Journal of Information-Based Decision Making Volume 1, Issue 1 (2026) 31-43

Journal of Information-Based Decision Making

Journal homepage: www.jibdm.journal-publishing.org
ISSN: XXXX-XXXX

Deformation-Scaled Conformal Prediction for Short-Term Landslide Displacement: A Granular Uncertainty Assessment Framework for Decision Support

Yongfei Wu[1], Gang Cheng[1,2,*], Haoran Zhang[1], Daisong Yang[1], Jingjing Xia[1]

[1] School of Computer and Information Security, University of Emergency Management, Beijing, China
[2] National Field Observation and Research Station (Hebei Cangzhou) for Groundwater and Land Subsidence, Cangzhou, China

ARTICLE INFO

Article history:
Received 31 May 2026
Received in revised form 2 July 2026
Accepted 3 August 2026
Available online 6 September 2026

Keywords:
Conformal Prediction; Landslide Displacement; Prediction Interval; Uncertainty Quantification; Granular Computing; Decision Support.

ABSTRACT

Deterministic forecasts conceal whether a predicted landslide displacement is sufficiently reliable for operational decision-making. This study develops a lightweight uncertainty-aware framework that represents a monitored slope through four spatial granules and three deformation-state granules. Daily records of displacement from global navigation satellite system (GNSS) monitoring points, rainfall, and reservoir water level at the Xinpu reservoir landslide were aggregated into four series corresponding to Zones I–IV. A LightGBM model predicted the seven-day displacement increment from a 30-day history, while split conformal prediction (Split-CP) and deformation-scaled conformal prediction (DSCP) constructed nominal 90% prediction intervals around the point forecasts. DSCP scales calibration residuals by the root-mean-square magnitude of the preceding 14-day displacement increments. During the chronological test period in 2021, LightGBM achieved an overall mean absolute error of 1.243 mm, 46.2% lower than that of persistence. Split-CP attained 86.20% empirical coverage with a mean width of 5.396 mm, while DSCP increased coverage to 89.69% and reduced the mean interval score by 7.2%, with a modest 4.5% increase in interval width. Interval width was strongly associated with forecast-origin deformation velocity (overall Spearman's $\rho = 0.768$). Conditional evaluation nevertheless revealed a failure mode: coverage in the rapid state of Zone III was only 68.0%. The results demonstrate that uncertainty should be evaluated across spatial and deformation-state granules rather than inferred from an aggregate accuracy score. The proposed framework provides a point forecast, prediction interval, and relative confidence index for monitoring prioritization without requiring modifications to the base predictor.

1. Introduction

Reservoir-bank landslides evolve under coupled geological and hydrometeorological forcing, while monitoring and emergency decisions must be made before the future deformation is observed. The Three Gorges Reservoir Area (TGRA) contains numerous reactivated slopes whose motion responds to seasonal rainfall and reservoir-level regulation [1, 2]. Modern remote-sensing and in-situ systems have expanded the temporal and spatial coverage of deformation observations [3], but forecast products used in practice are still commonly expressed as a single displacement value. A

* *Corresponding author.*
E-mail address: chenggang@ncist.edu.cn

point value can be accurate on average and yet be dangerously overconfident during a rapidly changing episode. For risk management, the operational question is therefore not only how much displacement is expected, but also what range of outcomes is plausible and whether that range is sufficiently informative for action.

Landslide displacement forecasting has progressed from empirical and decomposition-based models to machine-learning and deep-learning architectures. Studies in the TGRA have employed support vector regression and long short-term memory networks [4, 5], gated recurrent units and comparative machine-learning models [6, 7], and multi-site deep-learning architectures [8]. These models can capture nonlinear relations among antecedent displacement, rainfall, and reservoir water level, but comparisons across sites show that no algorithm dominates in every deformation setting. In particular, cumulative displacement can yield visually high coefficients of determination because the long-term trend dominates the variance; such scores may hide poor prediction of the short-term increment that matters for warning. Model ranking alone consequently offers an incomplete basis for decision making.

A smaller body of work has addressed interval forecasting. Neural lower-upper-bound estimation and bootstrap-based methods produce prediction intervals for step-like landslides [9, 10], while Monte Carlo analysis reveals sensitivity to training length and model randomness [11]. A recent conformalized graph-neural framework demonstrated the relevance of calibrated intervals for deformation monitoring [12]. However, highly parameterized interval models may be disproportionate when the objective is a concise, reproducible warning aid. More importantly, reporting only pooled coverage can conceal conditional failure in a particular slope region or deformation state.

Conformal prediction (CP) offers a model-agnostic route from point predictions to prediction sets. Under exchangeability, split CP provides finite-sample marginal coverage without assuming Gaussian residuals [13, 14]. Extensions can adapt interval width to covariate-dependent uncertainty [15] and improve robustness under distribution drift or other violations of exchangeability [16]. Landslide records, however, are serially dependent and seasonally nonstationary, so the classical guarantee cannot be transferred uncritically. In this study, CP coverage is therefore treated as an empirical property of a chronological test period, and conditional diagnostics are reported explicitly.

The decision framing is inspired by information granulation: a complex object is represented by groups formed through similarity, proximity, or function [17, 18]. Here, the 36 GNSS monitoring points are organized into four spatial granules to represent spatially heterogeneous deformation, and each forecast origin is assigned to a deformation-state granule (stable, active, or rapid) using only already-observed deformation. These granules do not create new physical boundaries; they organize heterogeneous evidence at a resolution that monitoring personnel can interpret.

The contributions are threefold. First, a reproducible seven-day forecasting workflow combines a lightweight LightGBM increment predictor with split conformal calibration. Second, deformation-scaled conformal prediction changes interval width using only recent displacement variability at the forecast origin, allowing uncertainty to expand during stronger motion without using future labels. Third, the study evaluates point accuracy and interval performance across spatial granules, and further examines conditional coverage, interval width, and point error across deformation-state granules, translating these diagnostics into a relative confidence index for monitoring prioritization. The approach is tested on 1,573 daily records from 36 GNSS monitoring points at the Xinpu landslide.

