

Hydrological proxy derived from InSAR coherence in landslide characterization

Yuqi Song^a, Xie Hu^{a,*}, Xuguo Shi^b, Yifei Cui^c, Chao Zhou^b, Yueren Xu^d

^a College of Urban and Environmental Sciences, Peking University, Beijing, China

^b School of Geography and Information Engineering, China University of Geosciences, Wuhan, China

^c Department of Hydraulic Engineering, Tsinghua University, Beijing, China

^d Institute of Earthquake Forecasting, China Earthquake Administration, Beijing, China

ARTICLE INFO

Keywords:

Landslide susceptibility
Hydrological forcing
InSAR coherence

ABSTRACT

Quantifying landslide susceptibility saves lives, especially in populous areas exposed to wet climates. However, available hydrological data sets such as precipitation and soil moisture are usually from reanalysis with a few to tens of kilometers' coarse resolution compared to the dimensions of landslides. Here we aim to seek substitutes to characterize hydrological features with finer spacing for landslide susceptibility assessment encompassing the tectonically active California. We synergize remote sensing big data and derivatives including topographic characteristics, vegetation index, hydrological variables, land cover, and geological units in different machine learning architectures. Our results illuminate that the interferometric coherence derived from synthetic aperture radar (SAR) can be an effective hydrological proxy, providing enhanced resolution by three orders of magnitude to tens of meters and presenting satisfactory performance, with recalls >85 % and AUCs >90 % in our landslide susceptibility models. The consequent spatially continuous landslide susceptibility map further demonstrates the effectiveness of high-resolution SAR products in compensating for limitations in traditional hydrological data sets. The map and our inferred relationship with the mélange and the distance to faults improve our ability in landslide hazard mitigation.

Editor: Jing M. Chen

1. Introduction

Landslides are general natural disasters worldwide and can lead to casualties and economic loss. Statistics show that from 1995 to 2014, a total of 3876 fatal landslide events occurred globally, resulting in 163,658 deaths (Haque et al., 2019). The annual loss from landslides can reach billions of dollars for a single country (Klose, 2015). Clarifying the likelihood of landslide occurrence (i.e., landslide susceptibility) is fundamental in landslide prevention and mitigation.

Landslide susceptibility modeling (LSM) largely relies on current landslide inventories and various geo-environmental information (e.g., morphological, hydrological, and land cover), assuming that mass movements will be more likely to occur under similar conditions resulting in past and present instability (Reichenbach et al., 2018). Landslide susceptibility assessment approaches include qualitative methods such as geomorphologic mapping (Reichenbach et al., 2005)

and landslide inventory analysis (Galli et al., 2008), and quantitative methods such as physically-based methods (Mergili et al., 2012) and the commonly used statistically-based modeling methods. For the target of landslide susceptibility assessment over a large area, statistically-based modeling methods, which aim at constructing the relationship between the likelihood of landslide occurrence and various geo-environmental factors, have been applied in cases from global (Tang et al., 2023), continental (Broeckx et al., 2018), to local scales (Yi et al., 2022). In the era of big data and artificial intelligence, machine learning methods have been intensively utilized in statistically-based damage assessment including landslide susceptibility quantification (Wang et al., 2021), leveraging geo-environmental factors associated with landslide development derived from multi-source remote sensing (RS) and geographic information system (GIS) products (Woodard et al., 2023; Youssef et al., 2023). However, the scales of environmental factor products and landslides are different, and the spatial resolution of each environmental factor may also be inconsistent. For example, precipitation and soil moisture are commonly used in landslide susceptibility

* Corresponding author.

E-mail address: hu.xie@pku.edu.cn (X. Hu).

modeling, but the spatial resolutions of these products are often on the order of kilometers, much greater than the dimensions of most landslides.

In the 1970s, Ulaby et al. conducted microwave experiments and found that the intensity of ground backscattering is sensitive to water content variation, illustrating the possibility of using microwaves to invert soil moisture (Ulaby and Batlivala, 1976). Since then, soil moisture retrieval has been conducted by both passive microwave remote sensing missions, e.g., SMAP (Soil Moisture Active Passive) and AMSR2 (Advanced Microwave Scanning Radiometer 2) (O'Neill et al., 2021; Owe et al., 2008), and active microwave remote sensing missions, e.g., METOP-A ASCAT (Advanced Scatterometer) (Bartalis et al., 2007). The Copernicus Sentinel-1A/B twin-satellite constellation performs C-band synthetic aperture radar (SAR) imaging, providing continuous high-resolution global observations over approximately the last 10 years, which has been frequently applied in soil moisture retrieval studies (e.g., Balenzano et al., 2021; Bauer-Marschallinger et al., 2019). Inspired by soil moisture retrieval experiments based on microwave remote sensing products, we hypothesize that high-resolution SAR-derived features may serve as hydrological proxies in landslide characterization, and test the hypothesis by establishing landslide susceptibility models over the plate-boundary-scale California using machine learning methods and identifying substitutes for hydrological variables from Sentinel-1 products. Landslide-prone California is the third largest and most populous state in the U.S. Rainstorms in this region may trigger fast-moving landslides and accelerate slow-moving landslides. The 1200-km-long San Andreas Fault (SAF) system, separating the Pacific Plate and the North American Plate at a rate of 20–35 mm/yr, encompasses much of California (Southern California Earthquake Data Center, 2013). Coastal cliff erosions along the coastal ranges cause bluff retreat and landslides as well (Hapke and Green, 2006). The United States Geological Survey (USGS) and California Geological Survey (CGS) have reported about 176k landslides in this state (Mirus et al., 2020). Populous counties like Los Angeles, Orange, Santa Clara, and Sonoma host landslides with a considerable total area, resulting in relatively high landslide risks

(Figs. 1b & 1c). Given the vast spatial extent, tectonically active setting, varied land cover, and extensive research on geological hazards, California serves as an ideal natural laboratory for developing data-driven models to illuminate landslide distribution and critical variables in determining landslide occurrences and activities.

This study aims to address the problem of inconsistent spatial scale between landslides and hydrological factors in landslide characterization. We evaluate the effectiveness of SAR-related features as hydrological proxies by substituting conventional low-resolution hydrological products with high-resolution SAR-derived indicators in landslide susceptibility modeling, which allows for a comparative analysis of model performance and susceptibility mapping results. The article is organized as follows: Section 1 reviews the literature on landslide susceptibility assessment and soil moisture retrieval from microwave products, points out the inconsistency issue and potential solution from microwave remote sensing products, and provides an overview of our study area. Section 2 presents the data derived from multi-source remote sensing and GIS products for machine learning. Section 3 describes the data preprocessing, the principles of applied machine learning models, and the design of our comparative experiments. Section 4 shows the main results, including the performance of all the landslide susceptibility models, contributions from geo-environmental factors and Sentinel-1 indicators, and landslide susceptibility maps generated from different feature sets. Section 5 discusses the performance of the Sentinel-1-derived products in representing the hydrological conditions, and other factors affecting the landslide activities in California. Finally, section 6 makes the conclusion.

2. Data

Landslide activities are governed by multiple processes including topography, geology, land cover, catchment, precipitation, and tectonics (e.g., dynamic shaking and aseismic creeping). Remote sensing data are beneficial to extract some of these critical parameters on a regional scale. Drawing on prior studies (e.g., Reichenbach et al., 2018;

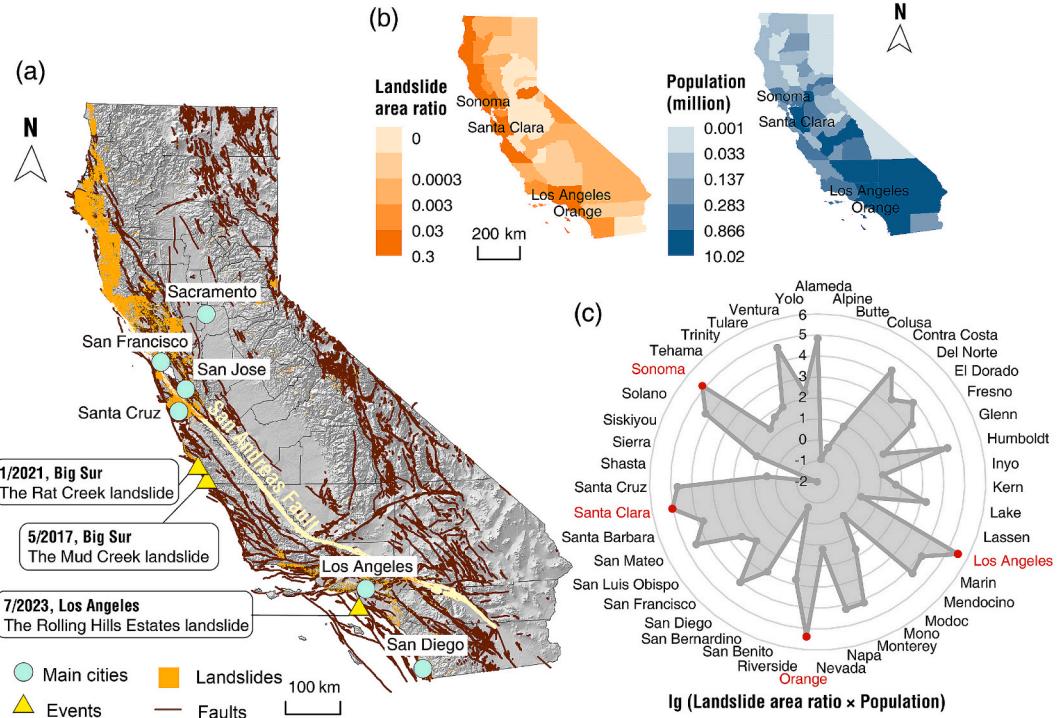


Fig. 1. Landslides, faults, and population distribution in California. (a) The geological settings and recent big landslide events. (b) The ratio between landslide area and entire area, and the population of each county. (c) The landslide risk index of each county with polygon landslide records generated by the landslide area ratio and the population.

Woodard et al., 2023), multiple remote sensing and GIS products were applied to obtain those critical features, including surface elevation, slope, aspect, and topographic wetness index (TWI) derived from SRTM digital elevation models (DEM), the average normalized differential vegetation index (NDVI) in summer derived from Landsat 8 OLI images, the average soil moisture from SMAP level 3 products, accumulated precipitation from GPM (The Global Precipitation Measurement Mission) products, the distance to rivers and faults calculated based on river and fault distribution GIS products (Lehner and Grill, 2013), the geological units and land cover maps compiled by USGS and other relevant departments (Dewitz and U.S. Geological Survey, 2021; Horton, 2017). Landslides were labeled referring to the second version of the landslide inventory compiled and released by USGS in 2022 (Belair et al., 2022; Mirus et al., 2020) (Table 1). It is worth noting that this inventory compiles both historical and recent landslides, recorded by multiple sources including USGS, NASA, local Geological Surveys, and published literature, and only ~20 % of records contain information about when landslides occurred. Therefore, in this study, we focus on the spatial distribution of potential landslides, rather than predicting their timing, which is consistent with the concept of landslide susceptibility (Reichenbach et al., 2018).

Here synthetic aperture radar (SAR) products of Sentinel-1 interferometric coherence and backscatter (amplitude), retrieved from Kellndorfer et al. (2022), are collected as possible hydrological proxies. Since most of California's precipitation occurs in the winter, we selected the median interferometric SAR (InSAR) coherence and mean amplitude images covering December 2019 to February 2020 in this data set, which originally contains derivatives of Sentinel-1 SAR images acquired over one year from December 2019 to November 2020. For comparison, we also sourced the average soil moisture and accumulated precipitation features over the same period.

