iterated version because a major drawback of the iterative one is its computational cost. A simplified version of the iterative procedure was run on a simple test case (see Figure 4.13 in the Appendix) The improvements were not that convincing with regards to the additional cost.

Following Da Veiga et al. [2009], another approach could have been (i) to estimate $\mathrm{Var}(Y)$ as far as to define the kernel smoother $\widetilde{m}_{\mathrm{h}}$ defined in (2), (ii) to perform the bias correction and selection of the bandwidth, and (iii) then to sample from the joint distribution of input vector $\mathbf{X}$ to estimate $\mathrm{Var}(\mathbb{E}(Y_j|X_i))$ using the bias corrected kernel smoother as a metamodel. By contrast, the method proposed is metamodel-free and belongs to the class of nonparametric given data (one sample) methods. In the following, our estimator with bias correction is denoted by `cnp` and the same estimator without bias correction, for which the bandwidth is selected by solving $\mathrm{argmin}_{h>0} \frac{1}{n} \sum_{k=1}^{n} \left(Y^k - \widehat{m}_{\mathrm{h},-k}(X^k)\right)^2$, is denoted by `np`.

4.2.3 Dimension reduction based on principal component analysis

If the output is high-dimensional or even functional, it may be computationally interesting to reduce the output dimension in a preliminary step. There are different reduction techniques. Principal component analysis, also known as Karhunen-Loève decomposition in the functional framework, consists in projecting the output on a new basis so that most information is concentrated in the first few components [see, e.g., Pearson, 1901, Loève, 1963]. Note that if the outputs are the discretization of more than one functional random variable (e.g., velocity and flow depth), simultaneous PCA may provide efficient reduction. For more details, see e.g., the work of Nanty et al. [2017] and reference therein.

In brief, simultaneous PCA consists in applying PCA to the matrix composed by the concatenation of the discretized functions divided by a normalization factor. Nanty et al. [2017] show that using the maximum of each functional as normalization factor provides the best results. Thus, in the avalanche application, we follow their procedure.

There exist dimension reduction tools specific to functional data, such as the regularized functional PCA described in Chapter 9 of Ramsay and Silverman [2005] [see also the work of Yao et al., 2005] which has been applied in different fields [e.g., Locantore et al., 1999, Antoniadis et al., 2012]. Also, it has been shown that in some applications, functional PCA provides better results than usual PCA applied after a discretization of the functional data [e.g., Viviani et al., 2005]. On our test cases, usual PCA applied on the discretized data set and functional PCA provided similar results. The results we show in this paper are the ones obtained by applying usual PCA on discretized data.

The PCA consists in decomposing the variance-covariance matrix of the output. More precisely, let us denote by Σ the variance-covariance matrix of the output vector $\mathbf{Y}$. The principal component decomposition of $\mathbf{Y}$ is based on the expansion:

$$\Sigma = \sum_{j=1}^{p} \lambda_j \mathbf{v}_j \mathbf{v}_j^T,$$

where $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ are the ordered eigenvalues and $\mathbf{v}_1, \ldots \mathbf{v}_p$ are the orthonormal eigenvectors of Σ. Therefore, $\mathbf{Y}$ can be decomposed as:

$$\begin{aligned} \mathbf{Y} &= \mathbb{E}(\mathbf{Y}) + \sum_{j=1}^{p} \big((\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^T \mathbf{v}_j\big)\mathbf{v}_j \\ &= \mathbb{E}(\mathbf{Y}) + \sum_{j=1}^{p} H_j \mathbf{v}_j \end{aligned}$$

where H_j is the jth principal component of $\mathbf{Y}$. Note that $\text{Var}(H_j) = \sqrt{\lambda_j}$. The output can be approximated using the $q \leq p$ first components which capture the major part of the output variance:

$$\mathbf{Y} \approx \mathbb{E}(\mathbf{Y}) + \sum_{j=1}^{q} H_j \mathbf{v}_j.$$

Then, the aggregated Sobol' indices can be computed in the first q principal components. For example, for input X_i:

$$GS_i \approx \frac{\sum_{j=1}^{q} \lambda_j S_i(H_j)}{\sum_{j=1}^{q} \lambda_j} \tag{4.10}$$

where $S_i(H_j)$ denotes the Sobol' index of H_j with respect to input X_i.

The estimation of the aggregated Sobol' indices will be done for the reduced output $\mathbf{H} = (H_1, \ldots, H_q)^T$.

4.3 Test cases

The `cnp` method is tested on three toy functions and then, it is applied to an avalanche dynamic model. The `cnp` accuracy is compared with `np` (bandwidth selected with cross-validation and without bias correction), the method of Solís [2019] denoted by `Solis2019` and implemented in R in the library `sobolnp`, and the method of Tissot and Prieur [2012], denoted by `Tissot2012`. Tissot and Prieur [2012] proposed a bias correction for the index estimation based on random balance designs. To perform step 3 of the bias correction algorithm, we used the function `npreg` of the R package `np` developed by Hayfield and Racine [2008].

To compare our method to the one proposed in Solís [2019], we use the Epanechnikov kernel of order 2:

$$K(u) = \frac{3}{4}(1 - u^2) I_{(|u| \leq 1)},$$

where I denotes the indicator function. We compared the estimation results with the ones obtained with a Gaussian kernel and the results were quite similar (see Figure 4.14 in the Appendix). All the computations were performed in R Core Team [2017].

4.3.1 Toy functions

For each toy function, to test the accuracy of the method, first-order Sobol' indices were estimated with $N = 100$ independent samples of sizes $n \in \{100, 200, 300\}$. For `cnp` and `Solis2019`, the number of bootstrap samples was set to $B = 100$.

Scalar g-Sobol function

The scalar g-Sobol function is defined as:

$$Y = f(X_1, \dots, X_d) = \prod_{i=1}^{d} \frac{|4X_i - 2| + a_i}{1 + a_i}, \quad a_i \in \mathbb{R}, i \in \{1, \dots d\}$$

where $(X_1, \dots, X_d) \sim \mathcal{U}([0,1]^d)$.

For this model, an analytical expression of the scalar Sobol' indices is available:

$$S_i = \frac{\frac{1}{3(1+a_i)^2}}{\prod_{i=1}^{d}\left(\frac{3(a_i)^2+6a_i+4}{3(1+a_i)^2}\right) - 1}, \qquad i \in \{1, \dots, d\}.$$

For our experiments, we chose $d = 8$ and $\mathbf{a} = (0, 1, 4.5, 9, 99, 99, 99, 99)$. First-order indices were estimated in $N = 100$ independent samples of sizes $n \in \{100, 200, 300\}$ using the four methods `cnp`, `np`, `Tissot2012` and `Solis2019`. The results can be seen in Figures 4.1 and 4.2.

We remark that the `cnp` procedure we have proposed in this work behaves well even for small data sets. The bias correction seems to outperform the usual `np` procedure, except maybe for $n = 100$.

The variances of both `cnp` and `np` are improved with respect to the nonparametric bagging approach in Solís [2019]. Tissot and Prieur [2012] and `cnp` methods provide good results for samples of size $n = 300$. Contrary to `Tissot2012`, the advantage of cnp is that it does not require a structured sample design. At least, we remark a decrease of the efficiency of our approach for very small Sobol' indices.

Multivariate g-Sobol function

We introduce now a multivariate version of the g-Sobol function. It is defined as:

$$(Y_1, \dots, Y_p)^T = f(X_1, \dots, X_d) \text{ with } (X_1, \dots, X_d) \sim \mathcal{U}([0,1]^d)$$

where

$$Y_j = \prod_{i=1}^{d} \frac{|4X_i - 2| + a_i^j}{1 + a_i^j}, \quad a_i^j, \text{ with } j \in \{1, \dots, p\}, i \in \{1, \dots d\}.$$

An analytical expression of aggregated Sobol' indices is available for this test case:

$$GS_i = \frac{\sum_{j=1}^{p} \frac{1}{3(1+a_i^j)^2}}{\sum_{j=1}^{p} \prod_{i=1}^{d}\left(\frac{3(a_i^j)^2+6a_i^j+4}{3(1+a_i^j)^2}\right) - 1}.$$

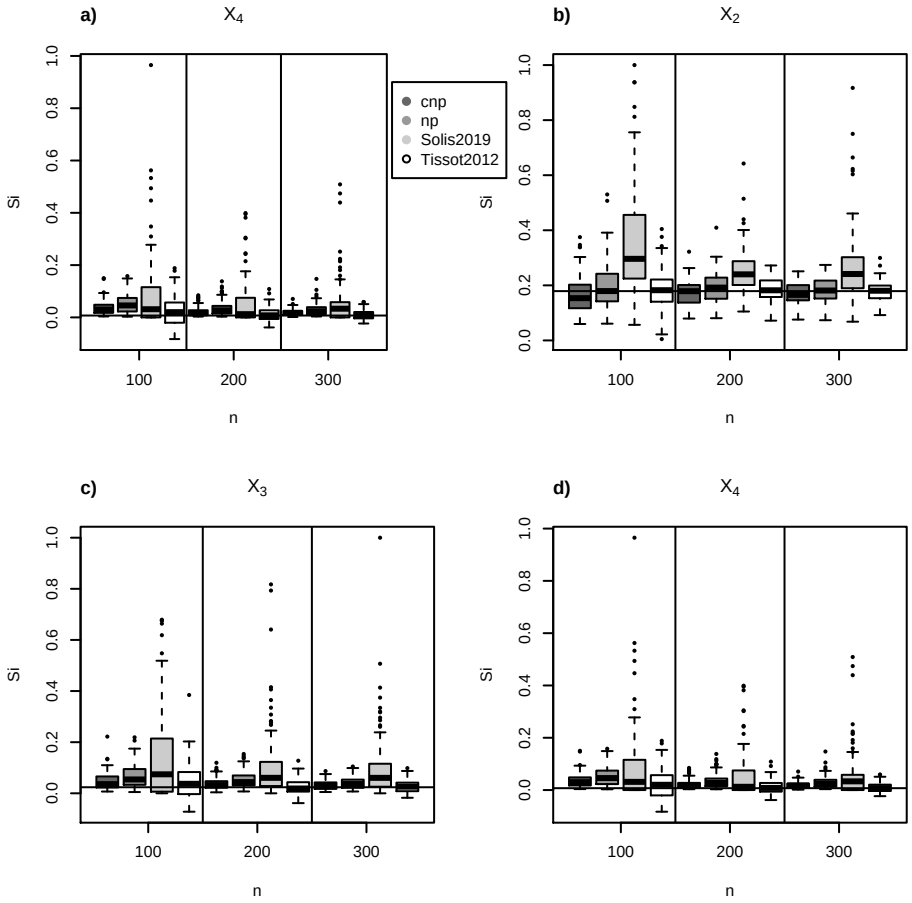

Figure 4.1: Scalar g-Sobol function: estimation of first-order Sobol' indices for X_1, X_2, X_3 and X_4 using four estimation methods: `cnp`, `np`, `Solis2019` and `Tissot2012`. To draw the boxplots, the indices were computed in $N = 100$ independent samples of sizes $n \in \{100, 200, 300\}$. For each iteration of `cnp` and `Solis2019`, we used $B = 100$ bootstrap samples.

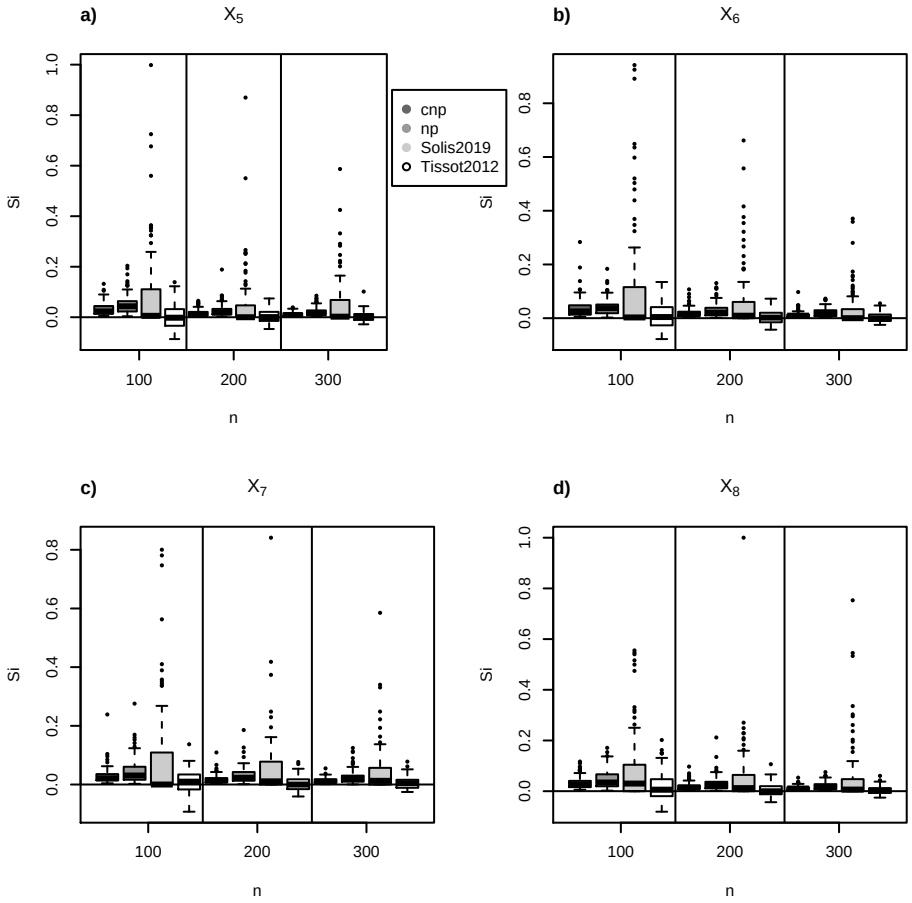

Figure 4.2: Scalar g-Sobol function: estimation of first-order Sobol' indices for X_5, X_6, X_7 and X_8 using four estimation methods: `cnp`, `np`, `Solis2019` and `Tissot2012`. To draw the boxplots, the indices were computed in $N = 100$ independent samples of sizes $n \in \{100, 200, 300\}$. For each iteration of `cnp` and `Solis2019`, we used $B = 100$ bootstrap samples.

Input	Description	Distribution
m	mass (kg)	$\mathcal{U}[10, 12]$
c	damping constant ($\mathrm{Nm}^{-1}\mathrm{s}$)	$\mathcal{U}[0.4, 0.8]$
k	spring constant (Nm^{-1})	$\mathcal{U}[70, 90]$
l	initial elongation (m)	$\mathcal{U}[-1, -0.25]$

Table 4.1: Mass spring model: Input description and uncertainty intervals.

We chose $p = 2$, $d = 6$ and the coefficients a_i^j, $i \in \{1, \ldots, 6\}$, $j \in \{1, 2\}$ were coded by the matrix $A_2 \in \mathbb{R}^{6\times 2}$:

$$A_2^T = \begin{pmatrix} 0 & 0.5 & 3 & 9 & 99 & 99 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Figure 4.3 shows the estimation of first-order Sobol' indices using `cnp`, `np`, `Solis2019` and `Tissot2012` in $N = 100$ independent samples of sizes $n \in \{100, 200, 300\}$. The bias accuracy of `cnp` compared to the other three methods is the best (see Figure 4.3). Furthermore, the variance of `cnp` estimation is lower than the other three methods (see Figure 4.3). Additionally, for this test case, the bias correction performance is more evident than in the previous section (see Figures 4.1 and 4.2 compared to Figure 4.3).

4.3.2 A functional example: the mass-spring model

Before applying our procedure to the avalanche application, we illustrate its performance on a test case with discretized functional output: the functional mass-spring model proposed by Gamboa et al. [2013], where the displacement of a mass connected to a spring is considered:

$$mx''(t) + cx'(t) + kx(t) = 0, \tag{4.11}$$

with initial conditions $x(0) = l$, $x'(0) = 0$, and $t \in [1, 40]$. There exists an analytical solution of Equation (4.11). This model has four inputs (see more details in Table 4.1). The model output is the vector:

$$\mathbf{Y} = f(\mathbf{X}) = (x(t_1), \ldots, x(t_{800})), \quad t_i = 0.05i \text{ with } i \in \{1, \ldots, 800\}.$$

The discretized output is high-dimensional ($p = 800$). Therefore, we first reduce the dimension of the output by using PCA before estimating the aggregated Sobol' indices in the first q principal components with $q \ll p$.

The indices were estimated in $N = 100$ independent samples of size $n \in \{100, 200, 300\}$. Each sample was also used for PCA. We tested if the results were more accurate if we used two independent samples, one for PCA and the other one for the estimation of Sobol' indices. At a constant cost in terms of model evaluations, the results were even better using only one sample (see Figure 4.12 in the Appendix Section 4.6).

The explained variance as a function of the number of PCs q is shown in Figure 4.4. Notice that 70% of the variance is explained with a small number of PCs ($q = 3$) and almost all the variance (95%) is explained by 6 PCs. Thus, the aggregated Sobol' indices were estimated using the first 6 PCs.

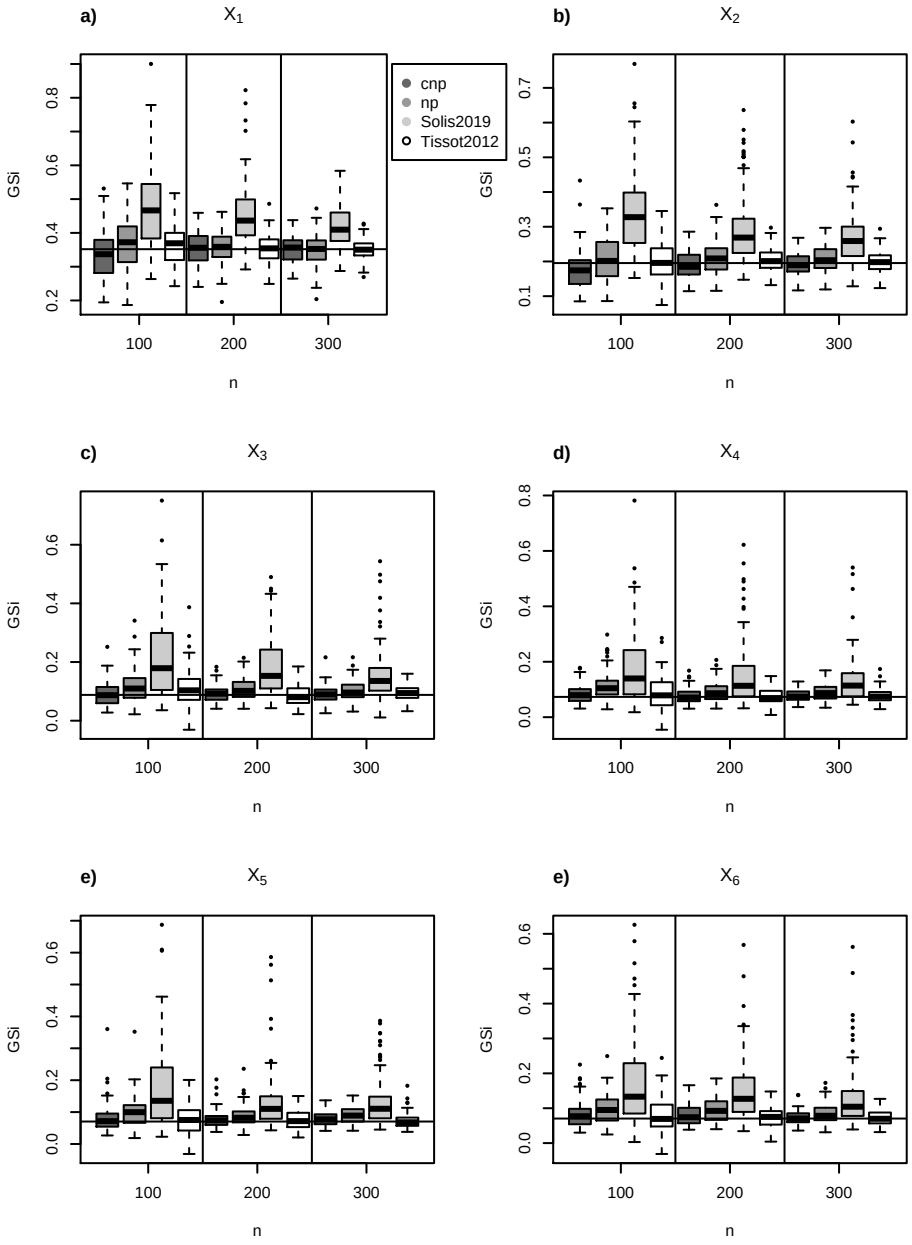

Figure 4.3: Multivariate g-Sobol function: estimation of the aggregated Sobol' indices using `cnp`, `np`, `Solis2019` and `Tissot2012`. Boxplots were drawn from N=100 independent samples of sizes $n \in \{100, 200, 300\}$. To use `cnp` and `Solis2019`, we set $B = 100$.

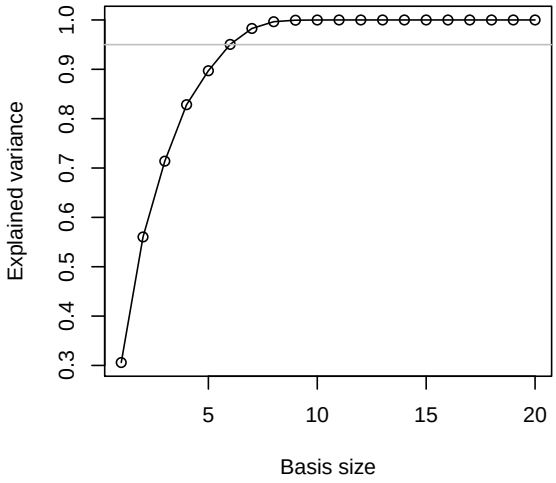

Figure 4.4: Mass-spring model: Explained variance as a function of the decomposition basis size (number of PCs). The gray line is displayed at the 95% of the variance explained which corresponds to 6 PCs.

The true aggregated Sobol' indices for this instance are unknown but they are considered close to the values estimated with a sample of large size $n = 100\,000$ using the pick-and-freeze method of Tissot and Prieur [2015].

The evolution of the estimation of the aggregated Sobol' indices using the four methods in function of the sample size $n \in \{100, 200, 300\}$ are shown in Figure 4.5. The `cnp` estimation procedure presents a much smaller variance than the np and `Solis2019` ones. The bias correction allows to correct the bias of the `np` estimation procedure, even if a residual bias seems to persist. The bias decreases as n increases, as expected. As it was mentioned before, our aim is factor prioritization and this objective is achieved with the `cnp` method in the sense that the most influential factor X_3 is clearly identified.

In general, the `cnp` method has a better accuracy compared to `np`, and `Solis2019` in all the test cases we considered. Even if `Tissot2012` shows better results than `cnp` for some of the first-order indices, it suffers from two drawbacks: (i) it requires a structured sampling design which makes impossible its application in a given data framework, (ii) it requires the choice of the number of harmonics M, for which no adaptive procedure is available yet (in our experiments M was fixed to 6, as recommended in the literature). For these reasons we apply in the following section the `cnp` method to estimate the aggregated Sobol' indices of an avalanche model.

4.4 Application: the avalanche model

The avalanche model used in this study represents the avalanche motion as a fluid using depth-averaged Saint-Venant equations [Naaim et al., 2004]. In short, the model considers only the dense layer of the avalanche. The depth of the flow is then small compared to its length. Under this assumption, shallow-water approximations of the mass and momentum equations can be used. Also, the model assumes the avalanche is flowing on a curvilinear

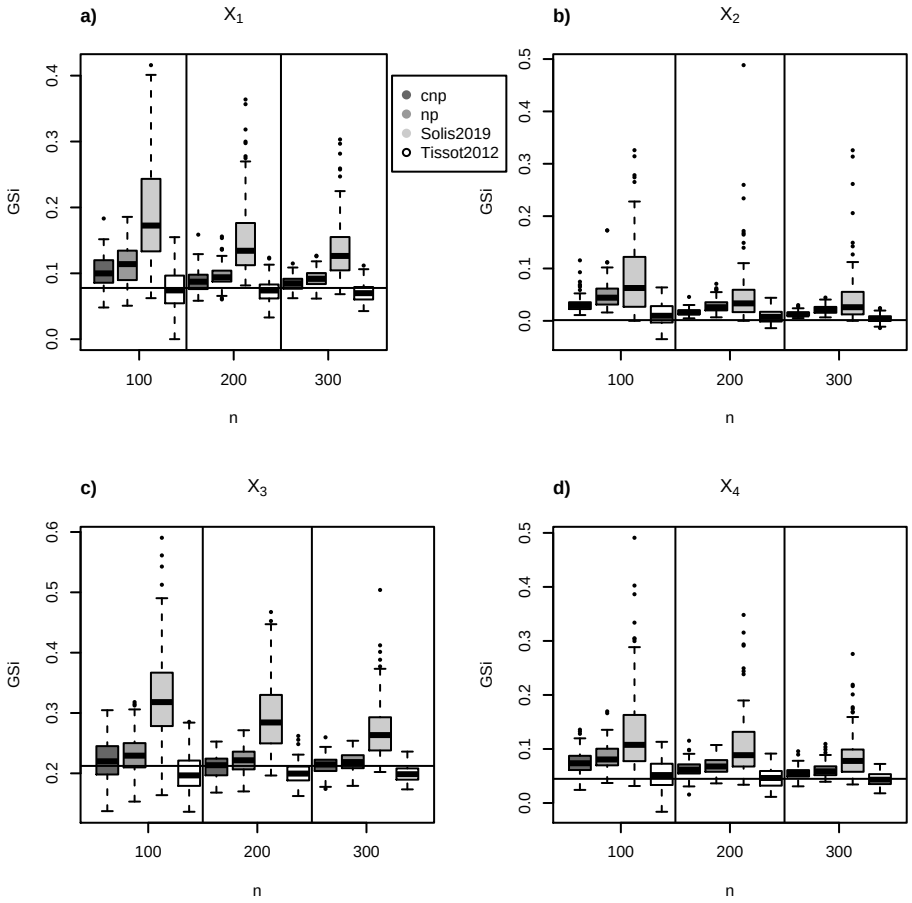

Figure 4.5: Mass-spring model: Estimation of the aggregated Sobol' indices using `np`, `Solis2019` and `Tissot2012`. Boxplots were drawn from N=100 independent samples of different sizes $n \in \{100, 200, 300\}$. To use `cnp` and `Solis2019`, we set $B = 100$.

Input	Description	Distribution
μ	Static friction coefficient	$\mathcal{U}[0.177, 0.498]$
ξ	Turbulent friction [m.s^{-2}]	$\mathcal{U}[306.97, 1475.67]$
$\mathrm{h_{start}}$	Snow depth within the release zone [m]	$\mathcal{U}[0.17, 0.33]$
$\mathrm{l_{start}}$	Length of the release zone [m]	$\mathcal{U}[24, 34]$
σ	Digital Elevation Model error [m]	$\mathcal{U}[0, 0.15]$

Table 4.2: Avalanche model: Input description and uncertainty intervals. $\mathcal{U}[a, b]$ denotes an uniform distribution in the interval $[a, b]$.

profile $z = l(x)$, where z is the elevation and x is the projected runout length distance measured from the starting abscissa point of the avalanche. The Voellmy friction law is considered. We denote hereafter the avalanche model by f.

The avalanche model depends on six inputs: two friction parameters μ and ξ, the length $\mathrm{l_{start}}$ of the avalanche release zone, the snow depth $\mathrm{h_{start}}$ within the release zone, the release position of the avalanche denoted by $\mathrm{x_{start}}$ and the discretized topography of the flow path, denoted by $D = (\mathbf{x}, \mathbf{z}) \in \mathbb{R}^{T\times 2}$ where $\mathbf{x} \in \mathbb{R}^T$ is the vector of projected runout length from the starting point of the avalanche and $\mathbf{z} = l(\mathbf{x}) \in \mathbb{R}^T$ is the elevation vector. T is the number of points of the discretized path. In addition to these parameters, we included the term σ to code the error of the digital elevation model $\mathbf{z}$ on the path's topography. To do so, a Gaussian error term is added to each element of $\mathbf{z}$ to construct a new topography $D' = (\mathbf{x}, \mathbf{z}')$ for each new model run, whose elevation is calculated as follows:

$$\mathbf{z}' = \mathbf{z} + \mathcal{N}(0, \mathbb{I}_T \sigma^2)$$

where $\mathbb{I}_T$ is the $T \times T$ identity matrix. This error codes the imprecision of the digital elevation model which allows analysing how it affects the simulation results.

In this study, we analyzed precisely the sensitivity of the model for the simulations of a single avalanche event. The input uncertainty intervals are those corresponding to an avalanche released at the Lautaret full-scale test-site on 13 February 2013. More details about this avalanche event can be found in [Pulfer et al., 2013, Heredia et al., 2020], thus parameter (x,z) is not uncertain anymore.

For this avalanche event, the release position of the avalanche $\mathrm{x_{start}}$ is precisely known because it was fixed by the experimental team (artificial avalanche release). Thanks to the sophisticated equipment of the test-site, the uncertainty of this input can therefore be neglected. The $\mathrm{l_{start}}$ and $\mathrm{h_{start}}$ input uncertainty intervals could be determined thanks to measures taken during the experiment. By contrast, the inputs μ and ξ cannot be measured. In another study, we developed a Bayesian calibration approach to infer these inputs from the measurements made along the flow. Thus, the uncertainty intervals of μ and ξ considered here correspond to the 95% credibility intervals of the posterior distribution we obtained [Heredia et al., 2020]. However, to avoid privileging some values in the sensitivity analysis, uniform distributions were chosen to model the $d = 5$ inputs, including the friction parameters (i.e. we did not directly use their posterior distribution). The uncertain inputs of the model and their uncertainty intervals are summarized in Table 4.2.

The outputs of the avalanche model are the functional maximal velocity $\vec{v}$ and the functional maximal flow depth $\vec{h}$ of the avalanche on the discretized grid corresponding to

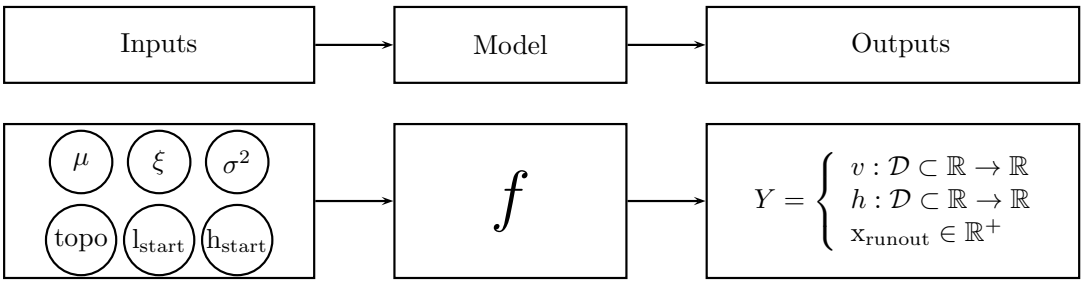

Figure 4.6: Avalanche model: Inputs and outputs. x_{start} is fixed. All other quantities are taken as random and considered in the sensitivity analysis.

