Abstract  Landslide run-out modeling is a powerful model-based decision support tool for landslide hazard assessment and mitigation. Most landslide run-out models contain parameters that cannot be directly measured but rely on back-analysis of past landslide events. As field data on past landslide events come with a certain measurement error, the community developed probabilistic calibration techniques. However, probabilistic parameter calibration of landslide run-out models is often hindered by high computational costs resulting from the long run time of a single simulation and the large number of required model runs. To address this computational challenge, this work proposes an efficient probabilistic parameter calibration method by integrating landslide run-out modeling, Bayesian inference, Gaussian process emulation, and active learning. Here, we present an extensive synthetic case study. The results show that our new method can reduce the number of necessary simulation runs from thousands to a few hundreds owing to Gaussian process emulation and active learning. It is therefore expected to advance the current practice of parameter calibration of landslide run-out models.

Keywords  Landslide run-out model · Bayesian inference · Gaussian process emulation · Active learning

Introduction

Landslides occur frequently in mountainous regions around the world, leading to many fatalities and economic losses. Process-based run-out modeling is a powerful tool to assess landslide hazards and risks, design mitigation strategies, and develop early warning systems. Due to the diverse and complex nature of landslides, semi-empirical landslide run-out models are more common than pure mechanistic counterparts (McDougall 2017). Examples are SHALTOP (Mangeney-Castelnau et al. 2003), TITAN2D (Pitman et al. 2003), DAN3D (Hungr and McDougall 2009), RAMMS (Christen et al. 2010), r avaflow (Mergili et al. 2017), etc. These models are on the one hand physics-based, namely based on mass and momentum balance of flow mass. On the other hand, they employ depth-averaging techniques and idealized rheological relationships instead of focusing on complex micromechanics of real landslides. Although approximate, such models have been proven to be able to simulate landslide bulk behavior to a satisfactory degree in controlled laboratory and field experiments (Savage and Hutter 1989; Mangeney-Castelnau et al. 2003; Medina et al. 2008; Hungr and McDougall 2009; George and Iverson 2014; Xia and Liang 2018), and in historic landslide events (Beguería et al. 2009; Christen et al. 2010; Mergili et al. 2017; Rauter et al. 2018; Xia and Liang 2018). Since rheological parameters in these semi-empirical models are rather conceptual than physical (Fischer et al. 2015), they are mostly calibrated by back-analyzing past landslide events.

Various calibration methods have been developed over past decades. They can be generally divided into two groups, namely deterministic and probabilistic methods. The aim of deterministic methods (classical inversion methods) is to find the best-fit parameter configuration that leads to simulation outputs as close as possible to observed data. It is traditionally done by subjective trial-and-error calibration, such as in Hungr and McDougall (2009); Frank et al. (2015); Schraml et al. (2015). More objective methods have recently been proposed by for example Calvello et al. (2017) and Aaron et al. (2019). They obtain the best-fit parameter configuration by minimizing the mismatch between simulation outputs and observed data using optimization theory. There are two issues with deterministic methods: first, different parameter configurations may lead to similar simulation outputs, known as the non-uniqueness or equifinality problem (McMillan and Clark 2009; Aaron et al. 2019); second, deterministic methods cannot account for measurement uncertainties.

Probabilistic methods, which avoid these two issues, aim to update prior knowledge of the calibration parameters to a posterior distribution based on observed data. They commonly require evaluating a run-out model at a large number of parameter configurations and depend on certain updating/selection rules. For example, Fischer et al. (2015) ran a depth-averaged flow model at 10,000 rheological parameter points from a Latin hypercube design, obtained reduced parameter combinations based on a user-defined selection rule, and approximated posteriors of rheological parameters using a frequency analysis. Brezzi et al. (2016) obtained posteriors of rheological parameters by running 2000 simulations based on a Monte Carlo design and by applying the Kalman filter. Moretti et al. (2020) and Heredia et al. (2020) approximated posteriors using 8000 and 50,000 Markov chain Monte Carlo (MCMC) iterations within the Bayesian inference framework respectively. Aaron et al. (2019) approximated posteriors of rheological parameters using a full grid search within the Bayesian inference framework (Aaron 2017). The main shortcoming of probabilistic methods is the high computational costs resulting from the large number of required simulation runs, as pointed out by many researchers (Fischer et al. 2015; Brezzi et al. 2016; Aaron 2017; Heredia et al. 2020).

As a well-established surrogate modeling technique for reducing computational costs, Gaussian process (GP) emulation has been extensively used for parameter calibration in the past decades. One type of GP emulation-based strategy, pioneered by Kennedy and O’Hagan (2001), emulates input-output relations of the simulation model; see Bayarri et al. (2007); Higdon et al. (2008); Gu and Wang (2018). This type of method may need to emulate a simulation model.
with high-dimensional outputs if observed data are high-dimensional. An alternative type of GP emulation-based technique directly emulates the loss/likelihood function that measures the mismatch between simulation outputs and observed data. It avoids the high-dimensional problem since the loss/likelihood function only has a scalar output. Examples are Oakley and Youngman (2017); Fer et al. (2018); Wang and Li (2018); Järvenpää et al. (2021). As for parameter calibration of landslide run-out models, Sun et al. (2021) built GP emulator for the landslide run-out model Massflow focusing on a scalar output (the run-out distance), and used the emulator for Bayesian inference of the model parameters. Navarro et al. (2018) used another surrogate modeling method, the polynomial chaos expansion, to approximate the landslide run-out model D-Claw, and used the surrogate for Bayesian inference of the model parameters. To our knowledge, no attempt has been made to directly emulate the loss/likelihood function regarding parameter calibration of landslide run-out models.