3. Methods

3.1. Preprocessing

Statistically-based landslide susceptibility models are usually obtained by fitting statistical models or training machine learning models with multi-source geo-environmental factors as features ("independent" variables) and landslides as labels ("dependent" variable) (Reichenbach et al., 2018). In this study, all the original data, including geo-environmental factors, landslide inventory, and SAR products, were rasterized, registered, and resampled at a resolution of 90 m, same as that of the original DEM. The landslide inventory was used to produce sample labels, and other variables constituted sample feature sets.

Continuous features were standard scaled, and the one-hot encoding was applied to the three categorical variables (i.e., aspect, geological unit, and land cover) so that the ranges of values are comparable for all features. Maps of all the prepared variables are shown in Fig. 2.

3.2. Machine learning workflow of initial models

Employing multiple machine learning methods can enhance landslide susceptibility assessments, leveraging algorithmic diversity to address model limitations, data complexity, and uncertainty, while enabling comprehensive validation of SAR-derived features as hydrological proxies. We selected three machine learning models: logistic regression (LR), support vector machine (SVM), and random forest (RF). Logistic regression is a classical binary classification method in machine learning, known for its efficiency and low computational requirements. The method's core is to apply a logistic sigmoid function on a linear function (Bishop, 2006). For a binary classification case (class C_1 and C_2), the posterior probability of class C_1 can be written as the logistic sigmoid $\sigma(a)$ acting on a linear function of the feature vector ϕ :

$$p(C_1|\phi) = y(\phi) = \sigma(\mathbf{w}^\top \phi) = \frac{1}{1 + \exp(-\mathbf{w}^\top \phi)} \quad (1)$$

$$p(C_2|\phi) = 1 - p(C_1|\phi)$$

where \mathbf{w} contains the weight parameter of features, and the model training is exactly the process of determining \mathbf{w} using features ϕ and the binary label (landslide or non-landslide) in the training set. In our study, the landslide susceptibility is given by $p(C_1|\phi)$ and $p(C_2|\phi)$.

For a binary classification problem, a linear model describing the relationship between prediction results and features is in the form of (Bishop, 2006):

$$y(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b \quad (2)$$

where $\phi(\mathbf{x})$ is the feature-space transformation, or kernel function (in this study we only used a linear kernel). The training set \mathbf{x} (including N samples $\mathbf{x}_1, \dots, \mathbf{x}_N$) corresponds to N target values t_1, \dots, t_N where $t_n \in \{-1, 1\}$. The objective of a support vector machine is to find a hyperplane (i.e., decision boundary) in the feature space that maximizes the minimum distance (a.k.a., margin) from each class (Cortes and Vapnik, 1995). The distance of a correctly classified sample point \mathbf{x}_n (which satisfies $t_n y(\mathbf{x}_n) > 0$) to the decision boundary is as follows:

$$\frac{t_n y(\mathbf{x}_n) - t_n (\mathbf{w}^\top \phi(\mathbf{x}_n) + b)}{\|\mathbf{w}\|} \quad (3)$$

Since rescaling \mathbf{w} and b synchronously won't change the distance,

Table 1
Data source and the original spatial resolution of features.

Data type	Feature	Source	Original spatial resolution	Period
Topographic data	Elevation, aspect, curvature, slope, TWI	SRTM DEM	30 m	February 2000
Optical images	NDVI	Landsat 8 OLI	30 m	June to September from 2014 to 2023
Hydrometeorological data	Precipitation	GPM	0.1° (~10 km)	December 2019 to February 2020 (daily accumulation)
	Soil moisture	SMAP	9 km	December 2019 to February 2020 (daily average)
River distribution	Distance to the nearest river	HydroRIVERS Version 1.0 (secondary products of HydroSHEDS v1)	Vector	The core data products of HydroSHEDS v1 were derived primarily from the SRTM DEM (2000)
Fault distribution	Distance to the nearest fault	Quaternary Fault and Fold Database of the United States	Vector	–
Land cover	Land cover categories	NLCD Land Cover (2019)	30 m	2019
Geological unit	Geological unit categories	SGMC Geodatabase of the Conterminous United States	Vector	–
SAR Products	Amplitude and coherence	Global seasonal Sentinel-1 interferometric coherence and backscatter data set by Kellndorfer et al. (2022)	90 m	December 2019 to February 2020
Landslide inventory	Landslide distribution	USGS Landslide Inventories across the United States (Version 2)	Vector	–

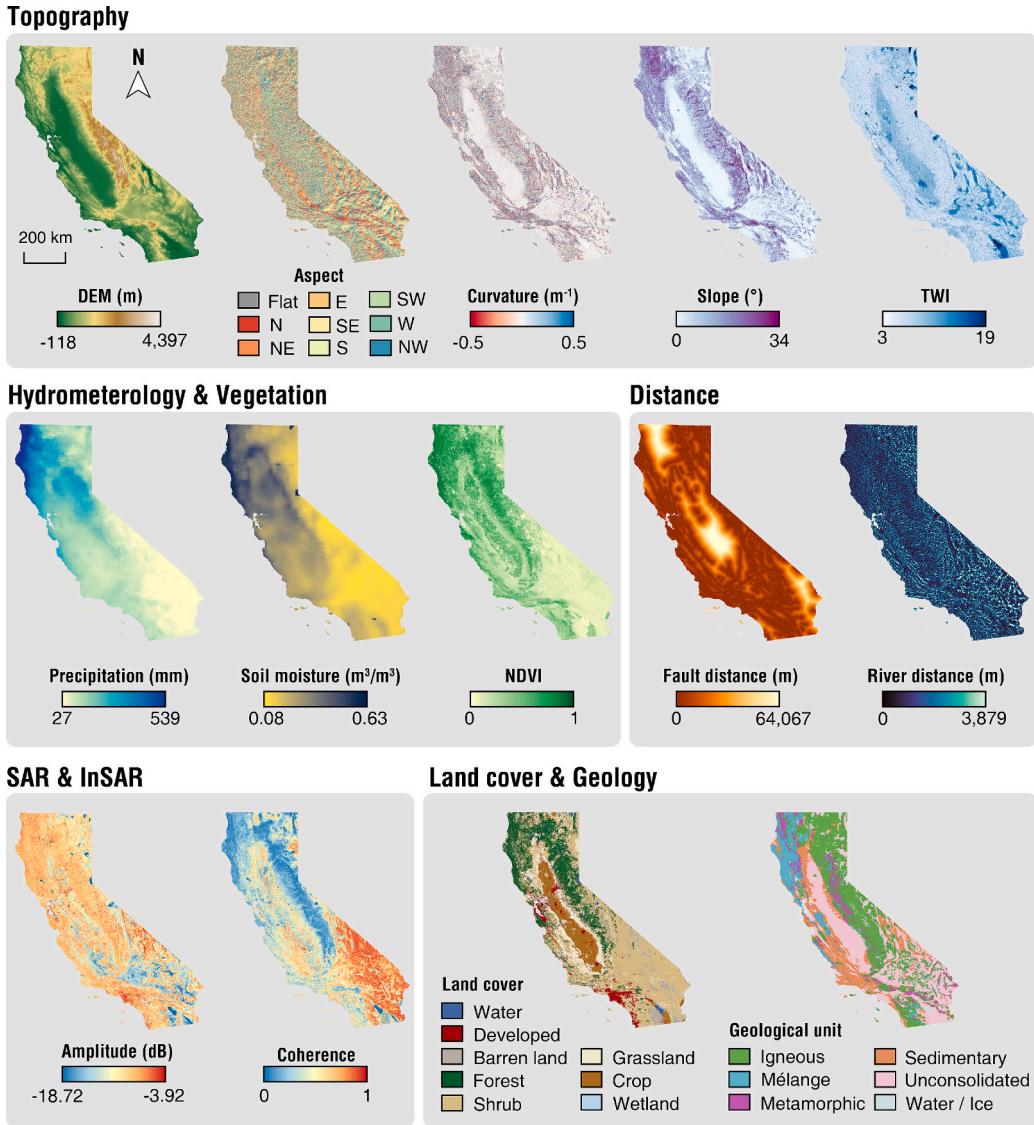


Fig. 2. Maps of multi-source geo-environmental variables and SAR products.

and given the constraint of $t_n y(\mathbf{x}_n) \geq 1$ (because the distances between the decision boundary and samples are equal or greater than the margin), the task of finding the best boundary is transformed into an optimization problem. To avoid overfitting, we allow some training data to be misclassified but with a certain penalty. Therefore, a slack variable $\xi_n \geq 0$ is introduced, and the optimization problem becomes:

$$\text{argmin} C \sum_{n=1}^N \xi_n + \frac{1}{2} \|\mathbf{w}\|^2 \quad (4)$$

where the parameter $C > 0$ is the trade-off controlling between penalty and margin. And the constraint becomes $t_n y(\mathbf{x}_n) \geq 1 - \xi_n$.

Random forest relies on the ensemble learning technique that integrates decision trees (DT) through bagging or bootstrap aggregating to achieve better performance than a single model (Breiman, 2001; James et al., 2021). Bagging or bootstrap aggregation is a sampling strategy that repeatedly resamples with replacement from the original data, so each classifier in the ensemble is trained using a subset of the whole input data set, which can effectively reduce the variance and improve the robustness of the model (James et al., 2021). A DT model consists of several nodes and directed edges, and a parent node is connected to child nodes by these edges. When applying a DT for classification, the

top and internal nodes yield a split based on a certain feature, and terminal nodes at the bottom represent classification results (James et al., 2021).

To build an RF classifier, first, the bootstrap resampling method is applied to the original data set to create B different training data sets. Then B different DT models are trained separately using those subsets. The final classification decision is made by taking a majority vote on all the B predictions. In addition, to decorrelate those trees and reduce variance, RF does not consider all the available features at each split, only a random sample of m features is chosen as split candidates from the full set of p features, and usually $m = \sqrt{p}$ (Hastie et al., 2009).

We initially attempted to develop landslide susceptibility models using all the conventional geo-environmental variables (i.e., no SAR products were included in the initial feature set). To avoid data skew and improve model performance (Zhou et al., 2018), the ratio of the number of positive (landslide) and negative (non-landslide) samples was set to be about 1:1 (~ 814 k landslide and non-landslide pixels, respectively). In order to obtain negative samples as representative as possible, the non-landslide pixels were randomly sampled in equal proportion from each land cover type in non-landslide areas. The data set was split into training and test sets in a 7:3 ratio, each with the same portion of positive and negative samples. After training, the model performance was

evaluated using the test set by metrics including accuracy, precision, recall, F_1 score and receiver operating characteristics curve and area under the curve (ROC curve and AUC), and the contributions of different features to the model output (i.e., feature importance) were also assessed. All the models were trained and tested using *sklearn* tool in Python, with the parameter “random_state” (which controls the randomization in *sklearn* algorithm) setting as 0. And to accelerate the process on our large data set, we applied *thundersvm* library (Wen et al., 2018) to enable GPU computing in SVM model training and prediction.

3.3. Feature importance

The weight w of features in the LR model can represent the importance of each feature (Eq. 1), with a larger absolute value implying a greater contribution of the corresponding feature. Similarly, in the linear SVM model, the w in Eq. (2) also assesses feature contribution, where larger absolute values indicate greater importance (Chang and Lin, 2008).