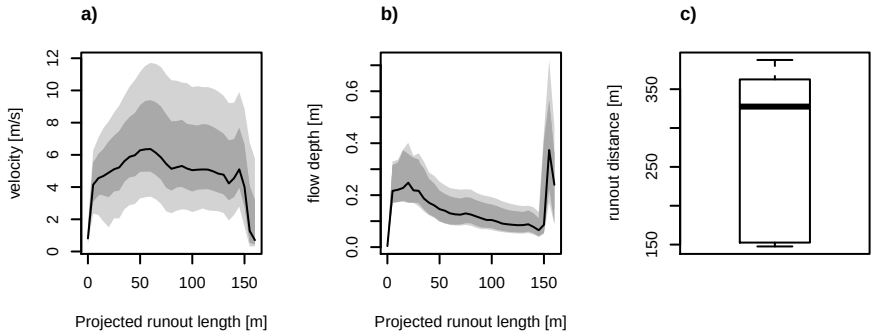

Figure 4.7: Avalanche model: **a)** functional high density region boxplots for the velocity **b)** functional high density region boxplots for the flow depth and, **c)** boxplot for the runout distance.

the topography D, and the runout distance of the avalanche denoted by x_{runout}. In other words, the model has two vectorial and one scalar outputs.

The vectors whose components are the evaluations on the discretized topography D of the velocity and flow depth functional outputs are denoted $\mathbf{v} = (v_1, \ldots, v_T)$ and $\mathbf{h} = (h_1, \ldots, h_T)$, respectively. Note that we have $p = 2 \times T + 1$ outputs. A brief summary of our framework is drawn in Figure 4.6.

The global sensitivity analysis is conducted only on a section of the path where all the avalanche simulations are flowing. More precisely, considering e.g., the velocity, our sensitivity analysis is performed on a subset of length n, $\mathbf{v} = \{v_1^\ell, \ldots, v_{T_1}^\ell\}_{\ell \in \{1,\ldots n\}}$, with $T_1 \leq T$ and such that $v_k^\ell > 0$ for all $k \in \{1, \ldots, T_1\}$.

Figure 4.7 shows the functional high density region (HDR) boxplots of 300 velocity (a) and snow depth curves (b) and the boxplot for the runout distance (c). The functional HDR plots are a tool for visualizing large amounts of functional data based on the estimation of the bivariate kernel density function of the two first components of the decomposition of the functional data [see Hyndman and Shang, 2010, for more details]. The HDR boxplots show the 50% HDR and the 100% HDR in light and dark gray, respectively. The modal curve is plotted with a solid line. These Figures were obtained using the R package `rainbow` developed by Hyndman and Shang [2010].

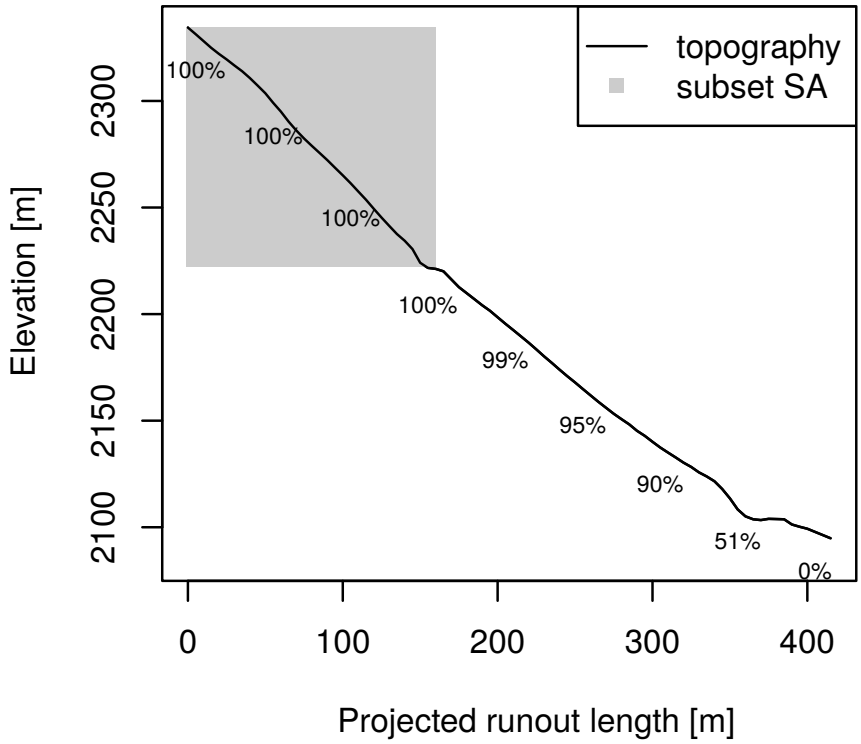

Figure 4.8: Avalanche model: path's topography (black line) with $T = 84$ discretization points and the subset of the path in gray box ($T_1 = 33$ discretization points) where the sensitivity analysis is performed. The panel highlights the percentage of avalanche flowing as function of the position within the path.

4.4.1 Sobol' indices

To estimate the sensitivity indices, we developed $n = 300$ model simulations using the random input distributions shown in Table 4.2. Figure 4.8 shows the topography of the avalanche path and the gray box shows the region where all the avalanche simulations are flowing. The output values in this subset are used in the sensitivity analysis. Also, in Figure 4.8, we show the percentage of avalanches flowing in the path as function of the position along the topography.

4.4.2 Scalar Sobol' indices

The scalar Sobol' indices calculated using $n = 300$ samples are shown in Figure 4.9. For the velocity output, in the first 50 m of the path, the input μ is the most important, then it is followed by ξ. For the rest of the path, the input ξ is the most important. For the flow depth output, the $\mathrm{h_{start}}$ input is the most important. The input σ^2 is the least important parameter for both outputs. The scalar sensitivity indices for the runout distance are shown in Figure 4.10. For this output, the h_{start} input is the most important. These three figures give us valuable information about the model. However, it could also be interesting to get a summarized information by computing aggregated sensitivity indices and this is done in the following section.

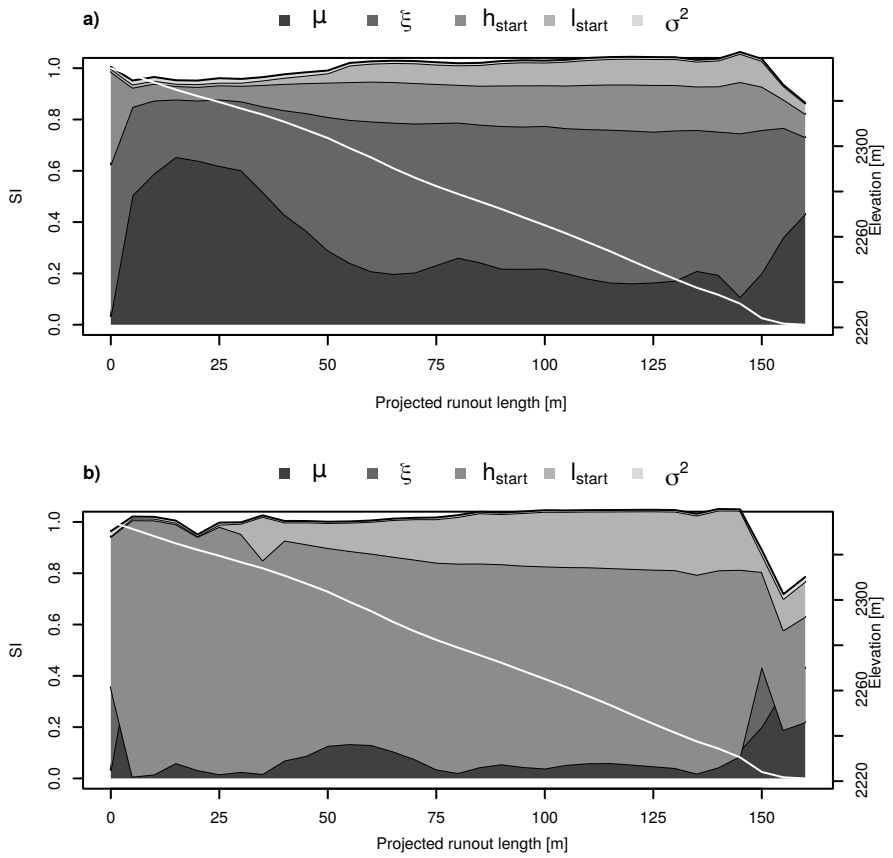

Figure 4.9: Avalanche model: **a)** scalar Sobol' indices for the velocity and **b)** flow depth outputs. The white line corresponds to the path's topography.

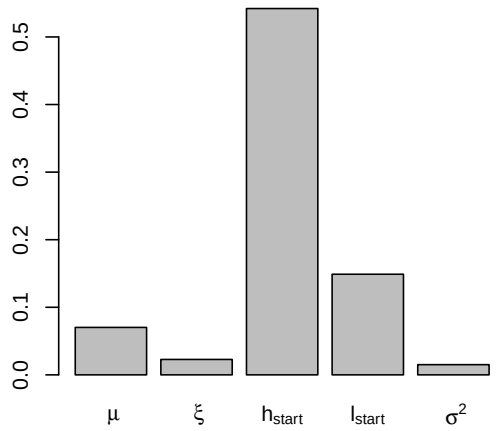

Figure 4.10: Avalanche model: scalar Sobol' indices for the runout distance output.

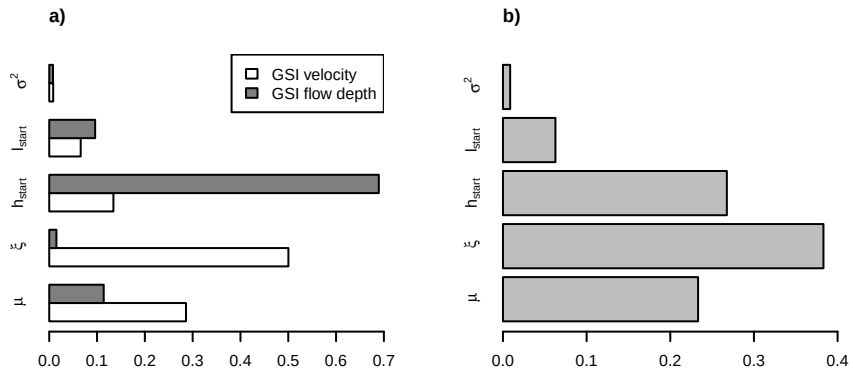

Figure 4.11: Avalanche model: aggregated indices estimated with two ways of aggregating the information. **a)** Two PCAs performed separately for the two functional outputs: one for the velocity output and the other one for the flow depth output. **b)** Simultaneous PCA to estimate the indices of the three outputs (the two functional ones and the scalar runout distance) altogether.

4.4.3 Aggregated Sobol' indices

To estimate the aggregated indices, we used two ways of aggregating the output information. In the first one, we performed two PCAs separately: one for the velocity output and the other one for the flow depth output. And in the second one, we used simultaneous PCA to estimate the indices of the three outputs. The results of the aggregated indices using two PCAs are shown in the panel **a)** of Figure 4.11 and using simultaneous PCA in the panel **b)** of Figure 4.11. This summarizes the importance of the inputs.

Let us analyze the results of the aggregated indices calculated when performing a PCA on each functional output: for the velocity output, the ξ and μ inputs are the most relevant and for the flow depth output, the h_{start} input is by far the most important. These results are consistent with findings of other studies [e.g., Barbolini and Savi, 2001, Borstad and McClung, 2009, Fischer et al., 2015].

Finally, let us analyze the results of the aggregated indices calculated when performing a simultaneous PCA for the whole set of outputs (the two functional ones and the scalar runout distance): the most important inputs are ξ, h_{start} and μ.

4.5 Conclusions and perspectives

In this work, we proposed a nonparametric method to estimate the aggregated Sobol' indices from a given random sample of small size, and called this method `cnp`. The method is based on the Nadaraya-Watson kernel smoother. Due to the small size of the sample at hand, the kernel estimation is biased. Therefore, to remove the bias of the estimation, we proposed a bias correction using bootstrapping samples based on the works of Racine [2001] and Solís [2019] before bandwidth selection based on cross-validation.

We tested the accuracy of the method on a scalar and two multivariate test cases. In general, the method `cnp` is accurate, even if the sample size is low ($n = 300$). The method was developed to estimate the aggregated indices for an avalanche model which has three outputs: the functional velocity and flow depth discretized on a given topography and

the runout distance of the avalanche. In this work, we developed a sensitivity analysis of a single avalanche event. The event corresponds to an avalanche released at the Lautaret test-site the 13 February 2013. The results have shown that for this particular avalanche event: for the velocity output, the μ and ξ inputs are the most relevant and for the snow depth output, h_{start} is the most important. Finally, for the runout distance output, h_{start} is the most important. We also showed how this information can be aggregated in one single set of indices summing up the model sensitivity which could be a very useful information for avalanche practitioners.

Note that in this work we proposed a method to quantify the importance of the inputs of a particular avalanche model but this method can be widely applied to other avalanche models. Moreover, the method could be applied to other avalanche events with the same model to generalize the results. Eventually, it could be adapted to various problems for which complex models with outputs which are both functional and scalar are employed.

Eventually, in this paper, we did not provide confidence intervals (CI) associated to the estimation of Sobol' or aggregated Sobol' indices. This would be an interesting task for a future work. However, bootstrap based CI are costly in that framework as they would involve two boostrap stages: one for the bias correction and one for the computation of CI.

4.6 Appendix

In this appendix, we compare the accuracy of Sobol' index estimation for the mass-spring test case when using a single sample of size $n = 200$ for both PCA and Sobol' index estimation or two independent samples of size $n = 100$, one for PCA, the other one for Sobol' index estimation. Boxplots presented in Figure 4.12 were obtained with $N = 100$ independent replications.

Figure 4.13 shows the comparison between the non iterative version of Racine [2001] algorithm (cnp) and the iterative version (ite. cnp) in the g-Sobol test case.

We also compare the estimation of the indices using an Epanechnikov of second order kernel (left panel) or a Gaussian kernel (right panel) in the g-Sobol test case. The results are shown in Figure 4.14.

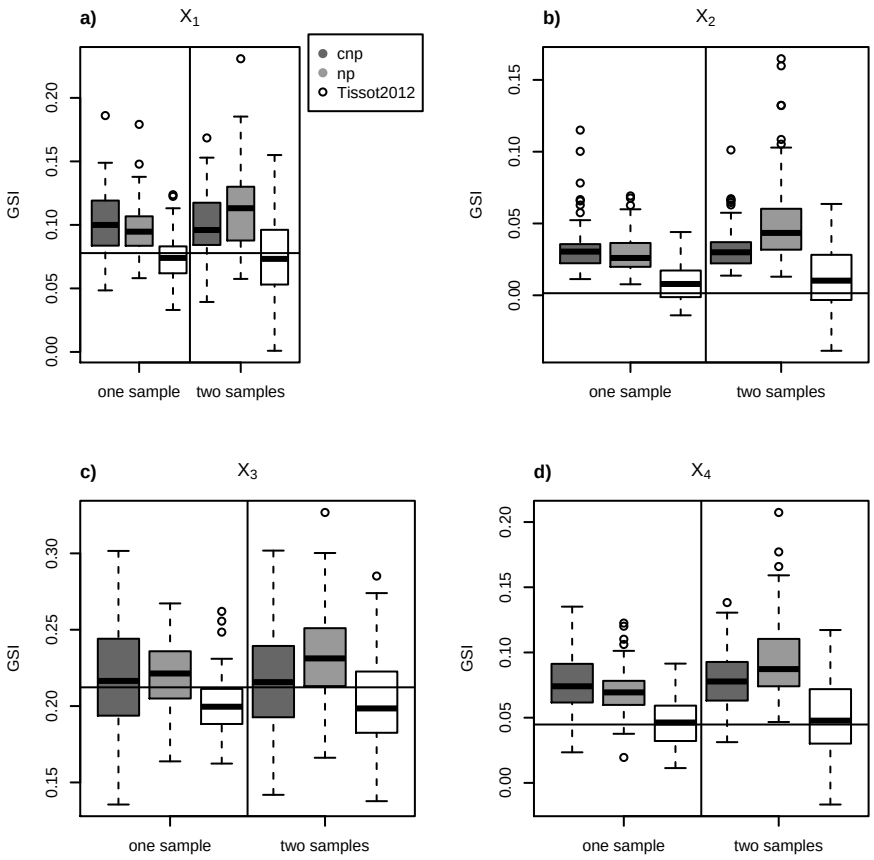

Figure 4.12: Mass-spring model: Estimation of the aggregated Sobol' indices using `cnp`, `np` and `Tissot2012`. Boxplots were drawn from N=100 independent samples of sizes 200. The estimation results using the same sample (one sample) are two different samples (two samples) to estimate Sobol' indices and PCs are shown. We set $B = 100$ for `cnp`.

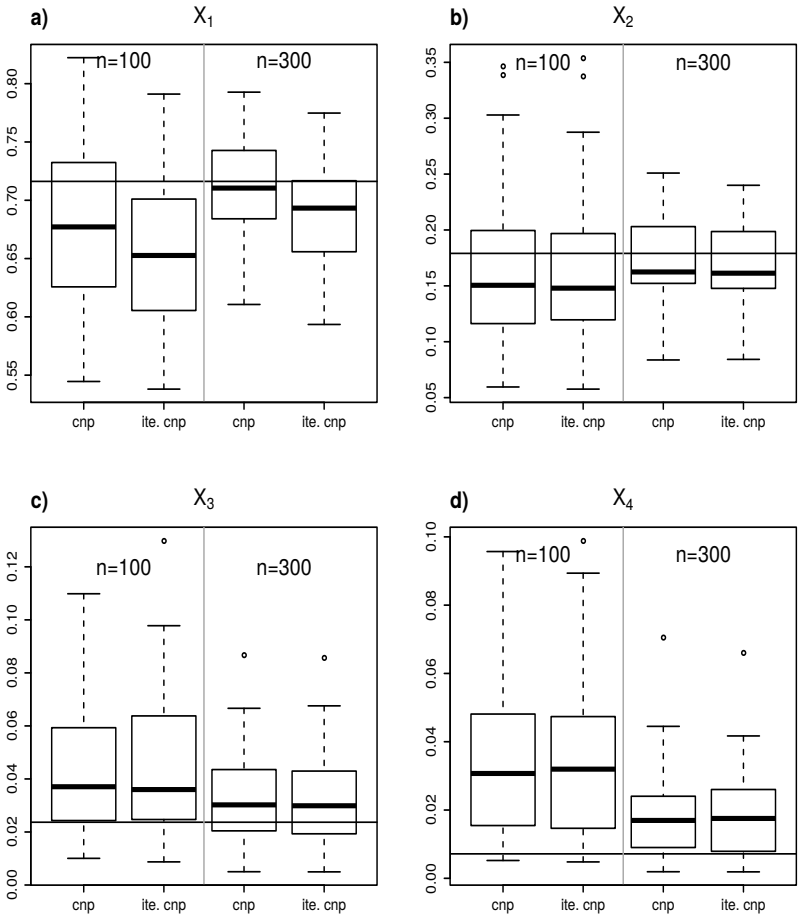

Figure 4.13: Boxplots of first-order Sobol' index estimation using the non iterative version of Racine [2001] algorithm (cnp) and the iterative version (ite. cnp) in the g-Sobol test case using $N = 30$ independent random samples of size $n = \{100, 300\}$.

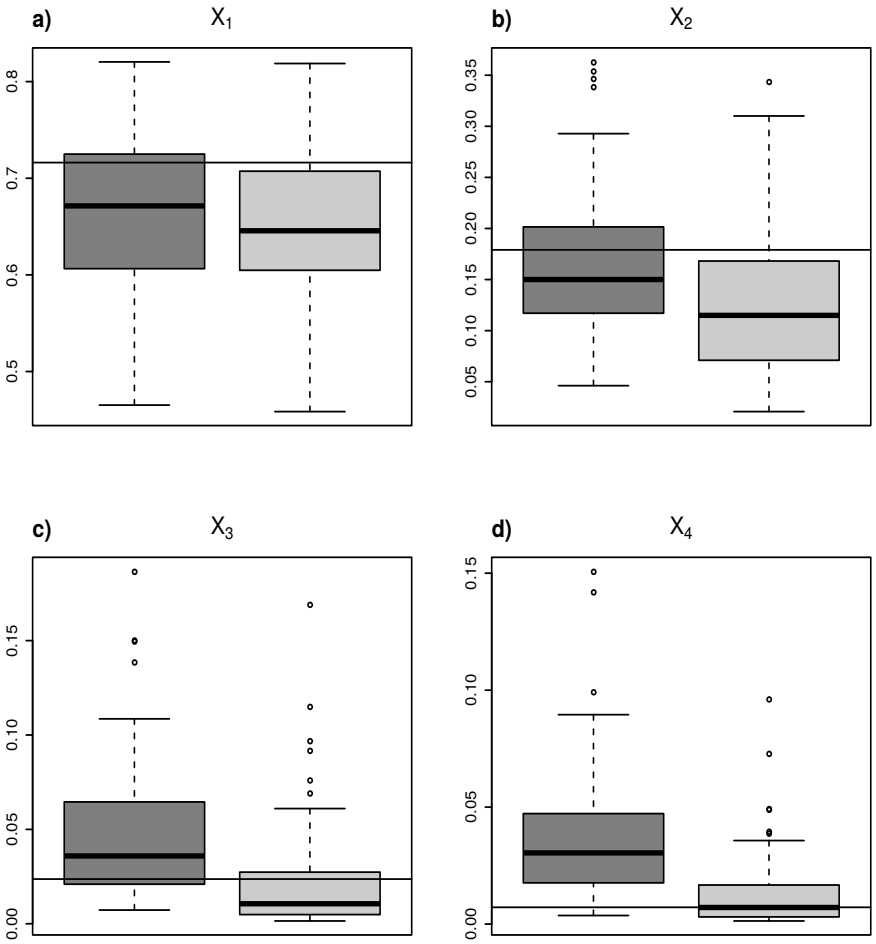

Figure 4.14: Boxplots of first-order Sobol' indices estimation using an Epanechnikov and a Gaussian kernel in the g-Sobol test case using $N = 100$ independent random samples of size $n = 100$.

5

Given data inference of aggregated Shapley effects, application to avalanche modeling.

María Belén Heredia, Clémentine Prieur, Nicolas Eckert, *Given data inference of aggregated Shapley effects, application to avalanche modeling.*

Contents

Abstract: Dynamic models are simplified representations of some real-world entities that change over time. They are essential analytical tools with significant applications, e.g., in environmental and social sciences. The outputs produced by dynamic models are typically time and/or space dependent. Due to physical constraints, their parameters are confined to a non-rectangular domain. Also, they can be significantly sensitive to variations of input parameters. A global sensitivity analysis (GSA) consists in modeling input parameters by a probability distribution which propagates through the model to the outputs. Then, input parameters are ordered according to their contribution on the model outputs by computing sensitivity measures. In this paper, we extend Shapley effects, a sensitivity measure well suited for dependent input parameters, to the framework

of dynamic models. Note that if input parameters are constrained to a non-rectangular domain, they cannot be modeled by independent random variables. We also propose an algorithm to estimate the so-called aggregated Shapley effects and to construct bootstrap confidence intervals for the estimation of scalar and aggregated Shapley effects. We measure the performances of the estimation procedure and the accuracy of the probability of coverage of the bootstrap confidence intervals on toy models. Finally, our procedure is applied to perform a GSA of an avalanche flow dynamic model, for which the sample is obtained by restricting the initial input/output sample to keep only combinations that are physically meaningful. More precisely, we analyze the sensitivity in two different settings: (i) little knowledge on the input parameter probability distribution, and (ii) well-calibrated input parameter distributions. Probative linkages between local slope and sensitivity indices demonstrate the usefulness of our approach for practical problems.

5.1 Introduction

Dynamic models are simplified representations of some real-world entity that change over time, in equations or computer code. These models are useful for the analysis of real-world phenomena, e.g., in environmental or social sciences. For a better understanding of a phenomenon or for forecasting purposes, it might be important to identify which input parameters entering in the formulation of such dynamic models are influential on the outputs of interest. Determining these influential parameters is one aim of global sensitivity analysis (GSA). A global sensitivity analysis (GSA) consists in modeling unknown input parameters by a probability distribution which propagates through the model to the outputs. Then, input parameters are ordered according to their contribution on the model outputs by computing sensitivity measures. In the literature, there exist different global sensitivity measures, e.g., variance based measures such as Sobol' indices [Sobol', 1993] or Shapley effects [Owen, 2014], density based measures [Borgonovo et al., 2016, Da Veiga, 2015], etc. A review of global sensitivity measures can be found in, e.g., Borgonovo and Plischke [2016] or Iooss and Lemaître [2015].

A particularity of dynamic models considered in this paper is that the outputs they produce are typically time and/or space dependent (see e.g., Alexanderian et al. [2020], Lamboni et al. [2011]). More specifically, the application that motivated our study is an avalanche flow dynamic model (proposed by Naaim et al. [2004]). The input parameters in our application cannot be considered as stochastically independent. Indeed, even if they are first sampled independently, we only analyze in the GSA the ones leading to snow avalanche simulations that are both realistic and of interest for risk purposes. For example, an avalanche simulation is rejected if its volume is inferior to 7000 m^3 as such small events do not represent a significant threat for settlements downslope. It is quite common that such physical constraints lead to dependent inputs by excluding infeasible sets of input values. Let us cite, e.g., Radaideh et al. [2019] dealing with the response of a nuclear reactor or López-Benito and Bolado-Lavín [2017] interested in a natural gas transmission model (see also Kucherenko et al. [2017]). For that reason, we develop a GSA methodology capable to handle dependent inputs, with unknown dependence structure, as far as dynamic models through functional outputs (or high-dimensional multivariate outputs as functional outputs are most of the time represented on a discrete grid).

Although the independence assumption on input parameters is unrealistic in many applications, it is traditionally required to interpret or to compute sensitivity measures.

Many alternatives to Sobol' indices [Sobol', 1993] based on functional ANOVA decomposition have been proposed in the literature to handle GSA in the framework of dependent input parameters. We refer the interested reader to the introduction in [Iooss and Prieur, 2019] and to references therein for a complete overview of these alternatives.

In this paper, we focus on Shapley effects introduced in [Owen, 2014], consisting in variance based sensitivity measures, still meaningful in the framework of dependent input parameters [Owen and Prieur, 2017]. This approach is based on the Shapley value which is a cooperative game theory concept. Briefly speaking, Shapley value ensures a fair distribution of a gain among team players according to their individual contributions. As a sensitivity measure, Owen [2014] adapted the Shapley value into the Shapley effects by considering model input parameters as players and the gain function as the output variance. The main advantage of such an approach is that it is possible to attribute a non negative sensitivity index to each parameter, and the sum of the indices is equal to one [Broto et al., 2020, Iooss and Prieur, 2019].

A first algorithm to estimate Shapley effects was proposed in [Song et al., 2016]. A block bootstrap procedure was proposed in [Benoumechiara and Elie-Dit-Cosaque, 2019] to add the construction of confidence intervals to the aforementioned algorithm. More recently, [Broto et al., 2020] proposed a subset aggregation procedure which leads to a significant reduction of the variance of Shapley effect estimation. Let us also cite the algorithm proposed in [Plischke et al., 2020] based on the Möbious inverse, which offers a computationally efficient alternative for the estimation of Shapley effects. The main advantage of the subset aggregation procedure introduced in [Broto et al., 2020] is that it has a version, based on nearest-neighbors, which does not require the ability to sample from the exact conditional distributions of the input parameters.

In the present paper, we propose an extension of Shapley effects to multivariate or functional outputs. Although it is possible to compute a Shapley effect for each component of a multivariate output, it leads to results difficult to interpret as the balance of sensitivities may change from one component to the other [Alexanderian et al., 2020] and to redundancies in the information, particularly when considering the discretization of a time and/or space dependent output [Lamboni et al., 2009]. For that reason, Lamboni et al. [2009] (see also Gamboa et al. [2013]) extended Sobol' indices to multivariate or functional outputs. Let us also cite Alexanderian et al. [2020] focused on time-dependent outputs. However, the so-called aggregated Sobol' indices cannot easily be interpreted if the input parameters are dependent. Following these papers, we introduce and study the properties of what we call aggregated Shapley effects, which can be interpreted in both the framework of independent and dependent input parameters. If the output dimension is high (as it is the case, e.g., when considering a fine discretization of a functional output), a dimension reduction can be applied as a preliminary step to the computation of aggregated Shapley effects. We use the Karhunen-Loève (KL) expansion as in [Lamboni et al., 2009, Alexanderian et al., 2020]. More precisely to perform KL expansion, we use the functional principal component analysis proposed by Yao et al. [2005]. Note that the extension of Shapley effects to multivariate outputs has been early used for nuclear safety application in [Delipei, 2019]. In this paper, we propose a proper definition of aggregated Shapley indices, we study their invariance properties and we propose a given data estimation procedure, extending the nearest neighbor approach introduced in [Broto et al., 2020] and a block bootstrap procedure to build confidence intervals. By given data procedure, we mean that a single sample of size n is needed for the estimation of all

Shapley effects. Eventually, we demonstrate the usefulness of this methodology in a complex avalanche application where the set of output simulations is constrained in order to keep only meaningful snow avalanche events. More precisely, we perform the GSA of our avalanche flow dynamic model in two different settings: (i) little knowledge on the input parameter probability distribution, and (ii) well-calibrated input parameter distribution (Eckert et al. [2010]). Notably, this methodology which could be applied to many other applications where samples are constrained to satisfy physical or user constraints opens a new field of applications to GSA.

The paper is organized as follows. In Section 5.2, aggregated Shapley effects and their main properties are described. In Section 5.3, we propose an estimator for aggregated Shapley effects in a given data framework by extending the Monte-Carlo nearest-neighbor estimator of scalar Shapley effects introduced in [Broto et al., 2020]. At the end of the section, we describe the functional principal components analysis algorithm to perform model dimension reduction proposed by Yao et al. [2005]. In Section 5.4, we propose a bootstrap algorithm to construct confidence intervals of the scalar and aggregated Shapley effect estimations based on [Benoumechiara and Elie-Dit-Cosaque, 2019]. In Section 5.5, we test our estimation procedure on two toy models: a multivariate linear Gaussian model and the mass-spring model. Finally in Section 5.6, our GSA procedure is applied to our avalanche dynamic model. We discuss our conclusions and perspectives in Section 5.7.

5.2 Aggregated Shapley effects

Shapley effects are sensitivity measures to quantify input importance proposed by Owen [2014]. These measures are particularly useful when inputs are dependent. Shapley effects are based in the concept of Shapley value, introduced in the framework of game theory [Shapley, 1953], which consists into dividing a game gain among a group of players in an equitable way. As sensitivity measures, Shapley effects consider model inputs as players and output variance as game function. Shapley effects can be naturally extended to multivariate output models by following the ideas presented in [Gamboa et al., 2013] and [Lamboni et al., 2009] to generalize Sobol' indices to multivariate output models (see also Alexanderian et al. [2020] for time-dependent models). We call these new sensitivity measures aggregated Shapley effects.

