Dear Reviewer,

We sincerely thank you for these detailed and constructive comments. In response, we have carried out a set of new analyses, which are reported in this letter, including landslide inventory declaration, a multi-resolution scale-sensitivity assessment, robustness and reproducibility using stability-based edge selection over repeated spatial partitions; and a SHAP-based pathway attribution scheme that replaces the earlier heuristic fusion to generate the contribution of the identified pathway. We also agree that parts of the reasoning, logical structure and wording require reorganization, and we specify below how each will be revised. We note that, as the journal does not permit submission of a revised manuscript at the current stage, we will systematically revise the manuscript in the next stage by incorporating your suggestions and will provide further clarification and additional information where appropriate.

Q: After carefully considering the authors' response and the current manuscript, I do not think the major concerns have been adequately addressed. The response mainly provides explanations or promises for future revision, but the fundamental weaknesses remain. The landslide inventory is still limited for supporting a highly parameterized DAG-GNN framework over a large and geologically complex basin, and the completeness, temporal consistency, mapping uncertainty, and independent validation of the inventory remain insufficiently demonstrated. The scale mismatch among the input factors is also a serious concern: several key climatic, geological, and seismic variables are derived from coarse-resolution or secondary datasets and then resampled to 30 m, but the effects of spatial uncertainty and scale dependency have not been rigorously assessed. The proposed "dependency pathways" remain largely statistical and hypothesis-generating rather than physically validated mechanisms. The final DAG still depends substantially on expert-guided correction, which weakens the claimed objectivity and reproducibility. In addition, the novelty is overstated, as graph-based and explainable AI methods have already been widely explored in landslide susceptibility studies. The discussion remains descriptive, and the manuscript still contains many problems in logic, expression, figure readability, and overinterpretation. Overall, the manuscript does not provide sufficient data reliability, methodological advance, or mechanism validation to meet the publication standard. Therefore, I do not recommend publication.

Reply: We thank the reviewer for this comment. Our clarifications and the corresponding revisions are given below.

Q1: Quality and representativeness of the landslide inventory. Relative to the size of the watershed and the complexity of its geological setting, the number of landslide samples remains limited. In addition, the completeness and temporal consistency of the inventory, mapping uncertainties, and independent validation have not been adequately documented.

(1) Quality, representativeness, completeness and temporal consistency of the landslide inventory

We have re-examined the inventory, which contains 1,198 landslide records. The attribute fields — landslide type, size class, recorded date of occurrence and triggering factor — are complete for all records; Table 1 lists the attribute fields of the inventory to document this completeness. Figure 1 presents the statistical characteristics of the inventory. In terms of size, the inventory is dominated by small- and medium-sized landslides. In terms of timing, landslide occurrence is concentrated in two periods: spring (March–May), associated with snowmelt, and July–October, associated with the rainy season; the recorded events span multiple years. The triggering factors are dominated by rainfall, earthquakes and snowmelt, together with their combined effects. These results indicate that the inventory records not only landslide locations but also key information on size, timing and triggering factors, and is therefore suitable for the subsequent susceptibility modelling and interpretative analysis.

More importantly, the inventory has undergone independent expert review and has been formally released on the official data-sharing platform of the Third Xinjiang Scientific Expedition Program of the Ministry of Science and Technology of China (https://cstr.cn/33110.11.XIEG.XJSEDATA.2024.00100302). The dataset is therefore representative of, and appropriate for, this region, and reflects the distribution characteristics of landslide hazards within this hazard-prone area.

The ratio of positive to negative samples is 1:1. Negative samples were selected using a 1 km buffer-zone approach in order to improve their plausibility and to reduce the risk of mislabelling unknown or unmapped landslides as negative samples (Guzzetti et al., 2012; Reichenbach et al., 2018).

In addition, the triggering-factor records obtained from the landslide inventory were used for external consistency checking of the directed spatial dependency pathways inferred by the DAG, i.e. to test whether different trigger types exhibit pathway characteristics consistent with their respective development processes.

For the dynamic conditioning factors (e.g. NDWI, NDVI, precipitation and snowmelt), values were extracted for the period corresponding to landslide occurrence, so that temporal consistency between the factors and the landslide records is ensured.

Table 1. Selected attributes of the landslide inventory.

ID	Elevation difference (m)	Aspect (°)	Landslide length (m)	Landslide width (m)	Area (m^2)	Occurrence time	Longitude	Latitude	Scale	Triggering factor
1	35	101.17	100	30	3402.49	2007.9	83° 9' 59.136" E	43° 53' 41.746" N	Small	Rainfall and earthquake

ID	Elevation difference (m)	Aspect (°)	Landslide length (m)	Landslide width (m)	Area (m²)	Occurrence time	Longitude	Latitude	Scale	Triggering factor
2	20	186.37	80	50	3265.21	2007.9	83° 11' 44.127" E	43° 53' 56.440" N	Small	Rainfall and earthquake
3	160	236.19	770	360	260501	2007.9	81° 24' 8.584" E	43° 14' 47.322" N	Medium	Rainfall and earthquake
4	165	229.45	950	210	251962	2007.9	81° 24' 11.973" E	43° 14' 36.964" N	Medium	Rainfall and earthquake
5	85	261.57	290	90	26190.7	2007.9	81° 24' 14.360" E	43° 15' 4.118" N	Small	Rainfall and earthquake
…	…	…	…	…	…	…	…	…	…	…
1198	140	30.02	376	518	127826	2005.3	83° 16′1.200″E	43° 18′42.707″N	Medium	Rainfall and engineering slope cutting