An RF model is an ensemble of several DTs. For each DT, the criterion determining the split of each internal node is called Gini impurity. In the classification, supposing there are K classes in total, for node m (i.e., feature m), let p_{mk} be the proportion of class k observations in this node, then Gini impurity is calculated by (Hastie et al., 2009):

$$Gini(p_{mk}) = \sum_{k=1}^K p_{mk}(1 - p_{mk}) \quad (5)$$

For a binary classification problem, if p is the proportion in the second class, then Gini impurity can be written as:

$$Gini(p) = 2p(1 - p) \quad (6)$$

The contribution or importance of a feature can be represented by the variation of Gini impurity after the node is split into two descendant nodes (a.k.a., Gini decrease or Gini gain):

$$VIM_m = Gini_m - (Gini_l + Gini_r) \quad (7)$$

in which $Gini_m$ is the Gini impurity at the node to be split, $Gini_l$ and $Gini_r$

are the Gini impurity at the descendant left and right nodes, respectively. VIM_m represents impurity decline caused by the split strategy. A large impurity decrease implies an appropriate selection of the feature and corresponding threshold, as well as a great contribution of this feature. In the RF model, Gini importance of a certain feature is the average “Gini decrease” of the same feature in all the “trees” (James et al., 2021). Feature importance evaluation results of the three initial models are shown in Fig. 3. Since all the geo-environmental factors were included, the feature importance rankings from initial LR and SVM models can be influenced by feature collinearity, while the RF model does not consider all the available features at each split, reducing the impact of collinearity. However, the initial rankings can still assist in preliminarily determining variable contributions.

3.4. Feature selection and combination for simplified models

In the initial training, we fed all the geo-environmental variables into three categories of machine learning models, tested their performance, and ranked the feature importance. Thereafter, to test the substitution of SAR products for conventional hydrological products, as well as avoid inaccurate evaluation of feature importance due to feature redundancy and collinearity, we conducted feature selection based on Pearson correlation coefficients and principal component analysis (PCA), while considering feature importance rankings from initial models to ensure the representativeness. In terms of Pearson correlation, soil moisture, precipitation, and NDVI are highly positively correlated, while InSAR coherence is negatively correlated with these factors (Fig. 4a). For continuous features, we considered the PCA component matrix and selected one or two original variables with high loading values (i.e., correlation coefficients between each original feature and principal component) in the same principal component (Fig. 4b). The first principal component mainly relates to hydrological conditions, indicated by high loading values of soil moisture, precipitation, and NDVI. InSAR coherence has a significant negative correlation with this component, further indicating its potential as a hydrological proxy. In the second principal component, DEM, TWI, and slope present higher loading values, implying a correlation with the topography. The third to sixth

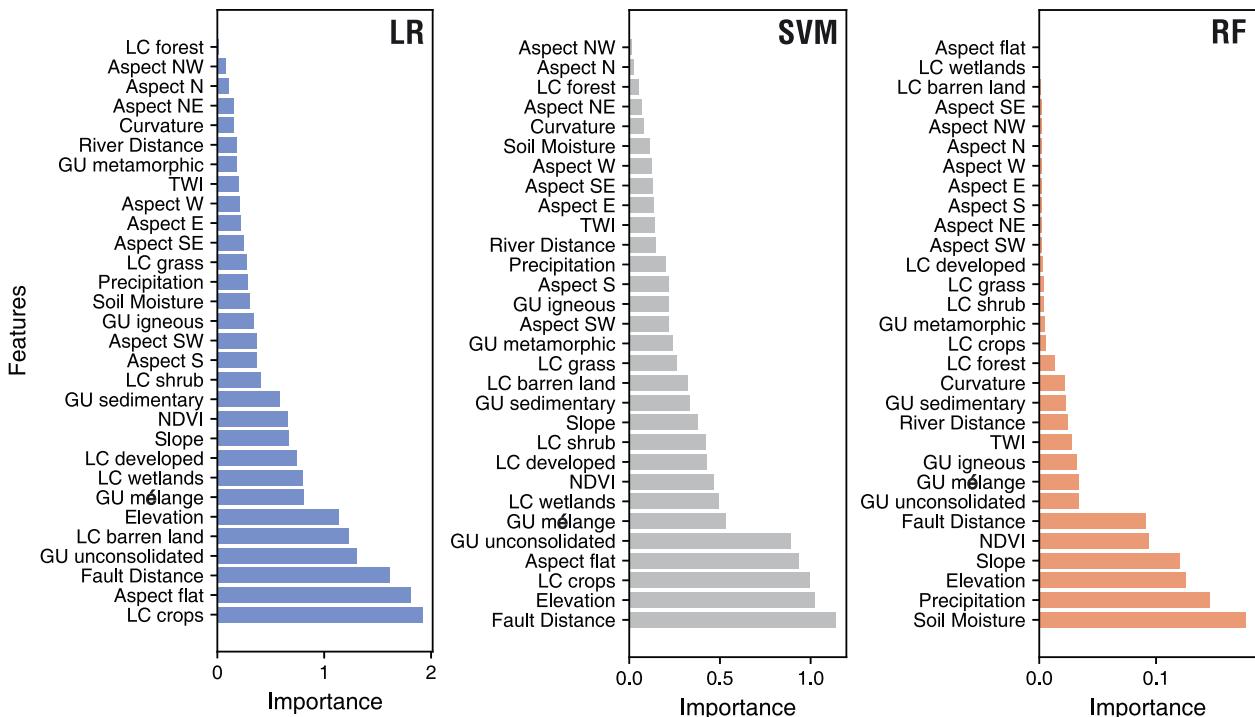


Fig. 3. Feature importance evaluations of initial models with all geo-environmental features.

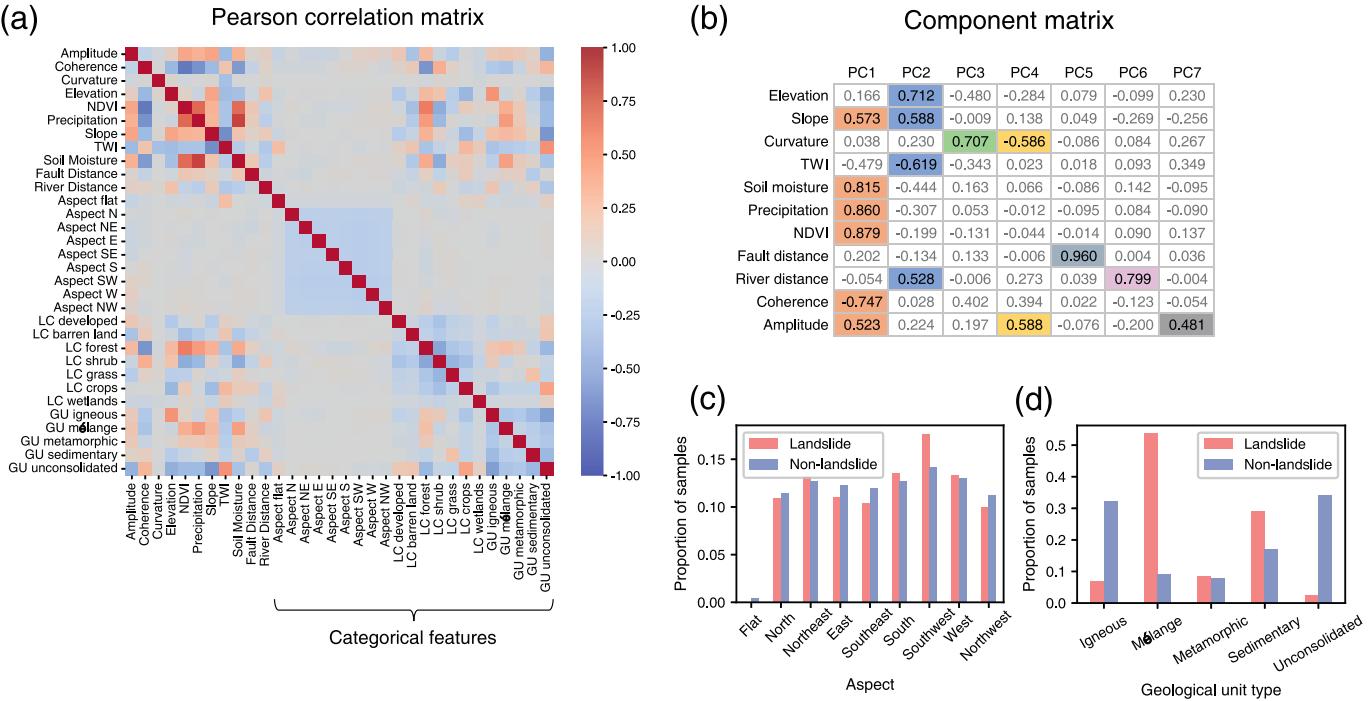


Fig. 4. Relationship between prepared features indicated by (a) Pearson correlation ('LC': land cover, 'GU': geological unit). (b) PCA component matrix. Loading values with the largest absolute values and those larger than 0.5 are highlighted. (c) Surface aspect and (d) geological unit distribution in landslide and non-landslide samples.

components capture surface curvature, fault distance, and river proximity. The cumulative initial eigenvalues of the first seven principal components account for more than 90 %, so we only kept the first seven columns of the component matrix. For categorical features, among different surface aspects, southwestern and northeastern directions were kept because they are predominant in landslide samples (Fig. 4c). Among the geological units, categories of igneous and mélange unique to California were selected for their relationship with landslides (Fig. 4d, see also Xu et al., 2021). Land cover types were all dropped since they have been applied in sampling negative samples, and some of them show correlation with existing features (e.g., forest land cover relates to hydrological features, and crop land cover is correlated with slope since fields are concentrated in the central valley in California).

Combined with geo-environmental feature importance revealed by initial models (Fig. 3), in addition to hydrological variables (i.e., precipitation or soil moisture) or SAR products (i.e., coherence or amplitude), we only considered elevation, slope, curvature, fault distance, and river distance as the continuous features, and southwestern aspect, northeastern aspect, igneous and mélange geological units as categorical variables in our simplified models. To assess the improvement of landslide susceptibility assessment by replacing coarse-resolution hydrological features with SAR-related features, we compared five feature combination strategies: i) using no hydrological or SAR features; ii) using precipitation and iii) soil moisture as hydrological features; iv) using InSAR coherence and v) InSAR amplitude as hydrological proxies. To be more inclusive and avoid bias in the sampling, negative samples were repeatedly sampled 100 times for each feature combination (by setting the random_state parameter in the sample function from 0 to 99, respectively), resulting in 100 different data sets with a 1:1 ratio of positive and negative samples. For each feature combination and each machine learning method, 100 different models were obtained from the 100 data sets. All the simplified models were trained and tested with the same train test split ratio as the three initial models (i.e., 7:3), with the random_state parameter setting from 0 to 99. The workflow of our experiments is illustrated in Fig. 5.

4. Results

The performance of landslide susceptibility models (measured by metrics including recall, AUC, etc.), along with mapping results obtained from different feature sets, illustrates how SAR-derived indicators can characterize hydrological conditions favorable for landslides from both quantitative and qualitative perspectives.

4.1. Landslide susceptibility model performance and feature importance

The metrics of three initial models and the average metrics of 15 types of simplified models are shown in Table 2 and Fig. 6. For each type of machine learning method and feature set, the metrics of simplified models were calculated by averaging the metrics of 100 different models trained by 100 different sample sets. Recalls of our initial landslide susceptibility models, which include all the collected geo-environmental features, are higher than 90 %, suggesting a low omission of true landslide samples. All the metrics of the RF model are evidently better than those of the other two machine learning methods, while the performance of LR and SVM models is comparable. The SVM model shows slightly higher recalls than that of LR (Table 2a).