5.2.1 Definition

Let us define $\mathbf{Y} = (Y_1, \dots, Y_j, \dots, Y_p) = f(\mathbf{X})$ the p multivariate output of a model f that depends on d random inputs $\mathbf{X} = (X_1, \dots, X_d)$. The inputs are defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\mathbf{X}})$ and $f \in \mathbb{L}^2(\mathbb{P}_{\mathbf{X}})$. For any $\mathfrak{u} \subseteq \{1, \dots, d\}$, let us define $-\mathfrak{u} = \{1, \dots, d\} \setminus \mathfrak{u}$ its complement. We set $\mathbf{X}_{\mathfrak{u}} = (X_i)_{i \in \mathfrak{u}}$. Note that the inputs are not necessary independent. In the framework of our application to avalanche modeling, the model produces outputs of the form $\mathbf{Y} = (Y_1 = f(s_1, \mathbf{X}), \dots, Y_p = f(s_p, \mathbf{X}))$, with $s_1, \dots, s_p \in \mathbb{R}$ the p discretization points along the avalanche corridor.

In this section we recall the definition and main properties of the Shapley value, on which the definition of Shapley effects is based. Given a set of d players in a coalitional game and a charateristic function $\text{val} : 2^d \to \mathbb{R}$, $\text{val}(\emptyset) = 0$, the Shapley value $(\phi_1, \dots, \phi_d)$ is the only distribution of the total gains $\text{val}(\{1, \dots, d\})$ to the players satisfying the

desirable properties listed below:

1. (Efficiency) $\sum_{i=1}^{d} \phi_i = \text{val}(\{1, \dots, d\})$.
2. (Symmetry) If $\text{val}(\mathfrak{u} \cup \{i\}) = \text{val}(\mathfrak{u} \cup \{\ell\})$ for all $\mathfrak{u} \subseteq \{1, \dots, d\} - \{i, j\}$, then $\phi_i = \phi_\ell$.
3. (Dummy) If $\text{val}(\mathfrak{u} \cup \{i\}) = \text{val}(\mathfrak{u})$ for all $\mathfrak{u} \subseteq \{1, \dots, d\}$, then $\phi_i = 0$.
4. (Additivity) If val and val' have Shapley values ϕ and ϕ' respectively, then the game with characteristic function val + val' has Shapley value $\phi_i + \phi'_i$ for $i \in \{1, \dots, d\}$.

It is proved in [Shapley, 1953] that according to the Shapley value, the amount that player i gets given a coalitional game (val, d) is:

$$\phi_i = \frac{1}{d} \sum_{\mathfrak{u} \subseteq -\{i\}} \binom{d-1}{|\mathfrak{u}|}^{-1} (\text{val}(\mathfrak{u} \cup \{i\}) - \text{val}(\mathfrak{u})) \ \ \forall i \in \{1, \dots, d\}. \tag{5.1}$$

The Shapley value also satisfies the linearity property:

5. (Linearity) Let $\lambda \in \mathbb{R}$, if λval and val have Shapley values ϕ' and ϕ, then $\phi'_i = \lambda \phi_i$ for all $i \in \{1, \dots, d\}$.

The linearity property is a direct consequence from (5.1) and it is used to prove some of the nice properties of aggregated Shapley effects (see Propositions 1 and 2 further).

The Shapley effects are defined by considering the characteristic function of the game as:

$$\text{val}_j(\mathfrak{u}) = \frac{\text{Var}(\mathbb{E}(Y_j | \mathbf{X}_\mathfrak{u}))}{\text{Var}(Y_j)}, \ \mathfrak{u} \subseteq \{1, \dots, d\} \tag{5.2}$$

in (5.1). Thus, the scalar Shapley effect of input i in output j is defined as:

$$\text{Sh}_i^j = \frac{1}{d \, \text{Var}(Y_j)} \sum_{\mathfrak{u} \subseteq -\{i\}} \binom{d-1}{|\mathfrak{u}|}^{-1} (\text{Var}(\mathbb{E}(Y_j | \mathbf{X}_{\mathfrak{u} \cup i})) - \text{Var}(\mathbb{E}(Y_j | \mathbf{X}_\mathfrak{u}))). \tag{5.3}$$

Shapley effects can be naturally extended to models with multivariate outputs following ideas from [Gamboa et al., 2013] and [Lamboni et al., 2009] where authors proposed to extend Sobol' indices to multivariate outputs. Aggregated Shapley effect of an input i is then defined as:

$$\text{GSh}_i = \frac{\sum_{j=1}^{p} \text{Var}(Y_j) \text{Sh}_i^j}{\sum_{j=1}^{p} \text{Var}(Y_j)}, \tag{5.4}$$

where Sh_i^j is the scalar Shapley effect of input X_i in output Y_j. This sensitivity measure is a weighted sum of the scalar Shapley effects where weights correspond to the proportion of the variance of each output over the sum of all individual variances.

5.2.2 Properties

In this section, we prove some nice properties of aggregated Shapley effects.

Proposition 1. *The aggregated Shapley effects* GSh_i, $i \in \{1, \dots, d\}$, *correspond to the Shapley value with characteristic function defined as:*

$$val(\mathfrak{u}) = \frac{\sum_{j=1}^{p} Var(Y_j) val_j(\mathfrak{u})}{\sum_{j=1}^{p} Var(Y_j)}, \ \mathfrak{u} \subseteq \{1, \dots, d\}. \tag{5.5}$$

Proof. The proof is straightforward. It is a direct consequence of the linearity and additivity properties of the Shapley value. Let $i \in \{1, \dots, d\}$ and $j \in \{1, \dots, p\}$. The characteristic function val_j (see (5.2)) has Shapley value Sh_i^j, $i \in \{1, \dots, d\}$. Thanks to the linearity and additivity properties (see properties 4. and 5. of the Shapley value), the characteristic function $\mathfrak{u} \subseteq \{1, \dots, d\} \mapsto \frac{\sum_{j=1}^{p} \text{Var}(Y_j)\text{val}_j(\mathfrak{u})}{\sum_{i=1}^{p} \text{Var}(Y_j)}$ leads to the Shapley value $\frac{\sum_{j=1}^{p} \text{Var}(Y_j)\text{Sh}_i^j}{\sum_{i=1}^{p} \text{Var}(Y_j)}$. □

The characteristic function (5.5) can be written in matricial form:

$$\text{val}(\mathfrak{u}) = \frac{\sum_{j=1}^{p} \text{Var}(Y_j)\text{val}_j(\mathfrak{u})}{\sum_{j=1}^{p} \text{Var}(Y_j)} = \frac{\sum_{j=1}^{p} \text{Var}(\mathbb{E}(Y_j|X_\mathfrak{u}))}{\sum_{j=1}^{p} \text{Var}(Y_j)} = \frac{tr(\Sigma_\mathfrak{u})}{tr(\Sigma)} \tag{5.6}$$

where $\Sigma_\mathfrak{u}$ is the covariance matrix of $\mathbb{E}(\mathbf{Y}|X_\mathfrak{u})$ and Σ is the covariance matrix of $\mathbf{Y}$. Note that the characteristic function val of aggregated Shapley effects corresponds to the definition of the aggregated Sobol' indices introduced in [Lamboni et al., 2009, Gamboa et al., 2013]. In the next proposition, we prove that aggregated Shapley effects accomplish the natural requirements for a sensitivity measure mentioned in Proposition 3.1 in [Gamboa et al., 2013].

Proposition 2. *Let $i \in \{1, \dots d\}$. The following items hold true.*

i. $0 \leq \text{GSh}_i \leq 1$.

ii. GSh_i *is invariant by left-composition by any nonzero scaling of f, which means, for any $\lambda \in \mathbb{R}$, the aggregated Shapley effect GSh'_i of $\lambda f(\boldsymbol{X})$ is GSh_i.*

iii. GSh_i *is invariant by left-composition of f by any isometry of $\mathbb{R}^p$, which means, for any $O \in \mathbb{R}^{p\times p}$ such that $O^tO = I$, the aggregated Shapley effect GSh'_i of $Of(\boldsymbol{X})$ is GSh_i for all $i \in \{1, \dots, d\}$.*

Proof. i. As for all $j \in \{1, \dots, p\}$ $0 \leq \text{Sh}_i^j \leq 1$ and as the sum of the non negative weights $\text{Var}(Y_j)/\sum_{\ell=1}^{p} \text{Var}(Y_\ell)$ is one, we deduce that $0 \leq \text{GSh}_i \leq 1$. *ii.* Note that GSh'_i can be written as $\text{GSh}'_i = \sum_{j=1}^{p} \text{Var}(\lambda Y_j)\text{Sh}_i'^j / \sum_{j=1}^{p} \text{Var}(\lambda Y_j)$, where $\text{Sh}_i'^j$ is the Shapley effect associated to the characteristic function val'_j. Note that $\text{val}'_j(\mathfrak{u}) = \text{Var}(\mathbb{E}(\lambda Y_j|X_\mathfrak{u}))/\text{Var}(\lambda Y_j) = \text{val}_j(\mathfrak{u})$ for all $\mathfrak{u} \subseteq \{1, \dots, d\}$. Thus, $\text{Sh}_i'^j = \text{Sh}_i^j$ from where $\text{GSh}'_i = \text{GSh}_i$ which means the aggregated Shapley effect is invariant by any nonzero scaling of f. *iii.* Let us write $g(\mathbf{X}) = Of(\mathbf{X}) = O\mathbf{Y} = \mathbf{U}$. The characteristic function associated to the aggregated Shapley effect GSh'_i of $\mathbf{U}$ is then (see Equation (5.6)) $\text{val'}(\mathfrak{u}) = tr(\Sigma_\mathfrak{u}^\mathbf{U})/tr(\Sigma^\mathbf{U})$ where $\Sigma_\mathfrak{u}^\mathbf{U}$ is the covariance matrix of $\mathbb{E}(\mathbf{U}|X_\mathfrak{u})$ and $\Sigma^\mathbf{U}$ is the covariance matrix of $\mathbf{U}$. Then,

$$\text{val'}(\mathfrak{u}) = \frac{tr(\Sigma_\mathfrak{u}^\mathbf{U})}{tr(\Sigma^\mathbf{U})} = \frac{tr(O\Sigma_\mathfrak{u}^\mathbf{Y}O^t)}{tr(O\Sigma^\mathbf{Y}O^t)} = \frac{tr(\Sigma_\mathfrak{u}^\mathbf{Y})}{tr(\Sigma^\mathbf{Y})} = \text{val}(\mathfrak{u}) \text{ for all } \mathfrak{u} \subseteq \{1, \dots, d\}.$$

As $\text{val}(\mathfrak{u})$ has a unique Shapley value GSh_i, $\text{val}'(\mathfrak{u})$ has Shapley value GSh_i which proves that $\text{GSh}'_i = \text{GSh}_i$ for all $i \in \{1, \dots, d\}$. □

In this section, we have proven that aggregated Shapley effects are sensitivity measures. In the next section, we describe the estimation procedure we propose for aggregated Shapley effects, based the estimation procedure of scalar Shapley effects proposed in [Broto et al., 2020, Section 6] when observing an i.i.d. sample of $(\mathbf{X}, \mathbf{Y})$. Such a procedure, which does not require a specific form for the design of experiments is also called given data procedure.

5.3 Estimation procedure for scalar and aggregated Shapley effects

The aggregated Shapley effect estimation procedure we propose in this section is based on the given data estimation procedure of the scalar Shapley effects introduced in [Broto et al., 2020, Section 6.1.1.]. In the application we consider in Section 5.6, the standard pick-freeze estimation procedure (see, e.g., Janon et al. [2014b]) cannot be used as it is based on a specific pick-freeze design of experiments. We recall indeed that input parameters in our application are first sampled independently, then some combinations are rejected as they do not lead to snow avalanche avalanche simulations that are both physically meaningful and of interest for risk purposes. It is the reason why we focus our attention to the given data estimation procedure of scalar Shapley effects introduced in [Broto et al., 2020, Section 6.1.1.]. For sake of clarity, we first present the estimation procedure for scalar Shapley effects in Subsection 5.3.1 before extending it to the estimation of aggregated Shapley effects in Subsection 5.3.2.

5.3.1 Given data estimator of scalar Shapley effects proposed by [Broto et al., 2020]

Let $\mathfrak{u} \subset \{1, \dots d\}$. Let $-\mathfrak{u}$ denote the complementary set of $\mathfrak{u}$ in $\{1, \dots, d\}$. As noticed in [Song et al., 2016, Theorem 1], replacing the characteristic function $\mathfrak{u} \mapsto \tilde{c}_j(\mathfrak{u}) = \mathrm{Var}(\mathbb{E}(Y_j|\mathbf{X}_\mathfrak{u}))$ by $\mathfrak{u} \mapsto c_j(\mathfrak{u}) = \mathbb{E}(\mathrm{Var}(Y_j|\mathbf{X}_{-\mathfrak{u}}))$ in Equation (5.3) does not change the definition of Shapley effects. The estimation of the Shapley effects Sh_i^j, $1 \leq i \leq d$, relies on the estimation of the $c_j(\mathfrak{u})$ (resp. $\tilde{c}_j(\mathfrak{u})$) for all $\mathfrak{u} \subseteq \{1, \dots, d\}$. To estimate $c_j(\mathfrak{u}) = \mathbb{E}(\mathrm{Var}(Y_j|\mathbf{X}_{-\mathfrak{u}}))$ (resp. $\tilde{c}_j(\mathfrak{u})$), one can use a double Monte Carlo estimator: an inner loop to estimate the variance (resp. the expectation) and an outer loop to estimate the expectation (resp. the variance). However, Song et al. [2016] pointed out that the double Monte Carlo estimator of $\tilde{c}_j(\mathfrak{u})$ can suffer from a non neglectable bias if the inner loop sample is small, while in contrast the double Monte Carlo estimator of $c_j(\mathfrak{u})$ is unbiased for any sample size. For that reason, we turned to the double Monte Carlo estimator of scalar Shapley effects introduced in [Broto et al., 2020, Section 6.1.1.] based on the double Monte Carlo estimation of $c_j(\mathfrak{u})$ with a subset aggregation procedure. They also introduce an adaptation of this estimator to the given data framework, well suited to our application as it is based on a given and unstructured sample.

The estimator works as follows: assume we are given a n independent random sample $(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})_{1 \leq k \leq n}$ of $(\mathbf{X}, \mathbf{Y})$, then for $1 \leq j \leq p$, $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1, \dots, d\}$, let $1 \leq N_\mathfrak{u} \leq n$ be an integer, let $(s_\ell)_{1 \leq \ell \leq N_\mathfrak{u}}$ be a sample of uniformly distributed integers in $\{1, \dots, n\}$ (with or without replacement). The inner Monte Carlo loop estimates the conditional variance $\mathrm{Var}(Y_j|\mathbf{X}_{-\mathfrak{u}})$ using N_I nearest neighbors of each $\mathbf{X}_{-\mathfrak{u}}^{s_\ell}$, $1 \leq \ell \leq N_\mathfrak{u}$. If all the $N_\mathfrak{u}$ were

chosen equal to n, the computation cost for the search of the nearest neighbors would be in $\mathcal{O}\big(n(2^d-2)\big)$. The integers $(N_{\mathfrak{u}})_{\emptyset\subsetneq\mathfrak{u}\subsetneq\{1,\dots,d\}}$ are related to the estimation accuracy of the conditional variances $(\mathrm{Var}(Y_j|\mathbf{X}_{-\mathfrak{u}}))_{\emptyset\subsetneq\mathfrak{u}\subsetneq\{1,\dots,d\}}$ and are referred as the accuracies hereafter. To reduce the estimation cost to a prescribed computational budget N_{tot}, the authors in [Broto et al., 2020] proposed to choose the set of accuracies that minimize the estimation variance for the prescribed total cost $N_{tot} = \sum_{\emptyset\subsetneq\mathfrak{u}\subsetneq\{1,\dots,d\}} N_{\mathfrak{u}}$. In the framework where it is possible to sample from the conditional distributions of the input vector, it can be proven under a strong assumption (see [Broto et al., 2020, Proposition 4.2., Remark 4.3.]) that there exists an optimal allocation of the accuracies, namely $N_{\mathfrak{u}}^* = \left\lfloor N_{tot}\binom{d}{|\mathfrak{u}|}^{-1}(d-1)^{-1}\right\rfloor$, for all $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1,\dots,d\}$. Even if this strong assumption, that all the conditional elements $\mathbb{E}(\mathrm{Var}(Y|\mathbf{X}_{-\mathfrak{u}}))$, $\mathfrak{u}\subseteq\{1,\dots,d\}$, $\mathfrak{u}\neq\emptyset$, $\mathfrak{u}\neq\{1,\dots,d\}$ can be estimated with the same variance, does not hold true in the given data framework, the authors present numerical experiments in [Broto et al., 2020, Section 6.3], showing that if one chooses the accuracies $N_{\mathfrak{u}}^*$ for all $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1,\dots,d\}$, one gets reasonable results, at least on the test case they consider. Note that the estimation cost in terms of number of model evaluations is n while the cost in terms of nearest-neighbor search is N_{tot}. In Section 5.8, we compare the choice of the accuracy $N_{\mathfrak{u}}^*$ for all $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1,\dots,d\}$ to a uniform allocation $N_{\mathfrak{u}} = \lfloor N_{tot}/(2^d-2)\rfloor$ for all $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1,\dots,d\}$ for a test case in dimension $d=10$. The allocation $N_{\mathfrak{u}}^*$ suggested in [Broto et al., 2020] leads to better results, at least on that experiment. Regarding the number of neighbors, we analyzed the impact of using different numbers of neighbors N_I on a linear model with $d=2$. The results are presented in Section 5.9. We see in Figure 5.14 that there is no clear evidence if $N_I \geq 3$ is always the best choice. However in the following we choose $N_I = 3$ following the recommendations in [Song et al., 2016], where the authors prove in the non given data framework that choosing $N_I = 3$ is better than choosing $N_I = 2$ and that the gain from increasing N_I from 3 decreases as N_I increases.

Let us write more precisely the inner Monte Carlo loop to estimate the conditional variance $\mathrm{Var}(Y_j|\mathbf{X}_{-\mathfrak{u}})$:

$$\widehat{E}^j_{\mathfrak{u},s_\ell} = \frac{1}{N_I-1}\sum_{i=1}^{N_I}\left(\sum_{v:\mathbf{X}^v_{-\mathfrak{u}}\in\mathcal{B}_{-\mathfrak{u},\ell}} Y_j^v - \frac{1}{N_I}\bar{Y}_{s_\ell}\right)^2 \text{ with } \bar{Y}_{s_\ell} = \frac{1}{N_I}\sum_{v:\mathbf{X}^v_{-\mathfrak{u}}\in\mathcal{B}_{-\mathfrak{u},\ell}} Y_j^v, \tag{5.7}$$

with $\mathcal{B}_{-\mathfrak{u},\ell}$ the set of N_I closest neighbors (chosen two by two distinct) of $\mathbf{X}^{s_\ell}_{-\mathfrak{u}}$.

The closest neighbors are computed using the Euclidean distance in $\mathbb{R}^{k_{\mathfrak{u}}}$ with $k_{\mathfrak{u}}$ denoting the cardinal of $-\mathfrak{u}$. Then, we compute the outer Monte Carlo loop to estimate the expectation:

$$\widehat{c}_j(\mathfrak{u}) = \frac{1}{N_{\mathfrak{u}}^*}\sum_{\ell=1}^{N_{\mathfrak{u}}^*}\widehat{E}^j_{\mathfrak{u},s_\ell}. \tag{5.8}$$

Note that $c_j(\emptyset) = 0$ and that $c_j(\{1,\dots,d\}) = \mathrm{Var}(Y_j)$, which is assumed to be known in [Broto et al., 2020], and that is estimated by the empirical variance in the present paper. In [Broto et al., 2020, Theorem 6.6.], it is proved that under mild assumptions, $\widehat{c}_j(\mathfrak{u})$ converges in probability to $c_j(\mathfrak{u})$ as n and $N_{\mathfrak{u}}^*$ go to ∞. Finally, the subset aggregation

procedure consists in estimating scalar Shapley effects by plugging (5.8) in Equation (5.3):

$$\widehat{\mathrm{Sh}}_i^j = \frac{1}{d\,\widehat{\sigma}_j^2} \sum_{\mathfrak{u} \subseteq -i} \binom{d-1}{|\mathfrak{u}|}^{-1} (\widehat{c}_j(\mathfrak{u} \cup \{i\}) - \widehat{c}_j(\mathfrak{u})) \tag{5.9}$$

where $\hat{\sigma}_j^2$ is the empirical estimator of $\mathrm{Var}(Y_j)$.

5.3.2 Estimator of the aggregated Shapley effects

Given scalar Shapley effect estimators whose definition is recalled in the previous section, we propose to estimate the aggregated Shapley effects by:

$$\widehat{\mathrm{GSh}}_i = \frac{\sum_{j=1}^p \widehat{\sigma}_j^2 \widehat{\mathrm{Sh}}_i^j}{\sum_{j=1}^p \widehat{\sigma}_j^2} = \frac{1}{d\sum_{j=1}^p \widehat{\sigma}_j^2} \sum_{j=1}^p \sum_{\mathfrak{u} \subseteq -i} \binom{d-1}{|\mathfrak{u}|}^{-1} (\widehat{c}_j(\mathfrak{u} \cup \{i\}) - \widehat{c}_j(\mathfrak{u})), \tag{5.10}$$

with $\hat{\sigma}_j^2$ the empirical estimator of $\mathrm{Var}(Y_j)$ and $\widehat{c}_j(\mathfrak{u})$ defined by (5.8).

5.3.3 Dimension reduction: functional principal component analysis

If model f is space or time-dependent, inspired by Alexanderian et al. [2020] and Lamboni et al. [2009], we perform a functional principal component analysis (fPCA) based on the Karhunen-Loève (KL) expansion to obtain a low-rank model representation. Then we compute the aggregated Shapley effects for the vectorial output composed with the first principal components. This parsimonious representation of the functional output usually allows easier interpretation for real applications. To perform KL expansion, we use the principal component analysis through conditional expectation (PACE) method introduced in Yao et al. [2005]. More precisely, we have a collection of n independent trajectories of a smooth random function $f(.,\mathbf{X})$ with unknown mean $\mu(s) = \mathbb{E}(f(s,\mathbf{X})), s \in \tau$, where $\tau \subseteq \mathbb{R}$ is a bounded and closed interval, and covariance function $G(s_1,s_2) = \mathrm{Cov}(f(s_1,\mathbf{X}), f(s_2,\mathbf{X})), s_1, s_2 \in \tau$. We assume that G has a L^2 orthogonal expansion in terms of eigenfunctions ξ_k and non increasing eigenvalues λ_k such that:

$$G(s_1,s_2) = \sum_{k\geq 1} \lambda_k \xi_k(s_1,\mathbf{X}) \xi_k(s_2,\mathbf{X}), s_1, s_2 \in \tau.$$

The KL orthogonal expansion of $f(s,\mathbf{X})$ is:

$$f(s,\mathbf{X}) = \mu(s) + \sum_{k\geq 1} \alpha_k(\mathbf{X}) \xi_k(s) \approx \mu(s) + \sum_{k=1}^q \alpha_k(\mathbf{X}) \xi_k(s), s \in \tau, \tag{5.11}$$

where $\alpha_k(\mathbf{X}) = \int_\tau f(s,\mathbf{X}) \xi_k(s) ds$ is the k-th functional principal component (fPC) and q is a truncation level. Following the authors in [Yao et al., 2005], we estimate first $\hat{\mu}(s)$ using local linear smoothers and $\widehat{G}(s_1,s_2)$ using local linear surface smoothers (Fan and Gijbels [1996]), then we estimate eigenfunctions and eigenvalues by solving the following integral equations:

$$\int_\tau \widehat{G}(s_1,s)\widehat{\xi}_k(s_1)ds_1 = \widehat{\lambda}_k\,\widehat{\xi}_k(s), s\in\tau,$$

with $\int_\tau \widehat{\xi}(s)ds = 1$ and $\int_\tau \widehat{\xi}_k(s)\widehat{\xi}_m(s) = 0$ for all $m \neq k \leq q$. The problem is solved by using a discretization of the smoothed covariance (see further details in [Rice and Silverman, 1991] and [Capra and Müller, 1997]). Finally, fPCs $\widehat{\alpha}_k(\mathbf{X}) = \int_\tau f(s,\mathbf{X})\widehat{\xi}_k(s)ds$ are solved by numerical integration.

Aggregated Shapley effects are finally estimated using the estimated low rank KL model representation with truncation level q, in other words, they are computed with only the q first estimated fPCs:

$$\widetilde{GSh_i} = \frac{1}{d\sum_{k=1}^q \widehat{\lambda}_k}\sum_{k=1}^q\sum_{\mathfrak{u}\subseteq -i}\binom{d-1}{|\mathfrak{u}|}^{-1}\left(\mathbb{E}(\mathrm{Var}(\widehat{\alpha}_k(\mathbf{X})|\mathbf{X}_{\mathfrak{u}\cup\{i\}})) - \mathbb{E}(\mathrm{Var}(\widehat{\alpha}_k(\mathbf{X})|\mathbf{X}_{\mathfrak{u}}))\right). \tag{5.12}$$

Remark. *(5.12) can be estimated using (5.10).*

Remark. *To avoid splitting our sample, we use the same sample to estimate the KL representation and to estimate aggregated Shapley effects. The comparison with the alternative procedure consisting in splitting the initial sample in two subsamples, one to estimate the KL representation and the other one to estimate aggregated Shapley effects is presented and discussed for the mass-spring model in Section 5.10. Based on these numerical results, we decided to apply the procedure without splitting in the following and in particular on our snow avalanche application.*

5.4 Bootstrap confidence intervals with percentile bias correction

Confidence intervals are a valuable tool to quantify uncertainty in estimation. We consider non parametric bootstrap confidence intervals with bias percentile correction (see, e.g., Efron [1981], Efron and Tibshirani [1986]). More precisely, we propose to construct confidence intervals, with a block bootstrap procedure, following ideas in [Benoumechiara and Elie-Dit-Cosaque, 2019]. Indeed, bootstrap by blocks is necessary to preserve the nearest-neighbor structure in Equation (5.7) and to avoid potential equalities in distance (see Assumption 6.3 in [Broto et al., 2020]). We describe in Algorithm 1 how to create B bootstrap samples for scalar Shapley effect estimators $\widehat{\mathrm{Sh}}_i^j$ and aggregated Shapley effect estimators $\widehat{\mathrm{GSh}}_i$, and then we describe the percentile bias correction method.

If model output is scalar, only Steps 1 to 3 of Algorithm 1 should be used. The block bootstrap procedure is described by Steps 3.1 to 3.3. Also, the same sample $(\mathbf{X},\mathbf{Y})$ is used to estimate the variance of the outputs Y_j, $1\leq j\leq p$, and the Shapley effects.

For $1\leq i\leq d$, $1\leq j\leq p$, let $\mathcal{R}_i = \{\widehat{\mathrm{GSh}}_i^{(1)},\ldots,\widehat{\mathrm{GSh}}_i^{(B)}\}$ and $\mathcal{R}_i^j = \{\widehat{\mathrm{Sh}}_i^{j,(1)},\ldots,\widehat{\mathrm{Sh}}_i^{j,(B)}\}$, the bias-corrected percentile method presented in [Efron and Tibshirani, 1986] is applied. Let us denote by Φ the standard normal cumulative distribution function and by Φ^{-1} its inverse. A bias correction constant z_0, estimated as

Algorithm 1 B bootstrap samples for $\widehat{\mathrm{Sh}}_i^j$ and $\widehat{\mathrm{GSh}}_i$

Inputs: (i) A n i.i.d. random sample $(\mathbf{X}^k, \mathbf{Y}^k)_{k\in\{1,\dots,n\}}$ with $\mathbf{X}^k \in \mathbb{R}^d$ and $\mathbf{Y}^k \in \mathbb{R}^p$. (ii) For each $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1,\dots,d\}$, a $N_\mathfrak{u}$ random sample $(s_\ell)_{1\le\ell\le N_\mathfrak{u}}$ from $\{1,\dots,n\}$.

Outputs: B bootstrap samples for $\widehat{\mathrm{Sh}}_i^j$ and $\widehat{\mathrm{GSh}}_i$.

for $b = 1$ **to** $b = B$ **do**

1. Create a n bootstrap sample $\mathbf{Y}^{(b)}$ by sampling with replacement from the rows of $\mathbf{Y}$.
2. Compute, for $1 \le j \le p$, $\widehat{\sigma}_j^{2,(b)}$ the empirical variance of $\mathbf{Y}_j^{(b)}$.
3. For each $j \in \{1,\dots,p\}$:
   3.1. For all $\mathfrak{u}$ and for all $(s_\ell)_{1\le\ell\le N_\mathfrak{u}}$ compute $\widehat{E}^j_{\mathfrak{u},s_\ell}$ using (5.7).
   3.2. For all $\mathfrak{u}$, create a $N_\mathfrak{u}$ bootstrap sample $\widehat{E}^{j,(b)}_{\mathfrak{u},s_\ell}$ by sampling with replacement from $\left(\widehat{E}^j_{\mathfrak{u},s_\ell}\right)_{1\le\ell\le N_\mathfrak{u}}$ computed in Step 3.1.
   3.3. Compute $\widehat{c_j}^{(b)}(\mathfrak{u}) = 1/N_\mathfrak{u} \sum_{\ell=1}^{N_\mathfrak{u}} \widehat{E}^{j,(b)}_{\mathfrak{u},s_\ell}$ for all $\mathfrak{u}$ using (5.8).
   3.4. Compute the b bootstrap sample of $\widehat{\mathrm{Sh}}_i^j$ according to (5.9):

$$\widehat{\mathrm{Sh}}_i^{j,(b)} = \frac{1}{d\,\widehat{\sigma}_j^{2,(b)}} \sum_{\mathfrak{u}\subseteq -i} \binom{d-1}{|\mathfrak{u}|}^{-1} \left(\widehat{c_j}^{(b)}(\mathfrak{u}\cup\{i\}) - \widehat{c_j}^{(b)}(\mathfrak{u})\right).$$

4. Compute the b bootstrap sample of $\widehat{\mathrm{GSh}}_i$ using (5.10):

$$\widehat{\mathrm{GSh}}_i^{(b)} = \frac{1}{d\sum_{j=1}^p \widehat{\sigma}_j^{2,(b)}} \sum_{j=1}^p \sum_{\mathfrak{u}\subseteq -i} \binom{d-1}{|\mathfrak{u}|}^{-1} \left(\widehat{c_j}^{(b)}(\mathfrak{u}\cup\{i\}) - \widehat{c_j}^{(b)}(\mathfrak{u})\right).$$

end for

$\hat{z}_0 = \Phi^{-1}(\#\{\widehat{\mathrm{GSh}}_i^{(b)} \in \mathcal{R}_i \text{ s. t. } \widehat{\mathrm{GSh}}_i^{(b)} \leq \widehat{\mathrm{GSh}}_i\}/B)$ is computed (similar for $\widehat{\mathrm{Sh}}_i^j$). Then, the corrected quantile estimate $\hat{q}(\beta)$ for $\beta \in]0,1[$ is defined as $\hat{q}_i(\beta) = \Phi(2\hat{z}_0 + z_\beta)$, where z_β satisfies $\Phi(z_\beta) = \beta$. Corrected bootstrap confidence interval of level $1-\alpha$ is estimated by the interval whose endpoints are $\hat{q}_i(\alpha/2)$ and $\hat{q}_i(1-\alpha/2)$.

