water

MDPI

Article

Process-Chain Analysis of Viscoplastic Landslide-Generated Impulse Waves from Initial Slide Conditions Through Water-Entry Dynamics to Wave Responses

Zhenxia Yuan [1,2,*], Yadong Bian [2,3], Song Yin [1], Xian Liu [4], Xinchen Pan [1], Zhenzhu Meng [5] and Bohan Cheng [1]

[1] School of Intelligent Construction and Civil Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China; 155320338@zut.edu.cn (S.Y.); 16665306130@163.com (X.P.); 18304058312@163.com (B.C.)
[2] Henan Environmental Geotechnical Engineering and Underground Engineering Disaster Control Engineering Research Center, Zhengzhou 450007, China; byd@hpu.edu.cn
[3] School of Civil Engineering, Henan Polytechnic University, Jiaozuo 454000, China
[4] College of Civil Engineering, Sun Yat-Sen University, Zhuhai 519082, China; liux597@alumni.sysu.edu.cn
[5] School of Hydraulic Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China; mengzhzh@zjweu.edu.cn
* Correspondence: 6570@zut.edu.cn

Abstract

A key scientific challenge in landslide-generated impulse-wave research is to determine how initial slide conditions are translated into wave responses through the intermediate water-entry process. Existing data-driven studies generally focus on direct prediction or variable-importance ranking and therefore provide limited insight into the pathways linking slide release, water entry, and wave generation. To address this issue, a three-stage process-chain framework was developed using 290 laboratory test cases, integrating initial slide conditions, water-entry dynamics, and wave responses. Cross-stage correlation analysis, standardized path modeling, mediation-effect decomposition, cascade prediction, bootstrap resampling, and diagnostic analyses were jointly applied. The results show that relative effective entry mass M is strongly correlated with relative wave amplitude A and relative wave height H, with Spearman coefficients of 0.96 and 0.91, respectively, while the corresponding coefficients for the entry Froude number Fr are 0.71 and 0.80. Within the present dataset and fitted model, mediation analysis indicates that the pathways through M have the largest modeled indirect contributions: the indirect effects of initial slide mass through M are 0.652 for A and 0.600 for H, whereas the corresponding effects of the Bingham number are -0.669 and -0.615. The cascade framework further shows that the realized entry state can be systematically reconstructed from the initial conditions and used as an explicit intermediate layer for wave-response prediction. Prediction comparisons further show that the observed entry-state model is best interpreted as an explanatory or diagnostic benchmark because it uses measured entry-state variables, whereas the direct initial-condition and cascade models address prospective prediction from prescribed initial-condition information. The main advantage of the cascade framework therefore lies in its physical interpretability rather than a substantial gain in predictive accuracy. Within the present dataset and fitted model, the associations between the initial slide conditions and wave responses are expressed mainly through mass-related and inertia-related pathways involving the realized entry state.

Keywords: landslide-generated waves; process-chain framework; water-entry dynamics; wave responses; effective entry mass

Check for updates

Academic Editor: Mariano Buccino

Received: 3 July 2026
Revised: 5 August 2026
Accepted: 31 August 2026
Published: 3 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Landslide-generated impulse waves are hazardous secondary phenomena that occur when rapidly moving landslides enter water bodies and transfer mass and momentum to the surrounding water. The resulting waves may cause severe run-up, overtopping, and impact damage. Their generation and propagation are influenced by landslide volume, impact velocity, slide geometry, water depth, slope angle, and basin geometry [1–4]. Understanding how landslide motion is transformed into wave response is therefore essential for impulse-wave prediction and hazard assessment.

Extensive laboratory and numerical studies have clarified many of the physical processes governing landslide-generated waves. Early experiments established relationships between slide characteristics and near-field wave responses, particularly the effects of impact velocity, slide volume and thickness, water depth, and slope geometry [3,5]. Subsequent studies examined scale effects, developed dimensionless impulse parameters, classified wave-generation regimes, and extended the analysis to wave propagation and three-dimensional water-body geometries [6–13]. Numerical models, including depth-integrated, multilayer, and coupled solid–fluid approaches, have further enabled simulations of slide motion, wave generation, propagation, and run-up under complex bathymetric and boundary conditions [14–21]. However, much of the existing knowledge has been developed for rigid-block, granular, or otherwise idealized slide materials [12,21–24].

Deformable and rheologically complex landslide materials remain more difficult to characterize [25–27]. Unlike rigid-block slides, natural landslides may deform, spread, or fragment during downslope motion and water entry [28–30]. Consequently, the mass, geometry, and momentum of the material that actually enters the water may differ substantially from those represented by the initial slide conditions. Experiments using viscoplastic materials have shown that material rheology can significantly influence impulse-wave generation [31–33]. Darvenne et al. (2024) [34] further demonstrated that the evolving configuration of a deformable slide during water entry affects the resulting wave response. These findings motivate treating water entry as a distinct process stage between landslide motion and wave generation.

Such cross-stage relationships are difficult to characterize using conventional empirical formulas alone. Data-driven methods have therefore been increasingly applied to geohazard modeling because of their ability to capture complex nonlinear relationships [35–37]. For landslide-generated waves, Lyu et al. (2025) [38] used machine learning models and large-scale physical-model data to predict surge-wave characteristics, while Mao et al. (2025) [39] investigated their spatiotemporal evolution using experimental data. Interpretable machine learning methods have also shown potential for identifying influential variables and supporting process-oriented analysis in broader landslide and geotechnical applications [40–43].

Despite these advances, most data-driven studies of landslide-generated waves rely on direct input–output mappings and focus primarily on predictive performance or variable-importance ranking [44–47]. Such frameworks do not generally distinguish whether an association with wave response is linked directly to the initial slide condition or is expressed through the realized water-entry state. A process-oriented framework is therefore needed to separate direct and entry-mediated associations and to distinguish the relative roles of effective mass entry, entry inertia, and entry geometry.

To address this gap, this study develops a three-stage process-chain framework using 290 laboratory test cases. The contribution of this study does not lie in proposing a new path-analysis, mediation-analysis, or gradient-boosting algorithm; rather, it lies in the physically structured integration of these established methods around the sequence from

prescribed initial slide conditions, through the realized water-entry state, to the resulting wave responses.

First, cross-stage correlation analysis is used to compare the associations among the three stages. Second, standardized path modeling and mediation-effect decomposition are applied to quantify direct and entry-mediated conditional associations and to compare the relative strengths of the modeled mass-related, inertia-related, and geometry-related pathways within the present dataset and fitted model. Finally, the direct initial-condition and cascade models are evaluated for prospective prediction from prescribed initial-condition information, whereas the observed entry-state model is used as an explanatory or diagnostic benchmark based on measured entry-state information.

2. Experimental Dataset and Variable System

2.1. Physical Model and Experimental Dataset

Landslide-generated impulse waves involve a continuous sequence of slide motion, water entry, momentum transfer, and wave propagation. After release, the landslide accelerates and deforms along the inclined plane under the combined effects of gravity and rheological resistance. Upon reaching the water surface, the entering material displaces the surrounding water and transfers mass and momentum to the water body, producing an initial surface disturbance that develops into an impulse wave. The resulting wave response therefore depends not only on the initial slide conditions, but also on the realized kinematic and geometric state of the slide at water entry.

Figure 1 presents the simplified physical model of landslide-generated impulse waves. Before release, the landslide material is retained in a container mounted on an inclined plane with a slope angle of θ. The downstream end of the container is equipped with a pneumatic gate. The initial landslide mass is denoted by m_i, while s_i denotes the initial thickness of the landslide material measured at the gate immediately before release. The sliding distance from the release position to the still-water surface is denoted by l_s. Once the pneumatic gate is opened, the landslide material moves downslope and subsequently enters the water body. At the instant when the front of the landslide reaches the still-water surface, the slide is characterized by the entry velocity u_0 and entry thickness s_0. The still-water depth is denoted by h_0 and was maintained at 0.20 m throughout the experiments. The generated impulse wave is characterized by the maximum wave amplitude a_m and maximum wave height h_m. The effective entry mass m_e is operationally defined as the mass of landslide material that has entered the water body at the instant when the maximum wave height h_m is attained. This event-based definition follows previous experimental studies of impulse waves generated by deformable and viscoplastic slides [31–33] and was retained to preserve consistency and comparability with the established experimental variable system.

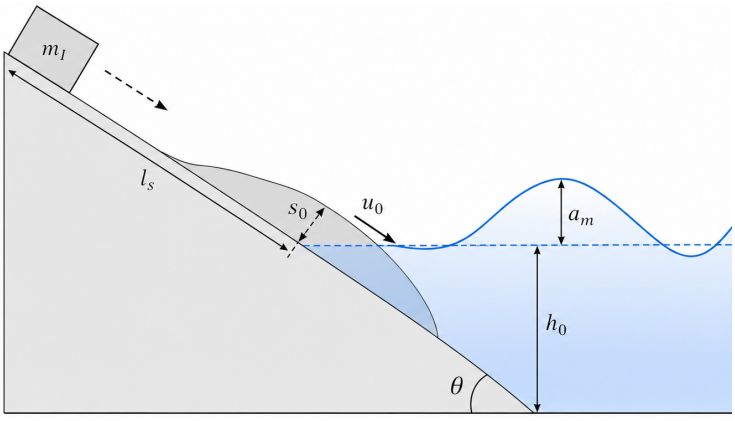


Figure 1. Conceptual physical model of landslide-generated impulse waves.

Laboratory experiments were conducted using a two-dimensional landslide-wave flume system. The experimental system consisted primarily of a landslide container, a pneumatic release gate, an adjustable inclined chute, a narrow transparent wave flume, and an image-acquisition system. The transparent flume was 2.5 m long and had a width of $B = 0.12$ m. The inclined chute was constructed from PVC, had the same width as the wave flume, and allowed the inclination angle to be adjusted between 30° and 50°. During each experiment, the Carbopol material was initially placed in the container at a prescribed position on the chute. Rapid opening of the pneumatic gate initiated the slide motion, after which the material moved downslope, entered the water, and generated impulse waves propagating along the flume. Figure 2a,b presents a schematic diagram and a photograph of the experimental system, respectively.

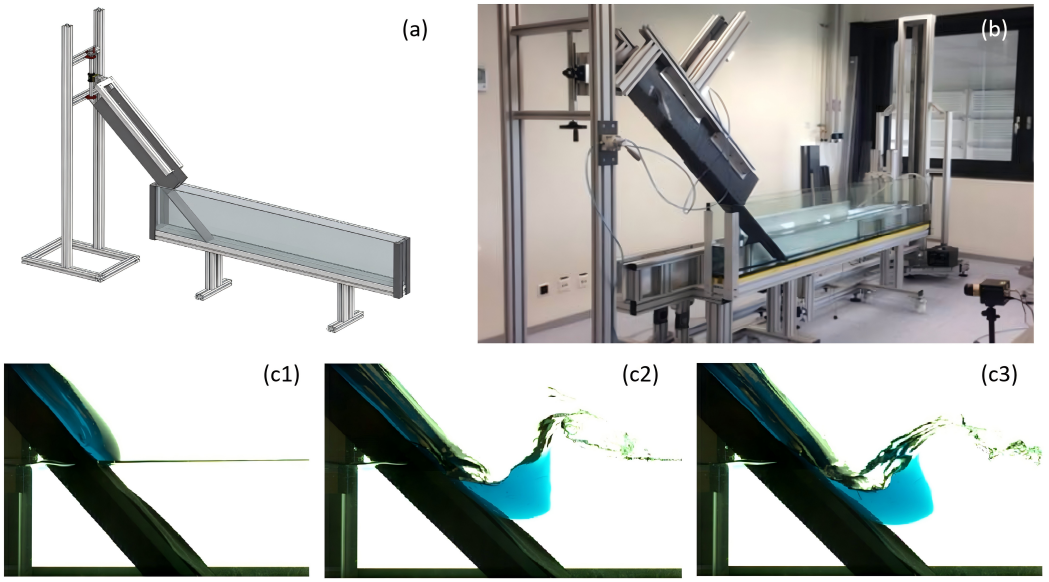


Figure 2. Experimental set-up and representative raw images: (**a**) schematic of the experimental facility; (**b**) photograph of the experimental system; and (**c1**–**c3**) representative images of the water-entry and subsequent wave-generation processes.

The slide motion, water-entry process, and free-surface evolution were recorded using a fixed high-speed camera at a frame rate of 400 fps and an image resolution of 600 × 800 pixels. Wave amplitude, wave height, entry thickness, entry velocity, and effective entry mass were subsequently extracted from the recorded images using image-processing procedures. Before testing, a calibration grid with known dimensions was placed in the measurement plane to establish the conversion between pixel coordinates and physical distances. The camera position and field of view were kept unchanged throughout the experiments. The field of view covered the slide-entry zone and the near-field wave-generation region, with the measurement locations shown in Figure 2(c1–c3). Wave amplitude was defined as the maximum upward displacement of the free surface relative to the still-water level, whereas wave height was measured as the vertical distance between the corresponding wave crest and trough. Entry thickness was extracted from the slide profile at the still-water shoreline, and entry velocity was calculated from the displacement of the slide front between successive frames. Effective entry mass was determined from the portion of the slide that had entered the water by the instant at which the maximum wave height occurred, based on the extracted slide contour and the known material density. Measurement uncertainty was evaluated according to the spatial resolution obtained from grid calibration and the temporal resolution of the high-speed camera. To assess test-to-test repeatability, a representative test condition was independently repeated three times under identical

experimental settings, and the variation among the repeated measurements was used to estimate repeatability uncertainty.