After feature selection, the simplified models show reduced metrics but still retain recalls higher than 85 % and AUC higher than 90 % (Table 2b). For these three categories of machine learning models, the performance of landslide susceptibility models using precipitation or soil moisture as hydrological features is very close (see the rows of (ii) Precipitation and (iii) SoilM in Table 2b). When InSAR coherence is applied to replace conventional hydrological features, the metrics of LR and SVM models are almost unchanged, but the performance of the RF model drops slightly (see the rows of (iv) Coherence in Table 2b). The slight decline might be attributed to the issues of obvious quadrangle boundaries in the landslide distribution map caused by diverse sources of this compiled inventory (Mirus et al., 2020). These unnatural spatial boundaries coincidentally align with the coarse-resolution hydrological products in space (e.g., Fig. 10a), resulting in slightly better performance of models using hydrological products than that using fine-resolution

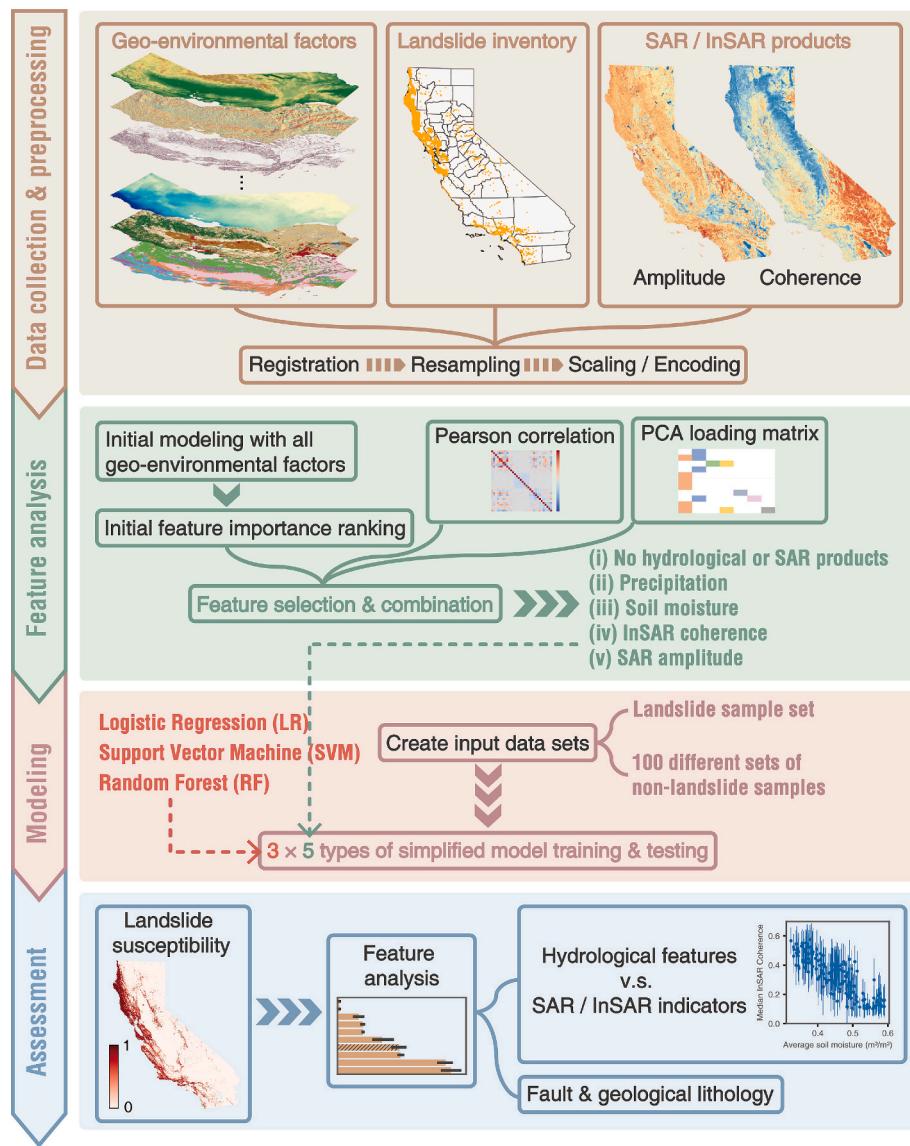


Fig. 5. Workflow on deriving the hydrological proxy from SAR / InSAR indicators in landslide characterization.

InSAR products. Nevertheless, the employment of InSAR coherence can improve performance over models without any hydrological or SAR features (see the rows of (i) w/o set in Table 2b). In contrast, adding SAR amplitude does not significantly enhance the model's ability to distinguish between landslides and non-landslides (see the rows of (v) Amplitude in Table 2b).

The absolute value of coefficients in LR and SVM models and Gini impurity in RF models were used to sort out the feature importance. Due to potential feature collinearity, the feature importance rankings of the initial models may be inaccurate. However, they can still generally show that hydrological features, fault distribution, and some topographic variables (e.g., elevation and slope) significantly contribute to the landslide susceptibility model, and mélange geological unit and southwestern aspect are of greater importance than other categories (Fig. 3).

After feature selection, precipitation, soil moisture, and InSAR coherence are equally and highly ranked in feature importance of their respective LR and SVM models. In RF models, while precipitation or soil moisture ranks first, the importance of topography and fault distance features comes to the fore when applying InSAR coherence as the hydrological feature proxy (Fig. 7).

4.2. Landslide susceptibility maps over California

Given the better performance of RF models compared to the other two categories of models, we obtained the landslide susceptibility maps over California using RF models with precipitation, soil moisture, and InSAR coherence features, respectively (Fig. 8). The model applied is the one closest to the median AUC for each feature combination. The maps made by models with precipitation and soil moisture show significant spatial discontinuities, as those unnatural “blocks” and abrupt jumps in zones A, B, and C in Fig. 8a and b, resulting from coarse resolution of precipitation and soil moisture products as well as the quadrangle boundaries in the original landslide inventory (Mirus et al., 2020). Replacing coarse-resolution hydrological products with high-resolution SAR products can eliminate those discontinuities, as shown in Fig. 8c.

To facilitate the comparison of landslide susceptibility results and reduce the impact of minor variations in the data (Barman and Das, 2024), we also conducted natural breaks (Jenks) classification on the three susceptibility maps, dividing the landslide susceptibility into five levels: very low, low, moderate, high, and very high. In the map generated from the RF model using precipitation (Fig. 8a), 80.21 %, 6.75 %, 4.24 %, 3.82 %, and 4.99 % of the study area fall within very low, low, moderate, high, and very high landslide susceptibility levels.

Table 2

(a) Performance of initial models, and (b) average performance of simplified models. The “Feature set” column in Table (b) corresponds to feature combination strategies in Section 3.4.

(a) Initial Models						
Model	Accuracy	Precision	Recall	F ₁ score	AUC	
LR	0.883	0.863	0.910	0.886	0.942	
SVM	0.888	0.863	0.923	0.892	0.942	
RF	0.962	0.941	0.987	0.963	0.988	

(b) Simplified models						
Model	Feature set	Accuracy	Precision	Recall	F ₁ score	AUC
LR	(i) w/o	0.857	0.850	0.867	0.858	0.917
	(ii) Precipitation	0.874	0.866	0.885	0.876	0.933
	(iii) SoilM	0.867	0.858	0.879	0.868	0.931
	(iv) Coherence	0.871	0.860	0.886	0.873	0.932
	(v) Amplitude	0.856	0.844	0.873	0.858	0.918
SVM	(i) w/o	0.857	0.849	0.867	0.858	0.917
	(ii) Precipitation	0.876	0.867	0.887	0.877	0.933
	(iii) SoilM	0.868	0.859	0.881	0.870	0.931
	(iv) Coherence	0.872	0.861	0.887	0.874	0.932
	(v) Amplitude	0.857	0.845	0.874	0.859	0.918
RF	(i) w/o	0.904	0.872	0.946	0.908	0.954
	(ii) Precipitation	0.952	0.926	0.983	0.954	0.985
	(iii) SoilM	0.952	0.930	0.979	0.954	0.985
	(iv) Coherence	0.920	0.891	0.957	0.923	0.965
	(v) Amplitude	0.907	0.876	0.948	0.910	0.957

For susceptibility assessed by the RF model with soil moisture (Fig. 8b), the proportions in the five levels are 78.89 %, 8.13 %, 4.67 %, 3.55 %, and 4.77 %, respectively. For the RF model with InSAR coherence (Fig. 8c), the corresponding proportions are 73.62 %, 8.82 %, 5.19 %, 5.26 %, and 7.11 %, respectively. The similar susceptibility distributions across different maps indicate the rationality of using InSAR coherence as a hydrological proxy.

5. Discussion

The landslide susceptibility model is essential for assessing disaster risk, which supports land use planning, infrastructure development, and emergency preparedness efforts. We explore the impact of precipitation and soil moisture on landslide susceptibility and how SAR features (especially InSAR coherence) characterize the hydrological context. Other factors affecting landslide occurrence, including fault distribution and geological units, are also discussed.

5.1. Precipitation and soil moisture

Hydrological features rank top among contributions of geo-environmental factors (Fig. 7). Landslide areas generally experience higher accumulated precipitation and average soil moisture than non-landslide areas (Fig. 9a). While the susceptibility to landslides increases with higher precipitation or soil moisture, they can also develop in drier areas (Fig. 9b).

Landslides may occur when shear stress on the shear plane exceeds the resisting strength (shear strength) (Lacroix et al., 2020b). Intense rainfall can increase water content in the sliding materials and the gravity of the landslide masses, thereby increasing the shear stress. It can also elevate pore water pressure, reducing the effective normal stress and cohesion of materials, which lowers shear strength. Together, these processes result in landslide instability (Handwerger et al., 2019; Lu and Kim, 2021). Precipitation significantly affects soil moisture, influencing landslide occurrences, which are also controlled by soil properties, vegetation, and human activities. For example, Lacroix et al. (2020a)

found that irrigation in southern Peru affected soil erosion and initiated slow-moving landslides. In addition, the seasonal landslide movement is often hydrologically driven, necessitating the establishment of thresholds to estimate the timing of landsliding based on rainfall and soil moisture (Zhou et al., 2022). For example, Hu et al. (2018) reported that the Cascade landslide complex in Washington showed initial subsidence when the cumulative precipitation reached 140 mm in 30 days, followed by more pronounced downslope movement when the precipitation accumulated by ~300 mm thereafter.

5.2. Implication of SAR amplitude and coherence to hydrological conditions

We aim to seek alternative hydrological variables when they are not available or at insufficient spatial resolution in landslide susceptibility assessments. We opted for SAR amplitude and coherence based on the intrinsic implication of SAR signals to hydrological conditions. Soil comprises particles, air, and water (both free water and bound water). The dielectric constant (ϵ) of soil is closely related to soil water content, generally increasing as water content rises due to the large contrast between liquid water ($\epsilon \approx 80$) and dry soil (usually $\epsilon < 10$) (Hallikainen et al., 1985). Given the relationship between dielectric constant and radar backscattering, it is feasible to characterize the soil moisture using SAR metrics, e.g., SAR amplitude (intensity of backscattering coefficient) and coherence (interferometric correlation) (Barrett et al., 2009).

A large number of model-based soil moisture retrieval approaches rely on SAR amplitude, including theoretical, empirical, and semi-empirical models. The integral equation model (IEM) is a widely used physically based radiative transfer model (Fung et al., 1992), which quantifies the backscattering coefficient as a function of the unknown soil moisture and surface roughness and the known radar configuration. Semi-empirical scattering models are usually site-independent and can also achieve satisfactory accuracy across various scenarios. The water cloud model (Attema and Ulaby, 1978) can represent backscatter from both vegetation canopy and underlying soil.

Regarding InSAR coherence, De Zan et al. (2014) first built a model to relate quantitatively SAR interferometric observables to soil moisture. Based on the case of plane wave oblique incidence on a lossy dielectric, coherence expression is given by:

$$\gamma = \frac{2j\sqrt{Im(k'_{z2})Im(k'_{z1})}}{k'_{z2}^* - k'_{z1}} \quad (8)$$

where k'_{z1} and k'_{z2} are wavenumber in the vertical direction of the two SAR acquisitions, which are a function of dielectric constant:

$$k'_z(\epsilon') = \sqrt{\omega^2 \epsilon' \mu - k_x^2} \quad (9)$$

where ω is the angular frequency of the SAR signal, μ is the magnetic conductivity, and k_x is the wavenumber along air-soil discontinuity interface. Variations in soil moisture influence the dielectric constant (ϵ'), which in turn affects wavenumber (k'_z) and thus InSAR coherence (γ).

