https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

A surrogate-assisted preliminary screening workflow for hypothetical fixed-location check-dam height scenarios in debris-flow mitigation

Jun Katagiri[1], Hidetaka Saomoto[1], and Takayuki Shinohara[1]

[1]Integrated Research Center for Resilient Infrastructure, National Institute of Advanced Industrial Science and Technology, Higashi 1-1-1, Tsukuba, Ibaraki, Japan 305-8567

Correspondence: Jun Katagiri (j-katagiri@aist.go.jp)

Abstract. Early-stage debris-flow mitigation planning often requires comparison of multiple countermeasure scenarios before detailed engineering design information is available. This study presents a surrogate-assisted preliminary screening workflow for comparing hypothetical fixed-location check-dam height scenarios using simplified proxy metrics. A depth-integrated debris-flow simulation model was applied to a terrain dataset from the area affected by the 2021 Atami debris-flow disaster in Japan. The analysis is not intended as a forensic reproduction of the event or as an assessment of existing or planned facilities. In total, 4,877 simulation cases were generated by varying the heights of six candidate check dams using Latin hypercube sampling. The simulation results were used to train surrogate models relating dam-height scenarios to the maximum downstream debris-flow depth, used here as a hazard-intensity proxy, while the sum of dam heights was used as a height-based proxy for construction-related effort. A multilayer perceptron provided the best coarse-grained approximation and enabled repeated surrogate-based searches under different trade-off weights. The screened scenarios formed a limited number of representative trade-off regimes. However, physics-based re-evaluation showed non-negligible discrepancies between surrogate-predicted and simulated depths, indicating that the workflow should be used for candidate screening rather than final quantitative design evaluation.

1 Introduction

Debris flows are among the most destructive hazards in mountainous regions. Their rapid onset and high destructive potential pose serious threats to downstream communities and infrastructure. Recent studies have suggested that rainfall-triggered slope failures, including debris flows and landslides, may become more frequent or intense under changing climatic conditions, further amplifying disaster risk in vulnerable regions (IPCC, 2021; Crozier, 2010; Gariano and Guzzetti, 2016).

Extensive research has been devoted to understanding and modeling debris-flow processes through laboratory experiments, field observations, and numerical simulations. In particular, physically based models, including depth-integrated, viscoplastic, and two-phase formulations, have significantly advanced the mechanistic understanding and predictive capability of debris-flow dynamics (Iverson, 1997; Takahashi, 2007; Pastor et al., 2009; Pudasaini, 2012). These models provide an important basis for evaluating debris-flow behavior and the potential performance of mitigation measures.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

However, the computational cost of such detailed numerical models often limits their direct use in engineering decision-making problems involving multiple design variables, such as the configuration and sizing of check-dam systems within a catchment. In practical mitigation planning, engineers and decision-makers must compare many design alternatives while balancing protective performance, construction effort, and site-specific constraints. Repeated full-scale simulations for all candidate configurations are often impractical in this context. This difficulty becomes more severe when multiple check dams are considered simultaneously, because systematic comparison of many candidate height combinations quickly becomes computationally prohibitive when each candidate design must be evaluated using a high-fidelity simulation model.

To address this challenge, surrogate modeling has been increasingly used to approximate simulation-derived response metrics with substantially reduced computational cost. Surrogate-assisted approaches have been applied in various engineering and environmental decision-making problems, including water resources management and infrastructure design (Forrester et al., 2008; Razavi et al., 2012; Giustolisi et al., 2015). More recently, surrogate models and machine-learning techniques have also been used to accelerate hazard assessment and simulation-based screening under limited computational budgets (Qiao and Myers, 2022; Lei et al., 2025; Yang et al., 2025; Ma et al., 2025; Hu et al., 2025). These developments suggest that surrogate modeling can be useful not only for prediction-oriented tasks, but also for systematic screening of candidate scenarios when direct simulation is too expensive for repeated evaluation.

In this study, we investigate a surrogate-assisted preliminary screening workflow for comparing hypothetical fixed-location check-dam height scenarios for debris-flow mitigation. Here, "layout" is used in a restricted sense: dam locations are fixed *a priori* based on practical planning considerations, and the screening focuses on the height configuration of the multiple dams. This simplification is adopted to keep the problem computationally tractable while enabling systematic comparison of height-scenario trade-offs under a realistic high-fidelity simulation budget; optimization of dam placement itself is outside the scope of this study and is left for future work. Protective performance is represented by the maximum debris-flow depth at a downstream evaluation section, which serves as a simplified indicator of hazard intensity. Because detailed cost information was not available, the sum of dam heights is used as a simplified proxy for construction-related effort, acknowledging that detailed cost models are typically held by infrastructure operators and were not available to the authors. These proxy metrics are not intended to fully capture real-world decision criteria; rather, they provide representative indicators for systematic comparison of alternative scenarios. When more detailed performance metrics (e.g., impact forces on structures) or refined cost models are available, the proposed workflow can be adapted by substituting the corresponding proxy metrics. This demonstration is not intended to reproduce the observed 2021 Atami debris-flow event, including its runout extent, arrival time, deposition pattern, or damage distribution. It is also not intended to assess the adequacy of existing or planned check-dam facilities in Atami City. Rather, the Atami terrain is used as a representative real-terrain setting for examining a simplified preliminary screening workflow under realistic topographic conditions.

By analyzing the distribution of screening-derived trade-off scenarios and the variability of height variables around them, we show how such information can support planning-stage scenario comparison and help identify variables that merit closer examination in subsequent physics-based analyses.

