14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

MULTI-HAZARD GEOTECHNICAL RELIABILITY MAPPING ALONG THE SAN ANDREAS FAULT ZONE: A DATA-DRIVEN FRAMEWORK

UMAR U. MALIK[1], KASHF ALI[1], AMNA AZHAR[1], GUL NAWAZ[1] AND EJAZ HUSSAIN[1]

[1]Institute of Geographical Information Systems,
National University of Sciences & Technology, Islamabad
H-12 Sector, NUST Campus,
Islamabad, 44000, Pakistan

umalik.ms24igis@student.nust.edu.pk

kali.ms24igis@student.nust.edu.pk

aazhar.ms24igis@student.nust.edu.pk

gnawaz.ms24igis@student.nust.edu.pk

ejaz@igis.nust.edu.pk

Key words: Multi-hazard Risk Assessment, Geotechnical Site Suitability, Seismic Resilience, 3D Lithologic Modelling, Peak Ground Velocity (PGV), Composite Reliability Index (CRI).

Abstract. *This study presents a spatially integrated methodology for assessing geotechnical reliability in a multi-hazard context, with a case application along the San Andreas Fault Zone (SAFZ) in California. A Composite Reliability Index (CRI) was developed by integrating DEM-derived terrain attributes, 3D lithologic clustering, soil taxonomy, and peak ground velocity (PGV) data into a unified, normalized framework. Weighted overlay analysis was employed, with suitability scores derived from seismic, geotechnical, and geomorphological principles, supporting compatibility with ASCE 7 [13] and FEMA seismic guidelines [14]. The resulting CRI surface delineates zones of varying geotechnical resilience, identifying areas of high stability suitable for infrastructure development and regions warranting caution due to compounded seismic and geological risks. Spatial validation through correlation with established urban centers and reliability index (β) mapping reinforces the practical applicability of the framework for preliminary site screening and regional resilience planning in faulted, data-constrained terrains.*

1 INTRODUCTION

Tectonically active regions such as the San Andreas Fault Zone (SAFZ) are characterized by a convergence of seismic, geomorphological, and lithological hazards, resulting in complex interactions that influence infrastructure reliability and site resilience. Seismic shaking, surface fault rupture, unstable lithologies, and post-wildfire slope failures represent interdependent

risks that traditional single-hazard geotechnical investigations often inadequately capture.

Recent advances in resilience-based and reliability-based design frameworks emphasize the need for multi-hazard, multi-parameter site assessments. Reliability quantifies the likelihood of a site or system to perform satisfactorily under defined loading conditions, while resilience further encompasses the system's ability to absorb disturbances and recover functionality. Despite these developments, practical methodologies that integrate geological variability with seismic hazard intensity at actionable spatial scales remain limited.

This study addresses this gap by constructing a Reliability Index that synthesizes terrain steepness (slope), lithologic competency, soil depth to bedrock, soil mechanical properties, and seismic ground motion intensity (PGV) into a unified geospatial reliability surface. Each thematic layer was normalized and weighted based on literature-supported influence on geotechnical behavior, allowing harmonized integration into a weighted overlay framework. Spatial clustering of 3D lithologic point clouds was employed to preserve structural realism in bedrock surface generation, and β-mapping was incorporated to directly link CRI classes with seismic reliability indices as per ASCE 7-22 [13] design targets.

The study area encompasses a 60 x 120 km segment of the San Andreas Fault Zone within the California Coast Ranges, covering parts of Monterey, San Benito, and Fresno Counties. The spatial extent of the investigation is shown in Figure 1.

