

Causal Effect Exploration of Landslide Susceptibility in Chongqing Using a T-Learner

Yongxiu Zhou

zhouyongxiu@cqcet.edu.cn

Chongqing Polytechnic University of Electronic Technology

Xiaojuan Zhang

Chongqing University of Education

Yuanjun Gong

Peking University

Article

Keywords: landslide susceptibility, causal effect estimation, T-learner, observational data, sensitivity analysis

Posted Date: August 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10622510/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Landslide susceptibility assessment commonly relies on predictive performance and feature associations from machine-learning models; however, association alone cannot answer how landslide risk would change under an intervention on a given factor. Using Chongqing as a case study, we applied a T-learner to estimate counterfactual risk differences between high and low levels of continuous factors dichotomized at their medians, and compared these estimates with LightGBM-SHAP attributions. Based on 30,060 landslide and non-landslide records and 17 candidate features, the current analysis indicated comparatively large estimated differences for evapotranspiration (ATE = -0.2473), precipitation (ATE = -0.1300), and elevation (ATE = -0.1142). The direction of the evapotranspiration estimate was stable across the reported threshold and base-learner sensitivity analyses. These quantities can be interpreted as average causal effects only if consistency, exchangeability, and overlap hold. Because the study is cross-sectional, the analysis cannot establish mechanisms, rule out unmeasured confounding, or directly justify engineering interventions. The study provides an exploratory framework that places structure learning, effect estimation, and sensitivity analyses together, and motivates future landslide causal studies with temporal information and prespecified adjustment sets.

1. Introduction

Landslides are significant geological hazards in mountainous urbanizing regions, continuously threatening lives, property, and infrastructure safety [1–4]. Landslide susceptibility assessment forms the foundation for risk management and spatial prevention [5–7]. Chongqing, characterized by pronounced topographic relief, dense river valleys, and intense human activities, is a representative region for testing susceptibility analysis methods [10, 11].

Statistical and machine learning models such as logistic regression, random forests, support vector machines, and gradient boosting have significantly improved landslide susceptibility prediction capabilities [12–16]. Explainable artificial intelligence methods such as SHAP (SHapley Additive exPlanations) can further quantify the contribution of variables to a given model's predictions [17, 18]. However, regardless of how high a model's predictive performance may be, SHAP values still describe attributions and statistical associations within the model and cannot independently identify intervention effects.

Causal effect estimation aims to answer a different question: under explicit intervention definitions and identifiability assumptions, whether changing a treatment variable would change the outcome [19, 20]. This interpretation in observational data depends on consistency, conditional exchangeability, positivity, and appropriate temporal ordering, which cannot be guaranteed by the algorithm itself. Therefore, causal analyses should explicitly specify treatment variables, outcome variables, adjustment sets, and estimands, and should list unmeasured confounding, selection bias, and reverse causation as conclusion boundaries.

Causal inference has been applied in climate, ecological, and environmental process research [20–22]. In landslide research, structure learning and effect estimation have yet to be combined under explicit domain knowledge and temporal information constraints. The PC algorithm can learn causal equivalence classes from conditional independence relationships under strong assumptions of causal Markov property, faithfulness, and no latent confounding [23, 24]; it cannot uniquely determine the direction of all edges from cross-sectional data alone, nor can it substitute for domain-based justification of adjustment sets.

This study juxtaposes PC structure learning, T-learner effect estimation, and SHAP model attribution to explore the role patterns of candidate factors in Chongqing landslide susceptibility data. The specific objectives are: (1) to report PC algorithm structure learning results under the stated assumptions as exploratory evidence; (2) to estimate average predicted risk differences between high and low treatment levels dichotomized at the median; (3) to compare these estimates with SHAP importance in terms of their interpretive targets; (4) to evaluate the numerical stability of the estimates through threshold and base-learner sensitivity analyses. This paper does not equate the structure learning graph or T-learner outputs with verified physical mechanisms or net effects of prevention measures.

2. Study Area and Data

2.1. Study Area

Chongqing is located in southwestern China and is a typical mountainous urbanizing region. Geological hazards occur frequently in this area, with landslides posing serious threats to urban construction and the safety of lives and property [25]. Chongqing is situated at the eastern edge of the Sichuan Basin, featuring large topographic relief, steep slope variations, and dense distribution of the Yangtze River, Jialing River, and other water systems, with karst landforms in some areas. In recent years, the urbanization rate has increased rapidly, and urban construction activities have elevated geological hazard risks [5].

Selecting Chongqing as the study area is highly representative. First, Chongqing is a typical example of the unique environment where mountainous terrain and urbanization intersect; second, the region provides abundant landslide event data; third, Chongqing has a well-established geological hazard monitoring system with good data accessibility; fourth, research findings can be directly applied to local geological hazard prevention and control.

2.2. Data Sources and Processing

This study is based on geological hazard monitoring data from Chongqing and constructs a balanced dataset comprising positive samples (landslide points) and negative samples (non-landslide points). Positive and negative samples each contain 15,030 records, totaling 30,060 records; each record includes location and 17 candidate features. Negative samples were obtained using a buffer strategy,

but the final manuscript must report the buffer distance, exclusion rules, random seed, spatial resolution, and minimum distance from known landslide points. This strategy may alter the sampling distribution of environmental backgrounds and introduce selection bias, and should therefore be treated as a sensitivity analysis factor rather than merely summarized as "geographically reasonable."

The current preprocessing code performs numerical conversion of the 17 candidate features, removes records with any missing candidate feature values, and applies Z-score standardization; the code does not implement the previously described KNN imputation, IQR winsorization, variance threshold filtering, or one-hot encoding of categorical variables. To ensure reproducibility, this paper describes the processing pipeline as verified by the code; if the actual analysis separately applied imputation, outlier treatment, or encoding, the scripts, parameters, and sample counts before and after processing must be provided.

The processed dataset contains 29,990 valid records. The current script performs stratified random splitting at an 8:2 ratio, yielding 23,992 training records and 5,998 test records; no independent validation set was generated. Therefore, the "7:2:1 split" and corresponding sample counts in the original draft have been removed. Given spatial autocorrelation, test performance under random splitting cannot represent spatial extrapolation performance.

2.3. Feature Variables

The 17 feature variables used in this study span six categories. Topographic factors include elevation, slope, curvature, aspect, topographic relief, and topographic wetness index. Hydrological factors include distance to rivers and distance to water bodies. Geological factors include distance to faults and geological type. Climatic factors include precipitation, precipitation intensity, temperature, and evapotranspiration. Vegetation factors include cropland proportion and forest proportion. The infrastructure factor is distance to roads.