To guarantee the validity of the previous BC corrected confidence interval, there must exist an increasing transformation g, $z_0 \in \mathbb{R}$ and $\tau > 0$ such that $g(\widehat{\mathrm{GSh}}_i) \sim \mathcal{N}(\mathrm{GSh}_i - \tau z_0, \tau^2)$ and $g(\widehat{\mathrm{GSh}}_i^*) \sim \mathcal{N}(\widehat{\mathrm{GSh}}_i - \tau z_0, \tau^2)$ where $\widehat{\mathrm{GSh}}_i^*$ is the bootstrapped $\widehat{\mathrm{GSh}}_i$ for fixed sample (see Efron [1981]). Normality hypothesis can be tested using traditional normality tests as Shapiro test or using graphical methods as empirical normal quantile-quantile plots. In our application and test cases, we observed that g can be chosen as the identity. The author in [Efron, 1987] advises the use of the bias correction to improve the accuracy of bootstrap intervals. In our application and test cases, there was no much difference between the percentile and BC intervals. However, as a good practice we performed the bias correction, especially since the algorithm is simple to implement.

To validate the coverage accuracy of the bootstrap intervals whose construction is described in Algorithm 1, we compute the empirical probability of coverage (POC) of simultaneous intervals using Bonferroni correction for the test cases where the true value of the effects is known. The POC with Bonferroni correction is the probability that the interval $[\hat{G}^{-1} \circ \hat{q}_i(\alpha/(2d)), \hat{G}^{-1} \circ \hat{q}_i(1-\alpha/(2d))]$, where $\hat{G}$ is the cumulative distribution function for $\widehat{\mathrm{GSh}}_i$, contains GSh_i for all $i \in \{1, \ldots, d\}$ simultaneously. To be more precise, if the confidence intervals are computed in N independent samples of size n of $(\mathbf{X}, \mathbf{Y})$, the POC is estimated as $\widehat{POC} = \sum_{k=1}^{N} w^k/N$, where w^k is equal to 1 if $\hat{G}^{-1} \circ \hat{q}_i(\alpha/(2d)) \leq \mathrm{GSh}_i \leq \hat{G}^{-1} \circ \hat{q}_i(1-\alpha/(2d))$ for all i, and 0 otherwise.

5.5 Test cases

In this section, we numerically study the performance of the estimation procedure and the probability coverage of the boostrap confidence intervals we introduced in the previous section. We consider two test cases: a multivariate linear Gaussian model and the functional mass-spring model proposed in the work of Gamboa et al. [2013]. To estimate the scalar Shapley effects, we use the function `shapleySubsetMc` of the `R` package `sensitivity` corresponding to the estimation procedure defined by (5.8), (5.7) and (5.9). Functional PCA is performed using the `R` package `FPCA` [Chen et al., 2020].

5.5.1 Multivariate linear Gaussian model in dimension $d = 2$

We consider a multivariate linear model with two Gaussian inputs based on the example from [Iooss and Prieur, 2019]. For this toy function, there is an analytical expression of the scalar and aggregated Shapley effects (see Iooss and Prieur [2019]).

The model f is defined as $\mathbf{Y} = f(\mathbf{X}) = B^T\mathbf{X}$ with $\mathbf{X} \sim \mathcal{N}(\mu, \Gamma)$, $\Gamma \in \mathbb{R}^{d\times d}$ a positive-definite matrix and $\mathbf{B} \in \mathbb{R}^{d\times p}$. In this example, we consider $d = 2$ and $p = 3$ which means $\mathbf{Y} = (Y_1, Y_2, Y_3)$. The variance of the centered random variables X_1 and X_2 are equal to $\sigma_1^2 = 1$ and $\sigma_2^2 = 3$, respectively and their correlation $\rho = 0.4$. Thus the covariance matrix of $\mathbf{X}$ is given by:

$$\Gamma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.69 \\ 0.69 & 3 \end{bmatrix},$$

and the coefficients of $B = (\beta_{ij}) \in \mathbb{R}^{2\times 3}$ are chosen as:

$$B = \begin{bmatrix} 1 & 4 & 0.1 \\ 1 & 3 & 0.9 \end{bmatrix}.$$

The variance of the output Y_j with $j \in \{1,2,3\}$ is $\sigma^2_{Y_j} = \beta^2_{1j}\sigma_1^2 + 2\rho\beta_{1j}\beta_{2j}\sigma_1\sigma_2 + \beta^2_{2j}\sigma_2^2$. The scalar Shapley effects are:

$$\begin{aligned} \sigma^2_{Y_j}\phi_1^j &= \beta^2_{1j}\sigma_1^2\left(1 - \frac{\rho^2}{2}\right) + \rho\beta_{1j}\beta_{2j}\sigma_1\sigma_2 + \beta_2^2\sigma_2^2\frac{\rho^2}{2}, \\ \sigma^2_{Y_j}\phi_2^j &= \beta^2_{2j}\sigma_2^2\left(1 - \frac{\rho^2}{2}\right) + \rho\beta_{1j}\beta_{2j}\sigma_1\sigma_2 + \beta_1^2\sigma_1^2\frac{\rho^2}{2}. \end{aligned}$$

Then, the aggregated Shapley effects for $i \in \{1,2\}$ are computed according to (5.10).

Let us first focus on scalar Shapley effect estimation and the associated confidence intervals, for example scalar Shapley effects for the first component of the output, Y_1. For Y_1, the most important input is X_2, with a Shapley effect around 0.66. In Figure 5.1, we analyze the evolution of the estimation accuracy through the mean absolute error (MAE) and the POC as a function of n and N_{tot}. The value of N_{tot} varies from 100 to $n(2^d - 2) = 2n$. For each combination of n and N_{tot}, $N = 300$ independent random samples are used for estimating the MAE and the POC. The bootstrap sample size B is fixed to 500. The evolution of the 5% and 95% quantiles of the MAE is also drawn. As expected, the estimation accuracy increases with n and N_{tot}. Also, for fixed n, it increases with N_{tot}. As N_{tot} reaches n, which means $N_{\mathfrak{u}} = n/2$ for all $\mathfrak{u}$ as $d = 2$, the decrease of the MAE slows down from exponential to linear. This behavior seems coherent with the result stated in [Broto et al., 2020, Corollary 6.8] which asserts that there exits $0 < C < +\infty$ such that, for $N_{\mathfrak{u}} \geq Cn^{1/(d-|\mathfrak{u}|)}$ and $\delta > 0$:

$$|\widehat{c}_1(\mathfrak{u}) - c_1(\mathfrak{u})| = o_p\left(\frac{1}{n^{1/(2(d-|\mathfrak{u}|)-\delta}}\right). \tag{5.13}$$

The bound in (5.13) does not depend on N_{tot} anymore. Finally, the behavior of the POC is as expected, around 0.9 whatever n and N_{tot}.

The evolution of the MAE and POC for the estimation of aggregated Shapley effects as function of n and N_{tot} is displayed in Figure 5.2. Similarly as for the estimation of scalar effects, the accuracy increases with n and N_{tot}. However for fixed n, the POC deteriorates for large N_{tot} (POC is around 0.8 for $N_{\mathfrak{u}} = n$). The reason for this deterioration is not well understood but is probably due to the use of bootstrap with nearest neighbors. In the following, we keep all the $N_{\mathfrak{u}}$ strictly below n to guarantee the POC.

We estimate Shapley effects if input correlation is higher ($\rho = 0.9$). POC and bias results are also satisfactory (not shown). In fact, POC values vary also around 0.9 and bias decreases and goes to 0 when n and N_{tot} increases. For this simple test case, we have shown that confidence intervals using Algorithm 1 reach accurate coverage probability and that bias reduces when n and N_{tot} increase. Nevertheless in this test case, estimation is effortless because $d = 2$.

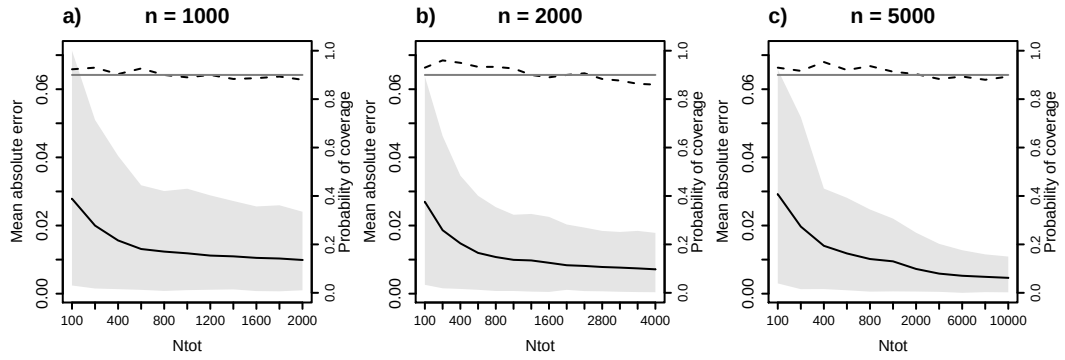

Figure 5.1: Linear Gaussian model in dimension $d = 2$: MAE for the estimation of scalar Shapley effects of the output Y_1 estimated from N=300 i.i.d. samples as a function of N_{tot} using different sample sizes a) $n = 1000$, b) $n = 2000$ and c) $n = 5000$. The 0.05 and 0.95 pointwise quantiles of the absolute error are drawn with gray polygons. The empirical coverage probability of the 90% bootstrap simultaneous confidence intervals is displayed with dotted lines. The theoretical coverage probability 0.9 is also shown with a plain gray line. The bootstrap sample size is fixed to $B = 500$.

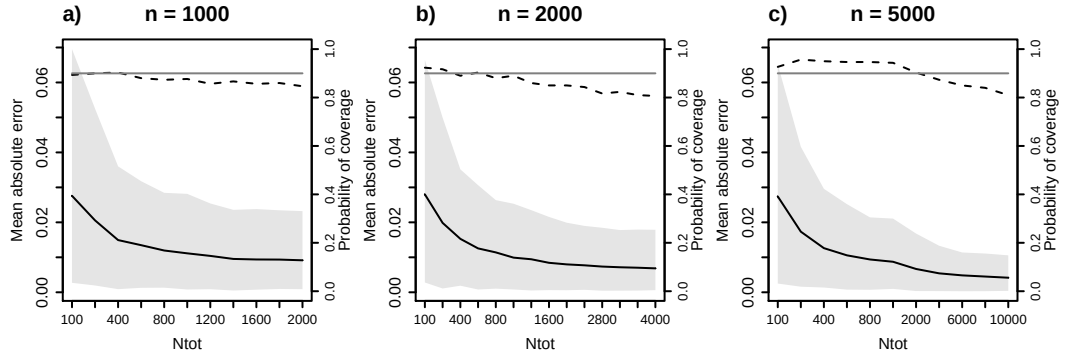

Figure 5.2: Linear Gaussian model in dimension $d = 2$: MAE for the estimation of aggregated Shapley effects estimated from N=300 i.i.d. samples as a function of N_{tot} using different sample sizes a) $n = 1000$, b) $n = 2000$ and c) $n = 5000$. The 0.05 and 0.95 pointwise quantiles of the absolute error are drawn with gray polygons. The empirical coverage probability of the 90% bootstrap simultaneous confidence intervals is displayed with dotted lines. The theoretical coverage probability 0.9 is also shown with a gray plain line. The bootstrap sample size is fixed to $B = 500$.

Input	Description	Distribution
m	mass (kg)	$\mathcal{U}[10, 12]$
c	damping constant ($\mathrm{Nm}^{-1}\mathrm{s}$)	$\mathcal{U}[0.4, 0.8]$
k	spring constant (Nm^{-1})	$\mathcal{U}[70, 90]$
l	initial elongation (m)	$\mathcal{U}[-1, -0.25]$

Table 5.1: Mass-spring model: Input description and uncertainty intervals. $\mathcal{U}$ denotes the uniform distribution. The inputs are independent from each other.

5.5.2 Mass-spring model

The method is illustrated on a test case with discretized functional output: the functional mass-spring model proposed by [Gamboa et al., 2013], where the displacement of a mass connected to a spring is considered:

$$m\ell''(t) + c\ell'(t) + k\ell(t) = 0, \tag{5.14}$$

with initial conditions $\ell(0) = l$, $\ell'(0) = 0$, and $t \in [0, 40]$. There exists an analytical solution to Equation (5.14). This model has four inputs (see more details in Table 5.1). The model output is the vector $\mathbf{Y} = f(\mathbf{X}) = (\ell(t_1), \dots, \ell(t_{800}))$, $t_i = 0.05i$ with $i \in \{1, \dots, 800\}$.

Inputs are considered independent. The true aggregated Shapley effects are unknown but they are approximated using a high sample size $n = 25\,000$ and $N_{tot} = 10\,000$. Then, the Shapley effects estimated are $\widehat{GS}_m = 0.38$, $\widehat{GS}_c = 0.01$, $\widehat{GS}_k = 0.51$ and, $\widehat{GS}_l = 0.09$. Given these results, inputs ranking is: k, m, l and c which corresponds to the same ranking obtained using Sobol' indices (see Table 3 of [Gamboa et al., 2013]).

The discretized output is high-dimensional ($p = 800$). We perform fPCA (see Subsection 5.3.3) to estimate the effects using the first $q \ll p$ fPCs. Figure 5.3 shows the POC and bias evolution if different values for n and N_{tot} are used for the aggregated effect estimation. We use the first 6 fPCs which explain 95% of the output variance (see Figure 5.3 a). For each n and N_{tot} combination, the aggregated Shapley effects are estimated for $N = 100$ independent samples and confidence intervals are estimated with $B = 500$ bootstrap samples. The accuracy is not very good if sample size is small $n = 1000$ (see Figure 5.3 b). However, it reduces drastically when the sample size n increases as expected. On our experiments, the smallest MAE is achieved for $n = 5000$ and $N_{tot} = 2002$ (see Figure 5.3 d). The confidence interval reaches the expected POC 0.9 as soon as n and N_{tot} are large enough (see Figure 5.3 b).

5.6 Snow avalanche modeling

5.6.1 Model

Our avalanche model is based on depth-averaged Saint-Venant equations and considers the avalanche as a fluid in motion. In more detail, it considers only the dense layer of the avalanche. The flow depth is then small compared to its length. The model assumes the avalanche is flowing on a curvilinear profile $z = l(x)$, where z is the elevation and x is the projected distance measured from the top of the avalanche path. Under these

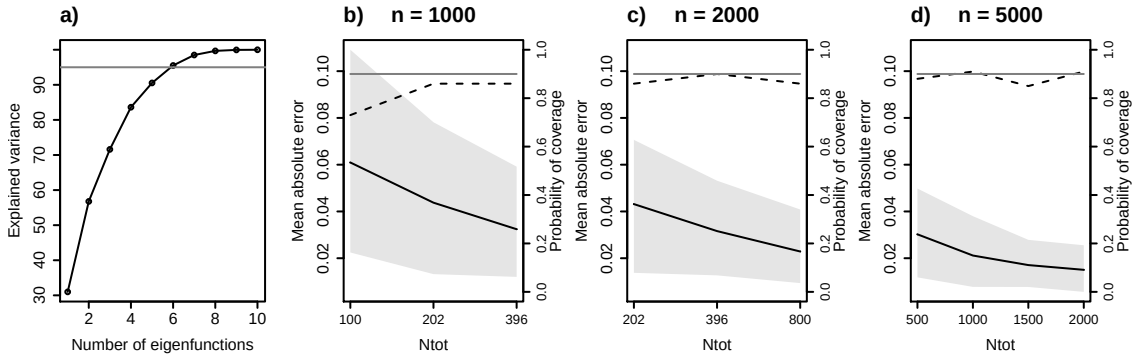

Figure 5.3: Mass-spring model: a) Explained variance as a function of the number of principal components. The gray line is displayed at 95% of the variance explained which corresponds to 6 components. The MAE computed from $N = 100$ independent realizations of the aggregated Shapley effect estimates, using the first 6 principal components, is drawn as a function of N_{tot} with b) $n = 1000$, c) $n = 2000$ and d) $n = 5000$. The 0.05 and 0.95 pointwise quantiles of the MAE are drawn with gray polygons. The probability of coverage of the 90% bootstrap simultaneous intervals is displayed with a dotted line. The theoretical POC 0.9 is highlighted with a plain gray line. The bootstrap sample size is fixed to $B = 500$.

assumptions, shallow-water approximations of the mass and momentum equations can be used:

$$\begin{aligned} \frac{\partial h}{\partial t} + \frac{\partial hv}{\partial x} &= 0 \\ \frac{\partial hv}{\partial t} + \frac{\partial}{\partial x}\left(hv^2 + \frac{h^2}{2}\right) &= h(g\sin\theta - \mathrm{F}) \end{aligned}$$

where $v = \|\vec{\mathbf{v}}\|$ is the flow velocity, h is the flow depth, θ is the local angle, t is the time, g is the gravity constant and $\mathrm{F} = \|\vec{\mathbf{F}}\|$ is a frictional force. The model uses the Voellmy frictional force $\mathrm{F} = \mu g cos\theta + g/(\xi h)v^2$, where μ and ξ are friction parameters. The partial differential equation (PDE) system is solved numerically using a finite volume scheme on the path curvilinear profile with a 5m resolution (see full details about the PDE scheme in [Naaim et al., 2004]). Hence, even if the PDEs are expressed in a two-dimensional frame, they in fact represent a one-dimensional flow on the curvilinear profile ([Greve et al., 1994]).

The numerical model depends on six inputs: the friction parameters μ and ξ, the length $\mathrm{l_{start}}$ of the avalanche release zone, the snow depth $\mathrm{h_{start}}$ within the release zone, the abscissa corresponding to the beginning of the release zone denoted by $\mathrm{x_{start}}$ and the discretized topography of the flow path, denoted by $D = (\mathbf{x}, \mathbf{z}) \in \mathbb{R}^{N_s \times 2}$ where $\mathbf{x} \in \mathbb{R}^{N_S}$ is the vector of projected abscissa positions and $\mathbf{z} = l(\mathbf{x}) \in \mathbb{R}^{N_S}$ is the elevation vector. N_s is the number of points of the discretized path. Subsection 5.6.1 shows an illustration of the avalanche model. We use for D the topography of a path located in Bessans, France. This particular path already considered in other works [Eckert et al., 2008a, 2010, Favier et al., 2014b, Eckert et al., 2018b] in well documented in the French avalanche database Bourova et al. [2016]. The model outputs are the flow velocity, flow depth trajectories in the path D and runout distance of an avalanche. Note that the model has two functional

Input	Description	Distribution
μ	Static friction coefficient	$\mathcal{U}[0.05, 0.65]$
ξ	Turbulent friction [m/s^2]	$\mathcal{U}[400, 10000]$
$\mathrm{l_{start}}$	Length of the release zone [m]	$\mathcal{U}[5, 300]$
$\mathrm{h_{start}}$	Mean snow depth in the release zone [m]	$\mathcal{U}[0.05, 3]$
$\mathrm{x_{start}}$	Release abscissa [m]	$\mathcal{U}[0, 1600]$

Table 5.2: Avalanche model, scenario 1: Input description and uncertainty intervals. In the the GSA, we consider $\mathrm{vol_{start}} = l_{start} \times h_{start} \times 72.3/\cos(35°)$ instead of $\mathrm{h_{start}}$ and $\mathrm{l_{start}}$.

and one scalar outputs and these three outputs are the objects of the GSA study. We develop our GSA in two contexts or scenarios: (i) little knowledge on the input parameter probability distribution, and (ii) well-calibrated input parameter distributions.

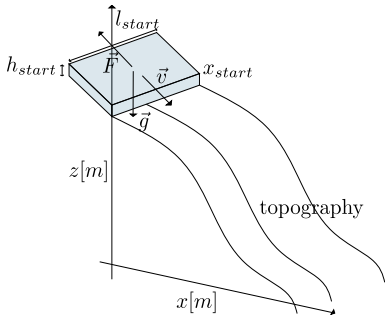

Figure 5.4: The avalanche model.

5.6.2 Scenario 1

Principle

We first determine the most influential input parameters sampled from uniform distributions. We thus expect from the GSA a better understanding of the numerical model. Inputs μ and ξ vary in their physical value ranges. Inputs $\mathrm{l_{start}}$ and $\mathrm{h_{start}}$ vary in their spectrum of reasonable values given the characteristics of the avalanche path. The $\mathrm{x_{start}}$ input distribution is determined by calculating the abscissa interval where the release zone average slope is superior to 30°. Indeed, the slope remains above 30° during the first 1600m of the path. A good approximation of avalanche release zones is commonly obtained this way. Since different studies [Bartelt et al., 2012, Brian Dade and Huppert, 1998] suggest that the volume of snow is a critical quantity that controls flow dynamics, we consider $\mathrm{vol_{start}}$ as input of the GSA instead of $\mathrm{h_{start}}$ and $\mathrm{l_{start}}$. The latter is evaluated as $\mathrm{vol_{start}} = l_{start} \times h_{start} \times 72.3/\cos(35°)$. The mean width and slope of the release zone equal to 72.3m and 35°, respectively. All uncertainty intervals are summed-up in Table 5.2. The input correlations are close to 0 since we assume they are a priori independent.

For a given avalanche simulation, its functional velocity and flow depth outputs have a high number of zeros because they are null above the beginning of the release zone and after the runout position. Also, there might be some avalanche simulations that are meaningless in physical terms and/or not useful to assess the related risk. Therefore to perform GSA, we select simulations that accomplish the following rules:

(i) avalanche simulation is flowing in the interval $[1600m, 2412m]$,

(ii) its volume is superior to 7000 m^3 and,

(iii) avalanche runout distance is inferior to 2500m which corresponds to the end of the path.

In the following, we use the terminology *constrained sample* to deal with the sample obtained after applying rules (i) to (iii).

The return period of avalanches in the interval $[1600m, 2412m]$ varies from 1 to 10 000 years according to the work of Eckert et al. [2010]. Roughly speaking, a return period is the mean time in which a given runout distance is reached or exceeded at a given path's position. Also, we focus on medium, large and very large avalanches which have a high potential damage.

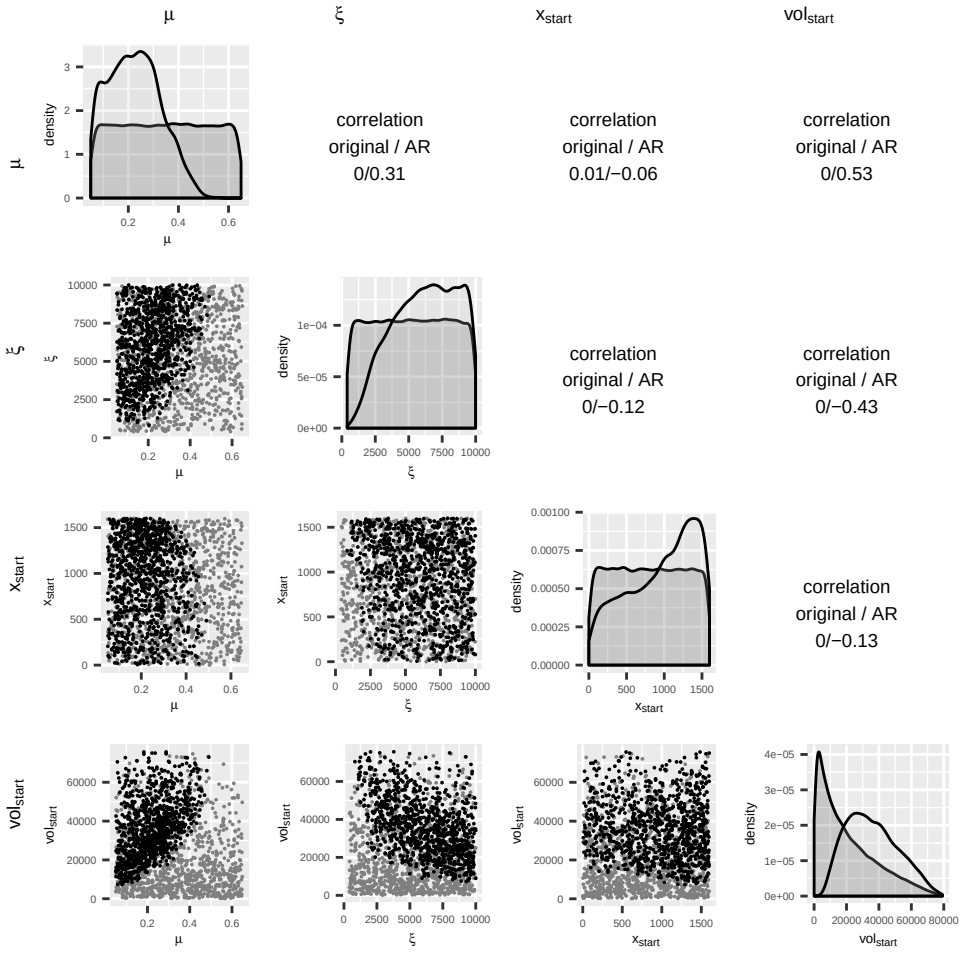

Figure 5.5: Avalanche model, scenario 1: scatter-plots of initial (black points) and constrained (gray points) samples. In the figure's diagonal, the density function of the initial (gray color) and constrained (transparent) samples are displayed. Input correlations of the original and constrained samples are shown. 1000 subsamples of original and constrained samples are used for illustration purpose.

Global sensitivity analysis results

We first ran $n_0 = 100\,000$ avalanche simulations from an i.i.d. sample of input distributions described in Table 5.2. Then, by applying (i) to (iii) the size of the *constrained sample* was reduced to $n_1 = 6152$ (Subsection 5.6.2). Even if the initial sample size is high ($n_0 = 100\,000$) and if the corresponding input parameter sample does not present any significant correlation structure, the size of *the constrained sample* is low and we can observe a correlation structure. For example, inputs μ and ξ were independent for the initial sample but the correlation computed after applying the constraints is 0.31. Note that the input parameter correlations induced by the constraints were the main motivation to compute Shapley effects and not Sobol' indices in this first scenario.

On Subsection 5.6.2 are plotted highest density region (HDR) boxplots for the velocity and flow depth, obtained by using the R package `rainbow` developed by Hyndman and Shang [2010]. The HDR boxplot is a vizualization tool for functional data based on kernel density estimation of the scores associated to the two first principal components of the functional data (see Hyndman [1996] for further details). In the data we consider in our study, the avalanche velocity ranges from 0.1ms^{-1} to 71.56ms^{-1} and avalanches are decelerating (see Subsection 5.6.2 a). Flow depths vary from 0.03m to 7.52m. The flow depth curves exhibit high fluctuations in the interval $[2100m, 2300m]$ (see Subsection 5.6.2 b) which corresponds to a region where path's topography is mostly convex. Runout distances vary from 2409m to 2484m (see Subsection 5.6.2 c).

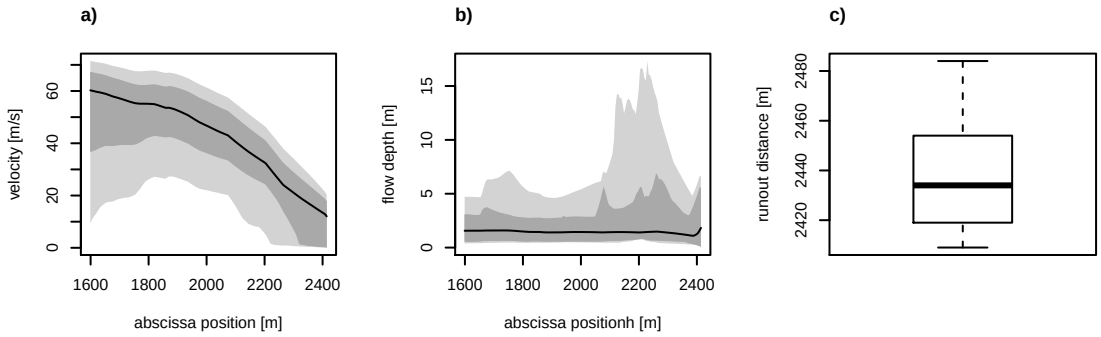

Figure 5.6: Avalanche model, scenario 1: a) and b) functional HDR boxplots of velocity and flow depth curves, resp. It is shown 50% HDR (dark gray), 100% HDR (light gray) and modal curve, the curve in the sample with the highest density (black line). c) runout distance boxplot. The constrained sample size is $n_1 = 6152$.

On Subsection 5.6.2 panels a and b, ubiquitous (pointwise) Shapley effects of velocity and flow depth curves are shown, respectively. Depending on the output, results are quite different. For velocity, x_{start} is the most relevant during a large part of the path but its importance decreases along the path and, conversely, the importance of the other inputs increases. For the flow depth output, the most important input is $\text{vol}_{\text{start}}$ since the corresponding Shapley effects vary from 0.4 to 0.2 along the path. Nevertheless, other inputs are not completely negligible. Input importance also varies according to the topography. In fact, the ubiquitous effect variation corresponds to local slope changes (see Subsection 5.6.2 a and b). Correlations between ubiquitous effects and local slope have been computed and are rather high. For example, for the velocity, the absolute value of the correlation is higher than 0.51 for all input parameters. This implies that local slope

changes play an important role on the input contribution to output variations, a nice results showing the relevance of the GSA analysis to understand the dynamical properties of the flow. Eventually, for runout distance, the four inputs are relevant. Our results are in accordance with previous works documenting the sensitivity of snow avalanche models to their input conditions using less formalised methods [Borstad and McClung, 2009]. They also relate to more general results regarding snow avalanche dynamics. Especially, it is well known that in the transit area of the path, an avalanche accelerates up to a maximal velocity and this limit velocity can be explicitly computed on a constant slope under the assumption of a Voellmy friction law [Salm et al., 1990, Gauer, 2014, McClung and Gauer, 2018]. It is therefore all the more logical that the influence of x_{start} on velocity decreases as the avalanche flows downslope. More generally, with depth-averaged flow equations, it is known that initial conditions corresponding to the avalanche release are important only in the early phases of the flow and are forgotten later, especially in the runout area where velocity and flow depths become primarily constrained by local slope.

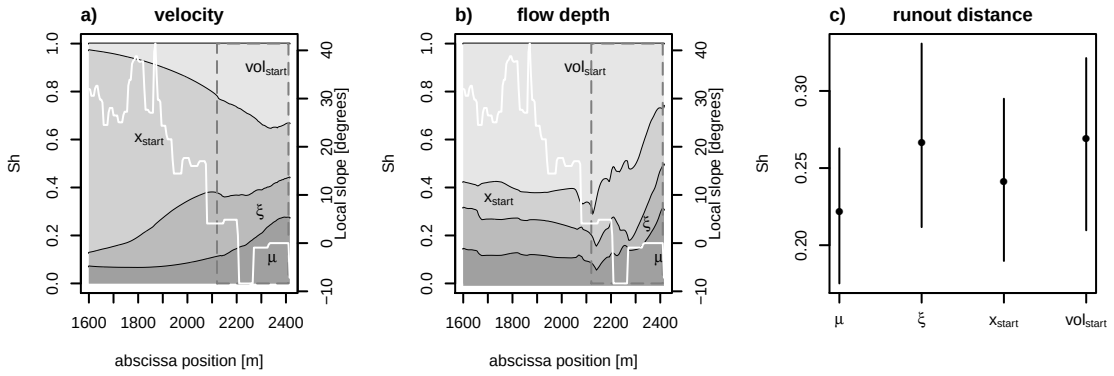

Figure 5.7: Avalanche model, scenario 1: a) and b) ubiquitous Shapley effects of velocity and flow depth curves, resp. and, c) runout distance Shapley effects. Shapley effects are estimated with a sample of size 6152 and Ntot=2002. The local slope is displayed with a white line. A gray dotted rectangle box is displayed at interval [2017, 2412] where snow avalanche return periods vary from 10 to 10 000 years. The bootstrap sample size is fixed to $B = 500$.