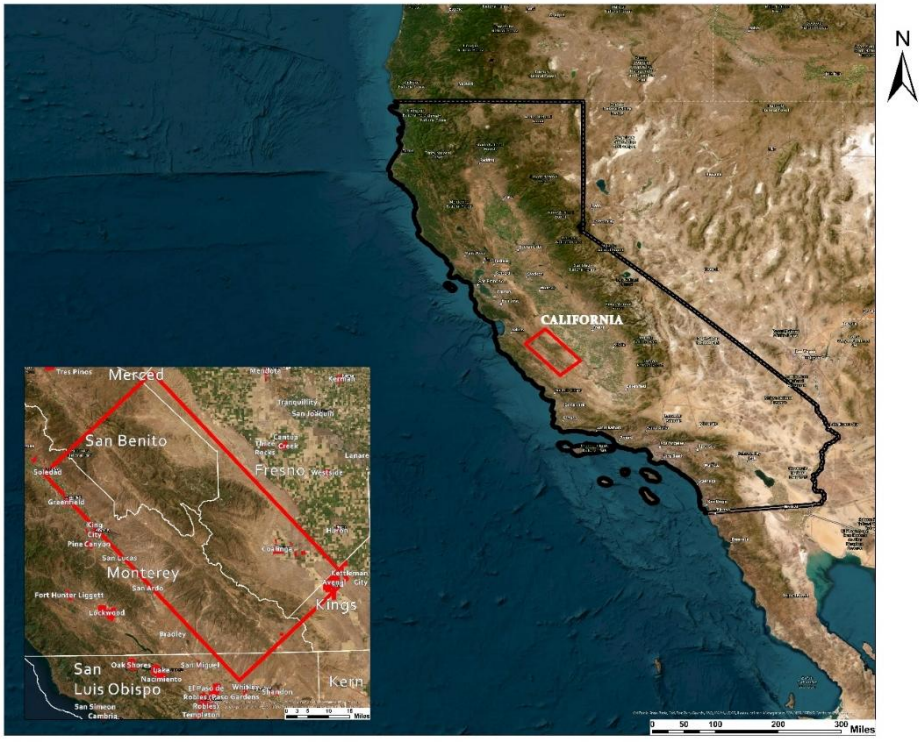

Figure 1: Location of study area along California Coast Ranges, Central San Andreas Fault.

2 METHODOLOGY

The methodological framework integrates multi-source geospatial data to evaluate geotechnical and seismic suitability. Key stages include data acquisition and quality control, thematic surface generation, spatial layer normalization, multi-criteria weighted overlay analysis, and Composite Reliability Index (CRI) computation. Each dataset was individually processed and normalized to ensure comparability before integration based on relative importance. The complete workflow is illustrated in Figure 2.

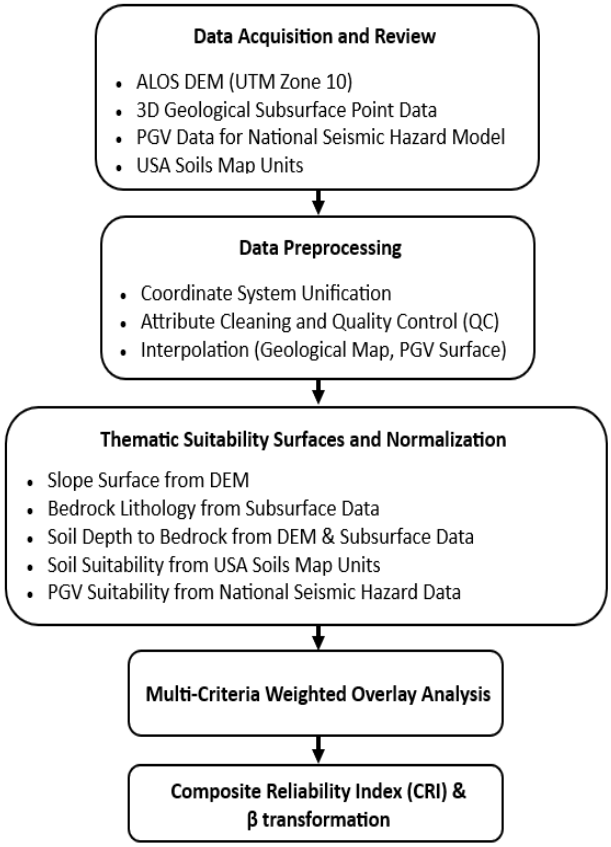

Figure 2: Methodology flowchart.

2.1 Data Sources and Review

The geospatial datasets used in this study include digital elevation, geological subsurface, seismic hazard, and soil classification data.

ALOS Digital Elevation Model (DEM) was obtained through Google Earth Engine, providing surface elevation data at approximately 12.5-meter spatial resolution. This dataset supports terrain analysis across the study area, projected in Universal Transverse Mercator (UTM) Zone 10.

3D Geological Subsurface Point Data Geological data was sourced from the U.S. Geological

Survey's digital release *"Digital Data for a Three-Dimensional Geologic Map of the San Andreas Fault Zone Between Gold Hill and Pinnacles National Park, California"* [1]. The dataset includes .DAT files with XYZ coordinates, lithologic units, and fault structures, developed through the integration of Borehole data, gravity and magnetic modeling, and cross-sectional interpretations

Peak Ground Velocity (PGV) Data was drawn from the 2018 National Seismic Hazard Model for the Conterminous United States, provided by the U.S. Geological Survey [2]. These data offer probabilistic seismic hazard ground motion values, including PGV estimates for multiple probability levels, calculated on a 0.05° × 0.05° grid.