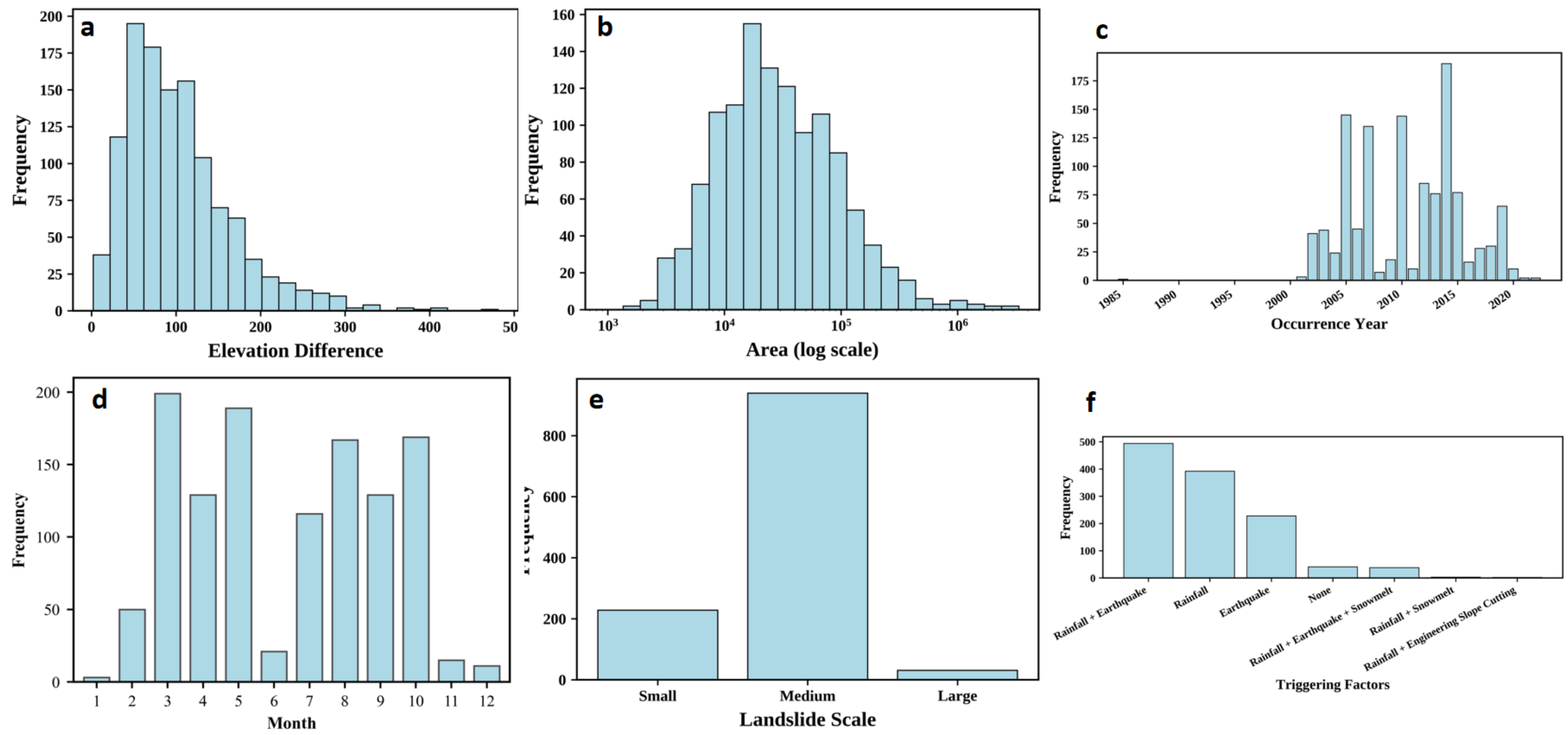

Figure 1. Statistical characteristics of the landslide inventory: (a) elevation difference, (b) area, (c) occurrence year, (d) occurrence month, (e) landslide scale, and (f) triggering factors.

(2) Mapping uncertainty and independent validation

The landslide inventory was compiled primarily through manual visual interpretation of geohazards using sub-metre high-resolution remote sensing imagery, including Google Earth imagery and GF-1/GF-2 imagery. Landslides were identified from their characteristic geomorphic features, including the head scarp, lateral flanks, landslide body and depositional zone. Landslides that were too small for their boundaries to be delineated reliably were not included in the inventory.

We acknowledge that manual visual interpretation is inherently subjective to some degree and therefore introduces a certain level of identification uncertainty. To improve the reliability of the inventory, field investigations, on-site interviews and cross-checks with local civil protection authorities were carried out for a subset of the mapped landslides (438 points were validated; Figure 2).