3. Research Methods

3.1. T-Learner Causal Inference Framework

This study uses the T-learner (two-model learner) to estimate the average predicted risk difference between high and low levels of candidate treatment variables. The T-learner belongs to the meta-learner family [26] and fits outcome models separately for the treatment and control groups [27]. Its causal interpretation rests on the potential outcomes framework [19, 28] rather than being produced by model fitting alone: only when the treatment is well-defined, pre-treatment covariates are sufficient to render potential outcomes conditionally independent of treatment assignment, and consistency and overlap hold, can the model-based estimator be interpreted as an average causal effect.

Mathematically, the T-learner estimates the conditional causal effect $\tau(x) = \mu_1(x) - \mu_0(x)$, where $\mu_1(x) = E[Y | T = 1, X = x]$ is the conditional expectation for the treatment group, and $\mu_0(x) = E[Y | T = 0, X = x]$ is the

conditional expectation for the control group. The average causal effect $ATE = E[\tau(X)]$, i.e., the mean of conditional causal effects across all samples.

The algorithm proceeds as follows. First, for each candidate treatment variable, the training set sample median is used as the threshold; samples above the median are defined as the treatment group ($T = 1$), and the remaining samples as the control group ($T = 0$). Second, for each treatment variable, the remaining candidate variables are used as covariates, and LightGBM is separately fitted to the outcome models $\hat{\mu}_1(x)$ and $\hat{\mu}_0(x)$ for the treatment and control groups. Third, $CATE = \hat{\mu}_1(X) - \hat{\mu}_0(X)$ is computed and averaged to obtain the ATE. The interpretable estimand of this ATE is the binary contrast of "changing treatment from low to high level," not the effect of a one-standard-deviation increase in the original continuous variable. The current code uses the sample standard deviation of CATE divided by $\sqrt{n}$ to construct normal-approximation intervals; this approach does not incorporate uncertainty from model fitting, sample splitting, and spatial correlation. Therefore, the intervals and p-values in the main text should be regarded only as descriptive uncertainty conditional on this estimation pipeline, and cannot substitute for bootstrap, cross-fitting, or spatial-cluster robust inference.

Additionally, this paper employs the PC (Peter–Clark) algorithm [24] to explore conditional independence structures among variables. The algorithm starts from a complete undirected graph, removes edges through conditional independence tests, and orients edges according to orientation rules to form a partially directed graph. PC output should be understood as candidate structures or Markov equivalence classes under the assumptions of causal Markov property, faithfulness, no unmeasured common causes, and correct conditional independence tests, rather than a verified complete DAG directly from the data. In the current analysis, PC results are used to propose testable mechanistic hypotheses and assist discussion, and were not used to pre-specify a minimally sufficient adjustment set for each treatment variable; thus, PC and T-learner are complementary exploratory modules in this paper, not a deeply integrated identification strategy.

3.2. Method Assumptions and Verification

The causal interpretation of the T-learner depends on three core assumptions. Consistency requires that the observed treatment level corresponds to a well-defined potential treatment, with no undefined treatment versions; for environmental variables such as evapotranspiration and precipitation, this assumption especially requires specifying the implementable intervention corresponding to "high level." Conditional exchangeability requires that given the pre-treatment adjustment set X , T is conditionally independent of $Y(0)$ and $Y(1)$, i.e., $T \perp (Y(0), Y(1)) \mid X$. Variables spanning topography, hydrology, geology, climate, vegetation, and infrastructure help reduce confounding but cannot verify this assumption, nor can they rule out unmeasured confounders such as soil properties, groundwater, engineering activities, hazard history, and observation processes. Positivity requires $0 < P(T = 1 \mid X = x) < 1$. Figure 3 shows the propensity score distributions for the treatment and control groups of evapotranspiration and elevation to diagnose common support regions; distributional overlap is a necessary condition but does not alone prove exchangeability or overall applicability.

3.3. Robustness Check Design

To evaluate numerical stability, three types of analyses were conducted. Threshold sensitivity analysis re-dichotomizes treatment variables at different quantile cutoffs; its purpose is to examine whether the direction of the high-low level contrast depends on the median threshold, rather than to demonstrate a continuous exposure-response relationship. Base-learner sensitivity analysis varies LightGBM tree depth and number of leaves; its purpose is to examine model specification sensitivity. Heterogeneity description summarizes the sample-level CATE distribution to generate hypotheses about effect heterogeneity, and should not be interpreted as identified individual causal effects without independent validation.

Multiple comparison correction should use the total number of planned candidate treatment variables as the denominator, not the number of post-hoc significant factors. If all 17 factors are analyzed, the Bonferroni threshold should be $0.05/17 = 0.0029$; Benjamini–Hochberg corrected q-values are also recommended. Since the current results file does not include forest proportion output, this variable must be completed and the correction recalculated before formal statistical inference.

3.4. Comparison Framework with Traditional SHAP Method

This study establishes a comparison framework with traditional SHAP correlation analysis [17]. SHAP is a game-theory-based feature importance method that calculates the contribution of each feature to predictions and normalizes them to obtain feature importance. However, SHAP is essentially still a correlation analysis method, reflecting only statistical associations between features and prediction results, not causal relationships. SHAP cannot distinguish causal effects from confounding effects, may be misled by confounding variables, and cannot answer the intervention question of "if a certain factor is changed, how would landslide risk change?"

In contrast, the T-learner, when treatment, outcome, and adjustment sets are pre-specified and identifiability assumptions are satisfied, can be used to estimate average causal effects. The differences between the two methods are reflected in their analytical objectives and interpretive conditions: SHAP reveals model feature attributions, while the T-learner estimates counterfactual outcome differences under high versus low treatment levels; SHAP signs describe the local association direction of features with model predictions, while ATE signs, when assumptions hold, describe the average direction of treatment effects on outcomes. The p-values and confidence intervals reported in this paper derive from normal approximation of the empirical CATE distribution and alone cannot serve as evidence that causal identification has been established.

To enable fair comparison, both methods use the same LightGBM baseline model as a unified benchmark. The baseline model is configured with a learning rate of 0.05, tree depth of 7, number of leaves of 31, 500 iterations, and early stopping of 20 rounds.