Carbopol was used as the landslide material in the experiments. Carbopol is an idealized viscoplastic material characterized by a well-defined yield stress and shear-thinning behavior. Its post-yield rheological behavior can be described by the Herschel–Bulkley model:

$$\tau = \tau_c + K\dot{\gamma}^n, \qquad \tau > \tau_c, \tag{1}$$

where τ is the shear stress, τ_c is the yield stress, K is the consistency coefficient, $\dot{\gamma}$ is the shear rate, and n is the flow-behavior index. When the applied shear stress does not exceed τ_c, the material remains in an unyielded state without continuous flow. The yield stress τ_c therefore represents the threshold resistance that must be overcome before sustained deformation and flow occur and was selected as the representative rheological parameter in the subsequent analysis. The preparation procedure, rheological characterization, and suitability of Carbopol as a laboratory-scale analogue for viscoplastic landslide materials have been described in detail in a previous study [31,33]. Figure 2(c1–c3) shows representative images of the water-entry and subsequent wave-generation processes.

The principal controlled parameters were the material yield stress τ_c, initial landslide mass m_i, sliding distance l_s, and slope angle θ. Their investigated ranges were 60–90 Pa, 2.0–5.6 kg, 0.55–0.75 m, and 40–45°, respectively. The entry thickness s_0, measured when the landslide reached the still-water surface, ranged from 0.026 to 0.050 m, while the corresponding entry velocity u_0 ranged from 1.12 to 2.36 m/s. The still-water depth was maintained at $h_0 = 0.20$ m for all tests. For each test, a_m and h_m were extracted separately from the same free-surface elevation extracted from the same calibrated camera field of view. The maximum wave amplitude a_m was defined as the maximum upward displacement of the water surface relative to the still-water level, whereas the maximum wave height h_m was defined as the maximum vertical distance between a wave crest and its associated trough. A total of 290 laboratory test cases were included in the present analysis. The 290 cases used for statistical modeling were generated by systematically varying the prescribed initial-condition variables at regularly spaced levels within their respective experimental ranges. Each test contained variables characterizing the initial slide conditions, the water-entry state, and the resulting wave response.

2.2. Dimensionless Variables

To reduce dimensional dependence and improve comparability among the experimental conditions, the dimensional variables were converted into dimensionless form. The initial slide conditions were characterized by the relative initial mass M_i, relative sliding distance L, and Bingham number Bi. The realized water-entry state was characterized by the entry Froude number Fr, relative entry thickness S, and relative effective entry mass M. The wave response was quantified by the relative maximum wave amplitude A and relative maximum wave height H.

The dimensionless variables describing the initial slide conditions are defined as

$$M_i = \frac{m_i}{\rho_w B h_0^2}, \qquad L = \frac{l_s}{h_0}, \qquad Bi = \frac{\tau_c}{\rho_s g s_i \sin\theta}, \tag{2}$$

where s_i is the initial landslide thickness measured at the pneumatic gate immediately before release. The Bingham number represents the ratio of the material yield resistance to the downslope gravitational stress associated with the initial slide configuration.

The dimensionless variables describing the realized water-entry and mass-transfer state are defined as

$$Fr = \frac{u_0}{\sqrt{gh_0}}, \qquad S = \frac{s_0}{h_0}, \qquad M = \frac{m_e}{\rho_w B h_0^2}, \tag{3}$$

where s_0 is the landslide thickness measured at the still-water surface during water entry.

The dimensionless wave-response variables are defined as

$$A = \frac{a_m}{h_0}, \qquad H = \frac{h_m}{h_0}, \tag{4}$$

where ρ_w is the water density, ρ_s is the bulk density of the landslide material, B is the flume width, and g is gravitational acceleration.

3. Methods

3.1. Process-Chain Framework

The dimensionless variables defined in Section 2.2 were organized into three process stages:

$$\mathbf{X}_0 = (M_i, L, Bi), \tag{5}$$

$$\mathbf{X}_e = (Fr, S, M), \tag{6}$$

and

$$\mathbf{Y} = (A, H), \tag{7}$$

where $\mathbf{X}_0$, $\mathbf{X}_e$, and $\mathbf{Y}$ denote the prescribed initial-condition variables, the realized entry-state variables, and the wave-response variables, respectively. These terms and notations are used consistently throughout the manuscript.

As illustrated in Figure 3, the proposed three-stage process chain is expressed as

$$\mathbf{X}_0 \rightarrow \mathbf{X}_e \rightarrow \mathbf{Y}. \tag{8}$$

This organization distinguishes the prescribed initial conditions from the entry state realized after downslope motion and water entry. It provides the basis for examining both the direct associations between $\mathbf{X}_0$ and $\mathbf{Y}$ and the entry-mediated associations transmitted through $\mathbf{X}_e$.

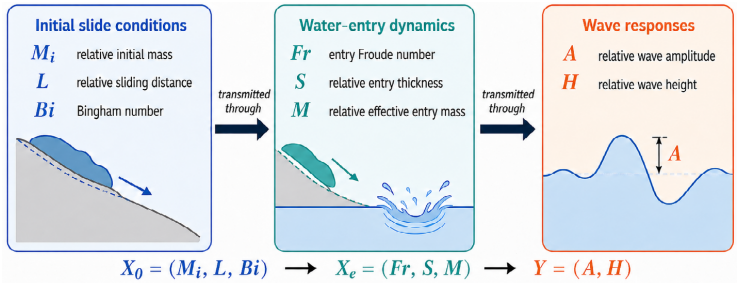


Figure 3. Three-stage process-chain framework linking the prescribed initial-condition variables, realized entry-state variables, and wave-response variables.

The overall analytical workflow is summarized in Figure 4. It comprises cross-stage correlation analysis, standardized path and mediation analysis, reconstruction of water-entry variables, comparison of wave-response prediction frameworks, mechanism mapping, and robustness diagnostics.

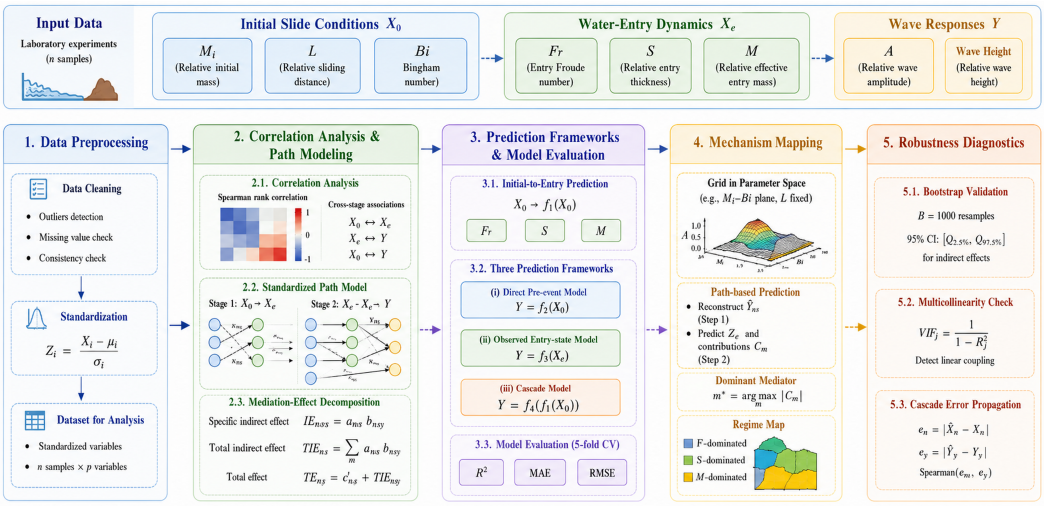


Figure 4. Methodological workflow for cross-stage correlation analysis, standardized path and mediation analysis, cascade prediction, mechanism mapping, and robustness diagnostics.

3.2. Cross-Stage Correlation Analysis

Descriptive statistics, including the mean, standard deviation, minimum, and maximum, were first calculated to characterize the ranges and distributions of the dimensionless variables.

Spearman rank correlation analysis was then used to quantify pairwise monotonic associations across the three process stages. Correlations were evaluated separately between initial slide conditions and water-entry dynamics, between water-entry dynamics and wave responses, and between initial slide conditions and wave responses. Statistical significance was denoted by $p < 0.05$, $p < 0.01$, and $p < 0.001$.

This analysis provided an initial assessment of whether the realized water-entry variables were more closely associated with the wave responses than the initial slide conditions alone.

3.3. Standardized Path and Mediation Analysis

Before path analysis, all variables were standardized as

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}, \tag{9}$$

where X_i is the original value of the ith variable, μ_i and σ_i are its mean and standard deviation, respectively, and Z_i is the corresponding standardized value.

A two-stage path model was established to quantify the associations across the process chain. The first-stage model relates the initial-condition variables to each water-entry variable:

$$Z_m = \alpha_{0m} + \sum_{k=1}^{3} a_{km} Z_k + \varepsilon_m, \tag{10}$$

where $Z_k \in \{Z_{M_i}, Z_L, Z_{Bi}\}$ denotes an initial-condition variable, $Z_m \in \{Z_{Fr}, Z_S, Z_M\}$ denotes a water-entry variable, a_{km} is the corresponding path coefficient, and ε_m is the residual term.

The second-stage model relates each wave response to both the initial-condition and water-entry variables:

$$Z_y = \beta_{0y} + \sum_{k=1}^{3} c'_{ky} Z_k + \sum_{m=1}^{3} b_{my} Z_m + \varepsilon_y, \tag{11}$$

where $Z_y \in \{Z_A, Z_H\}$ denotes a wave-response variable, c'_{ky} is the direct conditional association between initial variable k and response y after accounting for the entry-state variables, b_{my} is the path coefficient from mediator m to response y, and ε_y is the residual term.

The pathway-specific indirect effect through mediator m was calculated as

$$IE_{kmy} = a_{km} b_{my}. \tag{12}$$

The total indirect effect of initial variable k on response y was calculated as

$$TIE_{ky} = \sum_{m=1}^{3} IE_{kmy} = \sum_{m=1}^{3} a_{km} b_{my}, \tag{13}$$

and the total effect was obtained as

$$TE_{ky} = c'_{ky} + TIE_{ky}. \tag{14}$$

The association of each initial-condition variable with a wave response was decomposed into one direct component and three entry-mediated components through Fr, S, and M. Following conventional mediation terminology, these components are referred to as direct, indirect, and total effects; however, they represent conditional associations within the specified linear additive model rather than strictly causal effects. Although the temporal ordering of the experimental process provides a physical basis for the pathway structure, the estimated coefficients depend on the variables included in the model and may be sensitive to multicollinearity among the coupled initial-condition and water-entry variables. Interpretation therefore emphasizes coefficient direction, the overall pathway structure, and physical consistency rather than the precise magnitude of an individual coefficient.

3.4. Prediction Frameworks and Model Evaluation

3.4.1. Reconstruction of Water-Entry Variables

To evaluate whether the realized water-entry state could be inferred from the initial slide conditions, an initial-to-entry prediction model was established as

$$\hat{\mathbf{X}}_e = f_1(\mathbf{X}_0), \tag{15}$$

namely,

$$(\widehat{Fr}, \widehat{S}, \widehat{M}) = f_1(M_i, L, Bi). \tag{16}$$

Gradient Boosting Regression was adopted for the nonlinear prediction tasks because of its ability to capture complex relationships in relatively small experimental datasets [48,49]. The method constructs an additive ensemble by sequentially fitting regression trees to the residual errors of the preceding model, with each update controlled by a learning rate. Its general procedure is illustrated in Figure 5.

3.4.2. Comparison of Wave-Response Modeling Frameworks

Three modeling frameworks were compared. The direct initial-condition model predicts the wave responses directly from the initial slide conditions:

$$\hat{\mathbf{Y}} = f_d(\mathbf{X}_0). \tag{17}$$

The observed entry-state model estimates the responses from the measured entry-state variables:

$$\hat{\mathbf{Y}} = f_e(\mathbf{X}_e). \tag{18}$$

The cascade model first reconstructs the intermediate entry state and then predicts the wave responses:

$$\hat{\mathbf{Y}} = f_2[f_1(\mathbf{X}_0)]. \tag{19}$$

The direct initial-condition model and the cascade model use only prescribed initial-condition information at the prediction stage and are therefore evaluated in terms of prospective prediction. In contrast, the observed entry-state model uses measured Fr, S, and M and is treated as an explanatory or diagnostic benchmark rather than a strictly prospective model. All models were evaluated using five-fold cross-validation. Within each fold, model training was performed using only the training partition, and predictions were generated for the corresponding held-out partition. For the cascade model, both entry-state reconstruction and wave-response prediction were conducted within the same fold to avoid information leakage.