Figure 1 provides an overview of the preliminary screening workflow adopted in this study.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

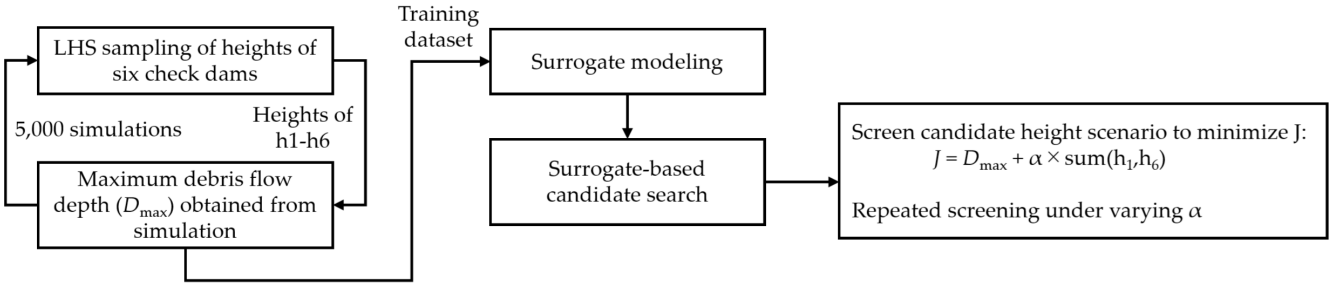

Figure 1. Overview of the preliminary screening workflow. Six check-dam heights are sampled by Latin hypercube sampling, evaluated with iRIC–Morpho2DH simulations, approximated by a surrogate model, and screened by particle-swarm search under different weighting parameters α.

2 Methods

2.1 Debris-flow simulation framework

In this study, debris-flow simulations were conducted using the free and open-source river analysis software package iRIC version 4. iRIC is an integrated simulation environment that combines multiple numerical solvers with a graphical user interface, and has been widely used in river engineering and sediment-related studies. Among the solvers implemented in iRIC, the Morpho2DH solver was employed to simulate debris-flow and mudflow behavior.

Morpho2DH is based on depth-integrated governing equations formulated in a generalized curvilinear coordinate system, and has been applied to debris-flow simulations in previous studies, including post-event reproduction analyses (Takebayashi et al., 2022; Takebayashi, 2023). The numerical formulation and implementation details of Morpho2DH are described in the official solver manual (iRIC Software, 2025).

2.2 Real-terrain demonstration setting and topographic data

A terrain dataset from the area affected by the July 3, 2021 Atami debris-flow disaster in Shizuoka Prefecture, Japan, was used as a real-terrain demonstration setting. The disaster was triggered by rainfall that persisted almost continuously from July 1, 2021. A large debris flow was initiated by the failure of a man-made fill slope in the upstream catchment of the Aizome River in the Izusan area of Atami City, where the slope had been developed around 2010. As of October 1, 2021, severe impacts were reported, including 26 fatalities and damage to 128 houses (Imaizumi et al., 2022). This disaster information is provided only to identify the geomorphic and hazard context of the demonstration terrain. The simulations are not calibrated against observed runout, arrival time, deposition, or damage records, nor are they intended to reproduce the observed event sequence. The demonstration terrain and computational domain are illustrated in Figure 2.

Topographic data were obtained from a 5-m resolution digital elevation model (DEM) provided by the Center for Spatial Information Science, The University of Tokyo (Center for Spatial Information Science, The University of Tokyo, 2025). For the

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

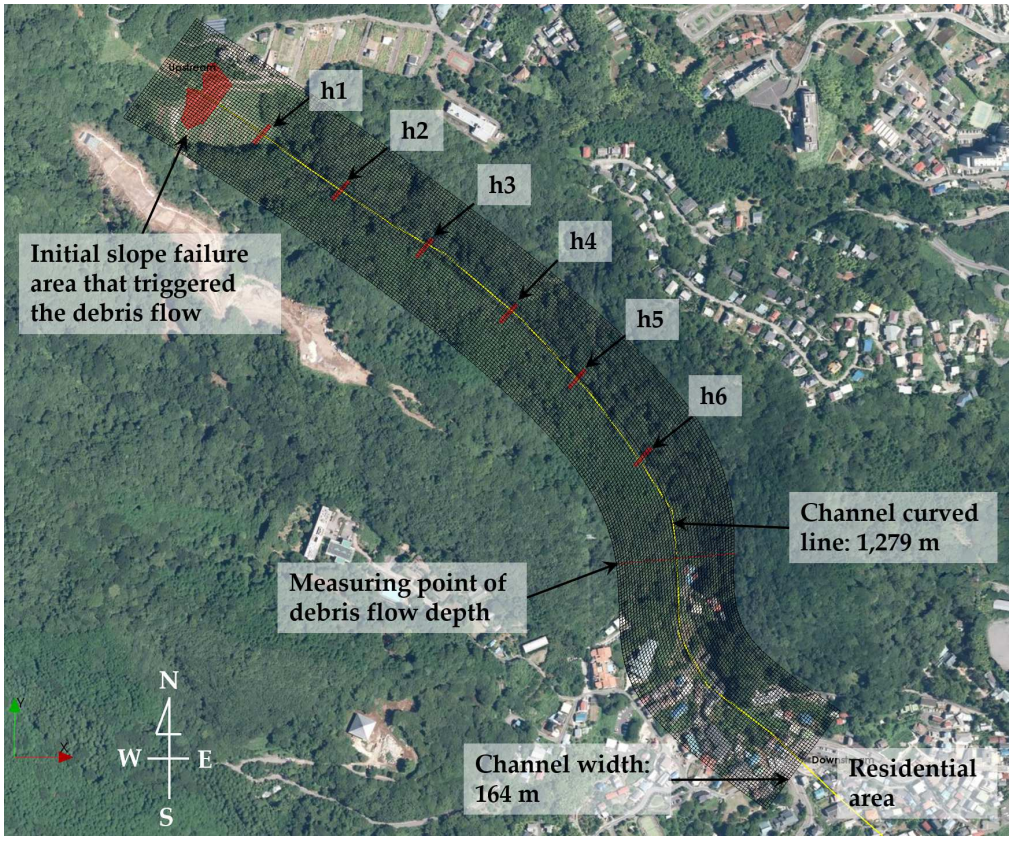

Figure 2. Demonstration terrain and computational setup. The red polygon denotes the approximated source area, h_1–h_6 denote hypothetical candidate check-dam locations, and the downstream point denotes the proxy section used to compare hazard-intensity scenarios. Background imagery: GSI Maps "Photo" layer, Geospatial Information Authority of Japan, accessed on 2026-02-04.

numerical demonstration, an initial source-area polygon was approximated from elevation differences between pre-event and post-event terrain data. The adopted source-area thickness and simulation parameters should therefore be regarded as simplified modelling assumptions rather than a calibrated reconstruction of the 2021 event. It is noted that iRIC allows direct import of terrain data from publicly available datasets such as the Geospatial Information Authority of Japan and SRTM through its built-in data acquisition functions.

