

Two methodologies to calibrate landslide runout models

Journal Article**Author(s):**

Aaron, Jordan; McDougall, Scott; Nolde, Natalia

Publication date:

2019-05-01

Permanent link:

<https://doi.org/10.3929/ethz-b-000342552>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

Landslides 16(5), <https://doi.org/10.1007/s10346-018-1116-8>

Landslides (2019) 16:907–920
 DOI 10.1007/s10346-018-1116-8
 Received: 23 May 2018
 Accepted: 29 November 2018
 Published online: 6 February 2019
 © Springer-Verlag GmbH Germany
 part of Springer Nature 2018

Jordan Aaron · Scott McDougall · Natalia Nolde

Two methodologies to calibrate landslide runout models

Abstract Extremely rapid, flow-like landslides pose a significant hazard worldwide; however, the analysis of the impact area and velocity of these flows is not routine. Semi-empirical numerical models are one tool that is available for performing this sort of analysis. These models are physically based; however, certain input parameters are determined through model calibration, using back-analysis of real landslide cases. Objective, repeatable calibration methods are needed for this approach to be useful for landslide runout prediction. The present analysis describes the application of optimization theory and Bayesian statistics to calibrate these types of models. Two complementary methods are presented. The first uses the Gauss-Marquardt-Levenberg optimization algorithm to efficiently determine a set of best-fit calibrated model parameters. The second uses a posterior analysis to quantify errors associated with parameter calibration, which can then be used for probabilistic forward analysis. Three case histories are presented to demonstrate how the new methods are able to rapidly calibrate a runout model, reduce subjectivity inherent in the calibration process, and provide information on parameter uncertainty.

Keywords Runout modeling · Flow-like landslides · Hazard mapping · Rock avalanche · Flowslide

Introduction

Some types of landslides, including rock avalanches, flowslides, debris avalanches, and debris flows (using the terminology of Hungr et al. (2014), which is based on Varnes (1978)), can attain high velocities, travel long distances, and impact large areas far from their source, where communities or critical infrastructure may be located (e.g., Boulton et al. 2006; Prochaska et al. 2008; Keaton et al. 2014; White et al. 2015). When a potential source of such a landslide is identified, a runout analysis (to estimate landslide travel distances, flow depths, and velocities) may be required as part of a hazard and/or risk assessment.

The destructive potential of these landslides is demonstrated by two recent case histories. The first is the Oso flowslide, which occurred on March 22, 2014 in Washington State (e.g., Keaton et al. 2014). Here, a terrace slope, which had been subject to previous limited displacement landslides, failed catastrophically. The resulting flowslide impacted the community of Steelhead Haven, killing 43 people. Later that year, the West Salt Creek rock avalanche occurred on the northern flank of Grand Mesa in Colorado (e.g., White et al. 2015). In this case, a slump block failed catastrophically, resulting in a long runout rock avalanche that killed three people (White et al. 2015). In both cases, the flow-like landslides attained high velocities and traveled long distances, which resulted in their tremendous destruction. Predicting the runout characteristics of such events before they occur is a crucial step in protecting society from these disasters.

Numerical models are one tool used to predict the motion of flow-like landslides. Although physically based, many of these models use simplified rheological relationships that require the

calibration of model parameters through back-analysis of real landslide cases. Such models are referred to as “equivalent fluid” models (Hungr 1995). As discussed in Hungr (1995, 2016), considering the complexity and heterogeneity of real landslides, this calibration-based approach is a practical method to analyze full-scale case histories.

When parameters are calibrated, the link between measurable material properties and calibrated model parameters is no longer obvious, and reestablishing this link is critical when using these models for both forensic back-analysis, and forward prediction. The present work presents two calibration methods that are useful for reestablishing this link. We hypothesize that the physical relevance of apparent parameters back-analyzed with equivalent fluid models can be identified by addressing two outstanding problems:

1. As will be described in the following sections, prior to the present work, calibration was generally performed using subjective, trial-and-error calibration. Although there are some notable exceptions to this (e.g., Galas et al. 2007; Cepeda et al. 2010; Brezzi et al. 2016; Calvello et al. 2017; McDougall 2017), no calibration methodology has yet addressed all potential problems associated with model calibration (discussed below). It is desirable to have a more objective, systematic methodology to calibrate runout models. This paper describes the application of techniques from inverse modeling and statistics to develop such a methodology.
2. There is currently no database of case histories, calibrated with a consistent methodology, that can be used to guide probabilistic forward analysis. Aaron (2017) describes the application of the calibration methodologies presented in this paper to a database of 24 rock avalanche case histories, and suggests a methodology for making probabilistic predictions of rock avalanche motion. With an appropriate dataset, this methodology can be extended to other types of extremely rapid, flow-like landslides.

In the present work, we apply methods developed in optimization theory and statistics to the problem of calibrating landslide runout models. Two methods are presented that address the weaknesses inherent in trial-and-error calibration. Three back-analyses that demonstrate the application of these methods are also presented.

Background and motivation

A variety of numerical models have been proposed to simulate the motion of flow-like landslides, as recently reviewed by McDougall (2017). Equivalent fluid models (defined in “Introduction” and “Overview of Dan3D”) have been used to reproduce the velocity, deposit distribution, and impact area of a wide variety of extremely rapid, flow-like landslides (see Aaron (2017) and the references contained therein). However, aside from the exceptions noted above, back-analysis of case histories with equivalent fluid models is performed through trial-and-error calibration. This approach involves manually adjusting the input basal resistance parameters

until a satisfactory reproduction of the simulation constraints is obtained. The suitability of a particular simulation is then subjectively assessed. This method of model calibration suffers from four main weaknesses:

1. It is very demanding of the user's time.
2. Model results are subjectively interpreted, meaning that different users could determine different parameters for the same case.
3. This method does not explore the entire parameter space, so there is no guarantee that the best-fit parameters have been determined. An additional difficulty that comes from this weakness is that, if the model conceptualization is wrong (for example, one material is used instead of two), the user will waste a lot of time varying parameters in search of a good fit, as opposed to changing the simulation configuration.
4. This method ignores parameter non-uniqueness. It is often impossible to determine a unique set of best-fit parameters because multiple parameter sets may produce very similar model outputs.

These four weaknesses can limit the use of back-analyzed parameters in forward analysis. The fact that trial-and-error back-analysis is subjective has been addressed by McDougall (2006), Galas et al. (2007), Cepeda et al. (2010), Brezzi et al. (2016), and Calvello et al. (2017). The matrix method, proposed by McDougall (2006), is a method to visually assess and compare calibration results. This method addresses parameter non-uniqueness; however, its utility is limited to model conceptualizations that only include two parameters. The receiver operating characteristic (ROC) method, proposed by Cepeda et al. (2010), can handle more than two parameters and provides an objective method to compare different parameter combinations; however, it is not guaranteed to explore the entire parameter space and may be demanding of the user's time. The method proposed by Calvello et al. (2017) uses a sensitivity analysis and the modified Gauss-Newton non-linear optimization algorithm to calibrate an equivalent fluid model using numerous estimates of deposit height. To the authors' knowledge, this procedure has so far not been generalized to cases where numerous point estimates of deposit heights are not available.

Methods

Overview of Dan3D

The equivalent fluid model that is used in this work is Dan3D (McDougall and Hungr 2004; McDougall 2006; Hungr and McDougall 2009; Aaron and Hungr 2016a). Although we only use Dan3D in this work for demonstration purposes, the methods presented are applicable to all landslide runout models that require parameter calibration.

Dan3D is a depth-averaged Lagrangian model that simulates landslide motion over complex, three-dimensional terrain. The model solves the following equations of motion using a numerical method based on Smoothed Particle Hydrodynamics (SPH):

$$\rho h \frac{Dv_x}{Dt} = \rho h g_x - k_x \sigma_z \frac{\partial h}{\partial x} + \tau_{zx} - \rho E v_x \quad (1)$$

$$\rho h \frac{Dv_y}{Dt} = \rho h g_y - k_y \sigma_z \frac{\partial h}{\partial y}, \quad (2)$$

where ρ is the density, h is the flow depth, v_x and v_y are the depth-averaged velocities in the local x and y directions, respectively, g_x and g_y are the x and y components of gravity, k_x and k_y are the x and y lateral stress ratios (ratio of lateral stress to bed-normal stress), σ_z is the bed-normal stress, τ_{zx} is the basal resistance stress, and E is the entrainment ratio that governs momentum transfer when accelerating stationary bed material to the velocity of the moving Lagrangian reference frame. The basal resistance stress and momentum transfer due to entrainment only occur in the x direction, which is aligned with the local direction of motion in Dan3D. As shown on Fig. 1, the first term of Eq. 1 is the downslope component of gravity, the second term represents the internal pressure gradients due to the inclination of the free surface, the third term is the basal resistance to motion, and the fourth term is the momentum transfer due to entrainment.

As mentioned previously, the parameters that govern the basal resistance stress (τ_{zx} in Eq. 1) are not true material properties, and can only be constrained through back-analysis of case histories. When performing a back-analysis using Dan3D, the objective is to analyze the runout behavior, assuming that catastrophic failure has occurred. Parameters are selected to simulate the bulk runout behavior, without consideration of pre-failure stability. One common rheology, used to calculate τ_{zx} , is the Voellmy rheology, given in Eq. (3) (e.g., Voellmy 1955):

$$\tau_{zx} = - \left(\sigma_z f + \frac{\rho g v_x^2}{\xi} \right), \quad (3)$$

where f is the friction coefficient and ξ is the turbulence coefficient (sometimes referred to as the turbulence "parameter"). Both f and ξ are calibrated parameters.