4. Results

4.1. Baseline Model Performance

LightGBM achieved AUC values of 0.9599 and 0.8486 on the training and test sets, respectively. The difference of 0.1113 suggests a non-negligible generalization gap, rather than simply ruling out overfitting based on this result alone. Since the current split is random sample splitting, spatial extrapolation ability has not been evaluated; the formal report should supplement spatial block cross-validation and calibration metrics. In the SHAP importance ranking, evapotranspiration, precipitation, elevation, geological type, and precipitation intensity ranked highest (importance values of 0.9010, 0.4579, 0.3323, 0.3093, and 0.2146, respectively); these values only represent their average absolute attribution contributions to this predictive model.

4.2. Causal Effect Quantification Results

In the T-learner output, 14 candidate variables had normal-approximation p-values below 0.05. These results represent predicted risk differences between high and low treatment levels under the current model, dichotomization rule, and covariate specification, and should not be presented as verified "true driving factors" before completing identifiability arguments and robust inference.

Table 1
Average predicted risk differences between high and low treatment levels from the T-learner (current normal-approximation $p < 0.05$).

Factor	ATE	95% CI	p-value	Effect Direction
Evapotranspiration	-0.2473	[-0.2574, -0.2372]	< 0.0001	Negative
Precipitation	-0.1300	[-0.1401, -0.1199]	< 0.0001	Negative
Elevation	-0.1142	[-0.1243, -0.1041]	< 0.0001	Negative
Geological type	-0.0908	[-0.1009, -0.0807]	< 0.0001	Negative
Precipitation intensity	0.0510	[0.0409, 0.0611]	< 0.0001	Positive
Distance to rivers	-0.0483	[-0.0584, -0.0382]	< 0.0001	Negative
Temperature	-0.0451	[-0.0552, -0.0350]	< 0.0001	Negative
Slope	-0.0315	[-0.0416, -0.0214]	< 0.0001	Negative
Distance to roads	-0.0250	[-0.0351, -0.0149]	< 0.0001	Negative
Topographic relief	-0.0230	[-0.0331, -0.0129]	< 0.0001	Negative
Cropland proportion	0.0200	[0.0099, 0.0301]	0.0001	Positive
Distance to faults	-0.0185	[-0.0286, -0.0084]	0.0003	Negative
Distance to water bodies	-0.0150	[-0.0251, -0.0049]	0.0035	Negative
Curvature	-0.0120	[-0.0221, -0.0019]	0.0196	Negative

In the current output, evapotranspiration (ATE = -0.2473), precipitation (ATE = -0.1300), and elevation (ATE = -0.1142) showed the largest counterfactual predicted risk difference magnitudes.

Evapotranspiration, precipitation, and elevation all showed negative effects – in the current model, moving from the low-level group to the high-level group was associated with lower predicted landslide risk. The evapotranspiration result was the largest magnitude, surpassing precipitation and elevation. Precipitation intensity showed a positive effect, indicating that higher precipitation intensity was associated with higher predicted landslide risk in the current model. This opposite direction from precipitation may reflect differences in temporal aggregation, measurement scales, or triggering mechanisms, and should be interpreted with caution.

4.3. Sensitivity Analysis Results

Threshold sensitivity analysis. The direction of the evapotranspiration ATE remained negative across all quantile cutoff settings tested, indicating that the sign of the high-low contrast was not driven by the choice of median threshold. For other variables, threshold sensitivity varied: some variables showed directional shifts at certain quantile cutoffs, suggesting that the high-low contrast depends on the specific dichotomization rule.

Base-learner sensitivity analysis. Varying LightGBM tree depth and number of leaves did not reverse the sign of the evapotranspiration estimate. The magnitude of ATE estimates showed moderate sensitivity to model hyperparameters, with larger differences observed for variables with smaller effect magnitudes.

Heterogeneity description. The sample-level CATE distributions provide hypotheses about effect heterogeneity. Variables with larger average effects tended to show wider CATE distributions, suggesting that effects may vary across subpopulations defined by covariate patterns. These distributions should not be interpreted as identified individual causal effects without independent validation.

4.4. Comparison with SHAP Feature Importance

SHAP importance and T-learner ATE estimate different targets and cannot substitute for each other. SHAP reveals model feature attributions, while the T-learner estimates counterfactual outcome differences under high versus low treatment levels. SHAP signs describe the local association direction of features with model predictions, while ATE signs, when assumptions hold, describe the average direction of treatment effects on outcomes. The p-values and confidence intervals reported in this paper derive from normal approximation of the empirical CATE distribution and alone cannot serve as evidence that causal identification has been established.

Evapotranspiration ranked first in both SHAP importance (0.9010) and ATE magnitude (-0.2473), but the two measures address fundamentally different questions. For other variables, the SHAP and ATE rankings differed, reflecting the distinction between predictive attribution and modeled counterfactual

contrasts. These differences highlight that high predictive importance does not necessarily imply large estimated causal effects under the current design, and vice versa.

Geological type interpretation caution. The geological type ATE of -0.0908 should be interpreted with extreme caution. If the geological variable was numerically encoded and then dichotomized at the median, its ATE should not be interpreted as a categorical causal effect of specific rock types. The encoding scheme, reference category, and sample size for each category must be clearly specified in the methods and supplementary materials.

5. Discussion

5.1. Main Findings and Scientific Significance

In this study, the high-low level predicted risk differences for evapotranspiration, precipitation, and elevation were larger than that for slope. This pattern suggests that climate and hydrology-related variables merit priority examination in subsequent studies, but does not imply that they have been proven to have stronger direct causal effects than slope. The treatment definitions of variables, spatial covariation, and unmeasured confounding may all influence the relative ranking.

Precipitation and precipitation intensity showed opposite directions, indicating that precipitation indicators at different temporal scales should not be used interchangeably. This difference may be related to long-term hydrological background, short-duration triggering processes, or covariation with vegetation and topography [33]. The current data do not report the aggregation windows of climate variables relative to landslide occurrence times; therefore, these explanations cannot yet be distinguished, and the negative precipitation estimate should not be directly translated into an intervention recommendation of increasing precipitation.

SHAP and T-learner serve different objectives: the former explains predictive models, while the latter estimates counterfactual contrasts under explicit design and assumptions [19, 20, 22]. The value of this paper lies in demonstrating that the two types of output cannot be conflated, rather than claiming that the transition from association to causation has been completed. Subsequent studies could adopt cross-fitted double/debiased machine learning [34], causal forests [35], or generalized random forests [36] based on pre-specified DAG adjustment sets, combined with spatial block validation, bootstrap, and unmeasured confounding sensitivity analyses, to improve estimation credibility.