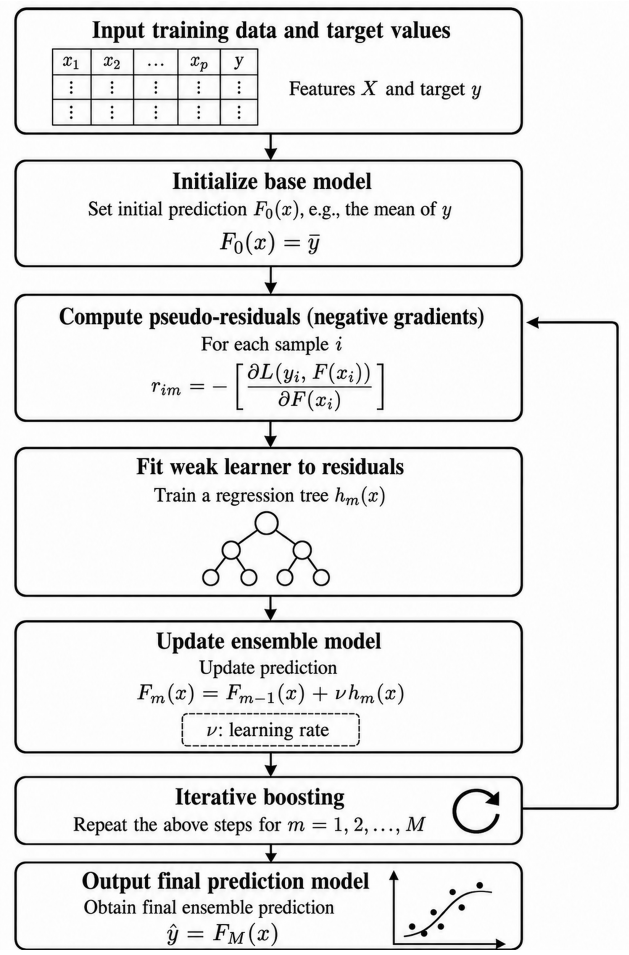


Figure 5. Flowchart of Gradient Boosting Regression used for the nonlinear prediction tasks in this study.

Predictive performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). Higher R^2 and lower MAE and RMSE indicate better predictive performance. For each modeling framework, R^2, MAE, and RMSE were calculated separately on each held-out fold and subsequently summarized as the mean and sample standard deviation across the five folds. This fold-wise evaluation was used to quantify the variability of model performance across different data partitions. No data-driven hyperparameter search was conducted as part of the final cross-validation evaluation. A single prespecified Gradient Boosting configuration was used, with `n_estimators = 250`, `learning_rate = 0.04`, `max_depth = 2`,

`min_samples_leaf = 5`, and `subsample = 0.85`. These settings were fixed before the final evaluation and were applied unchanged to all three modeling frameworks, response variables, and cross-validation folds. Consequently, the held-out folds were used exclusively for performance evaluation and played no role in model or hyperparameter selection. For the cascade model, the inner five-fold procedure was used only to generate out-of-fold entry-state reconstructions within each outer training partition and was not used for hyperparameter optimization.

3.5. Mechanism Mapping and Robustness Diagnostics

3.5.1. Path-Based Mechanism Mapping

To visualize how the dominant mediated association varied across the initial-condition space, a path-based mechanism map was constructed for H. A regular grid was generated in the M_i–Bi plane, while L was fixed at its median value.

For each grid point, the standardized entry-state variables were reconstructed using the first-stage path model:

$$\hat{Z}_m = \sum_{k=1}^{3} a_{km} Z_k. \tag{20}$$

The corresponding standardized wave response was calculated as

$$\hat{Z}_H = \sum_{k=1}^{3} c'_{kH} Z_k + \sum_{m=1}^{3} b_{mH} \hat{Z}_m. \tag{21}$$

The local contribution associated with mediator m was defined as

$$C_m = b_{mH} \hat{Z}_m. \tag{22}$$

The dominant mediator at each grid point was identified as

$$m^* = \arg \max_{m \in \{Fr, S, M\}} |C_m|. \tag{23}$$

The parameter space was consequently divided into inertia-mediated, thickness-mediated, and mass-mediated regions corresponding to Fr, S, and M, respectively. These regions represent model-derived mediation patterns rather than deterministic physical-regime boundaries.

3.5.2. Bootstrap and Multicollinearity Diagnostics

Bootstrap resampling was used to assess the stability of the pathway-specific indirect effects. In each iteration, n observations were drawn with replacement from the original dataset, and the path coefficients and indirect effects were re-estimated. A total of $B = 1000$ repetitions were performed.

The 95% bootstrap confidence interval was defined as

$$CI_{95\%} = [Q_{2.5\%}, Q_{97.5\%}], \tag{24}$$

where $Q_{2.5\%}$ and $Q_{97.5\%}$ are the corresponding percentiles of the bootstrap distribution. A pathway was considered statistically stable when its confidence interval did not include zero.

Multicollinearity was evaluated using the variance inflation factor:

$$VIF_j = \frac{1}{1 - R_j^2}, \tag{25}$$

where R_j^2 is obtained by regressing the jth predictor on the remaining predictors in the same model. Larger VIF values indicate stronger linear dependence among predictors. The VIF results were used to support cautious interpretation of the path coefficients as conditional associations within the specified model. Additional robustness analyses are provided in Appendix A.

3.5.3. Cascade Error-Propagation Analysis

To examine how reconstruction uncertainty propagated through the cascade framework, the absolute reconstruction error of entry-state variable m and the absolute prediction error of wave response y were defined as

$$e_m = |\hat{X}_m - X_m|, \tag{26}$$

and

$$e_y = |\hat{Y}_y - Y_y|. \tag{27}$$

Spearman correlation analysis was used to quantify the association between e_m and e_y. A stronger positive correlation indicates that reconstruction uncertainty in a particular entry-state variable is more closely associated with the final wave-response prediction error.

3.6. Computational Implementation and Reproducibility

All statistical analyses and machine learning models were implemented in Python 3.14.6 using NumPy 2.5.1, pandas 3.0.5, SciPy 1.18.0, scikit-learn 1.9.0, statsmodels 0.14.6, and Matplotlib 3.11.1. The path models were fitted using ordinary least squares after z-score standardization of all variables. The mediation analysis was conducted using 1000 bootstrap resamples with a random seed of 42, and the 2.5th and 97.5th percentiles of the bootstrap distributions were used to construct the 95% confidence intervals.

For the nonlinear modeling analysis, the `GradientBoostingRegressor` was combined with `MultiOutputRegressor`. The model settings were `n_estimators = 250`, `learning_rate = 0.04`, `max_depth = 2`, `min_samples_leaf = 5`, and `subsample = 0.85`. Input variables were standardized using `StandardScaler`, with the scaler fitted only to the training data within each cross-validation fold.

The direct initial-condition and observed entry-state models were evaluated using shuffled five-fold cross-validation with a random seed of 42. The cascade model was evaluated using an outer shuffled five-fold cross-validation with a random seed of 42. Within each outer training fold, an inner five-fold cross-validation with a random seed of 123 was used to generate out-of-fold entry-state reconstructions for training the second-stage wave-response model, thereby avoiding information leakage. Fold-specific random seeds of 43–47 and 143–147 were used for the first- and second-stage estimators, respectively. These settings were kept unchanged for all model comparisons.

4. Results

4.1. Variable Distributions and Cross-Stage Correlations

Figure 6 shows the grouping and distributions of the dimensionless variables, while Table 1 summarizes their basic statistics. The initial-condition variables cover relatively broad ranges, with M_i, L, and Bi varying from 0.357 to 1.120, 2.325 to 4.760, and 2.475 to 5.536, respectively. The water-entry variables have mean values of 1.136 for Fr, 0.221 for S, and 0.173 for M, while the mean values of the wave-response variables are 0.355 for A and 0.498 for H. The maximum values of A and H reach 0.585 and 0.833, respectively. The distributions of A and H are generally unimodal, whereas L, Bi, and several entry-state

variables exhibit more discrete patterns, reflecting the controlled parameter combinations adopted in the physical-model experiments.

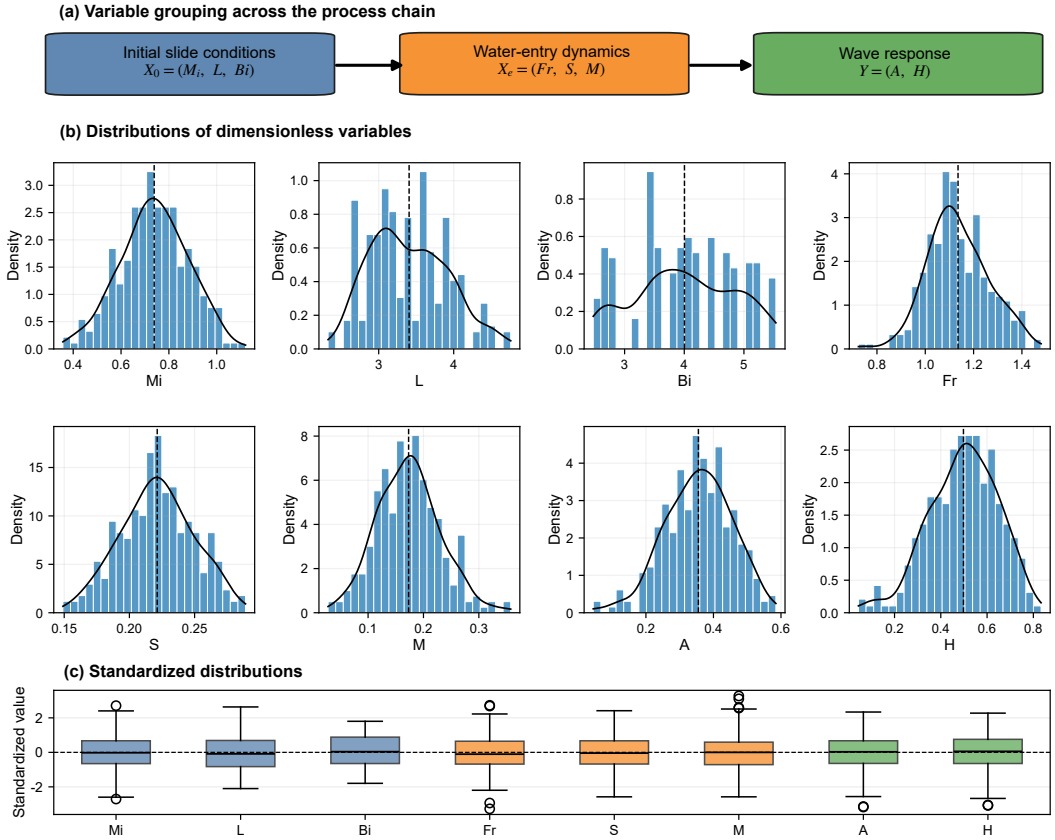


Figure 6. Variable grouping and basic distributions of dimensionless variables across the initial-condition, water-entry, and wave-response stages.

Table 1. Summary statistics of the dimensionless variables used in this study.

Variable	Mean	Standard Deviation	Minimum	Maximum
M_i	0.739	0.141	0.357	1.120
L	3.405	0.515	2.325	4.760
Bi	4.003	0.853	2.475	5.536
Fr	1.136	0.126	0.723	1.479
S	0.221	0.028	0.149	0.289
M	0.173	0.056	0.028	0.358
A	0.355	0.098	0.044	0.585
H	0.498	0.148	0.044	0.833

Figure 7 provides an initial overview of the statistical associations across the three process stages. Between the initial-condition and water-entry variables, M_i is positively correlated with S and M, with Spearman correlation coefficients of 0.61 and 0.36, respectively. The Bingham number Bi is strongly negatively correlated with Fr ($\rho_s = -0.79$) and weakly negatively correlated with M ($\rho_s = -0.29$). The sliding distance L is positively correlated with Fr ($\rho_s = 0.33$) but negatively correlated with S and M, with coefficients of -0.60 and -0.26, respectively.

Stronger associations are generally observed between the water-entry variables and the wave responses. The relative effective entry mass M is highly correlated with both A and H, with coefficients of 0.96 and 0.91, respectively. The entry Froude number Fr also shows strong positive correlations with A and H, particularly with H ($\rho_s = 0.80$).

By comparison, *S* exhibits moderate positive correlations with the two wave responses. The direct correlations between the initial-condition variables and wave responses are generally weaker, with *Bi* showing the strongest initial-to-response correlations ($\rho_s = -0.44$ for *A* and $\rho_s = -0.56$ for *H*). These cross-stage differences provide a preliminary statistical basis for the subsequent path analysis.

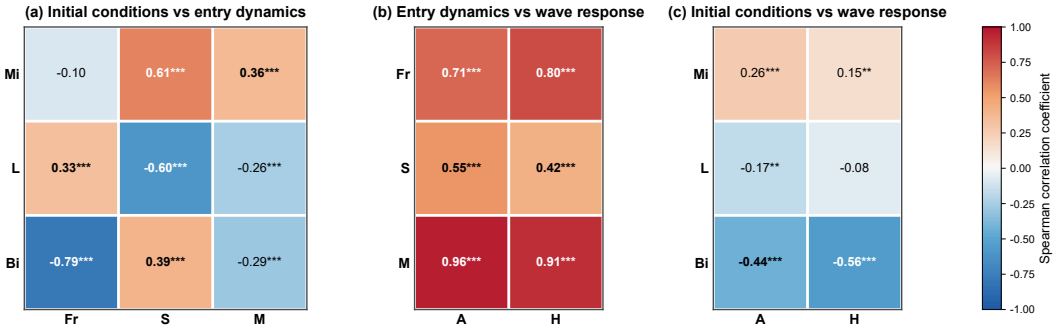


Figure 7. Cross-stage correlation heatmap among initial conditions, water-entry dynamics, and wave responses. Cell values are correlation coefficients; ** and *** indicate $p < 0.01$ and $p < 0.001$, respectively.