2.3 Computational domain and simulation conditions

The computational domain was defined along the main torrent from the vicinity of the approximated source area to a downstream zone near residential exposure. The downstream evaluation section was selected as a proxy location for comparing scenarios and was not intended to reproduce observed damage or official hazard zoning. A curvilinear centerline with a length

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

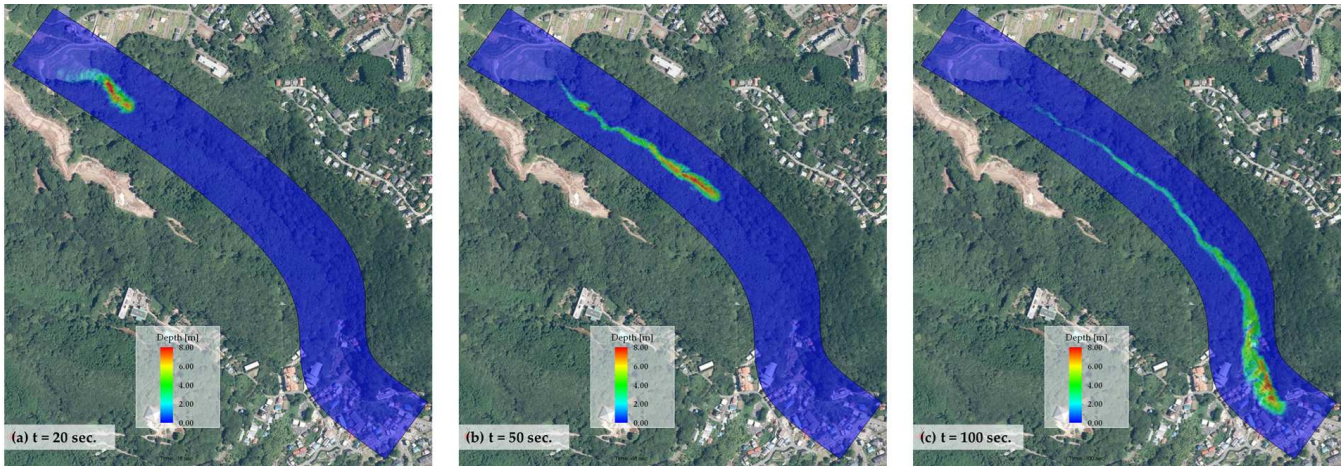

Figure 3. Simulated debris-flow depth for a representative case at (a) 20 s, (b) 50 s, and (c) 100 s after the start of the numerical release scenario. All panels use the same colour scale. Background imagery: GSI Maps "Photo" layer, Geospatial Information Authority of Japan, accessed on 2026-02-04.

of approximately 1279 m was traced along the channel, and a lateral extent of 82 m on both sides of the centerline was included in the computational domain.

The domain was discretized with an average grid spacing of approximately 4 m in both the longitudinal and transverse directions, resulting in a total of 13,398 grid cells. The simulation time step was set to 0.004 s, which was determined through preliminary trials to ensure numerical stability while maintaining computational efficiency. Each simulation was continued for a duration of 100 s. An example of the simulated debris-flow evolution is shown in Figure 3. These snapshots illustrate how hypothetical fixed-location height scenarios influence the temporal evolution and spatial attenuation of debris-flow depth, providing intuitive physical support for the surrogate-based analysis.

For several representative combinations of check dam heights, the simulation duration was confirmed to be sufficient to capture the main debris-flow behavior. However, it cannot be completely excluded that, under certain conditions, debris flow may not have reached the downstream area within this time frame. This limitation could be further examined by extending the simulation duration and is noted as a topic for future investigation.

The parameters used in the debris flow simulation are summarized in Table 1. These parameter values were adopted from the default settings commonly used in Morpho2DH, with reference to previous studies (Takebayashi et al., 2022; iRIC Software, 2025).

2.4 Generation of simulation cases

To construct a surrogate model described later, multiple simulation cases were generated by varying the heights of six check dams, denoted as h_1 to h_6. Each dam height was randomly sampled within the range of 0 to 12 m using Latin Hypercube Sampling (LHS) (McKay et al., 1979).

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

Table 1. Simulation parameters used in debris-flow simulation

Parameter	Value	Unit
Simulation time	100	s
Time step	0.004	s
Density of water	1,000	kg/m^3
Density of sediment	2,650	kg/m^3
Static sediment concentration	0.6	–
Fraction of sediment behaving as fluid	0.2	–
Minimum flow depth	0.01	m
Internal friction angle	34	deg
Laminar-layer thickness	constant	–
Ratio of laminar-layer thickness	0.4	–
Resistance coefficient	72	–
Bed material type	uniform sand	–
Mean grain size	0.01	m
Maximum erosion depth	2	m
Thickness of initial failure area	0.1	m

In this study, six hypothetical candidate locations were placed at approximately equal intervals along the torrent, with one location spatially referenced to an existing check-dam site for orientation. The resulting configuration does not represent an official design plan, a proposed construction project, or an evaluation of existing facilities. This configuration was adopted to control the dimensionality of the scenario space and maintain computational tractability, allowing the proposed surrogate-assisted screening workflow to be examined as a proof-of-concept rather than as an exhaustive scenario study.

A total of 5000 simulation cases were automatically executed using custom scripts, and the simulation results were exported as CSV files. Due to numerical instability or unexpected termination in a limited number of cases, valid simulation outputs were obtained for 4877 cases. These valid cases were used as the dataset for surrogate model construction.

A typical debris-flow simulation case required approximately 5 minutes on a workstation equipped with an Intel Core i9-13900 CPU. The full set of 5,000 simulation cases was executed in parallel by distributing the runs across one desktop PC and two laptop PCs, and the overall wall-clock time to complete all cases was approximately one week.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

2.5 Surrogate model construction

Using the simulation dataset, a surrogate model was constructed to approximate the relationship between the dam height vector $(h_1, h_2, \ldots, h_6)$ and the maximum debris-flow inundation depth $D_{\max}$. The dam heights were treated as input features, and $D_{\max}$ was treated as the output variable.