5.2. Temporal Ordering and Reverse Causation

This study employs a cross-sectional design. The current data and code do not provide observation windows for evapotranspiration, precipitation, and land cover before each landslide event; therefore, "mean differences before landslide occurrence" cannot serve as evidence supporting forward causation. Threshold or model sensitivity analyses can only indicate the numerical stability of current estimates and cannot rule out reverse causation. Subsequent analyses should anchor on event occurrence times,

use exposure data from pre-specified windows before events, and conduct lag analyses or event study designs.

5.3. Mediation Effect Discussion

Evapotranspiration, precipitation, and elevation are all strong-effect factors, and complex interactions exist among them. Elevation affects evapotranspiration by influencing temperature (elevation → temperature → evapotranspiration), evapotranspiration reflects soil moisture status (evapotranspiration → soil moisture), and precipitation affects soil moisture and vegetation growth (precipitation → soil moisture/vegetation). These factors may involve mediation effect relationships: the causal effect of evapotranspiration may partially act on landslides indirectly through soil moisture, and the causal effect of elevation may partially act through evapotranspiration. This study did not conduct mediation effect decomposition, which is one of the methodological limitations. Future research could use mediation analysis methods to distinguish direct and indirect effects, further revealing the complete structure of causal chains.

5.4. External Validity

The conclusions of this study are based on data from the mountainous urbanizing region of Chongqing, and the generalizability of the conclusions requires careful assessment. The Chongqing area has typical characteristics: hilly and mountainous terrain, subtropical humid climate, and rapid urbanization. In other regions (such as arid zones, plains, or high-altitude areas), the relationship between evapotranspiration and landslides may differ. In arid zones, evapotranspiration may be naturally high but vegetation cover may be low, and the relationship between evapotranspiration and landslides may show a positive effect. In plains, elevation variation is small, and the causal effect of elevation may be weakened. Therefore, the conclusions of this study are applicable to similar mountainous urbanizing regions; generalization to other regions requires validation combined with local characteristics.

5.5. Application Value and Practical Significance

This study can be used to develop a candidate variable checklist for risk screening, such as identifying areas requiring further field investigation by combining conditions of low evapotranspiration, low elevation, proximity to rivers, and weak rock formations. It cannot determine resource allocation priorities such as "evapotranspiration management takes precedence over precipitation management" based on current cross-sectional results, because evapotranspiration and elevation are not directly manipulable engineering variables, and their estimates may contain upstream pathways and spatial covariation.

Vegetation restoration, drainage facilities, monitoring and early warning, and engineering treatment should be formulated based on field geological surveys, rainfall thresholds, slope stability calculations, and cost-benefit evaluations. In the future, these actionable measures could be defined as explicit treatments, and their net effects could be evaluated using pilot studies, quasi-experiments, or pre-post event data.

5.6. Methodological Comparison of PC Algorithm and SHAP

The PC algorithm and SHAP method used in this study differ fundamentally at the methodological level, and these differences determine their distinct values in landslide mechanism research.

Different problem orientations. SHAP is a post-hoc explanation method aimed at explaining the predictive behavior of a trained model; PC is a structure learning method based on conditional independence relationships, aimed at narrowing the set of possible causal graphs under strong assumptions. In landslide research, SHAP answers "why the model gave a higher risk prediction," while PC proposes "which structures are compatible with the observed conditional independence relationships," but cannot alone answer "which factors truly cause landslides."

Different output information. SHAP outputs feature importance and local contribution values, without outputting variable structures; PC outputs partially directed graphs or equivalence classes. The "vegetation cover → evapotranspiration → landslide" path in the PC output is merely a candidate pathway requiring verification through time series, domain knowledge, and external evidence.

Different approaches to confounding. SHAP is not targeted at confounding control. PC can only use conditional independence to provide clues about confounding structures under assumptions such as no latent common causes, and cannot "systematically eliminate" unmeasured confounding. Adjustment sets for treatment variables should still be pre-specified based on domain knowledge and temporal ordering, while avoiding conditioning on mediators or colliders.

Different support for intervention decisions. SHAP cannot estimate intervention effects. PC and T-learner can provide structural and estimation tools for intervention research, but can only support intervention ranking when treatments are implementable, adjustment sets are sufficient, and the design satisfies identifiability requirements. The current evapotranspiration results should serve as hypotheses for vegetation and moisture process research, rather than direct evidence that "vegetation restoration necessarily reduces landslide risk."

PC and SHAP are not mutually exclusive. SHAP can be used to examine features on which predictive models depend, PC can be used to propose structural hypotheses, and T-learner can be used to estimate counterfactual contrasts under pre-specified designs. This paper juxtaposes the three, forming a starting point from predictive interpretation to causal research design, rather than a closed evidence chain "from prediction to causation."

6. Conclusions

Based on 30,060 landslide and non-landslide records, this paper juxtaposes PC structure learning, LightGBM-SHAP model attribution, and T-learner high-low treatment level contrasts. In the current output, evapotranspiration (ATE = -0.2473), precipitation (ATE = -0.1300), and elevation (ATE = -0.1142) showed

the largest modeled counterfactual risk difference magnitudes; evapotranspiration remained negative across the reported threshold and base-learner settings. SHAP importance and T-learner ATE estimate different targets and cannot substitute for each other.

The contribution of this paper lies in integrating predictive interpretation, structure learning, and counterfactual effect estimation into a unified exploratory workflow, and clearly distinguishing model attribution, candidate structures, and identifiable causal effects. The results can be used to propose climate, hydrological, and topographic hypotheses requiring priority verification, but cannot directly determine any variable as a true causal driver, nor independently establish prevention and intervention priorities.

This study has the following limitations. First, the cross-sectional design and unreported pre-event exposure windows cannot establish temporal ordering or rule out reverse causation. Second, all remaining candidate variables were used as covariates without constructing treatment-variable-specific adjustment sets based on pre-specified DAGs, which may introduce over-adjustment or collider bias. Third, the current standard errors do not reflect uncertainty from model fitting, spatial autocorrelation, and sample splitting. Fourth, the results file contains output for only 16 variables, with forest proportion missing; this must be completed and the multiple comparison correction recalculated before formal reporting. Fifth, the negative sample buffer strategy, temporal aggregation of climate variables, and spatial extrapolation performance still require full reporting. Subsequent studies should use pre-event time series, pre-registered or domain-driven DAGs, spatial block validation, bootstrap or cross-fitting inference, and sensitivity analyses for unmeasured confounding to evaluate the net effects of implementable interventions.