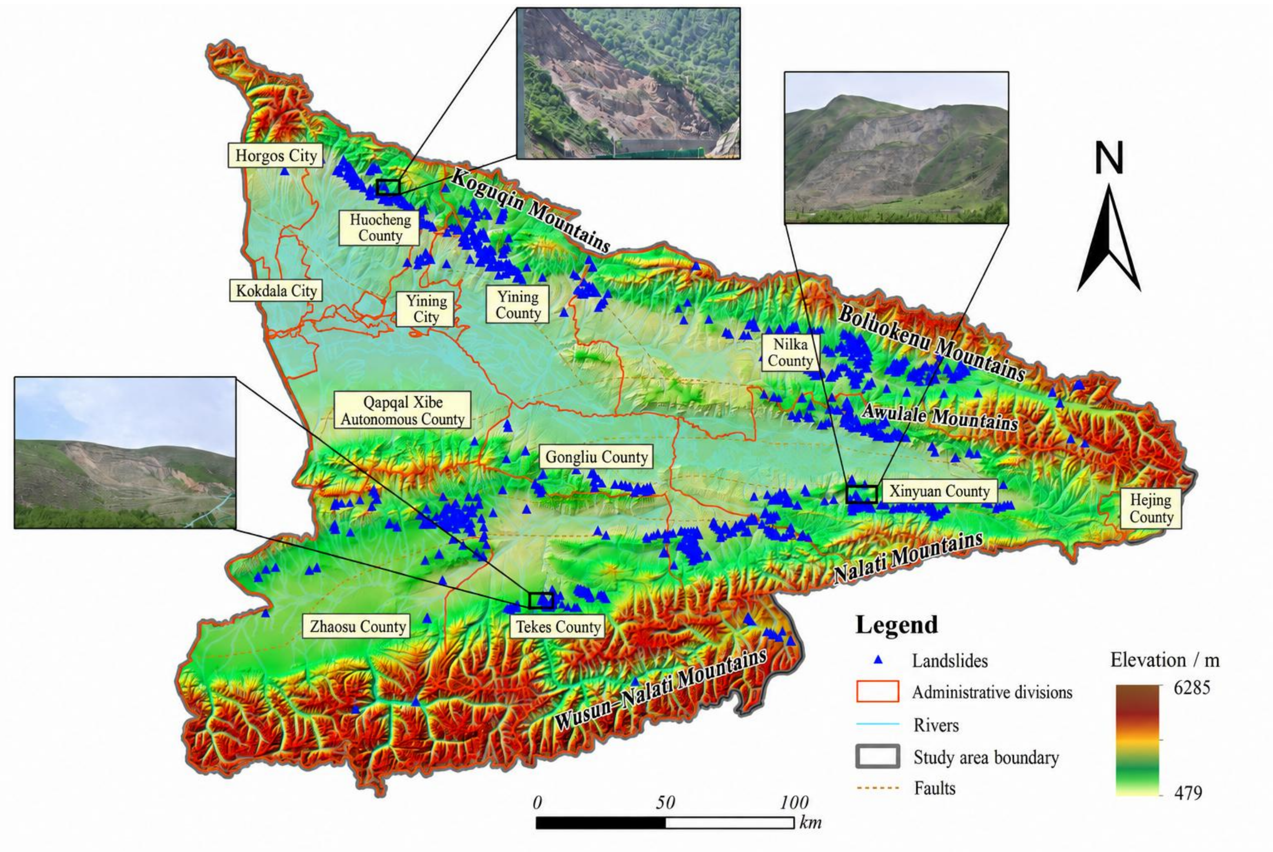

Figure 2. Independent field validation of the landslide inventory.

(3) Landslide inventory numbers issue

We would note that, for a large basin with complex geological conditions, the spatial representativeness of the landslide inventory is in fact more important than the sample number alone. The uncertainty in the assessment introduced by data sparsity was also taken into account in this study, and we therefore carried out data-ablation (sub-sampling) experiments, which confirm the stability of the results. Following the reviewer's suggestion, we also checked the relevant literature: in the original DAG-GNN study, the authors learned benchmark networks with 20, 37 and 441 nodes from 1,000 samples (Yu et al., 2019), which indicates that the graph structure used here does not constitute a high-dimensional parameterization problem. We therefore consider that the available data are sufficient to support an exploration of the dependency structure at the regional scale.

Q2: The scale mismatch among the input factors is also a serious concern: several key climatic, geological, and seismic variables are derived from coarse-resolution or secondary datasets and then resampled to 30 m, but the effects of spatial uncertainty and scale dependency have not been rigorously assessed.

Reply: We thank the reviewer for raising this point. Following this suggestion, we have analyzed the spatial uncertainty and the scale-dependency effects introduced by the resolution mismatch. Among the 13 conditioning factors used in this study, only precipitation, temperature and snow depth have a native resolution of 1 km, whereas the remaining factors have a native resolution of 30 m. We therefore designed the following experiment: the factors with a native resolution of 30 m were kept unchanged, and only the remaining factors were resampled to four resolutions by bilinear models (30 m, 100 m, 250 m and 500 m);

the differences in the mean values within the landslide areas were then computed across these resolutions. If the values obtained at the four resolutions differ substantially, this indicates a strong scale dependency; conversely, if the scale dependency is weak, unifying all factors to a 30 m resolution is feasible.