Several surrogate modeling approaches were examined, including conventional response surface methods such as polynomial regression, Kriging models, and machine-learning-based methods such as multilayer perceptrons (MLP), support vector regression (SVR), LightGBM, and XGBoost.

The polynomial response surface model (RSM) was trained using a pipeline consisting of standardization, polynomial feature expansion, and linear regression. Candidate polynomial degrees from 1 to 8 were examined. The degree was selected based on the in-sample Akaike information criterion (AIC), and the predictive performance of the selected model was then evaluated using 4-fold cross-validation (shuffle=True, seed=42) with RMSE, MAE, and R^2.

For the comparative evaluation of machine-learning surrogate models, the six dam-height variables (h_1–h_6) were used as inputs and $D_{\max}$ as the output. All models were evaluated using a common 4-fold cross-validation scheme (shuffle=True, seed=42), and the performance was assessed by RMSE, MAE, and R^2. In this comparative stage, the MLP model was implemented in a pipeline with input standardization (`StandardScaler`) and `MLPRegressor(max_iter=2000, random_state=0)`. As comparison baselines, we evaluated XGBoost and LightGBM (both with 300 estimators) and SVR under the same evaluation protocol.

For the subsequent surrogate-based candidate search, the selected MLP surrogate was retrained using a two-hidden-layer fully connected network with 50 and 50 neurons (activation: ReLU; L2 regularization coefficient: 10^{-4}; learning rate: 10^{-3}; maximum iterations: 4000; random seed: 0), again after input standardization. Because algorithm benchmarking was not the objective of this study, we avoided extensive hyperparameter searches.

Details of the comparative performance of these models are discussed in the Results section.

2.6 Screening objective and proxy metrics

In this study, a scalar screening score J was defined by combining two simplified proxy metrics: the maximum debris-flow inundation depth $D_{\max}$, used as a hazard-intensity proxy, and a construction-effort proxy C, represented by the sum of check-dam heights. Because detailed cost information was not available, C was used as a minimal proxy for relative construction-related effort under the simplified scenario setting.

The screening score was defined as

$$J = D_{\max} + \alpha C, \tag{1}$$

where α is a weighting parameter that controls the trade-off between the downstream-depth proxy and the height-sum proxy. In the present study, α was varied over the range from 0 to 0.5. The surrogate-based search was formulated to identify candidate height combinations with low values of J. This scalarized formulation was adopted for exploratory screening, aiming

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

to organize candidate scenarios under simplified proxy metrics rather than to provide a final optimal design or quantitative engineering decision.

Note that negative values of $D_{\max}$ estimated from the surrogate model were clipped to zero during the surrogate-based search because such values were physically meaningless.

2.7 Surrogate-based candidate search

Particle Swarm Optimization (PSO; Kennedy and Eberhart, 1995) was used as a practical search algorithm for generating candidate scenarios in the surrogate-proxy space. Although various general-purpose search algorithms, such as Bayesian optimization and genetic algorithms, are available, the role of PSO in this study was not to determine a final optimal design, but to sample candidate scenarios under varying trade-off weights.

The PSO hyperparameters were set based on preliminary trial-and-error tests. Specifically, the swarm size was set to 60 particles and the maximum number of iterations to 300. The inertia weight was $w = 0.72$, and the acceleration coefficients for the personal-best and global-best terms were both set to 1.49 ($c_1 = c_2 = 1.49$). The random seed was fixed to 42 for reproducibility. At each iteration, the design variables were constrained within $0 \leq h_i \leq 12$. When the surrogate prediction yielded a negative value, it was clipped to $D_{\max} = 0$ as a physical constraint before computing the screening score J.

Multiple values of α were independently examined to obtain a set of surrogate-screened candidate scenarios. The parameter α was varied over the range from 0 to 0.5 using 200 evenly spaced values, resulting in 200 independent surrogate-based search runs. By leveraging the computational efficiency of the surrogate model, repeated surrogate evaluations and search runs were conducted to investigate the sensitivity of surrogate-screened candidate scenarios to changes in the trade-off parameter.

Accordingly, the purpose of the surrogate-based search in this study is not to identify a single final scenario, but rather to organize candidate scenarios and examine their sensitivity under varying trade-off weights.

3 Results

3.1 Performance of surrogate models

Table 2 summarizes the predictive performance of the surrogate models evaluated in this study. The models were compared using cross-validated root mean square error (RMSE) and coefficient of determination (R^2).

Among the tested models, the multilayer perceptron (MLP) exhibited the highest predictive accuracy, achieving an RMSE of 0.99 m and an R^2 value of 0.73. Support vector regression (SVR) and tree-based ensemble models showed slightly lower accuracy, while the Kriging model provided comparable but inferior performance. In contrast, the polynomial response surface model demonstrated extremely poor predictive capability, indicating that the relationship between dam heights and debris-flow depth is highly nonlinear and not well represented by low-order polynomial functions. Based on these results, the MLP model was selected as the surrogate model for subsequent surrogate-based candidate-search analyses.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

Table 2. Cross-validated performance of surrogate models.

Model	RMSE (m)	R^2
MLP	0.99	0.73
SVR	1.03	0.71
LightGBM	1.08	0.68
Kriging	1.09	0.67
XGBoost	1.14	0.64
Polynomial RSM (8th order)	8.81	-20.50

Once trained, the surrogate model provides an estimate of $D_{\max}$ for arbitrary inputs of h_1–h_6 with virtually no waiting time. In contrast, as noted above, a single debris-flow simulation requires about 5 minutes. Although the surrogate is an approximation and therefore involves modeling errors, the substantial reduction in evaluation time is a practical advantage because it enables extensive scenario screening. In particular, the near-instantaneous surrogate evaluations make it feasible to perform many surrogate-based search runs under varying α conditions, as described in the following sections.

3.2 Trade-off structure in the surrogate-based screening results

Figure 4 illustrates the relationship between the downstream-depth proxy, represented by the maximum flow depth $D_{\max}$, and the height-sum proxy C, defined as the sum of dam heights. Each black marker corresponds to a surrogate-screened candidate scenario obtained for a specific value of the weighting parameter α in the screening score.