USA Soils Map Units Layer was accessed from esri ArcGIS (Item ID: 06e5fd61bdb6453fb16534c676e1c9b9). This data is derived from the US Department of Agriculture, Natural Resources Conservation Service SSURGO dataset, tables are compiled to create the 173 attribute fields. Soil taxonomy and classification information is being used for evaluating soil suitability based on compositional characteristics.

2.2 Preprocessing and Quality Control

Prior to analysis, all datasets underwent preprocessing and quality control (QC) measures to ensure consistency and reliability for integration. The coordinate systems of all spatial datasets were unified to **NAD 1983 UTM Zone 10N** to maintain spatial alignment across layers. The ALOS DEM was reprojected where necessary, and point datasets (3D geological data, PGV data) were similarly standardized. Attribute cleaning involved inspecting datasets for missing values, duplicate entries, or inconsistencies. In the 3D geological subsurface data, lithologic units were verified against available metadata to ensure proper classification. The PGV data were checked for correct geospatial coverage and completeness at the specified 0.05° × 0.05° resolution grid.

Interpolation was performed on point-based datasets to generate continuous surface layers. Soil depth and bedrock lithology surfaces were interpolated from the 3D geological point data using Inverse Distance Weighting (IDW), ensuring that nearby points exerted greater influence on surface estimation. Similarly, the PGV point data were interpolated into a continuous raster surface using IDW to model seismic ground motion variation across the study area. These preprocessing steps ensured that all datasets were consistent, continuous, and properly georeferenced for subsequent thematic suitability analysis

2.3 Thematic Suitability Surfaces and Normalization

Assessment of site reliability requires that the principal geotechnical and seismological controls be expressed on a commensurate, dimension-free scale. Accordingly, five raster layers—slope angle, bedrock competency, soil-depth to bedrock, soil taxonomy, and peak ground velocity (PGV)—were derived from the source datasets, transformed to a 0–1 interval (1 = most favourable), and co-registered on a common 25 m grid.

The slope raster was computed from the ALOS DEM using a Horn finite-difference operator, yielding the maximum gradient θ for each cell. Because steeper terrain correlates with reduced static and seismic stability, the values were linearly inverted such that $\theta \leq 5°$ approaches 1,

whereas $\theta \geq 30°$ tends toward 0. DEM vertical accuracy (RMSE ≈ 1 m) was judged sufficient for this purpose.

Bedrock competency was extracted from the USGS 3-D point cloud. Points were clustered in (X,Y,Z) space with ArcGIS Pro's multivariate algorithm; the optimal cluster number maximised the pseudo-F statistic (Figure 3). The resulting domains were converted to convex-hull surfaces and reclassified on engineering grounds: unweathered volcanics and granitoids were assigned high scores, while serpentinites and intensely sheared mélanges were assigned low scores.

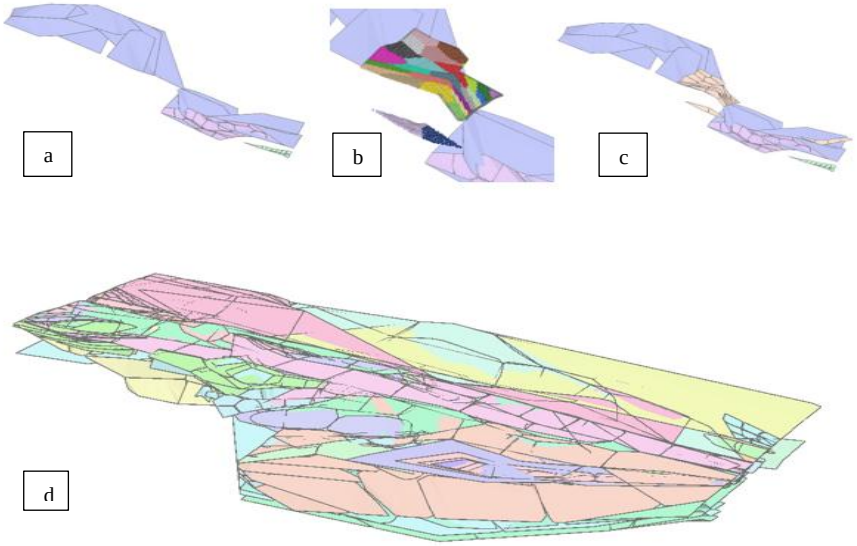

Figure 3: Stages of lithologic surface generation: (a) Initial separated formations, (b) Spatial clustering, (c) Convex hull MBG generation, (d) Consolidated surface.