2. Study Area and Data

2.1 Xinpu Reservoir Landslide

The Xinpu landslide is located in Xinpu Village, Fengjie County, Chongqing City, on the right bank of the Yangtze River in the central TGRA. It is a large, complex, north-facing reservoir landslide extending from approximately 90 m above sea level at the toe to 705 m at the crown. The mapped area is about 1.94 km², the reported volume is approximately 5.4 × 10^7 m³, the average slope angle is 18–28°, and the main sliding orientation is close to 347° [19]. Its toe is perennially affected by the operational reservoir range of approximately 145–175 m. Field and remote-sensing studies show strong spatial heterogeneity: rainfall is a major driver, while reservoir drawdown exerts a particularly important influence near the submerged toe, with delayed responses that vary across the slope [20, 21].

The monitoring archive used here contains daily rainfall, reservoir water level, and cumulative surface displacement at 36 GNSS monitoring points from 10 Aug 2017 to 29 Nov 2021. Rainfall is expressed in millimetres per day, reservoir water level in metres, and displacement in millimetres. Considering the documented spatial heterogeneity of deformation across the Xinpu landslide [22], the 36 GNSS monitoring points were organized into four spatial granules (Zones I–IV) according to their spatial locations and regional deformation characteristics. The regional series aggregate the monitoring points within these spatial granules rather than redefining geological units. Zone I contains 13 GNSS monitoring points in the slope front, Zone II contains 14 GNSS monitoring points in the middle slope area, Zone III contains eight GNSS monitoring points in the transition/rear areas, and Zone IV contains the single rear reference monitoring point GPS21. No missing values remained after temporal alignment, so the daily number of contributing monitoring points was fixed within each zone. For each day, the arithmetic mean of the monitoring points in a zone forms its regional displacement. Zone IV is therefore a single-point series and is interpreted cautiously. The regional deformation and hydrological forcing are shown in Figure 1.

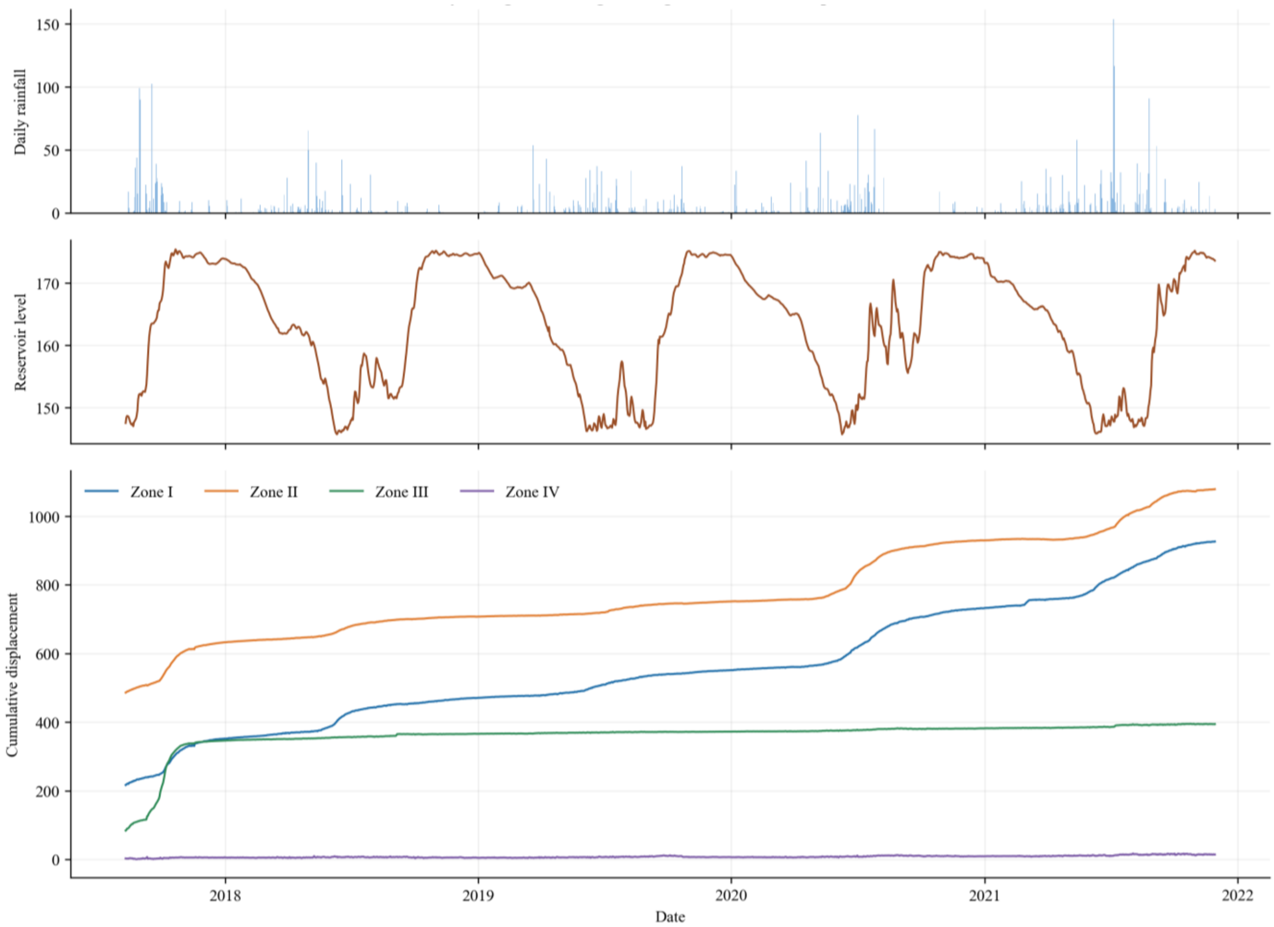

Fig. 1. Hydrological forcing and regional deformation at the Xinpu landslide from 10 Aug 2017 to 29 Nov 2021

2.2 Supervised Samples and Chronological Split

Each sample uses the preceding 30 days of regional displacement, daily displacement increment, rainfall, reservoir water level, and reservoir-level change. The LightGBM input is a 164-dimensional vector comprising 150 daily lag values (30 values for each of the five series) and 14 derived statistics: 1-, 3-, 7-, 14-, and 29-day displacement changes; the 30-day displacement trend; 3-, 7-, 14-, and 30-day rainfall totals; and 1-, 7-, 14-, and 29-day reservoir-level changes. The prediction target is the seven-day displacement increment, $z_t = D_{t+7} - D_t$. The cumulative displacement forecast is reconstructed by adding the predicted increment to D_t. This design prevents the cumulative trend from being the only source of apparent skill.