SAR-derived features have been effectively applied in generating soil moisture proxies regarding surface disturbances such as precipitation and tropical cyclones. Jordan et al. (2020) extracted decay timescales of soil moisture variations after precipitation events from Sentinel-1 InSAR coherence time series, showing that InSAR products enable high-resolution assessments of soil moisture variations over large areas, complementing sparse site observations. Similarly, Bürgi and Lohman (2021) derived soil moisture proxies from InSAR coherence after two cyclones. Compared to other coarse-resolution soil moisture products, high-resolution InSAR coherence data complement the vertical distribution of soil moisture, and accurately capture places most affected by storms in areas with great spatial heterogeneity. Generally, lower InSAR

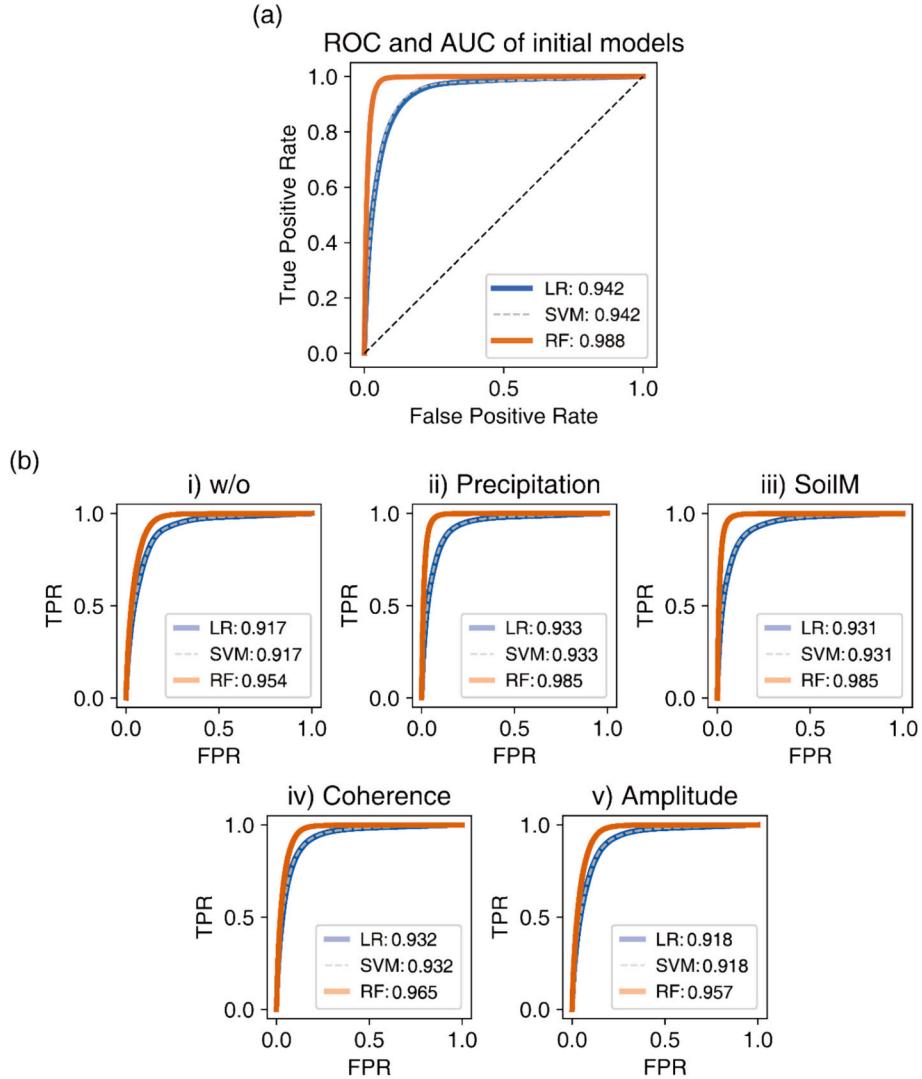


Fig. 6. ROC curves and corresponding AUC values (marked in the legend) for (a) the initial three models and (b) simplified models. In panel (b) all the 100 ROC curves for each model and feature set are plotted using lines with certain transparency, and the AUC values in the legend are the average values of corresponding models.

coherence indicates greater precipitation or higher soil moisture variation over the targeted ground. In Pearson correlation results, InSAR coherence is negatively correlated with precipitation and soil moisture. In PCA results, the three features all get high loading values in the first principal component (The sign of loading value of InSAR coherence is opposite to that of hydrologic features, indicating the negative correlation) (Figs. 4a & 4b). We can intuitively compare the spatial scales and correlations of the two variables by examining Zone B in Fig. 8. The landslide inventory is superimposed on the soil moisture map, and pixel boundaries of SMAP soil moisture products are outlined on the InSAR coherence map, suggesting that the spatial resolution of soil moisture products may be too coarse in landslide-related studies, while that of InSAR coherence is much more consistent with landslide scales (Fig. 10a). The values of soil moisture and average InSAR coherence in each soil moisture pixel reveal their negative correlation (Fig. 10b).

InSAR coherence values are usually lower in landslide areas than in non-landslide areas (Fig. 9a), indicating generally higher landslide susceptibility (Fig. 9b). However, Sentinel-1 satellites operate on C-band while SMAP soil moisture products are generated from L-band radiometer, which might cause inconsistency due to different interactions between ground targets and the two different microwave bands. Although SAR backscattering is not strictly linear with soil moisture and may be

influenced by factors like soil structure and environmental aridity (Ullmann et al., 2023), it still reflects the water content of the ground. In applications like landslide susceptibility assessment, the primary objective is not to retrieve accurate soil moisture or precipitation from InSAR coherence; however, applying high-resolution SAR products as a proxy for coarse-resolution hydrological products can indeed improve the quality of landslide susceptibility mapping (Fig. 8). In addition, here the hydrological and SAR data were acquired from December 2019 to February 2020, during which no major wildfires or earthquakes occurred in California, mitigating non-hydrological influences on SAR features.

Current global data sets of precipitation and soil moisture are available, yet these reanalysis or interpolated products are often low in spatial resolution (e.g., the finest grid of SMAP soil moisture product is only 3 km), limiting their effectiveness in local study sites and small-scale study objects. Substituting hydrological features with InSAR coherence generated from open-access images like Sentinel-1 can improve the resolution to tens of meters worldwide, providing more precise information about ground properties. Furthermore, with the future availability of NISAR (NASA-ISRO SAR) data with continuous temporal and wide spatial coverage, influences on SAR signals caused by factors like dense vegetation can be further eliminated due to its longer

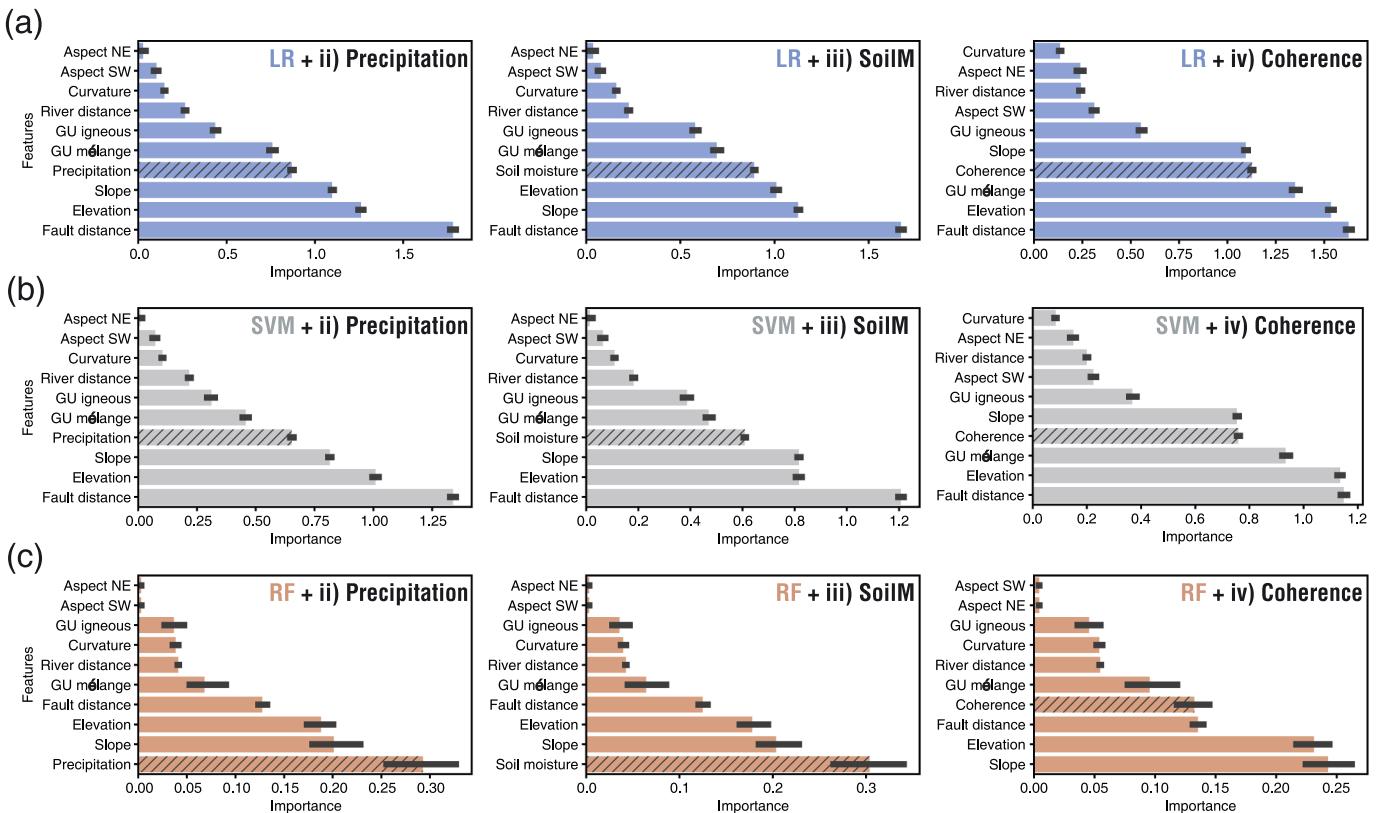


Fig. 7. Feature importance of simplified (a) LR, (b) SVM, and (c) RF models. The colored bars represent the median feature importance calculated from 100 models of each type. The thinner gray bars at the end of the colored bars represent the range of feature importance among the 100 models. The bars of hydrological and InSAR coherence features are filled with diagonal patterns.

operating wavelength (S-band and L-band), enabling more accurate characterization of surface hydrological conditions.

5.3. Faults

In addition to hydrological features, faults also contribute to landslide development. An enlarged view of San Jose, where the San Andreas Fault, Hayward Fault, and Calaveras Fault cut through, shows a correlation between landslides and fault distribution. Hydrological conditions, InSAR coherence, and geological units of this area are shown in Fig. 11.

A 50,000-m-long and 5000-m-wide buffer across the three faults was established and divided into 50 units (Fig. 12). The cross-section profiles in Fig. 12 show the predicted landslide susceptibility from RF models with soil moisture and InSAR coherence feature set of median AUC, respectively, as well as the number of landslides recorded in the original inventory in each unit. Landslides generally cluster along these faults, and the predicted landslide susceptibility results show the same trend (Fig. 12c). San Andreas Fault represents the transform boundary between the North American and the Pacific tectonic plates with tectonic strain accumulation (Hu et al., 2021). Hayward and Calaveras Fault, as two strike-slip faults of the SAF system, show potential for earthquake ruptures leading to $M > 7$ events, posing high landslide risks in the urbanized environment (Chaussard et al., 2015).