4.2. Standardized Path Coefficients Across the Process Chain

Figure 8 presents the standardized path coefficients linking the initial-condition, water-entry, and wave-response stages. Solid arrows denote paths involving the intermediate water-entry variables, whereas dashed arrows denote direct paths from the initial-condition variables to the wave responses. Statistical significance is indicated by $p < 0.05$, $p < 0.01$, and $p < 0.001$.

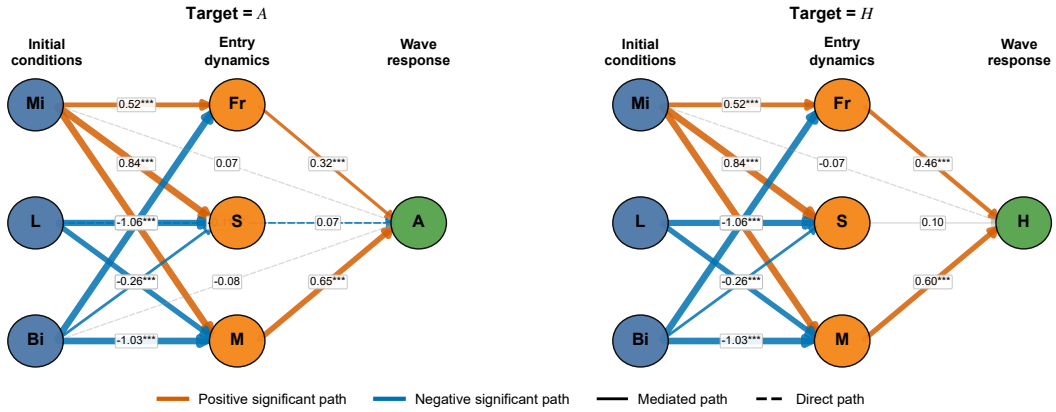


Figure 8. Standardized path coefficients linking initial conditions, water-entry dynamics, and wave responses. *** indicate $p < 0.001$.

In the first-stage models, M_i shows significant positive coefficients for all three water-entry variables, with values of 0.521 for *Fr*, 0.839 for *S*, and 1.006 for *M*. In contrast, *Bi* shows significant negative coefficients of −1.061, −0.261, and −1.031 for *Fr*, *S*, and *M*, respectively. The coefficient of *L* for *Fr* is small and statistically nonsignificant, whereas its coefficients for *S* and *M* are significantly negative, with values of −0.741 and −0.622, respectively. These results indicate that the relationships of the three initial-condition variables with the realized water-entry state differ substantially among *Fr*, *S*, and *M*.

In the second-stage models, *M* and *Fr* show the strongest positive coefficients for the wave responses. For *A*, the coefficients associated with *M* and *Fr* are 0.649 and 0.316, respectively. For *H*, the corresponding coefficients are 0.596 and 0.458. The coefficients

of S are relatively small and statistically nonsignificant for both responses, with values of 0.072 for A and 0.100 for H. After accounting for the water-entry variables, most direct paths from the initial-condition variables to the wave responses are also small; among these paths, only the coefficient from L to A is statistically significant (-0.114).

4.3. Direct, Indirect, and Total Effects

Figure 9 presents the decomposition of the standardized associations between the initial-condition variables and the wave responses into direct and entry-mediated components. For M_i, the total effects on A and H are positive, with values of 0.945 and 0.855, respectively. The largest components are the pathways mediated through M, which reach 0.652 for A and 0.600 for H. The corresponding pathways through Fr are positive but smaller, with values of 0.165 and 0.239.

The total effects of Bi on A and H are strongly negative, with values of -1.098 and -1.117, respectively. The largest negative components are associated with the pathways through M, with indirect effects of -0.669 for A and -0.615 for H. The pathways through Fr also make substantial negative contributions, particularly for H, for which the indirect effect reaches -0.486. In comparison, the direct components of Bi are small, with values of -0.075 for A and 0.010 for H.

The total effects of L are -0.560 for A and -0.471 for H. The largest components are the pathways through M, which contribute -0.403 and -0.371, respectively. Across the three initial-condition variables, the indirect pathways involving S are comparatively small, ranging from -0.074 to 0.084.

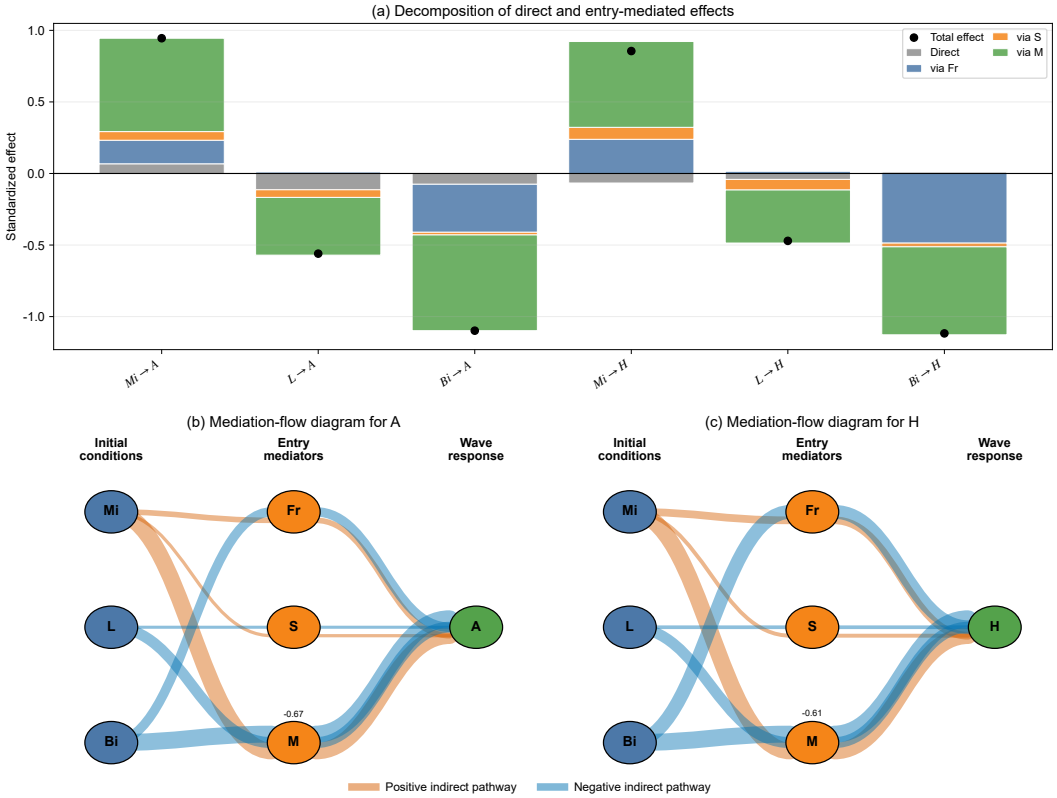


Figure 9. Direct, indirect, and total effects of initial-condition variables on wave responses: (**a**) decomposition of standardized effects into direct and entry-mediated components; and (**b**,**c**) dominant mediation-flow pathways for A and H, respectively.

Figure 10 provides a compact comparison of all pathway-specific indirect effects.

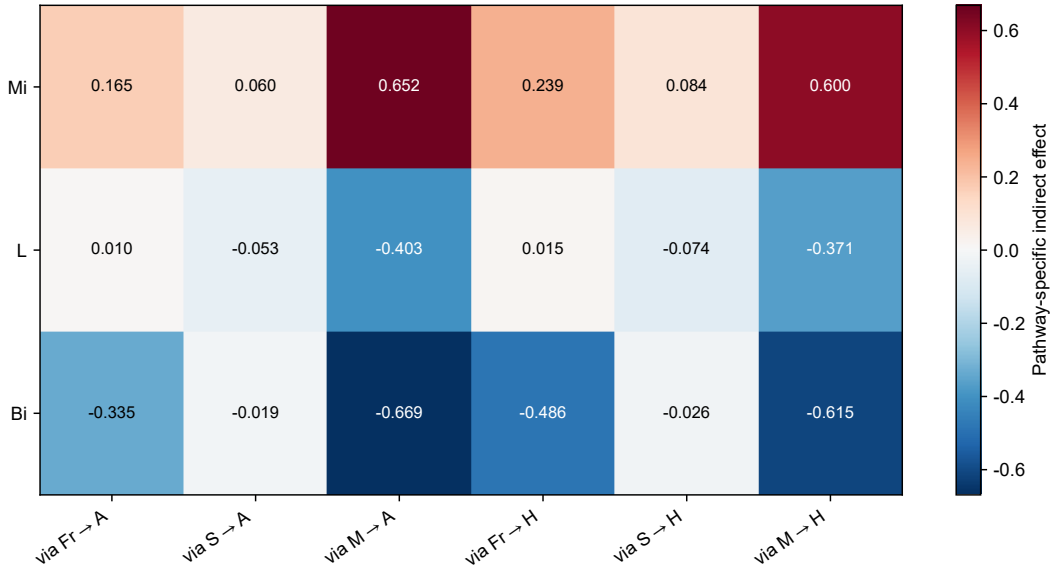


Figure 10. Heatmap of pathway-specific mediation contributions from initial-condition variables to wave responses through *Fr*, *S*, and *M*.

4.4. Predictive Implications of the Process-Chain Framework

4.4.1. Reconstruction of Water-Entry Variables

The initial-condition variables (M_i, L, Bi) were first used to reconstruct the intermediate water-entry variables *Fr*, *S*, and *M*. As shown in Figure 11, the reconstructed values are generally close to the corresponding observations. The models achieve R^2 values of 0.843 for *Fr*, 0.865 for *S*, and 0.824 for *M*. The associated MAE values are 0.038, 0.008, and 0.018, while the RMSE values are 0.050, 0.010, and 0.024, respectively.

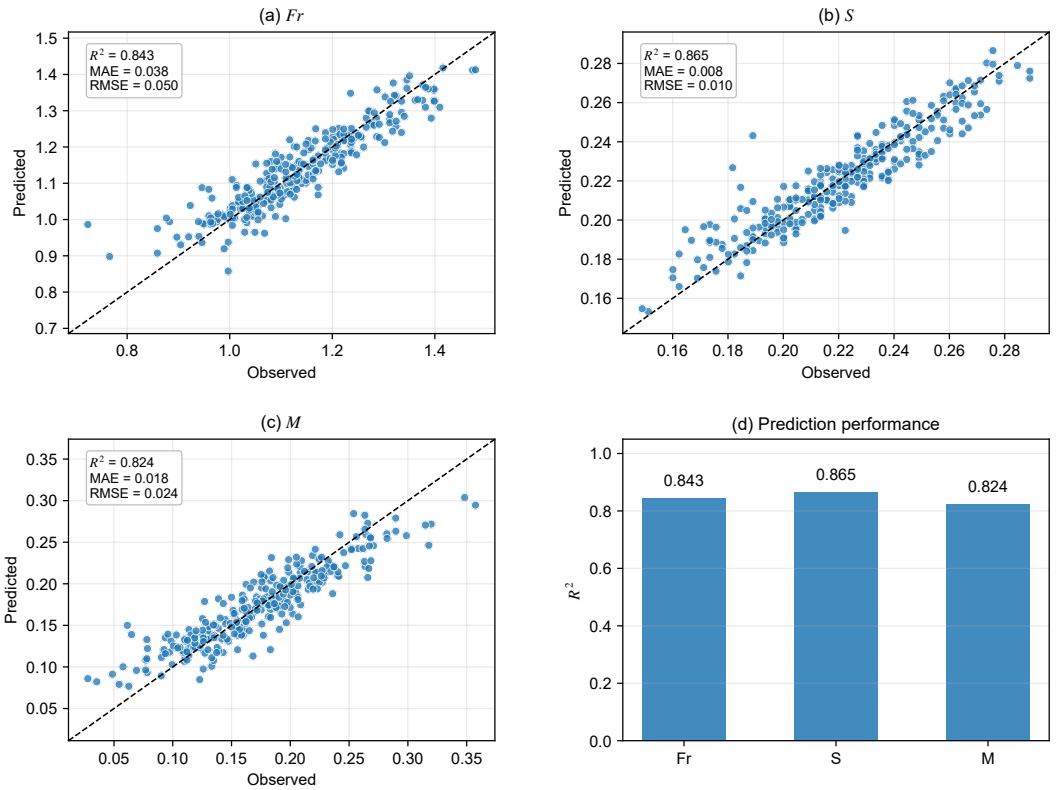


Figure 11. Prediction of water-entry variables from initial slide conditions: (**a**–**c**) observed and reconstructed values of *Fr*, *S*, and *M*, respectively; and (**d**) reconstruction performance in terms of R^2.

These results show that the realized water-entry variables can be systematically reconstructed from the initial-condition variables under the investigated experimental conditions, thereby enabling their use as an explicit intermediate layer in the cascade prediction framework.

4.4.2. Comparison of Wave-Response Modeling Frameworks

Figure 12 and Table 2 compare the fold-wise cross-validated performance of the three modeling frameworks. The direct initial-condition model achieves mean R^2 values of 0.822 ± 0.042 for A and 0.794 ± 0.045 for H. The observed entry-state model shows the highest and most stable performance, with mean R^2 values of 0.945 ± 0.008 for A and 0.924 ± 0.015 for H. Its MAE and RMSE values are also consistently lower than those of the other two frameworks.