The Voellmy rheology is empirical, and has been shown to reproduce the impact area and velocity distribution of many case

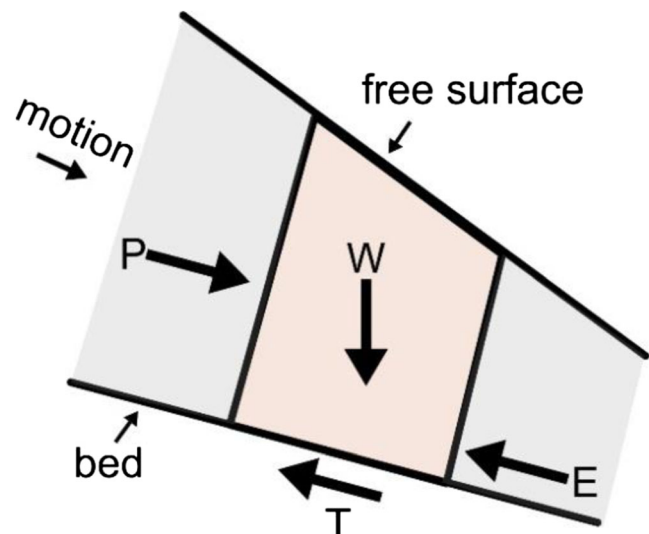


Fig. 1 Conceptual illustration of the forces resolved by Dan3D. The slice is oriented in the direction of motion. W is the weight, T is the basal resistance, P is the internal force due to free surface gradients, and E is the inertial resistance due to entrainment. Figure from Aaron et al. (2017), reprinted under Creative Commons Attribution License 4.0

histories (a summary is provided in Aaron (2017)). However, there may be a more fundamental interpretation. The first term on the right side of Eq. (3) accounts for frictional resistance to motion, and the use of a turbulent resisting stress in the second term may mimic the effect of pore pressure dissipation due to shear-induced dilation (Hung and Evans 2004). Further research is required to study this potential connection; however, it provides a possible explanation for the success of the Voellmy rheology when simulating extremely rapid, flow-like landslides.

Calibration framework

The following sections are based on the generic theory and methods described by Hsieh (2009) and Gregory (2010). As will be discussed, we have adapted these methods to the problem of calibrating equivalent fluid runout models.

Field investigations of a given landslide case history result in the observation of multiple features (or “back-analysis constraints”) that characterize an event (e.g., Keaton et al. 2014; White et al. 2015). These observed features may include the landslide impact area, velocity, deposit depth, and deposit distribution. The K observed features of an event can be represented by the vector:

$$\mathbf{y}^F = (y^{F,1}, \dots, y^{F,K})^T,$$

where the superscript T represents the transpose operator. Dan3D aims to reproduce the runout process of extremely rapid, flow-like landslides, based on a vector of input parameters. Given the value of the parameter vector \mathbf{b} , a Dan3D simulation can be performed. The parameter vector \mathbf{b} refers to the set of all model parameters used in a simulation. The simulation outputs simulated values of the K observed features of a given case, which can be represented by the vector:

$$\mathbf{y}^M(\mathbf{b}) = (y^{M,1}(\mathbf{b}), \dots, y^{M,K}(\mathbf{b}))^T,$$

where $y^{M,i}(\mathbf{b}) (i = 1, \dots, K)$ indicates the simulated value of a given observed landslide feature, for example, the velocity at a point.

The goal of model calibration is to determine the input parameter vector \mathbf{b} that minimizes the difference between the observed and simulated features of an event, \mathbf{y}^F and $\mathbf{y}^M(\mathbf{b})$. These differences (sometimes referred to as “residuals”), between \mathbf{y}^F and $\mathbf{y}^M(\mathbf{b})$, can be collected into the vector:

$$\mathbf{r}(\mathbf{b}) = (r^1(\mathbf{b}), \dots, r^K(\mathbf{b}))^T,$$

where, for example, $r^1(\mathbf{b})$ is the difference between $y^{F,1}$ and $y^{M,1}(\mathbf{b})$.

As will be discussed in more detail in “Selection of standard deviation of a measurement,” there is some uncertainty in both the measured and simulated value of observed landslide features. Therefore, even though the residuals may not be zero, ‘ \mathbf{b} ’ may be the ‘true’ parameter vector for a given case, with the model misfit due to statistical noise. A natural assumption is that the residuals follow a multivariate Gaussian (or normal) distribution with zero mean, and covariance matrix Σ (discussed in more detail in “Selection of standard deviation of a measurement”). Under this normality assumption, the following likelihood function, which

has a maximum when the residuals are zero, can be defined (e.g., Hsieh 2009):

$$L(\mathbf{b} | \mathbf{y}^F, \mathbf{y}^M(\mathbf{b})) = \frac{1}{\sqrt{(2\pi)^K |\Sigma|}} \exp\left(-\frac{1}{2} \mathbf{r}(\mathbf{b})^T \Sigma^{-1} \mathbf{r}(\mathbf{b})\right), \quad (4)$$

where $|\Sigma|$ is the determinant of the covariance matrix Σ . The likelihood function provides a means to compare how well different parameter vectors (“ \mathbf{b} ”) minimize the residuals (“ $\mathbf{r}(\mathbf{b})$ ”), while accounting for statistical uncertainty (quantified through the covariance matrix (Σ)). The parameter vector (“ \mathbf{b} ”) that best minimizes the difference between observed and simulated measurements maximizes this likelihood function. The expression above demonstrates the dependence of the residual vector \mathbf{r} on the parameter vector \mathbf{b} (which controls the numerical model output).

Two methodologies are proposed in this paper to determine the parameter set that maximizes Eq. (4). The first is based on an optimization algorithm, and is more efficient than the second proposed methodology (in terms of run time), at the expense of some transparency in the calibration process. The second is based on a Bayesian posterior analysis, and has the advantage of deriving an easily interpretable posterior probability density function for the calibrated parameters. An example calculation using these calibration methodologies is given in Appendix A.

Method 1: optimization approach using the Gauss-Marquart-Levenberg algorithm

Maximizing Eq. (4) is equivalent to minimizing the following least squares objective function:

$$\Phi(\mathbf{b}) = \mathbf{r}(\mathbf{b})^T \Sigma^{-1} \mathbf{r}(\mathbf{b}), \quad (5)$$

The selection of Σ is discussed in detail in “Selection of standard deviation of a measurement.” If errors can be assumed independent, then Σ is a diagonal matrix:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_K^2 \end{pmatrix}$$

Hence, Eq. (5) can be re-written as:

$$\Phi(\mathbf{b}) = \sum_{k=1}^K \frac{1}{\sigma_k^2} (r^k(\mathbf{b}))^2, \quad (6)$$

where σ_k is the standard deviation of the individual error component. There are many different algorithms that can be used to minimize the objective function (e.g., Nocedal and Wright 2006). The algorithm used in the present work is the Gauss-Marquart-Levenberg (GML) algorithm (Levenberg 1944; Marquardt 1963). This algorithm is described in Nocedal and Wright (2006) and Doherty (2010).

To run the GML algorithm, the user specifies the values of the observed features (\mathbf{y}^F), the components of the covariance matrix (Σ), discussed in “Selection of standard deviation of a measurement,” as well as an initial guess of the best-fit parameter values. The algorithm then runs the model multiple times to determine

the residuals ($\mathbf{r}(\mathbf{b})$), as well as the model sensitivities. Sensitivities are defined as the rate of change of the components of $\mathbf{r}(\mathbf{b})$ with respect to the parameters being calibrated. Based on the residuals and sensitivities, the algorithm then selects a new set of model parameters that reduces the value of the objective function (Eq. (5)). This process is repeated until the residuals are reduced to a user-specified value, the parameter change from one iteration to the next is below a user-specified threshold or the simulation is manually terminated.

The parameter estimation package PEST (Doherty 2010) has been used for optimization analysis detailed in this work. PEST is a model-independent parameter estimation package and contains an implementation of the GML algorithm that can handle both overdetermined (number of parameters greater than the number of constraints) and underdetermined (number of constraints greater than the number of parameters) inverse problems. Additionally, upper and lower bounds on calibrated parameters can be specified, to limit the parameter search space to a user defined area (similar to the Posterior Analysis approach described in “Method 2: posterior analysis of calibrated parameters”).

In addition to providing a set of optimized model parameters, the PEST algorithm outputs a parameter correlation matrix, which provides information on the uniqueness of the best-fit parameters. This matrix provides parameter correlation values between -1 and 1 . When parameter correlation is close to -1 or 1 , multiple combinations of the parameter values will give similar fit results. A negative parameter correlation indicates that similar fit results are obtained when one parameter is increased and the other is decreased. Details of how PEST calculates parameter correlations are found in Doherty (2010).

Method 2: Posterior analysis of calibrated parameters

When performing model calibration, we may have a prior belief about the possible values of the parameter vector \mathbf{b} . For example, based on previously successful back-analyses, we can restrict the range of plausible friction and turbulence coefficients when using the Voellmy rheology (Eq. (3)). We can express this prior belief in the form of a prior distribution, and then update this distribution to include the information gained from evaluating the likelihood function (Eq. (4)). The resulting distribution of the calibrated parameters is known as the *posterior distribution* (Gregory 2010). This sort of analysis, referred to as posterior analysis, is useful when combining multiple back-analysis results for use in a probabilistic forward analysis.

The posterior density of the calibrated parameters can be written using Bayes’ law (Gregory 2010). Let π_{prior} denote the prior density of \mathbf{b} , and π_{post} the posterior density after accounting for the data available about the case history. These data are comprised of the vector of residuals (“ $\mathbf{r}(\mathbf{b})$ ”) derived from field measurements of observed features (\mathbf{y}^F) and computer model outputs ($\mathbf{y}^M(\mathbf{b})$). We then have:

$$\pi_{post}(\mathbf{b}|\mathbf{r}) = \frac{L(\mathbf{b}|\mathbf{y}^F, \mathbf{y}^M(\mathbf{b})) \pi_{prior}(\mathbf{b})}{\int L(\mathbf{b}^*|\mathbf{y}^F, \mathbf{y}^M(\mathbf{b}^*)) \pi_{prior}(\mathbf{b}^*) d\mathbf{b}^*}, \quad (7)$$

where the denominator is a normalizing constant so that the posterior density integrates to one, and \mathbf{b}^* indicates that the integrand in the denominator is over the entire parameter space, distinct from \mathbf{b} in the numerator, which denotes a specific parameter set.