Another powerful technique to reduce computational costs is active learning. It has been recently used to improve inference quality. The essential idea is to iteratively run simulation at a new parameter point guided by all previous runs in order to increase our knowledge of the posterior the most (Cranmer et al. 2020). Various rules have been proposed regarding how a new parameter point may be selected. For instance, Zhang et al. (2016) adaptively constructed separate GP emulators for each model output, approximated the posterior using the MCMC method, and picked new parameter points by sampling from the approximated posterior. Kandasamy et al. (2017); Wang and Li (2018); Järvenpää et al. (2021) sequentially constructed GP emulators for certain variations of the likelihood function (log-likelihood, log-unnormalized-posterior, etc.), and chose new parameter points that reduce the posterior uncertainty the most. To date, no active learning technique has been employed concerning parameter calibration of landslide run-out models.

As pointed out in the recent review by Cranmer et al. (2020), the rapidly advancing frontier of simulation-based inference driven by machine learning (here GP emulation), active learning, and a few other factors is expected to profoundly impact many domains of science. Therefore, the main goal of our study is to develop an efficient parameter calibration method for landslide run-out models by leveraging Bayesian inference, GP emulation, and active learning. We present the new method in detail in "Methodology" and illustrate its efficiency using a synthetic case study based on the 2017 Bondo landslide event. The case design is given in "Case study" and the results and discussions are presented in "Results and discussions". In "Conclusions", the main conclusions are presented.

Methodology

Depth-averaged landslide run-out model

The governing system of depth-averaged landslide run-out models is derived from classical conservation laws of mass and momentum using continuum mechanics. It can be expressed in a surface-induced coordinate system {\( T_s, T_r, T_n \)} (Fischer et al. 2012) as

\[
\begin{align}
\partial_t h + \partial_x (hu_x) + \partial_y (hu_y) &= Q(x, y, t), \\
\partial_t (hu_x) + \partial_x (hu_x^2 + g_a k_{a/p} \frac{h^2}{2}) + \partial_y (hu_x u_y) &= g_x h - S_i, \\
\partial_t (hu_y) + \partial_x (hu_x u_y) + \partial_y (hu_y^2 + g_a k_{a/p} \frac{h^2}{2}) &= g_y h - S_y.
\end{align}
\]

Equation 1 represents the mass balance and Eqs. 2–3 represent the momentum balance in the surface tangent directions \( T_s \) and \( T_r \). They describe the time evolution of the state variables, namely flow height \( h \) and surface tangential flow velocity \( u = (u_x, u_y)^T \). The mass production source term \( Q(x, y, t) \) accounts for the entrainment process (Christen et al. 2010). Variables \( g_x, g_a, g_y, g_r \) are components of the gravitational acceleration in the \( T_s, T_r, T_n \) directions respectively. The variable \( k_{a/p} \) denotes the active/passive earth pressure coefficient, which was introduced by Savage and Hutter (1989) to account for the elongation/compression of the flow material in the surface tangent directions. The friction terms \( S_i \) and \( S_y \) describe the basal friction rheology. Various rheological models have been proposed and used in current practice, such as frictional, Voellmy, Bingham, and Poulilchen; see Naef et al. (2006); Pirulli and Mangeney (2008); Hungr and McDougall (2009) for an overview.

The entrainment process can greatly affect landslide flow behavior in highly erosive landslide events. We do not take it into account in this study and keep the synthetic case ("Case study") as simple as possible in order to keep the focus on the illustration of the new Bayesian active learning method. The new method, however, can be extended to landslide run-out models that incorporate entrainment processes by substituting the numerical core landslide run-out model in the workflow as shown in Fig. 1. Also, we demonstrate the new parameter calibration method by focusing on the Voellmy rheological model, which has been widely used in terms of flow-like landslides (Frank et al. 2015; Schraml et al. 2015; McDougall 2017). It should be noted that the new method can however be extended to account for any other rheological model without loss of generality. More specifically, parameters of any other rheological model can be calibrated using the new method by simply replacing the landslide run-out model in the workflow as shown in Fig. 1. The Voellmy rheological model is defined as

\[
S_i = \frac{u_i}{||u||} (\mu g_a h + \frac{g_r}{\xi} ||u||^2), \quad i \in \{x, y\},
\]

where \( \mu \) and \( \xi \) denote the dry-Coulomb friction coefficient and turbulent friction coefficient, respectively, and \( g \) denotes the norm of the gravitational acceleration. As mentioned in the introduction, parameters of the idealized rheological models are rather conceptual than physical. They often have to be calibrated by back-analyzing real landslide events where field observation data are available (Hungr 2016). It is typically not possible to measure them in the lab, except for the parameter in the frictional rheology which can be obtained by for example quasi-dynamic tilting tests (Manzella 2008; Hungr 2009).

Bayesian inference framework

Bayesian inference of rheological parameters aims to derive a posterior probability distribution of the rheological parameters in light of
observed data. The posterior distribution is computed using Bayes’ theorem, namely

$$\pi(x | d) = \frac{L(x | d)\pi(x)}{\int L(x | d)\pi(x)dx}. \quad (5)$$

Here, $x$ denotes the rheological parameters, namely $x = (\mu, \xi)^T \in \mathcal{X} \subset \mathbb{R}^2$ in the case of the Voellmy rheological model. The vector $d = (d_1, \ldots, d_k)^T \in \mathbb{R}^k$ denotes observed data, such as the impact area, deposit area and volume, and flow height and velocity at certain locations. $\pi(x)$ denotes the prior probability distribution of the rheological parameters. It encodes prior knowledge about the rheological parameters before even knowing the observed data. $L(x | d)$ is known as the likelihood. It is a function of $x$ which measures how likely $x$ takes certain values given the observed data $d$. In the context of parameter calibration of a landslide run-out model, the likelihood function $L(x | d)$ involves the observed data $d$ and corresponding landslide run-out model outputs $y = (y_1, \ldots, y_k)^T = f(x)$. The exact form of the likelihood function relies on the statistical ansatz used to model the residuals $\epsilon = (\epsilon_1, \ldots, \epsilon_k)^T$, where $\epsilon_i = d_i - y_i, i \in \{1, \ldots, k\}$. Commonly the residuals are assumed to fulfill a $k$-variate Gaussian distribution with zero mean and a $k \times k$ covariance matrix $\Sigma$, which leads to

$$L(x | d) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(d - y)^T \Sigma^{-1}(d - y) \right\}. \quad (6)$$

