

# Journal Pre-proof

Global optimization for data assimilation in landslide tsunami models

A.M. Ferreiro-Ferreiro, J.A. García-Rodríguez, J.G. López-Salas, C. Escalante, M.J. Castro

PII: S0021-9991(19)30774-0  
DOI: <https://doi.org/10.1016/j.jcp.2019.109069>  
Reference: YJCPH 109069

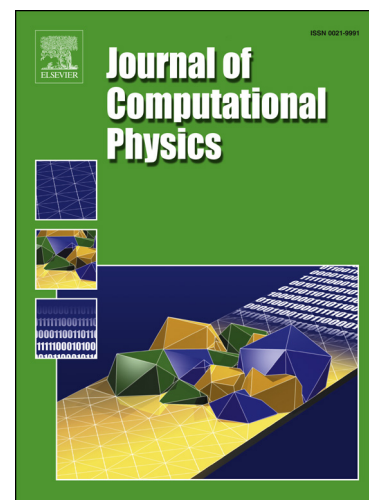
To appear in: *Journal of Computational Physics*

Received date: 25 April 2019  
Revised date: 24 October 2019  
Accepted date: 25 October 2019

Please cite this article as: A.M. Ferreiro-Ferreiro et al., Global optimization for data assimilation in landslide tsunami models, *J. Comput. Phys.* (2019), 109069, doi: <https://doi.org/10.1016/j.jcp.2019.109069>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.



## Highlights

- We develop a generic data assimilation framework for landslide tsunami models.
- The data assimilation problem is posed in a global optimization framework.
- Parallel and efficient global optimization algorithms are developed.
- We assess the identifiability of model parameters for a landslide tsunami model.
- The developed machinery is tested with real laboratory data.

# Global optimization for data assimilation in landslide tsunami models

A.M. Ferreiro-Ferreiro<sup>a,b</sup>, J.A. García-Rodríguez<sup>a,b,\*</sup>, J.G. López-Salas<sup>a,b</sup>,  
C. Escalante<sup>c</sup>, M.J. Castro<sup>c</sup>

<sup>a</sup>*Department of Mathematics, Faculty of Informatics, Campus Elviña s/n, 15071-A Coruña, Spain. Emails: ana.fferreiro@udc.es, jose.garcia.rodriguez@udc.es, jose.lsalas@udc.es*

<sup>b</sup>*CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain*

<sup>c</sup>*Department of Análisis Matemático, University of Málaga. Emails: escalante@uma.es, mjcastro@uma.es*

## Abstract

The goal of this article is to make automatic data assimilation for a landslide tsunami model, given by the coupling between a non-hydrostatic multi-layer shallow-water and a Savage-Hutter granular landslide model for submarine avalanches. The coupled model is discretized using a positivity preserving second-order path-conservative finite volume scheme. Then, the data assimilation problem is posed in a global optimization framework. Later, multi-path parallel metaheuristic stochastic global optimization algorithms are developed. More precisely, a multi-path Simulated Annealing algorithm is compared with a multi-path hybrid global optimization algorithm based on coupling Simulated Annealing with gradient local searchers.

**Keywords:** Tsunamis, Submarine avalanches, Finite volume methods, Data assimilation, Global optimization, Parallel computing

## 1. Introduction

The goal of this work is twofold. On the one hand, assessing the feasibility of performing data assimilation for models of tsunamis generated by submarine landslides, when only information of the fluid free surface is available. In other words, we aim at checking whether the data assimilation problem is well posed. This question is also known as the identifiability of model parameters. On the other hand, if the former is possible, we also aim at developing a generic data assimilation framework based on parallel and efficient global optimization algorithms which can deal with landslide tsunami models.

The tsunami hazard modeling is of great importance to prevent and forecast the consequences of such events, as they can cause a large number of casualties and huge financial losses. Tsunamis can be generated mainly by earthquakes, storm surges or landslides (subaerial or submarine). The majority of them are

\*Corresponding author  
Preprint submitted to Elsevier  
Email address: jose.garcia.rodriguez@udc.es (J.A. García-Rodríguez) October 29, 2019

caused by an offshore earthquake that pushes the ocean up or down. Nevertheless tsunamis can also be generated in other ways. Underwater landslides, also known as submarine mass failures (SMF), which might accompany an earthquake or occur independently, are a classic example. Traditional warning systems completely miss tsunamis from those types of sources. Once we have a model for these phenomena, the correct calibration of its parameters is of key importance for the accurate simulation of the tsunami. The calibration could be even done in real time just after the tsunami occurrence, by feeding the calibration machinery with the measures taken by the tide-gauges in the ocean. After this calibration stage, the model can be used in order to predict the trajectory of the tsunami and the impact areas.

Several types of models can be found in the literature for modeling landslide tsunamis. Their development focuses in three aspects: a physical model for landslide material, a hydrodynamic model that simulates the generation and propagation of resulting waves, and the coupling between both. The hydrodynamics of landslide-induced tsunamis has been extensively studied using numerical models based on different levels of simplification.

The simplest model contemplates the landslide as a rigid solid with fixed landslide shape [1]. Another approach to simulate landslide-induced tsunamis is to consider both the landslide and the water as two different fluids [2, 3, 4, 5, 6, 7]. This approach allows the landslide to deform, and to couple the landslide and the fluid. Although the two-fluid models described above can be reasonably successful in predicting tsunami wave generation, they may fail to determine the landslide motion from initiation to deposition.

Initial steps towards development of granular flow-based models for landslide behavior have usually been based on depth-integrated models pioneered by Iverson [8], Savage and Hutter [9], and others. These models were initially developed for application to shallow subaerial debris flows. In [10] a two-layer Savage-Hutter type model was proposed to simulate submarine landslides, where the hydrostatic pressure assumption is assumed to derive the model.

In [11] a two-phase model for granular landslide motion and tsunami wave generation is developed. On the one hand, the granular phase is modeled by a standard Savage-Hutter type model governed by Coulomb friction. On the other hand, the tsunami wave generation is simulated using a three-dimensional non-hydrostatic wave model. The latter is capable of capturing wave dispersion efficiently using a small number of discretized vertical layers.

In this article we follow a similar approach. A two-phase model is also considered, however the three-dimensional non-hydrostatic model is replaced by the multi-layer non-hydrostatic model recently proposed in [12]. This framework is briefly discussed in Section 2.

The previous model depends on a set of parameters that need to be calibrated in order to match real data. Note that, having a good model and a strong and reliable numerical method for solving the problem, is as important as performing a good parameters adjustment of the model according to physical measures. In other words, a good model together with a good numerical method, can lead to totally wrong results with poorly calibrated parameters. Data assimilation is the

74 tool for embedding reality in numerical simulation. Together with mathematical  
 75 modeling and development of proper numerical methods, it could be considered  
 76 as the third leg supporting the numerical simulation of processes in science and  
 77 engineering. Data assimilation allows the model to learn and profit from real  
 78 measured data. It is of key importance, among others, in atmospheric models for  
 79 weather forecasting [13] and models for geophysical fluids [14]. The pioneering  
 80 work about the mathematical basis of data assimilation and control was carried  
 81 out by Lions in [15].

82 Our work follows the classical approach to calibrate the parameters of a  
 83 model. They are adjusted in such a way that the behaviour of the model ap-  
 84 proximates, as closely and consistently as possible, the observed response of  
 85 a hydrologic system over some historical period of time. Ultimately, the best  
 86 parameters are those minimizing the simple least square objective function of  
 87 the residuals, which accounts for the differences between the model-simulated  
 88 output and the measured data. This is the right approach as long as the math-  
 89 ematical model is correct (realistic enough), and physical data are measured  
 90 without error. The uncertainty in the model prediction will be due to the un-  
 91 certainty in the parameter estimates.

92 There is a separate line of research [16] arguing that models have structural  
 93 errors arising from the aggregation of spatially distributed real-world processes  
 94 into mathematical models. Besides, due to this aggregation process, model pa-  
 95 rameters usually do not represent directly measurable entities. Therefore, they  
 96 must be estimated using measurements of the system inputs and outputs, thus  
 97 adding another source of errors. As a consequence, during the calibration pro-  
 98 cess one should also take into account input, output and model structural errors.  
 99 Several methods were firstly proposed to deal with model structural and data  
 100 errors, like the Bayesian approach, Recursive Parameter Estimation algorithms,  
 101 multiobjective calibration or stochastic input error models. Bayesian methods  
 102 treat model parameters as probabilistic variables, in contrast with Frequentists  
 103 approaches which consider model parameters fixed but unknown. Examples  
 104 of Bayesian methods in hydrology are the Generalized Likelihood Uncertainty  
 105 Estimation framework of Beven and Binley [17] and the Bayesian Recursive Esti-  
 106 mation approach of Thiemann [18]. Recursive Parameter estimation algorithms  
 107 help to identify model structural flaws by reducing the temporal aggregation  
 108 associated with traditional batch processing, like PIMLI and recursive Shuffled  
 109 Complex Evolution Metropolis algorithms (SCEM-UA) [19, 20]. Multiobjective  
 110 frameworks in order to better understand the limitation of the models, use com-  
 111plementary criteria in the optimization procedure and analyze the trade off in  
 112 the fitting of these criteria; MOCOM [21] and MOSCEM-UA [16] being exam-  
 113ples of these algorithms. Finally, realistic stochastic input error models, like the  
 114 Bayesian Total Error Analysis of Kavetski, only account for input errors.

115 These previously discussed methods were not successful to account for all the  
 116 referred sources of uncertainty in hydrologic modelling, i.e. parameter, input,  
 117 output and structural model errors. Later, sequential data assimilation (SDA)  
 118 techniques continuously update the parameters of the model when new mea-  
 119 surements are available. This is intended for improving the model forecast and

120 evaluate its accuracy. Kalman and extended Kalman filters represent SDA ap-  
 121 proaches for linear and nonlinear models, respectively. Recently, Vrugt et al. in  
 122 [16] enrich the classical calibration approach with SDA techniques. The authors  
 123 developed the so-called simultaneous parameter optimization and data assim-  
 124 ilation (SODA) method. This strategy combines the search efficiency and the  
 125 explorative capabilities of the Shuffled Complex Evolution Metropolis algorithm  
 126 [20], with the power of the ensemble Kalman filter. Therefore, the blending ac-  
 127 counts for the parameter, input, output and model structural uncertainties in  
 128 hydrologic modeling.

129 Another approach aiming to reduce the uncertainty of models and improve  
 130 their prediction skills consists on identifying the sensitive parameters and then  
 131 focus on reducing the error of these delicate parameters [22]. For example, in  
 132 [23], Yuan Shijin et al. studied the sensitivity of the wind stress, the viscosity  
 133 coefficient and the lateral friction for the simulation of the double-gyre varia-  
 134 tion in the Regional Ocean Modeling System [24]. This model can be used to  
 135 simulate global waters of any size from basins to oceans. Their sensitivity study  
 136 was carried out not only for single parameters, but also for the combination of  
 137 multiple parameters. To this end, the authors solved the Conditional Nonlinear  
 138 Optimal Perturbation related to Parameter (CNOP-P) problem [25] with the  
 139 help of a modified Simulated Annealing (SA) algorithm. These works exploring  
 140 optimal parameters using sensitivity experiments are not feasible for models  
 141 with large number of parameters. Indeed, the number of necessary experiments  
 142 increases exponentially with the involved number of model variables. A study  
 143 of the sensitivities of the parameters for a simplified version of the model we are  
 144 considering in this work was carried out by means of Multi-Level Monte Carlo  
 145 in [26], the fluid model component being hydrostatic with just one fluid layer.

146 In a general setting, the data assimilation problem for a given model, can be  
 147 posed as an unconstrained global optimization problem in a bounded domain.  
 148 Stochastic global metaheuristic algorithms are useful to solve this kind of prob-  
 149 lems. They have the advantage of needing little information of the function,  
 150 and also allow to escape from local optima. Their main disadvantage is the  
 151 slow rate of convergence, which is typical of Monte Carlo algorithms. Classi-  
 152 cal well known examples of these methods are Simulated Annealing [27, 28],  
 153 Particle Swarm (PS) [29, 30] or Differential Evolution (DE) [31]. Conversely,  
 154 local optimization algorithms are deterministic and use more information of  
 155 the function, thus being faster. Their main disadvantages are that, in general,  
 156 they require some regularity of the cost function, and even more important,  
 157 they do not guarantee reaching the global optimum, as they can get trapped  
 158 into a local minimum. They can be gradient free, for example Pattern Search  
 159 (PS) [32] or Nelder-Mead (NM) [33]; or gradient based, like steepest descent,  
 160 Newton method, Conjugate Gradient (CG), Nonlinear CG (NCG) [34, 35] or  
 161 Quasi-Newton methods, for example BFGS [36, 37, 38, 39], L-BFGS [40] or L-  
 162 BFGS-B [41]. One idea to profit from the good properties of stochastic (global)  
 163 and deterministic (local) algorithms, is to hybridize them: this can be done  
 164 by nesting the local search inside the global algorithm. One example is the  
 165 Basin Hopping (BH) algorithm [42, 43, 44]. In this work, in order to calibrate

the tsunami model, we follow this idea, using an in-house developed parallel multi-path version of the BH algorithm.

Data assimilation for shallow-water models has been addressed in many works. Usually gradient based local optimization methods, like the simplest steepest descent method, have been used to solve the resulting optimization problem. Due to the high computational cost, the gradient is computed by solving the adjoint problem, either by solving directly the adjoint system or by computing the adjoint using automatic differentiation (AD) [45, 46]. For example, in [47] the identification of Manning’s roughness coefficients in shallow-water flows is performed. The authors compared three local optimization algorithms: a n-trust region method, L-BFGS and L-BFGS-B minimizers. The gradients are computed by solving the adjoint equations. In [48] the variational data assimilation method (4D-VAR) is presented as a tool to forecast floods, in the case of purely hydrological flows. The cost function is a modification of the shallow-water equations to include a simplified sediment transport model. The steepest descent algorithm is then used to find the minimum. The initial and boundary conditions are calibrated. Besides, the gradient of the cost function is analytically computed by solving the adjoint equations of the model. In [49] the authors developed a 4D-VAR combining remote sensing data (spatially distributed water levels extracted from spatial images, SAR) and a 2D shallow-water model. They identified time-independent parameters (e.g. Manning coefficients and initial conditions) and time-dependent ones (e.g. inflow). In [50] the authors applied the technology developed in [49] in order to derive water levels with precision from satellite images of a real event. In [51] the authors presented a method to use Lagrangian data along with classical Eulerian observations. A variational data assimilation process for a river hydraulics using a 2D shallow-water model was considered. The trajectories of the particles advected by the flow and extracted from video images were employed. In all the cited works AD is applied for computing the gradients, and the data assimilation is performed using gradient local optimization algorithms.