Figure 3 presents the results of the scale-difference sensitivity analysis. Overall, the difference curves for precipitation, temperature and snow depth at the four resolutions (30 m, 100 m, 250 m and 500 m) show small mean relative errors, with values of approximately 0.20%, 0.74% and 1.32%, respectively.

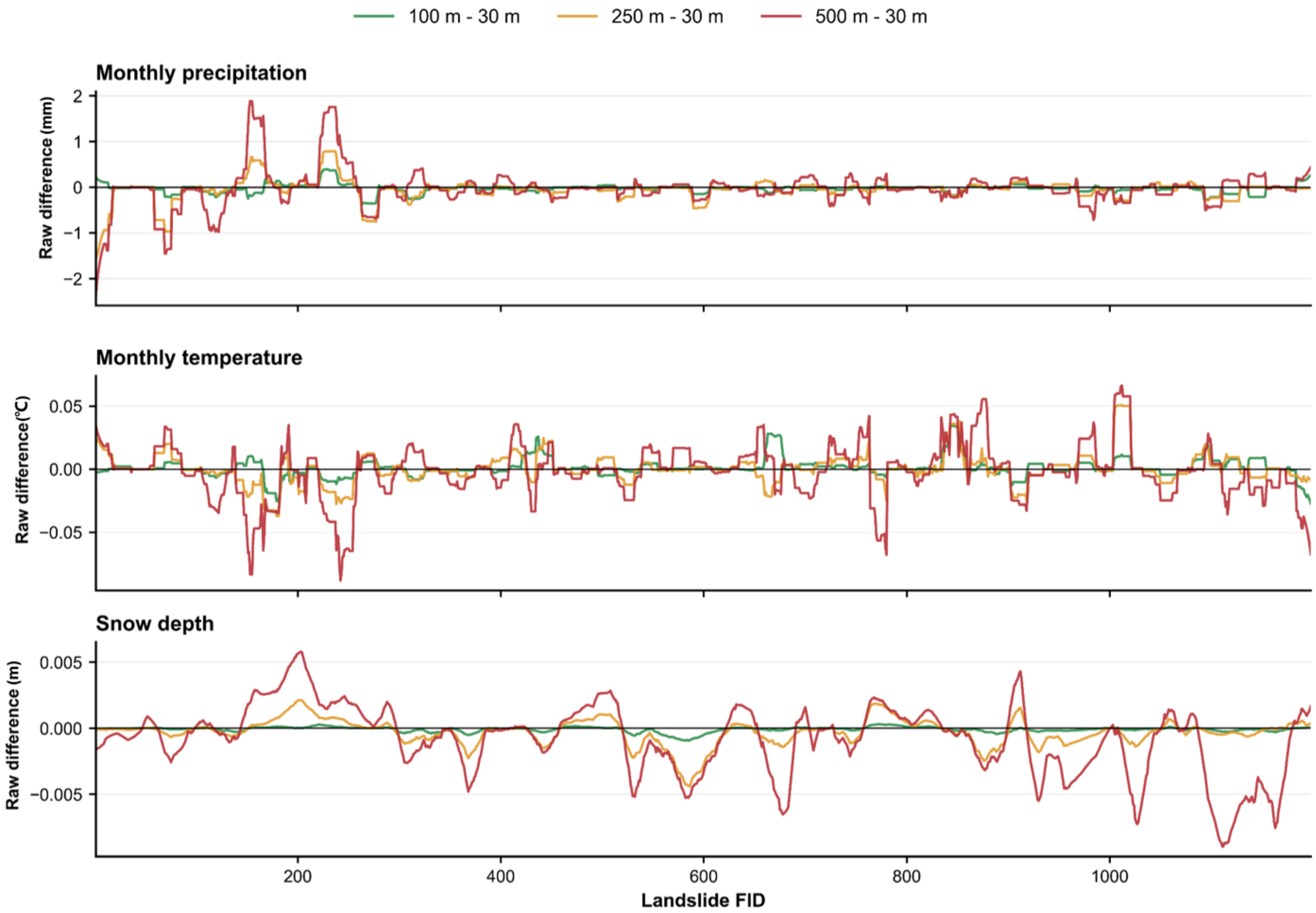

Figure 3. Comparison of conditioning-factor value errors across resolutions.

We acknowledge that a change in scale does introduce a certain degree of uncertainty; however, because an areal-averaging approach was adopted, the uncertainty and the scale dependency arising from the coarser-resolution data are attenuated. Figure 4 shows the distributions of precipitation, temperature and snow depth values over three representative landslide areas at the different resolutions. As the resolution becomes coarser, the local grid boundaries change from a fine mesh to a more evidently blocky representation, yet the mean value within the landslide area varies only slightly. This is mainly because the landslide attribute values were extracted with reference to the geometric centre of the landslide area. Areal averaging can, to a certain extent, attenuate the effects of individual pixel misregistration and of local resampling differences (Blöschl and Sivapalan, 1995; Hengl, 2006). This content will be added to the revised manuscript.