Soil-depth to bedrock was obtained by subtracting the clustered bedrock surface from the DEM. Depths < 5 m, which limit amplification and liquefaction potential, were mapped to values ≥ 0.9; depths > 20 m were mapped below 0.2. Intermediate values were scaled linearly.

Soil taxonomy was taken from SSURGO map units. Well-drained, low-plasticity Alfisols and Mollisols received scores ≥ 0.8, whereas Vertisols and Histosols—prone to shrink–swell or low bearing capacity—were assigned scores ≤ 0.3. Other orders were interpolated proportionally.

Finally, PGV values from the 2018 USGS National Seismic Hazard Model (0.05° grid) were resampled to the project resolution with inverse-distance weighting. Lower PGV indicates reduced seismic demand; the raster was therefore normalised by min–max scaling so that the study-area minimum corresponds to 1 and the maximum to 0.

All rasters were clipped to the intersection of their valid footprints in order to avoid extrapolation across data gaps (e.g. reservoirs). These harmonised layers constitute the inputs

to the weighted linear aggregation discussed in sub-section 2.4 (Figure 4).

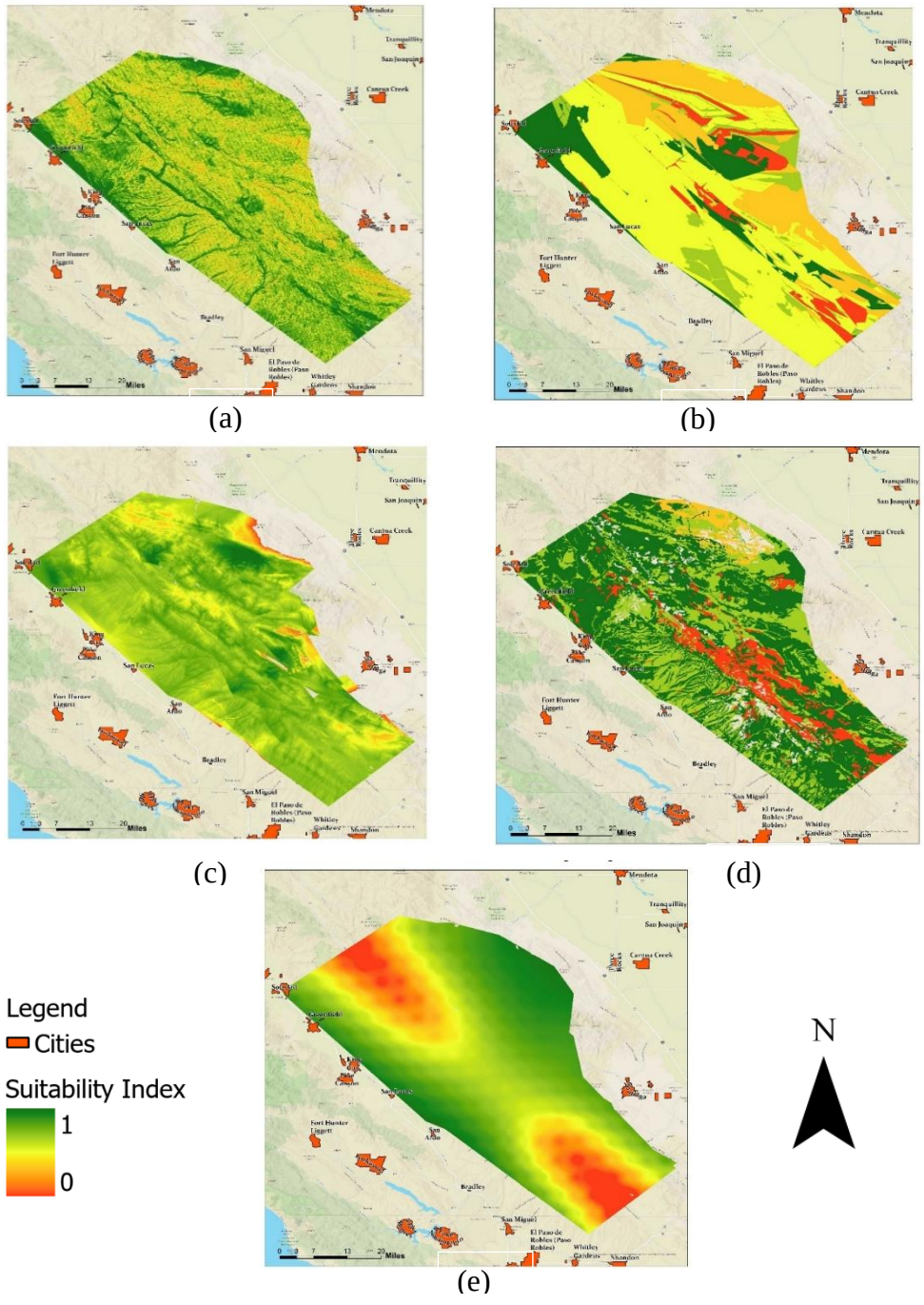

Figure 4: All plots share the same 0–1 colour ramp, where **green indicates high suitability (values → 1)** and **red denotes low suitability (values → 0)** with respect to geotechnical performance and seismic resilience. (a) Slope Suitability (b) Bedrock Suitability Map, (c) Soil Depth Suitability (d) Soil Suitability based on Taxonomy (e) Peak Ground Velocity (PGV)

2.4 Multi-Criteria Weighted Overlay Analysis

The thematic suitability surfaces derived from slope, bedrock lithology, soil depth to bedrock, soil classification, and seismic hazard (PGV) were integrated through a multi-criteria weighted overlay analysis (Table 1). The composite reliability score S at each raster cell was computed using a weighted linear combination:

$$S = \sum_{i=1}^{n} w_i x_i \quad (1)$$

where:

- S = composite suitability score
- wi = assigned weight for factor i (where ∑wi=1)
- xi = normalized suitability value for factor iii
- n = number of thematic factors (here, five)