The residuals $\epsilon_i, i \in \{1, \ldots, k\}$ are furthermore often assumed to be independent, for example in Navarro et al. (2018); Aaron et al. (2019); Moretti et al. (2020). In that case, the covariance matrix $\Sigma$ reduces to a diagonal matrix

$$\Sigma = \begin{pmatrix} \sigma_{\epsilon_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\epsilon_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{\epsilon_k}^2 \end{pmatrix}. \quad (7)$$

where $\sigma_{\epsilon_i}$ denotes the standard deviation of the residual $\epsilon_i, i \in \{1, \ldots, k\}$. The diagonal covariance matrix is used in this study. The proposed method can however be easily extended to account for correlations without loss of generality.

The posterior $\pi(x | d)$ in Eq. 5 cannot be computed in a closed form since a complex landslide run-out model is involved in the likelihood function. Various methods exist for approximating the posterior; see Gelman et al. (2013) for an overview. One type of method approximates the integral (the denominator in Eq. 5) using numerical integration. The term $L(x \mid d)\pi(x)$ needs to be evaluated at a large number of points $x_i, i \in \{1, \ldots, N\}$, which can be for example a set of evenly spaced full grid points. The posterior is then approximated by

$$\pi(x | d) \approx \frac{L(x | d)\pi(x)}{\sum_{i=1}^N L(x_i | d)\pi(x_i)\Delta x_i}. \quad (8)$$

This type of method works well when $x$ is low-dimensional, like two- or three-dimensional. When $x$ is high-dimensional, MCMC methods are often used to draw a set of samples $x_i, i \in \{1, \ldots, N\}$ from the unnormalized posterior.

---

**Fig. 1** A schematic illustration of the emulator-based Bayesian active learning method.
\[ \pi(x | d) \propto L(x | d)\pi(x). \] (9)

The posterior can then be estimated based on the MCMC samples \( x_i, i \in \{1, \ldots, N\} \) by for example kernel density estimation.

No matter for which method, the unnormalized posterior \( L(x | d) \pi(x) \) needs to be evaluated at a large number of input points \( x_i, i \in \{1, \ldots, N\} \) in order to approximate the posterior with reasonable accuracy. Each evaluation requires running the landslide run-out model at \( x_i \). The computational cost can be prohibitively high if a single run takes a relatively long time. In that case, it is promising to overcome the computational challenge by building a cheap-to-evaluate Gaussian process emulator to replace the expensive-to-evaluate unnormalized posterior.

**Gaussian process emulation**

Let \( z \) denote the logarithm of the unnormalized posterior, namely
\[ z = \ln[L(x | d)\pi(x)] = g(x). \] (10)

Note that the logarithmic form is commonly used to avoid computational overflows and underflows (Gelman et al. 2013). The function \( g(\cdot) \) represents a mapping from a \( p \)-dimensional parameter input space \( \mathcal{X} \subset \mathbb{R}^p \) to \( \mathbb{R} \). In terms of calibrating the two Voellmy rheological parameters, \( p \) equals 2. Since the function value at any \( x \) is unknown before running the landslide run-out model at \( x \), we can treat the function as an unknown function and model it by a Gaussian process, namely
\[ g(\cdot) \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot)). \] (11)

The Gaussian process is fully defined by its mean function \( m(\cdot) : \mathcal{X} \rightarrow \mathbb{R} \) and covariance kernel \( k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \). Conditioned on a set of training data \( \{x_i, z_i\}_{i=1}^n \), the function value \( z^* \) at any untried input point \( x^* \) follows a Gaussian distribution:
\[ z^* \sim \mathcal{N}(m(x^*), k(x^*, x^*)), \] (12)

\[ m(x^*) = m(x^*) + k^T(x^*)K^{-1}(z_1 - m(x_1), \ldots, z_n - m(x_n))^T, \] (13)

\[ k(x^*, x^*) = k(x^*, x^*) - k^T(x^*)K^{-1}k(x^*), \] (14)

where \( k(x^*) = [k(x^*, x_1), \ldots, k(x^*, x_n)]^T \) and \( K \) denotes the \( n \times n \) covariance matrix of which the \( (i, j) \)-th entry \( K_{ij} = k(x_i, x_j) \). The training input points \( \{x_i\}_{i=1}^n \) (at which the landslide run-out model is evaluated) are carefully selected as we will shortly see in "Active learning" or "Workflow". Each training data \( z_i \) is obtained by first running the landslide run-out model at the training input point \( x_i \) and then computing corresponding log-unnormalized posterior value according to Eq. 10. Equations 12–14 define the Gaussian process emulator, denoted as \( \hat{g}(x) \). It provides a prediction of the log-unnormalized posterior value at any untried input point \( x^* \) in the form of Eq. 13, together with an assessment of the prediction uncertainty in the form of Eq. 14.

Once the GP emulator \( \hat{g}(x) \) is built, we can approximate the computationally expensive term \( \exp(g(x)) \) in Eq. 5 by \( \exp(\hat{g}(x)) \), which results in

\[ \pi(x | d) \approx \frac{\exp(\hat{g}(x))}{\int_x \exp(\hat{g}(x))dx}. \] (15)

The posterior is then estimated by applying the grid approximation or MCMC methods to Eq. 15. It reduces the number of required landslide run-out model evaluations from \( N \) (usually thousands) to \( n \) (a few hundreds), and therefore greatly improves the computational efficiency.