Table 2. Fold-wise cross-validated performance of the three modeling frameworks. Values are reported as mean ± standard deviation across the five held-out folds.

Model	Target	R^2	MAE	RMSE
Direct initial-condition	A	0.822 ± 0.042	0.030 ± 0.003	0.041 ± 0.009
Direct initial-condition	H	0.794 ± 0.045	0.049 ± 0.006	0.066 ± 0.012
Observed entry-state	A	0.945 ± 0.008	0.017 ± 0.001	0.023 ± 0.003
Observed entry-state	H	0.924 ± 0.015	0.031 ± 0.001	0.040 ± 0.002
Cascade	A	0.831 ± 0.032	0.029 ± 0.003	0.040 ± 0.007
Cascade	H	0.794 ± 0.034	0.051 ± 0.005	0.066 ± 0.010

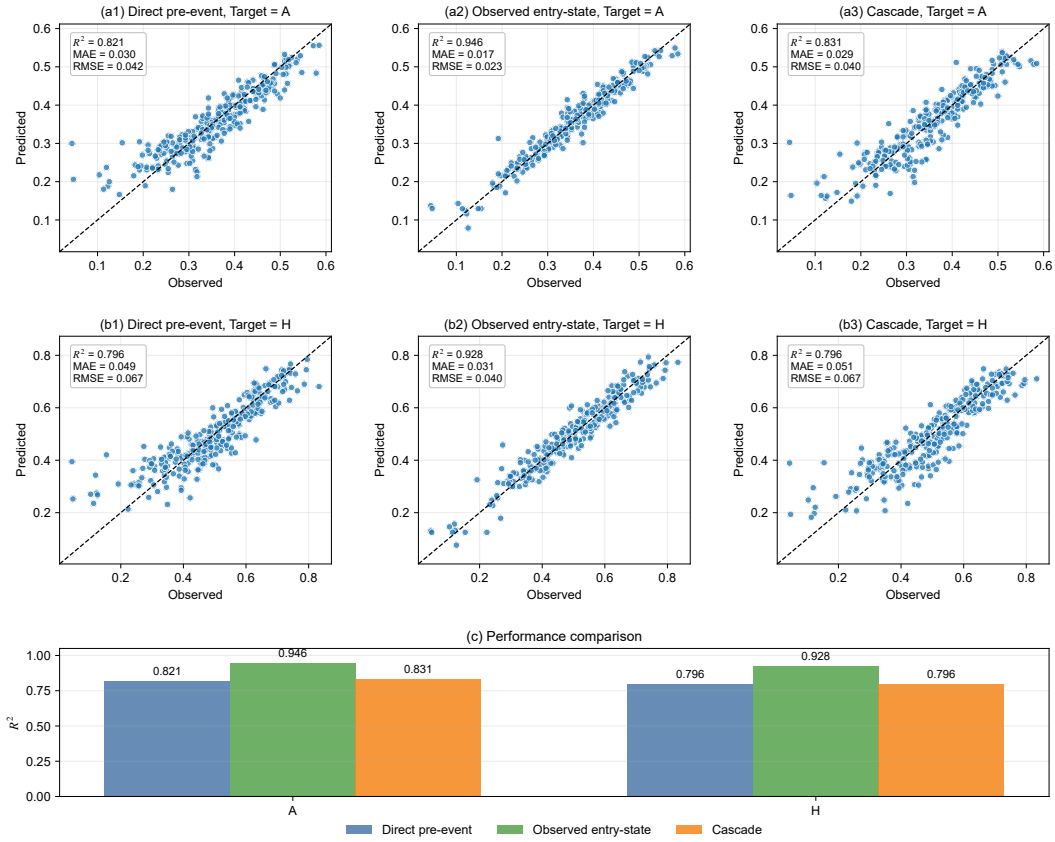


Figure 12. Prediction performance comparison of the direct initial-condition, observed entry-state, and cascade models for A and H. Panels (**a1**–**a3**) and (**b1**–**b3**) show observed-versus-predicted results for A and H, respectively, and panel (**c**) summarizes the corresponding R^2 values.

The cascade model achieves mean R^2 values of 0.831 ± 0.032 for A and 0.794 ± 0.034 for H. Its performance is therefore comparable to that of the direct initial-condition model. Although the cascade framework does not provide a substantial improvement in predictive

accuracy, it explicitly represents the reconstructed intermediate water-entry state and enables stage-specific interpretation of prediction errors.

Because the observed entry-state model uses variables measured during water entry, its higher cross-validated accuracy reflects the information content of the realized entry state and should not be interpreted as a strictly prospective predictive performance. The direct initial-condition and cascade models are the relevant prospective comparisons because both use only prescribed initial-condition information at the prediction stage.

4.5. Mechanism Mapping and Robustness Diagnostics

4.5.1. Path-Based Mechanism Map

As shown in Figure 13, within the fitted model and evaluated M_i–Bi grid, the pathway through M has the largest absolute model-derived contribution in approximately 79.4% of the grid cells, whereas the pathway through Fr has the largest contribution in approximately 20.6%. The pathway through S does not form a substantial region of largest contribution. These classifications describe relative pathway strengths within the fitted model and evaluated parameter space rather than deterministic physical regimes. The contour lines show the corresponding model-predicted values of H across the initial-condition space.

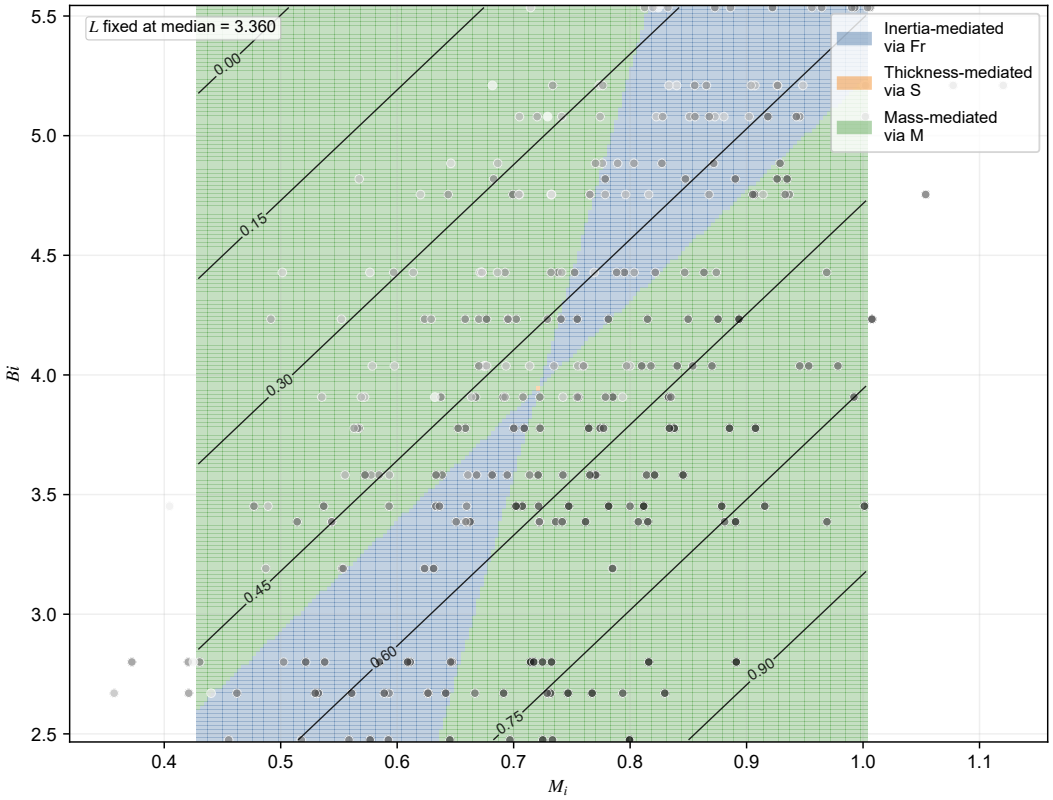


Figure 13. Path-based mechanism map in the M_i–Bi plane for predicting H, with L fixed at its median value. Background colors denote the mediator with the largest absolute model-derived contribution. The gray dots represent the experimental samples, with darker shades indicating larger observed values of H.

4.5.2. Bootstrap Stability of the Mediation Effects

Table 3 and Figure 14 show that the principal mass- and inertia-mediated pathways remain stable across bootstrap resampling, with their 95% confidence intervals excluding zero, whereas pathways involving S generally exhibit smaller indirect effects.

Table 3. Bootstrap confidence intervals of the dominant pathway-specific indirect effects.

Pathway	Mean	2.5%	97.5%
$M_i \rightarrow M \rightarrow A$	0.635	0.444	0.848
$M_i \rightarrow M \rightarrow H$	0.584	0.396	0.772
$Bi \rightarrow M \rightarrow A$	−0.653	−0.888	−0.459
$Bi \rightarrow M \rightarrow H$	−0.600	−0.805	−0.412
$Bi \rightarrow Fr \rightarrow H$	−0.486	−0.583	−0.394

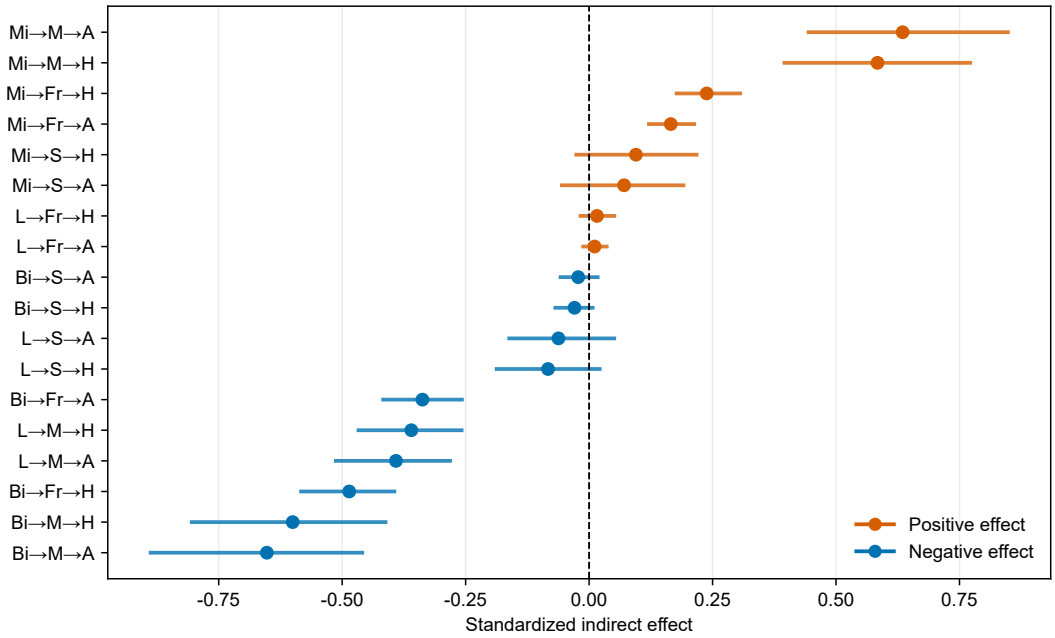


Figure 14. Bootstrap confidence intervals of pathway-specific mediation effects. Points denote mean standardized indirect effects, horizontal bars denote 95% bootstrap confidence intervals, and the vertical dashed line denotes zero effect.

4.5.3. Multicollinearity and Cascade Error Propagation

Figure 15 presents the multicollinearity diagnostics for the two stages of the path model. In the first-stage models, the VIF values of M_i, L, and Bi are 1.62, 1.18, and 1.72, respectively, indicating relatively weak linear dependence among the initial-condition predictors. In the second-stage models, the VIF values increase substantially and range from 6.16 to 11.72. The largest values are observed for Bi (11.72), S (11.53), and M (10.74), while the values for M_i, L, and Fr are 8.66, 6.34, and 6.16, respectively.

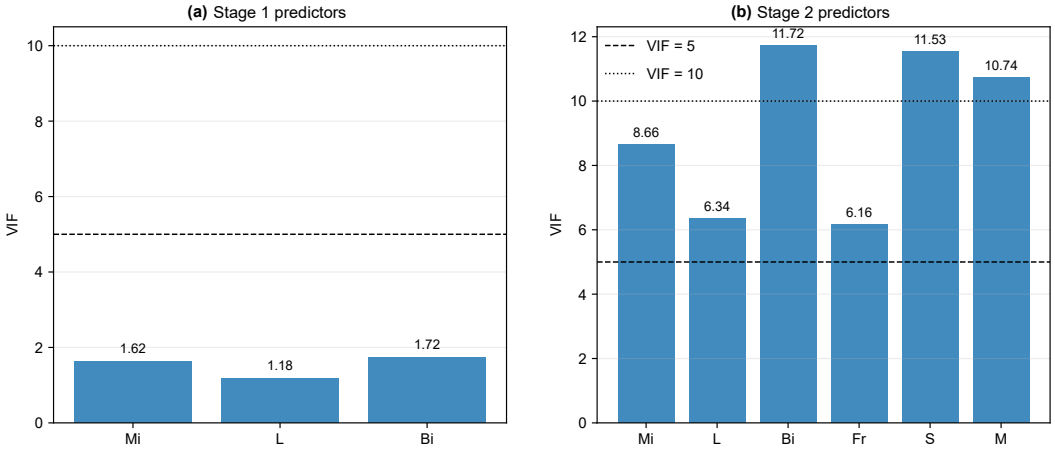


Figure 15. Variance inflation factor diagnostics for the first-stage and second-stage path models. Dashed and dotted horizontal lines denote reference VIF values of 5 and 10, respectively.