A strict chronological partition was applied independently to each zone: 60% training, 20% calibration, and 20% test. The 1,573 observations yield 1,537 supervised samples per zone: 922 training targets (15 Sep 2017-24 Mar 2020), 307 calibration targets (25 Mar 2020-25 Jan 2021), and 308 test targets (26 Jan-29 Nov 2021). No random shuffling was used. All point-model specifications and hyperparameters were fixed before calibration and test evaluation, and point-model fitting and internal validation used only the training period. Calibration and test labels were not used for point-model selection, hyperparameter tuning, or early stopping. The principal experimental settings are summarized in Table 1.

Table 1
Data partition and principal experimental settings

Component	Setting	Purpose / interpretation
Daily records	1,573; 2017-08-10 to 2021-11-29	Aligned displacement, rainfall, and reservoir water level series
Spatial granules	Zone I/II/III/IV: 13/14/8/1 GNSS monitoring points	Aggregation within four spatial granules
Input / horizon	30 days / 7 days	Past-only predictors and short-term decision horizon
Samples per zone	922 train / 307 calibration / 308 test	Chronological 60/20/20 split
Nominal coverage	90% (alpha = 0.10)	Target marginal prediction-interval coverage
DSCP scale	14-day RMS increment; training Q25 floor	Past-only local deformation difficulty
State thresholds	Training forecast-origin deformation velocity Q50 and Q85	Stable / active / rapid interpretation
Random seed	42 (+ zone index)	Reproducible model fitting

3. Methodology

3.1 Granular Uncertainty Assessment Framework

The granular uncertainty assessment framework in Figure 2 separates forecasting from uncertainty calibration. The spatial granule determines which regional series is modeled; the deformation-state granule summarizes the forecast-origin deformation velocity. A predictor produces a seven-day increment, the calibration set converts forecast errors into conformal scores, and the resulting interval is paired with a relative confidence index. This separation keeps the uncertainty layer auditable: changing the point model does not change the definition of calibration, and changing the relative confidence index does not alter the interval.

3.2 Point Forecasting Models

Persistence, LightGBM, and GRU are compared. Persistence sets the seven-day increment to zero and therefore provides a demanding baseline for slowly moving zones. LightGBM is used as the primary model because gradient-boosted decision trees efficiently represent nonlinear interactions among lagged and derived predictors [23]. The model contains 600 estimators, learning rate 0.025, 24 leaves, minimum child size 18, row subsampling 0.90, feature subsampling 0.85, L1 regularization 0.05, and L2 regularization 0.25; it minimizes an L1 regression objective. These settings were fixed before conformal calibration and held constant across zones. LightGBM was also fixed as the common base predictor for Split-CP and DSCP so that their differences isolate interval calibration rather than zone-specific predictor switching; this choice does not imply that LightGBM is the most accurate point predictor in every spatial granule. The model predicts z_t, and the cumulative displacement forecast is

$$\widehat{D}_{t+7} = D_t + \hat{z}_t \tag{1}$$

The GRU baseline receives a 30 × 5 sequence containing displacement level, displacement increment, rainfall, reservoir water level, and reservoir-level change. A single 32-unit GRU is followed by a 16-unit rectified linear layer and a scalar output. The final 15% of the chronological training subset (138 of 922 samples) is reserved as an internal validation tail for early stopping, while the preceding 784 training samples are used for weight optimization. Training uses AdamW, SmoothL1 loss, learning rate 0.002, batch size 128, a maximum of 120 epochs, gradient clipping at 1.0, and early stopping after 15 unimproved epochs. Neither the conformal calibration subset nor the test subset is used for GRU early stopping. The GRU is included as a representative recurrent model [24], not as the proposed uncertainty mechanism.

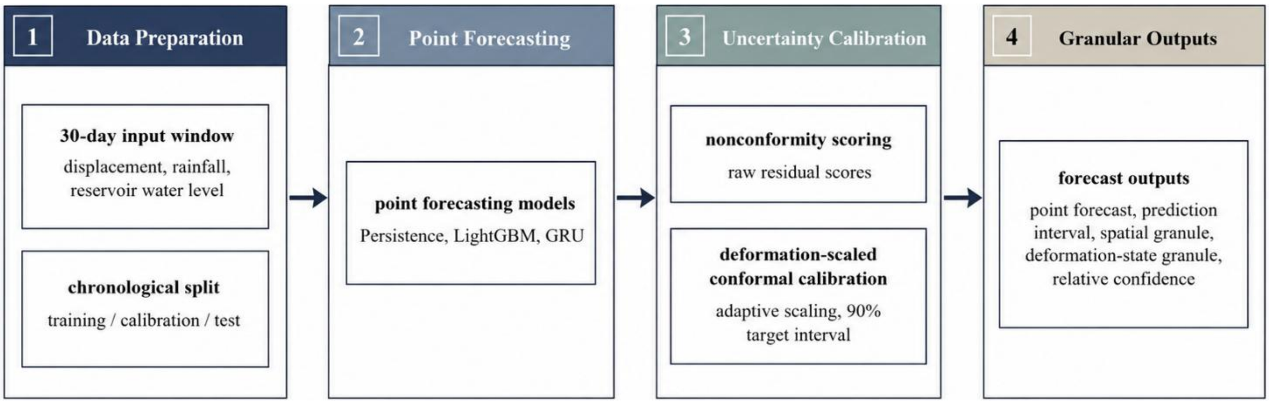

Fig. 2. Workflow of the proposed seven-day forecasting and granular uncertainty assessment framework

3.3 Split Conformal Prediction

Let $(y_i, \hat{y}_i)$ be calibration targets and LightGBM predictions, $i = 1, \dots, n_{\text{cal}}$. Split-CP uses the absolute residual as the nonconformity score, $a_i = |y_i - \hat{y}_i|$. For miscoverage level $\alpha = 0.10$, the finite-sample order-statistic rank is $k = \lceil (n_{\text{cal}} + 1)(1 - \alpha) \rceil$, clipped at n_{cal}. If q is the kth smallest calibration score, the interval for a test input x_t is

$$C_{\text{split}}(x_t) = \left[\widehat{D}_{t+7} - q, \widehat{D}_{t+7} + q\right] \tag{2}$$

The interval has constant width within a zone. Under exchangeability, this construction controls marginal miscoverage in finite samples. Because adjacent landslide samples overlap in their 30-day windows and seven-day horizons, exchangeability is not assumed here; performance is reported as empirical coverage on a later time block.