**Active learning**

The design of training data \( \{x_i, z_i\}_{i=1}^n \) for emulator-based Bayesian inference deserves particular attention. Due to information gained from the observed data, the posterior is often localized in a small portion of the input space where the likelihood has large values, and is close to zero elsewhere. Since the aim here is to estimate the posterior reasonably well with limited computational budget, the commonly used space-filling sampling techniques, such as the Latin hypercube design, are not efficient to determine training input points for the GP emulator \( \hat{g}(x) \). More specifically, many input points from a space-filling sampling scheme will be located in areas where the posterior value is close to zero and therefore do not provide much information on the posterior that we want to estimate.

Active learning, also known as sequential design, is a simple but very impactful idea to wisely choose training input points at which the landslide run-out model needs to be run (Cranmer et al. 2020). Instead of selecting all the training input points a priori like in a space-filling sampling scheme, active learning iteratively chooses new training input point that is expected to increase our knowledge on the posterior the most. The selection of each new training input point is guided by all previous simulation runs. Assume that \( b \) input points \( \{x_i\}_{i=1}^b \) have been chosen and \( \{z_i\}_{i=1}^b \) have been correspondingly computed based on the \( b \)-landslide run-out model evaluations and the observed data. Given \( \{x_i, z_i\}_{i=1}^b \), a GP emulator \( \hat{g}_b(x) \) can be built according to "Gaussian process emulation". It provides an approximation to the log-unnormalized posterior. The exponential term \( \exp(\hat{g}_b(x)) \) hence provides an approximation to the unnormalized posterior. It encodes our current knowledge about the posterior and can therefore be used to determine the next input point that is expected to provide the most information on the posterior.

A widely used strategy for active learning is to pick the parameter input point at which the approximate unnormalized posterior \( \exp(\hat{g}_b(x)) \) has the largest uncertainty. By running a new simulation at that parameter configuration, this uncertainty will be eliminated and the information gain on the posterior is expected to be the most. Many uncertainty indicators have been proposed in the literature, such as the variance (Kandasamy et al. 2017) or entropy (Wang and Li 2018) of the approximate unnormalized posterior \( \exp(\hat{g}_b(x)) \). In this study, we follow Wang and Li (2018). As presented in "Gaussian process emulation", \( \hat{g}_b(x) \) follows a Gaussian (normal) distribution at any untried input point \( x^* \), with a mean function \( m^*(x^*) \) (Eq. 13) and a kernel function \( K^*(x^*, x^*) \) (Eq. 14). The term \( \exp(\hat{g}_b(x^*)) \) therefore follows a log-normal distribution. The entropy \( H_b(x^*) \) of the log-normal distribution can be analytically computed as follows:
where $e$ is the Euler’s number. The optimal input point $x_{b+1}$ for the next simulation run is given by the input point that maximizes Eq. 16.

**Workflow**

Algorithm 1 presents the workflow of the proposed method for parameter calibration of landslide run-out models. A schematic illustration is given by Fig. 1. The method combines landslide run-out modeling, Bayesian inference, GP emulation, and active learning. An initial GP emulator for the log-unnormalized posterior is first built based on initial $b_0$ simulation runs. Then, the active learning strategy is employed to adaptively pick new input points, run simulations, and update training data. Last, the posterior probability distribution of the rheological parameters is estimated based on the final GP emulator using grid approximation or MCMC methods.

**Algorithm 1 Emulator-based Bayesian inference with active learning**

1: Define the number of total training input points $n$ and the number of initial training input points $b_0$.
2: Sample initial input points $\{x_i\}_{i=1}^{b_0}$ using a Latin hypercube sampling scheme.
3: Run the landslide run-out model at each $x_i$ and calculate $z_i$ based on the observed data and corresponding simulation outputs. Denote the training data as $D = \{x_i, z_i\}_{i=1}^{b_0}$
4: for $b = b_0, \ldots, (n - 1)$ do
5: Build the GP emulator $\hat{g}_b(x)$ based on $D$.
6: Determine the next input point $x_{b+1}$ by maximizing equation 16.
7: Run the landslide run-out model at $x_{b+1}$ and calculate $z_{b+1}$.
8: Augment training data $D = D \cup \{x_{b+1}, z_{b+1}\}$.
9: end for
10: Build the final GP emulator $\hat{g}_n(x)$ based on $D = \{x_i, z_i\}_{i=1}^{n}$.
11: Substitute $\hat{g}_n(x)$ into equation 15 and estimate the posterior using grid approximation or MCMC methods.

**Case study**

**Types of observed data**

Carrying out a parameter calibration requires the availability of observed data. Various types of observed data have been used in the literature, including the run-out distance (Sun et al. 2021), impact area, deposit distribution, deposit depth at specific locations, maximum flow velocity at specific locations (Aaron et al. 2019), time history of flow height at specific locations (Navarro et al. 2018), time history of the force exerted by flow mass onto the ground (Moretti et al. 2020), time history of flow velocity at the centre of flow mass (Heredia et al. 2020), etc. Among them, static data like the run-out distance, impact area, deposit distribution, and deposit height at specific locations can be obtained from pre- and post-event landscapes based on remote sensing, or post-event field investigation. The maximum flow velocity at certain locations can be obtained from for example post-event super-elevation measurements using the forced vortex equation (Prochaska et al. 2008; Scheidl et al. 2008).

**Fig. 2** Topography and release mass distribution of the 2017 Bondo landslide event. The impact area and deposit distribution are simulation results with Voellmy rheological parameters $\mu = 0.23$ and $\zeta = 1000 \text{m/s}^2$. The impact area and deposit area are determined using a cutoff value of 0.5 m.
The release volume is around 3 million m$^3$ of the event, as shown in Fig. 2, are used to generate synthetic data. Information on the topography and distribution of the release mass which has been introduced in detail by Mergili et al. (2020); Walter et al. (2020). The synthetic case is based on the 2017 Bondo landslide event, as shown in Fig. 2. Corresponding simulation results are extracted, are noted in Fig. 2. The time history of the force exerted by the flow mass onto the ground can be obtained from long-period seismic signals (Moretti et al. 2020). More specifically, a landslide event may generate long-period seismic waves as the flow mass moves down the topography, which can be recorded by seismic stations. By performing a waveform inversion using the seismic signals, the time history of the force can be obtained (Moretti et al. 2015).