Data assimilation for tsunamis forecasting and early warning is a very challenging problem. On top of that some data are even unknown, for example the geometry of the landslide or the bottom deformation related to earthquake. Real time data is available in Tsunami Early Warning Systems (TEWS). An example is the tide-gauges network of Deep-Ocean Assessment and Reporting of Tsunamis (DART) from National Data Buoy Center of the NOAA, or similar systems from other countries [52]. Tsunami buoys are not only intended to display the occurrence of the tsunami, but also to provide real time data that can be assimilated into the tsunami warning system. Thus, the accuracy of the tsunami forecasting can be improved. Real time data assimilation in tsunami models is mostly done using optimal interpolation (OI) and tsunami Green functions, which are calculated in advance with linear tsunami propagation models [53, 54]. Another alternative assimilation method, is to use Kalman filter techniques [55, 56] for wave field reconstructions and forecasts [57, 58]. In [59] data assimilation is done using a OI algorithm to both the real observations and virtual stations, in order to construct a complete wave front of tsunami



propagation. In [60] tsunami data assimilation of high-density offshore pressure gauges is performed. In [57] a Kalman filter technique is proposed and compared with OI. In [61] the assimilation of Lagrangian data into a primitive equations circulation model of the ocean at basin scale is performed. A four-dimensional variational technique and the adjoint method were used. In [62] retrospectively data assimilation is applied to the tsunami generated in 2011 off the Pacific coast by the Tohoku Earthquake (Mw 9.0). The data assimilation is done using an algorithm of near-field tsunami forecasting with tsunami data recorded at various offshore tsunami stations: these measures were taken between 5 and 10 minutes before the tsunami reached the coastal tide-gauge stations nearest to its origin.

Nevertheless, data assimilation in landslide generated tsunamis is not so well-developed. In this work we propose to use global optimization algorithms, that in general produce better results than the local ones. In fact, many times the calibrated parameters do not correspond to the global minimum of the involved cost function because the considered local optimizer got stuck in a local minima far from the global solution.

Our work lies in the same vein of the recent works [63, 64] of Sumata et al. In [64] the authors applied a global minimization algorithm in order to calibrate an Arctic Sea Ice-Ocean model. Their approach consists on minimizing, with a genetic algorithm, a cost function corresponding to the model-observation misfit of three sea ice quantities: the sea ice concentration, the drift and the thickness. The similarities between this work and our approach are twofold. The first one is to use a bound constrained global stochastic minimization algorithm. The second is the method to assess on the optimality of the achieved solution by using a pool of independent and randomly initialized minimization experiments. Nevertheless, the approach we are proposing differs from their strategy in several features. First of all, our goal is to calibrate a tsunami model involving less parameters than the 15 model variables of the sea ice-ocean model calibrated in their article. Besides, the different nature between this model and the tsunami model we are looking at, enforces a different optimization window: a large one (two decades) in their work versus a small one (a few hours at most) in our sketch. On top of that, Sumata et al. performed the optimization of the cost function on a discrete search space, while our approach, allowing a continuous parameter domain, is richer.

Based on their previous work [63], Sumata et al. in [64] support, as our work does, the statement that gradient descent local minimization algorithms are likely to get stuck at local minima for these complicated cost functions. Therefore, the authors impose the need to use stochastic global minimization algorithms. In fact, in [63] two types of optimization methods were applied to the calibration of a coupled ocean-sea ice model, and a comparison was made to assess the applicability and efficiency of both methods. One was a gradient descent method based on finite differences for computing the gradient, while the other was a genetic algorithm. Also a parallel implementation was carried out to speed up the optimization process. In the case of the gradient descent method, each component of the gradient was computed in parallel. They



precisely conclude that the global optimization GA is preferred. In fact, it yields a better optimum, since the gradient local optimizers could get trapped in local optima. This could happen even if several launches of the gradient algorithm are launched in a multi-start fashion. This statement exactly coincides with our forthcoming conclusions in Sections 4.1 and 4.2 (see Figures 4 and 11).

In our paper, we overcome this disadvantage by proposing for first time in this field, the use of a parallel hybrid local-global minimization algorithm. More precisely we develop a BH like algorithm. BH consists on hybridizing SA and local gradient searchers, allowing to benefit from both worlds: the global convergence properties of SA and the speed of local optimizers. We go even further by proposing a parallel version of the BH algorithm. For the local searcher ingredient of BH, we use a bounded version of the L-BFGS algorithm employed in [63], namely the L-BFGS-B algorithm. This setting is able to increase the convergence speed and the success rate of BH. The multi-start technique performed in [63] can be seen as computing only one temperature stage of our multi-path BH algorithm. Another advantage of our algorithm is its embarrassingly parallel nature, as we can map each search path to a different parallel thread. In [63] each CPU thread computes one component of the gradient, while in our case, each thread is responsible of one L-BFGS-B path. We show using an analytical test, that this algorithm improves the multi-start technique, as it is always able to find the global optimum. Besides, in our article we compare the efficiency of this multi-path BH with a multi-path SA. Additionally, we show that the use of gradient searches increases the convergence speed of a multi-path SA. As mentioned before, a SA algorithm was also used in [23] to effectively solve the CNOP-P of ROMS.

The organization of this paper is as follows. In Section 2 we pose the data assimilation problem. In Section 2.1 we describe the cost function, which is given by the measure of the mismatch between the free surface laboratory data and the computed one, that depends on the parameters we want to assimilate. The optimization of this cost function is a hard problem. On the one hand, the evaluation of the cost function is an expensive computational problem, because it relies in the solution of a time dependent system of partial differential equations. On the other hand, this data assimilation problem gives rise to a global optimization problem. In Section 2.2 we briefly describe the two-phase tsunami model and give some references about the numerical scheme we use. The physical parameters of the system, that need to be calibrated, are the ratio of densities between the grain and the fluid, the Coulomb friction angle, and the Manning friction coefficient. The evaluation of the cost function requires a numerical solution of this two-phase model, computed for a given set of parameters.

In Section 3, we recall the global optimization algorithms that we will use: multi-path Simulated Annealing and multi-path Basin Hopping algorithms. Both algorithms were proposed in [65, 66] for accelerating the convergence of SA and BH respectively. They are based in performing synchronized parallel Metropolis searches, or parallel gradient based local searches. The methods were assessed against hard benchmarks in the global optimization field. Besides, the

algorithms have been successfully applied to the calibration of models in finance, even in the case where the costly Monte Carlo method is the only alternative to price the involved financial products [67]. In this work we apply these strategies for data assimilation in landslide tsunami modeling. One of the objectives of this article is to show that this type of algorithms can be successfully applied for the parameters calibration on challenging geophysical problems.

In Section 4, some numerical experiments are presented. Section 4.1 is devoted to validating the methodology using synthetic tests. Indeed, the model is run for fixed sets of parameters, and files with the free surface information are generated. Then, this information is considered as data coming from laboratory. Next, by global optimization in a large domain, the parameters that were used to generate those data are recovered. After validating the methodology, in Section 4.2 the developed technique is applied to perform the data assimilation of real laboratory data.

## 2. Data assimilation problem

In general, the cost function measures the error, computed in some norm, between the real data and the solution produced by the numerical model. The model will depend on a set of parameters. For example, in the case of a one layer shallow-water model, they can be: one Manning coefficient for the whole domain, or also several Manning coefficients, one per subdomain; the initial conditions; the boundary conditions, etc. These parameters can be even time dependant (boundary conditions, for example).

### 2.1. Cost function

In this study, the cost function only depends on the free surface elevation because this quantity is easily measurable and perhaps the most important magnitude to predict the tsunami inundation. Thus, to carry out the data assimilation method we introduce the following cost function using the Hilbert space  $L^2(0, T; \Omega)$  norm:

$$f(\mathbf{p}) = \|\eta^{\mathbf{p}} - \eta^{obs}\|_{L^2(0, T; \Omega)} = \left( \int_0^T \|\eta^{\mathbf{p}}(\cdot, t) - \eta^{obs}(\cdot, t)\|_{L^2(\Omega)}^2 dt \right)^{1/2}, \quad (1)$$

where  $\Omega \subset \mathbb{R}$  is the spatial domain and  $[0, T]$  is the time domain. Besides,  $\eta^{\mathbf{p}}(x, t)$  is the free surface elevation at the point  $x$  and at time  $t$  computed with some model using the set of parameters  $\mathbf{p}$ . Finally,  $\eta^{obs}$  are the observed values, that can be obtained from SAR images, sea buoys or laboratory experiments. This leads to an unconstrained global optimization problem in a bounded domain. More precisely, we address problems that can be formulated as

$$\min_{\mathbf{p} \in D \subseteq \mathbb{R}^n} f(\mathbf{p}),$$

where  $f$  is a real valued function of the parameters  $\mathbf{p} \in \mathbb{R}^n$ . These parameters are defined on  $D = \prod_{i=1}^n [l_i, u_i]$ , with  $l_i$  and  $u_i$  being the lower and upper bounds

in direction  $i$ , respectively. The solution can be written as:

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in D \subseteq \mathbb{R}^n} f(\mathbf{p}).$$

In the discrete case, the cost function will have the following expression:

$$f(\mathbf{p}) = \sqrt{\sum_{k=1}^{N_T} \sum_{i=1}^N \left( \eta_{i,k}^{\mathbf{p}} - \eta_{i,k}^{obs} \right)^2},$$

where  $\eta_{i,k}^{\mathbf{p}} = \eta^{\mathbf{p}}(x_i, t_k)$  and  $\eta_{i,k}^{obs} = \eta^{obs}(x_i, t_k)$ .  $x_i$  the  $i$ -th measure point for  $i = 1, \dots, N$ , and  $t_k$  the  $k$ -th measure time for  $k = 1, \dots, N_T$ .

Note that the cost function depends on  $\eta^{\mathbf{p}}$ , which implicitly depends on the parameters to be calibrated. Therefore, a single evaluation of the cost function requires a realization of the numerical model for a given set of parameters. In the next section we present the equations of the two-phase model, pointing out what are the parameters to be calibrated. Some basic idea about the numerical scheme we use is also sketched.

## 2.2. Mathematical model

As discussed in the introduction, we use a two-phase model in order to describe the interaction between the submarine landslide and the fluid. In this work, a Savage-Hutter model [9] is considered for the kinematics of the submarine landslide, and a multi-layer non-hydrostatic shallow-water model is used for the evolution of the ambient water [12]. Both models are coupled through the boundary conditions at the sea-floor.

At this point, we suppose that the landslide is totally submerged and that the ratio of densities between the ambient fluid and the granular material is given by the parameter  $r$ . Usually

$$r = \frac{\rho_f}{(1 - \varphi)\rho_s + \varphi\rho_f},$$

where  $\rho_s$  is the typical density of the granular material,  $\rho_f$  is the density of the fluid ( $\rho_s > \rho_f$ ), and  $\varphi$  is the porosity ( $0 \leq \varphi < 1$ ). Here, we suppose that  $\varphi$  is constant on space and time, and therefore  $r$  is also constant. Note that  $0 < r < 1$ . Finally, let us remark that even on a uniform material,  $r$  is difficult to estimate as it depends on the porosity  $\varphi$ . Typical values of  $r$  are in the interval  $[0.3, 0.8]$ .

The 1D Savage-Hutter model considered in this article is written as follows:

$$\begin{cases} \partial_t z_s + \partial_x(z_s u_s) = 0, \\ \partial_t(z_s u_s) + \partial_x \left( z_s u_s^2 + \frac{g(1-r)}{2} z_s^2 \right) = g(1-r) z_s \partial_x H + \tau_C, \end{cases} \quad (2)$$

where  $g$  is the gravity acceleration ( $g = 9.81 \text{ m/s}^2$ );  $H(x)$  is the non-erodible bathymetry measured from a given reference level and unchanged during the

simulation;  $z_s(x, t)$  is the landslide depth at each point  $x$  at time  $t$ ; and  $u_s(x, t)$  is the averaged horizontal velocity.  $\tau_C$  is the Coulomb friction term given by:

$$\tau_C = -g(1 - r)\mu z_s \frac{\sqrt{u_s^2}}{u_s}.$$

Note that this term is multi-valuated when  $u_s = 0$ . The simplest friction law corresponds to a constant friction coefficient:

$$\mu = \tan(\theta), \quad (3)$$

where  $\theta$  is the friction angle, although more complex friction terms could be used to simulate natural subaerial or submarine landslides [68, 69]. Other definitions derived from experimental data have been proposed by Pouliquen [70], where the friction coefficient depends on the velocity and thickness of the granular layer. This law is widely used in the literature and involves at least three parameters to be calibrated [71].

The Coulomb friction term  $\tau_C$  is quite relevant, as it controls the motion of the landslide. In particular, it is defined in terms of the friction angle  $\theta$ , which is a parameter to calibrate in order to fit the simulation with the experimental data. Finally, let us mention that in the derivation of the previous model we have supposed a rigid-lid assumption with respect to the free surface of the ambient fluid: that is, the pressure variations induced by the fluctuation on the free surface of the ambient fluid over the landslide are neglected. Nevertheless, the buoyancy effects have been taken into account.