The fPCs convey useful information about the avalanche model and also, its sensitivity. Subsection 5.6.2 (resp., Subsection 5.6.2) shows the correlations between the fPCs and the velocity (resp., flow depth) output for scenario 1. The first fPC has a positive correlation with the velocity, always higher than 0.57, however the correlation decreases drastically from the top to the bottom of the path. By contrast, the second fPC is less correlated to the velocity, the correlation is positive at the top of the path, then the correlation decays and even becomes highly negative for $x \geq 2300$ m. Hence, these two components identify well two regions of the path where velocity behaves differently: the first fPC corresponds to the release and transit area of the path, where avalanches accelerate and maintain their velocity while the second fPC identifies the runout area where avalanches are decelerating and eventually stop. The behavior is similar for the flow depth, with a sharp peak in the second fPC corresponding to the location in the runout area where slope becomes null and then negative. To strengthen our physical interpretation, we estimated the scalar Shapley effects for each fPC with $n = 6152$ and $N_{tot} = 2002$. As far as velocity is concerned, for the first fPC, x_{start} is the most important input and for the second fPC, vol_{start}, x_{start} and μ are the most relevant. For the flow depth, for the first fPC, vol_{start} is the most

relevant. For the second fPC, the inputs x_{start}, vol_{start} and μ are the most important inputs. These results are in accordance with previous works documenting the sensitivity of snow avalanche models to their input conditions, and notably the high importance of the Coulomb friction coefficient μ and snow volume in the runout area [Dent and Lang, 1980, Borstad and McClung, 2009, Bartelt et al., 2012]. Similar results are obtained for scenario 2 (see Figures 5.16 and 5.17 in Section 5.11).

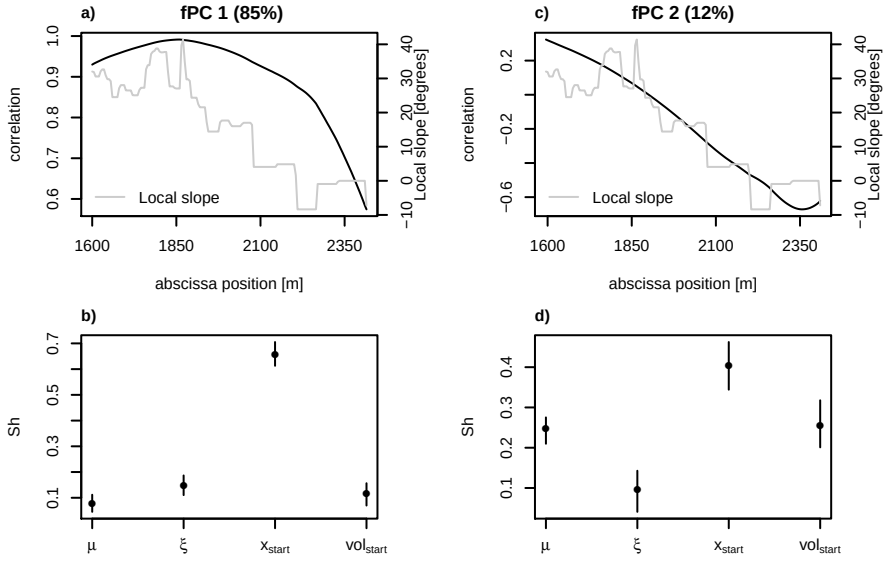

Figure 5.8: Avalanche model, scenario 1: a) and c) correlations between the first and second fPCs and the velocity. Scalar Shapley effects for b) fPC 1 and d) fPC 2 are estimated with $n = 6152$ and $N_{tot} = 2002$. For the confidence intervals, $B = 500$ is used. The local slope is displayed with a gray line.

Subsection 5.6.2 shows aggregated Shapley effects and 90% confidence intervals computed over space intervals $[x, 2412]$ where $x \in \{1600, 1700, \ldots, 2412\}$. The aggregated effects are computed in the first fPCs explaining more than 95% of the output variance. Aggregated effects seem more robust than ubiquitous effects, specially in zones where local slope shows high variations (see Subsection 5.6.2 compared to Subsection 5.6.2). For explaining more than 95% of the velocity output variance, 2 fPCs are required, while, for explaining more than 95% of the flow depth output variance, at most 4 fPCs are required, depending on x. For the velocity output, the most important input is x_{start} in the interval $[1600m, 2100m]$ but its importance decreases along the path. In the interval $[2017m, 2412m]$ where return periods are non trivial, x_{start} and vol_{start} are the most important followed by μ and ξ. For the flow depth output, vol_{start} is the most relevant but its importance decreases along the path. At the end of the path from 2300m to 2412m where return periods are high (between 100 to 10 000 years), confidence intervals intersect. It seems thus difficult to deduce a clear ranking of the inputs for these last portions of the path. Nevertheless, it seems that none of the inputs is negligible, even at the very end of the path. In summary, to estimate velocities with accuracy, the release zone and volume are the most important parameters and, for the flow depth, a good approximation of the volume released is essential.

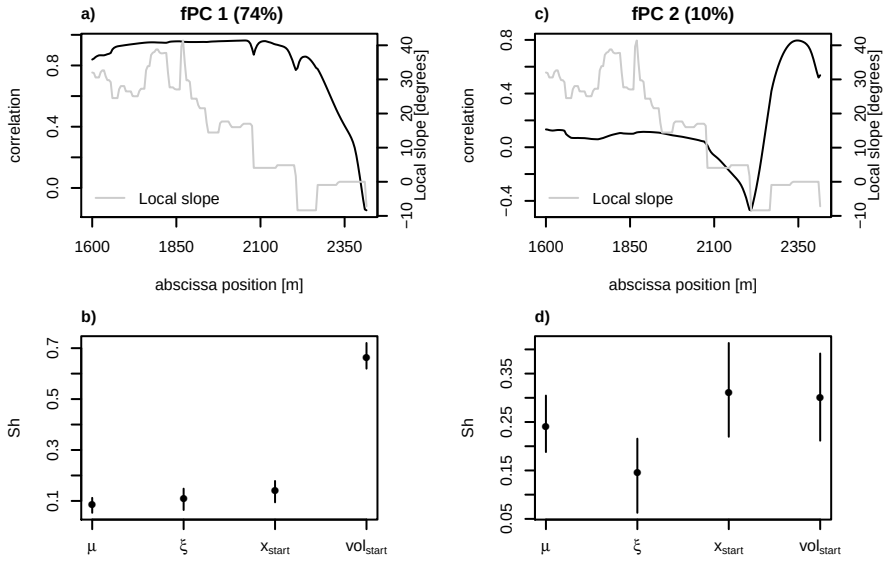

Figure 5.9: Avalanche model, scenario 1: a) and c) correlations between the first and second fPCs and the flow depth output. Scalar Shapley effects for b) fPC 1 and d) fPC 2 are estimated with $n = 6152$ and $N_{tot} = 2002$. For the confidence intervals, $B = 500$ is used. The local slope is displayed with a gray line.

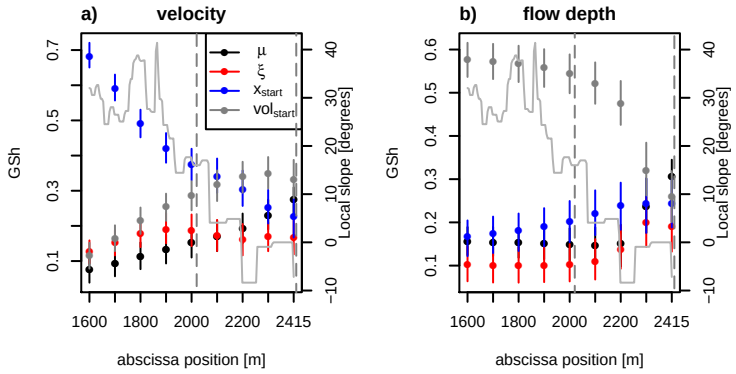

Figure 5.10: Avalanche model, scenario 1: a) and b) aggregated Shapley effects of velocity and flow depth curves calculated over space intervals $[x, 2412m]$ where $x \in \{1600m, 1700m, \ldots, 2412m\}$. Shapley effects are estimated with samples of size 6152 and Ntot=2002. Effects are estimated using the first fPCs explaining more than 95% of the output variance. The local slope is displayed with a gray line. A gray dotted rectangle is displayed at [2017m, 2412m] where snow avalanche return periods vary from 10 to 10 000 years. The bootstrap sample size is fixed to $B = 500$.

Input	Distribution
$\mathrm{x_{nstart}} = \frac{x_{start}}{1600}$	$\mathrm{Beta}(1.38, 2.49)$
$\mathrm{h_{start}}\|x_{\mathrm{nstart}}$	$\mathrm{Gamma}\left(\frac{1}{0.45^2}(1.52+0.03x_{nstart})^2, \frac{1}{0.45^2}(1.52+0.03x_{nstart})\right)$
$\mathrm{l_{start}}$	$31.25+87.5\mathrm{h_{start}}$
$\mu\|h_{\mathrm{start}}, x_{\mathrm{nstart}}$	$\mathcal{N}(0.449 - 0.013x_{nstart} + 0.025h_{start}, 0.11^2)$

Table 5.3: Avalanche model: Scenario 2. Input description and uncertainty intervals. $\mathrm{x_{nstart}}$ is a normalization of $\mathrm{x_{start}}$. A linear relationship between $\mathrm{h_{start}}$ and $\mathrm{l_{start}}$ inferred from the local data is used [Eckert et al., 2010]. Gamma distribution parameters are its shape and rate. In the the GSA, we consider $\mathrm{vol_{start}} = l_{start} \times h_{start} \times 72.3/\cos(35°)$ instead of $\mathrm{h_{start}}$ and $\mathrm{l_{start}}$.

5.6.3 Scenario 2

Principle

The aim is now to determine the most influential inputs in the context of strong knowledge regarding input distributions. In [Eckert et al., 2010], the authors developed a Bayesian framework to estimate input distributions from available avalanche observations. The objective is long-term avalanche hazard assessment in order to assess the related risk for buildings and people inside. In the avalanche literature, it is assumed that ξ depends on the path topography, so it is a parameter and not a variable varying from one avalanche to another in [Eckert et al., 2010]'s model. The input ξ is therefore fixed to its posterior estimate, 1300. Other input variables in this scenario are dependent. The dependence between $\mathrm{h_{start}}$ and $\mathrm{l_{start}}$ is modeled with a linear function $\mathrm{l_{start}} = 31.25 + 87.5\mathrm{h_{start}}$, and similarly as in scenario 1, we consider $\mathrm{vol_{start}}$ as input of the GSA instead of $\mathrm{h_{start}}$ and $\mathrm{l_{start}}$. The complete input distribution resulting from the Bayesian inference on the studied path is described in Table 5.3. Input correlations have been computed. As an example, the correlation between μ and $\mathrm{vol_{start}}$ is 0.8. As a preliminary step to GSA for this scenario, we apply the following constraints:

(i) avalanche is flowing in the interval [1600m, 2204m] where snow avalanche return periods vary from 10 to 300 years,

(ii) avalanche volume is superior to 7 000 m^3 and,

(iii) μ coefficient is inferior to 0.39 as we focus on dry snow avalanches.

Under these conditions, we sample the full set of dry snow avalanches that could cause strong material or human damages on the studied site.

Global sensitivity analysis results

We first ran $n_0 = 100\,000$ avalanches from an i.i.d. sample of input distribution following Table 5.3. After applying the constraints, the sample size was reduced to $n_2 = 1284$ and the input distribution was modified. For example, μ and $\mathrm{vol_{start}}$ correlation changes from 0.8 to 0.2 which is still non negligible. Ubiquitous Shapley effects are displayed on Subsection 5.6.3 panels a and b. For the velocity, the three inputs have a similar importance till 1900m, then $\mathrm{vol_{start}}$ importance decreases and μ and $\mathrm{x_{start}}$ importance increases (see Subsection 5.6.3 a). Similarly, as in scenario 1, the effects show fluctuations

which correspond to changes in local slope. In particular, for the flow depth, input effects suffer radical changes when the local slope decreases from 20° to 10° (see Subsection 5.6.3 b). For the runout distance, all inputs are relevant (see Subsection 5.6.3 c).

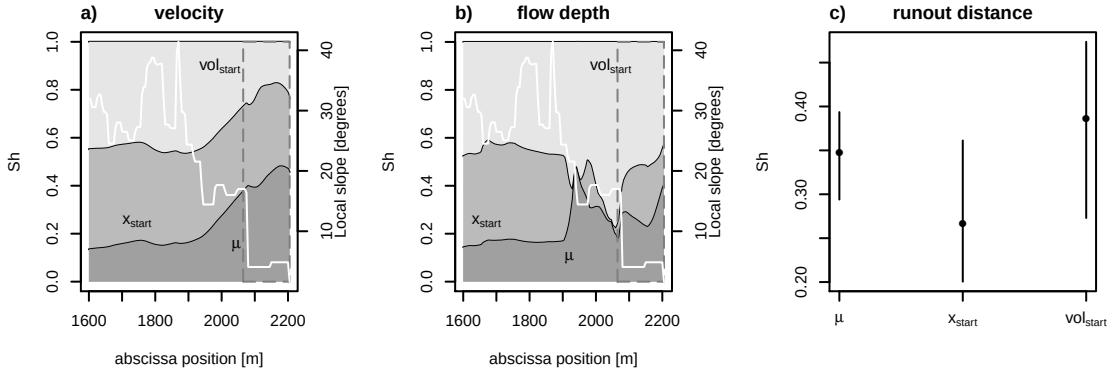

Figure 5.11: Avalanche model, scenario 2: a) and b) ubiquitous Shapley effects of velocity and flow depth curves, c) runout distance Shapley effects. Shapley effects are estimated with samples of size 1284 and Ntot=800. The local slope is displayed with a white line. A gray dotted rectangle shows the interval [2064, 2204] where return periods vary from 10 to 300 years. The bootstrap sample size is fixed to $B = 500$.

Aggregated effects (see Subsection 5.6.3) present less fluctuations and are easier to interpret (see Subsection 5.6.3). In summary, under this second scenario, it is fundamental to have a good approximation of the released volume and abscissa for velocity forecasting, while for flow depth forecasting, a good approximation of released volume is desirable. Nevertheless, none of the other inputs are negligible. Note that the uncertainty associated to the estimation of Shapley effects at 2204m is high (see the width of the corresponding confidence intervals on Subsection 5.6.3). To outperform the estimation accuracy at the end of the path generating a larger initial sample of avalanches is possible, but the computational burden is prohibitive.

5.7 Conclusions and perspectives

In this work, we extended Shapley effects to models with multivariate or functional outputs. We proved that aggregated Shapley effects accomplish the natural requirements for a GSA measure. For the estimation, we proposed to extend the subset aggregation procedure with double Monte Carlo given data estimator of Broto et al. [2020]. We proposed an analysis of the impact of the total budget N_{tot}, the number of neighbors N_I and the optimal allocation N_u^* on the accuracy of Shapley effect estimation. Also, we proposed an algorithm to construct bootstrap confidence intervals for scalar and aggregated Shapley effects based on the ideas of [Benoumechiara and Elie-Dit-Cosaque, 2019]. In test cases, the convergence of our estimator was empirically studied. Also, we empirically demonstrated that the bootstrap confidence intervals we proposed have accurate coverage probability. Estimation and bootstrap confidence interval algorithms well behave. Nevertheless, high sample sizes ($n = 5000$ and $N_{tot} = 2002$) are required to guarantee accurate results. Remark that it is well known that Shapley effects estimation is costly. It would be interesting to study theoretically the asymptotic properties of our estimator, but this

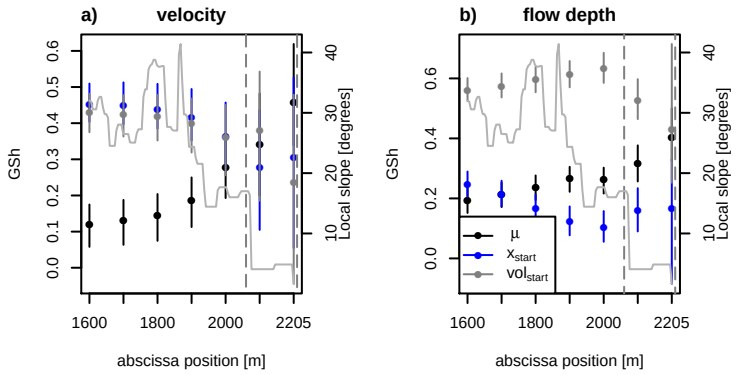

Figure 5.12: Avalanche model, scenario 2: a) and b) aggregated Shapley effects of velocity and flow depth curves calculated over space intervals $[x, 2204]$ where $x \in \{1600, 1700, \ldots, 2204\}$ and using the first fPCs which have 95% of output variance. Shapley effects are estimated with samples of size 1284 and Ntot=800. The local slope is displayed with a gray line. A gray dotted rectangle is displayed at [2017m, 2204m] where return periods vary from 10 to 300 years. The bootstrap sample size is fixed to $B = 500$.

study is out of the scope of this paper. Recently, in the R package `sensitivity` the function `sobolshap_knn` to estimate Shapley effects with n and N_{tot} from a given data sample has been implemented. This function uses a tree based technique to approximate nearest-neighbor search which reduces drastically computation times. The function is particularly attractive if n and N_{tot} are high, we could even use $N_{tot} = (2^d - 2) \times n$. However, we did not use this function in this is work because there is no paper that analyzes the convergence of the estimator and, this subject was out of the scope of this work. We rather used the `shapleySubsetMc` function which corresponds to the estimator introduced in [Broto et al., 2020] on which our estimator for aggregated Shapley effects is based.

We applied our GSA methods to an avalanche propagation model under two different settings. Due to physical modeling constraints input parameters were not confined in a rectangular region. For these reasons, it was not possible to consider independence of input parameters. Results showed probative linkages between local slope and sensitivity indexes. Notably, aggregated Shapley effects were more stable and easier to interpret than ubiquitous effects, as already observed by Alexanderian et al. [2020] in the case of aggregated Sobol' indices. This demonstrates the usefulness of our approach for many practical problems. Especially, it could be applied to other avalanche paths to generalize the results obtained in terms of respective weights of the inputs and interpret the sensitivity indices more deeply in terms of physical properties of avalanche flows. Eventually, application was challenging because constrained samples were of moderate size, for example, from the 100 000 initial sample, the physical constraints produced a 6000 to 1200 sample, depending on the setting. In a future work, it would be useful to construct a surrogate of the avalanche model to generate larger constrained samples, improve the accuracy of aggregated Shapley effect estimation and thus reduce confidence intervals width.

$\lvert\mathfrak{u}\rvert$	1	2	3	4	5	6	7	8	9
$N^*_{\mathfrak{u}}$	600	133	50	29	24	29	50	133	600
$N_{\mathfrak{u}} = \lfloor N_{tot}/(2^d-2) \rfloor$	53	53	53	53	53	53	53	53	53

Table 5.4: Input dimension $d = 10$: $N_{\mathfrak{u}}$ values for both strategies (i) and (ii) for $n = 10\,000$ (number of simulations) and $Ntot = 54\,000$ (total cost).

5.8 Comparison between the uniform allocation $N_{\mathfrak{u}} = \lfloor N_{tot}/(2^d-2) \rfloor$, and the one introduced in [Broto et al., 2020] $N^*_{\mathfrak{u}} = \left\lfloor N_{tot} \binom{d}{\lvert\mathfrak{u}\rvert}^{-1} (d-1)^{-1} \right\rfloor$, $\emptyset \subseteq \mathfrak{u} \subsetneq \{1,\dots,d\}$

The total cost N_{tot} to estimate all the $c(\mathfrak{u})$, $\emptyset \subsetneq \mathfrak{u} \subsetneq \{1,\dots,d\}$ could be uniformly divided as $N_{\mathfrak{u}} = \lfloor N_{tot}/(2^d-2) \rfloor$ (called $N_{\mathfrak{u}}$ uniform). To test if there is a difference in the estimation by using $N_{\mathfrak{u}}$ uniform or $N^*_{\mathfrak{u}}$ optimal, we consider the Gaussian linear model example from [Owen and Prieur, 2017]:

$$f(\mathbf{X}) = \sum_{i=1}^{d} \beta_i X_i,$$

where X_i are independent centered Gaussian variables with standard deviation $\sigma_i = i^2$ and, $\beta_i = 1$ for all $i \in \{1,\dots,d\}$. The scalar Shapley effects can be computed analytically (see e.g., Owen and Prieur [2017]): $\mathrm{Sh}_i = \beta_i^2\sigma_i^2 / \sum_{i=1}^d \sigma_i^2$. We used $N = 100$ independent samples of size $n = 10\,000$ and $N_{tot} = 54\,000$ to draw the boxplots shown in Figure 5.13. Table 5.4 summarizes the sample size used to estimate each conditional element $\mathbb{E}(\mathrm{Var}(Y|\mathbf{X}_{-\mathfrak{u}}))$ in function of $\lvert\mathfrak{u}\rvert$ in dimension $d = 10$. Figure 5.13 shows that in high dimension, at least for that toy model, the optimal allocation $N^*_{\mathfrak{u}}$ leads to a better accuracy than the uniform allocation. When dimension is low (for example $d = 4$ as in the mass-spring model), taking $N^*_{\mathfrak{u}}$ or $N_{\mathfrak{u}}$ give similar results because the two options are quite similar.

5.9 Comparison between different values of N_I

In Figure 5.14, we analyzed the impact of using different numbers of neighbors N_I in the Shapley effect estimation for the linear Gaussian model in dimension $d = 2$ test case described in Subsection 5.5.1. For the estimation we used $n = 5000$, $N_{tot} = 5000$ and $N^*_{\mathfrak{u}} = \left\lfloor N_{tot} \binom{d}{\lvert\mathfrak{u}\rvert}^{-1} (d-1)^{-1} \right\rfloor$, $\emptyset \subseteq \mathfrak{u} \subsetneq \{1,\dots,d\}$. Parameter N_I is varied from 2 to 300. We see in Figure 5.14 that there is no clear evidence if $N_I \geq 3$ is always the best choice. However in the following we choose $N_I = 3$ following the recommendations in [Song et al., 2016], where the authors prove in the non given data framework that choosing $N_I = 3$ is better than choosing $N_I = 2$ and that the gain from increasing N_I from 3 decreases as N_I increases.

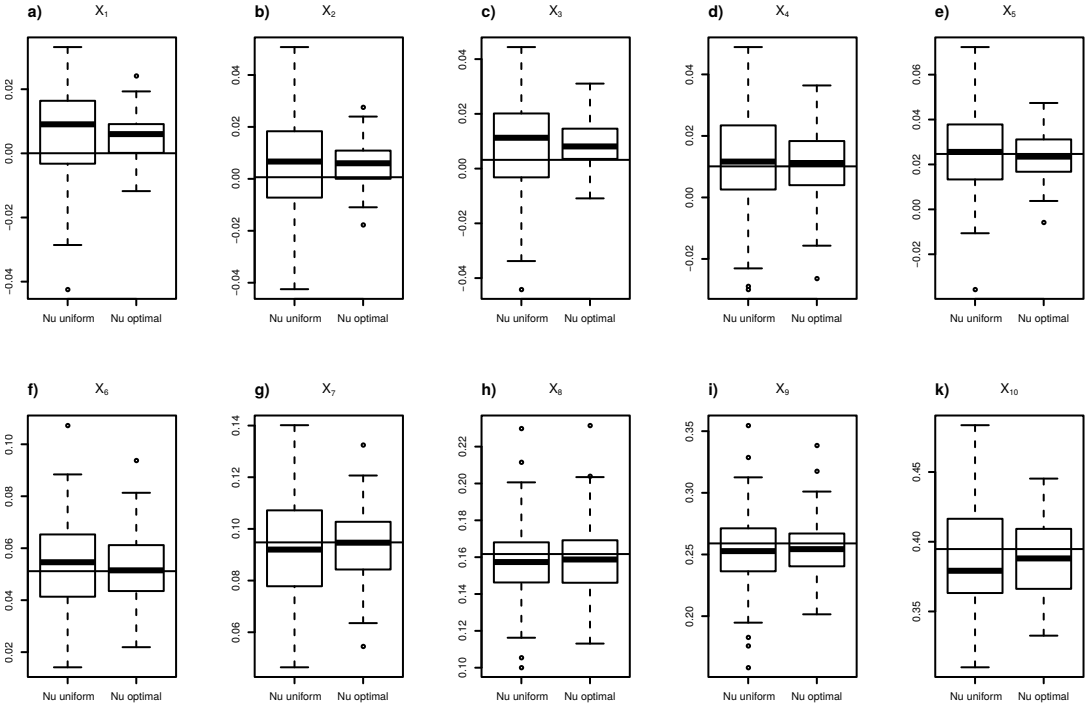

Figure 5.13: Linear Gaussian model with $d = 10$ inputs: comparison between $N_{\mathfrak{u}} = \lfloor N_{tot}/(2^d - 2) \rfloor$ for all $\mathfrak{u}$ (called $N_{\mathfrak{u}}$ uniform) and $N_{\mathfrak{u}} = N_{\mathfrak{u}}^*$ for all $\mathfrak{u}$ (called $N_{\mathfrak{u}}$ optimal) with $n = 10\,000$ and $N_{tot} = 54\,000$. The true scalar Shapley effects are displayed with black lines.

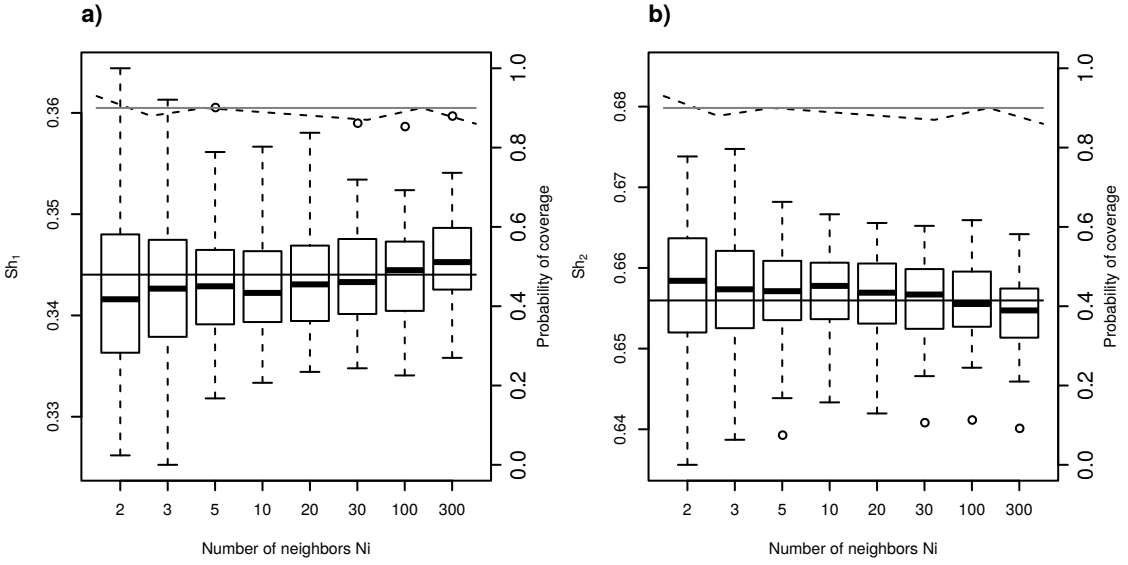

Figure 5.14: Linear Gaussian model in dimension $d = 2$: boxplots of scalar Shapley effects and evolution of the estimation by varying the number of closest neighbors $N_I = \{2, 3, 5, 10, 20, 30, 100, 300\}$. Boxplots are drawn with $N = 100$ independent random samples with $n = 5000$, $N_{tot} = 5000$ and $N_{\mathfrak{u}}^* = \left\lfloor N_{tot} \binom{d}{|\mathfrak{u}|}^{-1} (d-1)^{-1} \right\rfloor$, $\emptyset \subseteq \mathfrak{u} \subsetneq \{1, \ldots, d\}$.

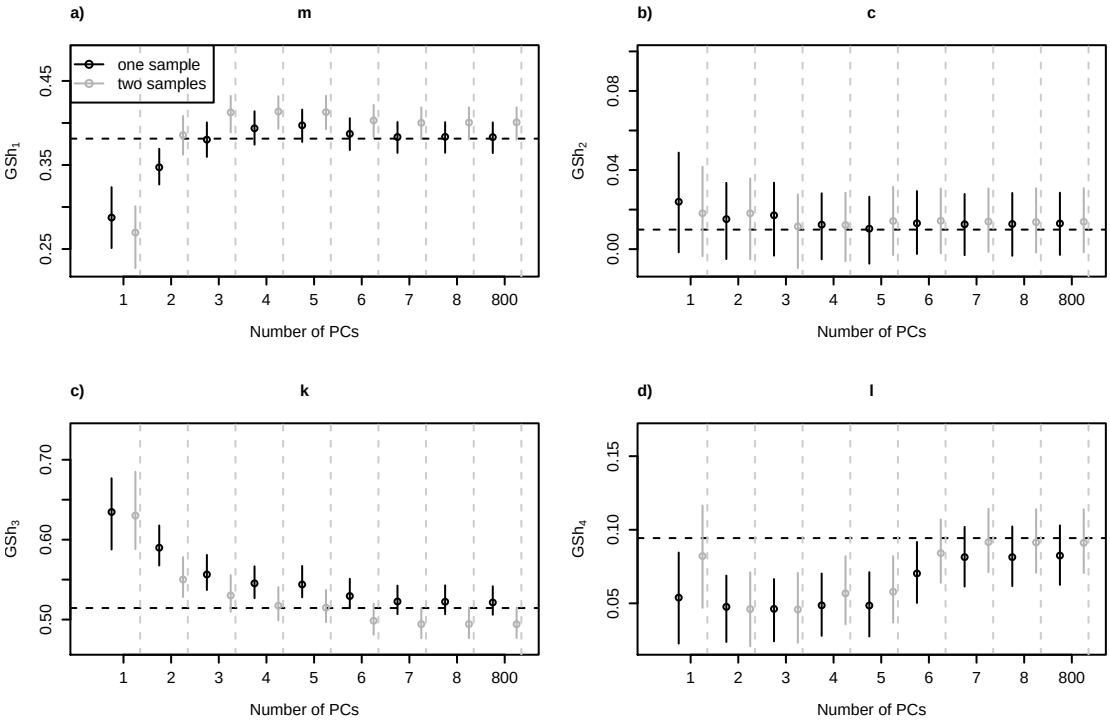

Figure 5.15: Mass-spring: estimation and associated confidence intervals for aggregated Shapley effects using a single sample of size $n = 5000$ to perform the dimension reduction and to estimate the effects or two samples (one to perform the dimension reduction of size $n_1 = 1000$ and another one to estimate the effects $n_2 = 4000$). The estimation as a function of the basis size (number of principal components) is displayed. All the pannels were produced with $N_{tot} = 2000$ and $B = 500$ (bootstrap sample size for confidence intervals).