Distribution histograms of the distance to the nearest fault between both landslide and non-landslide samples throughout California reveal that landslides are characterized by a more pronounced distribution close to faults, and about 2/3 of landslide samples are distributed within 5 km from a fault (Fig. 13). Tectonic activities play an important role in the formation and development of landslides. The spatial clustering of landslides during earthquakes is usually determined by the seismic magnitude and spacing between epicenters or ruptured faults, as well as

the fault type and slip rate (Huang and Fan, 2013). The 2002 Denali Fault earthquake in Alaska resulted in thousands of coseismic landslides concentrated in a narrow band of ~ 15 km on both sides of the rupture zone (Jibson et al., 2006). The 2008 Wenchuan earthquake nucleated on the Longmenshan thrust fault and triggered nearly 200 k landslides, mostly along the Yingxiu-Beichuan surface fault rupture, especially on the hanging wall (Xu et al., 2014). In the 2022 Luding earthquake, the detected 5007 coseismic landslides in the VII (and above)-degree areas were affected by both the seismogenic fault and other active faults (Xiao et al., 2023).

For slow-moving landslides, fault activities reduce the strength of rock and soil and may increase bedrock weathering rates and landslide instability. Scheingross et al. (2013) detected 150 previously undiscovered slow-moving landslides in the SAF zone in California and reported that $\sim 75\%$ of them were distributed within 2 km from the active fault. Bontemps et al. (2020) studied a typical slow-moving landslide in Peru and found that large earthquakes would reduce the rigidity of landslide materials, while a combining effect of small seismic activities and seasonal precipitation maintains the multi-annual slow movement of hillslopes.

5.4. Geological units of igneous and mélange

Igneous and mélange geological units are conducive to distinguishing landslide and non-landslide areas. Landslides on mélange zones account for over 53 % of all landslide areas, while those on igneous zones account for less than 7 %. The distribution of landslide susceptibility for each type of geological unit given by the RF model with soil moisture feature set of median AUC is shown in Fig. 14. In samples of igneous geological units, the landslide susceptibility is almost no more than 0.1, indicating greater stability, whereas that of mélange samples is notably higher than other types of geological units. Mélange is

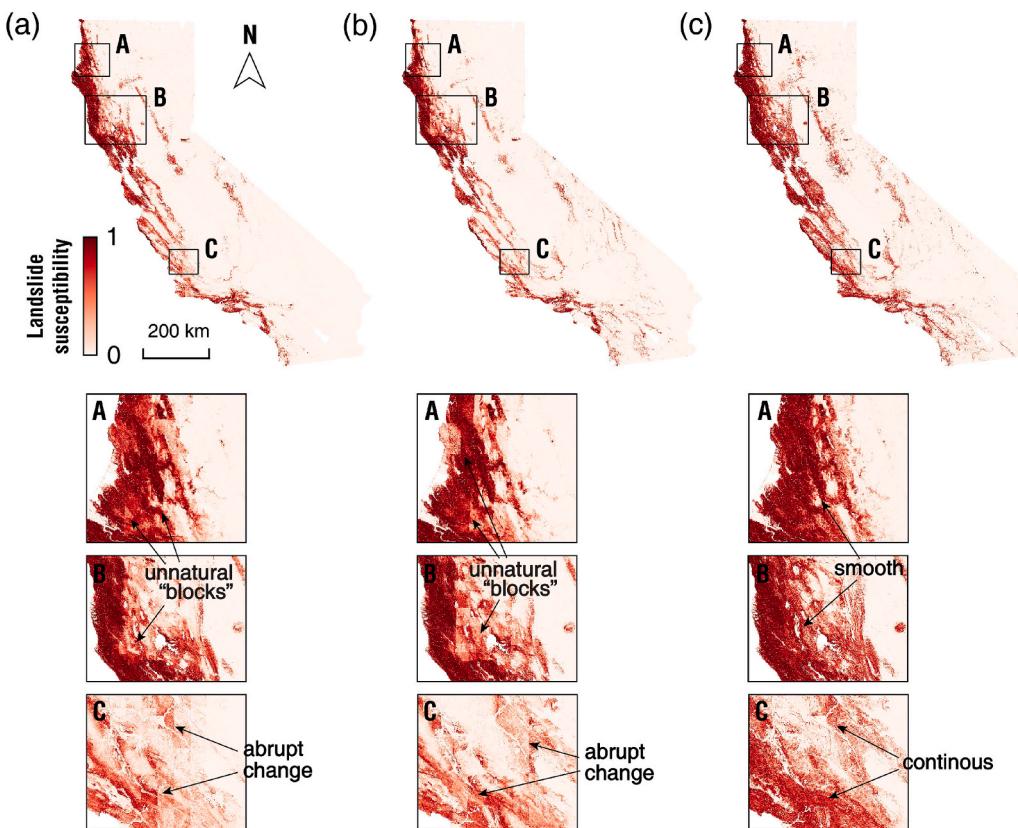


Fig. 8. Landslide susceptibility maps of California predicted by RF models with (a) precipitation, (b) soil moisture feature set, and (c) InSAR coherence feature set which get AUC closest to the median among 100 models for the corresponding feature combination.

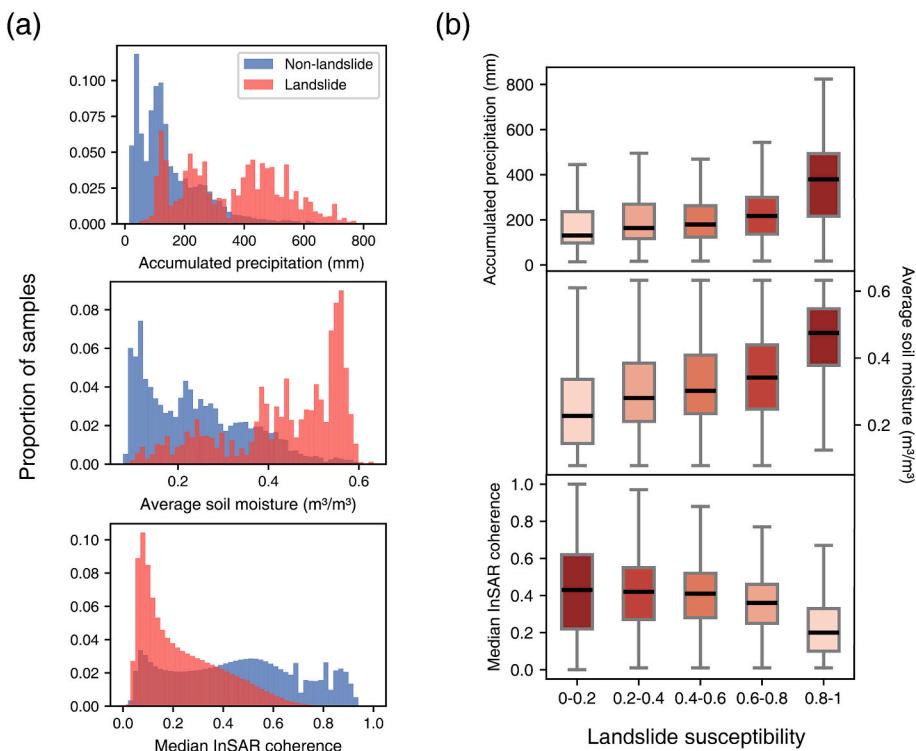


Fig. 9. Distribution of hydrological and InSAR coherence features compared between (a) landslide and non-landslide areas, and (b) different intervals of landslide susceptibility results.

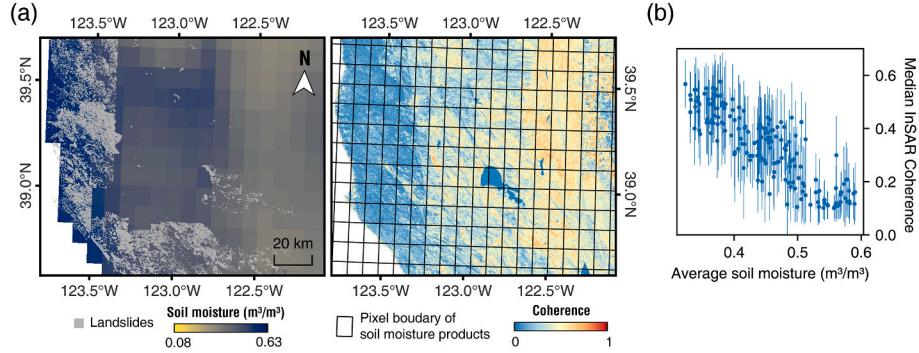


Fig. 10. Soil moisture and InSAR coherence within Zone B in Fig. 8. (a) Soil moisture and InSAR coherence maps, with landslide inventory and pixel boundaries of soil moisture products overlaid. (b) Values of soil moisture and average median InSAR coherence (with standard deviations shown by error bar) within each pixel of soil moisture products.

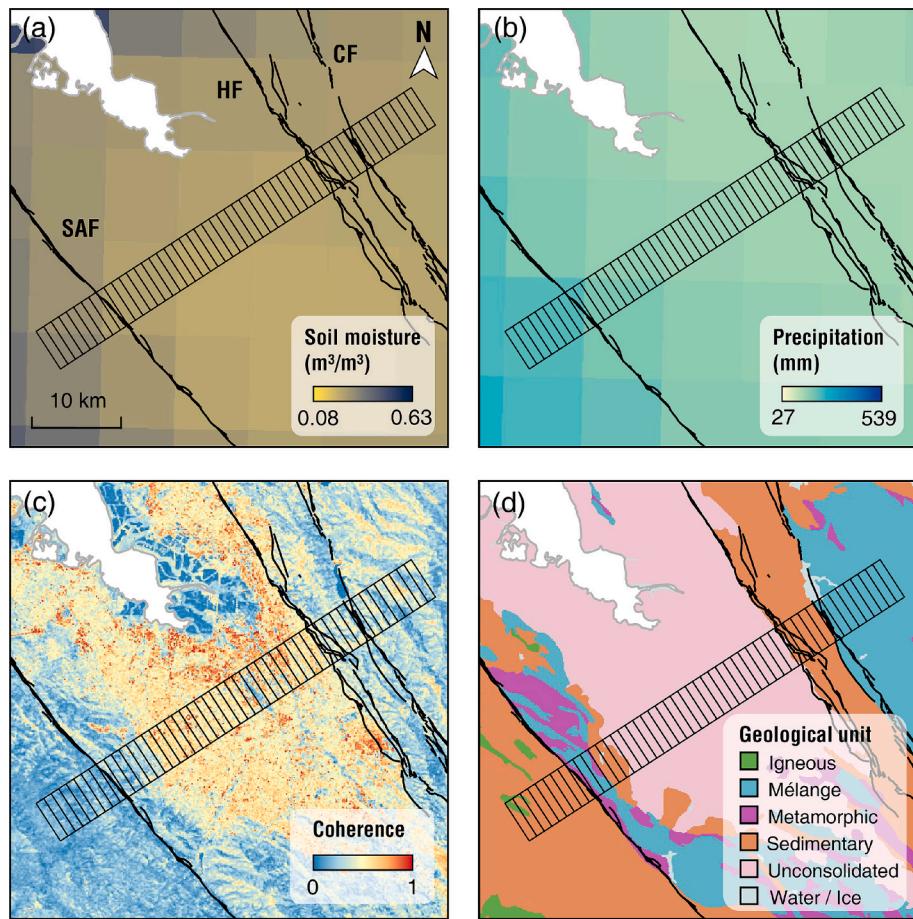


Fig. 11. (a) Soil moisture, (b) precipitation, (c) InSAR coherence, and (d) geological units of the area shown in Fig. 12. Three faults are shown in black lines and labeled with the names in panel (a) (SAF: San Andreas Fault, HF: Hayward Fault, and CF: Calaveras Fault), and the statistical buffer is shown by the grids.

characterized by the lack of internal continuity of contacts or strata and the inclusion of fragments and blocks of all sizes in a fragmented matrix of finer-grained material (Raymond, 1984). In California, mélange is mainly distributed along the northern coast, consisting of various rock types jumbled by shearing along faults during the subduction process.