While the methodology proposed in "Methodology" can be applied to any type of observed data and their combinations, the focus of this case study is put on parameter calibration using static data that are mostly available for real-world landslide events. More specifically, the impact area, deposit volume, deposit height at specific locations, and maximum flow velocity at specific locations will be used to calibrate the Voellmy rheological parameters based on a synthetic case. Here, the impact area is defined as the area where the maximum flow height is larger than a threshold value (0.5 m; see Fig. 2).

### Synthetic data generation

The synthetic case is based on the 2017 Bondo landslide event, which has been introduced in detail by Mergili et al. (2020); Walter et al. (2020). The topography and distribution of the release mass of the event, as shown in Fig. 2, are used to generate synthetic data in this study. The release volume is around 3 million m$^3$.

It should be noted that the purpose of the case study is not to calibrate the 2017 Bondo landslide event. Instead, the intention here is to test the proposed methodology and to study the impact of different types of observed data on calibration results. To this end, synthetic observed data are derived from given rheological parameters. Using these instead of field observations of the 2017 landslide event allows us to test the feasibility of the calibration in detail. The given rheological parameters serve as underlying truth which allows us to evaluate the proposed methodology.

The detailed procedure for generating synthetic observed data is as follows: First, we set the ranges of the Voellmy rheological parameters as $\mu \in [0.02, 0.3]$ and $\xi \in [100, 2200]$ m/s$^2$ following the typical ranges for rock avalanches, which are summarized by Zhao et al. (2021) based on a literature study. Then, a set of Voellmy rheological parameters, $\mu = 0.23$ and $\xi = 1000$ m/s$^2$, are picked from the ranges. Next, the landslide run-out model is run given the rheological parameters, topography, and distribution of release mass. For the landslide run-out model, we use the software r.avalflow developed by Mergili et al. (2017). The static data mentioned in "Types of observed data" are obtained by post-processing the simulation outputs. The impact area, deposit distribution, and locations where deposit height and/or maximum flow velocity are extracted, are noted in Fig. 2. Corresponding simulation results are summarized in Table 1. Last, synthetic observed data are generated by adding random noises to the simulation results. For each synthetic observed data (each row in Table 1 except the last two rows), the random noise is drawn from a Gaussian distribution with zero mean and an assumed standard deviation that equals 10% of the simulation result. Two additional synthetic observed data for the

### Table 1: Simulation results ($\mu = 0.23, \xi = 1000$ m/s$^2$) and synthetic observed data

| Quantity of interest          | Simulation result | Assumed standard deviation ($\sigma_\mu$) | Random noise | Synthetic observed data |
|------------------------------|-------------------|-----------------------------------------|--------------|------------------------|
| Maximum velocity at L1 (m/s) | 29.40             | 2.94                                    | 1.31         | 30.71                  |
| Maximum velocity at L3 (m/s) | 12.33             | 1.23                                    | 0.39         | 12.72                  |
| Deposit height at L2 (m)     | 5.53              | 0.55                                    | 0.37         | 5.91                   |
| Deposit height at L3 (m)     | 9.79              | 0.98                                    | 0.99         | 10.78                  |
| Impact area (million m$^2$)  | 2.48              | 0.25                                    | 0.12         | 2.59                   |
| Deposit volume (million m$^3$)| 2.80              | 0.28                                    | −0.09        | 2.71                   |
| Maximum velocity at L1 (m/s) | 29.40             | 5.88                                    | 2.62         | 32.03                  |
| Maximum velocity at L1 (m/s) | 29.40             | 8.82                                    | 3.94         | 33.34                  |

### Table 2: Cases for rheological parameter calibration

| Number | Used synthetic observed data                                   |
|--------|----------------------------------------------------------------|
| 1      | Maximum velocity at L1 ($\sigma_\mu = 2.94$)                   |
| 2      | Maximum velocity at L3                                         |
| 3      | Deposit height at L2                                           |
| 4      | Deposit height at L3                                           |
| 5      | Impact area                                                   |
| 6      | Deposit volume                                                |
| 7      | Maximum velocity at L1 ($\sigma_\mu = 5.88$)                   |
| 8      | Maximum velocity at L1 ($\sigma_\mu = 8.82$)                   |
| 9      | Maximum velocity at L1 ($\sigma_\mu = 2.94$) + deposit volume |
| 10     | Maximum velocity at L3 + deposit height at L3                 |
maximum flow velocity at L1 are generated using assumed standard deviations that equal to 20% and 30% of the simulation result respectively, as shown in the last two rows of Table 1. They are used to investigate the impact of the uncertainty of observed data.

For future parameter calibration using real-world observed data, the standard deviations \(\sigma_i\), \(i \in \{1, \ldots, k\}\); Eq. 7 can be determined by heuristics when there are only coarse observed data such as the impact area, point estimates of deposit depth, and point estimates of flow velocity in Aaron et al. (2019), or can be treated as additional calibration parameters in the Bayesian inference framework when there are rich observed data such as time history of flow velocity at the center of flow mass in Heredia et al. (2020).

**Parameter calibration setup**

The prior of the Voellmy rheological parameters is assumed to be a uniform distribution over the rectangular space defined by \(\mu \in [0.02, 0.3]\) and \(\xi \in [100, 2200] \text{ m/s}^2\). It means that no information about the rheological parameters, except the limiting values, is known before observing any data. This kind of prior is often used in the literature, such as Navarro et al. (2018); Aaron et al. (2019); Moretti et al. (2020). More informative prior, such as the Gamma distribution, could also be used when expert knowledge and well-known reference values are available (Heredia et al. 2020). A posterior obtained from previous calibration can also be used as the prior for a new calibration task when new observed data are available.