In our implementation using the two-parameter Voellmy model, we assume a uniform prior on the calibrated parameter vector over a rectangular region $B = I_{friction} \times I_{turbulence}$, where $I_{friction}$ and $I_{turbulence}$ are intervals of feasible ranges of the friction and turbulence coefficients, respectively. With normally distributed errors, for values of \mathbf{b} in region B , the posterior density has the form:

$$\pi_{post}(\mathbf{b}|\mathbf{r}) = \frac{\exp\left(-\frac{1}{2} \sum_{k=1}^K \frac{1}{\sigma_k^2} (r^k(\mathbf{b}))^2\right)}{\iint_B \exp\left(-\frac{1}{2} \sum_{k=1}^K \frac{1}{\sigma_k^2} (r^k(\mathbf{b}^*))^2\right) d\mathbf{b}^*}. \quad (8)$$

Due to the selected prior, $\pi_{post}(\mathbf{b}|\mathbf{r}) = 0$ for values of \mathbf{b} not contained in the rectangular region B .

Equation (8) shows that, with the assumption of normally distributed and independent errors, and a uniform prior, the posterior probability of the parameters is the likelihood function normalized so that $\int_B \pi_{post}(\mathbf{b}|\mathbf{r}) d\mathbf{b} = 1$, where the parameter space B is defined based on the prior distribution π_{prior} . Parameter combinations that, when used in Dan3D, well reproduce the field observations will be given high probabilities, and those that poorly reproduce the field observations will be given low probabilities.

Quantification of simulation constraints

In order to evaluate Eqs. (5) and (8), it is first necessary to define features that can be output by Dan3D and compared to field observations of extremely rapid, flow-like landslides. The following four quantities can be used to compare a Dan3D simulation to field data:

- Impact area;
- Estimates of volumes deposited in spatial areas;
- Point estimates of deposit depth; and
- Point estimates of flow velocity.

It should be noted that these four quantities are not the only features that can be used. However, they have proven useful for back-analysis of rock avalanches (examples of which are summarized in Aaron (2017)). For other types of flow-like landslides, additional landslide features may prove useful. The two calibration methodologies presented in “Method 1: optimization approach using the Gauss-Marquart-Levenberg algorithm” and “Method 2: posterior analysis of calibrated parameters” are able to incorporate any calibration constraint available for a given case history.

A procedure has been defined to automatically assess how well Dan3D reproduces these observations. A post-processor has been created to interpret Dan3D outputs and create a model output vector ($\mathbf{y}^M(\mathbf{b})$) at the conclusion of each simulation. It quantifies model outputs using the procedure detailed in the following sections.

Landslide impact area

When calibrating runout models that analyze motion along a path profile, an intuitive and widely used feature is the maximum runout distance. When performing model calibration over 3D terrain, however, the runout distance is difficult to define because the flow may bifurcate and deposit in multiple lobes. Instead of using runout distance as a metric, in 3D it is necessary to define a metric that quantifies the similarity between simulated and observed impact areas. Galas et al.

(2007), Cepeda et al. (2010), and Heiser et al. (2017) have proposed such metrics, and the following algorithm is based on these contributions.

Field or remote sensing investigations of long runout landslides commonly produce maps of the landslide impact area. These maps can be used to create a grid file (hereafter referred to as the trimline grid file) that has a value of one for areas that are within the impact area and a value of zero for areas that are outside the impact area. An example trimline grid file, created for the Zymoetz River Rock Avalanche (discussed in [Zymoetz River rock slide-debris avalanche](#)), is shown in Fig. 2a.

One of the outputs of Dan3D is a grid file that shows the maximum flow depth recorded at each node of the inputted sliding surface grid file during the simulation. An example of such a file is shown in Fig. 2b. As can be seen in this figure, the landslide is thickest in the source area, due to the fact that the sliding mass is concentrated in a small area. As the simulation is advanced in time, the sliding mass spreads out and, as a result, the flow depth thins as the sliding mass moves downslope.

In order to quantify how well a given simulation matches the landslide impact area, the maximum depth file is compared to the trimline grid file. These two grid files are overlapped to create a new grid (hereafter referred to as the trimline fitness grid) that has a value of zero where the trimline file contains a value of one and the maximum depth file contains a depth greater than a user-specified cutoff value (as summarized in McDougall (2006), a cutoff is necessary to realistically define the margins of the flow/deposit due to the

numerical solution method used by Dan3D). This result indicates agreement between Dan3D and the field investigation that an impact occurred at that node of the grid. This grid also contains a value of zero where the trimline file has a value of zero and the maximum depth is less than the user-specified cutoff value. This result indicates agreement between Dan3D and the field investigation that no impact occurred at that node of the grid. In areas where these conditions are not true, the grid has a value of one. This result indicates disagreement between Dan3D and the field investigation. The values of the trimline fitness grid are then summed to give a single fitness number. This number is zero if Dan3D perfectly simulates the trimline grid file, and gets bigger as the fitness gets worse.

Landslide deposit distribution

An important simulation constraint for some landslides is the deposit distribution. Investigations of long runout landslides often provide estimates of the volume of material deposited in different areas of the landslide path (e.g., Evans et al. 1994). In order to quantify this simulation constraint, the following routine has been implemented. A grid file of deposit zones based on the field investigation is created by the user and input into the post-processor. This grid file defines zones within the landslide path for which deposit volume estimates are available. At the end of each simulation, the volume deposited in each of the zones is calculated and output. These volumes can then be compared to known

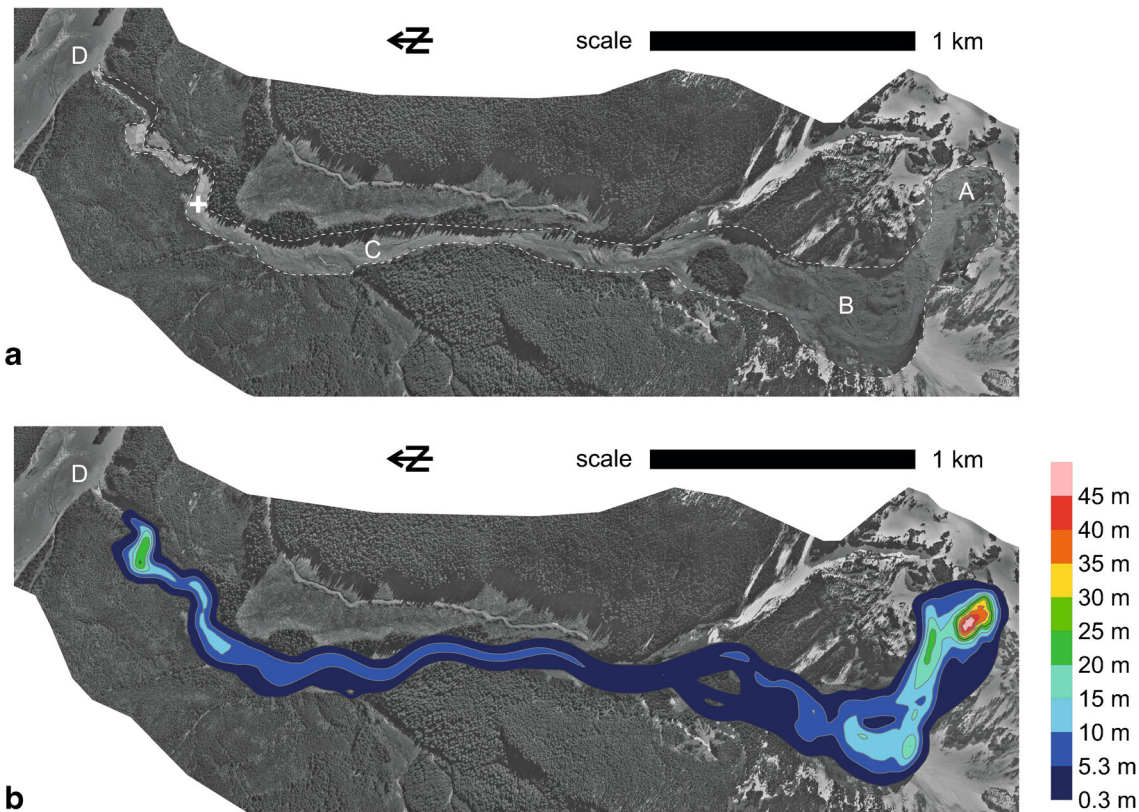


Fig. 2 Overview of the Zymoetz River rock avalanche. The simulation constraints used are the volume of the deposit near the "C" on the figure ($100,000 \text{ m}^3$), the velocity of the flow at the cross (17 m/s), and the impact area (extending from "A" to "D"). The rheology change was implemented in the channel, downstream of the location labeled "B." The top image (a) shows a trimline created based on a field investigation. This grid has a value of one for locations within the dashed white lines, and a value of zero for locations outside the dashed lines. The bottom image (b) shows an example of a maximum depth grid file output by Dan3D. Thickness values are reported in meters. Image modified from McDougall et al. (2006). Image: Province of British Columbia, Copyright © Province of British Columbia

volume estimates. An example deposit distribution constraint is shown on Fig. 3, where estimates of the volume of material deposited in two different zones can be seen.

Landslide deposit thickness estimates

In addition to inputting deposit zones, the user can input the spatial coordinates of points where the deposit depth is known. The post-processor will then output the deposit depth at these points at the conclusion of a simulation. The deposit depth can then be compared to field estimates of deposit depth.

Landslide velocities

Estimates of landslide velocities are generally made at specific spatial points. These estimates are commonly of the maximum landslide velocity at a given point in space, often based on super-elevation measurements (e.g., Hungr et al. 1984; Prochaska et al. 2008), seismic records (e.g., Allstadt 2013), or video evidence (e.g., Sosio et al. 2008). To quantify velocity constraints, the spatial coordinates of points where velocity estimates are available can be input into the post-processor. Dan3D tracks the maximum velocities simulated at each node during the simulation. At the end of each simulation, the maximum velocity at each of the user-specified points is output. These simulated velocities can then be compared to field estimates of velocity.

Selection of standard deviation of a measurement

The standard deviation of a simulation constraint is a statistic that quantifies the variability in the measurement error in the value of the constraint. In order to evaluate Eqs. (5) and (8), we need estimates of σ_k 's ($k = 1, \dots, K$), the standard deviations of the error components, which quantify the degree of variability in the measurement error associated with the value of a simulation constraint. For the simulation

constraints summarized above, selection of this statistic will require subjective judgment. However, some guidance can be given by exploring the source of errors. The two main sources of error are as follows:

1. The portion of the error that is a result of uncertainties in a measurement;
2. The portion of the error that results from uncertainties in input data as well as the model being an imperfect simulation of reality.