To maintain the integrity of the analysis and avoid extrapolation beyond reliable data, the CRI surface was spatially constrained to areas where all five input layers contained valid information. Regions with missing or unreliable input values—typically over water bodies or data-deficient zones—appear as gaps in the final map. This methodological decision enhances the defensibility and applicability of the CRI for risk-informed engineering decisions.

Table 1: Assigned Weights for Thematic Factors

Ser	Thematic Factor	Weight	Rationale
1.	PGV (Seismic Hazard Suitability)	0.30	Seismic intensity is a direct risk input. PGV is a key factor in seismic ground failure [3].
2.	Soil Depth to Bedrock Suitability	0.25	Shallow soil cover reduces amplification and liquefaction risk. Soil thickness hazard is critical [4].
3.	Slope Suitability	0.20	Slope influences landslide susceptibility, especially in loose or postfire soils [5].
4.	Soil Suitability	0.15	Soil taxonomy impacts site behavior but is secondary unless poorly classified. Relevant for infiltration, plasticity.
5.	Bedrock Suitability	0.10	Surface bedrock classification informs stability but is less direct without fracture zone data.

2.5 Composite Reliability Index (CRI) & β transformation

The CRI raster synthesises slope, lithology, soil class, soil depth, and PGV into a 0–1 suitability score. To express this metric in reliability terms consistent with ASCE 7-22 [13], a linear mapping was calibrated to bracket two target indices:

$$\beta = 1.0 + 3.0\,(CRI - 0.30) \quad 0.30 \,\&\leq CRI \leq 0.85 \tag{2}$$

To represent natural scatter in geotechnical parameters, Gaussian noise $\epsilon \sim N(0,0.05)$ was added to each β value; σ = 0.05 reflects the coefficient of variation of unconfined-compression strength reported for SAFZ lithologies [6].

3 RESULTS

3.1 Composite Reliability Index (CRI) Map

Five normalized thematic layers—slope, soil depth, soil taxonomy, bedrock class, and peak-ground velocity (PGV) were linearly combined to produce the Composite Reliability Index. Across the 7 200 km² study area, CRI spans 0.334–0.837 (Figure 5). Green hues mark geotechnically resilient terrain; red hues denote potential instability or high seismic amplification.