3.4 Deformation-Scaled Conformal Prediction

Constant-width intervals cannot express the engineering expectation that a rapidly changing slope is harder to forecast. DSCP therefore normalizes calibration residuals by a past-only local scale. For sample t, the raw scale is the root-mean-square value of the daily displacement increments in the preceding 14 days. The 14-day window, equal to two forecast horizons, was specified before test evaluation to balance recent responsiveness with short-range smoothing. A zone-specific floor λ is the 25th percentile of raw scales in the training set. This lower-quartile floor was specified as a robust training-period bound against near-zero widths during nearly static periods; neither the window nor the floor was selected from calibration or test coverage:

$$s_t = \max\left\{\lambda, \sqrt{\frac{1}{14}\sum_{j=0}^{13}\left(\Delta D_{t-j}\right)^2}\right\} \tag{3}$$

The scaled calibration score is $a_i^* = \frac{|y_i - \hat{y}_i|}{s_i}$. Let q^* be its conformal order statistic at the same 90% target. The DSCP interval is

$$C_{\mathrm{DSCP}}(x_t) = \left[\widehat{D}_{t+7} - q^* s_t, \widehat{D}_{t+7} + q^* s_t\right] \tag{4}$$

This normalization is a locally scaled conformal construction, but it should not be confused with online adaptive conformal inference: q^* is fixed after chronological calibration, while only the scale s_t changes using recent observations. The method remains a lightweight wrapper and requires no additional interval network.

3.5 Deformation-State Granules, Confidence, and Evaluation

The forecast-origin deformation velocity is $v_t = \frac{|D_t - D_{t-7}|}{7}$. Zone-specific thresholds are the 50th and 85th percentiles of v_t in the training set. These thresholds were specified before test evaluation: the median separates the lower half of training-period velocities, while the 85th percentile identifies the upper-velocity tail without producing an excessively small rapid-state subset. Forecast origins below the first threshold are labeled stable, those between thresholds active, and those above the second rapid. Because the thresholds are zone-specific, stable, active, and rapid denote relative deformation states within each spatial granule rather than common absolute velocity classes across zones. Deformation-state labels are used only for conditional evaluation and were not optimized against test coverage. The DSCP interval width is U_t. For concise communication, the relative confidence index is

$$C_t = \max\left(0, 1 - \frac{U_t}{U_{\mathrm{ref}}}\right) \tag{5}$$

where U_{ref} is the maximum DSCP width pooled across all zones during the calibration period and is fixed before test evaluation. The maximum operator truncates negative values at zero; because interval width is nonnegative, the resulting index is bounded within $[0,1]$. The index is a relative normalization, not a probability of correctness or a substitute for empirical coverage. Point forecasts are evaluated with mean absolute error (MAE), root-mean-square error (RMSE), and R^2 on both cumulative displacement and seven-day increments. Intervals are evaluated with prediction interval coverage probability (PICP), mean prediction interval width (MPIW), and the proper interval score of Gneiting and Raftery [25]. PICP alone rewards unnecessarily wide intervals, whereas the interval score penalizes both width and misses. Overall metrics were calculated by pooling all zone-time samples in the chronological test period. Gain-based LightGBM importance is normalized to sum to 100% within each zone and then aggregated into displacement, reservoir, and rainfall predictor groups. PICP and width terminology follows established prediction-interval evaluation practice [26].

Paired point-forecast comparisons use a moving-block bootstrap of the daily absolute-error difference, defined as persistence error minus LightGBM error. The procedure uses a 14-day block length, 2,000 replicates, and percentile 95% confidence intervals, and is applied to all four zones. The block length exceeds the seven-day forecast horizon and preserves short-range serial dependence in overlapping forecast samples.

4. Results

4.1 Point Forecast Accuracy Differs by Spatial Granule

Table 2 shows that LightGBM is clearly superior in the two strongly deforming zones. Its MAE is 1.754 mm in Zone I and 1.466 mm in Zone II, reductions of 59.6% and 58.0% relative to persistence, respectively. The 14-day moving-block bootstrap gives 95% confidence intervals of 1.681-3.698 mm for Zone I and 1.138-3.134 mm for Zone II when improvement is defined as persistence error minus LightGBM error. Both intervals remain positive under serial dependence at this block length. GRU does not improve on persistence in Zone I and offers only a modest gain in Zone II.

The ranking reverses in the low-motion zones. Persistence has the lowest MAE in Zone III (0.559 mm) and Zone IV (0.861 mm). LightGBM is worse by 0.234 and 0.098 mm, respectively. The corresponding bootstrap interval is -0.708 to 0.032 mm for Zone III and -0.150 to -0.036 mm for Zone IV; the Zone III difference is inconclusive, whereas the Zone IV interval remains negative. Reporting all four zones avoids selecting only favorable comparisons. These results demonstrate why a single pooled ranking is insufficient. Overall LightGBM MAE is 1.243 mm, 46.2% below persistence, but the gain is concentrated in Zones I and II.

Cumulative-displacement R^2 exceeds 0.99 for most models in the strongly deforming zones because the 2021 displacement range is large. Increment R^2 is more discriminating: LightGBM reaches 0.390 in Zone I, 0.621 in Zone II, -2.605 in Zone III, and -0.121 in Zone IV, with an overall value of 0.601. The apparently excellent cumulative R^2 in Zone III therefore does not indicate genuine seven-day increment skill. Overall cumulative-displacement R^2 is omitted from Table 2 because pooling zones with different displacement levels makes that metric unsuitable for model comparison; the pooled increment R^2 is retained.