Landslide mapping relying on geological units has historical precedent (Pachauri and Pant, 1992), and allows for quantitative analyses of the relationship between geological categories and landslide properties. Xu et al. (2021) used InSAR method to detect large slow-moving landslides over the west coast of the States, and noted that the density and extent of landslides in the region containing mélange and relatively

weak metamorphic rocks are twice as large as those nested over the sedimentary and igneous rocks. The homogeneous composition, discontinuity distribution, high clay content, and relatively low shear strength of mélange and metamorphic bedrock are responsible for this spatial coincidence.

5.5. Combination of multiple geo-environmental factors

Since a large amount of the data in the original landslide inventory lacks descriptions of triggers (e.g., precipitation, earthquakes), the susceptibility of landslides with different triggers was not modeled

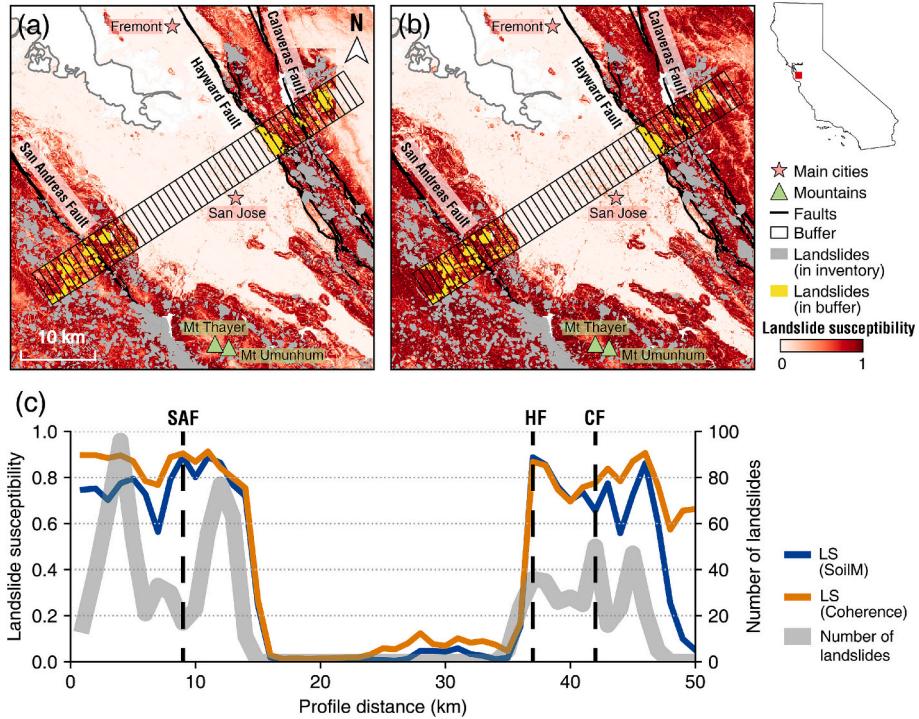


Fig. 12. Landslide susceptibility around San Jose derived from RF model with (a) soil moisture feature set and (b) InSAR coherence feature set of median AUC. (c) Profiles of mean landslide susceptibility and the number of landslides from the inventory within each buffer unit shown in panels (a) and (b). The positions of the three faults are marked using black dashed lines.

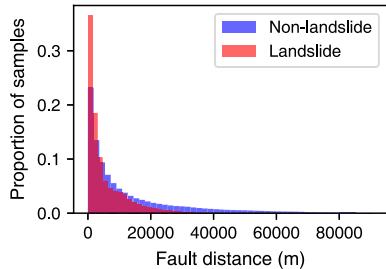


Fig. 13. Distribution of fault distance of both landslide and non-landslide samples.

separately in this study. However, the characteristics of landslides induced by different factors may vary, for example, many rainfall-triggered landslides are shallow and in small dimensions, while earthquake-triggered rapid failures are usually larger (Cui et al., 2020). Archived landslide cases in California with information about landslide areas and triggering factors account for 25,036 out of 132,234 records (~18.9 %), among which the rainfall-triggered landslides are mostly ~370 m² while the earthquake-triggered landslides are ~930 m²

(Fig. 15). Landslides usually develop due to interactions between multiple geo-environmental factors though triggered or accelerated by extreme events (Durand et al., 2018). For earthquake-induced landslides, precipitation may have reduced the mass strength and tectonic activities may have damaged the integrity of rock masses prior to a strong shaking (Nowicki Jesse et al., 2018). Meanwhile, precipitation-induced landslides are also more likely to occur on weakened rock masses (Xu et al., 2020). Therefore, landslide susceptibility assessment should consider all primary and secondary geo-environmental conditions inclusively.

6. Conclusions

In this study, we investigate appropriate high-resolution hydrological proxy in landslide characterization in California. To alleviate the problem that the spatial resolution of current commonly used hydrological products is too coarse to match the size of most landslides, we replaced coarse-resolution hydrological features with high-resolution InSAR products in landslide susceptibility models, leveraging the correlation between radar backscattering and soil moisture. All models achieve recalls over 85 % and AUCs exceeding 90 %, with RF models outperforming LR and SVM models. The satisfactory model performance

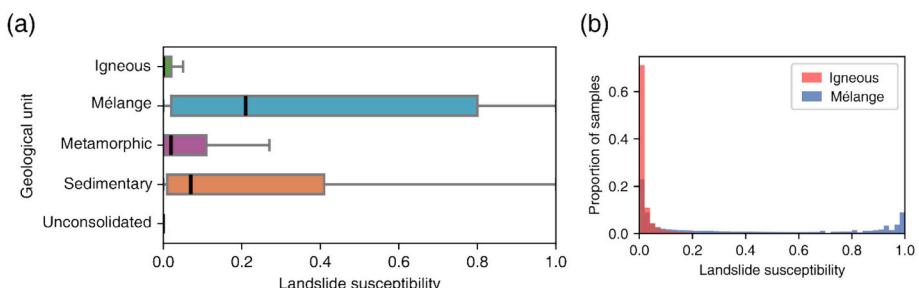


Fig. 14. Distribution of landslide susceptibility predicted by RF model with soil moisture feature set of median AUC in each type of geological unit.

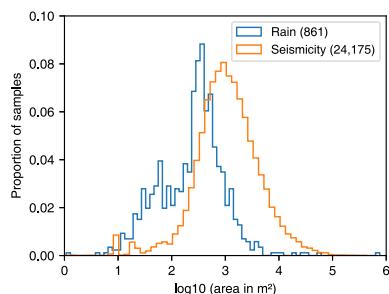


Fig. 15. Histograms of the area of landslides triggered by rainfall and earthquakes, respectively, from the landslide inventory (polygons). The amounts of landslide records of each type are marked in the legend. There are 132,234 landslide records in the polygon landslide inventory in total, but only 25,036 records (accounting for ~18.9 %) are labeled with triggering events.

and much more continuous landslide susceptibility map provided by models using InSAR coherence instead of precipitation and soil moisture suggest that fine-resolution InSAR products can act as alternative hydrological variables and compensate for the inadequate spatial scale of existing hydrological data. Given the landslides developed in active tectonic environments, the number of landslides decreased with an increasing distance from faults. In California, the landslide susceptibility is high in the mélange region and low in the igneous region. California's vast extent, environmental diversity, and high landslide frequency provided a robust validation of our hypothesis. While previous landslide susceptibility assessment studies relied on coarse-resolution hydrological products, we noted the inconsistency problem in spatial scale and addressed it by high-resolution hydrological proxy from InSAR products. The rise of remote sensing big data drives us to incorporate machine learning approaches to inventory landslides on a large scale and to seek alternatives when critical hydroclimate variables are unavailable or insufficient, which can also be considered in other hydrology-driven hazard characterization.

CRediT authorship contribution statement

Yuqi Song: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Xie Hu:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Data curation, Conceptualization. **Xuguo Shi:** Writing – review & editing, Validation, Investigation. **Yifei Cui:** Writing – review & editing, Investigation. **Chao Zhou:** Writing – review & editing, Investigation. **Yueren Xu:** Writing – review & editing, Investigation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Xie Hu reports financial support was provided by National Natural Science Foundation of China. Xie Hu reports a relationship with Peking University that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xie Hu reports financial support was provided by National Natural Science Foundation of China. Xie Hu reports a relationship with Peking University that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (42371078, 42474061, and 42371094), Technology Innovation Center for Geohazard Monitoring and Risk Early Warning, Ministry of Natural Resources (TICGM-2024-02), and the Distinguished Young Scholars of the National Natural Science Foundation of China (Overseas). MATLAB, ArcGIS and Adobe Illustrator software programs were applied to compile the raster and vector products and generate figures.

Data availability

SRTM 1 Arc-Second Global DEM and Landsat 8 OLI images are downloaded from USGS EarthExplorer (<https://earthexplorer.usgs.gov/>). Precipitation data (GPM_3IMERGDF.07) are available from <https://doi.org/10.5067/GPM/IMERGDF/DAY/07>. Soil moisture data (SPL3SMP_E) data are available from <https://doi.org/10.5067/4DQ54OUIJ9DL>. NDVI map is generated using Landsat 8 images at Google Earth Engine platform. River vectors are downloaded from HydroRIVERS (<https://www.hydrosheds.org/products/hydrorivers>). Fault vectors are downloaded from the Quaternary Fault and Fold Database of the United States (<https://www.usgs.gov/programs/earthquake-hazards/faul>). Global seasonal Sentinel-1 interferometric coherence and backscatter data set are available from Kellndorfer et al. (2022) (<https://doi.org/10.1038/s41597-022-01189-6>). Geological unit data are derived from The State Geologic Map Compilation (SGMC) Geodatabase of the Conterminous United States (<https://doi.org/10.5066/F7WH2N65>). Land cover types data are derived from the National Land Cover Database (NLCD) 2019 Products (<https://www.mrlc.gov/data/nlcd-2019-land-cover-conus>). Landslide Inventories across the United States (version 2) are available from <https://doi.org/10.5066/P9FZUX6N>.

References

- Attema, E.P.W., Ulaby, F.T., 1978. Vegetation modeled as a water cloud. *Radio Sci.* 13 (2), 357–364.
- Balenzano, A., Mattia, F., Satalino, G., Lovergne, F.P., Palmisano, D., Peng, J., et al., 2021. Sentinel-1 soil moisture at 1 km resolution: A validation study. *Remote Sens. Environ.* 263, 112554.
- Barman, J., Das, J., 2024. Assessing classification system for landslide susceptibility using frequency ratio, analytical hierarchical process and geospatial technology mapping in Aizawl district, NE India. *Adv. Space Res.* 74 (3), 1197–1224.
- Barrett, B.W., Dwyer, E., Whelan, P., 2009. Soil moisture retrieval from active spaceborne microwave observations: an evaluation of current techniques. *Remote Sens.* 1 (3), 210–242.
- Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., Anderson, C., 2007. Initial soil moisture retrievals from the METOP-A advanced Scatterometer (ASCAT). *Geophys. Res. Lett.* 34 (20).
- Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., et al., 2019. Toward global soil moisture monitoring with Sentinel-1: harnessing assets and overcoming obstacles. *IEEE Trans. Geosci. Remote Sens.* 57 (1), 520–539.
- Belair, G.M., Jones, E.S., Slaughter, S.L., Mirus, B.B., 2022. Landslide Inventories across the United States Version 2 [Dataset]. U.S. Geological Survey.
- Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York, NY.
- Bontemps, N., Lacroix, P., Larose, E., Jara, J., Taipe, E., 2020. Rain and small earthquakes maintain a slow-moving landslide in a persistent critical state. *Nat. Commun.* 11 (1), 780.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Broeckx, J., Vanmaercke, M., Duchateau, R., Poesen, J., 2018. A data-based landslide susceptibility map of Africa. *Earth Sci. Rev.* 185, 102–121.
- Bürgi, P.M., Lohman, R.B., 2021. High-resolution soil moisture evolution in hyper-arid regions: a comparison of InSAR, SAR, microwave, optical, and data assimilation systems in the southern Arabian peninsula. *J. Geophys. Res. Earth* 126 (12), e2021JF006158.
- Chang, Y.W., Lin, C.J., 2008. Feature Ranking Using Linear SVM. In: Paper Presented at Workshop on the Causation and Prediction Challenge at WCCI, Hong Kong, China. In: <http://proceedings.mlr.press/v3/chang08a.html>.
- Chaussard, E., Bürgmann, R., Fattah, H., Nadeau, R.M., Taira, T., Johnson, C.W., Johanson, I., 2015. Potential for larger earthquakes in the East San Francisco Bay Area due to the direct connection between the Hayward and Calaveras faults. *Geophys. Res. Lett.* 42 (8), 2734–2741.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.