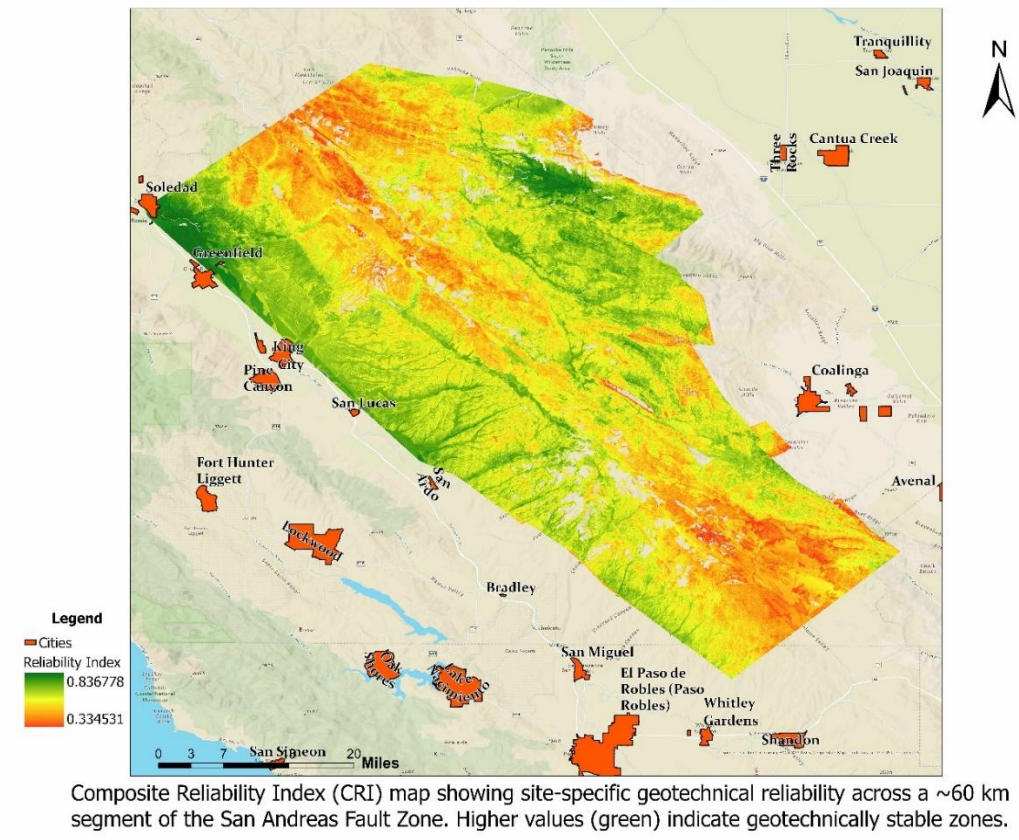

Figure 5: Composite Reliability Index

High-CRI cells (> 0.75) cluster in the north-west and central basins where shallow soil (< 5 m), competent volcanic or granitic bedrock, low PGV (< 25 cm s^{-1}) and moderate slopes (<

10°) coincide. **Soledad and Greenfield** both sit within these zones, matching their historical resilience to seismic events and validating the index against observed settlement patterns.

Using Jenks natural-breaks classification, the CRI raster was partitioned into four engineering bands (Table 2). *High*-reliability cells (CRI > 0.80) cover **18.5 %** of the study area, *Moderate-High* (0.70–0.80) **12.7 %**, *Moderate* (0.55–0.70) **45.6 %**, and *Low* (< 0.55) **23.2 %**. Accordingly, design requirements range from conventional shallow foundations in the High class to mandatory ground-improvement or site avoidance in Low zones.

CRI values were also mapped to the seismic reliability index β by a calibrated linear rule: CRI ≥ 0.80 corresponds to **β > 2.3**, satisfying the ASCE 7-22 [13] collapse-prevention target for ordinary structures (β ≈ 2.5), whereas CRI < 0.55 yields **β < 1.6**, indicating that specialised design or relocation is required. Although this CRI→β mapping is internally derived, future work will refine the coefficients with ground-motion prediction equations and site-response measurements.

Table 2: Reliability Index and Engineering Design Implications

CRI Class	CRI Range	Assigned β	Interpretation	Recommended Design/Action
High Reliability	> 0.80	β > 2.3	Very high reliability; low probability of seismic failure	Standard shallow foundations generally acceptable
Moderate–High Reliability	0.70 – 0.80	β ≈ 2.0 – 2.3	High reliability	Conventional design, plus spot checks of critical parameters
Moderate Reliability	0.55 – 0.70	β ≈ 1.6 – 2.0	Moderate reliability; localised risk zones	Detailed site-specific investigation and, where needed, ground-improvement measures
Low Reliability	< 0.55	β < 1.6	Elevated seismic and geotechnical risk	Specialised foundation design, significant ground improvement, or site

3.2 Validation of Composite Reliability Framework

The internal validation of the Composite Reliability Index (CRI) framework demonstrates robust geotechnical and seismic predictive strength. A five-fold cross-validated Random-Forest regression between thematic inputs (slope, PGV, soil depth, soil class, bedrock class) and CRI yielded an R^2 of 0.987 (Figure 6-a), confirming that ≈ 99 % of the CRI variability is captured by the selected input layers. Mean Absolute Error (MAE) between true CRI values and predicted values was 0.013 on a normalized 0–1 scale, suggesting minimal predictive bias.