Table 2
Seven-day point-forecast performance during the chronological test period

Zone	Model	MAE (mm)	RMSE (mm)	Cumulative R^2	Increment R^2
I	Persistence	4.336	5.459	0.9932	-1.663
I	LightGBM	1.754	2.614	0.9984	0.390
I	GRU	4.321	5.524	0.9931	-1.728
II	Persistence	3.491	5.083	0.9923	-0.746
II	LightGBM	1.466	2.369	0.9983	0.621
II	GRU	2.994	4.416	0.9942	-0.317
III	Persistence	0.559	0.911	0.9614	-0.096
III	LightGBM	0.793	1.653	0.8730	-2.605
III	GRU	0.640	0.975	0.9558	-0.254
IV	Persistence	0.861	1.143	0.7679	-0.010
IV	LightGBM	0.959	1.204	0.7424	-0.121
IV	GRU	0.881	1.133	0.7719	0.008
Overall	Persistence	2.312	3.800	—	-0.385
Overall	LightGBM	1.243	2.039	—	0.601
Overall	GRU	2.209	3.614	—	-0.253

Note: — indicates that pooled cumulative-displacement R^2 is not reported because between-zone displacement-level differences dominate the pooled variance.

4.2 Marginal Interval Coverage and Sharpness

Split-CP attains near-nominal coverage in Zone I (89.61%) and conservative coverage in Zone II (93.51%), but undercovers Zone III (83.44%) and Zone IV (78.25%). Pooling all zones gives 86.20% PICP. DSCP brings every zone-level PICP to within 1.04 percentage points of the 90% target: 88.96%, 90.58%, 89.61%, and 89.61% for Zones I-IV. Overall PICP is 89.69%.

Coverage improvement is not obtained by indiscriminate widening. DSCP narrows the average interval in Zones I and II by 5.5% and 4.3%, while widening Zones III and IV to correct severe Split-CP undercoverage. Overall MPIW increases from 5.396 to 5.641 mm (4.5%), but the mean interval score decreases from 10.019 to 9.298 (7.2%), indicating a better joint balance of coverage and sharpness. Figure 3 illustrates the temporal evolution of the DSCP intervals, which expand around stronger deformation episodes and contract during quieter periods. The comparative PICP, MPIW, and interval score of Split-CP and DSCP across the spatial granules and the pooled test set are shown in Figure 4.

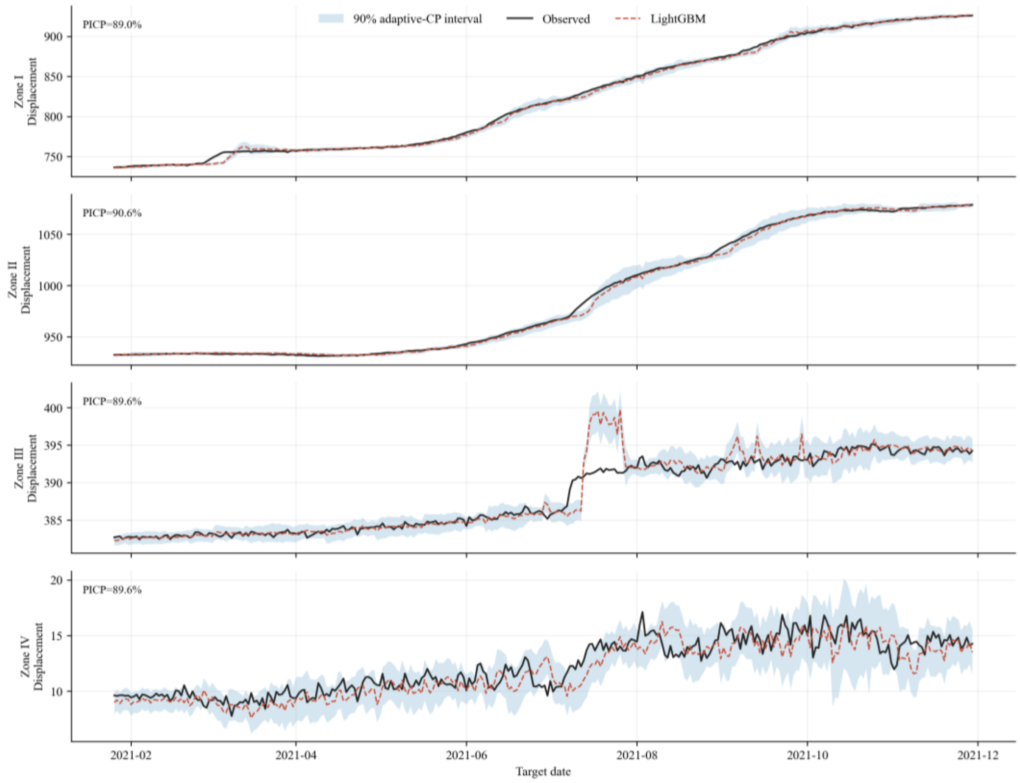

Fig. 3. Observed and predicted cumulative displacement with nominal 90% DSCP intervals during the test period

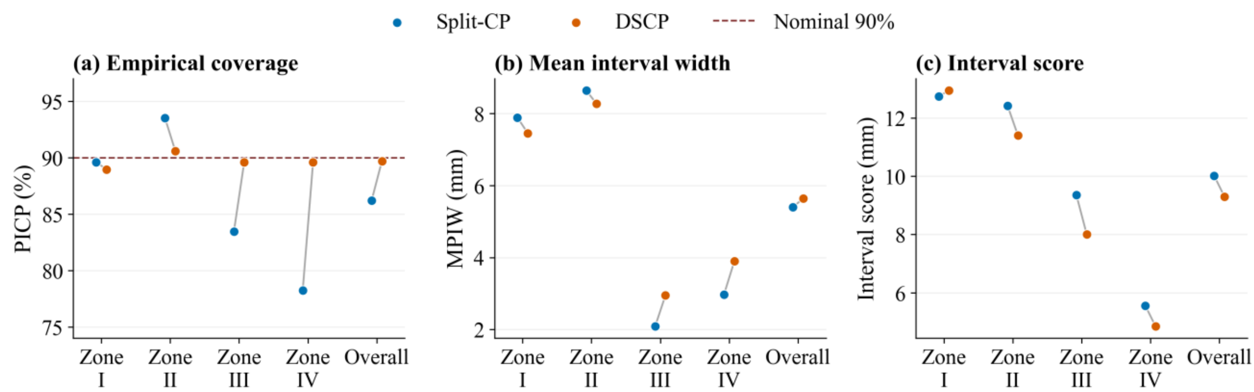

Fig. 4. Comparison of Split-CP and DSCP across spatial granules: (a) PICP, (b) MPIW, and (c) interval score