To understand the first source of error, consider the case of estimating the velocity of a landslide from a super-elevation measurement. Prochaska et al. (2008) show that velocity estimates derived from super-elevations are subject to many sources of error. These sources include estimating the radius of curvature as well as the difficulty of distinguishing splash marks from true super-elevations. For some cases presented by Prochaska et al. (2008), velocity ranges of 5 to 25 m/s appeared defensible. The standard deviation for this constraint should be selected to account for this source of error. For some super-elevation observations, such as the Zymoetz River Rock Avalanche (McDougall et al. 2006), the radius of curvature and banking angle appear well-constrained and the error associated with velocity estimates can be dramatically reduced.

An example of the second source of error is demonstrated in the simulation of the trimline of the Zymoetz River Rock Avalanche (Fig. 2b). Comparing Fig. 2a and b shows that Dan3D does not perfectly reproduce the observed trimline (the value of the trimline fitness is > 0). The observed trimline is derived from accurate remote sensing data, and it is unlikely that this residual is a result of measurement error. Instead, this residual results from a host of sources that are difficult to rigorously quantify, including uncertainties in pre-failure topography, volume and shape of the initial failure, distribution and character of path materials, as well

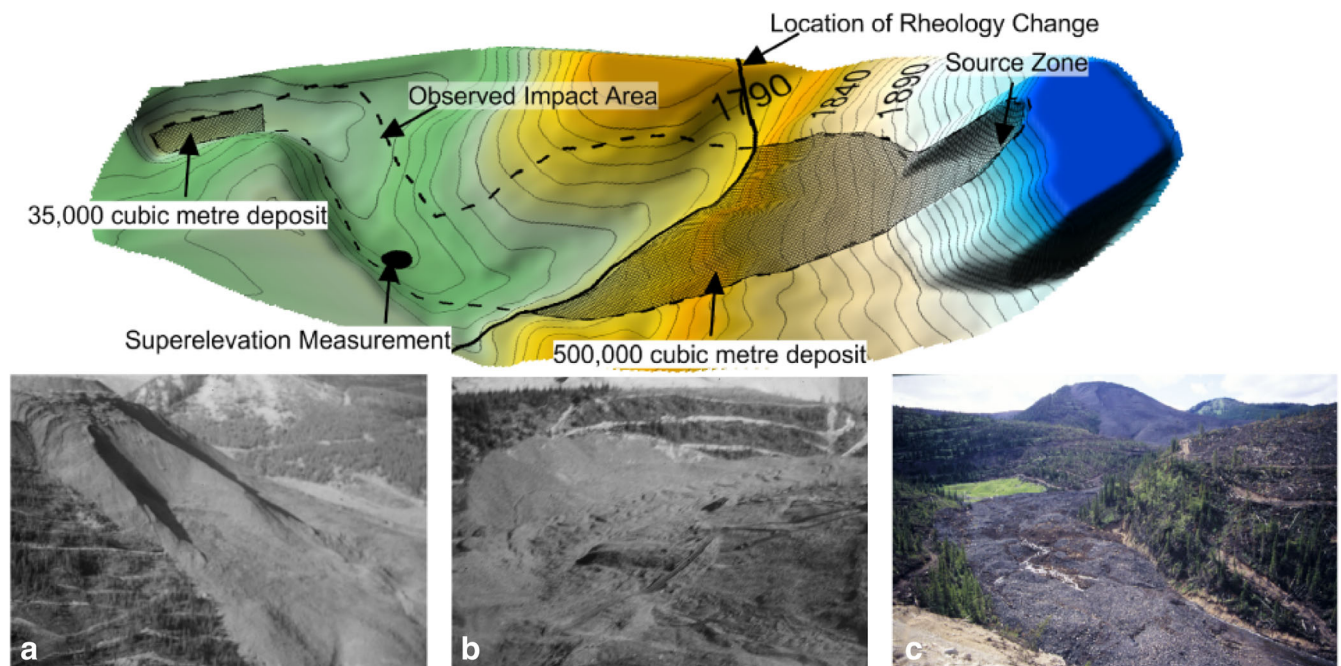


Fig. 3 Overview of the coal mine waste flowslide and the field observations used for model calibration. Three images of the flowslide area are shown: **a** shows an image of the source zone, **b** shows an image of the superelevation feature used to estimate flow velocity, and **c** shows an image of the highly mobile, channelized portion of the flow. Photos: Hungr (2017) reproduced with permission

as the assumptions used to derive the governing equations of Dan3D. All of these sources of error contribute to the non-zero trimline residual, detailed above.

Based on the discussion above, the following heuristics are suggested for selecting the standard deviation of the simulation constraints. The sensitivity of calibrated results to these values will be discussed in “Applications of three case histories” and “Discussion.” For quantities where simulation results and field estimates can be directly compared, such as velocity, deposit distribution, and deposit depth constraints, a best estimate and range of potential values should be determined. The standard deviation can then be estimated by dividing the difference between the upper and lower end of the range by four (with the assumption of normally distributed errors, this corresponds to a 95% probability that the estimated value lies within the estimated range).

For the trimline fitness metric, Aaron (2017) found that equivalent fluid models can typically reproduce an observed trimline with a residual of 25% (based on a database of 24 case histories). Therefore, the standard deviation of the trimline fitness metric can be estimated by:

$$\sigma_{trimline} = \frac{0.25 N_{ones}}{2}, \quad (9)$$

where $\sigma_{trimline}$ is the trimline standard deviation, and N_{ones} is the number of ones in the trimline grid file.

Application to three case histories

The following sections detail the application of the new calibration methodologies to three case histories. These include a coal mine waste dump flowslide, the Frank Slide, and the Zymoetz River Rock Avalanche. For all cases, the frictional rheology was used in the source zone, and the Voellmy rheology was used along the path. First, the coal mine waste dump flowslide is presented to demonstrate the optimization method. Next, the Frank Slide is presented to demonstrate the posterior analysis method. Finally, these two methods are compared using the example of the Zymoetz River Rock Avalanche. These cases demonstrate the application of the new calibration methodologies, and provide novel insights into the calibration process of equivalent fluid models. For each of the case histories, we only calibrated the parameters that govern the basal resistance, as entrainment volume was not significant relative to the source volume in any of the case histories, and previous work has shown that simulation results are insensitive to other user-specified parameters in Dan3D (Hung and McDougall 2009). The standard deviations used for the simulations, derived based on the methodology detailed in “Selection of standard deviation of a measurement,” are shown in Table 1. An overview of how each value in Table 1 was selected is provided in the description of each of the individual case histories.

Coal mine waste dump failure

There have been numerous long runout flowslides originating on coal mine waste dumps (e.g., Hungr 2017). Hungr et al. (2002) back-analyzed the runout characteristics of 44 coal mine waste dump flowslides using the equivalent fluid model Dan-W. They found that the majority of cases could be simulated using a back-analyzed friction angle of 21°. This value of bulk friction angle was explained as an average resistance composed of both liquefied and non-liquefied portions of the source mass. About 30% of the cases analyzed by Hungr et al. (2002) required a two-rheology simulation in order to accurately reproduce field observations. A frictional rheology was used in the source zone, and the Voellmy rheology was used along the path. The use of the Voellmy rheology was justified because the failed material overran loose saturated substrate, which can enhance mobility (Hung and Evans 2004).

A well-documented coal mine waste dump flowslide that occurred in British Columbia, Canada, has been back-analyzed using the GML algorithm. An overview of this event is shown in Fig. 3. The topography for this flowslide was digitized from a pre-event topographic map. The initial volume of the failed material was 700,000 m³. An estimated 500,000 m³ (estimated range of 440,000 to 560,000 m³) deposited near the source zone shown on Fig. 3, and 35,000 m³ (estimated range of 30,000 to 40,000 m³) deposited near the toe. Velocity estimates of 17 m/s (estimated range of 13 to 21 m/s) were made based on a superelevation measurement at the point shown on Fig. 3.

For this back-analysis, a two-rheology simulation was used, with the switch corresponding to the boundary between the source slope and the channel (shown on Fig. 3). This boundary corresponds to the location where the flowslide overran loose, organic-rich path material that was saturated by snowfall (Hungr 2017). This back-analysis required the calibration of three parameters (the friction angle in the source zone, and the two Voellmy parameters for the path), a difficult task to perform systematically using trial-and-error calibration. Three different initial starting guesses were tested in order to examine the uniqueness of the calibrated parameters. The results of the calibration using the standard deviations shown in Table 1 are shown in Table 2.

The standard deviations for this inverse analysis were selected based on the procedure detailed in “Selection of standard deviation of a measurement.” The trimline grid file contains 16,400 ones resulting in a trimline standard deviation of 2050 using Eq. (9). Using the deposit depth and velocity ranges given above, the standard deviations for these three constraints were selected as 33,000 m³, 3300 m³, and 2 m/s for the source zone deposit, toe deposit, and velocity constraint, respectively.

The best-fit simulation results are shown in Fig. 4. The best-fit parameters were found to be 24° for the source material, and a friction coefficient of 0.035 to 0.04 and turbulence coefficient of 500 to 1000 m/s² for the channelized portion of the runout path.