5.10 Estimation of Shapley effects using one or two samples

We compare the accuracy of Shapley effect estimation for the mass-spring test case when using a single sample of size $n = 5000$ or two independent samples of size $n_1 = 1000$ and $n_2 = 4000$ for estimating the KL representation and the aggregated Shapley effects. In Figure 5.15 we see that the procedure without splitting produces better results for three of the four inputs. For the last input, results with two samples are slightly better. Based on these results, we decided to apply the procedure without splitting on our snow avalanche application.

5.11 Functional principal components and Shapley effects for scenario 2

Figure 5.16 (resp., Figure 5.17) shows the correlations between the fPCs and the velocity (resp., flow depth) output for scenario 2.

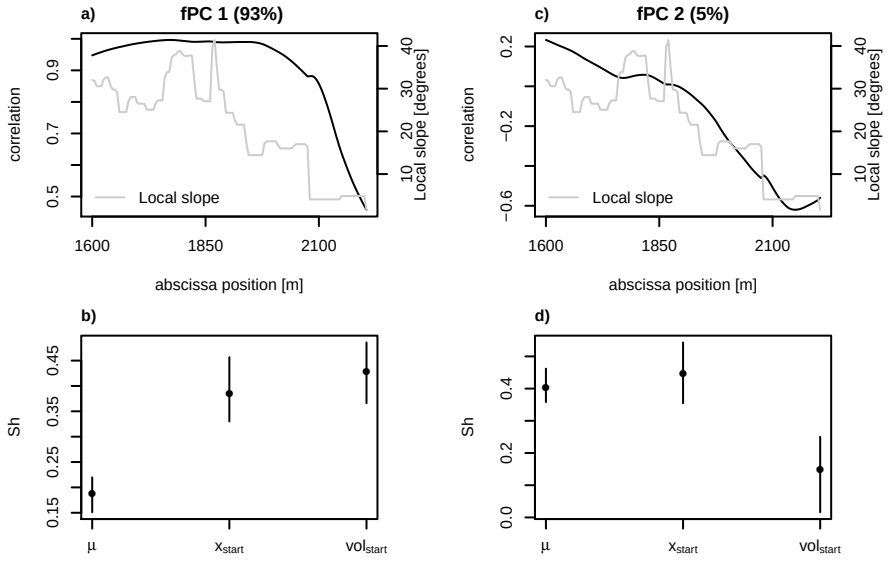

Figure 5.16: Avalanche model, scenario 2: a) and c) correlations between the first and second fPCs and the velocity output. Scalar Shapley effects for b) fPC 1 and d) fPC 2 are estimated with $n = 1284$ and $N_{tot} = 800$. For the confidence intervals, $B = 500$ is used. The local slope is displayed with a gray line.

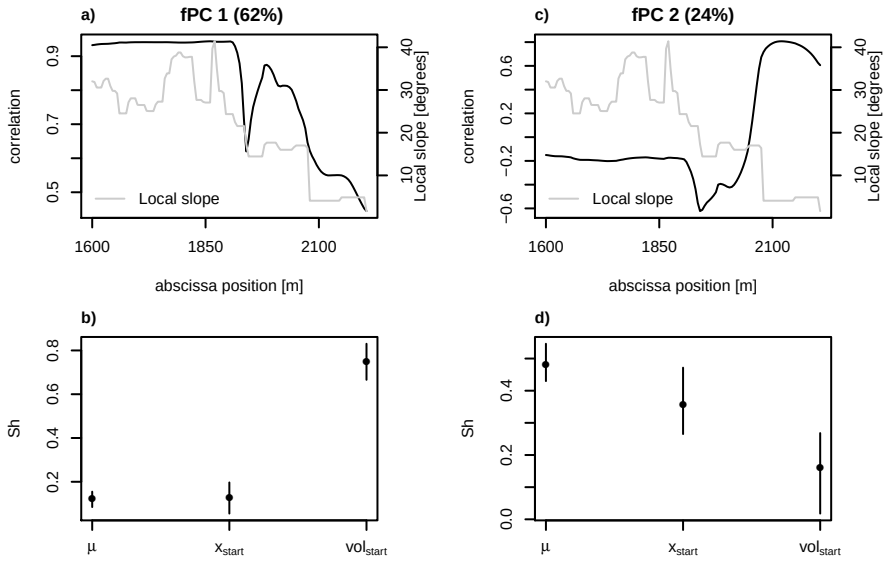

Figure 5.17: Avalanche model, scenario 2: a) and c) correlations between the first and second fPCs and the flow depth output. Scalar Shapley effects for b) fPC 1 and d) fPC 2 are estimated with $n = 1284$ and $N_{tot} = 800$. For the confidence intervals, $B = 500$ is used. The local slope is displayed with a gray line.

6 Conclusions and perspectives

In this thesis, we addressed the calibration and global sensitivity analysis of avalanche dynamic models. In general, we introduced and proposed advanced statistical methods for understanding better input parameters and outputs of snow avalanche models. All the techniques were illustrated with synthetic data and with real snow avalanche data to demonstrate their efficiency and usefulness for real applications. Our contributions may be useful for avalanche scientists in order to (i) account for the error structure in model calibration and (ii) to rank input parameters according to their importance in the output model.

- A framework based on Bayesian inference to calibrate avalanche dynamic models using high temporal resolution data.
- A nonparametric method based on Nadaraya-Watson kernel with bootstrap bias correction to estimate aggregated Sobol' indices from a given sample.
- A given data estimation method and associated confidence intervals for aggregated Shapley effects based on nearest-neighbor estimation.

The last two methods were applied to rank input parameters relatively to their impact on the variability of functional outputs of an avalanche dynamic model. In the next sections, we revisit in detail the contributions and conclusions of this manuscript and finally, we present the perspectives and open research questions.

6.1 Conclusions

Accounting for error structure in model calibration

In Chapter 3, we proposed a Bayesian inference framework to calibrate an avalanche dynamic model from an avalanche velocity time series with $\Delta t = 1s$. The time series was obtained from high rate positioning photogrammetry images of an avalanche released at the Lautaret test site on February 13 2013. We considered a sliding block model with Voellmy friction law. To perform the calibration under a Bayesian framework, assumptions about the model error distributions (defined as observations − model simulation) are required. Traditionally, the errors are assumed i.i.d. and normally distributed. However, as our data is a time series, the errors are autocorrelated. Thus, the objective of our work was to show how taking into account the autocorrelation might improve the friction

parameter estimation and the prediction of velocities. Within this purpose, we compared two statistical models: a first one that considers i.i.d. Gaussian centered errors and a second one that considers the errors modeled by an autoregressive process of order 1. We also generated synthetic data to illustrate how omitting the autocorrelation generates a bias for the estimation of friction parameters. Results showed that it is important to take into account the error structure (in our case, the autocorrelation) for model calibration to avoid bias.

Non parametric estimation method for aggregated Sobol' indices

In Chapter 4, our objective was to perform a GSA of an avalanche model. We considered the particular context of independent input parameters. The other constraint was that we had to propose a given data estimation procedure, accurate even for a sample of moderate size. Within this purpose, we proposed a nonparametric method to estimate the aggregated Sobol' indices from a given data sample. The nonparametric method was based on the Nadaraya-Watson kernel smoother. Since the sample size was moderate, the kernel estimation was biased and it required a bias correction. To remove such a bias, we proposed a bias correction based on the works of Racine [2001] and Solís [2019]. We tested the accuracy of the method on toy functions, and results were good even if sample sizes were moderate ($n = 300$). Then, we applied the method to estimate aggregated Sobol' indices of the depth averaged avalanche model proposed by Naaim et al. [2004]. In this study, we analyzed the sensitivity of the model for the avalanche event used in Chapter 3. More precisely, the friction parameter uncertainty intervals were given by the posterior Bayesian inference distributions and the uncertainty for other model input quantities were obtained from measures taken during the avalanche release. In this chapter, the input parameters were assumed to be independent. For this particular avalanche event, the input ξ, followed by μ, were the most important for the velocity output. For the flow depth and runout distance outputs, the snow depth within the released zone was fundamental.

Aggregated Shapley effects with bootstrap confidence intervals

In Chapter 5, our objective was to perform GSA of avalanche dynamic models in a more general framework than the one of Chapter 4. Indeed, in Chapter 4 we considered the input distribution associated to a specific avalanche event to guarantee the input independence. To perform GSA in a more general framework, we had to guarantee that all the samples were physically meaningful. Non physical sets of input parameters were then removed from the experimental design. For example, some combinations of friction parameters and released volume were preventing from an avalanche release or were producing small avalanches (volume below a threshold) which were not interesting for our study. Within this purpose, we developed a GSA method that can be applied to samples obtained from acceptance/rejection (AR) rules (corresponding to the physical constraints of the study). AR rules were leading to dependence among input parameters. Therefore we proposed to estimate Shapley effects extended to multivariate or functional outputs, which are more meaningful in the framework of dependent inputs. We called these sensitivity measures aggregated Shapley effects. For their estimation, we extended the given data estimator of Broto et al. [2020]. We also built confidence intervals by adapting ideas

of Benoumechiara and Elie-Dit-Cosaque [2019]. Then, we applied our GSA method to the avalanche dynamic model proposed by Naaim et al. [2004]. We found that the release position and volume were the most important inputs. Also, by looking at the evolution of ubiquitous Shapley effects along the path, we noticed that it might exist a relationship between the effects and the local slope.

6.2 Perspectives

6.2.1 For calibration using Bayesian inference

To reduce computational cost

A drawback of the framework proposed is that the Metropolis-Hasting algorithm, a MCMC algorithm used for sampling from the parameter posterior distribution, might require a high number of model runs to converge. In our application, we required at least 1000 model runs and computational cost was not a problem since, model runs were fast. However, for using Bayesian inference to calibrate a more complex model (with high number of parameters and high computational cost) some solutions could be: (i) to perform a GSA of the model with the aim of identifying the most important input parameters. Then, to only calibrate those ones and, fix the others to reference values. (ii) To construct a metamodel for the avalanche model and to perform the calibration with the metamodel. Users must also take into account the error induced by approximating the model with the metamodel. (iii) To apply other MCMC algorithms that might require less model runs to converge as for example the adaptive Metropolis-Hasting algorithm introduced by Haario et al. [2001]. In the adaptive MH, the proposal distribution q is updated in each iteration which makes convergence faster. Nevertheless, adaptive MH is non Markovian and some other theoretical nice properties of the MH algorithm might be lost.

To include other quantities within the calibration

In the study, we only used the avalanche velocity for the calibration. Thus, it could be interesting to add other quantities as for example, flow depth time series and runout distance. To do so, an appropriate likelihood function should be proposed (e.g., Gaussian process or Dirichlet process). However, a higher quantity of data would also be needed to optimize the new likelihood.

To apply the approach to several snow avalanches

It could be useful to apply the same approach to several avalanches to investigate, e.g., the relationship between friction parameters and snow conditions. Such findings could improve deeply our knowledge about the physics of the avalanche phenomenon. However, to do so, it would be necessary to acquire high-resolution avalanche data and the snow conditions of the snowpack for the same the avalanche events.

6.2.2 For the Global sensitivity analysis

To construct confidence intervals for non parametric indices estimation

It would be important to build confidence intervals for nonparametric bias corrected estimation of aggregated Sobol' indices. This task is not trivial because we cannot use bootstrap based confidence intervals since we already use bootstrap for correcting the bias and a double bootstrap is far too costly. Some interesting clues to explore could be (i) to use Bayesian inference to deduce credible intervals for the Sobol' indices [see Antoniano-Villalobos et al., 2019] and (ii) to prove asymptotic convergence properties of the aggregated Sobol' estimator. Nevertheless, the proof of such properties is a challenging task.

To apply GSA to other avalanche events and models

It could be useful to apply our GSA methods to other snow avalanche models and snow avalanche events to see if results could be generalized. The methods should be tested for example for 3-D spatio-temporal avalanche dynamic models [as for example the `r.avalflow` model of Mergili et al., 2017]. The output of such models are functional and dependent, thus the methods should be adjusted to this kind of outputs [see for example Marrel et al., 2011].

To include the topography as input parameter

In our studies, the topography was fixed but it could also be included in the GSA as a functional input. In such a case, our GSA method should be adapted to functional inputs [see for example, Iooss and Ribatet, 2009, Fruth et al., 2015]. Also, the fact of including the topography as input could provide more insights about the potential connection between the input parameters and the topographical characteristics of avalanche paths.

To compare our ranking with other input rankings

The GSA framework applied in Chapter 5 could be seen as a particular case of conditional sensitivity analysis (CSA) which aims at measuring the influence of the inputs on the output within a restricted domain [see Marrel and Chabridon, 2020]. In our application, the output domain was restricted by physical constraints. Thus it could be interesting to compare our rank to the input rank obtained by using the aggregated HSIC indices of Da Veiga [2015]. Moreover, Shapley effects might be used within the CSA framework, but this should be first study in test cases.

Appendices

A Given data estimation methods for Sobol' indices

A.1 Given data method

Plischke et al. [2013] [see also Borgonovo et al., 2016] proposed a given data method which relies on the notion of class-conditional densities, where a class is a sub-sample stemming from a suitable partition of the data. As its name suggests, the method does not require any specific sampling design to estimate the Sobol' indices (or other sensitivity measures). Hereafter, it is denoted by NSD (*non sampling design*).

Let us consider a scalar output $Y = f(\mathbf{X})$. Let $(x_1^j, \ldots, x_d^j, y^j)_{j\in\{1,\ldots,n\}}$ be a n sample of $(\mathbf{X}, Y)$. The support of the input X_i, denoted by $\mathcal{X}_i$, is partitioned into L disjoint classes $\{\mathcal{C}_\ell\}_{\ell=1,\ldots,L}$ such that $\bigcup_{\ell=1}^{L} \mathcal{C}_\ell = \mathcal{X}_i$ and $\mathcal{C}_{\ell_1} \cap \mathcal{C}_{\ell_2} = \emptyset$ for all $\ell_1 \neq \ell_2$. For notation simplicity, we remove i from the classes $\mathcal{C}_\ell$ but they depend on X_i. The estimation of S_i, the first-order scalar Sobol' index, proposed by Plischke et al. [2013] is:

$$\hat{S}_i = \frac{\sum_{\ell=1}^{L} n_\ell (\bar{y}_\ell - \bar{y})^2}{\sum_{j=1}^{n} (y^j - \bar{y})^2}$$

where n_ℓ counts the number of realizations of X_i in $\mathcal{C}_\ell$, $\bar{y}$ and s_y^2 denotes the empirical mean and variance of Y, *resp.*, $\bar{y}_\ell$ represents the within sample class mean and it is used to estimate the conditional mean of $Y|X_i \in \mathcal{C}_\ell$:

$$\bar{y}_\ell = \frac{1}{n_\ell} \sum_{j:x_i^j \in \mathcal{C}_\ell} y_j$$

The main drawback of this estimation method is the bias induced by the choice of the partition [Antoniano-Villalobos et al., 2019]. Moreover, in [Antoniano-Villalobos et al., 2019, Appendix A.2] different values of L were analyzed but till this moment, there is no universal valid rule to choose L. [Theorem 1 of Borgonovo et al., 2016] states that the estimator is consistent if the number of classes can be written as $L = g(n)$ where g is a monotonically increasing function such that:

$$\lim_{n\to\infty} \frac{n}{g(n)} = \infty.$$

For the test cases shown in Section A.3, we take $L = \sqrt{n}$.

A.2 The spectral approach EASI

The effective algorithm for computing global sensitivity indices (EASI) proposed by Plischke [2010] is a spectral method to estimate first-order Sobol' indices using the Fast Fourier Transform. Spectral approaches (as FAST and EASI) should be preferred if the model f has some regularity [Prieur and Tarantola, 2017]. Let $(x_1^\ell, \ldots, x_d^\ell, y^\ell)_{\ell\in\{1,\ldots,n\}}$ be a n sample of $(\mathbf{X}, Y)$. The sample vector $\mathbf{x}_i = (x_i^\ell)_{\ell\in\{1,\ldots,n\}}$ of input X_i is ordered increasingly to obtain a new vector $\mathbf{x}_i^{(\ell)}$ which elements satisfy $x_i^{(1)} \leq \ldots \leq x_i^{(n)}$. Then, all odd indices from $(x_i^{(\ell)})_{\ell\in\{1,\ldots,n\}}$ are ordered increasingly and followed by all even indices in decreasing order to obtain the vector $(x_i^{[\ell]})$:

$$x_i^{[\ell]} = \begin{cases} x_i^{(2\ell-1)} & \ell \leq \frac{n+1}{2}, \\ x_i^{2(n+1-\ell)} & \ell > \frac{n+1}{2} \end{cases}$$

The elements of $(x_i^{[\ell]})$ satisfy a zig-zag relation:

$$x_i^{[\ell]} \leq x_i^{[\ell+1]} \text{ if } \ell \leq \frac{n+1}{2} \text{ and } x_i^{[\ell]} \geq x_i^{[\ell+1]} \text{ if } \ell \geq \frac{n+1}{2}.$$

Let us denote by π the permutation of $\{1, \ldots, n\}$ such that $x_i^{\pi(\ell)} = x_i^{[\ell]}$
The first-order sensitivity index is then estimated by:

$$\tilde{S}_i = 2\frac{\sum_{m=1}^{M} |c_m|^2}{\sum_{m\neq 0} |c_m|^2} \tag{A.1}$$

where c_m is the m-complex coefficient of the discrete Fourier transform of $(y^{\pi(\ell)})_{\ell\in\{1,\ldots,n\}}$:

$$c_m = \sum_{\ell=1}^{n} y^{\pi(\ell)} e^{\frac{2i\pi m(\ell-1)}{n}},$$

with $m = 0, \pm 1, \pm 2, \ldots, \lfloor \frac{n}{2} \rfloor$ with $\lfloor . \rfloor$ denoting the integer part. The maximum harmonic M is usually taken as 4 or 6 [Plischke, 2010]. The estimation introduced in (A.1) is biased with respect to the maximal number of harmonics M, thus Plischke [2010] proposed a bias correction which we use in the test cases:

$$\check{S}_i = \frac{1}{n - 2M}(n\tilde{S}_i - 2M) \tag{A.2}$$

A.3 Test cases

We study the accuracy of NSD and EASI on two test cases. The methods are also compared to the pick-and-freeze method based on replicated orthogonal array proposed by Mara and Joseph [2008] and further studied in Tissot and Prieur [2015], we denoted it by *sobol-roalhs*. This last method requires specific designs thus is not a given data method. It is used as a benchmark approach. The R package `sensitivity` is used for its estimation.

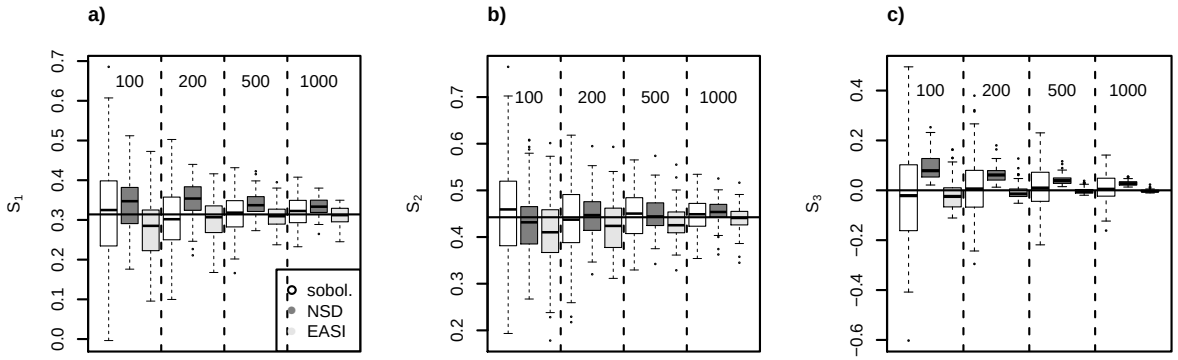

Figure A.1: *Ishigami function.* Estimation of Sobol' indices using sobolroalhs, NSD and EASI. For each sample size $n \in \{100, 200, 500, 1000\}$, $N = 100$ samples are used. The true sensitivity indices values are shown with black lines.

A.3.1 Ishigami function

The Ishigami function [Ishigami and Homma, 1990] is a classical test function which is nonlinear and non monotonic:

$$f(X_1, X_2, X_3) = \sin(X_1) + 7\sin^2(X_2) + 0.1X_3^4 \sin(X_1),$$

the inputs X_1, X_2, X_3 are uniformly distributed on $[-\pi, \pi]$.

Figure A.1 shows the estimation of the Sobol' indices using NSD, EASI and sobolroalhs. The true sensitivity indices of the Ishigami function are known and shown with black lines. In order to see the evolution of the accuracy of estimation, the number of model evaluations n varies from 100 to 1000 . For each sample size n, $N = 100$ random samples are used. The number of classes required by NSD was set to $\sqrt{n}$. The number of harmonics of EASI was set to 6 following the recommendation of Plischke [2010]. When sample sizes are small, the three methods show small bias. However, as the sample size increases the bias reduces. EASI shows accurate results even with small sample sizes for S_1 and S_3 ($n = 200$). NSD does not estimate accurately indices close to 0 (S_3).

A.3.2 Bartley et al. function

The second test function is the Bartley et al. function:

$$f(X_1, \ldots, X_6) = \sum_{i=1}^{6} \prod_{j=1}^{i} X_j, \tag{A.3}$$

$\mathbf{X} = (X_1, X_2, \ldots, X_6)$ is assumed to be uniformly distributed on $[0, 1]^6$.

Figure A.2 shows the estimation of Sobol' indices using the three methods for the Bartley et al. functions. Independently of the sample size, NSD estimation shows always a bias. EASI shows good results when the sample size increases.

In conclusion, within these two example we showed that EASI is a nice option to estimate Sobol' indices with small given data samples. The choice of M does not seem to

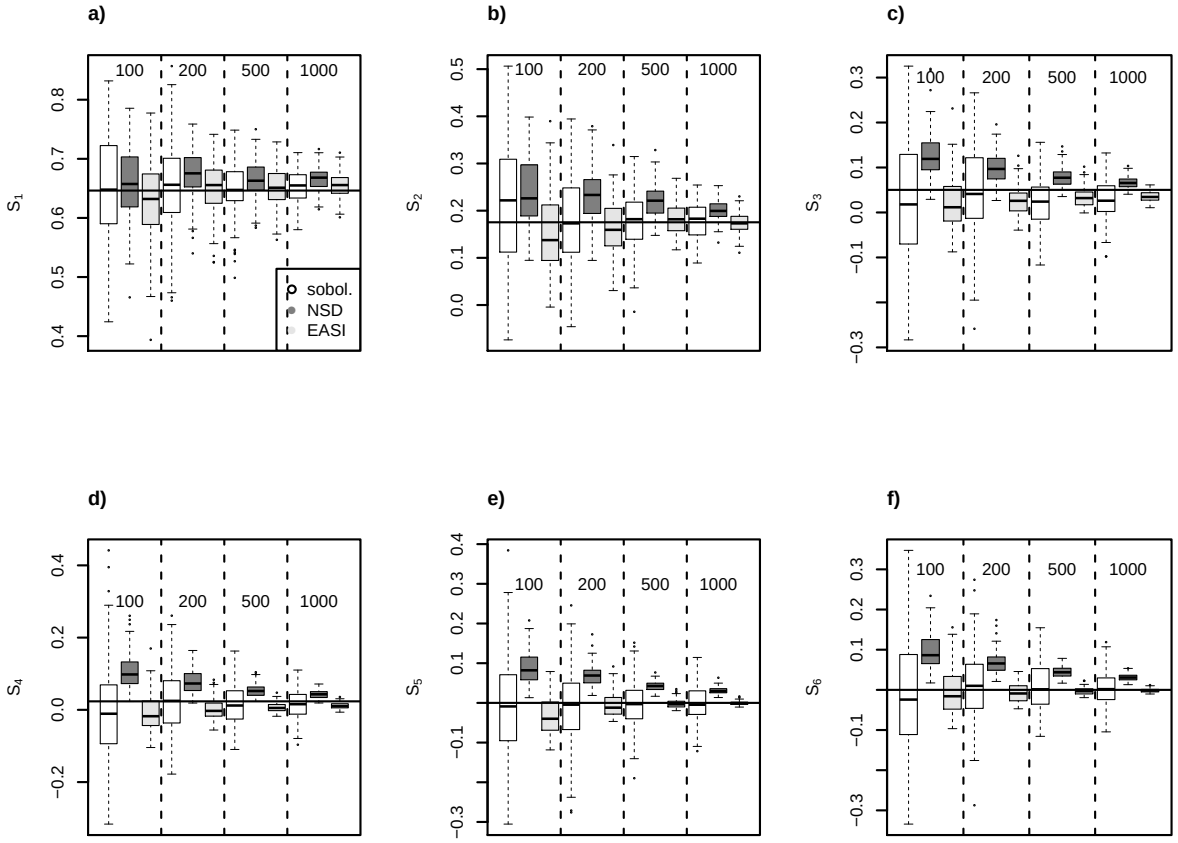

Figure A.2: *Bartley et al. function.* Estimation of Sobol' indices using sobolroalhs, NSD and EASI. For each sample size $n \in \{100, 200, 500, 1000\}$, $N = 100$ samples are used. The true sensitivity indices values are shown with black lines.

impact that much the results. However, for highly non regular models we could expect that $K = 6$ is too small, even if the reordering of the sample regularizes the functions. NSD is biased, and to our knowledge, there does not exist any selection procedure for the number of classes for small to moderate sample sizes.

B

Correlations between Shapley effects and local slope

In Chapter 5, we found that the correlations between Shapley effects and local slope were significant. The local slope term assigns the angle between the abscissa and the corridor. This fact might suggest that local slope changes play a role in the input importance. Figures B.1 and B.3 show the scatterplots of the ubiquitous Shapley effects versus the local slope for scenarios 1 and 2, respectively. Figures B.2 and B.4 show the scatterplots of the aggregated Shapley effects and local slope for scenarios 1 and 2, respectively. The correlations are displayed in Table B.1 and Table B.2 for scenarios 1 and 2, respectively. An interesting finding is that the correlations between μ and the local slope are always negative, this could suggest that if the local slope is high the importance of μ is low. However, to generalize the results, Shapley effects should be estimated in other corridors. It would also be interesting to include the topography in the global sensitivity analysis.

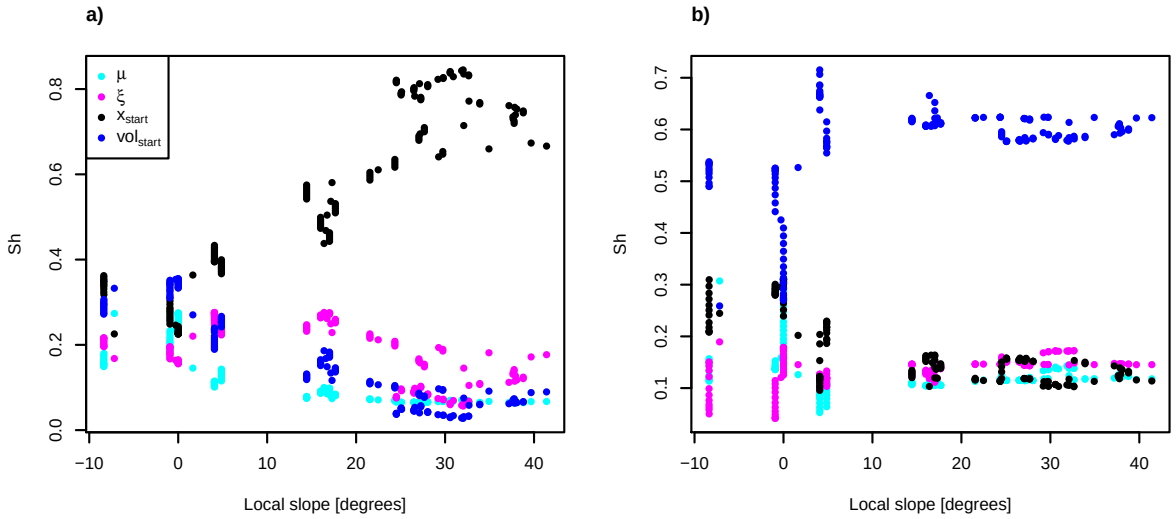

Figure B.1: *Scenario 1.* Scatterplots of ubiquitous Shapley effects and local slope for a) flow velocity and b) flow depth.

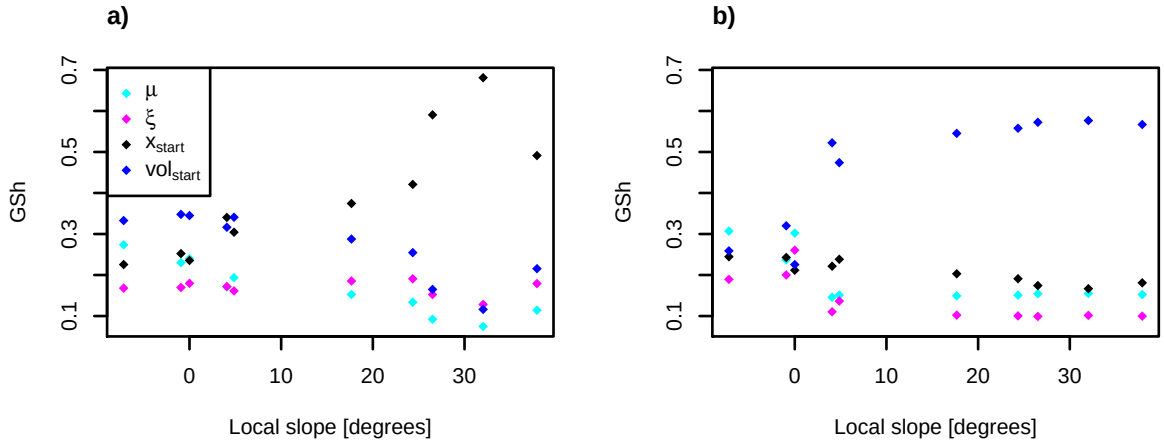

Figure B.2: *Scenario 1.* Scatterplots of aggregated Shapley effects and local slope for a) flow velocity and b) flow depth.

	Ubiquitous effects		Aggregated effects	
	Flow velocity [m s^{-1}]	Flow depth [m]	Flow velocity [m s^{-1}]	Flow depth [m]
μ	-0.79	-0.28	-0.93	-0.70
ξ	-0.50	0.54	-0.20	-0.75
x_{start}	0.92	-0.76	0.89	-0.92
vol_{start}	-0.92	0.49	-0.87	0.82

Table B.1: *Scenario 1.* Correlations between Shapley effects and local slope.