4.3 State-Conditional Uncertainty Reveals Hidden Failure

The expected relation between motion and uncertainty is strongest in Zones I and II, as shown in Figure 5(b–d). From stable to rapid states, mean DSCP width increases from 3.948 to 9.982 mm in Zone I and from 3.143 to 13.773 mm in Zone II. LightGBM MAE simultaneously rises from 1.014 to 2.319 mm and from 0.818 to 2.257 mm, respectively. At the pooled sample level, interval width has

Pearson and Spearman correlations of 0.887 and 0.768 with the forecast-origin deformation velocity; the corresponding correlations with realized future deformation velocity are 0.713 and 0.703. Because DSCP width and forecast-origin deformation velocity are both constructed from recent displacement increments, the width–velocity association is treated as a consistency check rather than independent validation. The pooled Spearman correlation between interval width and absolute LightGBM error is 0.473, with zone-specific values ranging from 0.295 to 0.409, indicating a moderate association with realized forecast difficulty.

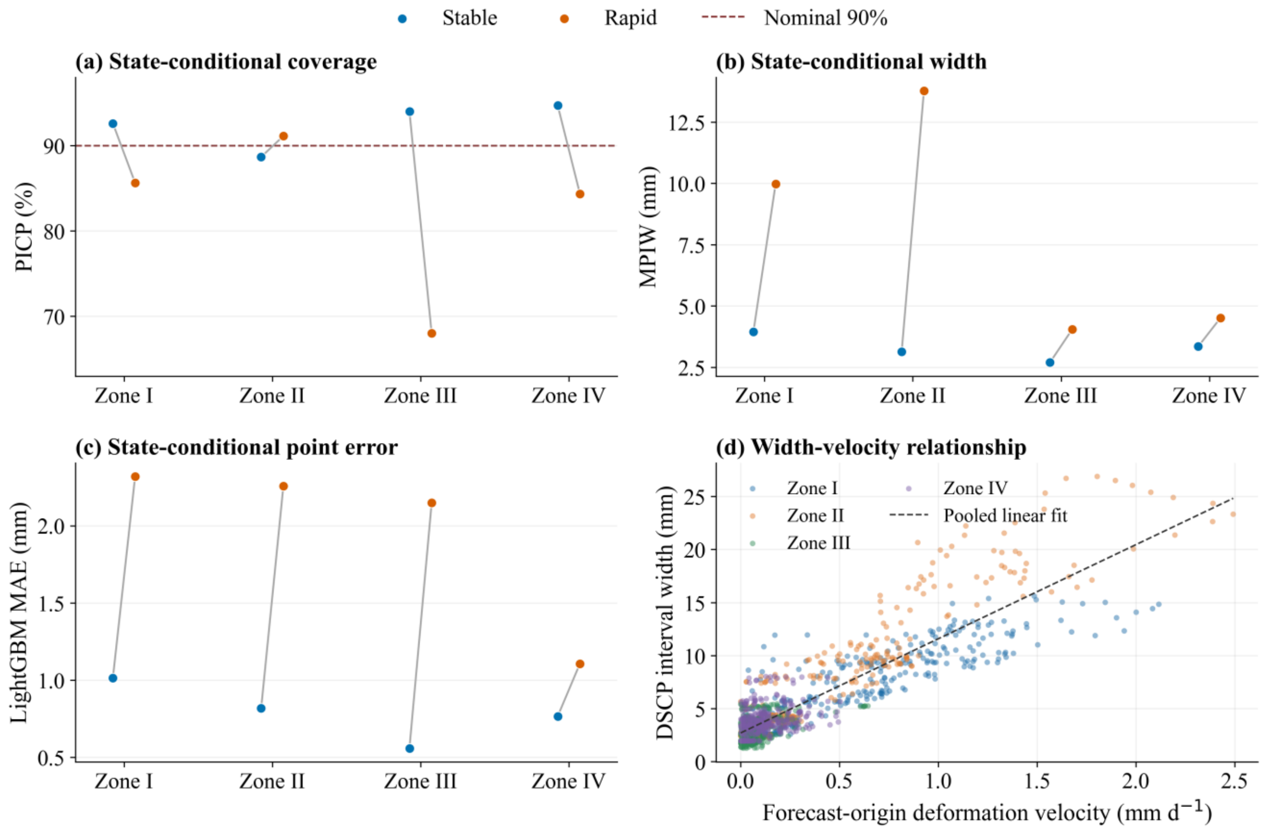

Fig. 5. State-conditional forecasting and uncertainty performance across spatial granules: (a) PICP, (b) MPIW, (c) LightGBM MAE, and (d) DSCP interval width versus deformation velocity

Marginal calibration nevertheless hides a critical conditional weakness, as illustrated in Figure 5(a–c). Active-state PICP is 92.54%, 92.11%, 93.60%, and 88.39% for Zones I–IV, respectively. The corresponding active-state MPIW/MAE pairs are 5.642/1.300, 5.024/0.886, 2.783/0.499, and 3.983/1.045 mm for Zones I–IV, respectively. Zone III has an overall PICP of 89.61%, but its rapid-state PICP decreases to only 68.0%, compared with 93.98% in the stable state. The principal cluster of misses occurs in July 2021, when the Zone III MAE rises to 3.755 mm and monthly PICP falls to 32.26%. Although the interval expands, the abrupt level change exceeds what a 14-day retrospective scale can anticipate. Zone IV also shows a decline in coverage, from 94.69% in the stable state to 84.34% in the rapid state. These results show that a nominally calibrated regional interval cannot be assumed to remain conditionally valid across all deformation states. For visual clarity, Figure 5 contrasts the stable and rapid deformation states, while active-state results are reported in the text.