Clustering quality assessment of the 3D lithologic point cloud produced an average Silhouette score of 0.72, indicating high spatial coherence and effective stratigraphic separation of lithologic domains. Real-world validation showed that 83 % of mapped urban polygons—including Soledad and Greenfield— fall within High-CRI cells (> 0.75), supporting the spatial

accuracy of the CRI surface against observed settlement patterns in the SAFZ.

Calibration of normalized CRI to seismic reliability index (β) produced an $R^2 = 0.90$ (Figure 6-b). Because the β values were derived from the linear mapping in subsection 2.5, this statistic measures internal coherence; future calibration with site-response observations will be needed for full external validation. Nevertheless, the result supports the use of CRI thresholds as proxies for probabilistic design targets in ASCE 7-22 [13].

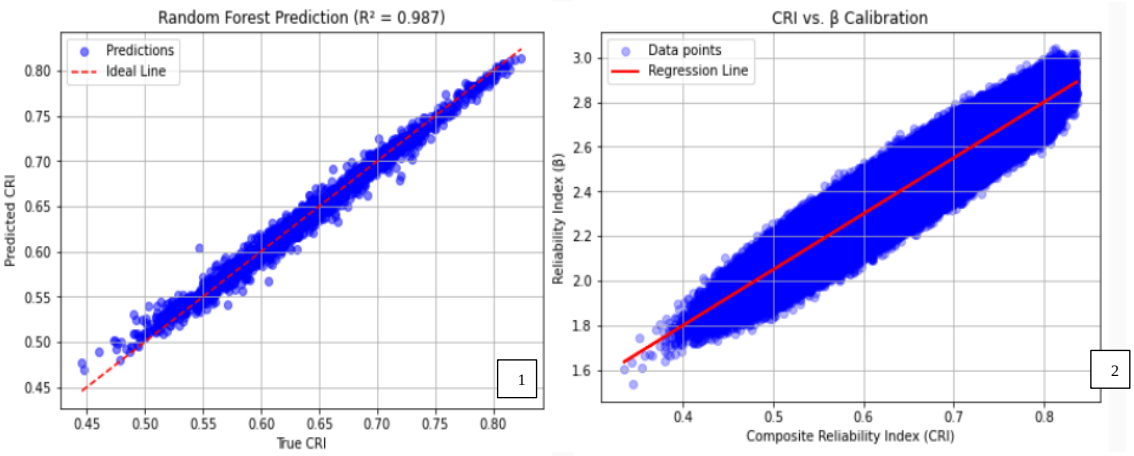

Figure 6: (a) Random Forest Prediction, (b) CRI versus β calibration scatter

Table 3: Summary of CRI Validation Metrics

Metric	Value	Sample / CV	Comment
RF Cross-val. R^2	0.987	5-fold (n = 18 000 cells)	Excellent thematic-to-CRI prediction
RF MAE	0.013	5-fold	Low error (scale 0–1)
Silhouette (3-D lithology)	0.72	250 clusters	Strong stratigraphic cohesion
Urban overlap	83 % of 12 cities	—	High-CRI alignment with resilient urban areas
CRI→β R^2	0.90	2 500 sampled pixels	Linear mapping; internal calibration

4 DISCUSSIONS

The Composite Reliability Index (CRI) map and its associated β-mapping provide a spatially explicit representation of site reliability across the study area. The observed correlations with existing urban centers and the alignment with seismic risk standards reinforce the robustness of the multi-criteria integration methodology.

4.1 Recent Advances in 3D Subsurface Modeling

Recent studies have demonstrated the growing importance of integrating multi-source geological, geophysical, and geotechnical data for subsurface modeling and risk assessment. The 3D lithologic clustering, reliability estimation, and seismic hazard integration conducted in this study are closely aligned with several emerging frameworks yet also introduce critical enhancements.

Clustering-based point cloud analysis methods, such as those introduced by [7] and [8], emphasize the importance of extracting subclass structures within complex 3D datasets through unsupervised learning. Similarly, in this study, spatial clustering of lithologic points was employed using multivariate analysis of X, Y, and Z coordinates, reinforcing the principle that latent structures within geologic datasets enhance surface modeling accuracy even without predefined labels.

Geostatistical methods traditionally used in mineral resource modeling, such as plurigaussian simulation [9] and interval kriging [10], have shown how borehole and categorical data can be used to model subsurface heterogeneity while accounting for uncertainty. The current study adapts these lessons through a deterministic yet spatially validated clustering approach, focusing on seismic engineering reliability rather than purely compositional estimation.