Table 1 Standard deviations used for the three back-analyses

Case History	Trimline (dimensionless)	Velocity (m/s)	Deposit Distribution (m ³)
Frank Slide	750/1,360	-	-
Zymoetz River	1,150	1	15,000
Coal Mine Waste Dump	2,050	2	33,300/3,300

Table 2 Calibrated parameters for the coal mine waste dump case history

Initial Parameters	Calibrated Parameters	Number of Model Runs
$\phi_b = 21^\circ$ $f = 0.04$ $\xi = 600 \text{ m/s}^2$	$\phi_b = 23^\circ$ $f = 0.042$ $\xi = 460 \text{ m/s}^2$	6
$\phi_b = 30^\circ$ $f = 0.1$ $\xi = 500 \text{ m/s}^2$	$\phi_b = 24.5^\circ$ $f = 0.036$ $\xi = 1072 \text{ m/s}^2$	48
$\phi_b = 15^\circ$ $f = 0.2$ $\xi = 500 \text{ m/s}^2$	$\phi_b = 24.8^\circ$ $f = 0.035$ $\xi = 845 \text{ m/s}^2$	41

As shown in Table 2, regardless of the initial starting condition, the model converged to similar parameter values, indicating that the parameters are well resolved in this case. This result is also shown in Table 3, which shows that the friction angle in the source zone and the Voellmy friction coefficient along the path are correlated, whereas the turbulence coefficient is well resolved (indicated by low correlation with the other parameters). This low correlation is due to the inclusion of a velocity constraint. The back-analysis of the Frank Slide, discussed below, shows that when no velocity constraints are available, the friction and turbulence coefficients can be strongly correlated.

Frank Slide

The Frank Slide occurred in 1903 in Alberta, Canada, and destroyed a portion of the town of Frank, claiming an estimated 70 lives (Cruden and Krahn 1978). An image of this event is shown in Fig. 5. The volume of this rock avalanche has been estimated as 36 Mm^3 (Cruden and Krahn 1978). This rock avalanche initiated as a planar slide along a slope parallel bedding plane on the eastern limb of Turtle Mountain (Cruden and Krahn 1978). BGC (2000) and Read et al. (2005) list a number of possible triggers for this event, including slope deformation due to coal mining and above average precipitation in the months preceding the event; however, no trigger has been definitively proven. After failure, the Frank Slide descended the slopes of Turtle Mountain and overran and entrained loose saturated sediments (Cruden and Hungr 1986).

Equivalent fluid models have been used previously to back-analyze the Frank Slide (e.g., McDougall 2006; Hungr et al. 2007). The topography files used in the present analysis are the same as those used in McDougall (2006) based on a post-event DEM modified to reflect the pre-1903 conditions. Dan3D-Flex (Aaron

and Hungr 2016a), which accounts for the initially coherent stage of motion, was used for this analysis. A coherent motion distance of 500 m was selected, corresponding to fragmentation as the rock avalanche vacated the source zone. A 15° bulk basal friction angle was used in the source zone. This friction angle was selected to ensure that the entire source mass vacates the source zone, and was not treated as a calibrated parameter in the back-analysis. Although empirical, this friction angle corresponds to the ultimate strength measured by Cruden and Krahn (1978) along flexural slip planes sampled from the debris of the Frank Slide. Both a posterior and GML analyses were used to determine the best-fit parameters for the floodplain material. The rheology change was implemented at the toe of the slope to correspond with the location where the mass encountered loose, saturated substrate.

The results of the posterior and GML analyses using the standard deviations shown in Table 1 are shown in Fig. 6. The standard deviation of 1360 for the trimline fitness was estimated based on Eq. (9), and the standard deviation of 750 was selected for comparison. Both the friction coefficient and the turbulence coefficient are constrained by the back-analysis; however, many different combinations result in the same trimline fitness. This is indicated on Fig. 6 by the shape of the high probability zone derived from the posterior analysis, as well as the dependence of the GML results on the initial guess. Therefore, for the case of the Frank Slide, the parameters are strongly correlated (parameter correlation is described in “Method 1: optimization approach using the Gauss-Marquart-Levenberg algorithm”). The reason for this correlation is that there is only one simulation constraint (the impact area); however, there are two calibrated parameters, resulting in non-unique calibration results. As will be discussed in “Discussion,” this has implications for using this case as a precedent for forward prediction. Additionally, Fig. 6 shows that the calibration

Table 3 Parameter correlation matrix calculated for the coal mine waste dump at the conclusion of the optimization analysis

	Friction Angle	Friction Coefficient	Turbulence Coefficient
Friction Angle	1	-0.80	0.30
Friction Coefficient	-0.80	1	0.17
Turbulence Coefficient	0.30	0.17	1

Table 4 Parameter correlation matrix calculated for the Zymoetz River Rock Avalanche at the conclusion of the inverse model. The moderate correlation between the two parameters reflects the fact that multiple parameter combinations can give similar fitness results

	Friction Coefficient	Turbulence Coefficient
Friction Coefficient	1	-0.59
Turbulence Coefficient	-0.59	1

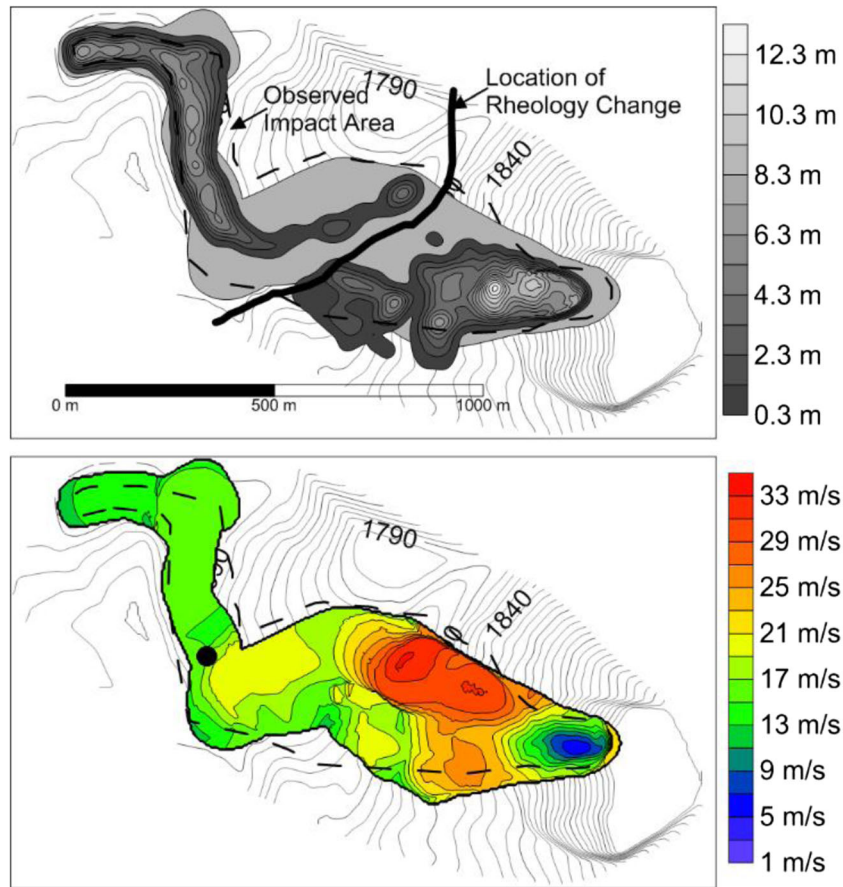


Fig. 4 Best-fit simulation results for the coal mine waste dump flowslide (“Discussion”). Top: deposit depths and impact area simulated by the model. The deposit in the source zone and the distal runout distance are well reproduced. Bottom: maximum velocities simulated by the model. The velocity at the superelevation is reproduced

results are relatively insensitive to the choice of standard deviation. With larger standard deviations, the uncertainty in the values of the calibrated parameters increases somewhat, due to larger uncertainty in the values of the constraints.

Zymoetz River rock slide-debris avalanche

The Zymoetz River Rock Avalanche (ZRRRA) occurred in 2002 in British Columbia, Canada, about 18 km east of the city of Terrace, B.C. An aerial photograph of this landslide is shown in Fig. 2. The



Fig. 5 Overview of the Frank Slide. The source zone as well as the blocky debris are visible

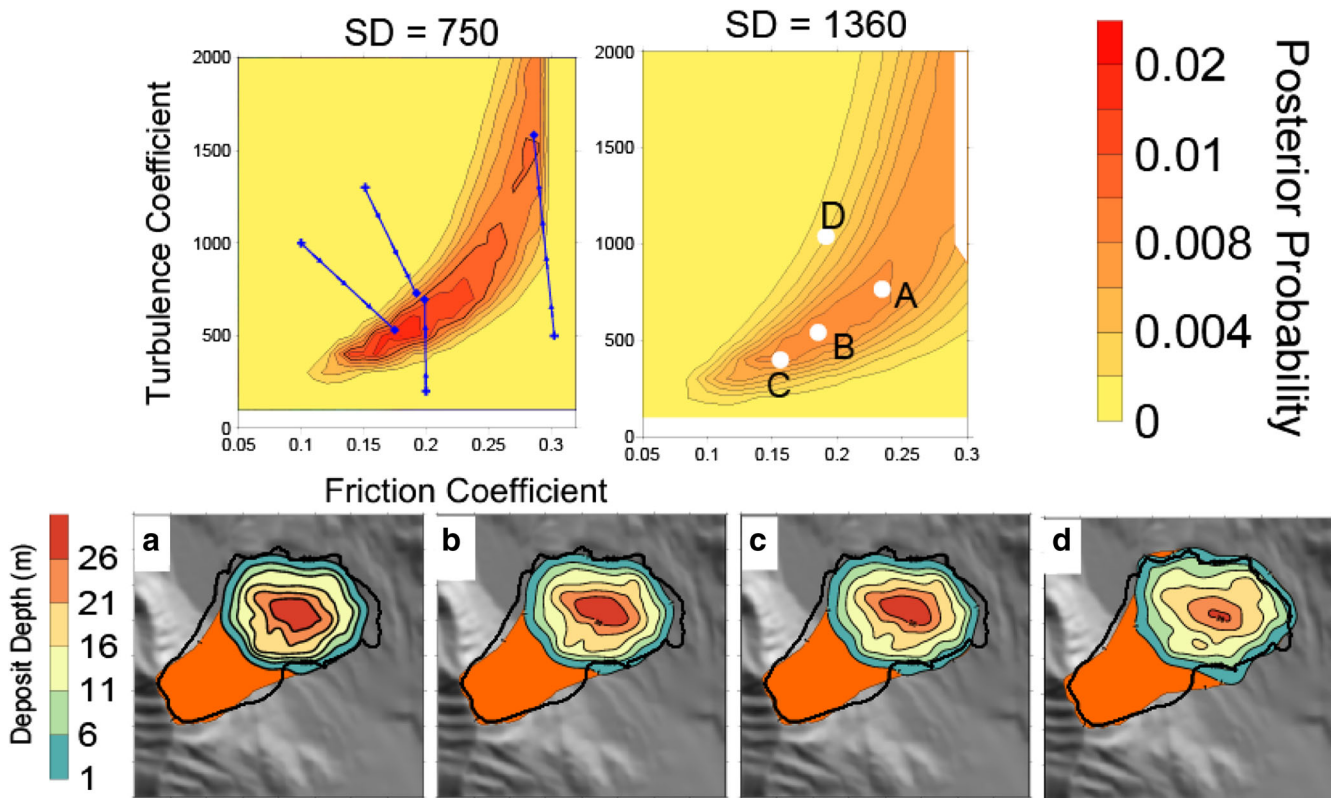


Fig. 6 Posterior probability density functions derived for the Frank Slide, as well as multiple simulations using relatively high probability parameters (A, B, C, and D). The black outline shows the observed impact area and the orange outline shows the simulated impact area. The blue crosses show the starting point for the GML analysis, and the blue diamonds show the final calibrated parameters

landslide initiated as a slide in volcanic bedrock, with an initial volume of about 900,000 m³. The landslide traveled about 4 km down a sinuous channel, entraining an additional 500,000 m³ of material (McDougall et al. 2006). This highly mobile landslide severed a gas pipeline, resulting in an estimated indirect cost of \$30 million (Schwab et al. 2003).