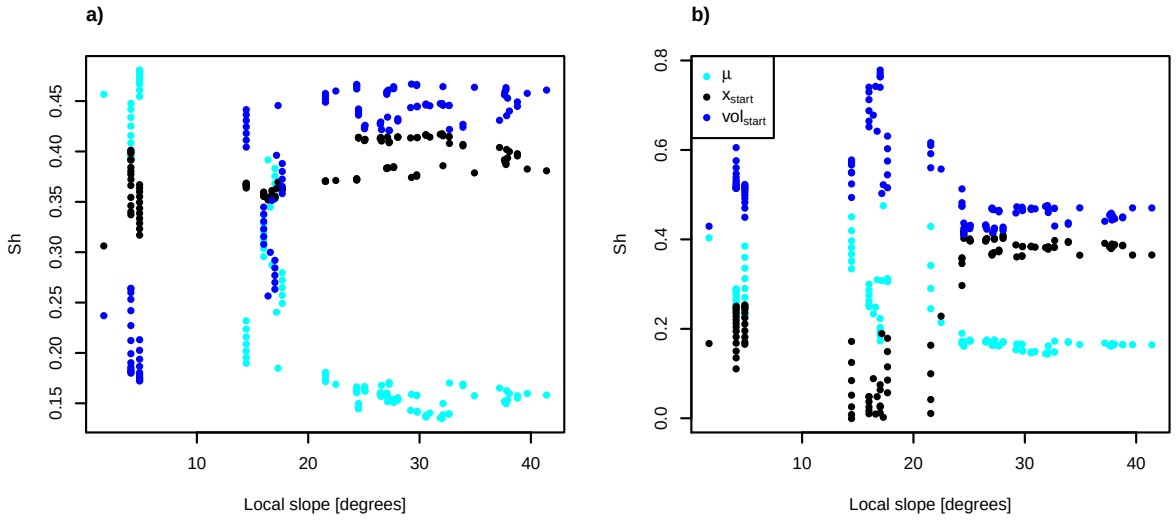

Figure B.3: *Scenario 2.* Scatterplots of ubiquitous Shapley effects and local slope for a) flow velocity and b) flow depth.

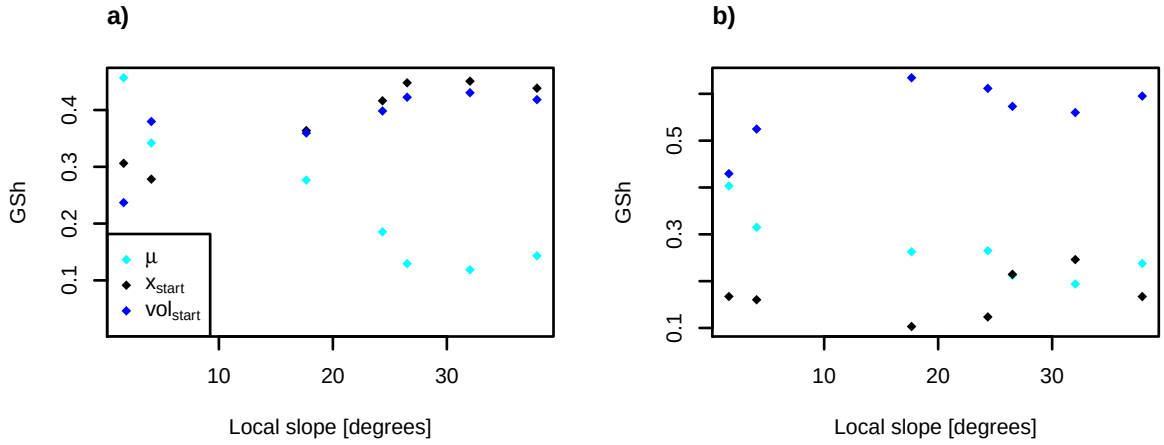

Figure B.4: *Scenario 2.* Scatterplots of aggregated Shapley effects and local slope for a) flow velocity and b) flow depth.

	Ubiquitous effects		Aggregated effects	
	Flow velocity [m s^{-1}]	Flow depth [m]	Flow velocity [m s^{-1}]	Flow depth [m]
μ	-0.89	-0.67	-0.96	-0.93
x_{start}	0.64	-0.76	0.95	0.59
vol_{start}	0.87	-0.42	0.83	0.54

Table B.2: *Scenario 2.* Correlations between Shapley effects and local slope.

Bibliography

WSL Institute for Snow and Avalanche Research SFL. https://www.slf.ch/en/avalanches/avalanche-protection/planning-measures.html. Accessed: 2020-08-20. (Cited on page 8.)

Alen Alexanderian, Pierre A. Gremaud, and Ralph C. Smith. Variance-based sensitivity analysis for time-dependent processes. *Reliability Engineering & System Safety*, 196:106722, 2020. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2019.106722. URL http://www.sciencedirect.com/science/article/pii/S0951832019303837. (Cited on pages 20, 22, 74, 75, 76, 81, and 97.)

C. Ancey. *Snow Avalanches*, pages 319–338. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. ISBN 978-3-540-45670-4. doi: 10.1007/3-540-45670-8_13. URL https://doi.org/10.1007/3-540-45670-8_13. (Cited on page 1.)

C. Ancey and M. Meunier. Estimating bulk rheological properties of flowing snow avalanches from field data. *Journal of Geophysical Research: Earth Surface*, 109(F1), 2004. doi: 10.1029/2003JF000036. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003JF000036. (Cited on pages 6, 25, 26, 27, 38, and 49.)

C. Ancey, M. Meunier, and D. Richard. Inverse problem in avalanche dynamics models. *Water Resources Research*, 39(4), 2003. doi: 10.1029/2002WR001749. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002WR001749. (Cited on pages 6, 25, and 27.)

Christophe Ancey. Monte Carlo Calibration of Avalanches Described as Coulomb Fluid Flows. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 363(1832):1529–1550, 2005. ISSN 1364503X. URL http://www.jstor.org/stable/30039672. (Cited on pages 7 and 25.)

Christophe Ancey. *Dynamique des avalanches*. Presses Polytechniques et Universitaires romandes & Cemagref, 2006. (Cited on pages 25 and 49.)

Christophe Ancey. Snow avalanches. *Oxford Research Encyclopedia*, 2016. doi: 10.1093/acrefore/9780199389407.013.17. URL http://infoscience.epfl.ch/record/217280. Published online. (Cited on page 25.)

Anestis Antoniadis, Céline Helbert, Clémentine Prieur, and Laurence Viry. Spatio-temporal metamodeling for West African monsoon. *Environmetrics*, 23(1):24–36, 2012. doi: 10.1002/env.1134. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1134. (Cited on page 55.)

Isadora Antoniano-Villalobos, Emanuele Borgonovo, and Xuefei Lu. Nonparametric estimation of probabilistic sensitivity measures. *Statistics and Computing*, Aug 2019. ISSN 1573-1375. doi: 10.1007/s11222-019-09887-9. URL https://doi.org/10.1007/s11222-019-09887-9. (Cited on pages 14, 50, 105, and 107.)

Benjamin Auder and Bertrand Iooss. Global sensitivity analysis based on entropy. In *In Safety, Reliability and Risk Analysis - Proceedings of the ESREL 2008 Conference*, pages 2107–2115. CRC Press, 2008. (Cited on page 49.)

M. Barbolini and F. Savi. Estimate of uncertainties in avalanche hazard mapping. *Annals of Glaciology*, 32:299–305, 2001. doi: 10.3189/172756401781819373. (Cited on pages 7, 29, 51, and 68.)

P. Bartelt, Y. Bühler, O. Buser, M. Christen, and L. Meier. Modeling mass-dependent flow regime transitions to predict the stopping and depositional behavior of snow avalanches. *Journal of Geophysical Research (Earth Surface)*, 117(F1):F01015, February 2012. doi: 10.1029/2010JF001957. (Cited on pages 26, 89, and 93.)

Perry Bartelt, Othmar Buser, Cesar Vera Valero, and Yves Bühler. Configurational energy and the formation of mixed flowing/powder snow and ice avalanches. *Annals of Glaciology*, 57(71):179–188, 2016. doi: 10.3189/2016AoG71A464. (Cited on page 25.)

Manel Baucells and Emanuele Borgonovo. Invariant probabilistic sensitivity analysis. *Management Science*, 59(11):2536–2549, 2013. doi: 10.1287/mnsc.2013.1719. URL https://doi.org/10.1287/mnsc.2013.1719. (Cited on page 19.)

Thomas Bayes. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, 53:370–418, 1763. (Cited on page 28.)

Nazih Benoumechiara and Kevin Elie-Dit-Cosaque. Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: ProcS*, 65: 266–293, 2019. doi: 10.1051/proc/201965266. URL https://doi.org/10.1051/proc/201965266. (Cited on pages 4, 17, 75, 76, 82, 96, and 104.)

L. Mark Berliner. Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003. doi: 10.1029/2002JD002865. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD002865. (Cited on page 28.)

José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009. doi: 10.1002/9780470316870. (Cited on page 28.)

Géraud Blatman and Bruno Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering & System Safety*, 95(11):1216 – 1229, 2010. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2010.06.015. URL http://www.sciencedirect.com/science/article/pii/S0951832010001493. (Cited on page 20.)

E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771 – 784, 2007. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2006.04.015. URL http://www.sciencedirect.com/science/article/pii/S0951832006000883. (Cited on pages 18 and 19.)

Emanuele Borgonovo and Elmar Plischke. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3):869–887, 2016. doi: 10.1016/j.ejor.2015.06.032. URL https://doi.org/10.1016/j.ejor.2015.06.032. (Cited on pages 49 and 74.)

Emanuele Borgonovo, William Castaings, and Stefano Tarantola. Moment independent importance measures: New results and analytical test cases. *Risk Analysis*, 31(3):404–428, 2011. doi: 10.1111/j.1539-6924.2010.01519.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2010.01519.x. (Cited on page 18.)

Emanuele Borgonovo, Gordon B. Hazen, and Elmar Plischke. A Common Rationale for Global Sensitivity Measures and Their Estimation. *Risk Analysis*, 36(10):1871–1895, 2016. doi: 10.1111/risa.12555. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.12555. (Cited on pages 20, 49, 74, and 107.)

Christopher P. Borstad and D.M. McClung. Sensitivity analyses in snow avalanche dynamics modeling and implications when modeling extreme events. *Canadian Geotechnical Journal*, 46(9):1024–1033, 2009. doi: 10.1139/T09-042. URL https://doi.org/10.1139/T09-042. (Cited on pages 7, 25, 29, 51, 68, 92, and 93.)

E. Bourova, E. Maldonado, J.B. Leroy, R. Alouani, N. Eckert, M. Bonnefoy-Demongeot, and M. Deschatres. A new web-based system to improve the monitoring of snow avalanche hazard in France. *Natural Hazards and Earth System Sciences*, 16(5):1205–1216, 2016. doi: 10.5194/nhess-16-1205-2016. URL https://hal.archives-ouvertes.fr/hal-01507671. (Cited on page 88.)

W Brian Dade and Herbert E Huppert. Long-runout rockfalls. *Geological Society of America*, 26(9):803–806, 1998. doi: 10.1130/0091-7613(1998)026⟨0803:LRR⟩2.3.CO;2. (Cited on page 89.)

Baptiste Broto. *Sensitivity analysis with dependent variables: Estimation of the Shapley effects for unknown input distribution and linear Gaussian models*. PhD thesis, Université Paris-Saclay, 2020. (Cited on page 16.)

Baptiste Broto, François Bachoc, Marine Depecker, and Jean-Marc Martinez. Sensitivity indices for independent groups of variables. *Mathematics and Computers in Simulation*, 163:19 – 31, 2019. ISSN 0378-4754. doi: https://doi.org/10.1016/j.matcom.2019.02.008. URL http://www.sciencedirect.com/science/article/pii/S0378475419300564. (Cited on page 15.)

Baptiste Broto, François Bachoc, and Marine Depecker. Variance Reduction for Estimation of Shapley Effects and Adaptation to Unknown Input Distribution. *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):693–716, 2020. doi: 10.1137/18M1234631. URL https://doi.org/10.1137/18M1234631. (Cited on pages ix, 4, 16, 17, 51, 52, 73, 75, 76, 79, 80, 82, 85, 96, 97, 98, and 103.)

R. Buhler, C. Argue, B. Jamieson, and A. Jones. Sensitivity Analysis of the RAMMS Avalanche Dynamics Model in a Canadian Transitional Snow Climate. ISSW, Innsbruck, Austria, 2018. (Cited on pages 7, 29, and 51.)

Yves Bühler, Marc Christen, Julia Kowalski, and Perry Bartelt. Sensitivity of snow avalanche simulations to digital elevation model quality and resolution. *Annals of Glaciology*, 52(58):72–80, 2011. doi: 10.3189/172756411797252121. (Cited on pages 7, 31, and 51.)

Katherine Campbell, Michael D. McKay, and Brian J. Williams. Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety*, 91(10):1468–1472, 2006. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2005.11.049. URL http://www.sciencedirect.com/science/article/pii/S0951832005002565. The Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004). (Cited on pages 21, 22, and 50.)

Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509 – 1518, 2007. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2006.10.004. URL http://www.sciencedirect.com/science/article/pii/S1364815206002805. Modelling, computer-assisted simulations, and mapping of dangerous phenomena for hazard assessment. (Cited on page 11.)

William B. Capra and Hans-Georg Müller. An accelerated-time model for response curves. *Journal of the American Statistical Association*, 92(437):72–83, 1997. doi: 10.1080/01621459.1997.10473604. URL https://doi.org/10.1080/01621459.1997.10473604. (Cited on page 82.)

M. Carmassi, P. Barbillon, M. Keller, E. Parent, and M. Chiodetti. Bayesian calibration of a numerical code for prediction. *ArXiv e-prints*, January 2018. (Cited on pages 8 and 25.)

G. Casassa, H. Narita, and N. Maeno. Measurements of friction coefficients of snow blocks. *Annals of Glaciology*, 13:40–44, 1989. doi: 10.3189/S0260305500007618. (Cited on page 25.)

W. Castaings, D. Dartus, F.-X. Le Dimet, and G.-M. Saulnier. Sensitivity analysis and parameter estimation for distributed hydrological modeling: potential of variational methods. *Hydrology and Earth System Sciences*, 13(4):503–517, 2009. doi: 10.5194/hess-13-503-2009. URL https://hess.copernicus.org/articles/13/503/2009/. (Cited on page 10.)

H. Castebrunet, N. Eckert, and G. Giraud. Snow and weather climatic control on snow avalanche occurrence fluctuations over 50 yr in the French Alps. *Climate of the Past*,

8(2):855–875, 2012. doi: 10.5194/cp-8-855-2012. URL https://www.clim-past.net/8/855/2012/. (Cited on page 25.)

Enrique Castillo, Ali S. Hadi, Antonio Conejo, and Alfonso Fernández-Canteli. A general method for local sensitivity analysis with application to regression models and other optimization problems. *Technometrics*, 46(4):430–444, 2004. ISSN 00401706. URL http://www.jstor.org/stable/25470891. (Cited on page 10.)

Javier Castro, Daniel Gómez, and Juan Tejada. Polynomial calculation of the shapley value based on sampling. *Computers & Operations Research*, 36(5):1726 – 1730, 2009. ISSN 0305-0548. doi: https://doi.org/10.1016/j.cor.2008.04.004. URL http://www.sciencedirect.com/science/article/pii/S0305054808000804. Selected papers presented at the Tenth International Symposium on Locational Decisions (ISOLDE X). (Cited on page 17.)

G. Chastaing, F. Gamboa, and C. Prieur. Generalized sobol sensitivity indices for dependent variables: numerical methods. *Journal of Statistical Computation and Simulation*, 85(7):1306–1333, 2015. doi: 10.1080/00949655.2014.960415. URL https://doi.org/10.1080/00949655.2014.960415. (Cited on page 15.)

Gaelle Chastaing, Fabrice Gamboa, and Clémentine Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis. *Electron. J. Statist.*, 6:2420–2448, 2012. doi: 10.1214/12-EJS749. URL https://doi.org/10.1214/12-EJS749. (Cited on page 15.)

Yaqing Chen, Cody Carroll, Xiongtao Dai, Jianing Fan, Pantelis Z. Hadjipantelis, Kyunghee Han, Hao Ji, Hans-Georg Mueller, and Jane-Ling Wang. *fdapace: Functional Data Analysis and Empirical Dynamics*, 2020. URL https://CRAN.R-project.org/package=fdapace. R package version 0.5.1. (Cited on page 84.)

M. Christen, J. Kowalski, and P. Bartelt. Ramms: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Regions Science and Technology*, 63(1):1 – 14, 2010a. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2010.04.005. URL http://www.sciencedirect.com/science/article/pii/S0165232X10000844. (Cited on page 26.)

Marc Christen, Perry Bartelt, and Julia Kowalski. Back calculation of the In den Arelen avalanche with RAMMS: interpretation of model results. *Annals of Glaciology*, 51(54): 161–168, 2010b. doi: 10.3189/172756410791386553. (Cited on page 6.)

James S. Clark. Why environmental scientists are becoming Bayesians. *Ecology Letters*, 8 (1):2–14, 2005. doi: 10.1111/j.1461-0248.2004.00702.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2004.00702.x. (Cited on page 28.)

Paul G. Constantine and Paul Diaz. Global sensitivity metrics from active subspaces. *Reliability Engineering & System Safety*, 162:1 – 13, 2017. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2017.01.013. URL http://www.sciencedirect.com/science/article/pii/S0951832016303052. (Cited on pages 19 and 20.)

R.I Cukier, H.B Levine, and K.E Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26(1):1 – 42, 1978. ISSN 0021-9991. doi: https://doi.org/10.1016/0021-9991(78)90097-9. URL http://www.sciencedirect.com/science/article/pii/0021999178900979. (Cited on page 14.)

Sebastien Da Veiga. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305, 2015. doi: 10.1080/00949655.2014.945932. URL https://doi.org/10.1080/00949655.2014.945932. (Cited on pages 19, 21, 74, and 105.)

Sébastien Da Veiga and Fabrice Gamboa. Efficient estimation of sensitivity indices. *Journal of Nonparametric Statistics*, 25(3):573–595, September 2013. doi: 10.1080/10485252.2013.784. (Cited on pages 3 and 14.)

Sebastien Da Veiga, Francois Wahl, and Fabrice Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463, 2009. doi: 10.1198/TECH.2009.08124. URL https://doi.org/10.1198/TECH.2009.08124. (Cited on pages 15, 50, 51, 52, and 55.)

Gregory Delipei. *Développement d'une méthodologie de Quantification d'Incertitudes pour une analyse Multi-Physique Best Estimate et application sur un Accident d'Éjection de Grappe dans un Réacteur à Eau Pressurisée*. PhD thesis, 2019. URL http://www.theses.fr/2019SACLX078. Thèse de doctorat dirigée par Garnier, Josselin Mathématiques appliquées Université Paris-Saclay (ComUE) 2019. (Cited on page 75.)

J. D. Dent and T. E. Lang. Modeling of snow flow. *Journal of Glaciology*, 26(94):131–140, 1980. doi: 10.3189/S0022143000010674. (Cited on pages 25, 35, and 93.)

J.D. Dent, K.J. Burrell, D. S. Schmidt, M.Y. Louge, E.E. Adams, and T.G. Jazbutis. Density, velocity and friction measurements in a dry-snow avalanche. *Annals of Glaciology*, 26:247–252, 1998. doi: 10.3189/1998AoG26-1-247-252. (Cited on pages 6 and 25.)

Kjell Doksum and Alexander Samarov. Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Ann. Statist.*, 23 (5):1443–1473, 10 1995. doi: 10.1214/aos/1176324307. (Cited on page 50.)

Lisa Dreier, Yves Bühler, Christian Ginzler, and Perry Bartelt. Comparison of simulated powder snow avalanches with photogrammetric measurements. *Annals of Glaciology*, 57(71):371–381, 2016. doi: 10.3189/2016AoG71A532. (Cited on page 26.)

Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222, 1987. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(87)91197-X. URL http://www.sciencedirect.com/science/article/pii/037026938791197X. (Cited on page 9.)

N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro. Anova kernels and rkhs of zero mean functions for model-based sensitivity analysis. *Journal of Multivariate Analysis*, 115:57 – 67, 2013. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2012.08.016. URL http://www.sciencedirect.com/science/article/pii/S0047259X1200214X. (Cited on pages 20 and 51.)

N. Eckert, E. Parent, and D. Richard. Revisiting statistical-topographical methods for avalanche predetermination: Bayesian modelling for runout distance predictive distribution. *Cold Regions Science and Technology*, 49(1):88 – 107, 2007. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2007.01.005. URL http://www.sciencedirect.com/science/article/pii/S0165232X07000080. Selected Papers from the General Assembly of the European Geosciences Union (EGU), Vienna, Austria, 25 April 2005. (Cited on pages 8, 25, and 27.)

N. Eckert, E. Parent, T. Faug, and M. Naaim. Optimal design under uncertainty of a passive defense structure against snow avalanches: from a general Bayesian framework to a simple analytical model. *Natural Hazards and Earth System Sciences*, 8(5):1067–1081, 2008a. doi: 10.5194/nhess-8-1067-2008. URL https://www.nat-hazards-earth-syst-sci.net/8/1067/2008/. (Cited on page 88.)

N. Eckert, E. Parent, M. Naaim, and D. Richard. Bayesian stochastic modelling for avalanche predetermination: from a general system framework to return period computations. *Stochastic Environmental Research and Risk Assessment*, pages 185–206, Feb 2008b. ISSN 1436-3259. doi: 10.1007/s00477-007-0107-4. URL https://doi.org/10.1007/s00477-007-0107-4. (Cited on pages 8, 25, and 49.)

N. Eckert, E. Parent, T. Faug, and M. Naaim. Bayesian optimal design of an avalanche dam using a multivariate numerical avalanche model. *Stochastic Environmental Research and Risk Assessment*, 23(8):1123–1141, Dec 2009. ISSN 1436-3259. doi: 10.1007/s00477-008-0287-6. URL https://doi.org/10.1007/s00477-008-0287-6. (Cited on page 25.)

N. Eckert, M. Naaim, and E. Parent. Long-term avalanche hazard assessment with a Bayesian depth averaged propagation model. *Journal of Glaciology*, 56:563–586, October 2010. doi: 10.3189/002214310793146331. (Cited on pages 7, 25, 34, 49, 76, 88, 90, and 95.)

N. Eckert, C.J. Keylock, D. Bertrand, E. Parent, T. Faug, P. Favier, and M. Naaim. Quantitative risk and optimal design approaches in the snow avalanche field: Review and extensions. *Cold Regions Science and Technology*, 79-80:1 – 19, 2012. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2012.03.003. URL http://www.sciencedirect.com/science/article/pii/S0165232X12000651. (Cited on page 25.)

Nicolas Eckert, Mohamed Naaim, Florie Giacona, Philomène Favier, Aurore Lavigne, Didier Richard, Franck Bourrier, and Eric Parent. Repenser les fondements du zonage réglementaire des risques en montagne récurrents. *La Houille Blanche*, (2):38–67, 2018a. (Cited on page 49.)

Nicolas Eckert, Naaim, Mohamed, Giacona, Florie, Favier, Philomène, Lavigne, Aurore, Richard, Didier, Bourrier, Franck, and Parent, Eric. Repenser les fondements du zonage règlementaire des risques en montagne " récurrents ". *La Houille Blanche*, (2):38–67, 2018b. doi: 10.1051/lhb/2018019. URL https://doi.org/10.1051/lhb/2018019. (Cited on pages 8 and 88.)

B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Statist.*, 9(3):586–596, 05 1981. doi: 10.1214/aos/1176345462. URL https://doi.org/10.1214/aos/1176345462. (Cited on pages 13 and 49.)

B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, 1(1):54–75, 02 1986. doi: 10.1214/ss/1177013815. URL https://doi.org/10.1214/ss/1177013815. (Cited on page 82.)

Bradley Efron. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9(2):139–158, 1981. doi: 10.2307/3314608. URL https://onlinelibrary.wiley.com/doi/abs/10.2307/3314608. (Cited on pages 82 and 84.)

Bradley Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987. ISSN 01621459. URL http://www.jstor.org/stable/2289144. (Cited on page 84.)

Bradley Efron and Gail Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983. ISSN 00031305. URL http://www.jstor.org/stable/2685844. (Cited on page 54.)

Guillaume Evin, Mark Thyer, Dmitri Kavetski, David McInerney, and George Kuczera. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, 50(3):2350–2375, 2014. ISSN 1944-7973. doi: 10.1002/2013WR014185. URL http://dx.doi.org/10.1002/2013WR014185. (Cited on pages 3, 9, and 26.)

W Faig, TY Shih, and XS Liu. A non-metric stereo system and its calibration. In *Proceeding of the ACSM-ASPRS Annual Convention, Denver, Colorado (USA) vol*, volume 5, pages 18–25, 1990. (Cited on page 31.)

Jianqing Fan and Irène Gijbels. *Local polynomial modelling and its applications*. Number 66 in Monographs on Statistics and applied Probability series. Chapman & Hall, London, 1996. ISBN 0412983214. doi: 10.1201/9780203748725. (Cited on page 81.)

T. Faug, B. Turnbull, and P. Gauer. Looking beyond the powder/dense flow avalanche dichotomy. *Journal of Geophysical Research: Earth Surface*, 123(6):1183–1186, 2018. doi: 10.1002/2018JF004665. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018JF004665. (Cited on page 25.)

P. Favier, D. Bertrand, N. Eckert, and M. Naaim. A reliability assessment of physical vulnerability of reinforced concrete walls loaded by snow avalanches. *Natural Hazards and Earth System Sciences*, 14(3):689–704, 2014a. doi: 10.5194/nhess-14-689-2014. URL https://www.nat-hazards-earth-syst-sci.net/14/689/2014/. (Cited on page 25.)

P. Favier, N. Eckert, D. Bertrand, and M. Naaim. Sensitivity of avalanche risk to vulnerability relations. *Cold Regions Science and Technology*, 108:163 – 177, 2014b. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2014.08.009. URL http://www.sciencedirect.com/science/article/pii/S0165232X14001426. (Cited on pages 25, 49, and 88.)

CRLA Fierz, Richard L Armstrong, Yves Durand, Pierre Etchevers, Ethan Greene, David M McClung, Kouichi Nishimura, Pramod K Satyawali, and Sergey A Sokratov. *The International Classification for Seasonal Snow on the Ground*, volume 5. UNESCO/IHP, 2009. (Cited on page 30.)

J.-T. Fischer. A novel approach to evaluate and compare computational snow avalanche simulation. *Natural Hazards and Earth System Sciences*, 13(6):1655–1667, 2013. doi: 10.5194/nhess-13-1655-2013. URL https://www.nat-hazards-earth-syst-sci.net/13/1655/2013/. (Cited on pages 25 and 49.)

Jan-Thomas Fischer, Reinhard Fromm, Peter Gauer, and Betty Sovilla. Evaluation of probabilistic snow avalanche simulation ensembles with Doppler radar observations. *Cold Regions Science and Technology*, 97(Supplement C):151 – 158, 2014. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2013.09.011. URL http://www.sciencedirect.com/science/article/pii/S0165232X13001481. (Cited on pages 7 and 25.)

Jan-Thomas Fischer, Andreas Kofler, Wolfgang Fellin, Matthias Granig, and Karl Kleemayr. Multivariate parameter optimization for computational snow avalanche simulation. *Journal of Glaciology*, 61(229):875–888, 2015. doi: 10.3189/2015JoG14J168. (Cited on pages 7, 25, and 68.)

Jan-Thomas Fischer, Andreas Kofler, Andreas Huber, Wolfgang Fellin, Martin Mergili, and Michael Oberguggenberger. Bayesian inference in snow avalanche simulation with r.avaflow. *Geosciences*, 10(5), 2020. ISSN 2076-3263. doi: 10.3390/geosciences10050191. URL https://www.mdpi.com/2076-3263/10/5/191. (Cited on page 7.)

Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922. doi: https://doi.org/10.1098/rsta.1922.0009. (Cited on page 28.)

J. Fruth, O. Roustant, and S. Kuhnt. Sequential designs for sensitivity analysis of functional inputs in computer experiments. *Reliability Engineering & System Safety*, 134:260 – 267, 2015. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2014.07.018. URL http://www.sciencedirect.com/science/article/pii/S0951832014001732. (Cited on page 105.)

F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *ArXiv e-prints*, November 2013. (Cited on pages 21, 76, 77, 78, 84, and 87.)

Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agnès Lagnoux. Sensitivity indices for multivariate outputs. *Comptes Rendus Mathematique*, 351(7):307 – 310, 2013. ISSN 1631-073X. doi: https://doi.org/10.1016/j.crma.2013.04.016. URL http://www.sciencedirect.com/science/article/pii/S1631073X13000824. (Cited on pages 20, 49, 52, 60, and 75.)

Fabrice Gamboa, Pierre Gremaud, Thierry Klein, and Agnès Lagnoux. Global sensitivity analysis: a new generation of mighty estimators based on rank statistics. *arXiv preprint arXiv:2003.01772*, 2020. (Cited on page 14.)

Peter Gauer. Comparison of avalanche front velocity measurements and implications for avalanche models. *Cold Regions Science and Technology*, 97:132 – 150, 2014. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2013.09.010. URL http://www.sciencedirect.com/science/article/pii/S0165232X1300147X. (Cited on pages 26, 38, and 92.)

Peter Gauer, Zenon Medina-Cetina, Karstein Lied, and Krister Kristensen. Optimization and probabilistic calibration of avalanche block models. *Cold Regions Science and Technology*, 59(2):251 – 258, 2009. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2009.02.002. URL http://www.sciencedirect.com/science/article/pii/S0165232X09000196. International Snow Science Workshop (ISSW) 2008. (Cited on pages 7 and 25.)

Johan Gaume, T Gast, J Teran, A van Herwijnen, and C Jiang. Dynamic anti-crack propagation in snow. *Nature communications*, 9:3047, 2018. doi: 10.1038/s41467-018-05181-w. (Cited on page 25.)

Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472, 1992. ISSN 08834237. URL http://www.jstor.org/stable/2246093. (Cited on pages 29 and 34.)

Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995. doi: https://doi.org/10.1201/b14835. (Cited on page 34.)

Christian Gouriéroux and Alain Monfort. *Séries temporelles et modèles dynamiques*. Economica, 1995. (Cited on page 43.)

Ralf Greve, Thilo Koch, and Kolumban Hutter. Unconfined flow of granular avalanches along a partly curved surface. i. theory. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 445(1924):399–413, 1994. (Cited on page 88.)

Urs Gruber and Stefan Margreth. Winter 1999: a valuable test of the avalanche-hazard mapping procedure in Switzerland. *Annals of Glaciology*, 32:328–332, 2001. doi: 10.3189/172756401781819238. (Cited on page 49.)

Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. ISSN 13507265. URL http://www.jstor.org/stable/3318737. (Cited on page 104.)

Carl Harbitz, Dieter Issler, and Chris Keylock. Conclusions from a recent survey of avalanche computational models. pages 128–135, 05 1998. (Cited on page 6.)

Joseph Hart and Pierre A. Gremaud. An approximation theoretic perspective of sobol' indices with dependent variables. *International Journal for Uncertainty Quantification*, 8(6):483–493, 2018. ISSN 2152-5080. doi: 10.1615/Int.J.UncertaintyQuantification.2018026498. (Cited on page 15.)

W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970. ISSN 00063444. URL http://www.jstor.org/stable/2334940. (Cited on pages 9 and 29.)

Tristen Hayfield and Jeffrey S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 2008. URL http://www.jstatsoft.org/v27/i05/. (Cited on page 56.)

J.C. Helton, J.D. Johnson, C.J. Sallaberry, and C.B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10):1175 – 1209, 2006. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2005.11.017. URL http://www.sciencedirect.com/science/article/pii/S0951832005002292. The Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004). (Cited on page 20.)

María Belén Heredia, Nicolas Eckert, Clémentine Prieur, and Emmanuel Thibert. Bayesian calibration of an avalanche model from autocorrelated measurements along the flow: application to velocities extracted from photogrammetric images. *Journal of Glaciology*, 66(257):373–385, 2020. doi: 10.1017/jog.2020.11. (Cited on pages 7 and 64.)