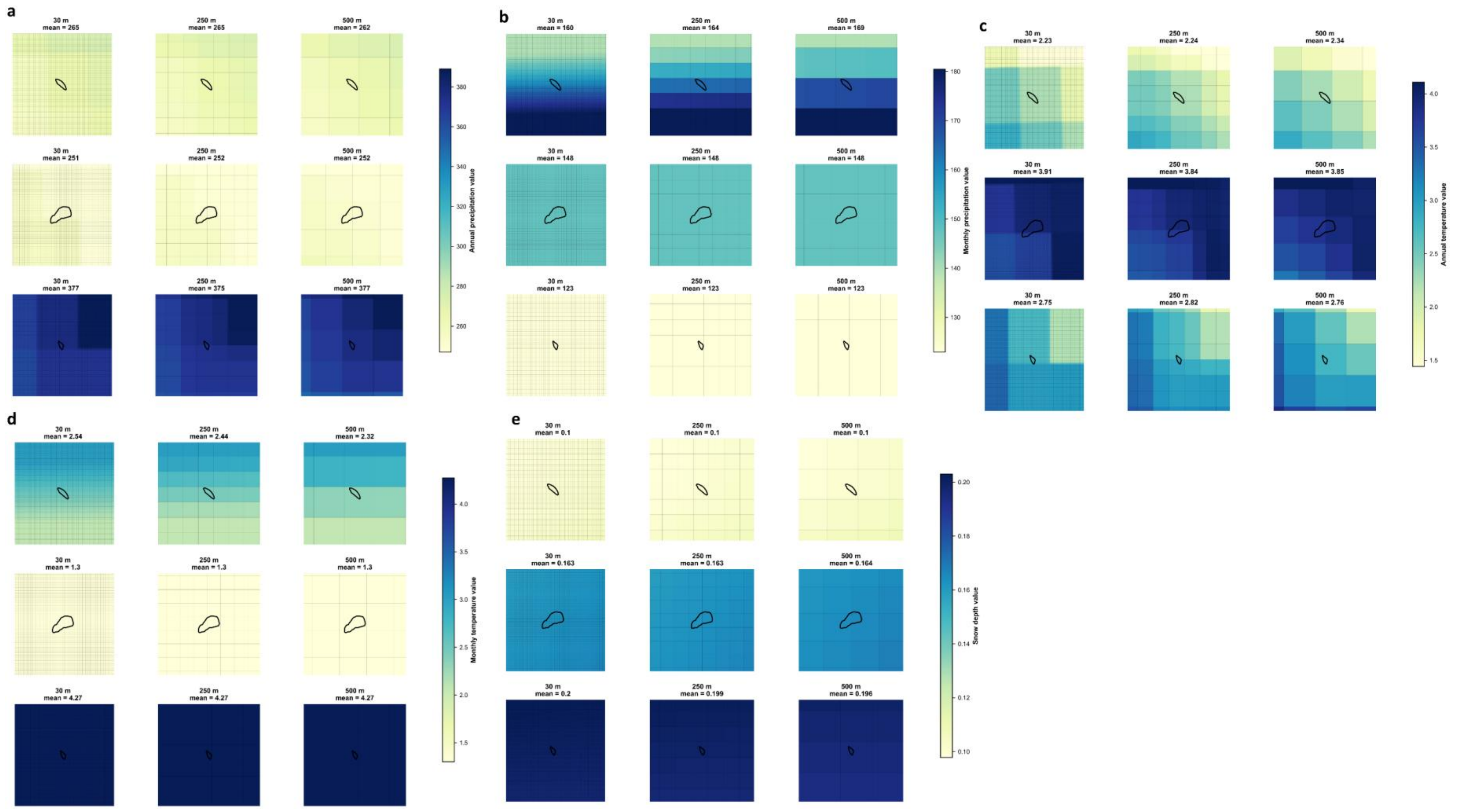

Figure 4. Representative landslide-area distributions of different factors across resolutions.

Q3: The proposed “dependency pathways” remain largely statistical and hypothesis-generating rather than physically validated mechanisms. The final DAG still depends substantially on expert-guided correction, which weakens the claimed objectivity and reproducibility.

Reply: We thank the reviewer for this comment and suggestion. Because landslide dynamic geotechnical parameters are difficult to obtain over large regions, and because of the scale differences involved, it is difficult to reveal the specific physical causal mechanisms of landsliding. The core objective of this study is therefore to construct a dependency-pathway neural network that identifies directional spatial dependency relationships among geological, topographic, hydrological and climatic factors at the large regional scale, rather than to determine physical mechanisms. In the revised manuscript we further clarify the objective and the intended contribution of the study.

Regarding the reviewer's concern about the objectivity and reproducibility of the model. A DAG obtained by purely data-driven structure learning is affected by data resolution, scale mismatch, sample distribution and other factors, and will inevitably contain incorrect edge directions. We consider that constraints based on expert knowledge are a necessary step for improving the plausibility of the dependency structure; however, such constraints must themselves be grounded in peer-recognized knowledge and be verifiable and operational, rather than relying on the subjective judgement of the individual researcher. We have therefore refined the structure-determination strategy, and the rule-based constraints now comprise the following steps.