This event has been described in Schwab et al. (2003), Boulton (2005), and Boulton et al. (2006). A previous Dan3D back-analysis of this event is provided in McDougall et al. (2006). In the present analysis, three simulation constraints were used. These constraints include the impact area, deposit distribution, and a velocity estimate. The impact area of the landslide was determined from post-event aerial imagery. A trimline grid file was created based on this information, and is shown in Fig. 2. It was estimated that a volume of 600,000 m³ deposited in the upper cirque basin, 200,000 m³ in the channel (about 100,000 ± 30,000 m³ deposited downstream of “C” on Fig. 2), and 600,000 m³ in the river at the bottom of the channel (see Fig. 2 for the location of these deposit zones) (Boulton 2005). Figure 2 shows the location where the flow rounded a bend and superelevated. Based on the forced vortex equation, a velocity of 17 ± 2 m/s was estimated (Boulton et al. 2006). A friction angle of 30° was used in the source zone, based on McDougall et al. (2006), and both an optimization and posterior analyses were performed to calibrate the parameters for the path material. The methodology detailed in Section “Selection of standard deviation of a measurement” was used to select the standard deviations given in Table 1, based on the best estimate and uncertainty values given above.

Application of the GML algorithm

The GML algorithm has been applied to the ZRRR to determine a set of calibrated parameters. Figure 7 shows a contour map of the values of the objective function (calculated from Eq. (5)), as well as the optimization steps taken by the GML algorithm. Lower values of the objective function indicate better agreement between model results and observations. The parameter correlation matrix is presented in Table 4. This matrix indicates that parameter correlation (defined in “Method 1: optimization approach using the Gauss-Marquart-Levenberg algorithm” as an estimate of the uniqueness of the calibrated parameters) is moderate. Figure 7 shows that a range of parameter values from friction coefficients of 0.095 to 0.11 and turbulence coefficients from 1300 to 1900 m/s² all give similar fitness values. The moderate parameter correlation is a reflection of this non-uniqueness.

Application of posterior analysis

The results of the posterior analysis using the standard deviations shown in Table 1 are summarized in Fig. 8. In Fig. 8, the calculated values of the residuals have been normalized to have a mean of zero and standard deviation of one. This operation, which was done by subtracting the mean of residuals and dividing by the standard deviation (Hsieh 2009), results in dimensionless residuals, facilitating easy comparison between the simulated values of the various observed landslide features.

The contours of the normalized residual for the volume simulated to deposit in the channel (Fig. 8c) show that this constraint is insensitive to the turbulence coefficient. This result is because,

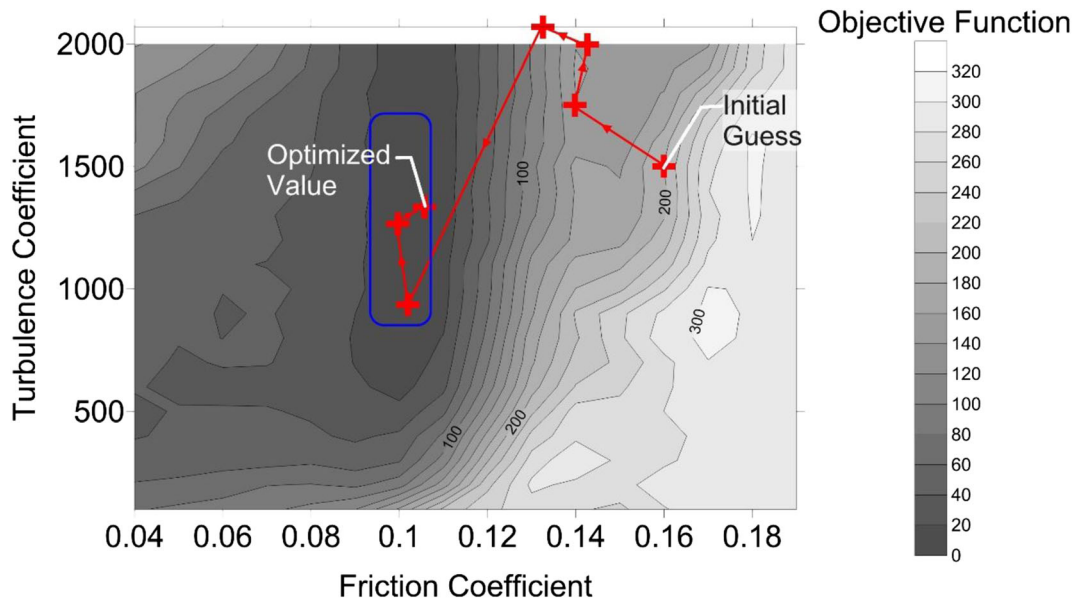


Fig. 7 Optimization steps taken by the GML algorithm for the Zymoetz River Rock Avalanche, overlaid on the contours of the objective function. The red crosses show the steps taken by the GML algorithm, described in “Method 1: optimization approach using the Gauss-Marquart-Levenberg algorithm.” For comparison, the rectangle denotes the high probability zone determined by the posterior analysis (see Fig. 8)

when the Voellmy rheology is used, the friction coefficient determines the slope angle where material deposits. The contours of the normalized residual of the velocity estimate (Fig. 8b) show that this constraint is sensitive to both the turbulence and friction coefficient, with the highest velocity corresponding to the lowest resistance parameters.

The posterior density, shown in Fig. 8a, defines a narrow parameter range, with three peaks. These peaks are due to the non-linearity in the velocity residual (Fig. 8b). If further simulation constraints were available for this case (for example, additional velocity estimates), the zone of best-fit parameters would likely shrink.

As can be seen by comparing Figs. 7 and 8, the results of the GML algorithm agree with that of the posterior analysis approach, as the GML algorithm determined the most probable parameters set. The GML algorithm takes 42 model runs in order to converge to the minimum value of the objective function, fewer than the 285 model runs needed to perform the posterior analysis.

Discussion

Both the GML algorithm and the posterior analysis approach can be used to effectively calibrate equivalent fluid models for a given case. In both procedures, model calibration is performed automatically, and only needs minimal user intervention. Additionally, these algorithms provide some assurance that the best set of parameters for a given model parameterization has been found. The GML algorithm can determine a set of best-fit parameters using relatively few model runs (as compared to the posterior analysis approach). This is because the GML algorithm makes use of model sensitivities (described in “Method 1: optimization approach using the Gauss-Marquart-Levenberg algorithm”) to efficiently explore the parameter space, whereas the posterior analysis approach relies on a grid search of the parameter space. The drawback of this approach is that parameter non-uniqueness is difficult to assess, as compared to the results of a posterior analysis.

The posterior analysis approach requires more model runs; however, it can be used to evaluate the posterior distribution of the calibrated parameters. This is useful as the posterior distribution provides information about the uncertainty in the calibrated parameters and can be used to combine the results of multiple back-analyses for probabilistic forward prediction (e.g., Aaron 2017). Therefore, when run times are feasible, it is recommended that the posterior analysis approach be used.

For all cases tested so far, the calibration results are insensitive to the methodologies used to quantify the simulation constraints. Velocity, deposit depth, and deposit distribution are calculated by the model, so these quantities can be directly compared with field estimates. For impact area, one potential factor which may lead results to be sensitive to the chosen methodology is a case where reproducing the distal runout can only be achieved by dramatic overprediction of the runout in other areas. In such cases, it may be useful to define separate metrics for over- and under-estimation of deposition, which could then be weighted separately. The calibration methodology presented in this paper can be easily modified to include such a constraint if it should prove useful for future analyses.