Wassily Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948. ISSN 00034851. URL http://www.jstor.org/stable/2235637. (Cited on pages 13 and 49.)

Rob J Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996. doi: 10.2307/2684423. (Cited on page 91.)

Rob J. Hyndman and Han Lin Shang. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45, 2010. doi: 10.1198/jcgs.2009.08158. URL https://doi.org/10.1198/jcgs.2009.08158. (Cited on pages 20, 65, and 91.)

B. Iooss and C. Prieur. Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications. *ArXiv e-prints*, July 2017. (Cited on page 49.)

Bertrand Iooss and Paul Lemaître. *A Review on Global Sensitivity Analysis Methods*, pages 101–122. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7547-8. doi: 10.1007/978-1-4899-7547-8_5. URL https://doi.org/10.1007/978-1-4899-7547-8_5. (Cited on pages 10, 11, 14, 20, 49, 50, and 74.)

Bertrand Iooss and Clementine Prieur. Shapley effects for sensitivity analysis with correlated inputs: Comparisons with Sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification*, 9(5):493–514, 2019. ISSN 2152-5080. doi: 10.1615/Int.J.UncertaintyQuantification.2019028372. (Cited on pages 16, 17, 75, and 84.)

Bertrand Iooss and Mathieu Ribatet. Global sensitivity analysis of computer models with functional inputs. *Reliability Engineering & System Safety*, 94(7):1194 – 1204, 2009. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2008.09.010. URL http://www.sciencedirect.com/science/article/pii/S0951832008002299. Special Issue on Sensitivity Analysis. (Cited on page 105.)

Bertrand Iooss and Andrea Saltelli. Introduction to sensitivity analysis. *Handbook of uncertainty quantification*, pages 1103–1122, 2017. (Cited on page 10.)

T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. In *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403, Dec 1990. doi: 10.1109/ISUMA.1990.151285. (Cited on page 109.)

Julien Jacques, Christian Lavergne, and Nicolas Devictor. Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System Safety*, 91(10):1126 – 1134, 2006. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2005.11.047. URL http://www.sciencedirect.com/science/article/pii/S0951832005002231. The Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004). (Cited on page 15.)

Bruce Jamieson, Stefan Margreth, and Alan Jones. Application and limitations of dynamic models for snow avalanche hazard mapping. *Proceedings of the ISSW 2008. Wistler, Canada*, 01 2008. (Cited on pages 6, 8, 49, and 51.)

Alexandre Janon, Thierry Klein, Agnès Lagnoux, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 2014a. doi: 10.1051/ps/2013040. (Cited on page 14.)

Alexandre Janon, Maëlle Nodet, and Clémentine Prieur. Uncertainties assessment in global sensitivity indices estimation from metamodels. *International Journal for Uncertainty Quantification*, 4(1):21–36, 2014b. doi: 10.1615/Int.J.UncertaintyQuantification.2012004291. (Cited on page 79.)

Robert E. Kass and Adrian E. Raftery. Bayes factors and model uncertainty. Technical report, University of Washington, Department of Statistics, 03 1993. (Cited on page 44.)

Robert E. Kass and Adrian E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995. ISSN 01621459. URL http://www.jstor.org/stable/2291091. (Cited on pages 29, 30, 39, and 41.)

Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001. doi: 10.1111/1467-9868.00294. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294. (Cited on page 8.)

Martin Kern, Perry Bartelt, Betty Sovilla, and Othmar Buser. Measured shear rates in large dry and wet snow avalanches. *Journal of Glaciology*, 55(190):327–338, 2009. doi: 10.3189/002214309788608714. (Cited on page 25.)

Christopher J. Keylock, David M. McClung, and Magnús Már Magnússon. Avalanche risk mapping by simulation. *Journal of Glaciology*, 45(150):303–314, 1999. doi: 10.3189/S0022143000001805. (Cited on page 25.)

A. Köhler, J. N. McElwaine, B. Sovilla, M. Ash, and P. Brennan. The dynamics of surges in the 3 february 2015 avalanches in vallée de la sionne. *Journal of Geophysical Research:*

Earth Surface, 121(11):2192–2210, 2016. doi: 10.1002/2016JF003887. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JF003887. (Cited on page 25.)

S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937 – 946, 2012. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2011.12.020. URL http://www.sciencedirect.com/science/article/pii/S0010465511004085. (Cited on pages 15, 17, and 18.)

S. Kucherenko, O.V. Klymenko, and N. Shah. Sobol' indices for problems defined in non-rectangular domains. *Reliability Engineering & System Safety*, 167:218 – 231, 2017. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2017.06.001. URL http://www.sciencedirect.com/science/article/pii/S0951832016301363. Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis. (Cited on page 74.)

G. Kuczera and E. Parent. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*, 211:69–85, November 1998. doi: 10.1016/S0022-1694(98)00198-X. (Cited on pages 3, 8, 9, 26, 34, and 44.)

George Kuczera. Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. *Water Resources Research*, 19(5):1151–1162, 1983. ISSN 1944-7973. doi: 10.1029/WR019i005p01151. URL http://dx.doi.org/10.1029/WR019i005p01151. (Cited on pages 3, 9, and 26.)

M. Lamboni, B. Iooss, A.-L. Popelin, and F. Gamboa. Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests. *Mathematics and Computers in Simulation*, 87:45 – 54, 2013. ISSN 0378-4754. doi: https://doi.org/10.1016/j.matcom.2013.02.002. URL http://www.sciencedirect.com/science/article/pii/S0378475413000141. (Cited on page 19.)

Matieyendou Lamboni, David Makowski, Simon Lehuger, Benoit Gabrielle, and Hervé Monod. Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, 113(3):312 – 320, 2009. ISSN 0378-4290. doi: https://doi.org/10.1016/j.fcr.2009.06.007. URL http://www.sciencedirect.com/science/article/pii/S0378429009001531. (Cited on pages 20, 21, 22, 49, 50, 52, 75, 76, 77, 78, and 81.)

Matieyendou Lamboni, Hervé Monod, and David Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4):450 – 459, 2011. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2010.12.002. URL http://www.sciencedirect.com/science/article/pii/S0951832010002504. (Cited on pages 22 and 74.)

L. Le Gratiet, S. Marelli, and B. Sudret. Metamodel-based sensitivity analysis: Polynomial chaos expansions and Gaussian processes. *arXiv e-prints*, art. arXiv:1606.04273, June 2016. (Cited on page 51.)

Genyuan Li, Herschel Rabitz, Paul E. Yelvington, Oluwayemisi O. Oluwole, Fred Bacon, Charles E. Kolb, and Jacqueline Schoendorf. Global sensitivity analysis for systems with independent and/or correlated inputs. *The Journal of Physical Chemistry A*, 114(19):6022–6032, 2010. doi: 10.1021/jp9096919. URL https://doi.org/10.1021/jp9096919. PMID: 20420436. (Cited on page 15.)

N. Locantore, J. S. Marron, D. G. Simpson, N. Tripoli, J. T. Zhang, K. L. Cohen, Graciela Boente, Ricardo Fraiman, Babette Brumback, Christophe Croux, Jianqing Fan, Alois Kneip, John I. Marden, Daniel Peña, Javier Prieto, Jim O. Ramsay, Mariano J. Valderrama, Ana M. Aguilera, N. Locantore, J. S. Marron, D. G. Simpson, N. Tripoli, J. T. Zhang, and K. L. Cohen. Robust principal component analysis for functional data. *Test*, 8(1):1–73, Jun 1999. ISSN 1863-8260. doi: 10.1007/BF02595862. URL https://doi.org/10.1007/BF02595862. (Cited on page 55.)

Michel Loève. *Probability Theory*. Springer, New York, 3rd edition, 1963. (Cited on page 55.)

Alfredo López-Benito and Ricardo Bolado-Lavín. A case study on global sensitivity analysis with dependent inputs: The natural gas transmission model. *Reliability Engineering & System Safety*, 165:11 – 21, 2017. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2017.03.019. URL http://www.sciencedirect.com/science/article/pii/S0951832017303447. (Cited on page 74.)

Thierry A. Mara and Stefano Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115 – 121, 2012. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2011.08.008. URL http://www.sciencedirect.com/science/article/pii/S0951832011001724. SAMO 2010. (Cited on page 15.)

Thierry A. Mara, Stefano Tarantola, and Paola Annoni. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental Modelling & Software*, 72:173 – 183, 2015. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2015.07.010. URL http://www.sciencedirect.com/science/article/pii/S1364815215300153. (Cited on page 15.)

Thierry Alex Mara and Onimihamina Rakoto Joseph. Comparison of some efficient methods to evaluate the main effect of computer model factors. *Journal of Statistical Computation and Simulation*, 78(2):167–178, 2008. doi: 10.1080/10629360600964454. URL https://doi.org/10.1080/10629360600964454. (Cited on page 108.)

Amandine Marrel and Vincent Chabridon. Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor. working paper or preprint, April 2020. URL https://hal.archives-ouvertes.fr/hal-02541142. (Cited on pages 19 and 105.)

Amandine Marrel, Bertrand Iooss, Béatrice Laurent, and Olivier Roustant. Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742 – 751, 2009. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2008.07.008. URL http://www.sciencedirect.com/science/article/pii/S0951832008001981. (Cited on pages 20 and 51.)

Amandine Marrel, Bertrand Iooss, Michel Jullien, Béatrice Laurent, and Elena Volkova. Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, 22(3):383–397, 2011. doi: 10.1002/env.1071. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1071. (Cited on page 105.)

Georgios Mavromatidis, Kristina Orehounig, and Jan Carmeliet. Uncertainty and global sensitivity analysis for the optimal design of distributed energy systems. *Applied Energy*, 214:219 – 238, 2018. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2018.01.062. URL http://www.sciencedirect.com/science/article/pii/S0306261918300710. (Cited on page 10.)

David McClung and Peter A Schaerer. *The avalanche handbook*. The Mountaineers Books, 2006. (Cited on page 49.)

DM McClung and Peter Gauer. Maximum frontal speeds, alpha angles and deposit volumes of flowing snow avalanches. *Cold Regions Science and Technology*, 153:78–85, 2018. (Cited on page 92.)

David McInerney, Mark Thyer, Dmitri Kavetski, Julien Lerat, and George Kuczera. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, 53(3):2199–2239, 2017. ISSN 1944-7973. doi: 10.1002/2016WR019168. URL http://dx.doi.org/10.1002/2016WR019168. (Cited on page 26.)

Martin Mergili, Jan-Thomas Fischer, Julia Krenn, and Shiva P. Pudasaini. r.avaflow v1, an advanced open-source computational framework for the propagation and interaction of two-phase mass flows. *Geoscientific Model Development*, 10(2):553–569, February 2017. doi: 10.5194/gmd-10-553-2017. (Cited on pages 7 and 105.)

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. URL https://doi.org/10.1063/1.1699114. (Cited on pages 9 and 29.)

Maurice Meunier, Christophe Ancey, and Jean-Michel Taillandier. Fitting avalanche-dynamics models with documented events from the Col du Lautaret site (France) using the conceptual approach. *Cold Regions Science and Technology*, 39(1):55 – 66, 2004. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2004.03.004. URL http://www.sciencedirect.com/science/article/pii/S0165232X04000242. (Cited on page 25.)

Anouar Meynaoui, Amandine Marrel, and Béatrice Laurent. New statistical methodology for second level global sensitivity analysis. *arXiv e-prints*, art. arXiv:1902.07030, February 2019. (Cited on page 19.)

Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991. ISSN 00401706. URL http://www.jstor.org/stable/1269043. (Cited on page 11.)

Mohamed Naaim, Florence Naaim-Bouvet, Thierry Faug, and Alexi Bouchet. Dense snow avalanche modeling: flow, erosion, deposition and obstacle effects. *Cold Regions Science and Technology*, 39(2):193 – 204, 2004. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2004.07.001. URL http://www.sciencedirect.com/science/article/pii/S0165232X04000643. Snow And Avalanches: Papers Presented At The European Geophysical Union Conference, Nice, April 2003. Dedicated To The Avalanche Dynamics Pioneer Dr. B. Salm. (Cited on pages 5, 6, 20, 26, 51, 62, 74, 88, 103, and 104.)

Mohamed Naaim, Thierry Faug, Florence Naaim, and Nicolas Eckert. Return period calculation and passive structure design at the taconnaz avalanche path, france. *Annals of Glaciology*, 51(54):89–97, 2010. (Cited on page 49.)

Mohamed Naaim, Yves Durand, Nicolas Eckert, and Guillaume Chambon. Dense avalanche friction coefficients: influence of physical properties of snow. *Journal of Glaciology*, 59(216):771–782, 2013. doi: 10.3189/2013JoG12J205. (Cited on pages 1, 7, 25, and 49.)

E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9: 141–142, 1964. (Cited on pages 15, 50, and 52.)

Simon Nanty, Céline Helbert, Amandine Marrel, Nadia Pérot, and Clémentine Prieur. Uncertainty quantification for functional dependent random variables. *Computational Statistics*, 32(2):559–583, Jun 2017. ISSN 1613-9658. doi: 10.1007/s00180-016-0676-0. URL https://doi.org/10.1007/s00180-016-0676-0. (Cited on pages 50 and 55.)

Jeremy Oakley and Anthony O'Hagan. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4):769–784, 2002. doi: 10.1093/biomet/89.4.769. URL http://dx.doi.org/10.1093/biomet/89.4.769. (Cited on page 25.)

Kenichi Oda, Katsuya Nakamura, Yoshikazu Kobayashi, and Jun-ichi Suzumura. Inverse simulation for extracting the flow characteristics of artificial snow avalanches based on computation fluid dynamics. *Geosciences*, 10(6), 2020. ISSN 2076-3263. doi: 10.3390/geosciences10060221. URL https://www.mdpi.com/2076-3263/10/6/221. (Cited on page 6.)

A. Owen. Sobol' Indices and Shapley Value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251, 2014. doi: 10.1137/130936233. URL https://doi.org/10.1137/130936233. (Cited on pages 13, 16, 51, 74, 75, and 76.)

Art B. Owen and Clémentine Prieur. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1), March 2017. URL https://doi.org/10.1137/16M1097717. (Cited on pages 16, 75, and 98.)

Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11): 559–572, 1901. (Cited on page 55.)

R. Perla, T. T. Cheng, and D.M. McClung. A Two-Parameter Model of Snow-Avalanche Motion. *Journal of Glaciology*, 26(94):197–207, 1980. doi: 10.3189/S002214300001073X. (Cited on page 5.)

Elmar Plischke. An effective algorithm for computing global sensitivity indices (EASI). *Reliability Engineering & System Safety*, 95(4):354 – 360, 2010. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2009.11.005. URL http://www.sciencedirect.com/science/article/pii/S0951832009002579. (Cited on pages 14, 50, 108, and 109.)

Elmar Plischke, Emanuele Borgonovo, and Curtis L. Smith. Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550, 2013. doi: 10.1016/j.ejor.2012.11.06. (Cited on pages 14, 15, 50, and 107.)

Elmar Plischke, Giovanni Rabitti, and Emanuele Borgonovo. Computing Shapley Effects for Sensitivity Analysis. *arXiv e-prints*, art. arXiv:2002.12024, February 2020. (Cited on pages 17 and 75.)

Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11, 2006. URL https://journal.r-project.org/archive/. (Cited on page 34.)

C. Prieur, L. Viry, E. Blayo, and J.-M Brankart. A global sensitivity analysis approach for marine biogeochemical modeling. *Ocean Modelling*, 139:101402, 2019. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2019.101402. URL http://www.sciencedirect.com/science/article/pii/S1463500318303688. (Cited on page 10.)

Clémentine Prieur and Stefano Tarantola. *Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms*, pages 1217–1239. Springer International Publishing, Cham, Switzerland, 2017. ISBN 978-3-319-11259-6. doi: 10.1007/978-3-319-11259-6_35-1. URL https://doi.org/10.1007/978-3-319-11259-6_35-1. (Cited on pages 14 and 108.)

Alexander Prokop. Assessing the applicability of terrestrial laser scanning for spatial snow depth measurements. *Cold Regions Science and Technology*, 54(3):155 – 163, 2008. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2008.07.002. URL http://www.sciencedirect.com/science/article/pii/S0165232X08001018. Snow avalanche formation and dynamics. (Cited on page 31.)

Alexander Prokop, Peter Schön, Florian Singer, Gaëtan Pulfer, Mohamed Naaim, Emmanuel Thibert, and Alvaro Soruco. Merging terrestrial laser scanning technology with photogrammetric and total station data for the determination of avalanche modeling parameters. *Cold Regions Science and Technology*, 110:223 – 230, 2015. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2014.11.009. URL http://www.sciencedirect.com/science/article/pii/S0165232X14002110. (Cited on pages 25 and 31.)

G. Pulfer, M. Naaim, E. Thibert, and A. Soruco. Retrieving avalanche basal friction law from high rate positioning of avalanches. In *International Snow Science Workshop (ISSW)*, pages p. 1418 – p. 1424, Grenoble – Chamonix Mont-Blanc, France, October 2013. Irstea, ANENA, Meteo France. URL https://hal.archives-ouvertes.fr/hal-00951328. (Cited on pages 31 and 64.)

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/. (Cited on pages 34 and 56.)

Jeff Racine. Bias-Corrected Kernel Regression. *Journal of Quantitative Economics*, 17 (1):25–42, 2001. (Cited on pages 3, 15, 50, 53, 55, 68, 69, 71, and 103.)

Majdi I. Radaideh, Stuti Surani, Daniel O'Grady, and Tomasz Kozlowski. Shapley effect application for variance-based sensitivity analysis of the few-group cross-sections. *Annals of Nuclear Energy*, 129:264 – 279, 2019. ISSN 0306-4549. doi: https://doi.org/10.1016/j.anucene.2019.02.002. URL http://www.sciencedirect.com/science/article/pii/S0306454919300714. (Cited on page 74.)

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition, June 2005. ISBN 038740080X. URL http://www.worldcat.org/isbn/038740080X. (Cited on pages 22, 50, and 55.)

John A. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243, sep 1991. doi: 10.1111/j.2517-6161.1991.tb01821.x. (Cited on page 82.)

Christian P. Robert. The Metropolis-Hastings algorithm. *ArXiv e-prints*, April 2015. (Cited on pages 9, 29, and 34.)

Christian P. Robert and George Casella. *Introducing Monte Carlo Methods with R (Use R)*. Springer-Verlag, Berlin, Heidelberg, 1st edition, 2009. ISBN 1441915753, 9781441915757. (Cited on pages 34 and 35.)

Pierre G. Rognon, François Chevoir, Hervé Bellot, Frédéric Ousset, Mohamed Naaïm, and Philippe Coussot. Rheology of dense snow flows: Inferences from steady state chute-flow experiments. *Journal of Rheology*, 52(3):729–748, 2008. doi: 10.1122/1.2897609. URL https://doi.org/10.1122/1.2897609. (Cited on page 25.)

Bruno Salm. A short and personal history of snow avalanche dynamics. *Cold Regions Science and Technology*, 39(2):83 – 92, 2004. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2004.06.004. URL http://www.sciencedirect.com/science/article/pii/S0165232X04000588. Snow And Avalanches: Papers Presented At The European Geophysical Union Conference, Nice, April 2003. Dedicated To The Avalanche Dynamics Pioneer Dr. B. Salm. (Cited on pages 5, 6, 8, 25, and 26.)

Bruno Salm, A. Burkard, and H. Gubler. Berenchnung von Fliesslanwinen, eine Anleitung für Paktiker mit Beispielen. Technical Report 47, Eidgenössches Institut für Schnee und Lawinenforschung, Davos, 1990. Technical report. (Cited on pages 5, 6, 34, and 92.)

A. Saltelli, S. Tarantola, and K. P.-S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999. doi: 10.1080/00401706.1999.10485594. URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1999.10485594. (Cited on pages 14 and 50.)

A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity Analysis: Gauging the Worth of Scientific Models*. Wiley Series in Probability and Statistics. Wiley, 2000. ISBN 9780471998921. (Cited on page 49.)

Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Halsted Press, New York, NY, USA, 2004. ISBN 0470870931. (Cited on pages 9, 10, and 49.)

Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis. The primer*. John Wiley, 2008. ISBN 9780470059975. (Cited on pages 10 and 49.)

Andrea Saltelli, Ksenia Aleksankina, William Becker, Pamela Fennell, Federico Ferretti, Niels Holst, Sushan Li, and Qiongli Wu. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, 114:29 – 39, 2019. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2019.01.012. URL http://www.sciencedirect.com/science/article/pii/S1364815218302822. (Cited on page 10.)

Fanny Sarrazin, Francesca Pianosi, and Thorsten Wagener. Global sensitivity analysis of environmental models: Convergence and validation. *Environmental Modelling & Software*, 79:135 – 152, 2016. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2016.02.005. URL http://www.sciencedirect.com/science/article/pii/S1364815216300251. (Cited on page 10.)

Bettina Schaefli and Dmitri Kavetski. Bayesian spectral likelihood for hydrological parameter inference. *Water Resources Research*, 53(8):6857–6884, 2017. ISSN 1944-7973. doi: 10.1002/2016WR019465. URL http://dx.doi.org/10.1002/2016WR019465. (Cited on page 26.)

Mark Schaer and Dieter Issler. Particle densities, velocities and size distributions in large avalanches from impact-sensor measurements. *Annals of Glaciology*, 32:321–327, 2001. doi: 10.3189/172756401781819409. (Cited on page 1.)

Romain Schläppy, Nicolas Eckert, Vincent Jomelli, Markus Stoffel, Delphine Grancher, Daniel Brunstein, Mohamed Naaim, and Michaël Deschatres. Validation of extreme snow avalanches and related return periods derived from a statistical-dynamical model using tree-ring techniques. *Cold Regions Science and Technology*, 99:12 – 26, 2014. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2013.12.001. URL http://www.sciencedirect.com/science/article/pii/S0165232X13001900. (Cited on pages 7, 8, and 25.)

Jürg Schweizer, J. Bruce Jamieson, and Martin Schneebeli. Snow avalanche formation. *Reviews of Geophysics*, 41(4), 2003. doi: 10.1029/2002RG000123. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002RG000123. (Cited on page 25.)

L. Shapley. A Value for n-Person Games. *Contributions to the Theory of Games (AM-28), Princeton: Princeton University Press.*, 2, 1953. ISSN 307-318. (Cited on pages 2, 13, 16, 76, and 77.)

I. M. Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1(4):407–414, 1993. (Cited on pages 2, 13, 14, 49, 50, 52, 74, and 75.)

I.M. Sobol' and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009 – 3017, 2009. ISSN 0378-4754. doi: https://doi.org/10.1016/j.matcom.2009.01.023. URL http://www.sciencedirect.com/science/article/pii/S0378475409000354. (Cited on page 19.)

Maikol Solís. Non-parametric estimation of the first-order Sobol indices with bootstrap bandwidth. *Communications in Statistics - Simulation and Computation*, pages 1–16, 2019. doi: 10.1080/03610918.2019.1655575. URL https://doi.org/10.1080/03610918.2019.1655575. (Cited on pages 3, 14, 50, 53, 56, 57, 68, and 103.)

E. Song, B. Nelson, and J. Staum. Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083, 2016. doi: 10.1137/15M1048070. URL https://doi.org/10.1137/15M1048070. (Cited on pages 17, 75, 79, 80, and 98.)

Alvaro Soruco, Emmanuel Thibert, Christian Vincent, Renaud Blanc, and Raphële Héno. Measurement of avalanche front velocity from high-speed terrestrial digital photogrammetry. In *Geophysical Research Abstracts*, volume 13, 2011. (Cited on page 30.)

Betty Sovilla, M. Schaer, M. Kern, and P. Bartelt. Impact pressures and flow regimes in dense snow avalanches observed at the vallée de la sionne test site. *Journal of Geophysical Research: Earth Surface*, 113(F1), 2008. doi: 10.1029/2006JF000688. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JF000688. (Cited on page 25.)

Adrien Spagnol, Rodolphe Le Riche, and Sébastien Da Veiga. Global sensitivity analysis for optimization with variable selection. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2):417–443, 2019. doi: 10.1137/18M1167978. URL https://doi.org/10.1137/18M1167978. (Cited on page 19.)

Daniel Sparkman, Harry R. Millwater, Jose Garza, and Benjamin P. Smarslok. *Importance Sampling-based Post-Processing Method for Global Sensitivity Analysis*. 2016. doi: 10.2514/6.2016-1440. (Cited on page 50.)

Walter Steinkogler, Betty Sovilla, and Michael Lehning. Influence of snow cover properties on avalanche dynamics. *Cold Regions Science and Technology*, 97:121 – 131, 2014. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2013.10.002. URL http://www.sciencedirect.com/science/article/pii/S0165232X13001535. (Cited on page 25.)

Walter Steinkogler, Johan Gaume, Henning Löwe, Betty Sovilla, and Michael Lehning. Granulation of snow: From tumbler experiments to discrete element simulations. *Journal of Geophysical Research: Earth Surface*, 120(6):1107–1126, 2015. doi: 10.1002/2014JF003294. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JF003294. (Cited on page 25.)

Charles J. Stone. The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics*, 22(1):118–171, 1994. ISSN 00905364. URL http://www.jstor.org/stable/2242446. (Cited on page 15.)

Daniel Straub and Adrienne Grêt-Regamey. A Bayesian probabilistic framework for avalanche modelling based on observations. *Cold Regions Science and Technology*, 46(3):192 – 203, 2006. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldrccegions.2006.08.024. URL http://www.sciencedirect.com/science/article/pii/S0165232X06001297. (Cited on pages 7 and 25.)

Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964 – 979, 2008. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2007.04.002. Bayesian Networks in Dependability. (Cited on page 51.)

Ruochen Sun, Huiling Yuan, and Xiaoli Liu. Effect of heteroscedasticity treatment in residual error models on model calibration and prediction uncertainty estimation. *Journal of Hydrology*, 554:680 – 692, 2017. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2017.09.041. URL http://www.sciencedirect.com/science/article/pii/S0022169417306455. (Cited on page 26.)

Kunio Takezawa. *Introduction to nonparametric regression*, volume 606. John Wiley & Sons, 2005. (Cited on page 53.)

T. Tang, P. Reed, T. Wagener, and K. van Werkhoven. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences Discussions*, 3(6):3333–3395, November 2006. (Cited on page 10.)

S. Tarantola, D. Gatelli, and T.A. Mara. Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety*, 91(6):717 – 727, 2006. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2005.06.003. URL http://www.sciencedirect.com/science/article/pii/S0951832005001444. (Cited on pages 14 and 50.)

F. Techel, F. Jarry, G. Kronthaler, S. Mitterer, P. Nairz, M. Pavšek, M. Valt, and G. Darms. Avalanche fatalities in the european alps: long-term trends and statistics. *Geographica Helvetica*, 71(2):147–159, 2016. doi: 10.5194/gh-71-147-2016. URL https://gh.copernicus.org/articles/71/147/2016/. (Cited on page 1.)

Emmanuel Thibert, Hervé Bellot, Xavier Ravanat, Frédéric Ousset, Gaëtan Pulfer, Mohamed Naaim, Pascal Hagenmuller, Florence Naaim-Bouvet, Thierry Faug, Koichi Nishimura, Yoichi Ito, Djebar Baroudi, Alexander Prokop, Peter Schon, Alvaro Soruco, Christian Vincent, Ali Limam, and Raphaele Héno. The full-scale avalanche test-site at Lautaret Pass (French Alps). *Cold Regions Science and Technology*, 115:30 – 41, 2015. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2015.03.005. URL http://www.sciencedirect.com/science/article/pii/S0165232X15000622. (Cited on pages 1, 26, 30, 31, and 51.)

Jean-Yves Tissot and Clémentine Prieur. A randomized Orthogonal Array-based procedure for the estimation of first- and second-order Sobol' indices. *Journal of Statistical Computation and Simulation*, 85(7):1358–1381, 2015. doi: 10.1080/00949655.2014.971799. URL https://hal.archives-ouvertes.fr/hal-00743964. (Cited on pages 14, 15, 62, and 108.)

Jean-Yves Tissot and Clémentine Prieur. Bias correction for the estimation of sensitivity indices based on random balance designs. *Reliability Engineering & System Safety*, 107:205 – 213, 2012. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2012.06.010. SAMO 2010. (Cited on pages 56 and 57.)

Franck Torre, Jean-Jacque Boreux, and Eric Parent. The Metropolis-Hastings algorithm, a handy tool for the practice of environmental model estimation: illustration with biochemical oxygen demand data. *Cybergeo : European Journal of Geography [En ligne]*, (187):9447–9463, 2001. doi: 0.4000/cybergeo.4750. URL http://cybergeo.revues.org/4750. (Cited on page 34.)

H.K. Truong, C.J. Keylock, N. Eckert, H. Bellot, and M. Naaïm. Refining the processing of paired time series data to improve velocity estimation in snow flows. *Cold Regions Science and Technology*, 151:75 – 88, 2018. ISSN 0165-232X. doi: https://doi.org/10.1016/j.coldregions.2018.03.004. URL http://www.sciencedirect.com/science/article/pii/S0165232X17306304. (Cited on page 25.)

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st edition, 2008. ISBN 0387790519, 9780387790510. (Cited on page 53.)

B. Turnbull and P. Bartelt. Mass and Momentum Balance Model of a Mixed Flowing/Powder Snow Avalanche. *Surveys in Geophysics*, 24(5):465–477, Nov 2003. ISSN 1573-0956. doi: 10.1023/B:GEOP.0000006077.82404.84. URL https://doi.org/10.1023/B:GEOP.0000006077.82404.84. (Cited on page 26.)

Roberto Viviani, Georg Grön, and Manfred Spitzer. Functional principal component analysis of fMRI data. *Human Brain Mapping*, 24(2):109–129, 2005. doi: 10.1002/hbm.20074. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.20074. (Cited on page 55.)

A Voellmy. On the destructive force of avalanches (translation by R. E. Tate). *Alta Avalanche Study Center Wasatch National Forest*, 1964. (Cited on pages 5, 25, and 26.)

N. M. Vriend, J. N. McElwaine, B. Sovilla, C. J. Keylock, M. Ash, and P. V. Brennan. High-resolution radar measurements of snow avalanches. *Geophysical Research Letters*, 40(4):727–731, 2013. doi: 10.1002/grl.50134. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50134. (Cited on page 25.)

Geoffrey S. Watson. Smooth regression analysis. *Sankhyā Ser.*, 26:359–372, 1964. (Cited on pages 15, 50, and 52.)

Philippe Wiederkehr. Global sensitivity analysis with dependent inputs, 2018. (Cited on page 16.)

Chonggang Xu and George Zdzislaw Gertner. Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, 93(10): 1563 – 1573, 2008. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2007.06.003. URL http://www.sciencedirect.com/science/article/pii/S0951832007001652. (Cited on page 15.)

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470): 577–590, 2005. doi: 10.1198/016214504000001745. URL https://doi.org/10.1198/016214504000001745. (Cited on pages 22, 55, 75, 76, and 81.)

Kaichao Zhang, Zhenzhou Lu, Lei Cheng, and Fang Xu. A new framework of variance based global sensitivity analysis for models with correlated inputs. *Structural Safety*, 55:1 – 9, 2015. ISSN 0167-4730. doi: https://doi.org/10.1016/j.strusafe.2014.12.005. URL http://www.sciencedirect.com/science/article/pii/S0167473015000181. (Cited on page 15.)