The ambient fluid is supposed to be modeled by a multi-layer non-hydrostatic shallow-water system recently proposed in [12]. This system is obtained by a process of depth-averaging of the incompressible Euler equations. More precisely, it can be seen as a particular semi-discretization with respect to the vertical variable of the incompressible Euler equations. Total pressure is decomposed into a sum of a hydrostatic and a non-hydrostatic component. In this process, vertical velocities are assumed to have a linear vertical profile, whilst the horizontal velocities are supposed to have a constant vertical profile. The resulting multi-layer model admits an exact energy balance. Furthermore, when the number of layers increases, the linear dispersion relation of the linear model converges to the same of Airy's theory. The model proposed in [12] can be written in compact form as

$$\left\{ \begin{array}{l} \partial_t h + \partial_x(h\bar{u}) = 0, \\ \partial_t(hu_\alpha) + \partial_x\left(hu_\alpha^2 + \frac{g}{2}h^2\right) - gh\partial_x(H - z_s) \\ \quad + u_{\alpha+1/2}\Gamma_{\alpha+1/2} - u_{\alpha-1/2}\Gamma_{\alpha-1/2} = -h(\partial_x p_\alpha + \sigma_\alpha \partial_z p_\alpha) - \tau_\alpha, \\ \partial_t(hw_\alpha) + \partial_x(hu_\alpha w_\alpha) + w_{\alpha+1/2}\Gamma_{\alpha+1/2} - w_{\alpha-1/2}\Gamma_{\alpha-1/2} = -h\partial_z p_\alpha, \\ \partial_x u_{\alpha-1/2} + \sigma_{\alpha-1/2}\partial_z u_{\alpha-1/2} + \partial_z w_{\alpha-1/2} = 0, \end{array} \right. \quad (4)$$

for  $\alpha \in \{1, 2, \dots, L\}$ , being  $L$  the number of layers. In the previous system, we have used the following notation:

$$\begin{aligned} u_{\alpha+1/2} &= \frac{1}{2}(u_{\alpha+1} + u_{\alpha}), & \partial_z u_{\alpha+1/2} &= \frac{1}{h\Delta s}(u_{\alpha+1} - u_{\alpha}), \\ w_{\alpha+1/2} &= \frac{1}{2}(w_{\alpha+1} + w_{\alpha}), & \partial_z w_{\alpha+1/2} &= \frac{1}{h\Delta s}(w_{\alpha+1} - w_{\alpha}), \\ p_{\alpha} &= \frac{1}{2}(p_{\alpha+1/2} + p_{\alpha-1/2}), & \partial_z p_{\alpha} &= \frac{1}{h\Delta s}(p_{\alpha+1/2} - p_{\alpha-1/2}), \\ \sigma_{\alpha} &= \partial_x(H - z_s - h\Delta s(\alpha - 1/2)), & \sigma_{\alpha-1/2} &= \partial_x(H - z_s - h\Delta s(\alpha - 1)). \end{aligned} \quad (5)$$

As depicted in Figure 1, the flow depth  $h$  is split along the vertical axis into  $L \geq 1$  layers and  $\Delta s = 1/L$ .  $u_{\alpha}$  and  $w_{\alpha}$  are the depth averaged velocities in the  $x$  and  $z$  directions respectively, and  $g$  is the gravitational acceleration. The term  $p_{\alpha+1/2}$  is the non-hydrostatic pressure at the interface  $z_{\alpha+1/2}$ . The free surface elevation measured from the still-water level is  $\eta = h - H + z_s$ , where again  $H(x)$  is the unchanged non-erodible bathymetry measured from a given reference level.  $\tau_{\alpha} = 0$ ,  $\alpha > 1$  and  $\tau_1$  is the Manning friction term that is only present at the lowest layer ( $\alpha = 1$ ) given by

$$\tau_1 = gh \frac{n^2}{h^{4/3}} u_1 |u_1|.$$

Finally, for  $\alpha = 1, \dots, L-1$ ,  $\Gamma_{\alpha+1/2}$  account for the mass transfer across interfaces and are defined by

$$\Gamma_{\alpha+1/2} = \sum_{\beta=\alpha+1}^L \partial_x(h\Delta s(u_{\beta} - \bar{u})), \quad \bar{u} = \sum_{\alpha=1}^L \Delta s u_{\alpha}.$$

Here we suppose that  $\Gamma_{1/2} = \Gamma_{L+1/2} = 0$ , i.e. there is no mass transfer through the bottom nor the free-surface.

In order to close the system, the following boundary conditions are considered:  $p_{L+1/2} = 0$ ,  $u_0 = 0$  and  $w_0 = \partial_t z_s$ . Note that the last two conditions enter into the incompressibility condition for the lowest layer ( $\alpha = 1$ ), given by

$$\partial_x u_{1/2} + \sigma_{1/2} \partial_z u_{1/2} + \partial_z w_{1/2} = 0.$$

Observe that both models are coupled through the unknown  $z_s$ , present in the equations and in the boundary condition ( $w_0 = \partial_t z_s$ ).

Note that the two-phase model depends on three coefficients, that are the ones to be calibrated. More precisely, the vector of coefficient is  $\mathbf{p} = (r, \theta, n)$ , where  $r$  is the ratio of densities between the fluid and the granular phase,  $\theta$  the Coulomb friction angle, and  $n$  the friction (Manning) coefficient. In particular the first two are quite relevant for the landslide motion and therefore, for the induced tsunami water waves.

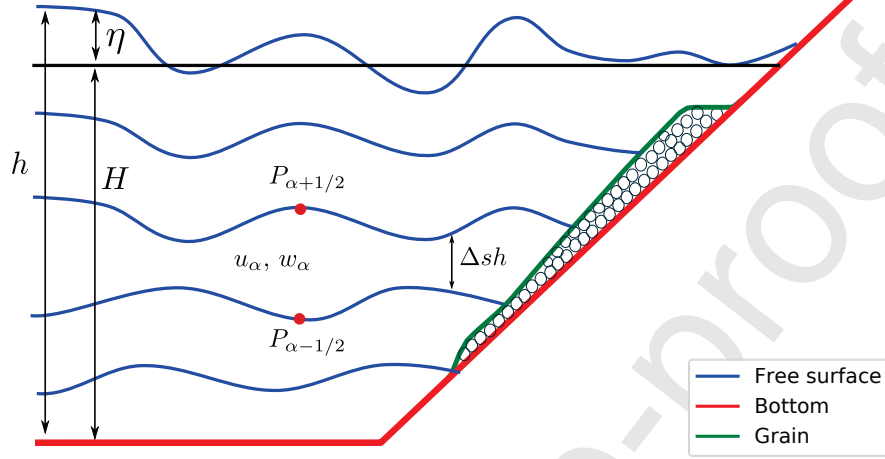


Figure 1: Sketch of the model.

414 System (2) could be written in the following compact way:

$$\partial_t U_s + \partial_x F_s(U_s) = G_s(U_s) \partial_x H - S_s(U_s), \quad (6)$$

415 being

$$U_s = \begin{bmatrix} z_s \\ u_s z_s \end{bmatrix}, \quad F_s(U_s) = \begin{bmatrix} z_s u_s \\ z_s u_s^2 + \frac{g(1-r)}{2} z_s^2 \end{bmatrix},$$

416

$$G_s(U_s) = \begin{bmatrix} 0 \\ g(1-r) z_s \end{bmatrix}, \quad S_s(U_s) = \begin{bmatrix} 0 \\ \tau_C \end{bmatrix}.$$

417 The multi-layer non-hydrostatic shallow-water system could also be expressed  
418 in a similar way:

$$\begin{cases} \partial_t U_f + \partial_x F_f(U_f) + B_f(U_f) \partial_x U_f = G_f(U) \partial_x (H - z_s) + \mathcal{T}_{NH} - S_f(U_f), \\ B(U_f, (U_f)_x, H, H_x, z_s, (z_s)_x) = 0, \end{cases} \quad (7)$$

where

$$U_f = \begin{bmatrix} h \\ hu_1 \\ \vdots \\ hu_L \\ hw_1 \\ \vdots \\ hw_L \end{bmatrix}, F_f(U_f) = \begin{bmatrix} h\bar{u} \\ hu_1^2 + \frac{1}{2}gh^2 \\ \vdots \\ hu_L^2 + \frac{1}{2}gh^2 \\ hu_1w_1 \\ \vdots \\ hu_Lw_L \end{bmatrix}, G_f(U_f) = \begin{bmatrix} 0 \\ gh \\ \vdots \\ gh \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

419 Additionally,  $B_f(U_f)\partial_x(U_f)$  contains the non-conservative products involving  
 420 the momentum transfer across the interfaces

$$B_f(U_f)\partial_x(U_f) = \begin{bmatrix} 0 \\ u_{3/2}\Gamma_{3/2} \\ u_{5/3}\Gamma_{5/2} - u_{3/2}\Gamma_{3/2} \\ \vdots \\ -u_{L-1/2}\Gamma_{L-1/2} \\ w_{3/2}\Gamma_{3/2} \\ w_{5/3}\Gamma_{5/2} - w_{3/2}\Gamma_{3/2} \\ \vdots \\ -w_{L-1/2}\Gamma_{L-1/2} \end{bmatrix}.$$

421 Besides,  $S_f(U_f)$  contains the Manning friction term

$$S_f(U_f) = \begin{bmatrix} 0 \\ \tau_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$



422 The non-hydrostatic corrections in the momentum equations are given by

$$423 \quad \mathcal{T}_{NH} = \mathcal{T}_{NH}(h, h_x, H, H_x, z_s, (z_s)_x, p, p_x) = - \begin{bmatrix} 0 \\ h(\partial_x p_1 + \sigma_1 \partial_z p_1) \\ \vdots \\ h(\partial_x p_L + \sigma_L \partial_z p_L) \\ h \partial_z p_1 \\ \vdots \\ h \partial_z p_L \end{bmatrix}. \quad 424$$

423 Finally, the operator related with the incompressibility condition at each layer  
424 is given by:

$$B(U_f, (U_f)_x, H, H_x, z_s, (z_s)_x) = \begin{bmatrix} \partial_x u_{1/2} + \sigma_{1/2} \partial_z u_{1/2} + \partial_z w_{1/2} \\ \vdots \\ \partial_x u_{L-1/2} + \sigma_{L-1/2} \partial_z u_{L-1/2} + \partial_z w_{L-1/2} \end{bmatrix}.$$

425 The discretization of systems (6)-(7) becomes difficult. In this article, we  
426 have considered the natural extension of the numerical schemes proposed in  
427 [72, 73], where a splitting technique was described. Firstly, the systems (6)-(7)  
428 can be expressed as the following non-conservative hyperbolic system:

$$\begin{cases} \partial_t U_s + \partial_x F_s(U_s) = G_s(U_s) \partial_x H, \\ \partial_t U_f + \partial_x F_f(U_f) + B_f(U_f) \partial_x (U_f) = G_f(U_f) \partial_x (H - z_s). \end{cases} \quad (8)$$

429 Both equations are solved simultaneously using the same *time step*, by means of  
430 a second order HLL, positivity-preserving and well-balanced, path-conservative  
431 finite volume scheme [74]. The synchronization of time steps is done taking into  
432 account the CFL condition of the complete system (8). A first order estimation  
433 of the maximum of the wave speed for system (8) is the following:

$$\lambda_{\max} = \max(|u_s| + \sqrt{(g(1-r)h_s}, |\bar{u}| + \sqrt{gh}).$$

434 Next, the non-hydrostatic pressure corrections  $p_{1/2}, \dots, p_{L-1/2}$  at the ver-  
435 tical interfaces are computed from

$$\begin{cases} \partial_t U_f = \mathcal{T}_{NH}(h, h_x, H, H_x, z_s, (z_s)_x, p, p_x), \\ B(U_f, (U_f)_x, H, H_x, z_s, (z_s)_x) = 0. \end{cases}$$

436 This requires the discretization of an elliptic operator by means of standard sec-  
437 ond order central finite differences. The resulting linear system is solved using  
438 an iterative Scheduled Jacobi method [75]. Finally, the horizontal and vertical  
439 momentum at each layer are updated using the computed non-hydrostatic cor-  
440 rections. At this stage, the frictions  $S_s(U_s)$  and  $S_f(U_f)$  are also discretized fol-  
441 lowing [72, 73]. We refer the reader to [10] for the discretization of the Coulomb  
442 friction term.

### 3. Multi-path BH global optimization

In this section we describe the optimization algorithms multi-path SA ( $SA_M$ ) and multi-path BH ( $BH_M$ ). They can be seen as a modification of the sequential BH algorithm, introducing a parallel multi-path searching technique.

The BH algorithm is a hybrid between the Metropolis algorithm and some kind of gradient local optimization method. Therefore the minimizer profits from the speed and accuracy of the local optimizer, while retaining the global convergence properties of the stochastic one. The seminal idea was presented by Navon and Robertson et al. in [42, 43] for finding the global minimum of Potential Energy Surfaces (PES) related to structures of mixed Argon-Xenon clusters. The authors developed the finite-temperature lattice based Monte Carlo method. Additionally, they compared the use of three different limited memory Quasi-Newton CG methods as local minimizers, L-BFGS against two others, being L-BFGS the better performing one. Seven years later, a similar idea was also successfully applied by Wales and Doye [44] in order to minimize the PES, for finding Lennard-Jones clusters. In this work the Polak Ribière [35] NCG method was used as the local optimizer. The authors named the method Basin-Hopping. This name became widely accepted for referring to this kind of global optimization methods. Nowadays, the term BH encompasses a family of algorithms obtained by combining different local (NCG, BFGS, ...) and global stochastic algorithms (Metropolis or SA). Quasi-Newton methods (BFGS and descendants) are the most common choice for the local component. BH methods have been extensively studied by Locatelli et al. [76, 77, 78, 79] and Leary [80]. In the BH method, the local optimizer can be seen as an operator that transforms the original function  $f$ , returning a new piece-wise constant function  $L$ :  $L(\mathbf{x}) = f(\mathcal{LS}(\mathbf{x}))$ , being  $\mathcal{LS}(\mathbf{x})$  the point where a local minimum of  $f$  is obtained from a starting point  $\mathbf{x}$ . The resulting global optimization problem for the function  $L$ , is much more tractable for the global optimizer, as the barriers between local minima have been softened.