- Cui, S., Yang, Q., Pei, X., Huang, R., Guo, B., Zhang, W., 2020. Geological and morphological study of the Daguangbao landslide triggered by the Ms. 8.0 Wenchuan earthquake, China. *Geomorphology* 370, 107394.
- De Zan, F., Parizzi, A., Prats-Iraola, P., López-Dekker, P., 2014. A SAR interferometric model for soil moisture. *IEEE Trans. Geosci. Remote Sens.* 52 (1), 418–425.
- Dewitz, J., U.S. Geological Survey, 2021. National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021) [Dataset]. U.S. Geological Survey.
- Durand, V., Mangeney, A., Haas, F., Jia, X., Bonilla, F., Peltier, A., et al., 2018. On the link between external forcings and slope instabilities in the piton de la Fournaise summit crater, Reunion Island. *J. Geophys. Res. Earth* 123 (10), 2422–2442.
- Fung, A.K., Li, Z., Chen, K.S., 1992. Backscattering from a randomly rough dielectric surface. *IEEE Trans. Geosci. Remote Sens.* 30 (2), 356–369.
- Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. *Geomorphology* 94 (3), 268–289.
- Hallikainen, M.T., Ulaby, F.T., Dobson, M.C., El-rayes, M.A., Wu, L., 1985. Microwave dielectric behavior of wet soil-part 1: empirical models and experimental observations. *IEEE Trans. Geosci. Remote Sens.* GE-23(1), 25–34.
- Handwerger, A.L., Fielding, E.J., Huang, M.-H., Bennett, G.L., Liang, C., Schulz, W.H., 2019. Widespread initiation, reactivation, and acceleration of landslides in the northern California coast ranges due to extreme rainfall. *J. Geophys. Res. Earth* 124 (7), 1782–1797.
- Hapke, C.J., Green, K.R., 2006. Coastal landslide material loss rates associated with severe climatic events. *Geology* 34 (12), 1077.
- Haque, U., da Silva, P.F., Devoli, G., Pilz, J., Zhao, B., Khaloua, A., et al., 2019. The human cost of global warming: deadly landslides and their triggers (1995–2014). *Sci. Total Environ.* 682, 673–684.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition. Springer, New York, NY.
- Horton, J.D., 2017. The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (ver. 1.1, August 2017) [Dataset]. U.S. Geological Survey Data Release.
- Hu, X., Lu, Z., Pierson, T.C., Kramer, R., George, D.L., 2018. Combining InSAR and GPS to determine transient movement and thickness of a seasonally active low-gradient translational landslide. *Geophys. Res. Lett.* 45 (3), 1453–1462.
- Hu, X., Bürgmann, R., Fielding, E.J., Xu, X., Zhen, L., 2021. Machine-learning characterization of tectonic, hydrological and anthropogenic sources of ground deformation in California. *J. Geophys. Res. Solid Earth* 126, e2021JB022373.
- Huang, R., Fan, X., 2013. The landslide story. *Nat. Geosci.* 6 (5), 325–326.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. *An Introduction to Statistical Learning: With Applications in R*, second edition. Springer, New York, NY.
- Jibson, R.W., Harp, E.L., Schulz, W., Keefer, D.K., 2006. Large rock avalanches triggered by the M 7.9 Denali fault, Alaska, earthquake of 3 November 2002. *Eng. Geol.* 83 (1–3), 144–160.
- Jordan, T.E., Lohman, R.B., Tapia, L., Pfeiffer, M., Scott, C.P., Amundson, R., Godfrey, L., Riquelme, R., 2020. Surface materials and landforms as controls on InSAR permanent and transient responses to precipitation events in a hyperarid desert, Chile. *Remote Sens. Environ.* 237, 111544.
- Kellndorfer, J., Cartus, O., Lavalle, M., Magnard, C., Milillo, P., Oveisgharan, S., et al., 2022. Global seasonal Sentinel-1 interferometric coherence and backscatter data set. *Sci. Data* 9 (1), 73.
- Klose, M., 2015. *Landslide Databases as Tools for Integrated Assessment of Landslide Risk*. Springer, Cham, Switzerland.
- Lacroix, P., Deheq, A., Taipe, E., 2020a. Irrigation-triggered landslides in a Peruvian desert caused by modern intensive farming. *Nat. Geosci.* 13 (1), 56–60.
- Lacroix, P., Handwerger, A.L., Bièvre, G., 2020b. Life and death of slow-moving landslides. *Nat. Rev. Earth & Environ.* 1 (8), 404–419.
- Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Process.* 27 (15), 2171–2186.
- Lu, Z., Kim, J., 2021. A framework for studying hydrology-driven landslide hazards in northwestern US using satellite InSAR, precipitation and soil moisture observations: early results and future directions. *GeoHazards* 2 (2), 17–40.
- Mergili, M., Schratz, K., Ostermann, A., Fellin, W., 2012. Physically-based modelling of granular flows with open source GIS. *Nat. Hazards Earth Syst. Sci.* 12 (1), 187–200.
- Mirus, B.B., Jones, E.S., Baum, R.L., Godt, J.W., Slaughter, S., Crawford, M.M., et al., 2020. Landslides across the USA: occurrence, susceptibility, and data limitations. *Landslides* 17 (10), 2271–2285.
- Nowicki Jessee, M.A., Hamburger, M.W., Allstadt, K., Wald, D.J., Robeson, S.M., Tanyas, H., Hearne, M., Thompson, E.M., 2018. A global empirical model for near-real-time assessment of seismically induced landslides. *J. Geophys. Res. Earth* 123 (8), 1835–1859.
- O'Neill, P.E., Chan, S., Njoku, E.G., Jackson, T., Bindlish, R., Chaubell, J., Colliander, A., 2021. SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 Km EASE-Grid Soil Moisture, Version 5 [Dataset]. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA.
- Owe, M., de Jeu, R., Holmes, T., 2008. Multisensor historical climatology of satellite-derived global land surface moisture. *J. Geophys. Res. Earth* 113 (F1).
- Pachauri, A.K., Pant, M., 1992. Landslide hazard mapping based on geological attributes. *Eng. Geol.* 32 (1–2), 81–100.
- Raymond, L.A., 1984. Classification of melanges. In: Raymond, L.A. (Ed.), *Melanges: Their Nature, Origin, and Significance*. Geological Society of America, Boulder, CO, pp. 7–20.
- Reichenbach, P., Galli, M., Cardinali, M., Guzzetti, F., Ardizzone, F., 2005. Geomorphologic mapping to assess landslide risk: Concepts, methods and applications in the Umbria region of Central Italy. In: Glade, T., Anderson, M.G., Crozier, M.J. (Eds.), *Landslide Risk Assessment*. John Wiley, London, UK, pp. 429–468.
- Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth Sci. Rev.* 180, 60–91.
- Scheingross, J.S., Minchew, B.M., Mackey, B.H., Simons, M., Lamb, M.P., Hensley, S., 2013. Fault-zone controls on the spatial distribution of slow-moving landslides. *GSA Bull.* 125 (3–4), 473–489.
- Southern California Earthquake Data Center, 2013. San Andreas Fault Zone. <http://scsed.caltech.edu/earthquake/sanandreas.html>.
- Tang, G., Fang, Z., Wang, Y., 2023. Global landslide susceptibility prediction based on the automated machine learning (AutoML) framework. *Geocarto Int.* 38 (1), 2236576.
- Ulaby, F.T., Batlivala, P.P., 1976. Optimum radar parameters for mapping soil moisture. *IEEE Trans. Geosci. Electron.* 14 (2), 81–93.
- Ullmann, T., Jagdhuber, T., Hoffmeister, D., May, S.M., Baumhauer, R., Bubenzier, O., 2023. Exploring Sentinel-1 backscatter time series over the Atacama Desert (Chile) for seasonal dynamics of surface soil moisture. *Remote Sens. Environ.* 285, 113413.
- Wang, H., Zhang, L., Luo, H., He, J., Cheung, R.W.M., 2021. AI-powered landslide susceptibility assessment in Hong Kong. *Eng. Geol.* 288, 106103.
- Wen, Z., Shi, J., Li, Q., He, B., Chen, J., 2018. ThunderSVM: a fast SVM library on GPUs and CPUs. *J. Mac. Lear. Res.* 19 (1), 797–801.
- Woodard, J.B., Mirus, B.B., Crawford, M.M., Or, D., Leshchinsky, B.A., Allstadt, K.E., Wood, N.J., 2023. Mapping landslide susceptibility over large regions with limited data. *J. Geophys. Res. Earth* 128 (5), e2022JF006810.
- Xiao, Z., Xu, C., Huang, Y., He, X., Shao, X., Chen, Z., Xie, C., Li, T., Xu, X., 2023. Analysis of spatial distribution of landslides triggered by the Ms 6.8 Luding earthquake in China on September 5, 2022. *Geoenviron. Disasters* 10 (1), 3.
- Xu, C., Xu, X., Yao, X., Dai, F., 2014. Three (nearly) complete inventories of landslides triggered by the May 12, 2008 Wenchuan mw 7.9 earthquake of China and their spatial distribution statistical analysis. *Landslides* 11 (3), 441–461.
- Xu, Y., Lu, Z., Schulz, W.H., Kim, J., 2020. Twelve-year dynamics and rainfall thresholds for alternating creep and rapid movement of the Hooskanaden landslide from integrating InSAR, pixel offset tracking, and borehole and hydrological measurements. *J. Geophys. Res. Earth* 125 (10), e2020JF005640.
- Xu, Y., Schulz, W.H., Lu, Z., Kim, J., Baxstrom, K., 2021. Geologic controls of slow-moving landslides near the US west coast. *Landslides* 18 (10), 3353–3365.
- Yi, Y., Zhang, W., Xu, X., Zhang, Z., Wu, X., 2022. Evaluation of neural network models for landslide susceptibility assessment. In: *J. Digi. Earth* 15 (1), 934–953.
- Youssef, K., Shao, K., Moon, S., Bouchard, L.-S., 2023. Landslide susceptibility modeling by interpretable neural network. *Communicat. Earth & Environ.* 4 (1), 162.
- Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., Pourghasemi, H.R., 2018. Landslide susceptibility modeling applying machine learning methods: a case study from Longji in the three gorges reservoir area, China. *Comput. Geosci.* 112, 23–37.
- Zhou, C., Cao, Y., Hu, X., Yin, K., Wang, Y., Catani, F., 2022. Enhanced dynamic landslide hazard mapping using MT-InSAR method in the three gorges reservoir area. *Landslides* 19 (7), 1585–1597.