Although a single scalar screening objective was employed, varying α produced a set of scenarios that were non-dominated with respect to the two surrogate-based proxy metrics. Five representative non-dominated scenarios were identified in the surrogate-proxy space, indicating that the search repeatedly converged to a limited number of characteristic height combinations despite the continuous variation of α.

This result suggests that the surrogate-screened scenarios can be grouped into a limited number of representative trade-off regimes under the present proxy definitions, rather than forming a continuous spectrum of candidate scenarios. For detailed re-simulation and spatial comparison, three representative scenarios (P1–P3) were selected from these five non-dominated scenarios, as shown in Fig. 4. We then conducted the debris-flow simulations using the corresponding combinations of the hypothetical check dam heights.

3.3 Post-screening full-simulation check of representative scenarios

Because surrogate-based search can exploit local approximation errors, selected candidate scenarios should be checked using the original physics-based simulation before any quantitative interpretation. To illustrate this point, three representative surrogate-screened scenarios were re-simulated using the full debris-flow model.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

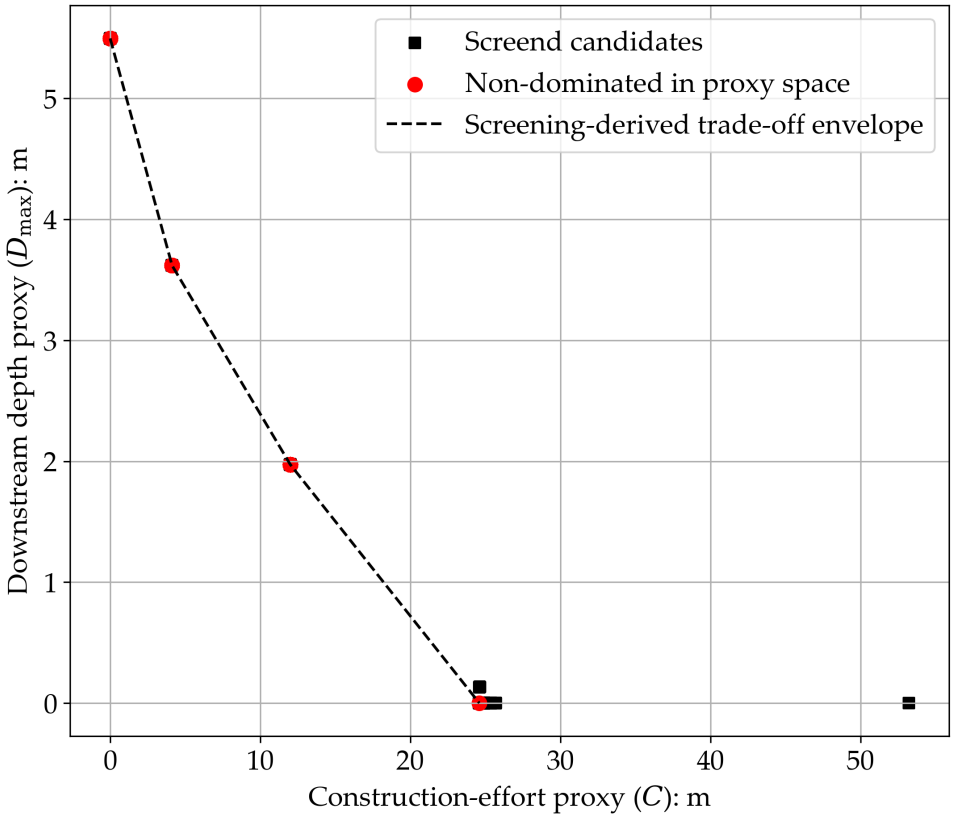

Figure 4. Screening-derived trade-off relationship between the downstream-depth proxy ($D_{\max}$) and the height-sum proxy (C). Red markers denote scenarios that are non-dominated in the surrogate-proxy space.

For the three representative surrogate-screened scenarios (P1–P3), the surrogate-predicted $D_{\max}$ values were compared with the corresponding values obtained from full debris-flow simulations (Table 3). For clarity, the surrogate-predicted value of $D_{\max}$ for P3 is reported as 0.0 because negative surrogate predictions were clipped to zero during the surrogate-based search. The surrogate systematically underestimated $D_{\max}$ for these representative scenarios, with the largest discrepancy observed for the low-$D_{\max}$-proxy scenario P3. This discrepancy is a central observation of the present preprint: surrogate-based search can be useful for organizing candidate scenarios, but the resulting candidates cannot be interpreted quantitatively without subsequent physics-based re-evaluation. In particular, P3 was predicted as nearly zero by the surrogate after clipping, whereas the full simulation still produced a non-negligible downstream depth. This illustrates why the workflow should be used for preliminary screening rather than final design evaluation.

Figure 5 compares the water-depth distributions for P1–P3 under the same time instant ($t = 100$ s), computational domain, and color scale. From P1 (low height-sum proxy) to P3 (low-$D_{\max}$-proxy), the extent and localization of high-depth zones in

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

Table 3. Post-screening full-simulation check of three representative surrogate-screened scenarios. The discrepancies between surrogate-predicted and simulated $D_{\max}$ values illustrate the need for physics-based re-evaluation before quantitative interpretation.