(1) Macro-hierarchical constraint framework. Based on a review of the literature and on the established understanding of the development of landslides in Tianshan mountain landscape evolution, an assumption was proposed that tectonic activity and geological conditions shape the first-order topographic pattern of a region, that topography in turn modulates the regional climate, and that climatic conditions govern hydrological, freeze–thaw and vegetation processes near the surface, which together control the spatial variation of slope stability (Densmore et al., 1998; Roe et al., 2003). On this basis, we defined the macro-scale orientation.

The factors are constrained hierarchically following the sequence "tectonic–geological → topographic → meteorological–environmental → surface process → landslide". Cross-level and same-level connections are permitted, whereas connections in the reverse direction are not, and any such edges are removed. In addition, the principle of "propagation from base variables to derived variables" is applied, for example, slope and curvature are computed as the first- and second-order spatial derivatives of the DEM, so that, for example, "elevation → slope → curvature" is permitted, whereas reverse connections such as "curvature → slope" or "slope → elevation" are not.

(2) Structural stability constraint based on spatial cross-validation. Building on the constraints above, and in order to further reduce the influence of the randomness of a single training run and of the spatial sample partition on the learned structure, the constrained DAG-GNN was trained repeatedly under spatial cross-validation, and the occurrence frequency and the connection weight of each directed edge across the different training results were recorded. A candidate edge was considered to exhibit adequate structural stability, and was retained, only when its occurrence frequency across the repeated runs exceeded 60% and its mean connection weight exceeded 0.3. Connections with an occurrence frequency not exceeding 60% or a mean weight not exceeding 0.3 were regarded as unstable pathways that are sensitive to the sample partition or to model training, and were removed according to this uniform criterion (Figure 5).

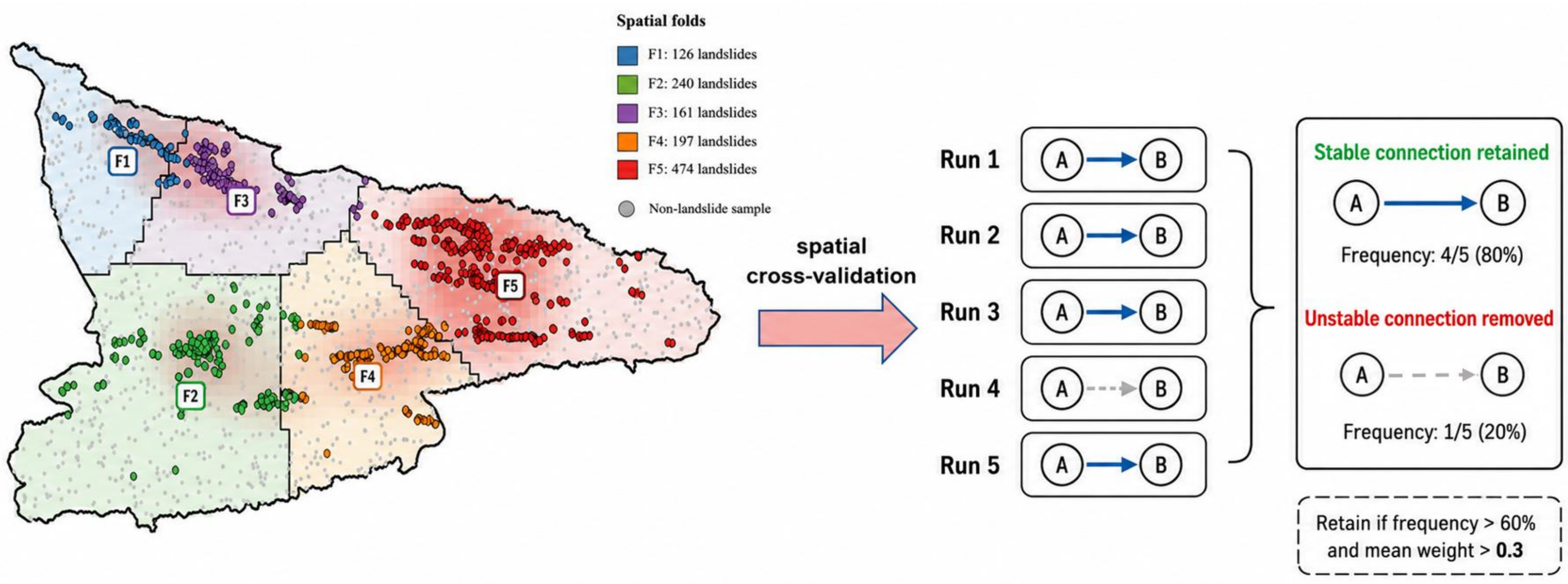

Figure 5. Stability-based edge screening under spatial cross-validation.

Through this procedure, the determination of the final DAG structure changes from the subjective manual judgement of individual connections used in the original manuscript to a uniform screening based on macro-scale rule constraints and on spatially independent repeated training, thereby improving the robustness, objectivity and reproducibility of the resulting dependency pathways.