These results indicate stronger predictor coupling when the initial-condition and water-entry variables are included simultaneously in the second-stage models. Several second-stage VIF values exceed 10, suggesting that multicollinearity may affect the numerical stability and independent interpretability of individual path coefficients. This coupling is partly physically understandable because the two sets of variables describe closely related aspects of the same slide-entry process. Accordingly, the coefficients are interpreted as conditional associations within the specified process-chain structure, with greater emphasis on their direction, the overall pathway pattern, and consistency with the underlying physical process than on small differences in magnitude.

The cascade error-propagation results are shown in Figure 16. For the prediction of A, the reconstruction errors of Fr, S, and M are positively correlated with the final prediction error, with Spearman coefficients of 0.24, 0.22, and 0.42, respectively. For the prediction of H, the corresponding coefficients are 0.26, 0.12, and 0.35. The association between the reconstruction error of S and the prediction error of H is relatively weak ($\rho_s = 0.12$, $p = 0.042$), whereas the remaining associations are statistically significant at $p < 0.001$.

Among the three reconstructed water-entry variables, the error in M shows the strongest association with the prediction errors of both A and H. The reconstruction error of Fr shows the second-strongest association with the prediction error of H, while the associations involving S are generally weaker.

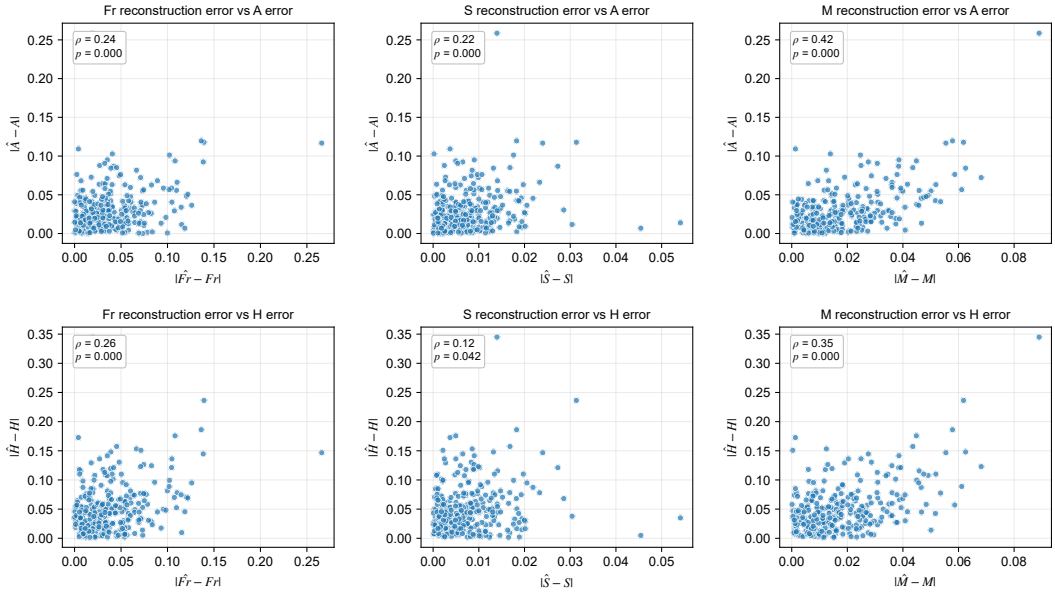


Figure 16. Cascade error propagation from reconstructed water-entry variables to wave-response prediction errors. Scatter plots show the relationships between the absolute reconstruction errors of Fr, S, and M and the absolute prediction errors of A and H.

5. Discussion

5.1. Process-Chain Interpretation and Relative Pathway Strengths

The combined correlation, path, and mediation results support interpreting viscoplastic landslide-generated impulse waves as a sequential process,

$$(M_i, L, Bi) \rightarrow (Fr, S, M) \rightarrow (A, H),$$

rather than solely as a direct response to the initial slide conditions. The stronger associations of the water-entry variables with the wave responses, together with the generally small direct paths after adjustment, indicate that much of the association between the initial conditions and wave responses is expressed through the realized entry state. This temporal

ordering is physically plausible, although the estimated pathways represent conditional associations within the specified model rather than strictly causal effects.

Among the intermediate variables, the relative effective entry mass M exhibits the strongest conditional association with both wave responses. Unlike the initial mass M_i, M represents the material effectively participating in water displacement during the principal wave-generation stage. It therefore incorporates the consequences of deformation, spreading, rheological resistance, and temporal redistribution before and during water entry. Slides with similar initial masses may consequently generate different waves when different proportions of their material participate effectively in water displacement.

Entry inertia, represented by Fr, provides a secondary but important pathway, particularly for H. The relative importance of the corresponding paths suggests that wave height is comparatively more sensitive to rapid momentum transfer, whereas wave amplitude is more closely associated with effective mass entry. The negative associations of Bi with both M and Fr are physically consistent with the role of the Bingham number as a measure of yield resistance relative to the downslope driving stress. Greater rheological resistance can reduce both entry inertia and the amount of material effectively participating in water entry.

The influence of sliding distance L is more complex. Its association with Fr is weak, whereas its paths to S and M are negative. Under the investigated conditions, a longer travel path may promote spreading, thinning, and internal redistribution before entry, thereby reducing the effective entry mass. This interpretation is specific to the present material, geometry, and parameter range and should not be generalized to other landslide settings.

Although S shows moderate marginal associations with the wave responses, its conditional contribution becomes small after M and Fr are included. The correlation between S and M suggests that part of the geometric information represented by entry thickness overlaps with that contained in effective entry mass. Within the present model, S is therefore better interpreted as a secondary geometric mediator than as an independent dominant control.

The mechanism map further illustrates the relative balance among these pathways. Within the fitted model and evaluated parameter space, the pathway through M has the largest absolute contribution over most of the M_i–Bi grid, whereas the pathway through Fr has the largest contribution over a smaller region and the pathway through S does not form a substantial region. These classifications represent relative pathway strengths within the fitted model and evaluated parameter space rather than deterministic physical regimes.

5.2. Predictive Implications and Robustness

The successful reconstruction of the water-entry variables from the initial conditions indicates that the realized entry state retains systematic information inherited from the initial slide. At the same time, the superior performance of the observed entry-state model confirms that variables measured during water entry are more proximal to wave generation than the initial conditions alone.

This distinction is important for interpreting the prediction results. The observed entry-state model uses measured Fr, S, and M and is therefore most appropriately viewed as an explanatory or diagnostic model of the slide-entry process. When the entry state must first be reconstructed from the initial conditions, the cascade model performs comparably to the direct initial-condition model rather than providing a substantial improvement in accuracy. For rapid prospective prediction based only on the initial slide conditions, the direct model is therefore simpler. The principal advantage of the cascade framework lies instead in its ability to represent the intermediate water-entry process and separate entry-state reconstruction errors from wave-response prediction errors. The direct and cascade models should thus be regarded as complementary.

The error-propagation analysis identifies the reconstruction of M as the main source of uncertainty in the cascade predictions. Because M is also the most prominent second-stage predictor, improving its reconstruction is likely to provide the greatest benefit to future cascade models. This may require additional descriptors of slide deformation, spreading, and the temporal evolution of water entry.

Bootstrap resampling preserves the principal mass- and inertia-related pathways, supporting their qualitative stability. However, the elevated second-stage VIF values indicate substantial coupling among the initial-condition and water-entry variables. Such coupling is physically understandable because these variables describe related aspects of the same slide-entry process, but it limits the independent interpretability of individual coefficients. The robustness results therefore support the direction and overall organization of the identified pathways more strongly than precise differences in their numerical magnitudes.

From a computational perspective, the framework is relatively lightweight because it uses a limited number of dimensionless variables and combines linear path analysis with tree-based regression. Once trained, it can provide rapid wave-response estimates together with pathway-oriented interpretation at substantially lower computational cost than high-fidelity CFD or particle-based simulations. It cannot, however, reproduce detailed free-surface evolution, velocity fields, local pressures, or complex three-dimensional effects.

5.3. Limitations and Future Work

Several limitations should be acknowledged. First, the analysis is based on controlled experiments using Carbopol material, a fixed flume geometry, and restricted parameter ranges. Although the general process-chain organization, $\mathbf{X}_0 \rightarrow \mathbf{X}_e \rightarrow \mathbf{Y}$, is conceptually transferable to other experimental or numerical datasets, the present path coefficients and predictive relationships require recalibration for other materials, geometries, and scales. Natural landslides involve additional complexities, including material heterogeneity, fragmentation, entrainment, pore-pressure evolution, saturation, three-dimensional topography and spreading, and scale effects. Field application would therefore require appropriately defined process variables and independent validation against site-specific observations.

Second, the effective entry mass m_e was defined as the amount of material that had entered the water when the maximum wave height was attained. Although this definition characterizes the material participating in the principal wave-generation stage, it links the timing of the mediator to the observed wave response. Consequently, models using the measured effective entry mass should not be interpreted as strictly pre-event predictive models. Future studies should examine the sensitivity of the identified pathways to independent definitions based on fixed times after first contact, prescribed penetration depths, or predefined entry stages.

Third, the path analysis assumes a linear additive structure. Although the tree-based prediction models can represent nonlinear relationships, they do not establish that the individual path coefficients are invariant to alternative model specifications. Future work should evaluate nonlinear and interaction effects and compare the present results with regularized, dimension-reduction, and alternative mediator models.

Finally, the present analysis considers only relative maximum wave amplitude and height. The framework could be extended to wave period, attenuation, run-up, impact pressure, and overtopping volume. Validation through larger-scale experiments, numerical simulations, and field observations would help determine whether the identified mass- and inertia-related pathways remain stable under more complex conditions.

6. Conclusions

This study developed a three-stage process-chain framework, $(M_i, L, Bi) \rightarrow (Fr, S, M) \rightarrow (A, H)$, to examine how the initial conditions of viscoplastic landslides are associated with impulse-wave responses through the intermediate water-entry process. Evidence from 290 physical-model tests supports the following conclusions:

1. The realized water-entry state is more closely associated with wave generation than the initial slide conditions alone. Among the entry-state variables, the relative effective entry mass M shows the strongest conditional association with both wave amplitude and wave height, while Fr provides a secondary but important inertia-related pathway. The independent contribution of S is comparatively limited after M and Fr are considered.
2. Within the present dataset and fitted model, the mass-related and inertia-related pathways account for the largest modeled contributions linking the initial slide conditions to the wave responses. The largest modeled contribution associated with initial slide mass occurs through effective entry mass, whereas increasing Bi is associated with reductions in both effective entry mass and entry inertia. The pathway through M has the largest absolute model-derived contribution over most of the evaluated parameter space. These pathways represent conditional associations within the present dataset and fitted model rather than deterministic causal relationships.
3. The observed entry-state model, which uses measured entry-state variables, provides an explanatory or diagnostic benchmark and achieves higher cross-validated accuracy than models using initial-condition information alone. For prospective prediction from prescribed initial conditions, the cascade model performs comparably to the direct initial-condition model. Its principal advantage therefore lies in process interpretation and stage-specific error diagnosis rather than improved predictive accuracy. Reconstruction of M is the main source of uncertainty in the cascade predictions.

Overall, the framework distinguishes the mass-related and inertia-related routes through which initial slide conditions are associated with impulse-wave responses. Its applicability beyond the present material, geometry, and experimental scale requires independent validation.

Author Contributions: Conceptualization, Z.Y. and Y.B.; methodology, Z.Y. and Y.B.; validation, S.Y., X.L. and X.P.; formal analysis, Z.Y. and B.C.; data curation, Z.M.; writing—original draft preparation, Z.Y.; writing—review and editing, Z.Y.; supervision, Y.B.; project administration, Y.B.; funding acquisition, Z.Y. and Y.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Henan Provincial Key Research and Development Program (Grant No. 251111231500), Central Guidance Funds for Science and Technology Local Development Projects (Grant No. 2025ZY01091).

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author, as the data are currently being used in ongoing follow-up research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following symbols and abbreviations are used in this manuscript:

m_i	Initial slide mass (kg)
s_i	Initial slide thickness at the gate (m)
l_s	Sliding distance (m)
τ_c	Yield stress (Pa)

u_0	Entry velocity (m s^{-1})
s_e	Entry thickness at the still-water surface (m)
m_e	Effective entry mass (kg)
h_0	Still-water depth (m)
a_m	Maximum wave amplitude (m)
h_m	Maximum wave height (m)
M_i	Relative initial slide mass (dimensionless)
L	Relative sliding distance (dimensionless)
Bi	Bingham number (dimensionless)
Fr	Entry Froude number (dimensionless)
S	Relative entry thickness (dimensionless)
M	Relative effective entry mass (dimensionless)
A	Relative maximum wave amplitude (dimensionless)
H	Relative maximum wave height (dimensionless)
ρ_s	Spearman's rank correlation coefficient
GBR	Gradient boosting regression
VIF	Variance inflation factor
MAE	Mean absolute error
RMSE	Root mean square error
R^2	Coefficient of determination

Appendix A. Supplementary Path-Model Robustness Analyses

Appendix A.1. Alternative Specifications and Diagnostic Procedures

All variables were standardized before analysis. Ridge regression was fitted to each component equation using 161 candidate penalties from 10^{-4} to 10^{4}, with the optimal penalty selected by five-fold cross-validation. The OLS and Ridge coefficients are compared in Table A1. The second-stage equations were also evaluated by omitting Fr, S, or M individually (Table A2) and by adding quadratic and two-way interaction terms among the mediators (Table A3).