Given the synthetic observed data generated in "Synthetic data generation" (Table 1), the following cases are set up, as summarized in Table 2. Note that L1, L2, and L3 refer to point locations as shown in Fig. 2. In cases 1–6, the Voellmy rheological parameters \(\mu\) and \(\xi\) are calibrated based on a single synthetic observed data using the proposed methodology. They are used to study the impact of different types of observed data on calibration results. The impact of the standard deviation \(\sigma_i\) is investigated based on cases 1, 7, and 8. Cases 9 and 10 are designed to investigate the impact of combinations of observed data.

The number of initial training input points \(b_0\) (step 1 in Algorithm 1) is set as \(40\). Initial training input points \((\mu_i, \xi_i)^T\) \(i = 1, \ldots, 40\) are chosen using the maximin Latin hypercube design. All the cases share the same 40 initial training input points, meaning that the initial 40 simulation runs only need to be conducted once. The number of total training input points \(n\) is set to be 120, meaning that 80 adaptive simulation runs are used to actively learn the respective posterior in each case. It should be noted that the 80 adaptive simulation runs are distinct for each case since they are tailored to a specific observed dataset in each case by active learning. It leads to a total 80 adaptive simulation runs for all the cases. For each case, the posterior is computed using grid approximation based on a \(100 \times 100\) grid (step 11 of Algorithm 1).

**Results and discussions**

**Active learning process and the impact of \(\sigma_i\)**

Figure 3a–d show the estimated posteriors at four different iteration steps in case 1 (see Table 2). They provide a direct impression on how the active learning works and the impact of the number of iteration steps on the calibration result. For each given iteration number \(i\), a GP emulator \(\mathcal{g}_{\text{iso}}(x)\) is built based on the simulation runs up to the given iteration, namely 40 initial runs and \(i\) adaptive runs. Then the posterior at that iteration number can be computed by substituting \(\mathcal{g}_{\text{iso}}(x)\) into Eq. 15 and applying grid approximation. From Fig. 3a–d, the following observations can be made:

- The quality of the estimated posterior is quite poor without any adaptive simulation run as expected. It gets gradually improved with increasing iteration number, owing to the information gain from simulation runs at adaptively chosen input points.
- The underlying truth of the rheological parameters locates very close to the high-probability regions of the final posterior (Fig. 3d). It implies that the proposed methodology is capable of correctly calibrating the rheological parameters.
- The final posterior has high values not only in regions near the underlying truth, but also in regions far from the underlying truth. It highlights the non-uniqueness or equifinality problem associated with deterministic calibration methods as mentioned in the introduction. Namely, different configurations of the rheological parameters may lead to similar simulation outputs. Probabilistic calibration methods should be therefore used whenever possible.
- The final posterior occupies only a small portion of the parameter space and has values close to zero elsewhere. In this case, the active learning scheme which adaptively determines input points has great advantage, since it allocates more computational resources for exploring the high-probability regions. In other words, the active learning scheme can provide better approximation of the posterior than a pure space-filling design scheme with the same computational budget.

Figure 3c, f show the final posteriors of case 7 and case 8 (based on synthetic maximum velocity at location L1 with \(\sigma_1 = 5.88\) and \(\sigma_1 = 8.82\) respectively). Comparing them with Fig. 3d, it can be seen that the underlying truth of the rheological parameters is still close to the high-probability regions, but the shape of the posterior becomes flat with the increase of the standard deviation \(\sigma_i\). This result is expected and is similar to the findings of Aaron et al. (2019); Sun et al. (2021). More specifically, increasing \(\sigma_i\) means increasing uncertainty of the observed data. The information gained from the observed data (encoded in the likelihood function) therefore decreases with increasing \(\sigma_i\). The posterior accordingly relies more on the prior information, here a uniform distribution.

In order to investigate the convergence behavior of the active learning scheme, the change of the total variation distance with respect to the number of iterations for cases 1, 7, and 8 is plotted in Fig. 4. The total variation distance measures the difference between two probability density distributions \(p(x)\) and \(p'(x)\), and is defined as (Järvenpää et al. 2021)

\[
TV(p, p') = \frac{1}{2} \int |p(x) - p'(x)| dx.
\]  

In each case, the total variation distance \(TV(p_i, p_{i+10})\) is iteratively calculated after each 10 adaptive runs, where \(p_i\) and \(p_{i+10}\) denote
the estimated posteriors based on simulation runs up to the $i$-th and $(i-10)$-th iterations respectively. It can be seen from Fig. 4 that $\text{TV}(p_i, p_{i-10})$ for all three cases generally decreases with increasing number of adaptive runs and remains at a relatively low value after a certain number of adaptive runs. It implies that the estimated posterior for each case reaches a stable stage. It should be noted that the total variation distance, or other quantities that measure the difference between two probability distributions like the Kullback-Leider
divergence, can be used to design early stopping criteria for Algorithm 1; see for example Wang and Li (2018).

Based on above results, we can conclude that the proposed emulator-based Bayesian active learning method is able to correctly calibrate the rheological parameters. Compared to commonly used probabilistic methods without emulation techniques as mentioned in the introduction, the proposed method greatly improves the computational efficiency by reducing the number of necessary simulation runs from thousands (even tens of thousands) to a few hundreds. Compared to emulator-based Bayesian inference without active learning, the proposed method can provide better approximation of the posterior if the computational budget is the same by wisely allocating computational resources.