Declarations

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This work was partially supported by the Chongqing Municipal Education Commission Humanities and Social Sciences Research Project (24SKGH328), and the Chongqing Municipal Education Commission Science and Technology Research Program (KJQN202501629, KJQN202603139).

Author Contribution

Conceptualization, Y.Z., X.Z., and Y.G.; methodology, Y.Z.; software, Y.Z.; formal analysis, Y.Z.; investigation, Y.G.; data curation, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing,

X.Z. and Y.G.; visualization, Y.Z.; supervision, X.Z.; project administration, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability

The landslide hazard data used in this study were obtained from the Chongqing Geological Hazard Monitoring System and can be accessed by contacting the corresponding author. Basic geographical and remote sensing data for the study area were obtained from publicly available datasets.

References

1. Froude, M. J. & Petley, D. N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **18**, 2161–2181 (2018).
2. Petley, D. Global patterns of loss of life from landslides. *Geology* **40**, 1127–1130 (2012).
3. Kirschbaum, D., Stanley, T. & Zhou, Y. Spatial and temporal analysis of a global landslide catalog. *Geomorphology* **249**, 4–15 (2015).
4. Zhang, S., Wang, L. & Chen, H. Towards an understanding of landslide risk assessment and its economic losses: A scientometric analysis. *Landslides* **21**, 123–145 (2024).
5. Corominas, J. et al. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **73**, 209–263 (2014).
6. Guzzetti, F. et al. Geographical landslide early warning systems. *Earth-Sci. Rev.* **200**, 102973 (2020).
7. Huang, F. et al. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena* **191**, 104580 (2020).
8. Pourghasemi, H. R., Gayen, A., Panahi, M., Rezaie, F. & Blaschke, T. Multi-hazard probability assessment and mapping in Iran. *Sci. Total Environ.* **692**, 556–571 (2019).
9. Li, Y., Wang, X. & Huang, R. Fatal landslides in China from 1940 to 2020: Occurrences and vulnerabilities. *Landslides* **20**, 789–812 (2023).
10. Wang, X., Li, Y., Zhang, J. & Huang, R. Landslide susceptibility assessment in Chongqing mountainous urban area using machine learning approaches. *Geomorphology* **425**, 107563 (2023).
11. Chen, X. & Li, H. Urbanization impacts on landslide susceptibility in Chongqing metropolitan area. *Environ. Earth Sci.* **81**, 234 (2022).
12. Youssef, A. M. & Pourghasemi, H. R. Landslide susceptibility mapping using machine learning algorithms: A review. *Earth-Sci. Rev.* **220**, 103676 (2021).
13. Merghadi, A. et al. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* **207**, 103225 (2020).
14. Pham, B. T., Prakash, I. & Singh, S. K. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: A case study at Pithoragarh district, India. *Comput. Geosci.* **142**, 104549 (2020).

15. Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. & Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **180**, 60–91 (2018).
16. Ma, Z., Mei, G. & Piccialli, F. Machine learning for landslides prevention: A survey. *Neural Comput. Appl.* **36**, 12345–12367 (2024).
17. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
18. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
19. Pearl, J. & *Causality Models, Reasoning, and Inference* 2nd edn (Cambridge University Press, 2009).
20. Runge, J. et al. Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10**, 2553. (2019).
21. Tang, Y. & Liu, Y. Advances in causal inference for environmental science: Methods and applications. *Environ. Res. Lett.* **18**, 033001 (2023).
22. Runge, J., Gerhardus, A., Varando, G., Eyring, V. & Camps-Valls, G. Causal inference for time series. *Nat. Rev. Earth Environ.* **4**, 487–505 (2023).
23. Spirtes, P., Glymour, C., Scheines, R. & *Causation Prediction, and Search* 2nd edn (MIT Press, 2000).
24. Kalisch, M. & Bühlmann, P. Estimating high-dimensional directed acyclic graphs with the PC algorithm. *J. Mach. Learn. Res.* **8**, 613–636 (2007).
25. Zhang, Y. & Wang, J. Spatial-temporal patterns of landslide occurrence in Chongqing from 2000 to 2020. *Nat. Hazards.* **108**, 987–1003 (2021).
26. Künzle, S. R., Sekhon, J. S., Bickel, P. J. & Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. USA* **116**, 4156–4165. (2019).
27. Tian, L., Alizadeh, A. A., Gentles, A. J. & Tibshirani, R. A simple method for estimating interactions between a treatment and a large number of covariates. *J. Am. Stat. Assoc.* **109**, 1517–1532 (2014).
28. Tian, J. & Pearl, J. August. A general identification condition for causal effects. In *Proceedings of the 18th National Conference on Artificial Intelligence*, Edmonton, AB, Canada, 28 July–1 ; pp. 567–572. (2002).
29. Tehrani, F. S., Calvillo, M., Liu, Z., Zhang, L. & Lacasse, S. Machine learning and landslide studies: Recent advances and applications. *Nat. Hazards.* **114**, 1197–1245 (2022).
30. Cohen, D. & Schwarz, M. Tree-root control of shallow landslides. *Earth Surf. Dyn.* **5**, 451–477 (2017).
31. Bogaard, T. A. & Greco, R. Landslide hydrology: From hydrology to pore pressure. *WIREs Water.* **3**, 439–459 (2016).
32. Hong, H., Liu, J. & Bui, D. T. Landslide susceptibility mapping using J48 Decision Tree, AdaBoost and Random Forest methods: A comparative study. *Geocarto Int.* **34**, 1395–1412 (2019).
33. Sidle, R. C. & Bogaard, T. A. Dynamic earth system and ecological controls of rainfall-initiated landslides. *Earth-Sci. Rev.* **159**, 275–291 (2016).

34. Chernozhukov, V. et al. Double/debiased machine learning for treatment and structural parameters. *Econom J.* **21**, C1–C68 (2018).
35. Athey, S. & Imbens, G. W. Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **113**, 7353–7360. (2016).
36. Athey, S., Tibshirani, J. & Wager, S. Generalized random forests. *Ann. Stat.* **47**, 1148–1178 (2019).