Unlike trial-and-error calibration, in which the user subjectively interprets the results, the user of the proposed methods is required to subjectively specify a covariance matrix as part of the input procedure (guidelines for specifying this matrix are given in “Selection of standard deviation of a measurement”). This matrix explicitly quantifies uncertainty in measured values, allowing for transparency when defining that a given model residual is probable (a process implicit in subjectively interpreting model results). At present, heuristics (detailed in “Selection of standard deviation of a measurement”), supplemented by judgment, can be used to select this matrix. Figure 6 suggests that, for the case of the Frank Slide, the calibration results are relatively insensitive to the choice of standard

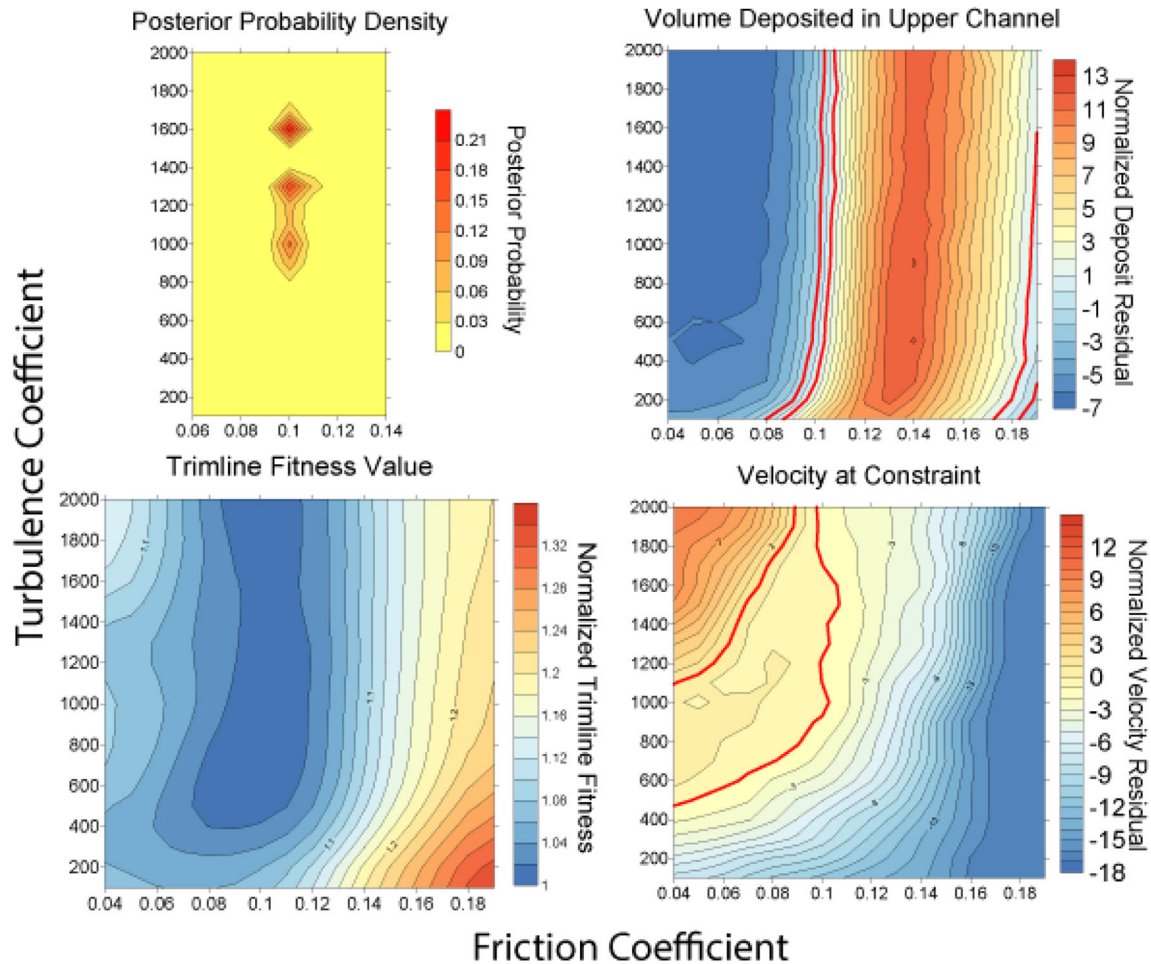


Fig. 8 Zymoetz River Rock Avalanche calibration results. The posterior probability density function and normalized residuals of each simulation constraint are shown. High probability zones in the parameter space correspond to areas where all three residuals are near zero. These zones are highlighted in red for the plots of normalized velocity and volume residuals

deviation. Future work could focus on developing an objective methodology to define the uncertainty in the measured values of the various types of back-analysis constraints. Additionally, the added objectivity of the present approach comes at the expense of more complexity in understanding and interpreting the results of model calibration. However, these calibration methodologies address the main limitations of trial-and-error calibration (listed in “Background and motivation”).

The covariance matrix can be used to incorporate expert judgment into the calibration process. This may become necessary in the future, as the increased availability of pre- and post-event digital elevation data could potentially result in a large number of deposit depth measurements. Such a large number of measurements may dominate the calibration process; however, the covariance matrix can be selected to counteract this by giving each measurement a high standard deviation. Additionally, features can be assigned low standard deviations if it is desirable to more accurately reproduce them, at the expense of accurate simulation of other features.

The calibration results presented in Figs. 7, 8, and Table 2 for the coal mine waste dump flowslide and Zymoetz River rock avalanche have important implications for the use of equivalent fluid models for both forensic back-analyses and hazard prediction. By combining site investigation data with the back-analyzed parameters, information can be inferred about the movement mechanisms of these two extremely rapid, flow-like landslides. As summarized in McDougall et al. (2006), the mobility of the Zymoetz River rock avalanche can be explained based on two mechanisms. In the source zone, strength was typical of that expected for dry fragmented rock, resulting in a best-fit friction angle of 30° . Along the path, interaction with saturated sediments enhanced mobility. The analysis of the Zymoetz River case presented here shows that these back-analyzed strengths are unique, supporting the interpretation detailed above. Similarly, the back-analyses of the coal mine waste dump case study support the hypothesis advanced by Hungr et al. (2002) and Hungr (2017) that two mechanisms resulted in its dramatic runout. The first is partial liquefaction of the material in the source zone, resulting in a bulk

friction angle of $\sim 24^\circ$. The second is interaction with saturated organic material, which resulted in low strengths in the distal zone, dramatically increasing the mobility of the event.

In both these cases, the calibration methodologies presented here provided an efficient means of calibrating the model for different parameterizations, and an objective way of comparing the results. Therefore, more confidence can be placed in the conclusions regarding the movement mechanisms of case histories calibrated with these methodologies, as they are based on a systematic, repeatable procedure. This approach was used in Aaron et al. (2017) to infer movement mechanisms of the West Salt Creek Rock Avalanche.

As shown by the Frank Slide example (Fig. 6), back-analysis results with equivalent fluid models are often non-unique when only the impact area is available as a constraint (a common scenario in published back-analyses (e.g., Aaron and Hungr 2016b)). Figure 7 and Table 4 show that the inclusion of a velocity constraint dramatically reduces this parameter non-uniqueness. Parameter non-uniqueness presents a challenge when using these models for forward analysis, as it may be unclear which parameters should be used to make a forward prediction. This is a problem that has been mentioned in the literature (e.g., Körner 1976; Hungr et al. 2005), but to our knowledge this problem has not been quantitatively investigated. Through the use of a posterior analysis (Fig. 6 and Fig. 8) or by interpreting the results of the GML algorithm (Tables 2, 3, and 4), the calibration methodologies presented in this paper can help model users identify and quantify parameter non-uniqueness. Ongoing work is focused on explicitly accounting for this source of uncertainty when predicting rock avalanche motion.

One potential limitation of the two proposed calibration methodologies is that they are ill suited for calibrating a large number of parameters. Aaron (2017) applied these techniques to calibrate as many as five parameters, however if calibration of 10's of parameters were desired then run times may get very slow. This problem could be overcome with the use of the GML algorithm and more sophisticated techniques to calculate model sensitivities (e.g., Nocedal and Wright 2006).

Conclusions

Calibration of landslide runout models is conventionally performed using a trial-and-error approach. This approach suffers from four major weaknesses: it is time consuming, subjective, may not explore the entire parameter space, and does not acknowledge parameter non-uniqueness. This paper presents two calibration methodologies, one based on an optimization algorithm and the other based on a posterior analysis, that address these four weaknesses. With these methodologies, model calibration is more efficient, repeatable, and objective. Ongoing work is focused on applying this calibration methodology to a wide variety of case histories, which will likely lead to the implementation of new calibration constraints that are appropriate for a wide variety of extremely rapid, flow-like landslides.

Acknowledgements

This paper was greatly improved by many insightful conversations with Oldrich Hungr. We are grateful to two anonymous reviewers for providing thorough and constructive comments that improved the manuscript.

Funding information Financial support for this work was provided by a graduate scholarship given by the Natural Science and Engineering Research Council of Canada (NSERC), as well as scholarships given by the Department of Earth, Ocean and Atmospheric Sciences at the University of British Columbia.

Appendix A: Example Calculation

To illustrate the methodology introduced in “Methods,” consider the example of the back-analysis of the Mt. Meager case history presented by McDougall (2017). For clarity, we will ignore the impact area feature in this example. Therefore, the vector of features based on the field investigation comprises the values of the two point estimates of velocity: $y^F = (62 \frac{m}{s}, 73 \frac{m}{s})^T$. For the present example, we will use the model results obtained when the parameter vector is $\mathbf{b} = (f = 0.05, \xi = 500 \frac{m}{s^2})^T$, where model parameters that are not being calibrated have been omitted. When Dan3D is run using these parameters, the simulated velocities can be compiled in the vector $y^M(\mathbf{b}) = (70 \frac{m}{s}, 75 \frac{m}{s})^T$. By differencing $y^M(\mathbf{b})$ and y^F , we can create the error vector $r(\mathbf{b}) = (8 \frac{m}{s}, 2 \frac{m}{s})^T$.

To evaluate Eqs. (5) and (8), the covariance matrix Σ must be specified. Selection of the covariance matrix is discussed in “Selection of standard deviation of a measurement,” but for now assume that the covariance matrix is given by:

$$\Sigma = \begin{pmatrix} \left(5 \frac{m}{s}\right)^2 & 0 \\ 0 & \left(5 \frac{m}{s}\right)^2 \end{pmatrix}$$

Given the normality assumption, the interpretation of this covariance matrix for the first super elevation bend is as follows. We believe the best estimate of the velocity is 63m/s; however, there is some measurement error associated with this estimate. We are 70% sure that the velocity is between 57 and 67m/s, and 95% sure that the velocity is between 52 and 72m/s. This measurement error is quantified by using a standard deviation of 5m/s. The results of evaluating the likelihood function (Eq. (4)) and the GML objective function (Eq. (5)) for this parameter combination are values of $1.6e-3$ and 2.72 , respectively.