4.4 Predictor Contribution and Confidence Communication

Gain-based LightGBM importance was normalized within each zone and aggregated according to the predictor groups defined in Section 3.5. Averaged across the four zone-specific models, displacement predictors account for 52.6% of total gain, reservoir water level and reservoir-level change for 36.8%, and rainfall lags and accumulated rainfall for 10.5%. The corresponding zone ranges are 49.5-56.3%, 33.3-39.2%, and 10.0-11.3%. Because lagged predictors within each group are strongly correlated, these importances indicate predictive allocation rather than causal effects. They are consistent with the observed joint role of reservoir regulation, antecedent motion, and seasonal rainfall at Xinpu [27].

Using the calibration-period reference width of 37.534 mm, the mean relative confidence index is 0.801, 0.780, 0.921, and 0.896 for Zones I-IV, respectively. These values indicate that the absolute intervals are generally broader in the strongly deforming zones, but they do not guarantee local coverage. In particular, most Zone III samples retain high relative confidence because their widths are small compared with Zone II, even though the Zone III rapid-state interval undercovers. Operational use should therefore display the relative confidence index together with the spatial granule, deformation-state granule, and recent conditional PICP; the index must never be presented as a probability that the forecast is correct.

5. Discussion

5.1 Why Granularity Matters for Intelligent Decisions

The principal finding is not that one learner is universally best, but that decision quality depends on the granularity at which performance is inspected. At the site-wide level, LightGBM appears decisively superior and DSCP achieves near-nominal marginal coverage at the pooled level. Evaluation across spatial granules shows that persistence is more competitive in slowly moving rear zones, while evaluation across deformation-state granules exposes severe rapid-state undercoverage in Zone III. An intelligent warning interface should retain these distinctions instead of collapsing them into one accuracy score.

The resulting product supports a simple three-part communication protocol. First, report the seven-day point forecast as the expected operational estimate. Second, report the DSCP interval as the empirically calibrated range. Third, attach the current deformation-state granule and relative confidence index as an attention cue. A stable, well-covered state can remain under routine observation; an active state or widening interval warrants enhanced review; and a rapid state, low recent conditional coverage, or abrupt change should trigger expert inspection and possibly higher monitoring frequency. These are decision semantics rather than validated alarm thresholds. Deployment thresholds must be coupled to local displacement-rate criteria, asset exposure, monitoring noise, and agency procedures.

DSCP improves interval utility without a new neural architecture. The overall interval score decreases even though average width increases slightly, because width is reallocated: intervals in Zones I and II narrow on average, whereas low-motion zones receive sufficient width to recover marginal coverage. This is preferable to multiplying all intervals by one conservative factor. The association between interval width and forecast-origin deformation velocity is an expected consistency check because both quantities use recent displacement increments; the moderate association with absolute forecast error provides a separate empirical diagnostic. However, the July Zone III event demonstrates the limit of retrospective scaling when a structural shift has little precursor in the chosen window.

5.2 Methodological Boundaries

Four limitations define the scope of the conclusions. First, classical split-CP validity depends on exchangeability, which overlapping landslide time-series samples do not satisfy. The reported 89.69% coverage is an empirical result for the chronological test period, not a distribution-free guarantee for arbitrary future sequences. Block conformal calibration, online adaptation, or rolling weighted calibration should be tested when longer operational records are available.

Second, the study evaluates one landslide over a single chronological test period in 2021. The 2021 period contains substantial activity and is therefore informative, but transferability to different lithology, instrumentation, climate, and reservoir operations remains untested. Third, Zone IV contains only one GNSS monitoring point, so its zone label does not imply spatial averaging. A future multi-site study should preserve point-level uncertainty before aggregation within spatial granules. Fourth, the relative confidence index uses one reference width pooled across zones during calibration. Although this removes test-period look-ahead, a common reference can still overstate relative confidence in naturally low-amplitude zones. A calibrated conditional reliability estimate or rolling coverage dashboard would be a safer operational complement.

Additional robustness work should evaluate alternative lookback and forecast horizons, repeated or ensemble training of stochastic baselines, missing-data scenarios, and ablations of the deformation scale. The present feature-importance analysis is predictive and must not be interpreted as hydrological causality. Physical variables such as pore pressure and groundwater level, when available, could improve anticipation of abrupt changes that surface displacement alone does not reveal.

6. Conclusions

This study formulated short-term landslide forecasting as an uncertainty-aware decision problem rather than a point-accuracy contest. For the Xinpu landslide, a 30-day LightGBM model reduced overall seven-day MAE by 46.2% relative to persistence, mainly through gains in Zones I and II. Deformation-scaled conformal prediction achieved 89.69% empirical coverage for a nominal 90% interval and improved the interval score by 7.2% relative to constant-width Split-CP, with only a 4.5% increase in average width. The interval-width association with forecast-origin deformation velocity is consistent with the construction of the deformation scale, while its moderate association with absolute forecast error supports continued error-based validation.

The aggregate result is not sufficient for safety communication. Zone III rapid-state coverage fell to 68.0% despite near-nominal regional coverage. This conditional failure is itself an actionable result: uncertainty products should be audited across spatial granules and deformation-state granules, and rapid-state misses should escalate monitoring rather than be hidden by pooled statistics. The proposed point forecast, prediction interval, and relative confidence index are computationally light and suitable for a compact monitoring workflow, provided that the relative confidence index is interpreted as a relative measure and empirical coverage is continuously rechecked.

This research was funded by the National Natural Science Foundation of China (42377200), the Natural Science Foundation of Hebei Province, China (D2025508013) and the State Key Laboratory for Tunnel Engineering (TESKL202435).