Different observed data and their combinations

Figure 5a–f show the final posteriors for cases 1–6 respectively. The underlying truth of the rheological parameters and 80 adaptively chosen input points are also plotted in the figures. The following observations can be made:

- The underlying truth of the rheological parameters closely locates near the high-probability regions, no matter which synthetic observed data is used for parameter calibration. It further validates that the proposed method is able to correctly calibrate the rheological parameters.
- The posteriors obtained based on different synthetic observed data significantly differ from one another. It means that the information on the rheological parameters gained from different types of observed data can be greatly different.
- Location-wise observed data, such as maximum velocity and deposit height at specified locations, can better constrain the rheological parameters than aggregated overall observed data, like the impact area and deposit volume.
- A single observed data alone is not enough to constrain the rheological parameters. It implies that different types of observed data should be combined in order to effectively calibrate the rheological parameters.

Based on the results shown in Fig. 5a–f, it can be presumed that in our case the rheological parameters can be better constrained if the calibration relies on a combination of complementary observed data, such as maximum velocity at L1 and deposit volume or maximum velocity at L3 and deposit height at L3. In order to validate the presumption, the rheological parameters are calibrated using the combination of maximum velocity at L1 and deposit volume, and the combination of maximum velocity and deposit height at L3, respectively. The results are shown in Fig. 6a–b. It can be seen that the resulted posteriors are better constrained in each case as expected. In the future, such an analysis can serve as a starting point to optimize data acquisition, e.g., which sensors to use and where to place them, such that we can maximize the knowledge return of field observations.

In regard to using the calibrated parameters in the prediction of future landslide events, case-specific calibrated results have limited use, as pointed out by McDougall (2017). Instead, calibrated results of similar landslide events should be combined. Here, the similarity can relate to for example types of landslide and path material. The calibrated result of a single landslide event using the Bayesian active learning method is given as a probability density function of the rheological parameters, namely the posterior distribution. Combining calibrated results of a group of landslide events means combining a group of probability density functions of the rheological parameters. This can be done by for example the modified kernel density estimation which results in an overall probability density function of the rheological parameters; see Aaron (2017) for more detail. The combined probability density function could subsequently be used for probabilistic prediction of a future landslide event which shares similarities to the group of landslide events, by for example Monte Carlo methods. Risk assessment and mitigation design can be accordingly conducted based on results of the probabilistic prediction, such as a probabilistic hazard map.
Fig. 5 (a)–(f) estimated posteriors at 80 adaptive simulation runs for cases 1–6 (see Table 2). In each panel, the black cross shows the underlying true values of $\mu$ and $\xi$; the black circles denote the 40 initial training input points; the red diamonds represent the 80 input points that are adaptively determined by active learning; the color map shows the posterior of the rheological parameters which is estimated based on the 40 initial and 80 adaptive simulation runs.
Conclusions

Probabilistic parameter calibration of landslide run-out models is challenging due to the long run time of a single simulation and the large number of required simulations. We have proposed an efficient probabilistic parameter calibration method to address this computational challenge in this paper. The new method is developed by integrating landslide run-out modeling, Bayesian inference, Gaussian process emulation, and active learning. We have demonstrated its feasibility and efficiency based on a synthetic case study that allowed us to test the quality of the calibration against ground truth data. The new method can reduce the number of necessary simulation runs from thousands (even tens of thousands) to a few hundreds. It is therefore expected to advance the current practice of parameter calibration of landslide run-out models.

The impact of different types of observed data is also studied based on the case study using the proposed method. It is found that the information gained from different types of observed data can greatly differ from one another. Location-wise observed data like maximum velocity and deposit height at specific locations provide better constraint for the Voellmy rheological parameters than aggregated overall observed data like the impact area and deposit volume. In addition, a single observed data alone cannot effectively constrain the Voellmy rheological parameters. Different types of observed data should be combined in order to improve the quality of the posterior.

Acknowledgements

The authors greatly thank for the support of the China Scholarship Council (grant number: 201706260262) and the Helmholtz Graduate School for Data Science in Life, Earth and Energy.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aaron J (2017) Advancement and calibration of a 3D numerical model for landslide runout analysis. PhD thesis, The University of British Columbia, Vancouver

Aaron J, McDougall S, Nolde N (2019) Two methodologies to calibrate landslide runout models. Landslides 16:907–920. https://doi.org/10.1007/s10346-018-1116-8

Bayarri MJ, Berger JO, Paulo R, Sacks J, Cafeo JA, Cavendish J, Lin CH, Tu J (2007) A framework for validation of computer models. Technometrics 49(2):138–154. https://doi.org/10.1198/004017007000000092

Beguiera S, Van Asch TWJ, Malet JP, Gröndahl S (2009) A GIS-based numerical model for simulating the kinematics of mud and debris flows over complex terrain. Nat Hazards Earth Syst Sci 9(6):1897–1909. https://doi.org/10.5194/nhess-9-1897-2009

Brezzi L, Gabrieli F, Marcato G, Pastor M, Cola S (2016) A new data assimilation procedure to develop a debris flow run-out model. Landslides 13:1083–1096. https://doi.org/10.1007/s10346-015-0625-y
The role of observations in the inverse analysis of landslide propagation. Comput Geotech 92:11–21. https://doi.org/10.1016/j.compgeo.2017.07.011

Christen M, Kowalski J, Bartelt P (2010) RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain. Cold Reg Sci Technol 63(1):1–14. https://doi.org/10.1016/j.coldregions.2010.04.005

Cramer K, Brehmer J, Louppe G (2020) The front of simulation-based inference. Proc Natl Acad Sci 117(48):30055–30062. https://doi.org/10.1073/pnas.1912789117

Fer I, Kelly R, Moorcroft PR, Kelly R, Moorcroft PR, Dietze MC (2016) The importance of entrainment and bulking on debris flow runout modeling: examples from the Swiss Alps. Nat Hazards Earth Syst Sci 15(11):2569–2583. https://doi.org/10.5194/nhess-15-2569-2015