The idea of BH is to use a temperature process like in SA.  $\mathcal{T}$  and  $\mathcal{T}_{min}$  denote the current and minimum temperatures, respectively. The temperature reduction schedule  $\mathcal{T}_{k+1} = \rho \mathcal{T}_k$  is considered, being  $\rho \in (0, 1)$  the cooling rate. Besides, a Metropolis process with  $N$  steps at each temperature level is performed. More precisely, at temperature  $\mathcal{T}_k$  and starting point  $\mathbf{x}_k$ , a random neighbor  $\mathbf{y}_k$  is randomly generated inside a ball centered at  $\mathbf{x}_k$  with radius  $r_k$ ,  $\mathbf{y}_k \in B(\mathbf{x}_k, r_k)$ . Next, in order to obtain a local minimum, a gradient local search starting from  $\mathbf{y}_k$  is launched. Then, the obtained point is accepted or discarded according to Boltzmann's law. Finally, the process advances to the next temperature level. The algorithm stops when the temperature reaches  $\mathcal{T}_{min}$ , or the number of successive rejections exceeds  $J$ . The radius  $r_k$  is periodically updated by using the 50% acceptance rule [81]. A nice property is that BH can also be seen as a generalization of SA: SA can be recovered by skipping the local optimization phase in BH.

In [65, 66] the authors proposed a synched multiple Metropolis path approach for SA and BH algorithms, respectively. The idea is to perform not one, but

488 M Metropolis searches at each temperature level  $\mathcal{T}_k$ . M is also referred as the  
 489 number of searching paths. In the simplified case with  $N = 1$ , the algorithm  
 490 consists of launching M gradient local searchers. The  $l$ -th local search (with  
 491  $l = 1, \dots, M$ ) starts from a random neighbor  $\mathbf{y}_k^l$  of the current minimum point  
 492  $\mathbf{x}_k$ . Thus, in this setting the Metropolis searches are entirely replaced by local  
 493 searches. After performing the Metropolis process at each path, and before  
 494 advancing to the next temperature level, the final information is gathered. The  
 495 best of the attained minima is kept, i.e.  $\mathbf{x}_{k+1}^{best} = \min(\mathbf{x}_*^l)$ . We will refer to this  
 496 algorithm as BH<sub>M</sub>. A sketch of the method is shown in Algorithm 1 and Figure  
 497 2. Note that BH<sub>1</sub> corresponds with the classical BH with only one Metropolis  
 498 path, or just one local search. Also, if the local search operator  $\mathcal{LS}$  is replaced  
 499 with the identity  $id$ , then the multi-path SA algorithm SA<sub>M</sub> [65] is recovered.  
 500 Once more, SA<sub>1</sub> corresponds to the classical single path SA. Furthermore, multi-  
 501 path BH<sub>M</sub> algorithms have two interesting properties. On the one hand, they  
 502 are highly parallelizable. On the other hand, they improve the convergence  
 503 speed and success rate of the classical SA and BH.

---

**Algorithm 1:** Synched multi L-BFGS-B BH, pseudocode.

---

```

y = random uniform in  $D$ ;
Set # successive rejections:  $j = 0$ ;
Iteration number:  $k = 0$ ;
Initial position:  $\mathbf{x}_0 = \mathbf{x}_* = \mathcal{LS}(\mathbf{y})$ ;
while ( $j < J$ ) or ( $\mathcal{T} < \mathcal{T}_{min}$ ) do
  for  $l = 1:M$  do
    for  $i = 1:N$  do
       $\mathbf{y}_i^l$  = random uniform in  $B(\mathbf{x}_i^l, r_k)$ ;
       $\mathbf{u}$  = random uniform in  $[0, 1]$ ;
       $\Delta = L(\mathbf{y}_i^l) - L(\mathbf{x}_*^l)$ ;
      if  $\mathbf{u} < \exp(-\Delta/\mathcal{T})$  then
         $\mathbf{x}_*^l = \mathbf{x}_{i+1}^l = \mathcal{LS}(\mathbf{y}_i^l)$ ;
         $j = 0$ ;
      else
         $j = j + 1$ ;
      end if
    end for
  end for
  Synchronization:  $\mathbf{x}_{k+1}^{best} = \min(\mathbf{x}_*^l)$ ;
  for  $l = 1:M$  do
     $\mathbf{x}_*^l = \mathbf{x}_{k+1}^{best}$ ;
  end for
   $k = k + 1$ ;
  Update  $r_k$ ;
   $\mathcal{T} = \rho \cdot \mathcal{T}$ ;
end while

```

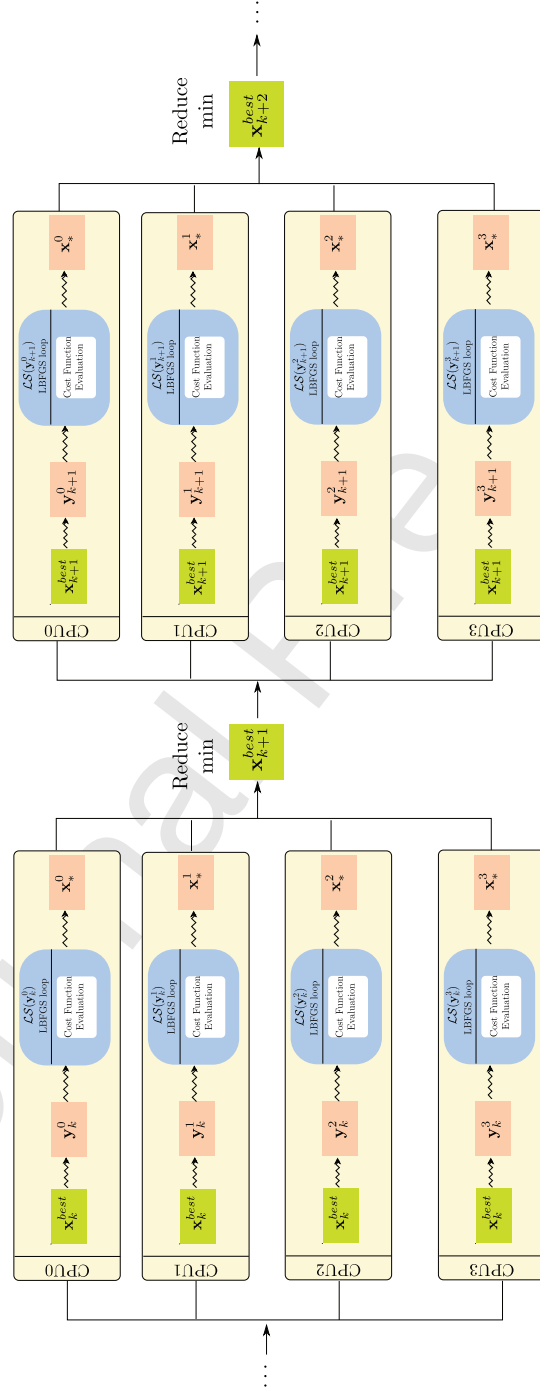
---

This multi-path approach has the advantage of being easily parallelizable: the search paths can be computed asynchronously at the same time. Under a multi-CPU architecture, each CPU thread can take care of computing one local search. Then, the achieved minima have to be synchronized (see Figure 2). This is the hardware setting considered in the present article.

Regarding the convergence properties, in [66] the study of the optimal number of multiple searches is done empirically, both from the convergence and the success rate viewpoints. According to the results, increasing the number of searchers improves the convergence rate, although this increase in convergence rate is not unlimited. For example, if the problem is simple and/or the dimension is low, by increasing the number of searchers one would only obtain a marginal increase in convergence speed. Nevertheless, even in those cases, the computing time is lower if evaluations are done in parallel. Thus, this increase in the number of search paths comes almost for free. Even if the problem is computationally hard, it always comes a point where the optimal convergence rate is achieved and a further increase in the number of searchers will not have any advantage. Usually this number of searches for obtaining an optimal convergence rate is moderate: the optimization problem has to be really tough in order to demand a high number of local searchers. The good properties of the proposed algorithm also apply to the success rate and the same conclusions can be obtained. Generally, it comes a point when a 100% success rate is achieved: more searchers will not yield any advantage. Besides, the number of searches for obtaining this 100% success rate is, once more, normally moderate. For tough problems, the advantage of performing a large number of local searches becomes more evident.

In this work, for the local optimizer we will use the very robust L-BFGS-B algorithm. This minimizer is intended for problems in which information on the Hessian matrix is difficult to obtain. It was presented by Nocedal in [82] as an extension of the L-BFGS minimizer. It allows to solve nonlinear optimization problems with restrictions given by simple bounds on the variables of the function to be optimized. Besides, it is categorized as a limited-memory quasi-Newton algorithm, since it does not need to store the Hessian matrix. In our work we used this bounded minimizer in order to ensure that the optimizer would never explode by following a wrong path outside the physical domain. If one uses a non bounded gradient local optimizer, some search paths could reach points outside the physical domain, where the equations could stop making sense. In that case the evaluation of the cost function (a finite volume solver) may explode, either by crashing or by entering in a very low  $\Delta t$  state (imposed by the CFL condition). As a consequence the assimilation process will crash or never end. We preferred to stay safe with the bounded algorithm, as it has almost the same computational cost as the unbounded L-BFGS version.

In order to compute the gradient of the objective function, one can use either algorithms based on the so-called adjoint method or the standard finite-difference approximation. Both techniques have their own advantages and disadvantages. In this article we opted for the finite difference procedure attending to the reasons that will be discussed hereafter.

Figure 2: Schematic visualization of the BH<sub>M</sub> algorithm (with M = 4).

550 There are two different approaches for tackling the adjoint problem. One  
 551 technique is the classical approach developed by Lions [15], and applied by Lai  
 552 and Monnier in [49] to the simpler 2D one layer shallow water model. It con-  
 553 sists in computing the adjoint PDE system, and then solving it by numerical  
 554 methods. This is a very challenging problem even for the simpler shallow water  
 555 model assimilated by Lai and Monnier, and even much more for our problem  
 556 at hand. We emphasize that we are dealing with a coupled model involving an  
 557 arbitrary number of fluid layers of “shallow-water type systems”, along with the  
 558 Savage-Hutter equations. This results in a large hyperbolic system of coupled  
 559 conservation laws. The mentioned system can only be numerically approximated  
 560 by means of very involved finite volume numerical discretizations. Therefore,  
 561 dealing with the corresponding stability problems related to the high nonlinear-  
 562 ities involved in hyperbolic problems along with spatial-temporal discretization  
 563 issues is mandatory. As a consequence, the adjoint method will lead to a system  
 564 of conservation laws with source terms and non-conservative products, for which  
 565 it would not be clear the hyperpolicity region. Besides, the numeric approxima-  
 566 tion of this adjoint system will be very sophisticated. One wonders if all this  
 567 challenging work, even if feasible, is worth it for calibrating just this particular  
 568 model. On the other hand, a way to circumvent those difficulties and avoid  
 569 computing the adjoint system, is to compute the partial derivatives by means  
 570 of Automatic Differentiation (AD). Note that in the close future we pretend to  
 571 tackle real two dimensional problems. They involve much higher computational  
 572 cost, and consequently even more for the adjoint AD procedure. In this sce-  
 573 nario, speeding up on GPUs the cost function evaluation becomes compulsory.  
 574 Additionally, the automatic differentiation algorithm should be carried out in  
 575 the GPU side. Therefore, an AD library for GPUs is needed, something that  
 576 can be an obstacle due to the fact that these tools are not always available,  
 577 specially for massively parallel architectures like GPUs. Furthermore, the code  
 578 should be rewritten from the very basics using the overloaded operators pro-  
 579 vided by the AD library. On top of that, more memory will be needed in this  
 580 adjoint setting, which is again an issue in GPUs.

581 Having in mind all the previously discussed issues, in this article we opted  
 582 for the direct numerical approximation of the partial derivatives involved in the  
 583 gradient using finite differences. In our case this has several advantages when  
 584 compared to the adjoint computation. First of all, one of our goals is to de-  
 585 velop a generic data assimilation machinery for landslide tsunami models. This  
 586 framework should be directly applicable if one wants to enrich the here con-  
 587 sidered model with further characteristics, or even fully replace it with other  
 588 models. The tool should endow us with a way to compare the accuracy of  
 589 (possibly quite) different models. This is a strong reason for not developing  
 590 an algorithm that is too tailored for a particular model or numerical scheme.  
 591 In this sense, by computing the gradient via finite differences we gain general-  
 592 ity. In fact, the method can be easily applied to models of all kinds without  
 593 changes in the calibration procedure, in the same vein of [63]. One will just  
 594 need to invoke it by plug in the new model solver (no changes are needed in  
 595 the solver, unlike with the adjoint method). Hence we are well positioned in

order to face the calibration of the previously mentioned oncoming richer two dimensional model to real data. Additionally, our technique is able to cope with the strongly nonlinear relation between model state and parameters, for which other approaches based on Kalman filter have difficulties. Finally, if the number of optimization variables is short, the finite-difference method is not much more computing time demanding than the adjoint method. Indeed, this is the case we are dealing with. Our goal is to calibrate three parameters, namely the ratio  $r$  of densities between the fluid and the granular phase, the Coulomb friction angle  $\theta$ , and the friction Manning coefficient  $n$ . Last but not least, nowadays, thanks to the available high computational power, the numerical computation of the gradient could be directly addressed making use of parallel codes. The right approach is to combine a multi-CPU version of the optimizer with a multi-GPU implementation of the numerical solver used to evaluate the cost function.

All in all, the gradient of the cost function will be numerically computed, using first order progressive finite differences

$$\frac{\partial f}{\partial p_i}(\mathbf{p}) = \frac{f(\mathbf{p} + \varepsilon \mathbf{e}_i) - f(\mathbf{p})}{\varepsilon},$$

with  $\varepsilon = 10^{-6}$ , being  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$  the unitary vector of direction  $i$ .