Point	Type	α	$\sum h_i$ (m)	Surrogate $D_{\max}$ (m)	Simulated $D_{\max}$ (m)	Abs. error (m)
P1	low height-sum proxy	0.4573	0.0000	5.4969	7.6776	2.1807
P2	intermediate	0.4397	4.1270	3.6187	7.0658	3.4471
P3	low-$D_{\max}$-proxy	0.1206	24.6025	0.0000	3.7626	3.7626

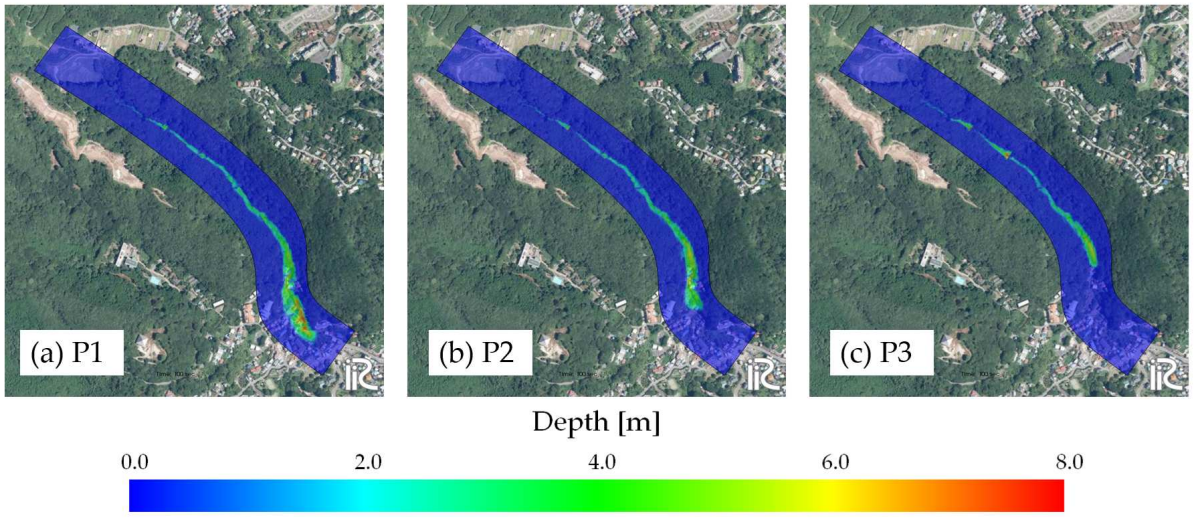

Figure 5. Simulated debris-flow depth distributions at $t = 100$ s for three representative surrogate-screened scenarios: (a) P1, (b) P2, and (c) P3. All panels use the same colour scale and computational domain. Background imagery: GSI Maps "Photo" layer, Geospatial Information Authority of Japan, accessed on 2026-02-04.

the downstream area change, indicating spatial differences that are not fully captured by the scalar trade-off between $D_{\max}$ and the height-sum proxy ($C = \sum_i h_i$). This comparison suggests that, in scenario screening, it is useful to examine simulated spatial patterns of flow and inundation in addition to scalar proxy metrics.

3.4 Variability of screened height scenarios

Figure 6 shows the distribution of dam-height configurations obtained around a representative screening-derived trade-off scenario. Gray lines represent individual surrogate-screened candidate scenarios converging to the same non-dominated point, while the red line indicates the representative screening-derived trade-off scenario.

The variability of dam heights differs markedly among locations. Upstream dams (e.g., h_1 and h_2) exhibit relatively large variability, suggesting that these height variables are less tightly constrained within the present surrogate-proxy setting. In

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

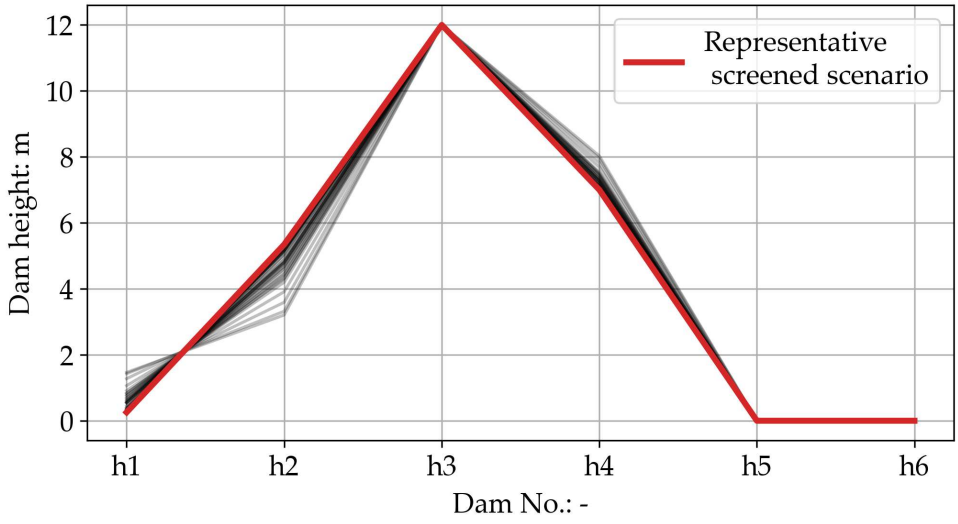

Figure 6. Candidate height combinations associated with a representative screened scenario. Gray lines indicate surrogate-search results with similar proxy values; the red line denotes the selected representative scenario.

contrast, the midstream dam (h_3) shows consistently low variability, indicating that this height variable repeatedly converged to a narrower range in the surrogate-based search.

This pattern suggests that some height variables are more influential than others within the present surrogate-proxy setting, whereas others appear less tightly constrained in the screened scenarios. Such information may be useful for planning-stage discussion because variables with low variability can be flagged for closer examination in subsequent physics-based simulations or site-specific engineering studies.

4 Discussion

4.1 Positioning of the surrogate model and interpretation of coarse-graining

In recent years, an increasing number of studies have combined numerical simulations with surrogate models or machine learning techniques in the field of natural hazards (e.g., (Qiao and Myers, 2022; Lei et al., 2025)). Many of these studies primarily aim to improve predictive accuracy for hazard indicators such as inundation depth, landslide displacement, or slope stability.

In contrast, the surrogate model constructed in this study was not intended to reproduce detailed physical processes with high fidelity. The target output, the maximum debris-flow depth $D_{\max}$, was approximated from dam-height configurations to capture overall trends in the scenario space. It should be noted that the debris-flow simulation itself is based on depth-integrated governing equations, in which complex processes such as flow–structure interaction, deposition, and re-entrainment

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

are already simplified to some extent (Pastor et al., 2009). Accordingly, the surrogate model should be regarded as a coarse-grained representation of the design–response relationship rather than a high-precision predictive tool.

From this perspective, the moderate prediction accuracy obtained in this study (approximately $R^2 \approx 0.7$) is considered acceptable, as the primary objective is not exact prediction but preliminary scenario comparison and characterization of trade-off tendencies. The repeated convergence of surrogate-based search results toward similar screening-derived trade-off scenarios across a wide range of α values suggests that the identified trade-off structure appears informative for global design-space screening, despite local surrogate approximation errors.