Figures

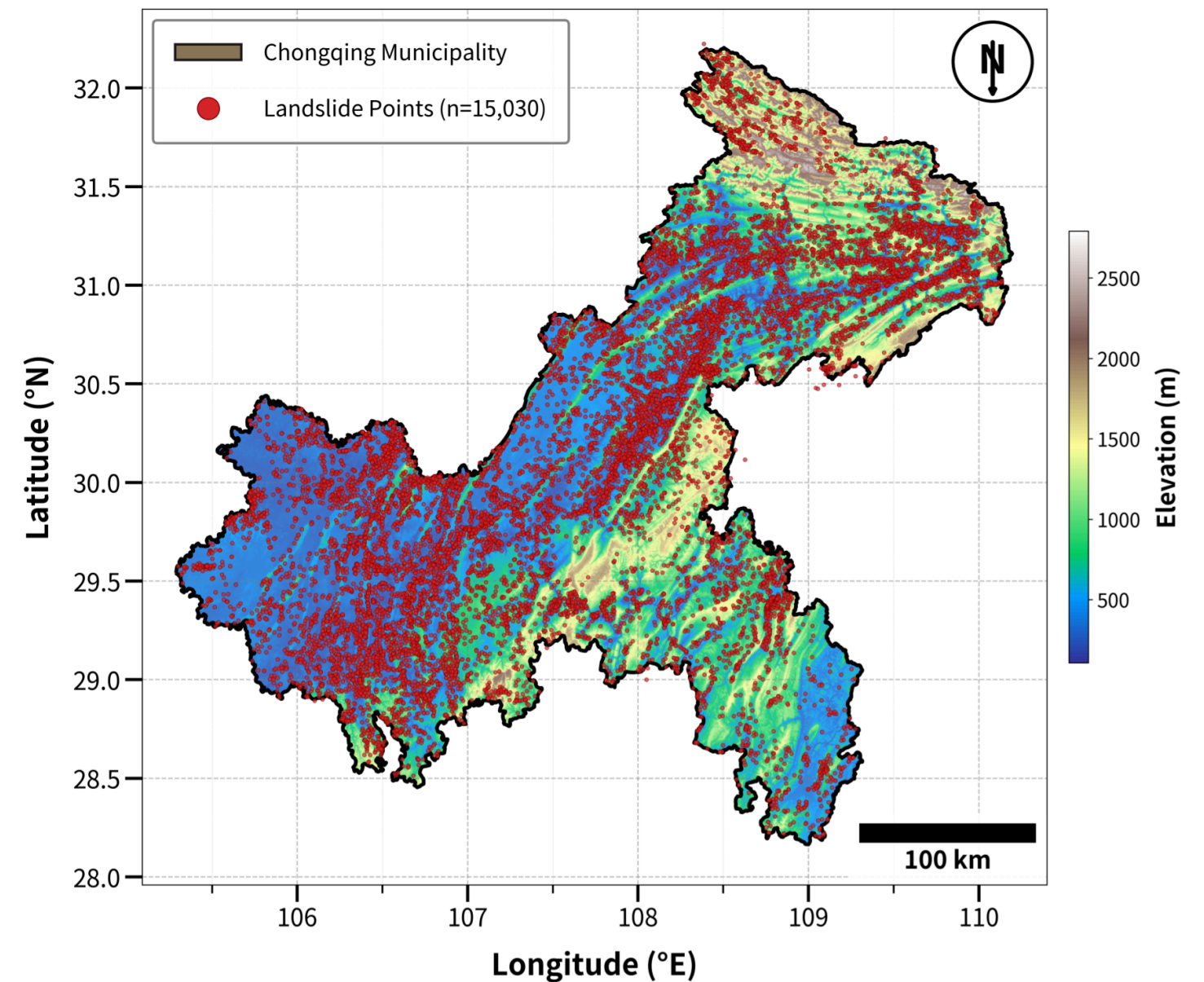


Figure 1

Digital Elevation Model (DEM) and spatial distribution of landslide hazard points in Chongqing. The background is a 30 m resolution DEM, with colors representing elevation; red points are landslide hazard

points, mainly distributed in low-elevation river valley zones.

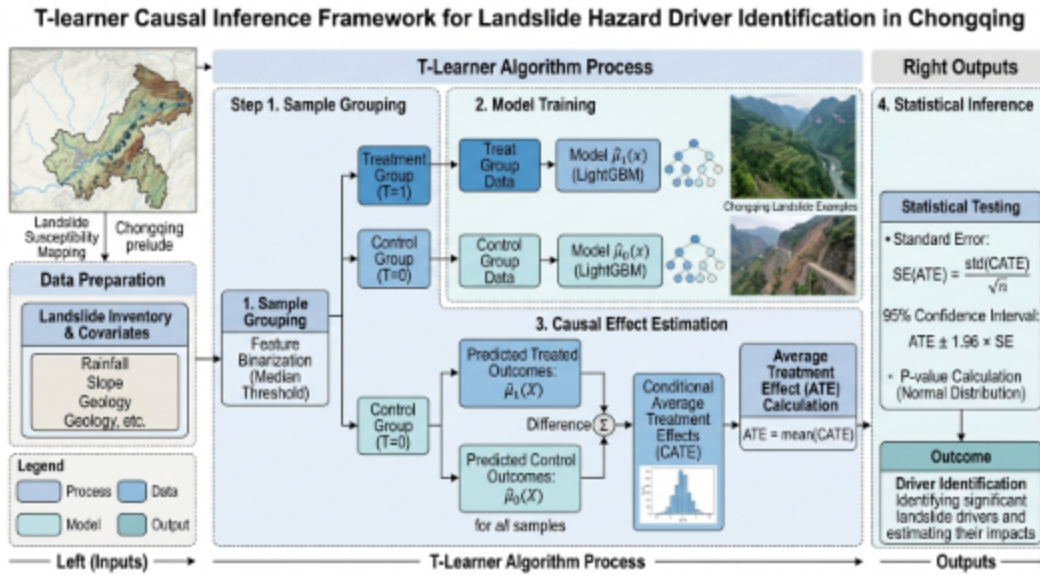


Figure 2

Schematic diagram of the T-learner causal inference framework.