Regarding the implementation of the algorithms, the whole code of the cost function, its gradient, and the optimization algorithms is custom made. Both algorithms have been integrated in an efficient code using C++, and OpenMP is used for the parallel implementation of the optimization codes (see Figure 2). Also we want to emphasize that the cost function is integrated with the optimization tool, so that it is called on the fly for each set of parameters during the optimization process. Therefore no intermediate results need to be discharged from RAM to the hard drive for computing the value of the cost function, thus resulting in an efficient code. Furthermore, during the whole optimization process the laboratory data is read only once at the beginning.

#### 4. Numerical results

In this section we present two sets of numerical examples. The first one in Section 4.1 is a pool of synthetic tests with known solutions. They are used to validate the proposed algorithms and methodology, to discuss about the identifiability of the problem, and to show the convergence results and computational speedup. The second one in Section 4.2 shows an application of the proposed methodology to the assimilation of real laboratory data.

The laboratory experiment that will be calibrated in this article was presented in [83], and the data can be accessed at [84]. In that work, the authors designed different laboratory experiments and performed numerical simulations to validate a landslide tsunami model. Additionally, they assessed how tsunami hazard from SMFs is affected by slide kinematics and rheology.

In [85] the Tsunami-HySEA model is used to perform some of the numerical benchmark problems proposed in [83]. The obtained results are documented in



the “Proceedings and results of the 2011 NTHMP Model Benchmarking Workshop”.

In our article we focus in one of the experiments performed in [84]: the benchmark 4 (deformable submarine landslide). For both the analytical and the laboratory experiments, the physical conditions of this benchmark are considered. The length of the channel is 6 meters, and its sketch can be seen in Figure 3. The initial condition is water at rest with  $\eta = 0$  and a triangular block of sediments, whose geometry is depicted in Figure 3. In our numerical results, we will have four tide-gauges,  $N = 4$ , where laboratory measures have been taken each 5 milliseconds, thus generating four tidal series. These buoys are located at the positions 1.87, 2.87, 3.87 and 4.87 meters, and they are depicted in Figure 3. We take  $g = 9.81 \text{ m/s}^2$  and  $L = 5$  layers of fluid in the model.

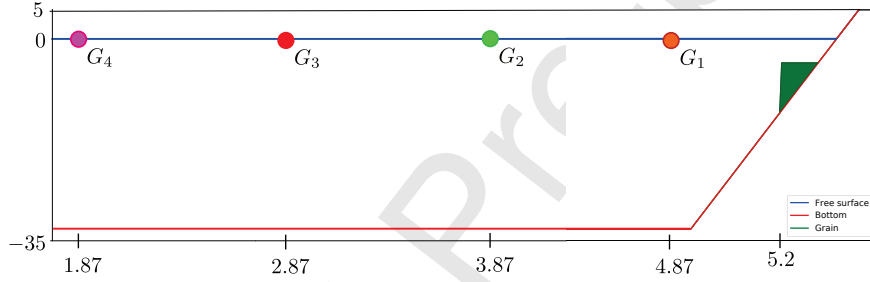


Figure 3: Sketch of the channel, initial condition and position of the tide-gauges.

The calibration tests are run until  $T = 8$  seconds both for the synthetic test and the laboratory experiment. For the finite volume method we consider 200 cells in the analytical test and 800 cells in the laboratory essay with  $CFL = 0.5$ .

We recall that the parameters are three,  $\mathbf{p} = (r, \theta, n)$ , where  $r$  is the ratio of densities between the fluid and the sediment,  $\theta$  the Coulomb angle, and  $n$  the friction coefficient. The search domain for all the experiments in this section is  $D = [0.3, 0.8] \times [5, 45] \times [10^{-5}, 10^{-3}]$ , which is quite a broad domain.

Concerning the hardware configuration, all tests have been performed in a server with 16 CPU cores (two Intel Xeon E5-2620 v4 clocked at 2.10GHz, accounting 32 logical threads) and 16 GB of RAM.

#### 4.1. Synthetic test

The notion of identifiability addresses the question of whether it is at all possible to obtain unique solutions of the inverse problem for unknown parameters of interest in a model from data collected in the spatial and temporal domains [86, 87]. As we have seen, data assimilation problems deal in the end with the search of the global minimum of a cost function. The exploration of the global minimum is a nontrivial task as long as the cost function has a complicated structure. On top of that, ensuring that the involved cost function has a unique global minimum is a extremely difficult goal. This is mainly due

to the fact that sophisticated numerical methods are needed to simulate from landslide tsunami models being able to recover realistic physical phenomena, as discussed in Section 2.2.

Analyzing parameters identifiability is precisely one of the aims of this work. We seek to check whether the data assimilation problem for landslide tsunami models is well posed when using only information of the fluid free surface. In fact, our goal in this first set of numerical experiments is precisely to empirically discuss the problem of identifiability and uniqueness of the here proposed parameters calibration strategy. We will observe that the parameters are identifiable using only data from the free surface. This is something unexpected and eye catching. At first sight, one could expect that information of the lower layers, the sediment layer or the speed of the fluid should be required in order to assimilate the data into the model. Nevertheless, the information of the free surface proves to be enough in practice.

In this work, as in the article [64], the uniqueness of the minimum of the cost function will be discussed by invoking results from the following additional optimization experiments.

#### 4.1.1. Synthetic test 1

First of all, we designed a synthetic experiment, where given the unknown set of parameters, we created the observations numerically, which were then assimilated into the model to retrieve the original set of parameters. The values of the parameters were set at  $r = 0.55$ ,  $\theta = 12^\circ$  and  $n = 0.0002$ . The test was run for 8 seconds. With these data we computed the simulation and stored the series corresponding with the free surface at each measure point in an interval of 0.005 seconds. Then, we supposed that the parameters were unknown and tried to recover them using our optimization algorithms. There is no doubt about the uniqueness of the global minimum: the value of the cost function at this unique global minimum is zero, since the observations are perfect because they arise from the model. This problem has a very similar level of complexity from the optimization point of view to the real one we want to tackle, although it has the advantage of being easier to handle, as the exact solution is known. Moreover, this benchmark allowed us to test and compare the different algorithms with different number of parallel search paths.

First we show that if only a local optimization algorithm is applied, no convergence to global minimum is obtained. In fact, after executing a local L-BFGS-B searcher starting from a random point of the search domain, the obtained set of parameters is  $(r, \theta, n) = (0.682698, 10.688417^\circ, 8.178492 \times 10^{-4})$ , the value of the cost function being  $5.27 \times 10^{-2}$ . The simulation obtained with this set of parameters is shown and compared with the exact solution at Figure 4. Therefore, a robust global optimization algorithm should be used to seek the global minimum of this problem.

Figure 5 shows the results obtained if hybrid multi-path SA or BH algorithms are applied. Now, the parameters are computed exactly (see Table 1), and a perfect agreement between the signals is observed.

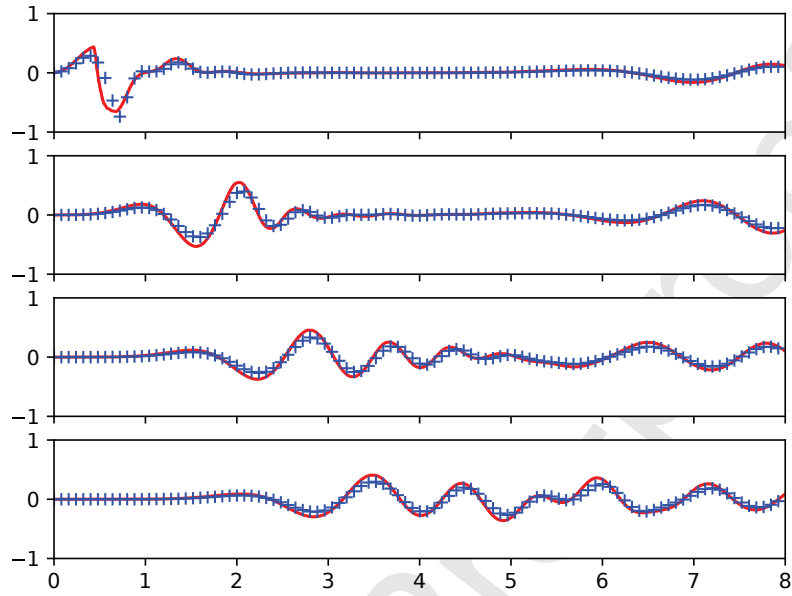


Figure 4: Synthetic generated series vs calibrated ones with the multi-start L-BFGS-B. Target series in red, simulated series in blue.

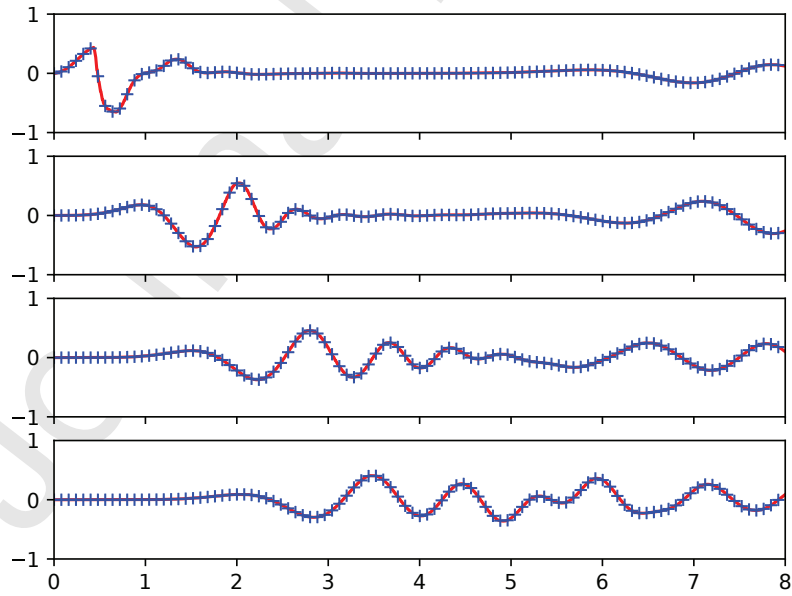


Figure 5: Synthetic generated series vs calibrated ones. Target series in red, simulated series in blue.

	$r$	$\theta$	$n$
Target values	0.55	$12^\circ$	0.0002
Obtained values	0.55	$12^\circ$	0.0002

Table 1: Target and obtained values of the parameters.

We can also use this benchmark to assess the convergence and efficiency of the two proposed hybrid multi-path global optimization algorithms. In Figures 6 and 7, a comparison of the convergence of  $SA_M$  and  $BH_M$  algorithms is shown, respectively, when using different number of paths ranging from 1 to 16. At each temperature, the value of the cost function at the best point visited so far by the algorithm is shown. Note that the current state of the minimizer at each stage could be different to the referred best visited point owing to the stochastic nature of the SA and BH algorithms.

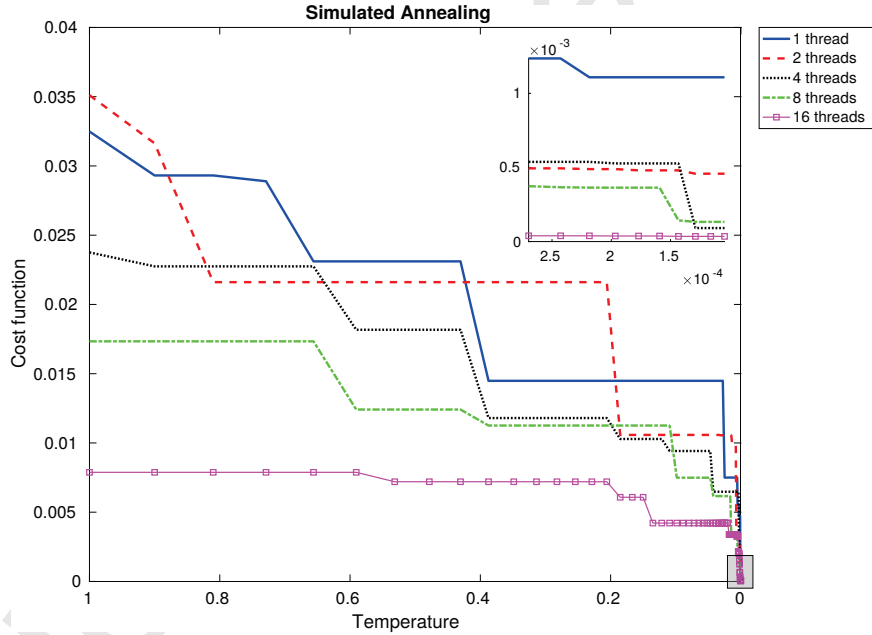


Figure 6: Convergence of multi-path SA with 1,2,4,8 and 16 search paths.

In Table 2 we show the convergence of the multi-path algorithms when increasing the number of paths. The convergence speed is shown in terms of the number of cost function evaluations performed by the algorithm. This number of evaluations is shown at different levels of temperatures in the annealing process,  $\mathcal{T} = 1, 0.48, 10^{-4}$ , and for different number of search paths, ranging from 1 to 16. The computing time of each function evaluation for this test is 2.8 seconds in our hardware configuration. Note that when doing more than

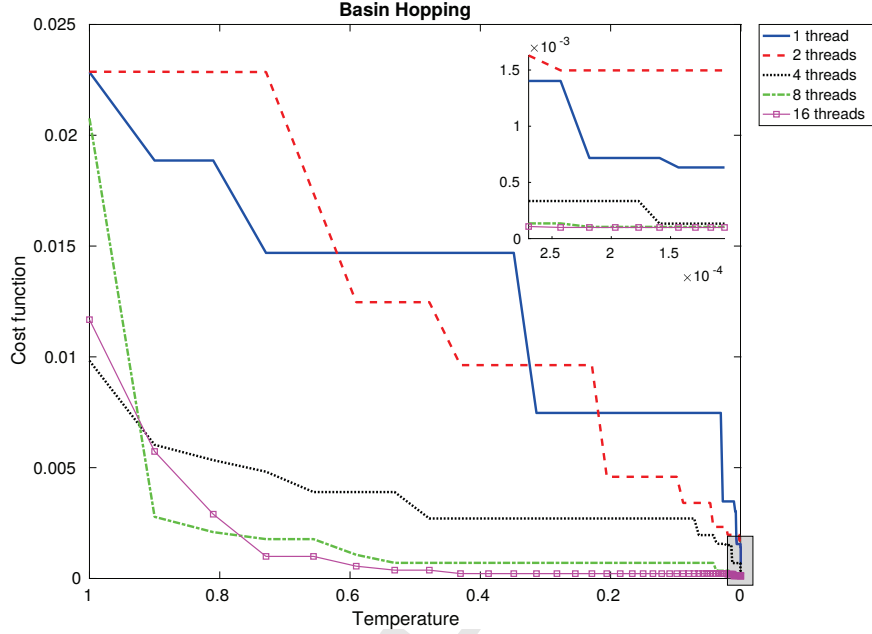


Figure 7: Convergence of multi-path BH with 1,2,4,8 and 16 search paths.

one search, the searches are distributed among the number of CPU cores. Besides, the number of function evaluations performed by  $BH_M$  include the three extra evaluations for computing the gradient. In Table 3 we show the parallel computational efficiency, in terms of the speedup, when using multiple cores for performing 16 search paths, with a number of threads ranging from 1 to 16.