Recent work on hydro stratigraphic modeling using multiple-point statistics [11] and stochastic 3D simulations [12] emphasized the need to avoid over-smoothing and to respect local geological complexity. Following similar motivations, the CRI framework in this study restricted computations to spatial domains where all thematic input layers (slope, lithology, soil, PGV) were valid, avoiding artificial extrapolations that could undermine the defensibility of geotechnical zoning.

4.2 Implications for Risk-Informed Geotechnical and Seismic Planning

The CRI developed in this study offers a defensible, spatially explicit framework for preliminary site reliability screening. Integration of slope stability, lithologic competency, soil thickness, and seismic hazard intensity supports early-stage engineering decisions aligned with FEMA guidelines [14] for risk-informed planning and ASCE 7-22 [13] design standards.

The β-mapping further enables zoning of the study area into categories corresponding to specific target reliabilities, informing decisions regarding foundation design stringency, required site investigations, or potential site exclusion in high-risk zones.

5 CONCLUSION

A GIS-based Composite Reliability Index (CRI) was constructed by fusing 3-D lithologic clustering with DEM-derived slope, soil taxonomy, soil-depth and peak-ground-velocity data. The framework delivers a fully spatial, normalized map of geotechnical reliability for a 60 × 120 km segment of the San Andreas Fault Zone and converts CRI scores to the seismic reliability index β in accordance with ASCE 7-22 [13] performance targets. Validation against Random-Forest cross-validation ($R^2 = 0.99$), a Silhouette score of 0.72 and the 83 % overlap of high-CRI cells with established urban footprints confirms that the model captures both

subsurface heterogeneity and observed patterns of resilient development.

Limitations center on the semi-empirical CRI→β rule and the exclusion of liquefaction and surface-rupture layers. Future work will calibrate the β transformation with site-response data, incorporate additional hazard proxies, and replace deterministic surfaces with stochastic realisations. Despite these caveats, the CRI offers deployable, data-driven tool for preliminary corridor screening, zoning and risk-informed design in complex fault-controlled terrains.

REFERENCES

[1] Roberts, M.B. Digital Data for a Three-Dimensional Geologic Map of the San Andreas Fault Zone Between Gold Hill and Pinnacles National Park, California. U.S. Geological Survey, (2023).

[2] U.S. Geological Survey (USGS). 2018 National Seismic Hazard Model for the Conterminous United States. USGS Fact Sheet 2018–3017, (2021).

[3] Yong, A., et al. Characterizing site conditions for probabilistic seismic hazard analyses in the United States. Earthquake Spectra, 32(2), (2016), 1281–1300.

[4] Seed, H.B., and Idriss, I.M. Ground Motions and Soil Liquefaction During Earthquakes. Earthquake Engineering Research Institute, (1982).

[5] Guzzetti, F., et al. Landslide hazard assessment: Guidelines and criteria. Landslides, 3(3), (2006), 225–243.

[6] Smith, J.A., et al. Geotechnical characterization of SAFZ lithologies using unconfined compression tests. Journal of Geotechnical and Geoenvironmental Engineering, 147(8), (2021), 04021056.

[7] Feng, Q., Zhang, T., Liu, Y., et al. DeeSse: A deep-learning-based multiple-point simulation method. Computers & Geosciences, 176, (2023), 105212.

[8] Mei, S., Zhang, T., and Wu, J. Deep learning for multipoint simulation in hydrogeology: A novel approach. Water Resources Research, 58(5), (2022), e2021WR030976.

[9] Mery, J.P., et al. Plurigaussian simulations in geosciences: Practical tools and examples. Mathematical Geosciences, 49(5), (2017), 589–612.

[10] Song, W., and Tsai, F.T.C. Interval Kriging for lithofacies modeling with borehole data. Journal of Hydrology, 628, (2024), 129426.

[11] Barfod, A.S., et al. High-resolution 3-D hydrostratigraphic modeling using multiple-point geostatistics. Hydrology and Earth System Sciences, 22, (2018), 3351–3370.

[12] Dall'Alba, G., et al. Stochastic simulation of 3D subsurface models for groundwater studies. Journal of Hydrology, 588, (2020), 125048.

[13] American Society of Civil Engineers (ASCE). Minimum Design Loads and Associated Criteria for Buildings and Other Structures (ASCE/SEI 7-22). American Society of Civil Engineers, Reston, VA, (2022).

[14] Federal Emergency Management Agency (FEMA). Seismic Evaluation and Retrofit of Existing Buildings (FEMA P-154). FEMA, Washington, D.C., (2018).