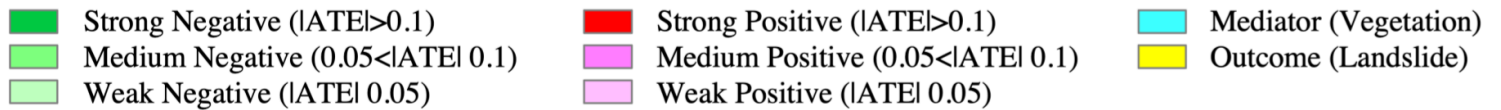
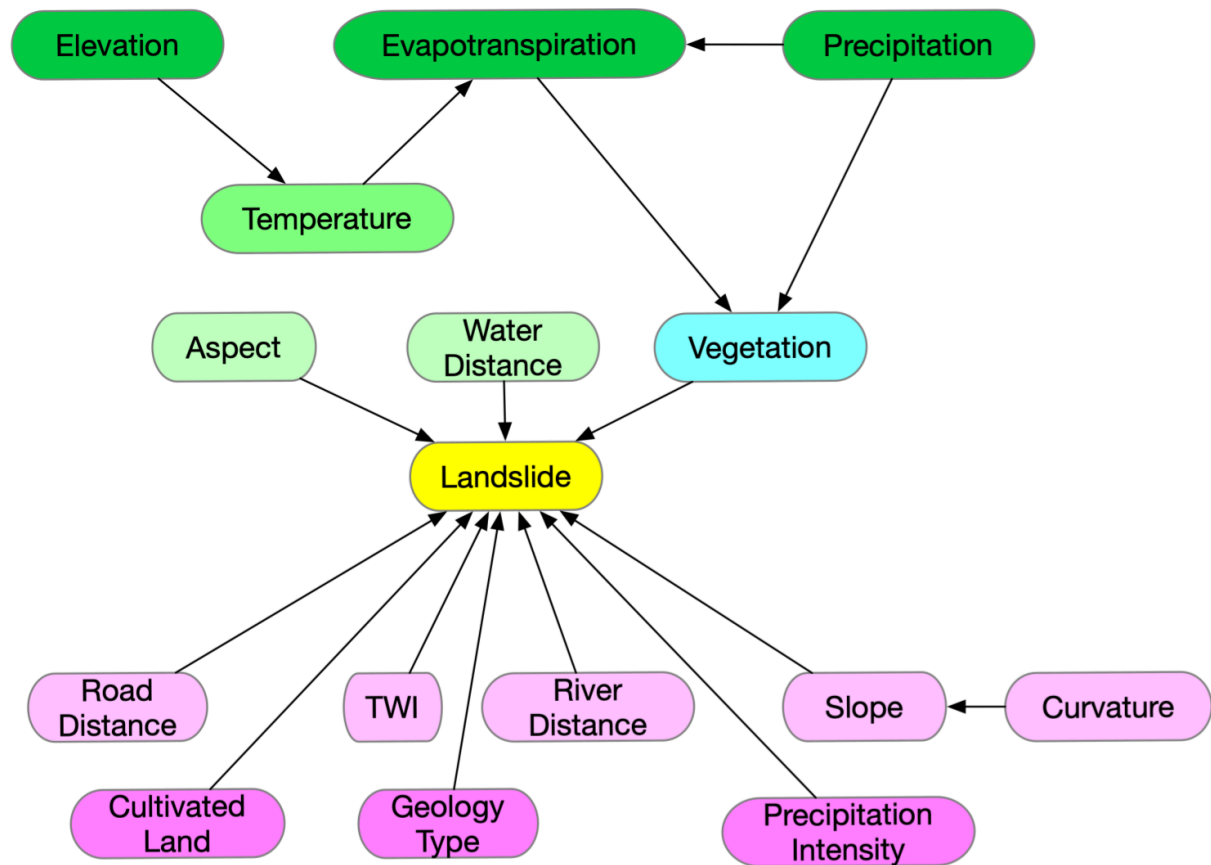


Figure 3

Exploratory structure learning results from the PC algorithm. Edges and directions in the graph depend on assumptions of causal sufficiency, faithfulness, and conditional independence tests, and should be viewed as candidate structures to be tested through temporal information, domain knowledge, and external data, rather than verified causal DAGs.

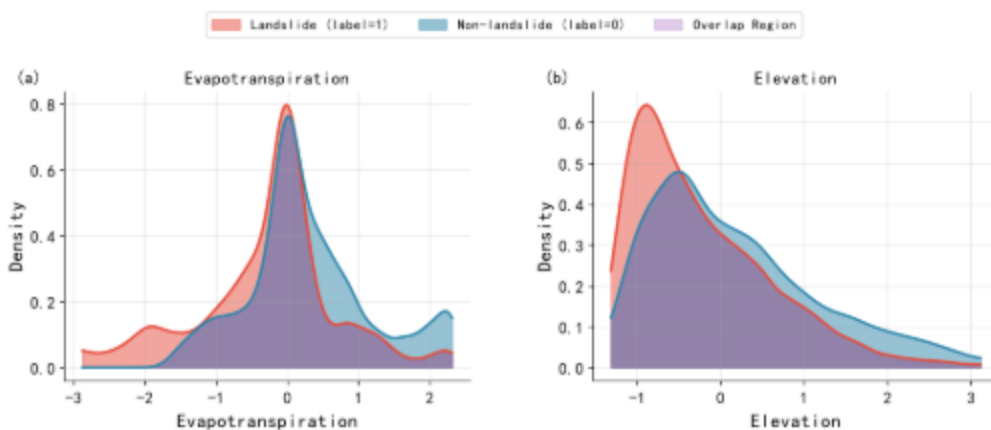


Figure 4

Propensity score distribution verification (evapotranspiration and elevation). The propensity score distributions of the treatment and control groups overlap substantially, satisfying the overlap assumption.