Q4: In addition, the novelty is overstated, as graph-based and explainable AI methods have already been widely explored in landslide susceptibility studies.

Reply: We thank the reviewer for this comment. Although graph neural networks and related interpretable models have already been applied to landslide susceptibility assessment, existing studies have focused mainly on improving predictive performance and on explaining factor contributions and their interactions. For example, Zhang et al. (2024) applied graph neural networks to landslide susceptibility modelling in order to improve predictive accuracy, whereas Pradhan et al. (2023), Youssef et al. (2023) and Lv et al. (2024) used SHAP, an intrinsically interpretable neural network and spatially explicit SHAP, respectively, to analyse factor importance, response relationships, variable interactions and their spatial variation. Current interpretability analyses based on methods such as SHAP focus primarily on the contributions of individual factors and on variable interactions, and have rarely gone further to characterise directional, dependency pathways or to quantify the importance of such pathways. Bayesian networks have also been used for structural modelling and probabilistic inference in landslide susceptibility, mainly to describe conditional dependence relationships among variables and to estimate landslide occurrence probability (Song et al., 2012; Khabiri et al., 2023). However, the present study goes further by examining which dependency connections appear stably under different spatial training samples, how these stable connections combine into multi-level dependency pathways, how much each pathway contributes to susceptibility, and how these pathway contributions are distributed in space; this is further extended into pathway-specific susceptibility mapping. To the best of our knowledge, studies oriented towards the quantification of dependency-pathway contributions and their spatial decomposition have rarely been carried out in existing landslide susceptibility assessment. The comparison is summarised in Table 2.

Table 2. Differences between representative GNN/XAI studies and the present study.

No.	Representative reference	Existing research	Problem addressed
1	Zhang et al., 2024, Gondwana Research	Used GraphSAGE to aggregate geographic-node features.	Linked geographically distant but environmentally similar areas, improving spatial information propagation and prediction in data-sparse areas.
2	Pradhan et al., 2023, Applied Soft Computing	Applied SHAP to machine-learning-based LSA to explain variable contributions and response	Identified influential factors and how different values increase or decrease susceptibility.

No.	Representative reference	Existing research	Problem addressed
		characteristics.	
3	Youssef et al., 2023, Communications Earth & Environment	Proposed an intrinsically interpretable SNN that decomposes predictions into univariate functions and selected interactions, such as slope-precipitation coupling.	Provided factor-contribution functions and selected interactions directly, extending beyond conventional post-hoc XAI.
4	Lv et al., 2024, Journal of Environmental Management	Extended SHAP from local and global interpretation to spatial interpretation by mapping the spatial distribution of environmental-factor contributions.	Identified where a given factor contributes more strongly to susceptibility.
5	Song et al., 2012, Computers & Geosciences	Used a Bayesian network to represent directed conditional dependencies among landslide factors and infer susceptibility from conditional probabilities.	Explained how factors are conditionally dependent and the associated landslide probability.
6	Khabiri et al., 2023, Remote Sensing	Compared naive Bayes and tree-augmented naive Bayes structures; the latter permits tree-structured dependencies among conditioning factors.	Showed that accounting for factor dependencies and selecting the network structure can affect LSA results.

In addition, we searched the relevant keywords in the Web of Science, and the results are shown in Figure 6. Figure 6a indicates that research related to GNN, SHAP and XAI in landslide susceptibility assessment has grown markedly, increasing from only 1 publication in 2011–2014 to 57 publications in 2019–2022 and further to 278 publications in 2023–2026. By comparison, research related to DAGs and causal discovery, shown in Figure 6b, also displays an upward trend but in clearly smaller numbers, with only 1 publication in each of 2011–2014 and 2015–2018, 12 publications in 2019–2022 and 52 publications in 2023–2026. These results indicate that graph models and interpretable machine learning represent an emerging research trend in landslide susceptibility research, whereas the interpretation of directed dependency pathways based on DAGs or causal discovery remains relatively uncommon.