Acknowledgement

This research was not funded by any grant.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Yin, Y., Huang, B., Wang, W., Wei, Y., Ma, X., Ma, F., & Zhao, C. (2016). Reservoir-induced landslides and risk control in Three Gorges Project on Yangtze River, China. Journal of Rock Mechanics and Geotechnical Engineering, 8, 577-595. https://doi.org/10.1016/j.jrmge.2016.08.001

[2] Tang, H., Wasowski, J., & Juang, C. H. (2019). Geohazards in the Three Gorges Reservoir Area, China: Lessons learned from decades of research. Engineering Geology, 261, 105267. https://doi.org/10.1016/j.enggeo.2019.105267
[3] Casagli, N., Intrieri, E., Tofani, V., Gigli, G., & Raspini, F. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. Nature Reviews Earth & Environment, 4, 51-64. https://doi.org/10.1038/s43017-022-00373-x
[4] Jiang, H., Li, Y., Zhou, C., Hong, H., Glade, T., & Yin, K. (2020). Landslide displacement prediction combining LSTM and SVR algorithms: A case study of Shengjibao landslide from the Three Gorges Reservoir Area. Applied Sciences, 10, 7830. https://doi.org/10.3390/app10217830
[5] Yang, B., Yin, K., Lacasse, S., & Liu, Z. (2019). Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides, 16, 677-694. https://doi.org/10.1007/s10346-018-01127-x
[6] Zhang, W., Li, H., Tang, L., Gu, X., Wang, L., & Wang, L. (2022). Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks. Acta Geotechnica, 17, 1367-1382. https://doi.org/10.1007/s11440-022-01495-8
[7] Wang, Y., Tang, H., Huang, J., Wen, T., Ma, J., & Zhang, J. (2022). A comparative study of different machine learning methods for reservoir landslide displacement prediction. Engineering Geology, 298, 106544. https://doi.org/10.1016/j.enggeo.2022.106544
[8] Nava, L., Carraro, E., Reyes-Carmona, C., Puliero, S., Bhuyan, K., Rosi, A., Monserrat, O., Floris, M., Meena, S. R., Galve, J. P., & Catani, F. (2023). Landslide displacement forecasting using deep learning and monitoring data across selected sites. Landslides, 20, 2111-2129. https://doi.org/10.1007/s10346-023-02104-9
[9] Lian, C., Zeng, Z., Wang, X., Yao, W., Su, Y., & Tang, H. (2020). Landslide displacement interval prediction using lower upper bound estimation method with pre-trained random vector functional link network initialization. Neural Networks, 130, 286-296. https://doi.org/10.1016/j.neunet.2020.07.020
[10] Zheng, Z., Li, Y., Zhang, Y., Wen, L., Wang, T., & Sun, X. (2025). Failure probability assessment of step-like landslide using a hybrid interval prediction method under uncertain conditions. Journal of Rock Mechanics and Geotechnical Engineering, 17, 7265-7287. https://doi.org/10.1016/j.jrmge.2025.08.008
[11] Wang, L., Xiao, T., Liu, S., Zhang, W., Yang, B., & Chen, L. (2023). Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation. Gondwana Research, 123, 27-40. https://doi.org/10.1016/j.gr.2023.03.006
[12] Ren, X., Liu, W., Yang, W., Mao, L., & Li, H. (2025). Landslide deformation uncertainty quantification using conformalized graph neural networks: A case study in Sichuan Province, China. IEEE Access, 13, 86719-86731. https://doi.org/10.1109/ACCESS.2025.3568273
[13] Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Springer. https://doi.org/10.1007/b106715
[14] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113, 1094-1111. https://doi.org/10.1080/01621459.2017.1307116
[15] Lei, J., & Wasserman, L. (2014). Distribution-free prediction bands for non-parametric regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76, 71-96. https://doi.org/10.1111/rssb.12021
[16] Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. The Annals of Statistics, 51, 816-845. https://doi.org/10.1214/23-AOS2276
[17] Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems, 90, 111-127. https://doi.org/10.1016/S0165-0114(97)00077-8
[18] Pedrycz, W. (2013). Granular computing: Analysis and design of intelligent systems. CRC Press. https://doi.org/10.1201/9781315216737
[19] Ye, X., Zhu, H.-H., Cheng, G., Pei, H.-F., Shi, B., Schenato, L., & Pasuto, A. (2024). Thermo-hydro-poro-mechanical responses of a reservoir-induced landslide tracked by high-resolution fiber optic sensing nerves. Journal of Rock Mechanics and Geotechnical Engineering, 16, 1018-1032. https://doi.org/10.1016/j.jrmge.2023.04.004
[20] Zheng, W., Sun, Q., Hu, J., Lu, Z., Zhu, K., Ye, X., Huang, G., Hu, M., Zhu, J., & Li, Z. (2023). Investigating kinematics and triggers of slow-moving reservoir landslide using an improved MT-InSAR method. Geomatics, Natural Hazards and Risk, 14, 2289835. https://doi.org/10.1080/19475705.2023.2289835
[21] Zhu, W., He, Q., Wu, J., Wang, Y., Li, Z., Zhan, J., & Zhang, B. (2025). Identifying reservoir landslide deformation evolution stages with time series InSAR: Application to the Xinpu landslide in China's Three Gorges region. Journal of Hydrology, 663, 134136. https://doi.org/10.1016/j.jhydrol.2025.134136
[22] Zhou, C., Ye, M., Xia, Z., Wang, W., Luo, C., & Muller, J.-P. (2025). An interpretable attention-based deep learning method for landslide prediction based on multi-temporal InSAR time series: A case study of Xinpu landslide in the TGRA. Remote Sensing of Environment, 318, 114580. https://doi.org/10.1016/j.rse.2024.114580

[23] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29, 1189-1232. https://doi.org/10.1214/aos/1013203451

[24] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1724–1734. https://doi.org/10.3115/v1/D14-1179

[25] Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102, 359-378. https://doi.org/10.1198/016214506000001437

[26] Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. IEEE Transactions on Neural Networks, 22, 1341-1356. https://doi.org/10.1109/TNN.2011.2162110

[27] Chang, F., Dong, S., Yin, H., Ye, X., Wu, Z., Zhang, W., & Zhu, H. (2025). 3D displacement time series prediction of a north-facing reservoir landslide powered by InSAR and machine learning. Journal of Rock Mechanics and Geotechnical Engineering, 17, 4445-4461. https://doi.org/10.1016/j.jrmge.2024.10.033