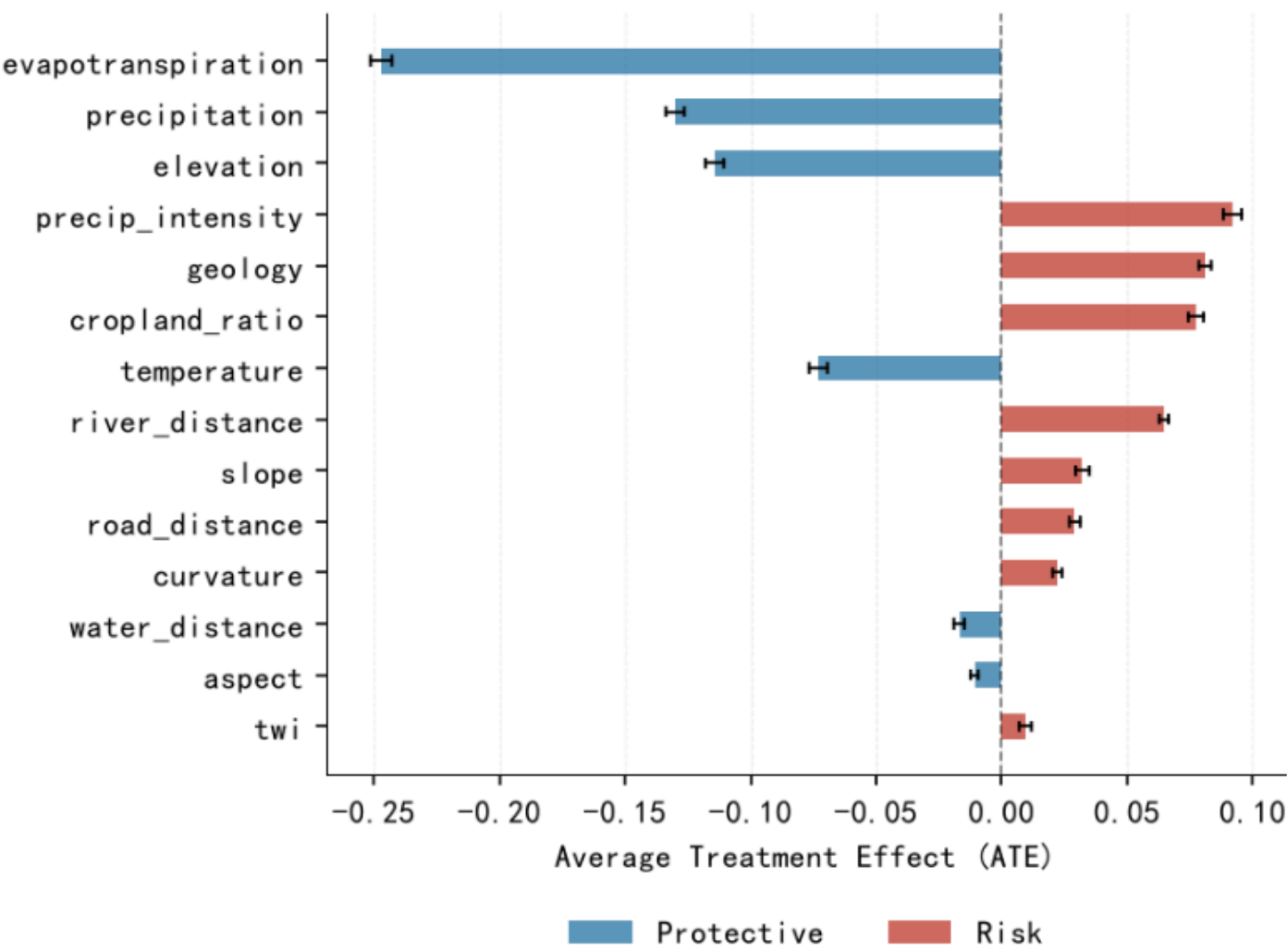


Figure 5

Forest plot of average treatment effects (ATE) for all candidate factors. Error bars represent 95% confidence intervals derived from normal approximation of the empirical CATE distribution.

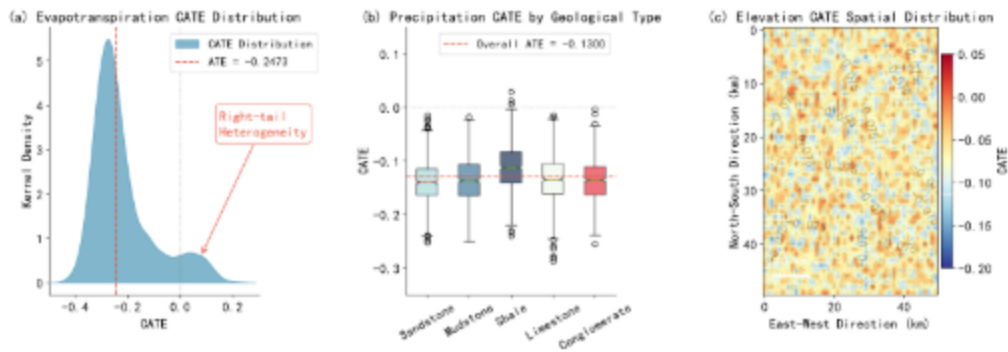


Figure 6

Distribution of conditional average treatment effects (CATE) for selected factors. These distributions provide hypotheses about effect heterogeneity and should not be interpreted as identified individual causal effects.

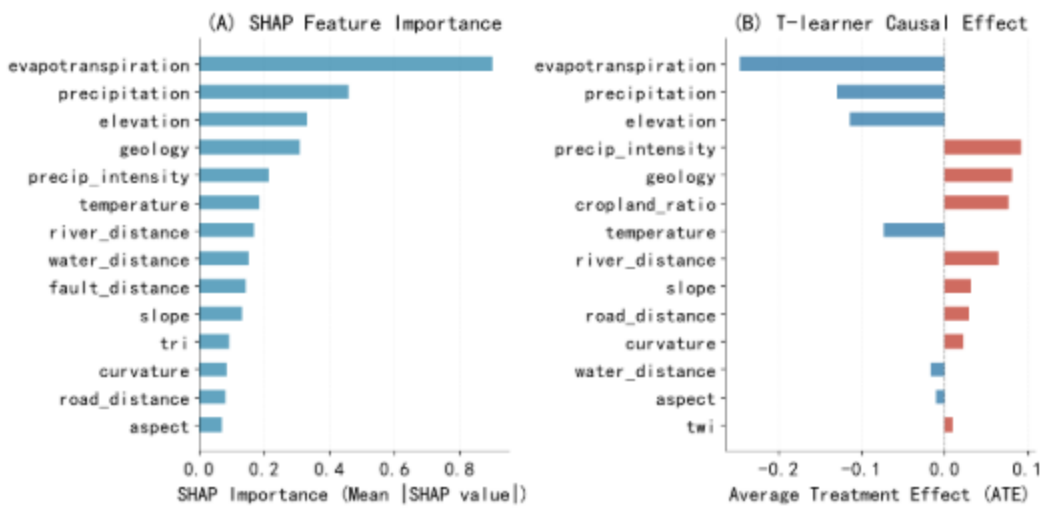


Figure 7

Comparison of SHAP feature importance and T-learner ATE estimates. The two measures address fundamentally different questions: SHAP reflects model attribution, while ATE estimates counterfactual risk differences under explicit treatment definitions.