HC3-robust standard errors were calculated for the OLS equations (Table A4). Residual and influence diagnostics included heteroskedasticity, normality and specification tests, studentized residuals, leverage, Cook's distance, and DFBETAs (Table A5 and Figure A1). The adopted thresholds were $|r_i| > 3$, leverage $> 2p/n$, Cook's distance $> 4/n$, and $|\text{DFBETA}| > 2/\sqrt{n}$. Influence sensitivity was further assessed by temporarily excluding Cook-flagged observations and refitting each equation (Table A6). The overall coefficient comparison is shown in Figure A2.

Appendix A.2. Results of the Robustness Analyses

Ridge regularization retained the directions of 20 of the 21 OLS paths (Table A1). The only sign change involved the weak and nonsignificant $Bi \rightarrow H$ path. The leave-one-mediator-out results showed that removing M or Fr reduced the predictive performance of the A and H equations, whereas removing S had little effect (Table A2). This confirms the dominant contributions of M and Fr.

The augmented specification produced modest improvements in cross-validated R^2 while retaining the positive effects of M and Fr (Table A3). HC3 inference likewise preserved the significance of the principal paths, whereas several weaker paths remained nonsignificant (Table A4).

Although the diagnostics identified heteroskedasticity, non-normal residuals, and influential observations, particularly for A and H (Table A5 and Figure A1), the Cook-based refitting generally preserved the coefficient directions and the dominant effects of M and Fr (Table A6). Overall, Figure A2 indicates that the principal pathway structure was

broadly stable across the alternative specifications, although weaker paths should be interpreted cautiously.

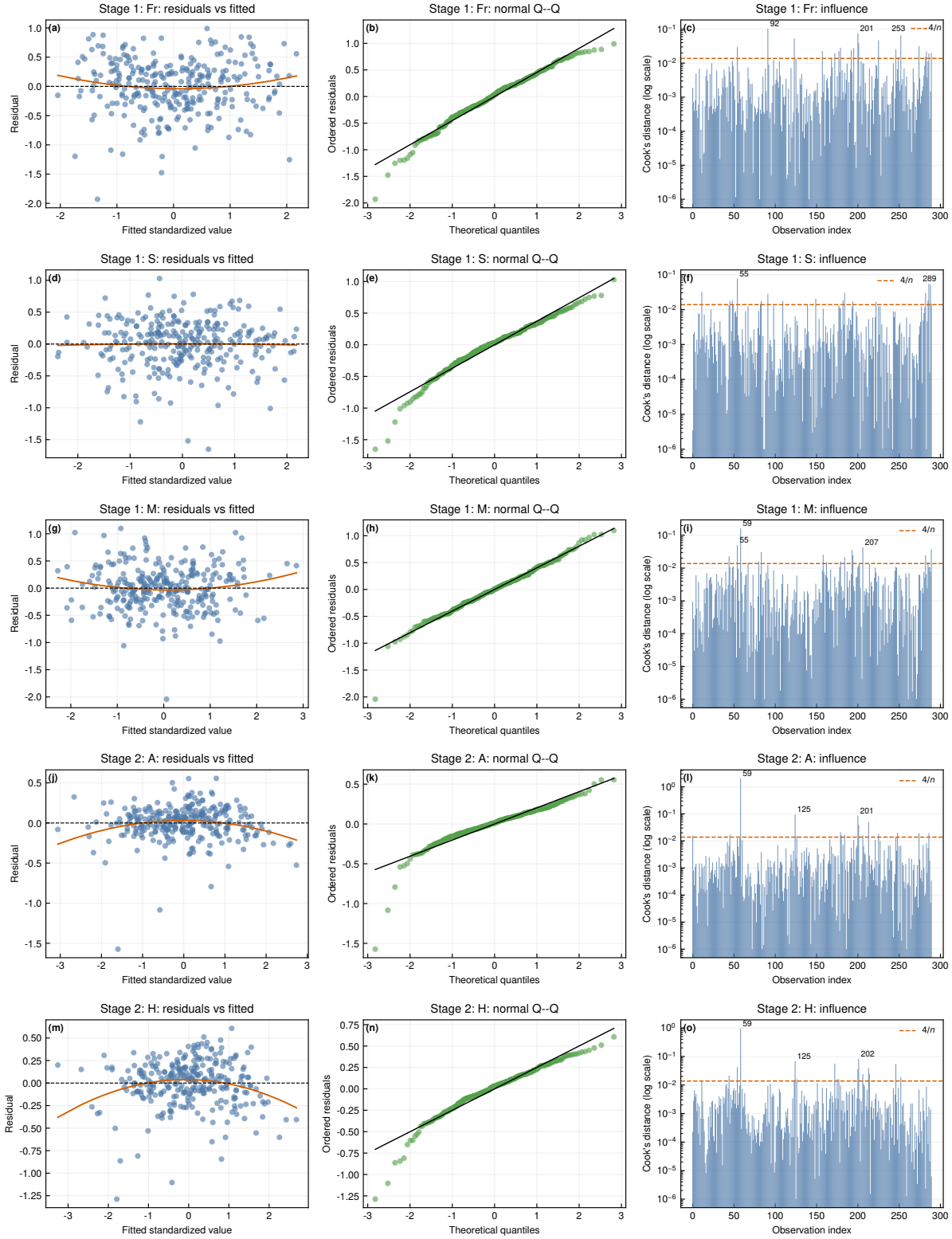


Figure A1. Residual and influence diagnostics for the five component path equations. The first column shows standardized residuals versus fitted values with a quadratic smooth; the second column shows normal Q–Q plots; and the third column shows Cook's distance on a logarithmic scale. The dashed horizontal line in the influence panels denotes $4/n$, and labels identify the three largest Cook's-distance values in each equation.

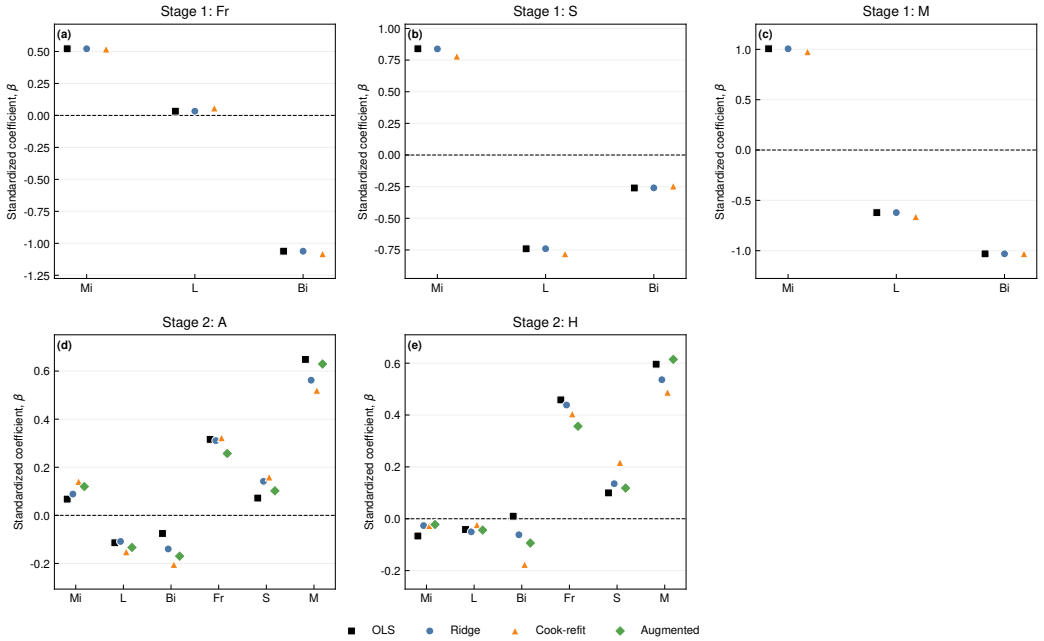


Figure A2. Sensitivity of the standardized path coefficients to Ridge regularization, temporary exclusion of Cook-flagged observations, and the augmented nonlinear specification. The augmented specification applies only to the second-stage equations.

Table A1. Comparison of OLS and Ridge standardized path coefficients.

Equation	Predictor	OLS β	OLS p	Ridge β	λ	Same Sign
Fr	M_i	0.521	<0.001	0.521	0.0001	Yes
Fr	L	0.033	0.259	0.033	0.0001	Yes
Fr	Bi	−1.061	<0.001	−1.061	0.0001	Yes
S	M_i	0.839	< 0.001	0.838	0.1585	Yes
S	L	−0.741	<0.001	−0.740	0.1585	Yes
S	Bi	−0.261	<0.001	−0.260	0.1585	Yes
M	M_i	1.006	<0.001	1.005	0.0447	Yes
M	L	−0.622	<0.001	−0.621	0.0447	Yes
M	Bi	−1.031	<0.001	−1.031	0.0447	Yes
A	M_i	0.068	0.073	0.089	8.9125	Yes
A	L	−0.114	<0.001	−0.108	8.9125	Yes
A	Bi	−0.075	0.086	−0.140	8.9125	Yes
A	Fr	0.316	<0.001	0.311	8.9125	Yes
A	S	0.072	0.099	0.142	8.9125	Yes
A	M	0.649	<0.001	0.562	8.9125	Yes
H	M_i	−0.067	0.138	−0.026	5.0119	Yes
H	L	−0.041	0.280	−0.050	5.0119	Yes
H	Bi	0.010	0.850	−0.062	5.0119	No
H	Fr	0.458	<0.001	0.439	5.0119	Yes
H	S	0.100	0.055	0.135	5.0119	Yes
H	M	0.596	<0.001	0.536	5.0119	Yes

Table A2. Leave-one-mediator-out results. Cross-validation values are means ± one standard deviation across five held-out folds.

Target	Specification	β_{Fr}	β_S	β_M	Adjusted R^2	CV R^2	CV RMSE
A	Full	0.316	0.072	0.649	0.953	0.943 ± 0.012	0.0230 ± 0.0045
A	Without *Fr*	–	0.133	0.773	0.936	0.926 ± 0.014	0.0262 ± 0.0041
A	Without *S*	0.323	–	0.686	0.952	0.944 ± 0.008	0.0229 ± 0.0038
A	Without *M*	0.462	0.431	–	0.913	0.911 ± 0.043	0.0286 ± 0.0094
H	Full	0.458	0.100	0.596	0.933	0.924 ± 0.014	0.0399 ± 0.0056
H	Without *Fr*	–	0.188	0.776	0.899	0.889 ± 0.020	0.0483 ± 0.0059
H	Without *S*	0.469	–	0.647	0.932	0.924 ± 0.012	0.0399 ± 0.0043
H	Without *M*	0.593	0.430	–	0.900	0.895 ± 0.041	0.0468 ± 0.0118

Table A3. Comparison of the linear-additive and augmented quadratic/interaction specifications.

Target	Specification	Adjusted R^2	AIC	BIC	CV R^2	CV RMSE	β_M	β_{Fr}	HC3 Joint p
A	Linear additive	0.953	−54.20	−28.51	0.943 ± 0.012	0.0230 ± 0.0045	0.649	0.316	–
A	Quadratic + interactions	0.963	−119.54	−71.83	0.947 ± 0.015	0.0222 ± 0.0052	0.629	0.257	<0.001
H	Linear additive	0.933	46.77	72.46	0.924 ± 0.014	0.0399 ± 0.0056	0.596	0.458	–
H	Quadratic + interactions	0.945	−6.59	41.12	0.938 ± 0.011	0.0360 ± 0.0034	0.614	0.357	<0.001

Table A4. OLS and HC3-robust inference for the standardized path coefficients.

Equation	Predictor	β	OLS SE	OLS p	HC3 SE	HC3 p
Fr	M_i	0.521	0.034	<0.001	0.042	<0.001
Fr	*L*	0.033	0.029	0.259	0.039	0.393
Fr	*Bi*	−1.061	0.035	<0.001	0.046	<0.001
S	M_i	0.839	0.028	<0.001	0.032	<0.001
S	*L*	−0.741	0.024	<0.001	0.022	<0.001
S	*Bi*	−0.261	0.029	<0.001	0.029	<0.001
M	M_i	1.006	0.030	<0.001	0.035	<0.001
M	*L*	−0.622	0.026	<0.001	0.028	<0.001
M	*Bi*	−1.031	0.031	<0.001	0.034	<0.001
A	M_i	0.068	0.038	0.073	0.067	0.315
A	*L*	−0.114	0.032	<0.001	0.049	0.022
A	*Bi*	−0.075	0.044	0.086	0.100	0.452
A	*Fr*	0.316	0.032	<0.001	0.042	<0.001
A	*S*	0.072	0.043	0.099	0.097	0.457
A	*M*	0.649	0.042	<0.001	0.151	<0.001
H	M_i	−0.067	0.045	0.138	0.061	0.279
H	*L*	−0.041	0.038	0.280	0.050	0.405
H	*Bi*	0.010	0.052	0.850	0.092	0.915
H	*Fr*	0.458	0.038	<0.001	0.048	<0.001
H	*S*	0.100	0.052	0.055	0.087	0.251
H	*M*	0.596	0.050	<0.001	0.129	<0.001

Table A5. Residual and influence diagnostics.