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis. CRC Press, Boca Raton

Gu MY, Wang L (2018) Scaled Gaussian stochastic process for computer computing granular avalanches and landslides. Phys Fluids 15(12):3638–3646. https://doi.org/10.1063/1.1614253

Heredia MB, Eckert N, Prieur C, Thibert E (2020) Bayesian calibration of an avalanche model from autocorrelated measurements along the flow: application to velocities extracted from photogrammetric images. J Glaciol 66(237):373–385. https://doi.org/10.1017/jog.2020.11

Higdon D, Gattiker J, Williams B, Rightley M (2008) Computer model calibration using high-dimensional output. J Am Stat Assoc 103(482):570–583. https://doi.org/10.1198/016214507000008888

Hung O (2009) Numerical modelling of the motion of rapid, flow-like landslides for hazard assessment. KSCE J Civ Eng 13:281–297. https://doi.org/10.1007/s10346-007-0102-3

Hung O (2016) A review of landslide hazard and risk assessment methodology. In: Aversa S, Cascini L, Picarelli L, Scavia C (eds) Landslides: a case study of Heifangtai Terrace, Gansu Province, China. Geomorphology 226:1483–1492. https://doi.org/10.1016/j.geomorph.2016.05.011

Järvenpää M, Guttmann MU, Vehtari A, Marttinen P (2021) Parallel Bayesian inference with expensive likelihood functions. Neural Comput 33(5):106933. https://doi.org/10.1093/cogps/ptaa020

Kandek M, Schneider J, Póczos B (2017) Query efficient posterior estimation in scientific experiments via Bayesian active learning. Artif Intell 243:45–56. https://doi.org/10.1016/j.artint.2016.11.002

Kennedy MC, O’Hagan A (2001) Bayesian calibration of computer models. J R Stat Soc Ser B Stat Methodol 63(3):425–464. https://doi.org/10.1111/1467-9868.00294

McMillan H, Clark M (2009) Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. Water Resour Res 45(4). https://doi.org/10.1029/2008WR007288

McMillan H, Clark M (2009) Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. Water Resour Res 45(4). https://doi.org/10.1029/2008WR007288

Medina V, Hürlimann M, Bateman A (2008) Application of FLATModel, a 2D finite volume code, to debris flows in the northeastern part of the Iberian Peninsula. Landslides 5:127–142. https://doi.org/10.1007/s10346-007-0102-3

Mergili M, Fischer JT, Krenn J, Pudasaini SP (2017) r.avaflow v1, an advanced open-source computational framework for the propagation and interaction of two-phase mass flows. Geosci Model Dev 10(3):553–569. https://doi.org/10.5194/gmd-10-553-2017

Mergili M, Jaboyedoff M, Pullarello J, Pudasaini SP (2020) Back calculation of the 2017 Piz Cengalo–Bondo landslide cascade with r.avaflow: what we can do and what we can learn. Nat Hazards Earth Syst Sci 20(2):505–520. https://doi.org/10.5194/nhess-20-505-2020

Moretti L, Allstadt K, Mangeney A, Capdeville Y, Ruffi M, Ruffi M, Ruffi M, Bouchut F (2015) Numerical modeling of the Mount Meager landslide constrained by its force history derived from seismic data. J Geophys Res Solid Earth 120(4):2579–2599. https://doi.org/10.1002/2014JB01426

Bouchut F (2005) Comparison of flow resistance relations for debris flows using a one-dimensional finite element simulation model. Nat Hazards Earth Syst Sci 6(1):155–165. https://doi.org/10.5194/nhess-6-155-2006

Ruffi M, Mangeney A, Walter F, Capdeville Y, Bouchut F, Bouchut F (2015) Constraining landslide characteristics with Bayesian inversion of field and seismic data. Geophys J Int 221(2):1341–1348. https://doi.org/10.1093/gji/gga056

Ruffi M, Mangeney A, Walter F, Capdeville Y, Bouchut F (2015) Constraining landslide characteristics with Bayesian inversion of field and seismic data. Geophys J Int 221(2):1341–1348. https://doi.org/10.1093/gji/gga056

Sun XP, Zeng P, Li TB, Wang S, Jimenez R, Feng XD, Xu Q (2021) From the Iberian Peninsula. Landslides 5:127–142. https://doi.org/10.1007/s10596-018-9765-1

Sun XP, Zeng P, Li TB, Wang S, Jimenez R, Feng XD, Xu Q (2021) From the Iberian Peninsula. Landslides 5:127–142. https://doi.org/10.1007/s10596-018-9765-1

Tognacca C, Clinton J, Diehl T, Bonanomi Y (2020) Direct observations with granular materials and blocks at small scale. PhD thesis, École polytechnique fédérale de Lausanne, Switzerland

Wang HQ, Li JL (2018) Adaptive Gaussian process approximation for Bayesian inference with expensive likelihood functions. Neural
Xia XL, Liang QH (2018) A new depth-averaged model for flow-like landslides over complex terrains with curvatures and steep slopes. Eng Geol 234:174–191. https://doi.org/10.1016/j.enggeo.2018.01.011

Zhang JJ, Li WX, Zeng LZ, Wu LS (2016) An adaptive Gaussian process-based method for efficient Bayesian experimental design in groundwater contaminant source identification problems. Water Resour Res 52(8):5971–5984. https://doi.org/10.1002/2016WR018598

Zhao H, Amann F, Kowalski J (2021) Emulator-based global sensitivity analysis for flow-like landslide run-out models. Landslides. https://doi.org/10.1007/s10346-021-01690-w

Hu Zhao
Chair of Methods for Model-based Development in Computational Engineering, RWTH Aachen University, Schinkelstr. 2a, 52062 Aachen, Germany
Email: zhao@mbd.rwth-aachen.de

Julia Kowalski
Chair of Methods for Model-based Development in Computational Engineering, RWTH Aachen University, Elfschornsteinstr. 18, 52062 Aachen, Germany
Email: kowalski@mbd.rwth-aachen.de