The simulation duration was fixed at 100 s based on preliminary runs without check dams, which indicated that debris flow can reach the downstream evaluation section within this time window. Nevertheless, for some height combinations, a reduced $D_{\max}$ may reflect delayed arrival rather than complete interception. Hence, $D_{\max}$ evaluated over a limited time horizon should be interpreted as a planning-stage proxy, and it does not by itself allow a definitive separation between delayed propagation and full blockage. Extending the simulation time and/or incorporating auxiliary descriptors (e.g., arrival time) is left for future work.

In addition, $D_{\max}$ is an easily interpretable representative indicator, but it may not adequately summarize spatial features such as localized concentration of high-depth zones or deposition patterns. Because the proposed workflow is agnostic to the choice of proxy metrics, it can be directly extended to alternative simulation-derived metrics (e.g., deposited thickness/volume, impact-related indicators, and/or arrival time) by replacing $D_{\max}$ without changing the overall surrogate-assisted screening workflow.

The post-screening full-simulation check for the three representative scenarios (P1–P3) showed that the surrogate systematically underestimated the simulated $D_{\max}$ values. This discrepancy suggests that local errors can remain non-negligible even when the surrogate is adequate for preliminary scenario comparison. In particular, because the surrogate-based search procedure clips negative surrogate predictions to zero, the low-$D_{\max}$-proxy side of the screening-derived trade-off scenario set may be affected more strongly by such approximation errors. Accordingly, the present surrogate should be interpreted as a planning-stage screening tool rather than as a high-precision predictor for final quantitative design evaluation.

4.2 Interpretation of the screening-derived trade-off structure

Surrogate-based search was used as an exploratory tool to generate representative height scenarios that characterize trade-offs in the surrogate-proxy space, rather than to determine a single final scenario.

In this study, surrogate-based search was conducted using the scalar screening score defined in Section 2.6, with the weighting parameter α varied continuously. Although a single scalar screening score was employed, the resulting surrogate-screened candidate scenarios did not form a continuous distribution but instead converged to a limited number of non-dominated scenarios in the surrogate-proxy space.

Such behavior has been reported in surrogate-based search or optimization problems, particularly when the response surface is highly nonlinear or non-convex, and when abrupt changes in the response occur over certain regions of the scenario space

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

(Yang et al., 2025). Under these conditions, some feasible scenarios may be difficult to identify, and the obtained screened scenario set may reflect both the structure of the underlying problem and the characteristics of the surrogate approximation.

Therefore, the screening-derived trade-off structure observed in this study should not be interpreted as a complete representation of all possible surrogate-screened candidate scenarios. Rather, it provides a practical approximation of dominant trade-off tendencies between the downstream-depth proxy and the height-sum proxy under the present simplified definitions.

4.3 Variability around screening-derived trade-off scenarios and implications for planning-stage scenario comparison

Analysis of screened height combinations around representative non-dominated scenarios revealed differences in the variability of candidate dam heights depending on their locations. Upstream height variables exhibited relatively large variability among scenarios with similar proxy values, whereas one midstream height variable repeatedly converged to a narrower range.

Within the present demonstration setting and proxy definitions, the variability of screened heights differs among candidate locations. This should not be interpreted as direct evidence of structural priority or required reinforcement, but as a hypothesis-generating indicator for subsequent sensitivity analysis. In particular, variables with consistently narrow screened ranges may warrant closer examination using additional physics-based simulations, alternative performance metrics, or site-specific engineering constraints.

Such information cannot be obtained from a single surrogate-screened candidate scenario alone. Instead, it may help organize subsequent analyses by indicating which height variables and candidate scenarios merit closer examination in planning-stage discussion.

4.4 Comparison with existing studies and contribution of this work

Existing studies that integrate numerical simulations with machine learning or surrogate models have largely focused on improving the accuracy and efficiency of hazard prediction (e.g., (Qiao and Myers, 2022; Lei et al., 2025)). While these approaches are valuable for risk assessment, they often provide limited information regarding the structure of the scenario space or the variability of alternative countermeasure scenarios.

In contrast, the present study emphasizes the exploratory use of surrogate models to investigate trade-off structures and variability among the surrogate-screened candidate scenarios. The surrogate model is treated not as a replacement for physics-based simulation, but as a tool to enable systematic screening of candidate scenarios at the pre-design or planning stage.

This workflow may support preliminary comparison of hypothetical countermeasure scenarios and may help identify variables that require closer examination in subsequent physics-based or site-specific engineering studies. As such, the proposed approach complements existing high-accuracy prediction studies and offers a different perspective on the use of surrogate modeling in debris-flow mitigation planning.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

5 Conclusions

This study presented a surrogate-assisted preliminary screening workflow for comparing hypothetical fixed-location check-dam height scenarios using debris-flow simulation results and simplified proxy metrics. A physics-based debris-flow simulation was combined with data-driven surrogate modeling to investigate the relationship between dam-height scenarios, a downstream-depth proxy, and a simplified height-sum proxy for construction-related effort. Because the Atami terrain was used only as a real-terrain demonstration setting, the results should not be interpreted as a reproduction of the 2021 event, an assessment of actual check-dam facilities, or a basis for site-specific safety judgement.

Among the tested surrogate models, a multilayer perceptron provided the best coarse-grained approximation of the simulation-derived response. The surrogate enabled repeated candidate searches under varying trade-off weights, and the screened scenarios formed a limited number of representative trade-off regimes in the surrogate-proxy space.

However, comparison with full simulations for representative surrogate-screened scenarios indicated that local quantitative errors remained non-negligible. In particular, the low-$D_{\max}$-proxy side was sensitive to negative surrogate predictions that had been clipped to zero. This result supports the interpretation of the workflow as a planning-stage screening tool that must be complemented by physics-based re-evaluation before quantitative interpretation.

Analysis of screened height combinations showed that the variability of candidate heights differed among locations under the present proxy definitions. This result should be interpreted as a hypothesis-generating indicator for subsequent sensitivity analysis, not as evidence of structural priority, required reinforcement, or site-specific engineering importance.