Next, the convergence of the algorithm to the global optimum is assessed when a lower number of measure points is used. We made the experiment of considering the time series of the free surface only at tide-gauge G4, or only at tide-gauges G3-G4. In Table 4 the value of the cost function together with the obtained set of parameters using only data from G4 are shown. The same information is also shown when calibrating against tide-gauges G3-G4. As expected, the value of the cost function is better when taking the four tide-gauges.

#### 4.1.2. Synthetic test 2

Secondly, we ran a pool of 20 independent optimization experiments with our setup. Each optimization started from a different initial parameter configuration randomly chosen in the search domain. Furthermore, every single test used a different seed for the creation of the random numbers consumed by the algorithm in order to explore the search domain. Therefore, all experiments performed a seek of the minimum from a different starting point along a different search path. Figure 8 shows the evolution of the cost function and its spread for the 20 optimization experiments. The spread is defined by the range of the maximum

Simulated Annealing	#Threads	$\mathcal{T}$	#Func. Evals.	Cost Function
	1	1	101	$3.25 \times 10^{-2}$
		0.48	$3.61 \times 10^3$	$2.31 \times 10^{-2}$
		$10^{-4}$	$3.92 \times 10^5$	$1.11 \times 10^{-3}$
	2	1	201	$3.51 \times 10^{-2}$
		0.48	$7.21 \times 10^3$	$2.16 \times 10^{-2}$
		$10^{-4}$	$7.83 \times 10^5$	$4.58 \times 10^{-4}$
	4	1	401	$2.38 \times 10^{-2}$
		0.48	$1.44 \times 10^4$	$1.82 \times 10^{-2}$
		$10^{-4}$	$1.57 \times 10^6$	$9.11 \times 10^{-5}$
	8	1	801	$1.73 \times 10^{-2}$
		0.48	$2.88 \times 10^4$	$1.24 \times 10^{-2}$
		$10^{-4}$	$3.13 \times 10^6$	$1.33 \times 10^{-4}$
	16	1	1601	$7.88 \times 10^{-3}$
		0.48	$5.76 \times 10^4$	$7.20 \times 10^{-3}$
		$10^{-4}$	$6.27 \times 10^6$	$3.50 \times 10^{-5}$
Basin Hopping	#Threads	$\mathcal{T}$	#Func. Evals.	Cost Function
	1	1	82	$2.28 \times 10^{-2}$
		0.48	$1.72 \times 10^3$	$1.47 \times 10^{-2}$
		$10^{-4}$	$1.96 \times 10^5$	$6.33 \times 10^{-4}$
	2	1	103	$2.29 \times 10^{-2}$
		0.48	$3.48 \times 10^3$	$1.24 \times 10^{-2}$
		$10^{-4}$	$3.15 \times 10^5$	$1.50 \times 10^{-3}$
	4	1	285	$9.83 \times 10^{-3}$
		0.48	$6.99 \times 10^3$	$2.71 \times 10^{-3}$
		$10^{-4}$	$6.49 \times 10^5$	$1.33 \times 10^{-4}$
	8	1	369	$2.08 \times 10^{-2}$
		0.48	$1.30 \times 10^4$	$7.04 \times 10^{-4}$
		$10^{-4}$	$1.38 \times 10^6$	$1.04 \times 10^{-4}$
	16	1	825	$1.17 \times 10^{-2}$
		0.48	$2.51 \times 10^4$	$3.75 \times 10^{-4}$
		$10^{-4}$	$2.67 \times 10^6$	$9.92 \times 10^{-6}$

Table 2: Parallel SA ( $SA_M$ ) vs. parallel BH ( $BH_M$ ). The column labeled as “Cost Function” shows the value of the cost function at the best point visited so far by the minimization algorithm.

Number of cores	1	2	4	8	16
Time (seconds)	872.64	1640.57	3035.07	5493.48	9338.92
Speedup	1	1.88	3.47	6.30	10.70

Table 3: Multi-path  $BH_{16}$ : speedup using multi-CPU implementation.

745 value of the cost function and its minimum in the set of the 20 optimizations at  
746 each temperature step. The average of the 20 realizations was also computed.

Gauges	Parameters			Cost func.
	$r$	$\theta$	$n$	
$G1-G2-G3-G4$	0.55	$12^\circ$	$2 \times 10^{-4}$	$9.92 \times 10^{-6}$
$G3-G4$	0.549343	$11.840494^\circ$	$2.045592 \times 10^{-4}$	$6.23 \times 10^{-4}$
$G4$	0.552934	$11.250767^\circ$	$2.132503 \times 10^{-4}$	$1.64 \times 10^{-3}$

Table 4: Calibrated parameters and value of cost function.

747 More precisely at each temperature, the worse, the best, and the average of  
 748 the best points visited by the twenty minimizers up to the current temperature  
 749 are shown. All experiments show an asymptotic reduction of the values of  
 750 the cost function toward the same zero value, and none of the optimizations  
 751 ends up in a local minimum. Therefore, this study clearly shows that this  
 752 stochastic approach (hybrid local-global optimization) is suitable to find the  
 753 global minimum of a structurally complicated cost function.

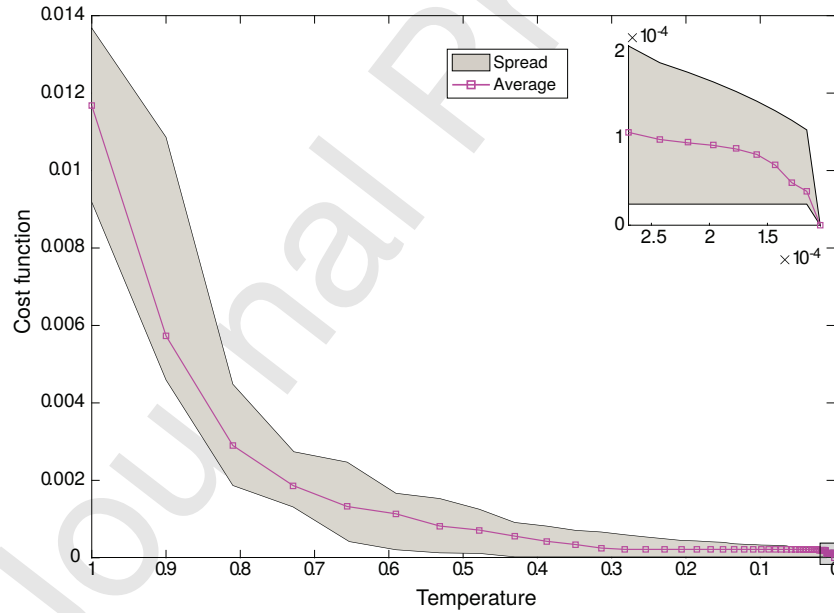


Figure 8: Evolution of the cost function for 20 optimization experiments. The gray shade denotes the spread (the range of maximum-minimum cost) and the squared line the average of the 20 experiments.

#### 754 4.1.3. Synthetic test 3

755 Finally, we sampled the search domain with 18 sets of parameters. For all of  
 756 them, once more we generated the corresponding synthetic tests and performed  
 757 a successful data assimilation. These experiments are summarized on Table 5.



We note that with this pool of data we swimmingly calibrated the model to all types of waves varying from those with very high amplitudes to the flat ones, see Figure 9. Notice that this smoothing effect was obtained by increasing more and more the ratio of densities  $r$  and the Coulomb angle  $\theta$ . Therefore, the issue of identifiability has been accomplished for the very different types of possible waves.

$r$	$\theta$	$n$	Cost func.
0.35	12	0.0002	$1.19 \times 10^{-5}$
0.35	12	0.0004	$9.34 \times 10^{-5}$
0.35	25	0.0002	$2.90 \times 10^{-5}$
0.35	25	0.0004	$6.53 \times 10^{-5}$
0.35	37	0.0002	$3.50 \times 10^{-5}$
0.35	37	0.0004	$6.41 \times 10^{-5}$
0.55	12	0.0002	$1.89 \times 10^{-5}$
0.55	12	0.0004	$8.75 \times 10^{-5}$
0.55	25	0.0002	$6.89 \times 10^{-5}$
0.55	25	0.0004	$2.75 \times 10^{-5}$
0.55	37	0.0002	$2.06 \times 10^{-5}$
0.55	37	0.0004	$8.33 \times 10^{-5}$
0.75	12	0.0002	$5.49 \times 10^{-5}$
0.75	12	0.0004	$8.61 \times 10^{-5}$
0.75	25	0.0002	$4.81 \times 10^{-5}$
0.75	25	0.0004	$8.96 \times 10^{-5}$
0.75	37	0.0002	$4.62 \times 10^{-5}$
0.75	37	0.0004	$2.21 \times 10^{-5}$

Table 5: Values of the cost function for several data assimilations.

#### 4.2. Application to a laboratory test with real data

In this experiment we performed the data assimilation for a real situation where laboratory series of the free surface for four measure points are given. The experiment was performed at École Centrale de Marseille (IRPHE), France, [83]. The positions of the measure buoys are once more 1.87, 2.87, 3.87 and 4.87 meters. The time series for these points are shown in Figure 10. These time series, together with the description of the experiment and some videos, are available in the web page [84].

One more time, first we show that the results obtained with a multi-start algorithm are worse than those obtained with a hybrid multi-path algorithm. If a multi-start L-BFGS-B minimizer is applied to this problem, the obtained solution does not match adequately the laboratory data (see Figure 11). This experiment corresponds with launching only one temperature step of  $BH_M$  with 32 paths. The set of obtained parameters is  $(r, \theta, n) = (0.463274, 9.480647^\circ, 8.176018 \times 10^{-1})$ , for which the value of the cost function is  $1.32 \times 10^{-1}$ .

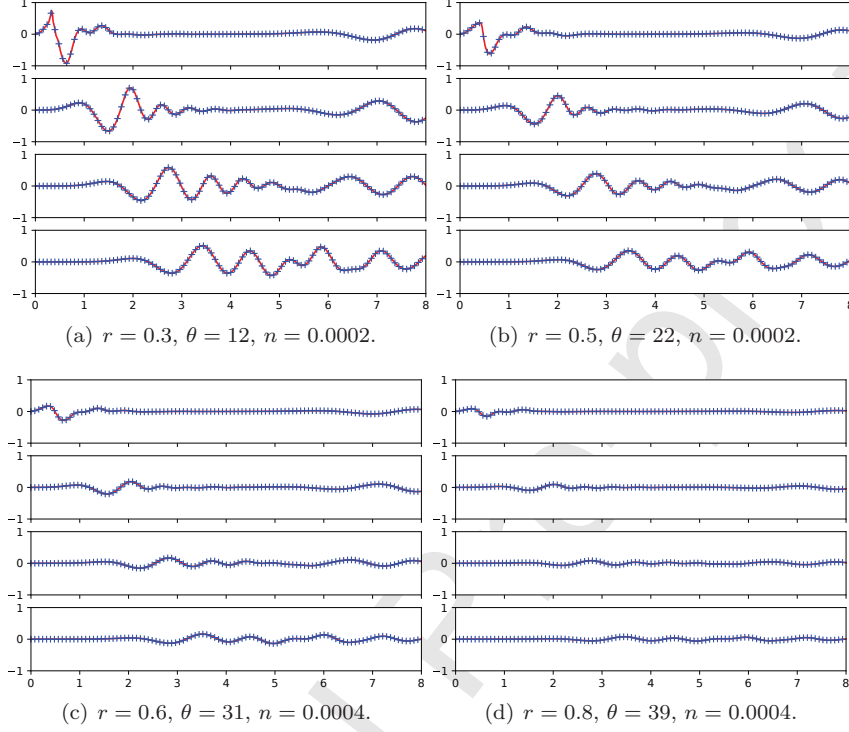


Figure 9: Synthetic generated series vs calibrated ones. In red, target series a priori generated with a set of known parameters. In blue, simulated series obtained with the assimilated parameters achieved with the global optimizer.

After global calibration with  $BH_M$ , the results can be seen in Figures 12 and 13. The obtained values for the parameters are shown in Table 6 and the value of the cost function is  $1.22 \times 10^{-1}$ .