References

- Aaron J (2017) Advancement and calibration of a 3D numerical model for landslide runout analysis, PhD Thesis, University of British Columbia
- Aaron J, Hungr O (2016a) Dynamic simulation of the motion of partially-coherent landslides. *Eng Geol* 205:1–11. <https://doi.org/10.1016/j.enggeo.2016.02.006>
- Aaron J, Hungr O (2016b) Dynamic analysis of an extraordinarily mobile rock avalanche in the Northwest Territories, Canada. *Can Geotech J* 53(6):899–908. <https://doi.org/10.1139/cgj-2015-0371>
- Aaron J, McDougall S, Moore JR, Coe JA, Hungr O (2017) The role of initial coherence and path materials in the dynamics of three rock avalanche case histories. *Geoenviron Dis* 4(1):5. <https://doi.org/10.1186/s40677-017-0070-4>
- Allstadt K (2013) Extracting source characteristics and dynamics of the August 2010 Mount Meager landslide from broadband seismograms. *J Geophys Res Earth Surf* 118:1472–1490. <https://doi.org/10.1002/jgrf.20110>
- BGC Engineering Inc. (2000) Geotechnical hazard assessment of the south flank of Frank Slide, Hillcrest, Alberta. Report to Alberta Environment, No 00–0153
- Boulbee N (2005) Characterization of the Zymoetz River Rock Avalanche, Masters Thesis. Simon Fraser University

- Boulton N, Stead D, Schwab J, Geertsema M (2006) The Zymoetz River rock avalanche, June 2002, British Columbia, Canada. *Eng Geol* 83(1–3):76–93. <https://doi.org/10.1016/j.enggeo.2005.06.038>
- Brezzi L, Bossi G, Gabrieli F, Marcato G, Pastor M, Cola S (2016) A new data assimilation procedure to develop a debris flow run-out model. *Landslides* 13(5):1083–1096. <https://doi.org/10.1007/s10346-015-0625-y>
- Calvello M, Cuomo S, Ghasemi P (2017) 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>
- Cepeda J, Chávez JA, Cruz Martínez C (2010) Procedure for the selection of runout model parameters from landslide back-analyses: application to the metropolitan area of San Salvador, El Salvador. *Landslides* 7(2):105–116. <https://doi.org/10.1007/s10346-010-0197-9>
- Cruden DM, Hungr O (1986) The debris of the Frank Slide and theories of rockslide–avalanche mobility. *Can J Earth Sci* 23(3):425–432. <https://doi.org/10.1139/e86-044>
- Cruden D, Krahn J (1978) Frank Rockslide, Alberta, Canada. In: Voight B (ed) *Rockslides and avalanches, Vol 1 Natural phenomena*. Elsevier Scientific Publishing, Amsterdam, pp 97–112
- Doherty J (2010) PEST, model-independent parameter estimation (Vol. 2005). Watermark Numerical Computing
- Evans SG, Hungr O, Enegegn E (1994) The Avalanche Lake rock avalanche, Mackenzie Mountains, Northwest Territories, Canada: description, dating and dynamics. *Can Geotech J* 31(5):749–768
- Galas S, Dalbey K, Kumar D, Patra A, Sheridan M (2007) Benchmarking TITAN2D mass flow model against a sand flow experiment and the 1903 Frank Slide. In: Ho K, Li V (eds) *2007 International forum on landslide disaster management*. Geotechnical Division, The Hong Kong Institution of Civil Engineers, Hong Kong, pp 899–917
- Gregory P (2010) Bayesian logical data analysis for the physical sciences: a comparative approach with Mathematica® support. Cambridge University Press, Cambridge
- Heiser M, Scheidl C, Kaitna R (2017) Evaluation concepts to compare observed and simulated deposition areas of mass movements. *Comput Geosci* 21(3):335–343. <https://doi.org/10.1007/s10596-016-9609-9>
- Hsieh WW (2009) *Machine learning methods in the environmental sciences: neural networks and kernels*. Cambridge University Press, Cambridge
- Hungr O (1995) A model for the runout analysis of rapid flow slides, debris flows and avalanches. *Can Geotech J* 32(4):610–623. <https://doi.org/10.1139/t95-063>
- Hungr O (2016) A review of landslide hazard and risk assessment methodology. In: Aversa S, Cascini L, Picarelli L, Scavia C (eds) *Landslides and engineered slopes. Experience, theory and practice: Proceedings of the 12th International Symposium on Landslides (Napoli, Italy, 12–19 June 2016)*. CRC press, London, pp 3–27
- Hungr O (2017) Chapter 9 - runout analysis. In: Hawley M, Cunniff J (eds) *Guidelines for mine waste dump and stockpile design*. CSIRO Publishing
- Hungr O, Evans SG (2004) Entrainment of debris in rock avalanches: an analysis of a long run-out mechanism. *Geol Soc Am Bull* 116(9):1240–1252. <https://doi.org/10.1130/B25362.1>
- Hungr O, McDougall S (2009) Two numerical models for landslide dynamic analysis. *Comput Geosci* 35(5):978–992. <https://doi.org/10.1016/j.cageo.2007.12.003>
- Hungr O, Morgan GC, Kellerhals R (1984) Quantitative analysis of debris torrent hazards for design of remedial measures. *Can Geotech J* 21(4):663–677. <https://doi.org/10.1139/t84-073>
- Hungr O, Dawson RF, Kent A, Campbell D, Morgenstern NR (2002) Rapid flow slides of coal-mine waste in British Columbia, Canada. In: Evans SG, DeGraff JV (eds) *Catastrophic landslides: effects, occurrence, and mechanisms, vol XV*. Geological Society of America Reviews in Engineering Geology, Boulder, pp 191–208
- Hungr O, Corominas J, Eberhardt E (2005) Estimating landslide motion mechanism, travel distance and velocity—state of the art report. In: Hungr O, Fell R, Couture R, Eberhardt E (eds) *Landslide risk management*. A.A. Balkema, Amsterdam, pp 99–128
- Hungr O, Morgenstern NR, Wong HN (2007) Review of benchmarking exercise on landslide debris runout and mobility modelling. In: Ho, Li (eds) *Proceedings of the International Forum on Landslide Disaster Management*. Hong Kong Geotechnical Engineering Office, pp 775–812
- Hungr O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. *Landslides* 11(2):167–194
- Keaton JR, Wartman J, Anderson S, Benoit J, DeLaChapelle J, Gilbert R, Montgomery DR (2014) The 22 March 2014 Oso Landslide, Snohomish County, Washington. Geotechnical extreme event reconnaissance association report GEER-036. <https://doi.org/10.18118/G6V884>
- Körner HJ (1976) Reichweite und Geschwindigkeit von Bergstürzen und fleisschnee-lawinen. *Rock Mech* 8:225–256
- Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Q Appl Math* 2(2):164–168
- Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11(2):431–441
- McDougall S (2006) A new continuum dynamic model for the analysis of extremely rapid landslide motion across complex 3D terrain, PhD Thesis, University of British Columbia
- McDougall S (2017) 2014 Canadian Geotechnical Colloquium: landslide runout analysis—current practice and challenges. *54(5):605–620*. <https://doi.org/10.1139/cgj-2016-0104>
- McDougall S, Hungr O (2004) A model for the analysis of rapid landslide motion across three-dimensional terrain. *Can Geotech J* 41(6):1084–1097. <https://doi.org/10.1139/t04-052>
- McDougall S, Boulton N, Hungr O, Stead D, Schwab JW (2006) The Zymoetz River landslide, British Columbia, Canada: description and dynamic analysis of a rock slide–debris flow. *Landslides* 3(3):195–204. <https://doi.org/10.1007/s10346-006-0042-3>
- Nocedal J, Wright SJ (2006) Numerical optimization. In: Mikosch TV, Resnick SI, Robinson SM (eds) *Springer series in operations research and financial engineering*, 2nd edn. Springer, New York
- Prochaska AB, Santi PM, Higgins JD, Cannon SH (2008) A study of methods to estimate debris flow velocity. *Landslides* 5(4):431–444. <https://doi.org/10.1007/s10346-008-0137-0>
- Read RS, Langenberg W, Cruden D, Field M, Stewart R, Bland H, Chen Z, Froese CR, Cavers DS, Bidwell AK, Murray C, Anderson WS, Jones A, Chen J, McIntyre D, Kenway D, Bingham DK, Weir-Jones I, Seraphim J, Freeman J, Spratt D, Lamb M, Herd E, Martin D, McLellan P, Pana D (2005) Frank Slide a century later: the Turtle Mountain monitoring project. In: Hungr O, Fell R, Couture RR, Eberhardt (eds) *Landslide risk management*. Balkema, Rotterdam, pp 713–723
- Schwab JW, Geertsema M, Evans SG (2003) Catastrophic rock avalanches, west-central B.C., Canada. In *3rd Canadian Conference on Geotechnique and Natural Hazards*, Edmonton, pp 252–259
- Sosio R, Crosta GB, Hungr O (2008) Complete dynamic modeling calibration for the Thurwieser rock avalanche (Italian Central Alps). *Eng Geol* 100(1–2):11–26. <https://doi.org/10.1016/j.enggeo.2008.02.012>
- Varnes DJ (1978) Slope movement types and processes. In: Schuster RL, Krizek RJ (eds) *Landslides, analysis and control, special report 176: transportation research board*. National Academy of Sciences, Washington, DC, pp 11–33
- Voellmy A (1955) Über die Zerstörungskraft von Lawinen. *Schweizerische Bauzeitung* 73:212–285
- White JL, Morgan ML, Berry KA (2015) The West Salt Creek landslide: a catastrophic rockslide and rock/debris avalanche in Mesa County, Colorado. *Bulletin 55*. Colorado geological survey, Golden, Colorado. Retrieved from <http://store.coloradogeologicalsurvey.org/product/the-west-salt-creek-landslide-a-catastrophic-rockslide-and-rockdebris-avalanche-in-mesa-county-colorado/>

J. Aaron

Department of Earth Sciences, Geological Institute, ETH Zürich, Sonneggstrasse 5, 8048, Zürich, Switzerland
Email: jordan.aaron@erdw.ethz.ch

S. McDougall

Department of Earth, Ocean and Atmospheric Sciences, The University of British Columbia, Vancouver, Canada
e-mail: smcdouga@eoas.ubc.ca

N. Nolde

Department of Statistics, The University of British Columbia, Vancouver, Canada
e-mail: natalia@stat.ubc.ca