The objective function adopted in this study represents a simplified proxy formulation. The maximum flow depth was used as an easily interpretable hazard-intensity proxy, while the total dam height served as a minimal height-sum proxy for construction-related effort. The workflow can be extended to alternative simulation-derived metrics, such as arrival time, deposited thickness or volume, or impact-related indicators, when such outputs and evaluation criteria are available.

The number of simulation cases used in this study does not guarantee complete coverage of the scenario space. The objective of the surrogate model is exploratory rather than exhaustive, aiming to capture dominant trends and organize candidate scenarios relevant to planning-stage discussion. Therefore, the proposed workflow may be useful for preliminary scenario comparison and pre-screening, where multiple alternatives must be compared under uncertainty and practical constraints.

Future work should address the influence of sampling strategies on surrogate accuracy and explore the use of more realistic performance and construction-effort indicators. The workflow should therefore be regarded as a transparent preliminary screening procedure for organizing hypothetical mitigation scenarios, not as a substitute for event reproduction, site-specific hazard assessment, engineering design, or legal and technical decision-making.

Code and data availability. The topographic dataset used in this study is available from the Center for Spatial Information Science, The University of Tokyo, as cited in the References. The iRIC software and Morpho2DH solver are available from the iRIC project website. The simulation-derived dataset and scripts used for Latin hypercube sampling, surrogate-model training, candidate search, and figure generation are available from the corresponding author upon reasonable request, subject to removal of environment-specific file paths and dependencies.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

Author contributions. JK designed the study, performed the simulations and surrogate analyses, prepared the figures, and wrote the initial manuscript draft. HS and TS contributed to the interpretation of the results and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

References

Center for Spatial Information Science, The University of Tokyo: Drone-based LiDAR survey data of the Atami debris-flow disaster on July 3, 2021, https://www.geospatial.jp/ckan/dataset/20210703-atami-dronelazer, accessed August 5, 2025, 2025.

Crozier, M. J.: Landslides and climate change, in: Landslides and Climate Change, edited by McInnes, R., Jakeways, J., Fairbank, H., and Mathie, E., pp. 1–22, CRC Press, Boca Raton, 2010.

Forrester, A. I. J., Sobester, A., and Keane, A. J.: Engineering Design via Surrogate Modelling, Wiley, Chichester, 2008.

Gariano, S. L. and Guzzetti, F.: Landslides in a changing climate, Earth-Science Reviews, 162, 227–252, 2016.

Giustolisi, O., Laucelli, D., and Berardi, L.: Operational and tactical management of water and energy resources by means of surrogate models, Environmental Modelling & Software, 67, 1–14, 2015.

Hu, J., Pang, A., and Deng, C.: Urban flood risk analysis using presence-only machine learning approach: an integrated MaxEnt-cloud model framework in Harbin, China, Natural Hazards, 121, 16 827–16 856, https://doi.org/10.1007/s11069-025-07452-4, 2025.

Imaizumi, F., Osanai, N., Kato, M., Koike, Y., Kosugi, K., Sakai, Y., Sakaguchi, H., Satofuka, Y., Takayama, S., Tanaka, T., and Nishi, Y.: Debris-flow disaster in Atami City, Shizuoka Prefecture, Japan, July 2021, Journal of the Japan Society of Erosion Control Engineering, 74, 34–42, in Japanese, 2022.

IPCC: Climate Change 2021: The Physical Science Basis, 2021.

iRIC Software: Morpho2DHver2 Solver Manual for Debris and Mud Flow Simulation (iRIC v4), https://i-ric.org/, in Japanese, 2025.

Iverson, R. M.: The physics of debris flows, Reviews of Geophysics, 35, 245–296, 1997.

Kennedy, J. and Eberhart, R.: Particle swarm optimization, Proceedings of the IEEE International Conference on Neural Networks, pp. 1942–1948, 1995.

Lei, D., Ma, J., Zhang, G., Wang, Y., Deng, X., and Liu, J.: Bayesian ensemble learning and Shapley additive explanations for fast estimation of slope stability with a physics-informed database, Natural Hazards, 121, 2941–2970, https://doi.org/10.1007/s11069-024-06917-2, 2025.

Ma, G., Zhi, M., Zang, X., Chen, S., Wu, D., and Huang, X.: Slope stability prediction based on AutoML for multiple failure mechanisms, Natural Hazards, 121, 15 297–15 330, https://doi.org/10.1007/s11069-025-07397-8, 2025.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979.

Pastor, M., Blanc, T., and Pastor, M. J.: A depth-integrated viscoplastic model for dilatant saturated cohesive-frictional fluidized mixtures: Application to fast catastrophic landslides, Journal of Non-Newtonian Fluid Mechanics, 158, 142–153, 2009.

Pudasaini, S. P.: A general two-phase debris flow model, Journal of Geophysical Research: Earth Surface, 117, F03 010, 2012.

Qiao, C. and Myers, A. T.: Surrogate modeling of time-dependent metocean conditions during hurricanes, Natural Hazards, 110, 1545–1563, https://doi.org/10.1007/s11069-021-05002-2, 2022.

Razavi, S., Tolson, B. A., and Burn, D. H.: Surrogate modeling for water resources management: A review, Water Resources Research, 48, W07 401, 2012.

Takahashi, T.: Debris Flow: Mechanics, Prediction and Countermeasures, CRC Press, Boca Raton, 2007.

Takebayashi, H.: Numerical simulation of debris and mud flows using iRIC-Morpho2DH, Sabo, 134, 1–10, in Japanese, 2023.

Takebayashi, H., Fujita, M., and Ohgushi, K.: Numerical modeling of debris flows using basic equations in a generalized curvilinear co-ordinate system and its application to debris flows in the Kinryu River Basin, Saga City, Japan, Journal of Hydrology, 605, 128 636, 2022.

https://doi.org/10.5194/egusphere-2026-2765
Preprint. Discussion started: 2 June 2026

Yang, Y., Ji, F., Gao, Y., and Liang, P.: Slope stability analysis using a surrogate model with varying sampling precision: a case study of open-pit mine dump slopes, Natural Hazards, 121, 10 963–10 988, https://doi.org/10.1007/s11069-025-07239-7, 2025.