In the aforementioned figures we can see that with the calibrated set of parameters a good agreement in the signals amplitudes and pulses is obtained, between laboratory and simulated data. The approximation is even better up to the 4th second (see Figure 12). The matching is quite good at initial seconds, and it becomes worse as time evolves. Also a better agreement for the farthest tide-gauges  $G3$  and  $G4$  is observed. However, it becomes worse the closer we are to the initial position of the landslide, close to tide-gauge  $G1$ . The amplitudes of the signal are very well captured by the model. The period (pulses, maximums and minima of the signal) is well captured for the last three tide-gauges until the fourth second, and there is a little gap from that time on. The first tide-gauge is difficult to be captured by the model. Further investigation should be done. In fact, at this early stage, compaction and dilatancy effects are quite

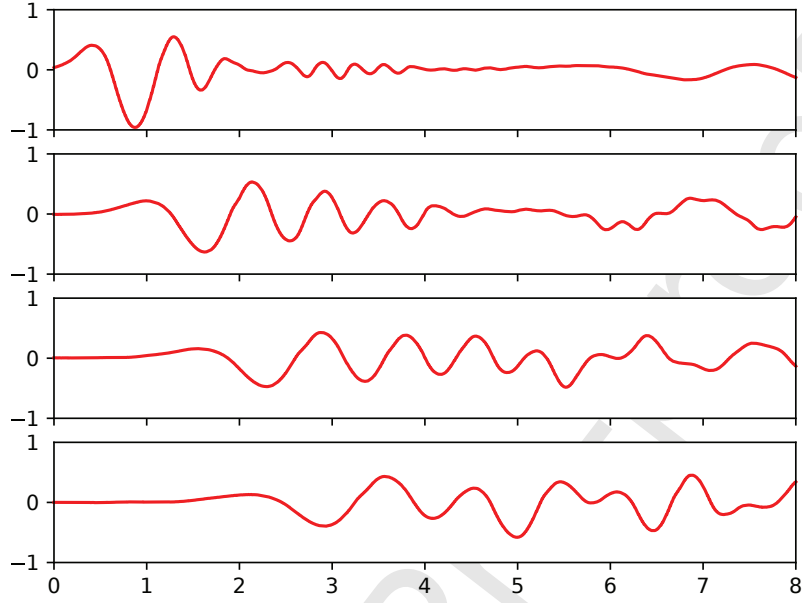


Figure 10: Series measured in laboratory experiments.

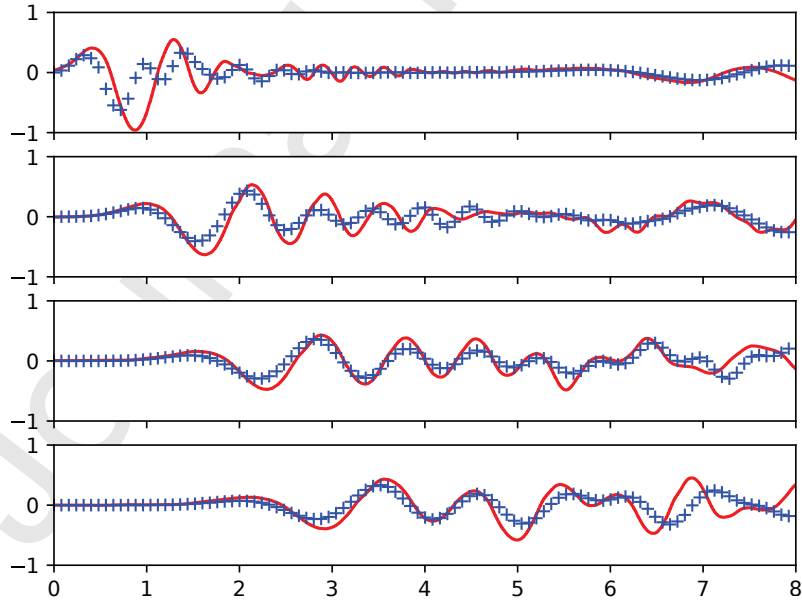


Figure 11: Multi-start solution with 32 L-BFGS-B local searches: computed signals in blue and laboratory data in red.

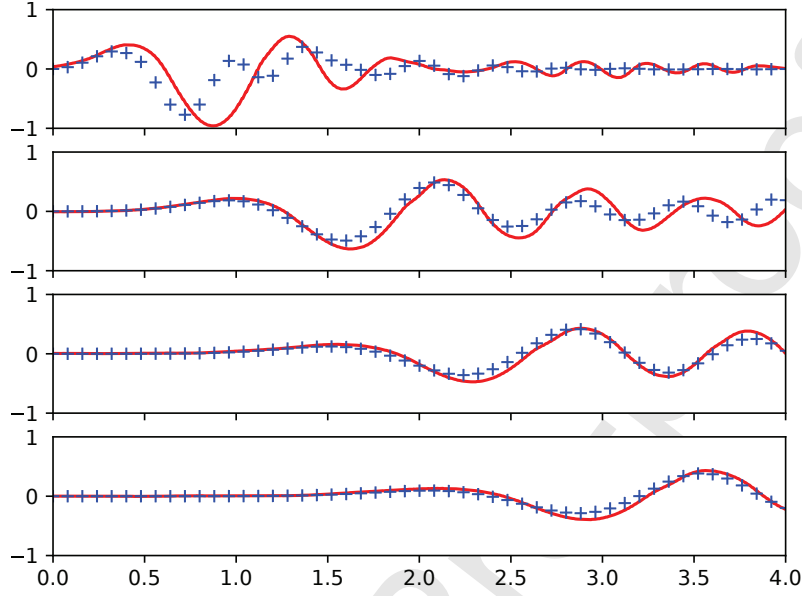


Figure 12: Laboratory series vs calibrated ones. Laboratory series in red, simulated series in blue. From top to bottom, free surface at tide-gauges  $G1$ ,  $G2$ ,  $G3$  and  $G4$ .

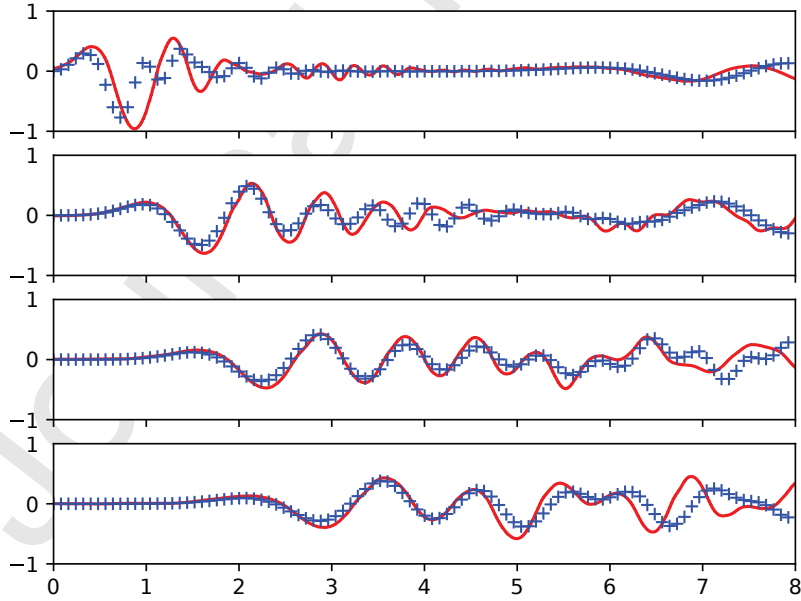


Figure 13: Laboratory series vs calibrated ones. Laboratory series in red, simulated series in blue. From top to bottom, free surface at tide-gauges  $G1$ ,  $G2$ ,  $G3$  and  $G4$ .

Gauges	Parameters			Cost func.
	$r$	$\theta$	$n$	
$G1-G2-G3-G4$	0.650116	$6.035102^\circ$	$4.369000 \times 10^{-4}$	$1.22 \times 10^{-1}$
$G3-G4$	0.708088	$5.387702^\circ$	$3.144702 \times 10^{-4}$	$1.24 \times 10^{-1}$
$G4$	0.763357	$5.162403^\circ$	$2.397312 \times 10^{-4}$	$1.31 \times 10^{-1}$

Table 6: Obtained values of the parameters and value of cost function.

important, and they are not taken into account in the here considered landslide model. Therefore, a more accurate model for the landslide motion is needed to better simulate this early stages of the landslide motion.

Newly in this laboratory experiment we repeated the practice of using a lower number of measure points. We made the test of considering the free surface series only at tide-gauge  $G4$ , or only at tide-gauges  $G3-G4$ . The results are shown in Table 6. The obtained error considering the four series until time  $T = 8$  seconds, using the parameters calibrated with only the last tide-gauge  $G4$ , is  $1.31 \times 10^{-1}$ . Besides, the obtained error considering the four series, using the parameters calibrated with only the last two tide-gauges  $G3-G4$  is  $1.24 \times 10^{-1}$ . The result is not too poor when considering only the measures of the last tide-gauge, nevertheless it is off course much better when considering  $G3-G4$ . Finally, note that in this case, the free surface series are quite close to the best obtained result using the four tide-gauges.

## 5. Conclusions

We have shown that hybrid multi-path global optimization algorithms are suitable for solving the data assimilation problem for SMF models. Besides, we have assessed the identifiability of the model parameters if only data of the free surface is available.

Additionally, we have discussed that using a local optimizer or a multi-start technique produces poor results: global optimization algorithms are more suitable for this kind of problems. We have also exhibited that the problem can be solved using gradient-based numerical optimization algorithms in the local part.

The here developed calibration technique results also interesting because it allows to measure the quality of the model. In fact, the quality of two different models can be quantitative (not only qualitative) compared attending to the result of the calibration. It provides us with a machinery for comparing the good properties of different models. The one with the lowest minimum, can be quantitative said to better approximate the real physical problem.

The laboratory experiment is quite challenging. The obtained results look promising, although a perfect match between laboratory data and the calibrated model has not been achieved due to limitations of the underlying model. In any case, we have shown that the multi-path BH algorithm can be used to calibrate this kind of problems.

As a conclusion, this work opens the door for tackling data assimilation of real SMF problems using 2D models. Besides, this machinery can help in the development of better models for simulating submarine landslide tsunamis.

## 6. Acknowledgments

The authors want to acknowledge the designers of the experiment [83], for making the data publicly available. The authors also wish to thank the anonymous reviewers for their through review of the article and their constructive advises.

This research has been financially supported by Spanish Government Ministerio de Economía y Competitividad through the research projects MTM2016-76497-R and MTM2015-70490-C2-1-R.

## References

- [1] S. T. Grilli, P. Watts, Tsunami Generation by Submarine Mass Failure. I: Modeling, Experimental Validation, and Sensitivity Analyses, *J. Waterw Port Coast* 131 (6) (2005) 283–297.
- [2] I. V. Fine, A. B. Rabinovich, B. D. Bornhold, R. E. Thomson, E. A. Kulikov, The Grand Banks landslide-generated tsunami of November 18, 1929: preliminary analysis and numerical modeling, *Mar. Geol.* 215 (1) (2005) 45 – 57.
- [3] A. Skvortsov, B. Bornhold, Numerical simulation of the landslide-generated tsunami in Kitimat Arm, British Columbia, Canada, 27 April 1975, *J. Geophys. Res.-Earth* 112 (F2) (2007) 1–12.
- [4] S. M. Abadie, J. C. Harris, S. T. Grilli, R. Fabre, Numerical modeling of tsunami waves generated by the flank collapse of the Cumbre Vieja Volcano (La Palma, Canary Islands): Tsunami source and near field effects, *J. Geophys. Res.-Oceans* 117 (C5) (2012) 1–26.
- [5] J. Horrillo, A. Wood, G.-B. Kim, A. Parambath, A simplified 3-D Navier-Stokes numerical model for landslide-tsunami: Application to the Gulf of Mexico, *J. Geophys. Res.-Oceans* 118 (12) (2013) 6934–6950.
- [6] S. Assier Rzedkiewicz, C. Mariotti, P. Heinrich, Numerical simulation of submarine landslides and their hydraulic effects, *J. Waterw Port Coast* 123 (4) (1997) 149–157.
- [7] G. Ma, J. T. Kirby, F. Shi, Numerical simulation of tsunami waves generated by deformable submarine landslides, *Ocean Model.* 69 (2013) 146–165.
- [8] R. M. Iverson, The physics of debris flows, *Rev. Geophys.* 35 (3) (1997) 245–296.

- [9] S. B. Savage, K. Hutter, The motion of a finite mass of granular material down a rough incline, *J. Fluid Mech.* 199 (1989) 177–215.
- [10] E. Fernández-Nieto, F. Bouchut, D. Bresch, M. C. Díaz, A. Mangeney, A new Savage-Hutter type model for submarine avalanches and generated tsunamis, *J. Comput. Phys.* 227 (16) (2008) 7720–7754.
- [11] G. Ma, J. T. Kirby, T.-J. Hsu, F. Shi, A two-layer granular landslide model for tsunami wave generation: Theory and computation, *Ocean Model.* 93 (2015) 40–55.
- [12] E. Fernández-Nieto, M. Parisot, Y. Penel, J. Sainte-Marie, A hierarchy of dispersive layer-averaged approximations of Euler equations for free surface flows, *Commun. Math. Sci.* 16 (5) (2018) 1169–1202.
- [13] E. Kalnay, Atmospheric modeling, data assimilation and predictability, Cambridge: Cambridge University Press, 2003.
- [14] J. Blum, F.-X. L. Dimet, I. M. Navon, Data Assimilation for Geophysical Fluids, in: R. M. Temam, J. J. Tribbia (Eds.), Special Volume: Computational Methods for the Atmosphere and the Oceans, Vol. 14 of *Handb. Numer. Anal.*, Elsevier, 2009, pp. 385 – 441.
- [15] J. Lions, Optimal control of systems governed by partial differential equations, Springer Verlag, 1971.
- [16] J. A. Vrugt, C. G. H. Diks, H. V. Gupta, W. Bouten, J. M. Verstraten, Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.* 41 (1) (2005) 1–17.
- [17] K. Beven, A. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.* 6 (3) (1992) 279–298.
- [18] M. Thiemann, M. Trosset, H. Gupta, S. Sorooshian, Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.* 37 (10) (2001) 2521–2535.
- [19] J. A. Vrugt, W. Bouten, H. V. Gupta, S. Sorooshian, Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.* 38 (12) (2002) 48–1–48–13.
- [20] J. A. Vrugt, H. V. Gupta, W. Bouten, S. Sorooshian, A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.* 39 (8) (2003) 1–14.
- [21] P. O. Yapo, H. V. Gupta, S. Sorooshian, Multi-objective global optimization for hydrologic models, *J. Hydrol.* 204 (1) (1998) 83 – 97.