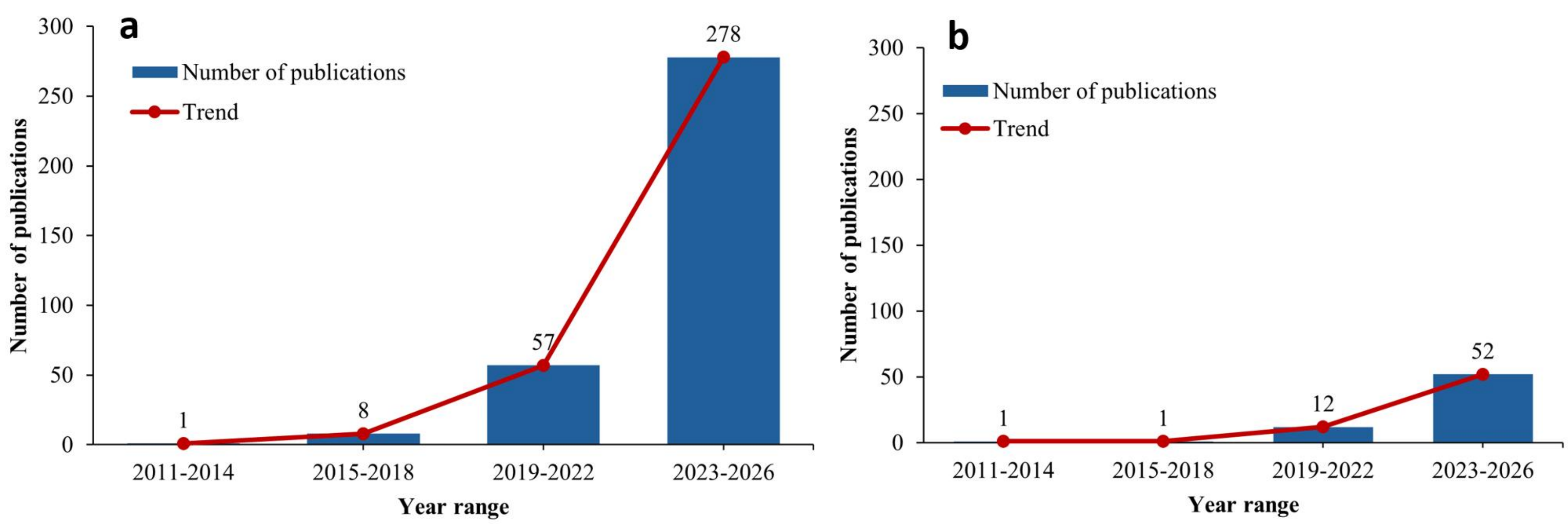

Figure 6. Publication trends for GNN/XAI and causal discovery in landslide susceptibility assessment: (a) landslide susceptibility AND (GNN-related OR XAI-related); (b) landslide susceptibility AND causal-discovery-related.

Q5: The discussion remains descriptive, and the manuscript still contains many problems in logic, expression, figure readability, and overinterpretation.

Reply: We agree that the original manuscript still has shortcomings in the depth of the discussion, in logical coherence, in language and expression, in figure readability and in the interpretation of some of the results. We believe that our paper has some innovation in methodology and applications in landslide susceptibility assessment. We will revise the manuscript to address these problems. Specifically, we will define the focus of the paper as the physical background of the main spatial dependency pathways, their range of applicability and their comparison with existing studies, and we will further discuss the possible reasons for the spatial differences among the pathways. We will discuss the potential value of statistical directed dependence. The English expression and the use of terminology throughout the manuscript will be comprehensively checked in order to improve accuracy and readability. Finally, we will re-examine and improve the main figures, including fonts, legends, annotations, resolution and inset enlargements, so as to improve their legibility. Through these revisions, we hope to satisfy the requirement of this prestigious journal.

References

Blöschl, G., and Sivapalan, M., 1995. Scale issues in hydrological modelling: A review. Hydrological Processes, 9(3-4), 251-290.

Densmore, A. L., Ellis, M. A., and Anderson, R. S., 1998. Landsliding and the evolution of normal-fault-bounded mountains. Journal of Geophysical Research: Solid Earth, 103(B7), 15203-15219. https://doi.org/10.1029/98JB00510.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.-T., 2012. Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, 112(1-2), 42-66.

Hengl, T., 2006. Finding the right pixel size. Computers & Geosciences, 32(9), 1283-1298.

Khabiri, S., Crawford, M. M., Koch, H. J., Haneberg, W. C., and Zhu, Y., 2023. An assessment of negative samples and model structures in landslide susceptibility characterization based on Bayesian network models. Remote Sensing, 15, 3200. https://doi.org/10.3390/rs15123200.

Lv, J., Zhang, R., Shama, A., Hong, R., He, X., Wu, R., Bao, X., and Liu, G., 2024. Exploring the spatial patterns of landslide susceptibility assessment using interpretable Shapley method: Mechanisms of landslide formation in the Sichuan-Tibet region. Journal of Environmental Management, 366, 121921. https://doi.org/10.1016/j.jenvman.2024.121921.

Pradhan, B., Dikshit, A., Lee, S., and Kim, H., 2023. An explainable AI (XAI) model for landslide susceptibility modeling. Applied Soft Computing, 142, 110324.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Reviews, 180, 60-91.

Roe, G. H., Montgomery, D. R., and Hallet, B., 2003. Orographic precipitation and the relief of mountain ranges. Journal of Geophysical Research: Solid Earth, 108(B6), 2315. https://doi.org/10.1029/2001JB001521.

Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., et al., 2012. Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. Computers & Geosciences, 42, 189-199. https://doi.org/10.1016/j.cageo.2011.09.011.

Youssef, K., Shao, K., Moon, S., and Bouchard, L.-S., 2023. Landslide susceptibility modeling by interpretable neural network. Communications Earth & Environment, 4, 162.

Yu, Y., Chen, J., Gao, T., and Yu, M., 2019. DAG-GNN: DAG structure learning with graph neural networks. Proceedings of the 36th International Conference on Machine Learning, PMLR 97, 7154-7163.

Zhang, Q., et al., 2024. A landslide susceptibility assessment method considering the similarity of geographic environments based on graph neural network. Gondwana Research, 132, 323-342.