Equation	BP p	White p	JB p	RESET p	$\lvert r_i \rvert > 3$	High Leverage	$D_i > 4/n$	Max D_i
Fr	0.002	<0.001	<0.001	0.079	2	17	21	0.104
S	0.019	0.003	<0.001	0.837	3	17	19	0.078
M	0.133	0.005	<0.001	0.053	1	17	17	0.162
A	<0.001	<0.001	<0.001	<0.001	3	16	17	2.044
H	<0.001	<0.001	<0.001	<0.001	5	16	17	0.967

Table A6. Summary of the Cook-refit influence-sensitivity analysis.

Equation	Cases Temporarily Omitted	Directions Retained	Max $\lvert\Delta\beta\rvert$	Refit β_M	Refit β_{Fr}
Fr	21	3/3	0.024	–	–
S	19	3/3	0.059	–	–
M	17	3/3	0.041	–	–
A	17	6/6	0.129	0.520	0.322
H	17	5/6	0.186	0.488	0.404

References

1. Ward, S.N. Landslide tsunami. *J. Geophys. Res. Solid Earth* **2001**, *106*, 11201–11215. [CrossRef]
2. Walder, J.S.; Watts, P.; Sorensen, O.E.; Janssen, K. Tsunamis generated by subaerial mass flows. *J. Geophys. Res. Solid Earth* **2003**, *108*, 2236. [CrossRef]
3. Panizzo, A.; De Girolamo, P.; Petaccia, A. Forecasting impulse waves generated by subaerial landslides. *J. Geophys. Res. Oceans* **2005**, *110*, C12025. [CrossRef]
4. Løvholt, F.; Pedersen, G.; Harbitz, C.B.; Glimsdal, S.; Kim, J. On the characteristics of landslide tsunamis. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *373*, 20140376. [CrossRef]
5. Slingerland, R.L.; Voight, B. Occurrences, properties, and predictive models of landslide-generated water waves. In *Developments in Geotechnical Engineering*; Elsevier: Amsterdam, The Netherlands, 1979; Volume 14, pp. 317–394.
6. Heller, V.; Hager, W.H.; Minor, H.E. Scale effects in subaerial landslide generated impulse waves. *Exp. Fluids* **2008**, *44*, 691–703. [CrossRef]
7. Heller, V.; Hager, W.H. Impulse product parameter in landslide generated impulse waves. *J. Waterw. Port Coast. Ocean Eng.* **2010**, *136*, 145–155. [CrossRef]
8. Heller, V.; Hager, W.H. Wave types of landslide generated impulse waves. *Ocean Eng.* **2011**, *38*, 630–640. [CrossRef]
9. Lindstrøm, E.K.; Pedersen, G.K.; Jensen, A.; Glimsdal, S. Experiments on slide generated waves in a 1:500 scale fjord model. *Coast. Eng.* **2014**, *92*, 12–23. [CrossRef]
10. Heller, V.; Spinneken, J. On the effect of the water body geometry on landslide–tsunamis: Physical insight from laboratory tests and 2D to 3D wave parameter transformation. *Coast. Eng.* **2015**, *104*, 113–134. [CrossRef]
11. Evers, F.M.; Hager, W.H. Spatial impulse waves: Wave height decay experiments at laboratory scale. *Landslides* **2016**, *13*, 1395–1403. [CrossRef]
12. Zheng, F.; Liu, Q.; Shen, L.; Zhai, X.; Zhang, X.; Xu, J. Experimental study on the characterization of landslide-generated waves in water bodies with rigid vegetation. *Ocean Eng.* **2025**, *317*, 120026. [CrossRef]
13. Du, Z.h.; Zhou, J.W.; Zhang, S.C.; Zhong, Q.M.; Li, H.B.; Hu, Y.X.; Li, C.J.; Zhang, J.Y. Experimental analysis on breaching mechanism of earth-rock dam induced by landslide generated waves. *Eng. Geol.* **2025**, *346*, 107913. [CrossRef]
14. Lynett, P.; Liu, P.L.F. A numerical study of submarine–landslide–generated waves and run–up. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **2002**, *458*, 2885–2910. [CrossRef]
15. Ma, G.; Kirby, J.T.; Hsu, T.J.; Shi, F. A two-layer granular landslide model for tsunami wave generation: Theory and computation. *Ocean Model.* **2015**, *93*, 40–55. [CrossRef]
16. Hu, Y.x.; Yu, Z.y.; Zhou, J.w. Numerical simulation of landslide-generated waves during the 11 October 2018 Baige landslide at the Jinsha River. *Landslides* **2020**, *17*, 2317–2328. [CrossRef]
17. Feng, X.; Cheng, L.; Dong, Q.; Qi, X.; Xiong, C. Numerical study of hydraulic characteristics of impulse waves generated by subaerial landslides. *AIP Adv.* **2022**, *12*, 125118. [CrossRef]
18. Yavari-Ramshe, S.; Ataie-Ashtiani, B. Numerical modeling of subaerial and submarine landslide-generated tsunami waves—recent advances and future challenges. *Landslides* **2016**, *13*, 1325–1368. [CrossRef]
19. Li, Y.; Zhong, Q.; Wu, H.; Liu, J.; Mei, S.; Wang, H. Numerical simulation of landslide-generated waves based on multiphase smoothed particle hydrodynamics method with a case study of Wangjiashan landslide. *Environ. Earth Sci.* **2025**, *84*, 35. [CrossRef]
20. Li, N.; Hu, X.; Zheng, H.C.; Li, Y.; Wu, Z.; Xu, C.; Li, W. A Bayesian method to optimize empirical wave height model in subaerial landslide-generated wave hazard probabilistic assessment. *Bull. Eng. Geol. Environ.* **2025**, *84*, 1–20. [CrossRef]
21. Liu, Y.; Wang, H.; Li, J.; Wu, Y.; Wang, W. A SPH-DEM coupling method based on mixture theory for landslide-generated waves. *Comput. Geotech.* **2026**, *191*, 107764. [CrossRef]
22. Ma, H.; Wang, H.; Xu, W.; Chen, S.; Liu, Y.; Chen, G. Numerical estimation of landslide-generated impulse waves induced by high-elevation landslides in mountain reservoirs using a coupled discrete element and smoothed particle hydrodynamics method. *Phys. Fluids* **2025**, *37*, 116601. [CrossRef]
23. Wang, W.; Xia, H.; Chai, B.; Lu, X. SPH-Based Numerical Investigation of Energy Transformation and Prediction of Wave Characteristics in Granular Landslide–Generated Impulsive Waves. *J. Hydrol. Eng.* **2025**, *30*, 04025009. [CrossRef]

24. Shi, C.; Wang, X.; Liu, Q. Wave generation, energy conversion, and sediment accumulation in a real-scale channel-like reservoir by landslide using a soil-water coupling SPH model. *Landslides* **2025**, *22*, 2769–2783. [CrossRef]
25. Petley, D.N.; Higuchi, T.; Petley, D.J.; Bulmer, M.H.; Carey, J. Development of progressive landslide failure in cohesive materials. *Geology* **2005**, *33*, 201–204. [CrossRef]
26. Bird, R.B.; Dai, G.; Yarusso, B.J. The rheology and flow of viscoplastic materials. *Rev. Chem. Eng.* **1983**, *1*, 1–70. [CrossRef]
27. Qin, P.; Meng, Z.; Su, H.; Cheng, C. A Novel Permutation Entropy–Based Method for Assessing the Stability of Seawalls on Soft Soils. *Struct. Control Health Monit.* **2026**, *2026*, 3016498. [CrossRef]
28. Kaitna, R.; Rickenmann, D.; Schatzmann, M. Experimental study on rheologic behaviour of debris flow material. *Acta Geotech.* **2007**, *2*, 71–85. [CrossRef]
29. Chen, C.L. Generalized viscoplastic modeling of debris flow. *J. Hydraul. Eng.* **1988**, *114*, 237–258. [CrossRef]
30. Milledge, D.G.; Bellugi, D.; McKean, J.A.; Densmore, A.L.; Dietrich, W.E. A multidimensional stability model for predicting shallow landslide size and shape across landscapes. *J. Geophys. Res. Earth Surf.* **2014**, *119*, 2481–2504. [CrossRef]
31. Meng, Z. Experimental study on impulse waves generated by a viscoplastic material at laboratory scale. *Landslides* **2018**, *15*, 1173–1182. [CrossRef]
32. Meng, Z.; Zhang, J.; Hu, Y.; Ancey, C. Temporal Prediction of Landslide-Generated Waves Using a Theoretical–Statistical Combined Method. *J. Mar. Sci. Eng.* **2023**, *11*, 1151. [CrossRef]
33. Meng, Z.; Ancey, C. The effects of slide cohesion on impulse-wave formation. *Exp. Fluids* **2019**, *60*, 151. [CrossRef]
34. Darvenne, A.; Viroulet, S.; Lacaze, L. Physical model of landslide-generated impulse waves: Experimental investigation of the wave-granular flow coupling. *J. Geophys. Res. Ocean* **2024**, *129*, e2024JC021145. [CrossRef]
35. Hibert, C.; Noël, F.; Toe, D.; Talib, M.; Desrues, M.; Wyser, E.; Brenguier, O.; Bourrier, F.; Toussaint, R.; Malet, J.P.; et al. Machine learning prediction of the mass and the velocity of controlled single-block rockfalls from the seismic waves they generate. *Earth Surf. Dyn.* **2024**, *12*, 641–656. [CrossRef]
36. Yuan, Z.; Bian, Y.; Liu, W.; Qi, F.; Ma, H.; Zheng, S.; Meng, Z. Predicting the Deformation of a Slope Using a Random Coefficient Panel Data Model. *Fractal Fract.* **2024**, *8*, 429. [CrossRef]
37. Hu, Y.; Wang, Y.; Wei, B.; Meng, Z.; Yuan, D.; Jin, T. An interpretable dynamic evaluation framework fusing multi-dimensional data for assessing the operational safety of concrete dams. *Expert Syst. Appl.* **2025**, *299*, 130225. [CrossRef]
38. Lyu, C.; Xu, W.; Huang, Q.; Tian, L.; Shi, H.; Chen, H.; Liu, Y.; Lei, J. Predicting landslide surge waves from large-scale physical experimental using machine learning. *Phys. Fluids* **2025**, *37*, 036605. [CrossRef]
39. Mao, P.; Lei, J.; Tian, L. Research on the Spatiotemporal Evolution Patterns of Landslide-Induced Surge Waves Based on Physical Model Experiments. *Water* **2025**, *17*, 685. [CrossRef]
40. Zhou, X.; Wen, H.; Li, Z.; Zhang, H.; Zhang, W. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. *Geocarto Int.* **2022**, *37*, 13419–13450. [CrossRef]
41. Lin, N.; Zhang, D.; Feng, S.; Ding, K.; Tan, L.; Wang, B.; Chen, T.; Li, W.; Dai, X.; Pan, J.; et al. Rapid landslide extraction from high-resolution remote sensing images using SHAP-OPT-XGBoost. *Remote Sens.* **2023**, *15*, 3901. [CrossRef]
42. Inan, M.S.K.; Rahman, I. Explainable AI integrated feature selection for landslide susceptibility mapping using TreeSHAP. *SN Comput. Sci.* **2023**, *4*, 482. [CrossRef]
43. Merabet, K.; Di Nunno, F.; Granata, F.; Kim, S.; Adnan, R.M.; Heddam, S.; Kisi, O.; Zounemat-Kermani, M. Predicting water quality variables using gradient boosting machine: Global versus local explainability using SHapley Additive Explanations (SHAP). *Earth Sci. Inform.* **2025**, *18*, 1–34. [CrossRef]
44. Xu, X.; Qin, P.; Li, Z.; Wang, J.; Zhou, Y.; Zheng, S.; Meng, Z. Interpretation Analysis of Influential Variables Dominating Impulse Waves Generated by Landslides. *J. Mar. Sci. Eng.* **2025**, *13*, 2223. [CrossRef]
45. Açıkkar, M.; Aydın, B. Deep learning-based landslide tsunami run-up prediction from synthetic gage data. *Appl. Ocean Res.* **2025**, *154*, 104360. [CrossRef]
46. Li, N.; Hu, X.; Wang, J.; Huang, J.; Zheng, H.; Li, W. A physics-based probabilistic method for assessment of landslide-induced wave run-up hazards. *J. Rock Mech. Geotech. Eng.* **2025**, *18*, 3675–3687. [CrossRef]
47. Ding, Q.; Ma, G.; Guo, C.; Zhang, G.; Mei, J.; Zhou, W. Optimizing reservoir landslide susceptibility mapping with physics-enhanced data-driven models. *Bull. Eng. Geol. Environ.* **2026**, *85*, 64. [CrossRef]
48. Fang, H.; Tan, K.; Hooker, G. Statistical Inference for Gradient Boosting Regression. *Adv. Neural Inf. Process. Syst.* **2026**, *38*, 32746–32788. [CrossRef]
49. Rizkallah, L.W. Enhancing the performance of gradient boosting trees on regression problems. *J. Big Data* **2025**, *12*, 35. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.