- [22] X. Yin, B. Wang, J. Liu, X. Tan, Evaluation of conditional non-linear optimal perturbation obtained by an ensemble-based approach using the Lorenz-63 model, *Tellus Ser. A-Dyn. Meteorol. Oceanol.* 66 (1) (2014) 22773.
- [23] S. Yuan, H. Zhang, M. Li, B. Mu, CNOP-P-based parameter sensitivity for double-gyre variation in ROMS with simulated annealing algorithm, *J. Oceanol. Limnol.* 37 (3) (2019) 957–967.
- [24] D. B. Haidvogel, H. G. Arango, K. Hedstrom, A. Beckmann, P. Malanotte-Rizzoli, A. F. Shchepetkin, Model evaluation experiments in the North Atlantic Basin: Simulations in nonlinear terrain-following coordinates, *Dyn. Atmos. Oceans* 32 (2000) 239–281.
- [25] M. Mu, W. Duan, Q. Wang, R. Zhang, An extension of conditional nonlinear optimal perturbation approach and its applications, *Nonlinear Process Geophys.* 17 (2) (2010) 211–220.
- [26] C. Sánchez-Linares, M. de la Asunción, M. Castro, S. Mishra, J. Šukys, Multi-level Monte Carlo finite volume method for shallow water equations with uncertain parameters applied to landslides-generated tsunamis, *Appl. Math. Model.* 39 (23) (2015) 7211–7226.
- [27] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by Simulated Annealing, *Science* 220 (1983) 671–680.
- [28] E. Aarts, P. van Laarhoven, Statistical cooling: A general approach to combinatorial optimization problems, *Philips J. Res.* 40 (1985) 193–226.
- [29] A. I. F. Vaz, L. N. Vicente, A particle swarm pattern search method for bound constrained global optimization, *Int. J. Comput. Math.* 39 (2) (2007) 197–219.
- [30] A. I. F. Vaz, L. N. Vicente, PSwarm: a hybrid solver for linearly constrained global derivative-free optimization, *Optim. Method Softw.* 24 (4–5) (2009) 669–685.
- [31] R. Storn, K. Price, Differential Evolution - A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, *J. Glob. Optim.* 11 (4) (1997) 341–359.
- [32] R. Hooke, T. A. Jeeves, “Direct Search” Solution of Numerical and Statistical Problems, *J. ACM* 8 (2) (1961) 212–229.
- [33] J. A. Nelder, R. Mead, A simplex method for function minimization, *Comput. J.* 7 (1965) 308–313.
- [34] R. Fletcher, C. M. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964) 149–154.

- [35] R. G. Polak, E., Note sur la convergence de méthodes de directions conjuguées, ESAIM-Math. Model. Numer. Anal.-Model. Math. Anal. Numer. 3 (R1) (1969) 35–43.
- [36] C. G. Broyden, The convergence of a class of double rank minimization algorithms: 2. The new algorithm, IMA J. Appl. Math. 6 (3) (1970) 222–231.
- [37] R. Fletcher, A new approach to variable metric algorithms, Comput. J. 13 (1970) 317–322.
- [38] D. Goldfarb, A family of variable metric methods derived by variational means, Math. Comput. 24 (109) (1970) 23–26.
- [39] D. F. Shanno, Conditioning of quasi-Newton methods for function minimization, Math. Comp. 24 (111) (1970) 647–650.
- [40] D. C. Liu, J. Nocedal, On the Limited Memory Method for Large Scale Optimization, Math. Program. 45 (1-3) (1989) 503–528.
- [41] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. 16 (5) (1995) 1190–1208.
- [42] D. Robertson, B. Brown, I. Navon, Determination of the Structure of Mixed Argon-Xenon Clusters Using a Finite-Temperature, Lattice-Based Monte-Carlo Method, J. Chem. Phys. 90 (1989) 3221–3229.
- [43] I. Navon, F. Brown, D. Robertson, A Combined Simulated-Annealing and Quasi-Newton-Like Conjugate Gradient Method for Determining The Structure of Mixed Argon-Xenon Clusters, Comput. Chem. 14 (1990) 305–311.
- [44] D. J. Wales, J. P. K. Doye, Global optimization by Basin-Hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 Atoms, J. Phys. Chem. A 101 (1997) 5111–5116.
- [45] R. E. Wengert, A simple automatic derivative evaluation program, Commun. ACM 7 (8) (1964) 463–464.
- [46] Automatic Differentiation Engine, TAPENADE, <https://www-sop.inria.fr/tropics/tapenade.html>, [Online; accessed 26-09-2019].
- [47] Y. Ding, Y. Jia, S. S. Y. Wang, Identification of Manning’s Roughness Coefficients in Shallow Water Flows, J. Hydraul. Eng.-ASCE 130 (6) (2004) 501–510.
- [48] E. Bélanger, A. Vincent, Data assimilation (4D-VAR) to forecast flood in shallow-waters with sediment erosion, J. Hydrol. 300 (1) (2005) 114–125.

- [49] X. Lai, J. Monnier, Assimilation of spatially distributed water levels into a shallow-water flood model. Part I: Mathematical method and test case, *J. Hydrol.* 377 (1) (2009) 1–11.
- [50] R. Hostache, X. Lai, J. Monnier, C. Puech, Assimilation of spatially distributed water levels into a shallow-water flood model. Part II: Use of a remote sensing image of Mosel River, *J. Hydrol.* 390 (3) (2010) 257–268.
- [51] M. Honnorat, J. Monnier, F.-X. Le Dimet, Lagrangian data assimilation for river hydraulics simulations, *Comput. Vis. Sci.* 12 (5) (2009) 235–246.
- [52] E. Bernard, V. Titov, Evolution of tsunami warning systems and products, *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* 373 (2053) 20140371.
- [53] Y. Wang, K. Satake, T. Maeda, A. R. Gusman, Data assimilation with dispersive tsunami model: a test for the Nankai Trough, *Earth Planets Space* 70 (131) (2018) 1–9.
- [54] Y. Wang, K. Satake, T. Maeda, A. R. Gusman, Green’s Function-Based Tsunami Data Assimilation: A Fast Data Assimilation Approach Toward Tsunami Early Warning, *Geophys. Res. Lett.* 44 (20) (2017) 10282–10289.
- [55] J. Li, D. Xiu, On numerical properties of the ensemble Kalman filter for data assimilation, *Comput. Meth. Appl. Mech. Eng.* 197 (43–44) (2008) 3574 – 3583.
- [56] A. Narayan, Y. Marzouk, D. Xiu, Sequential data assimilation with multiple models, *J. Comput. Phys.* 231 (19) (2012) 6401 – 6418.
- [57] Y. Yang, E. M. Dunham, G. Barnier, M. Almquist, Tsunami Wavefield Reconstruction and Forecasting Using the Ensemble Kalman Filter, *Geophys. Res. Lett.* 46 (2) (2019) 853–860.
- [58] T. Takagi, K. Inamoto, M. Kawahara, Estimation of Wave Propagation using a Kalman Filter, *Int. J. Comput. Fluid Dyn.* 9 (1) (1998) 77–84.
- [59] Y. Wang, T. Maeda, K. Satake, M. Heidarzadeh, H. Su, A. F. Sheehan, A. R. Gusman, Tsunami Data Assimilation Without a Dense Observation Network, *Geophys. Res. Lett.* 46 (4) (2019) 2045–2053.
- [60] A. R. Gusman, A. F. Sheehan, K. Satake, M. Heidarzadeh, I. E. Mulia, T. Maeda, Tsunami data assimilation of Cascadia seafloor pressure gauge records from the 2012 Haida Gwaii earthquake, *Geophys. Res. Lett.* 43 (9) (2016) 4189–4196.
- [61] Maëlle Nodet, Variational assimilation of lagrangian data in oceanography, *Inverse Probl.* 22 (1) (2006) 245–263.

- [62] H. Tsushima, K. Hirata, Y. Hayashi, Y. Tanioka, K. Kimura, S. Sakai, M. Shinohara, T. Kanazawa, R. Hino, K. Maeda, Near-field tsunami forecasting using offshore tsunami data from the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets and Space* 63 (7) (2011) 56.
- [63] H. Sumata, F. Kauker, R. Gerdes, C. Köberle, M. Karcher, A comparison between gradient descent and stochastic approaches for parameter optimization of a sea ice model, *Ocean Sci.* 9 (4) (2013) 609–630.
- [64] H. Sumata, F. Kauker, M. Karcher, R. Gerdes, Simultaneous Parameter Optimization of an Arctic Sea Ice-Ocean Model by a Genetic Algorithm, *Mon. Weather Rev.* 147 (6) (2019) 1899 – 1926.
- [65] A. M. Ferreiro, J. A. García-Rodríguez, J. G. López-Salas, C. Vázquez, An efficient implementation of parallel Simulated Annealing algorithm in GPUs, *J. Glob. Optim.* 57 (3) (2013) 863–890.
- [66] A. M. Ferreiro, J. A. García-Rodríguez, L. Souto, C. Vázquez, Basin Hopping with synched multi L-BFGS local searches. Parallel implementation in multi-CPU and GPUs, *Appl. Math. Comput.* 356 (2019) 282–298.
- [67] A. M. Ferreiro, J. A. García-Rodríguez, J. G. López-Salas, C. Vázquez, SABR/LIBOR market models: Pricing and calibration for some interest rate derivatives, *Appl. Math. Comput.* 242 (2014) 65 – 89.
- [68] A. Mangeney, F. Bouchut, N. Thomas, J. P. Vilotte, M. O. Bristeau, Numerical modeling of self-channeling granular flows and of their levee-channel deposits, *J. Geophys. Res.-Earth Surf.* 112 (F2) (2007) 1–21.
- [69] M. Pirulli, A. Mangeney, Results of Back-Analysis of the Propagation of Rock Avalanches as a Function of the Assumed Rheology, *Rock Mech. Rock Eng.* 41 (1) (2008) 59–84.
- [70] O. Pouliquen, Scaling laws in granular flows down rough inclined planes, *Phys. Fluids* 11 (3) (1999) 542–548.
- [71] M. Brunet, L. Moretti, A. Le Friant, A. Mangeney, E. D. Fernández Nieto, F. Bouchut, Numerical simulation of the 30–45 ka debris avalanche flow of Montagne Pelée volcano, Martinique: from volcano flank collapse to submarine emplacement, *Nat. Hazards* 87 (2) (2017) 1189–1222.
- [72] C. Escalante, T. Morales, M. Castro, Non-hydrostatic pressure shallow flows: GPU implementation using finite-volume and finite-difference scheme, *Appl. Math. Comput.* 338 (2018) 631–659.
- [73] C. Escalante, E. D. Fernández-Nieto, T. Morales de Luna, M. J. Castro, An Efficient Two-Layer Non-hydrostatic Approach for Dispersive Water Waves, *J. Sci. Comput.* 79 (1) (2019) 273–320.

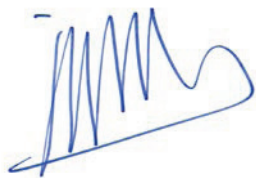
- [74] M. Castro Díaz, E. Fernández-Nieto, A Class of Computationally Fast First Order Finite Volume Solvers: PVM Methods, *SIAM J. Sci. Comput.* 34 (4) (2012) A2173–A2196.
- [75] J. Adsuaara, I. Cordero-Carrión, P. Cerdá-Durán, M. Aloy, Scheduled relaxation Jacobi method: Improvements and applications, *J. Comput. Phys.* 321 (2016) 369–413.
- [76] M. Locatelli, On the multilevel structure of global optimization problems, *Comput. Optim. Appl.* 30 (2005) 5–22.
- [77] M. Locatelli, F. Schoen, *Global Optimization: Theory, Algorithms, and Applications*, MOS-SIAM Series on optimization, SIAM, 2013.
- [78] B. Addis, M. Locatelli, F. Schoen, Local optima smoothing for global optimizations, *Optim. Method Softw.* 20 (2005) 417–437.
- [79] B. Addis, *Global optimization using local searches*, Ph.D. thesis, Università degli studi di Firenze (2004).
- [80] R. H. Leary, Global optimization on funneling landscapes, *J. Glob. Optim.* 18 (4) (2000) 367–383.
- [81] W. L. Goffe, SIMANN: A Global Optimization Algorithm using Simulated Annealing, *Stud. Nonlinear Dyn. Econom.* 1 (3) (1996) 1–9.
- [82] C. Zhu, R. H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-scale Bound-constrained Optimization, *ACM Trans. Math. Softw.* 23 (4) (1997) 550–560.
- [83] S. T. Grilli, M. Shelby, O. Kimmoun, G. Dupont, D. Nicolsky, G. Ma, J. T. Kirby, F. Shi, Modeling coastal tsunami hazard from submarine mass failures: effect of slide rheology, experimental validation, and case studies off the US East Coast, *Nat. Hazards* 86 (1) (2017) 353–391.
- [84] Landslide tsunami model benchmarking workshop, <http://www1.udel.edu/kirby/landslide/problems.html>, [Online; accessed 26-02-2019].
- [85] J. Macías, M. J. Castro, S. Ortega, C. Escalante, J. M. González-Vida, Performance Benchmarking of Tsunami-HySEA Model for NTHMP’s Inundation Mapping Activities, *Pure Appl. Geophys.* 174 (8) (2017) 3147–3183.
- [86] I. M. Navon, Practical and Theoretical Aspects of Adjoint Parameter Estimation and Identifiability in Meteorology and Oceanography, *Dyn. Atmos. Oceans* 27 (1 - 4) (1998) 55 – 79.
- [87] D. G. Cacuci, I. M. Navon, M. Ionescu-Bujor, *Computational Methods for Data Evaluation and Assimilation*, Chapman and Hall/CRC, 2013.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

The corresponding author J. A. García-Rodríguez on behalf of the authors of the article.

A handwritten signature in blue ink, consisting of a series of vertical strokes followed by a horizontal line and a small loop.

## Declaration of interest

April 24th, 2019

Editorial Department of Journal of Computational Physics

Dear Editor,

I confirm that the manuscript is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. We also have no conflicts of interest to disclose.

Sincerely,

José Antonio García Rodríguez  
Associate Professor  
University of A Coruña  
Department of Mathematics  
Faculty of Informatics  
E-mail: jose.garcia.rodriguez@udc.es